



4-23-2013

Neolithic Mitochondrial Haplogroup H Genomes and the Genetic Origins of Europeans

Paul Brotherton

Wolfgang Haak

Jennifer E. L Templeton

Guido Brandt

Julien Soubrier

See next page for additional authors

Follow this and additional works at: http://repository.upenn.edu/anthro_papers

 Part of the [Anthropology Commons](#), and the [Genetics and Genomics Commons](#)

Recommended Citation

Brotherton, P., Haak, W., Templeton, J. E., Brandt, G., Soubrier, J., Adler, C. J., Richards, S. M., Sarkissian, C., Ganslmeier, R., Friederich, S., Dresely, V., van Oven, M., Kenyon, R., Van der Hoek, M. B., Korlach, J., Luong, K., Ho, S. Y., Quintana-Murci, L., Behar, D. M., Meller, H., Alt, K. W., Cooper, A., Adhikarla, S., Ganesh Prasad, A., Pitchappan, R., Varatharajan Santhakumari, A., Balanovska, E., Balanovska, O., Bertranpetit, J., Comas, D., Martinez-Cruz, B., Melé, M., Clarke, A. C., Matisoo-Smith, E., Dulik, M. C., Gaieski, J., Owings, A. C., Schurr, T. G., Vilar, M. G., Hobbs, J., Hoobs, A., Soodiyall, H., Javed, A., Parida, L., Platt, D. E., Royyuru, A. K., Jin, L., Li, S., Kaplan, M. E., Merchant, N. C., Mitchell, R., Renfrew, C., Lacerda, D. R., Santos, F. R., Soria-Hernanz, D. F., Wells, R., Swamikrishnan, P., Tyler-Smith, C., Vieira, P. R., Ziegler, J., & Genographic Consortium (2013). Neolithic Mitochondrial Haplogroup H Genomes and the Genetic Origins of Europeans. *Nature Communications*, 4 1764. <https://doi.org/10.1038/ncomms2656>

Neolithic Mitochondrial Haplogroup H Genomes and the Genetic Origins of Europeans

Abstract

Haplogroup H dominates present-day Western European mitochondrial DNA variability (>40%), yet was less common (~19%) among Early Neolithic farmers (~5450 BC) and virtually absent in Mesolithic hunter-gatherers. Here we investigate this major component of the maternal population history of modern Europeans and sequence 39 complete haplogroup H mitochondrial genomes from ancient human remains. We then compare this 'real-time' genetic data with cultural changes taking place between the Early Neolithic (~5450 BC) and Bronze Age (~2200 BC) in Central Europe. Our results reveal that the current diversity and distribution of haplogroup H were largely established by the Mid Neolithic (~4000 BC), but with substantial genetic contributions from subsequent pan-European cultures such as the Bell Beakers expanding out of Iberia in the Late Neolithic (~2800 BC). Dated haplogroup H genomes allow us to reconstruct the recent evolutionary history of haplogroup H and reveal a mutation rate 45% higher than current estimates for human mitochondria.

Keywords

Europe, Haplogroup, Haplotype, human mitochondria

Disciplines

Anthropology | Genetics and Genomics | Life Sciences | Social and Behavioral Sciences

Author(s)

Paul Brotherton, Wolfgang Haak, Jennifer E. L. Templeton, Guido Brandt, Julien Soubrier, Christina J. Adler, Stephen M. Richards, Clio Der Sarkissian, Robert Ganslmeier, Susanne Friederich, Veit Dresely, Mannis van Oven, Rosalie Kenyon, Mark B. Van der Hoek, Jonas Korlach, Khai Luong, Simon Y.W. Ho, Lluís Quintana-Murci, Doron M. Behar, Harald Meller, Kurt W. Alt, Alan Cooper, Syama Adhikarla, Arun Kumar Ganesh Prasad, Ramasamy Pitchappan, Arun Varatharajan Santhakumari, Elena Balanovska, Oleg Balanovska, Jaume Bertranpetit, David Comas, Begoña Martínez-Cruz, Marta Melé, Andrew C. Clarke, Elizabeth Matisoo-Smith, Matthew C. Dulik, Jill Bennett Gaieski, Amanda C. Owings, Theodore G. Schurr, Miguel G. Vilar, Jennifer L. Hobbs, Angela Hoobs, Himla Soodyall, Asif Javed, Laxmi Parida, Daniel E. Platt, Ajay K. Royyuru, Li Jin, Shilin Li, Matthew E. Kaplan, Nirav C. Merchant, R. John Mitchell, Colin Renfrew, Daniela R. Lacerda, Fabrício R. Santos, David F. Soria-Hernanz, R. Spencer Wells, Pandikumar Swamikrishnan, Chris Tyler-Smith, Pedro Paulo R. Vieira, Janet Ziegler, and Genographic Consortium

Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans

Paul Brotherton^{#1,2,*}, Wolfgang Haak^{#1,*}, Jennifer Templeton¹, Guido Brandt³, Julien Soubrier¹, Christina Jane Adler^{1,12}, Stephen M. Richards¹, Clio Der Sarkissian^{1,13}, Robert Ganslmeier⁴, Susanne Friederich⁴, Veit Dresely⁴, Mannis van Oven⁵, Rosalie Kenyon⁶, Mark B. Van der Hoek⁶, Jonas Korfach⁷, Khai Luong⁷, Simon Y. W. Ho⁸, Lluís Quintana-Murci⁹, Doron M. Behar¹⁰, Harald Meller⁴, Kurt W. Alt³, Alan Cooper³, and The Genographic Consortium

¹The Australian Centre for Ancient DNA, University of Adelaide, Adelaide, South Australia 5005, Australia ²Archaeogenetics Research Group, School of Applied Sciences, University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK ³Institute of Anthropology, Colonel-Kleinmann Weg 2, Johannes Gutenberg University, Mainz, D-55128 Mainz, Germany ⁴State Office for Heritage Management and Archaeology Saxony-Anhalt / State Museum for Prehistory Halle, Richard-Wagner-Straße 9, D-06114 Halle/Saale, Germany ⁵Department of Forensic Molecular Biology, Erasmus MC, University Medical Centre, Rotterdam, 3000 CA Rotterdam, The Netherlands ⁶SA Pathology, Adelaide, South Australia 5000, Australia ⁷Pacific Biosciences, USA ⁸School of Biological Sciences, The University of Sydney, New South Wales 2006, Australia ⁹Institut Pasteur, Paris, France ¹⁰Rambam Medical Centre, 31096 Haifa, Israel

These authors contributed equally to this work.

*Correspondence and requests for materials should be addressed to W.H. (wolfgang.haak@adelaide.edu.au).

¹²Current address: Institute of Dental Research, Westmead Centre for Oral Health, The University of Sydney, New South Wales 2145, Australia.

¹³Current address: Centre for Geogenetics, Natural History Museum of Denmark, 1350 Copenhagen, Denmark.

Members of The Genographic Consortium:

Syama Adhikarla, ArunKumar GaneshPrasad, Ramasamy Pitchappan, Arun Varatharajan Santhakumari Madurai Kamaraj University, Madurai, Tamil Nadu, India; **Elena Balanovska, Oleg Balanovsky** Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, Russia; **Jaume Bertranpetit, David Comas, Begoña Martínez-Cruz, Marta Melé**, Universitat Pompeu Fabra, Barcelona, Spain; **Andrew C. Clarke, Elizabeth A. Matisoo-Smith**, University of Otago, Dunedin, New Zealand; **Matthew C. Dulik, Jill B. Gaieski, Amanda C. Owings, Theodore G. Schurr, Miguel G. Vilar**, University of Pennsylvania, Philadelphia, Pennsylvania, United States; **Angela Hobbs, Himla Soodyall**, National Health Laboratory Service, Johannesburg, South Africa; **Asif Javed, Laxmi Parida, Daniel E. Platt, Ajay K. Royyuru**, IBM, Yorktown Heights, New York, United States; **Li Jin, Shilin Li**, Fudan University, Shanghai, China; **Matthew E. Kaplan, Nirav C. Merchant**, University of Arizona, Tucson, Arizona, United States; **R. John Mitchell**, La Trobe University, Melbourne, Victoria, Australia; **Lluís Quintana-Murci**, Institut Pasteur, Paris, France; **Colin Renfrew**, University of Cambridge, Cambridge, United Kingdom; **Daniela R. Lacerda, Fabrício R. Santos**, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; **David F. Soria Hernandez, R. Spencer Wells**, National Geographic Society, Washington, District of Columbia, United States; **Pandikumar Swamikrishnan**, IBM, Somers, New York, United States; **Chris Tyler-Smith**, The Wellcome Trust Sanger Institute, Hinxton, United Kingdom; **Pedro Paulo Vieira**, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil; **Janet S. Ziegler**, Applied Biosystems, Foster City, California, United States.

Author contributions:

P.B., W.H., and A.C. conceived and designed the project. P.B. designed and developed the DNA extraction, DNA library construction and hybridisation-based DNA-capture protocols (with assistance from J.T.). P.B., J.T., and W.H. generated and analysed the data. S.M.R., C.D., R.K., and M.B.v.d.H. contributed experimental steps and C.J.A., J.S., S.Y.W.H., J.K., and K.L. contributed analytical steps. G.B., R.G., S.F., V.D., M.v.O., L.Q., D.M.B., H.M., and K.W.A. provided ancient samples, contextual information, radiocarbon dating, and access to critical population data. P.B., W.H. and A.C. wrote the manuscript with input from C.J.A., J.S., S.Y.W.H., S.M.R., and J.K. All authors discussed the paper and gave comments.

Accession codes: The complete consensus mt genome sequences have been deposited to NCBI GenBank under accession numbers KC553980 to KC554018.

Competing financial interests: The authors declare no competing financial interests.

Abstract

Haplogroup (hg) H dominates present-day Western European mitochondrial (mt) DNA variability (>40%), yet was less common (~19%) amongst Early Neolithic farmers (~5450 BC) and virtually absent in Mesolithic hunter-gatherers. Here we investigate this major component of the maternal population history of modern Europeans and sequence 39 complete hg H mitochondrial genomes from ancient human remains. We then compare this 'real-time' genetic data with cultural changes taking place between the Early Neolithic (~5450 BC) and Bronze Age (~2200 BC) in Central Europe. Our results reveal that the current diversity and distribution of hg H were largely established by the Mid-Neolithic (~4000 BC), but with substantial genetic contributions from subsequent pan-European cultures such as the Bell Beakers expanding out of Iberia in the Late Neolithic (~2800 BC). Dated hg H genomes allow us to reconstruct the recent evolutionary history of hg H and reveal a mutation rate 45% higher than current estimates for human mitochondria.

Introduction

A key unanswered issue in human prehistory is the extent to which cultural change identifiable in the archaeological record can be ascribed to the movements of people, as opposed to the movements of just their ideas and artefacts. The Central European archaeological record identifies a succession of profound cultural and economic changes between the last hunter-gatherers of the Mesolithic and the first farmers of the Early Neolithic, through to the socially stratified chiefdoms of the Early Bronze Age¹⁻³. The exact nature and genetic context of the transformative changes that took place over these four millennia remain unclear^{4,5}, although current genetic patterns of mtDNA hg distribution suggest a complex series of events in European prehistory⁴⁻⁹ and hint at multiple inputs from outside Central Europe^{4,10,11}.

Phylogeographic studies suggest that mitochondrial (mt) haplogroup H (hg H) arrived in Europe from the Near East prior to the Last Glacial Maximum (22,000 BP), and survived in glacial refugia in Southwest Europe before undergoing a postglacial re-expansion^{4,12}. Haplogroup H now accounts for over 40% of mtDNA variation in anatomically modern humans across much of Western Eurasia, with declining frequencies south and east to ~10–30% in the Near East and Caucasus¹⁰. However, it remains uncertain when and how H became the dominant European haplogroup. Traditional approaches (including ancient DNA studies) have been unable to resolve either the phylogeny or phylogeographic distribution of H sub-haplogroups (sub-hgs)⁶, however they have generally relied on sequencing only 300–400 bp of the mitochondrial D-loop or control region^{10,13}. A number of studies based on complete 16.6 kb human mt genomes have revealed a complex evolutionary history for hg H (e.g. ^{12,14-18}, with phylogenetic analyses recognizing 87 H sub-hg¹⁹). These complete mt genomes revealed that 71% of hg H polymorphic diversity is located outside the D-loop, in the coding region²⁰ and, as a result, this diversity has not yet been exploited at the population genetics level.

To investigate the relationship between the European genetic and archaeological records we sequenced whole hg H mt genomes from skeletal remains directly assigned to distinct Central European archaeological cultures. Due to its excellently preserved human skeletal remains, forming a continuous record across a series of archaeological cultures since Palaeolithic times, the Mittelbe-Saale region of Saxony-Anhalt (Germany) provided a unique opportunity to address this issue. We analysed a time transect spanning the >3500 years of the Central European Neolithic period (Table 1, Supplementary Table S1), from the first farmers of the Early Neolithic *linear pottery culture* LBK (5450–4775 BC), through the subsequent Rössen (4625–4250 BC), Schöningen (4100–3950 BC), Baalberge (3950–3400 BC) and Salzmünde (3400–3025 BC) cultures. These were followed by two of the first pan-

European Late Neolithic cultural complexes, the Corded Ware (CWC, 2800-2050 BC) and Bell Beaker (BBC, 2500-2050 BC) cultures, before the emergence of the Early Bronze Age with the Unetice culture (2200-1575 BC). We chose to focus on hg H because of its recent dramatic rise in frequency to become the dominant hg in Europe, due to its presence in all Neolithic cultures in the Mittelbe-Saale region, and the potential it provided to explore detailed genetic structure on a sub-hg level. Overall, our results suggest that the broad foundations of the Central European mtDNA pool, here approximated via hg H, were formed during the Neolithic rather than the post-glacial period.

Results

Sequence and network analyses

From a collection of over 400 European prehistoric human archaeological remains we selected 37 Mittelbe-Saale individuals, as well as two samples from Italy (Supplementary Table S1), previously assigned to hg H by simplex and multiplex PCR⁷. Work was independently replicated for two samples per individual (Supplementary Methods). We designed and optimised a hybridisation-based DNA-capture system to sequence complete mt genomes on the Affymetrix MitoChip v2.0²¹ (Supplementary Methods, Supplementary Fig. S1-S2, and Supplementary Tables S2-S3) via immortalised libraries prepared from the highly damaged and degraded endogenous DNA recovered from archaeological remains^{22,23}. Six of the 39 target-enriched libraries were also analysed via a single-molecule, real-time (SMRT²⁴) Pacific Biosciences *RS* sequencing platform (Supplementary Table S4, Supplementary Dataset). In addition, 35/391 (9%) of all SNPs identified via the MitoChip were independently confirmed by direct PCR and Sanger sequencing (Supplementary Methods, Supplementary Tables S5 and S6). Mt genomes from all 39 individuals were unambiguously assignable to individual sub-hgs of hg H²⁰, confirming that a single human was typed in each case (Table 1). The mitochondrial hypervariable region I sequences matched those previously determined for each individual. The ancient hg H mt genomes were highly diverse, with 34 distinct haplotypes attributed to 20 major sub-hgs (gene diversity $H = 0.997 \pm 0.0071$; nucleotide diversity 0.000421 ± 0.000225), including three novel lineages (provisionally named H88-H90).

Phylogenetic network analysis of these ancient mt genomes reveals evidence of dynamic changes in the composition of H sub-hgs over the ~3500-year time transect (Figure 1). Importantly, sequences from older samples (and cultures) tend to represent basal lineages, only one to three mutations away from the ancestral root of hg H, while younger samples (after ~4000 BC) largely comprise more derived haplotypes appearing on longer branches. This temporal relationship provides further support for the authenticity of the ancient mt genomes.

Network analysis (Figure 1) reveals pronounced differences in the composition of sub-hgs between the Early Neolithic (ENE) cultures (LBK, Rössen, Schönningen), and those of the Mid/Late Neolithic to Early Bronze Age (MNE/LNE: Baalberge, Salzmünde, Corded Ware, Bell Beaker, Unetice). Early Neolithic (and in particular LBK) mt genomes are either rare today (H16, H23, H26), extinct or have not yet been observed in present-day populations (H46b, H88, H89). In sharp contrast, most of the later H sub-hgs are more common in present-day European populations (e.g. hg H3, H4, H6, H7, H11, and H13)^{12,14-16}. Of the 39 haplotypes detected, only three (within the common, basal, sub-hg H1) were shared between ENE and MNE/LNE cultures. Since the observed gene diversity is high, we might expect the number of shared haplotypes within and between cultures to be low. However, since the MNE/LNE haplotypes are on different sub-hg branches from the ENE haplotypes, these patterns combined show minimal local genetic continuity over this time period (Table 1).

Genetic distances

To further examine these apparent temporal shifts in sub-hg distribution, we tested whether hg H individuals represent different meta-populations by pooling them into different cultural and/or temporal groups of Early versus Late Neolithic cultures (Table 2, Supplementary Table S7). When pooled in four groups (ENE, MNE, LNE and Bronze Age), pairwise population comparisons via F_{ST} values based on sequence data showed that genetic distances increased with time over the duration of the Neolithic, reaching a significant value ($F_{ST} = 0.08722$; $p=0.00386\pm 0.0006$) between the Early Neolithic and the early Bronze Age (Table 2). This suggests a transformation of hg H diversity during the Neolithic period. This effect was less apparent (non-significant F_{ST} values) when samples from various sites were pooled in larger temporal groups (Table 2). However, non-parametric multivariate analysis of variance (NP-MANOVA, $p=0.0072$) also confirmed a significant difference between pooled groups of ENE and LNE individuals when comparisons were based on the presence or absence of sub-hgs (Table 2).

Genetic affinities

To examine potential geographic origins for Neolithic cultures (Supplementary Table S1) and to assess their contribution to present-day Central European mtDNA diversity, we used Principal Component Analysis (PCA) to investigate genetic affinities between three ancient culturally/temporally pooled groups (LBK, MNE, BBC) and 37 present-day Western Eurasian populations (Supplementary Table S8). PCA of the frequencies of the 15 most common H sub-hg showed that the present-day populations form three significantly supported geographic clusters (a grouping which was also supported using NP-MANOVA, $p<0.0001$; Table 2a): (i) Iberia in the west; (ii) the Caucasus, the Near East, and Anatolia; and (iii) Central and Eastern Europe from the Urals to France (Figure 2a). This particular number of clusters was also the best supported in a model-based test on sub-hg H frequencies followed by Ward clustering (Figure 2c-d). We also used Procrustes analysis to quantify the relationship between hg H substructure and the geographic locations of both the present-day Western Eurasian and the Mittelbe-Saale ancient populations. For this analysis, we superimposed the PCA co-ordinates on the geographic map of the present-day and ancient sampling locations. We found a striking resemblance between the genetic and geographic maps, with a highly significant Procrustes similarity score ($t_0 = 0.733$) obtained for the comparison ($p < 10^{-6}$; 100,000 permutations). The analysis supported a clustering of the transformed genetic data from present-day populations into the three major groups described above (Figure 2b). In contrast, Procrustes analysis clearly showed that the genetic data for LBK and BBC samples were not related to their geographic location. Although all three ancient groups were sampled from the same Central European location only the MNE group genetically resembles present-day populations from this region.

The combined set of analyses (PCA, Procrustes and Ward clustering) revealed that Mittelbe-Saale's earliest farmers (LBK; $n=9$) cluster with present-day Caucasus, Near Eastern, and Anatolian populations, as previously noted⁷. In contrast, individuals from the successor series of regional post-LBK (and Mid Neolithic) Rössen, Schöningen, Baalberge, and Salzmünde cultures (ca. 4625-3025 BC, MNE; $n=10$) cluster with present-day Central European populations (Figure 2). Mitochondrial genomes from BBC individuals in Mittelbe-Saale (BBC; $n=7$) display close genetic affinities to present-day Iberian populations (Figure 2). The component loadings of the PCA biplot indicate that this is largely based on high frequencies of sub-hgs H1 and H3, which are thought to have spread from a glacial Iberian refugium¹³ and which have also been reported from ancient Neolithic sites from France and Spain^{8,25}. Other Late Neolithic samples add further to the genetic complexity. Individuals from the Corded Ware culture (CWC; 2800-2200 BC), which has archaeological associations towards North-Eastern Europe, produced two distinct mt

genomes (H1_TBD and H6a1a) which have not been found in their contemporaneous Bell Beaker neighbours, nor in preceding Central European cultures. Similarly, data from the subsequent Early Bronze Age Unetice culture revealed haplotypes with genetic affinities to both the East (sub-hg H2a, H7 and H11) and the West (sub-hg H3 and H4), based on frequency distributions of these sub-hg in present-day populations¹³. We also included two individuals from outside Central Europe (Sardinia and South Tyrol) and from different time periods (Nuragic Bronze Age and Iron Age, respectively) to further investigate genetic diversity within hg H and to test the power of resolution of complete mt genomes. Both individuals had mt genomes that are not found in samples from the Mittelbe-Saale region. The Iron Age sample from South Tyrol produced another new sub-hg (provisional H90) and the Bronze Age individual from Sardinia a new H1 haplotype (H1aw1).

Reconstructing the demographic history of mtDNA hg H

It has previously proved difficult to use present-day data alone to determine when hg H became the predominant haplogroup in Europe, as archaeogenetic and palaeodemographic reconstructions have very large uncertainties^{4,26}. However, as our 39 dated ancient mt genome sequences provide precise temporal calibration points, we performed a Bayesian skyride analysis with 200 random present-day mt genome sequences to reconstruct the lineage history of hg H through time (with the caveat of assuming a continuous and panmictic population). The resulting skyride plot (Figure 1b) is the first real-time estimation of the European hg H population size (and consequently its contribution to Europe's effective population size and demographic history) with a broad temporal coverage over ~3500 years of the Neolithic period in Central Europe (5500–2000 BC). Hg H shows a consistent and strong exponential growth over the entire course of the Neolithic. The estimated population size tracks the European census size²⁷ and population density estimates from archaeological sites²⁸ in the Late Holocene, but also provides detailed estimates for prehistoric times for which data points remain very scarce (Figure 1b).

Another major advantage of the temporal calibration points provided by ancient hg H mt genomes is that the data allow a relatively precise estimate of the evolutionary substitution rate for human mtDNA. The temporal dependency of evolutionary rates predicts that rate estimates measured over short timespans will be considerably higher than those using deep fossil calibrations, such as the human/chimpanzee split at ~6 million years²⁹. The rate calibrated by the Neolithic and Bronze Age sequences is 2.4×10^{-8} substitutions/site/year (1.7 to 3.2×10^{-8} ; 95% High Posterior Density) for the entire mt genome, which is 1.45× (44.5%) higher than current estimates based on the traditional human/chimp split (e.g. 1.66×10^{-8} for the entire mt genome³⁰ and 1.26×10^{-8} for the coding region³¹). Consequently, the calibrated 'Neolithic' rate infers a considerably younger coalescence date for hg H (10.9 to 19.1 kya) than those previously reported (19.2 to 21.4 kya for HVSI¹⁰, 15.7 to 22.5 kya for the mt coding region³¹ or 14.7-22.6 kya when corrected for purifying selection³⁰).

Discussion

Despite recent successes in sequencing portions of nuclear genomes from Meso- and Neolithic samples^{11,32,33}, mtDNA remains the most widely studied and best described marker in population genetics. While its interpretation is limited to the matrilineal genetic history^{4,13,19}, this can be an important socio-cultural and demographic signal additional to that gained from autosomal loci^{34,35}. Our results clearly demonstrate that high-resolution full mt genome-typing, combined with the ability to analyse large numbers of individuals from multiple cultural layers, can provide highly resolved temporal views that are not yet practical with nuclear DNA studies.

The phylogenetic network analysis of our chronological hg H mt genome dataset (Figure 1a) provides the first detailed real-time view of mutations in human mtDNA. It has enabled the direct observation of the mutation rate over thousands of years and revealed a distinct temporal distribution pattern of hg H diversity. While a temporal pattern could be expected in an expanding population with stable/increasing hg H frequencies (Figure 1b, Supplementary Fig. S3), ENE and MNE/LNE/Bronze Age samples clearly show a mutually exclusive sub-hg distribution with the exception of sub-hg H1, which is the most common and basal sub-hg within H¹⁴⁻¹⁶. Under an assumption of genetic continuity we would expect MNE/LNE and Bronze Age individuals to be on the same sub-hg branches as Early Neolithic individuals. Instead, ENE mt genomes are generally either rare today¹⁹ or have not yet been observed in present-day populations, possibly due to subsequent extinction of these lineages. In contrast, most MNE/LNE and Bronze Age sub-hgs are still common today. This suggests that individuals from the Early Neolithic made a marginal contribution to Late Neolithic and present-day hg H diversity. Although the relatively small sample numbers from each time period limit detailed analyses of the causes of the distribution shifts, we interpret this phylogenetic pattern as a genetic discontinuity between Early and subsequent Neolithic cultures in Europe, potentially mirroring genetic structure in Neolithic European populations. Genetic drift could also have played a role in generating discrepant hg distributions over time and space. However, if drift was the sole cause we would expect a random distribution across all sub-hgs rather than a clear distinction between ENE and MNE/LNE/Bronze Age mt genomes.

Our genetic distance data also indicate minimal local genetic continuity between the Early Neolithic and the Mid/Late Neolithic in Central Europe (Figure 1; Table 1), again suggesting that Early Neolithic lineages were largely superseded during the Mid/Late Neolithic (~4100-2200 BC) in a previously unrecognised major genetic transition. This pronounced genetic changeover between Early and Mid/Late Neolithic cultures is comparable to other known major genetic transition thus far revealed by ancient DNA and coalescent simulations (between indigenous European hunter-gatherers and incoming early farmers from the Near East during the initial Meso-Neolithic transition from ~7500 BC in Central Europe)^{6,7}. When compared to hg H diversity in present-day Central Europe^{14,15,18,36,21}, the network in Figure 1 suggests that much of the present-day diversity can be attributed to the incorporation of new lineages in the Mid/Late Neolithic and emerging Bronze Age (from 2200 BC). The Late Neolithic in particular is known to have been a period of profound cultural and economic change³⁷, with newly emerging pan-European cultures such as the Bell Beaker phenomenon in Western Europe and the Corded Ware culture in north eastern Europe. It therefore seems likely that these pan-European cultures were associated with the introduction of lineages from outside Central Europe. Fortunately, the ranges of both these groups overlapped in the Mittelbe-Saale sample area (Supplementary Methods), allowing this possibility to be further investigated.

Our data on genetic affinities (PCA, Procrustes and Ward clustering) revealed that Mittelbe-Saale's earliest farmers (LBK; n=9) cluster with present-day Caucasus, Near Eastern, and Anatolian populations. These findings are consistent with a highly detailed archaeological record tracing the temporal and spatial spread of agriculture into Central Europe; beginning initially in Anatolia and the Near East, where farming originated ~12,000 years ago^{7,1}.

Our observation that individuals from the successor series of regional post-LBK and Mid Neolithic cultures (Rössen, Schöningen, Baalberge, and Salzmünde) cluster with present-day Central European populations could be explained by a loss of lineages from the Early Neolithic LBK period during a short phase of population decline in the centuries after 5000 BC (as proposed in some archaeological models)³⁸. However, our results suggest that

mtDNA H sub-hg diversity established during the Mid Neolithic is still present in Central European populations today. This is consistent with independent archaeological evidence of a phase of more localised cultural development during the MNE period, potentially involving influences from contemporaneous Mid Neolithic cultures outside Mittelbe-Saale, which (perhaps in concert with LBK population decline) could have resulted in a replacement of most Early Neolithic H sub-hgs. Together, the genetic and archaeological evidence highlight the complexities of both the formative and consolidation phases in Central Europe.

From around 2800 BC, the Late Neolithic Bell Beaker culture emerged from the Iberian Peninsula to form one of the first pan-European archaeological complexes. This cultural phenomenon is recognised by a distinctive package of rich grave goods including the eponymous bell-shaped ceramic beakers. The genetic affinities between Central Europe's Bell Beakers and present-day Iberian populations (Figure 2) is striking and throws fresh light on long-disputed archaeological models³. We suggest these data indicate a considerable genetic influx from the West during the Late Neolithic. These far-Western genetic affinities of Mittelbe-Saale's Bell Beaker folk may also have intriguing linguistic implications, as the archaeologically-identified eastward movement of the Bell Beaker culture has recently been linked to the initial spread of the Celtic language family across Western Europe³⁹. This hypothesis suggests that early members of the Celtic language family (e.g. Tartessian)⁴⁰ initially developed from Indo-European precursors in Iberia and subsequently spread throughout the Atlantic Zone; before a period of rapid mobility, reflected by the Beaker phenomenon, carried Celtic languages across much of Western Europe. This idea not only challenges traditional views of a linguistic spread of Celtic westwards from Central Europe during the Iron Age, but also implies that Indo-European languages arrived in Western Europe substantially earlier, presumably with the arrival of farming from the Near East⁴¹.

Other Late Neolithic population movements appear to have added further genetic complexity, as exemplified by the Corded Ware culture (CWC; 2800-2200 BC), which preceded the Bell Beaker culture in Mittelbe-Saale and has archaeological associations with North-Eastern Europe. A genetic affinity to eastern populations is consistent with two distinct CWC mt genomes (H1_TBD and H6a1a) not identified in either their contemporaneous Bell Beaker neighbours or in preceding Central European cultures. The subsequent Early Bronze Age Unetice culture, associated with emerging metallurgy and increasingly stratified societies^{37,42}, marks a consolidation of social and cultural systems in Mittelbe-Saale that were established during the Late Neolithic by the two pan-European Bell Beaker and Corded Ware cultures. The Unetice culture appears contemporaneously with the last Neolithic horizon (~2200 BC) in areas where elements of both the Bell Beaker and Corded Ware cultures are present, sometimes overlapping at the same sites. It is therefore not surprising that individuals ascribed to the newly emerging Unetice culture carry mt genomes with both Western (sub-hgs H3 and H4) and Eastern (sub-hgs H2a, H7 and H11) associations.

The demographic reconstruction, which is based on direct calibration points, has major implications for understanding post-glacial human history in Europe. Our new estimate is incompatible with traditional views that the majority of present-day hg H lineages were carried into Central, Northern, and Eastern Europe via a post-glacial human population expansion prior to the Holocene (12 kya)¹³. Our data complements a recent study, based on present-day mt genomes, which describes a pronounced population increase at ~7000 BC (interpreted as a Neolithic expansion into Europe), but followed by a slow population growth until the present day²⁶. By including ancient DNA data from across the critical time points in question, our skyride plot corrects for missing temporal data and suggests

substantial growth of hg H from the beginning of the Neolithic and continuing throughout the entire Neolithic period. This emphasizes the role of farming practices and cultural developments in the demographic expansions inferred in subsequent time periods, which have not yet been explored genetically.

Although an expansion of hg H could in principle be compatible with a post-glacial resettling of Northern and Central Europe from southwestern refugia^{12,16} (as indicated by our population skyline and PCA plots), we instead propose that the rise of hg H to become the predominant mtDNA branch in Europe was mediated by subsequent demographic events during the Neolithic, as shown by a general increase in hg H frequency and strong population growth during this period (Figure 1b). Support for this position comes from data suggesting that hg H was virtually absent amongst Central and Northern European hunter-gatherers^{6,43} and formed only 19% in LBK individuals, most likely introduced from Southeast Europe and/or the Near East⁷. In our updated dataset from Mittelbe-Saale, hg H appears to have been established by the LBK period and increased in frequency after 4000 BC (Supplementary Fig. S3). Interestingly, Mid/Late Neolithic cultures with cultural associations to the North and Northeast, such as the Bernburg and Corded Ware cultures, show reduced hg H frequencies and hg H only moved northwards into southern Scandinavia during the Neolithisation of Northern Europe around the Middle Neolithic, as exemplified by individuals from the Funnel Beaker Culture^{11,43}. However, hg H appears to have been generally more frequent in prehistoric Western Europe: at 20% from a Middle Neolithic (3030-2890 calBC) site in France⁹; at ~25% from Iberian (Epi-)Cardial Neolithic samples^{8,25}; at 36% from a Neolithic site in Catalonia⁴⁴; and at 44% from Neolithic sites from the Basque Country and Navarre⁴⁵. Importantly, a recent study on Iberian hunter-gatherers revealed the presence of hg H there in Mesolithic times⁴⁵. In Mittelbe-Saale, the Bell Beaker samples signpost a significant increase in hg H frequency (the 95% confidence intervals do not overlap with earlier LBK and Schöningen Neolithic cultures; Supplementary Fig. S3). In conclusion, the Western European Neolithic and the widespread pan-European Bell Beaker phenomenon appear to be important factors in driving the spread of H sub-hgs throughout large parts of Western Europe. In particular, high proportions of sub-hgs H1 and H3 seem to have made substantial contributions to the hg H diversity that exists in Western and Central Europe today¹⁶. Having reached significant levels, and assuming a generally higher rate of population growth in southern and western Europe in post-Neolithic times²⁷, these Neolithic processes appear to have been the major factor in hg H becoming the predominant European mtDNA haplogroup.

Overall, our results suggest that the broad foundations of the Central European mtDNA pool, here approximated via hg H, were formed during the Neolithic rather than the post-glacial period. Early Neolithic hg H mt lineages brought in from the Near East by Central Europe's first farmers do not appear to have contributed significantly to present-day Central Europe's hg H diversity, instead being largely superseded during the Mid and Late Neolithic (with the process starting around 4000 BC), after which there appears to have been substantial genetic continuity to the present-day in Central Europe. These developments have been revealed by comparative full mt genome sequencing and would have remained obscure using standard HVS I data.

In conclusion, demographic changes across the Mid Neolithic, followed by the widespread Bell Beaker cultural phenomenon, are likely to have been the key factors in the expansion of hg H across Western Europe and the eventual rise of hg H to become the predominant mtDNA hg. However, Late Neolithic Corded Ware and Early Bronze Age data suggest a complex series of additional genetic contributions, which require further investigation.

Methods

Ancient DNA analyses

DNA was extracted from two independent samples for each individual (Supplementary Methods). HVS I was amplified using a minimum of four short overlapping primer pairs, following established protocols and authentication criteria as described previously^{7,46}. Multiplex SNP typing of 22 haplogroup informative SNPs (GenoCoRe22) was carried out using a SNaPshot based protocol as described previously⁷.

Ancient DNA Library preparation

Ancient DNA extract polishing/phosphorylation, adaptor ligation and polymerase ‘fill-in’ reactions were used sequentially to create fully double-stranded adaptor-tagged aDNA libraries (Figure 3). Following every step, DNA was purified using MinElute spin columns (Qiagen) as per the manufacturer’s instructions. PCR amplification reactions were then performed to create ‘primary’ DNA libraries, ready for DNA-capture hybridisation steps, and amplification products were sized and quantified (Supplementary Methods).

Hybridisation-based enrichment of human mtDNA

The basic conceptual design for the hybridisation of tracer DNA (aDNA library) to biotinylated driver DNA sequences (human mitochondrial probe) was previously described⁴⁷ and the overall scheme is outlined in Figure 3. The two library-specific PCR primers were included as part of the hybridisation mix as blocking oligonucleotides to minimise unwanted hybridisation between the adaptor-tagged flanking regions of otherwise unrelated single-stranded library DNA molecules⁴⁸. A key innovation of this methodology was the use of a DNA polymerase with strand-displacing activity after post-hybridisation stringency washes. This allowed primer extension from the bound library (blocking) primers to disrupt the double-stranded region of stable hybridisation between human mitochondrial probe DNA sequences and single-stranded library DNA molecules that had inserts with complementary sequences. These mtDNA-enriched library DNA molecules captured in the hybridisation step could thereby be cleanly separated from biotinylated probe molecules, which remained bound to magnetic Streptavidin beads. PCR reamplification reactions from the mt-enriched library DNA molecules comprised the ‘first enrichment’ DNA libraries. In general, we used three cycles of hybridisation / enrichment / reamplification to produce DNA libraries highly enriched for short endogenous mtDNA sequence fragments ready for genotyping (Supplementary Methods).

Affymetrix Mitochip v2.0 array typing and Pacific Biosciences SMRT® sequencing

MtDNA-enriched libraries underwent biotin labelling using Terminal Deoxynucleotidyl Transferase (TdT) as per the Affymetrix GeneChip Whole Transcript Sense Target Labelling Assay Manual (P/N 701880, rev. 4). Biotin-labelled DNA libraries were hybridised to Affymetrix GeneChip Human Mitochondrial Resequencing 2.0 Arrays for 17 hours at 49°C. Arrays were washed, stained and scanned as per the GeneChip CustomSeq Resequencing Array Protocol (P/N 701231, rev. 5). Affymetrix GeneChip Command Console software (v3.2) was used to generate CEL files, which were then analysed using GeneChip Sequence Analysis Software (GSEQ v4.1, Affymetrix) and validated using the software Geneious⁴⁹ (Supplementary Fig. S1 and S2, Supplementary Tables S2 and S3). Six of the mt-enriched DNA libraries were also converted to SMRTbell template libraries for sequencing on a Pacific Biosciences *RS* platform (Supplementary Methods).

Network analyses

A median joining network of all ancient hg H mt genomes (Figure 1a) was constructed manually using the most up-to-date version of the mitochondrial phylogenetic tree (PhyloTree.org, mtDNA tree Build 14) as a scaffold on which to place the observed hg H lineages^{19,20}. This version included a revised version of the hg H sub-tree comprising 1203 sequences in total. As per convention, insertions at np 309.1C(C), 315.1C, 523-524d (aka 522-523d), 16182C, 16183C, 16193.1C(C) and mutation 16519 were not considered for phylogenetic reconstruction²⁰.

Procrustes-based Principal Component analysis and Ward Clustering

PCA was used to describe and visualise the maternal genetic relationships amongst the Neolithic cultures investigated, as well as to 37 present-day European and Near Eastern populations (Figure 2a). PCA was performed on the frequency of H sub-hg taken from the literature (Supplementary Table S8). To minimise statistical noise caused by rare sub-clades and to allow for data compatibility across published studies we considered only the following 15 most common H sub-hg in Europe and the Near East: H*, H1, H1a, H1b, H2, H2a1, H3, H4, H5, H5a, H6, H6a, H7, H8 and H11. PCAs were performed and visualised in R version 2.11.1⁵⁰ using a customised script based on the function *prcomp*.

Ancient hg H individuals were pooled into three different groups based on the numbers of samples available: two for ‘pan-European’ archaeological phenomena/cultures alongside hypothesised geographic origins (LBK, n=9 and BBC, n=7); and a temporally transitional group pooling regional (mostly Mid Neolithic) cultures (MNE, n=10). Small sample sets such as the Corded Ware (n=2) and later Bronze Age Unetice (n=5) were excluded. To test whether the clustering pattern observed in the PCA was significantly supported, we performed a number of statistical tests including Ward clustering, Procrustes analysis and NP-MANOVA (as described below). First, we performed model-based cluster tests to identify the number of clusters via the model with the best support (highest Bayes Information criterion; BIC) followed by Ward hierarchical clustering of sub-hg H frequencies using the packages *mclust*, *pvclust* (for bootstrap values) and *hclust* in R, respectively. Procrustes analysis was also performed in R using the package *vegan* based on PCA scores and geographic coordinates (Supplementary Table S8) and the function *protest* to calculate the similarity score (100,000 permutations).

Summary statistics

Population pairwise F_{ST} , Slatkin’s linearised F_{ST} and haplotype diversity were calculated in Arlequin version 3.5⁵¹. We used jMODELTEST 0.1.1⁵² in order to find the best fitting evolutionary model and, if required, to estimate a discrete gamma shape parameter for our 39 non-partitioned mt genomes. Based on the resulting scores for each model (AIC and BIC), we subsequently used the Tamura & Nei model and a gamma value of 0.049 for our calculations of population distances in Arlequin. The ancient hg H individuals were pooled into different temporal/cultural groups in order to calculate genetic diversity indices and to test for genetic differentiation (Table 2, Supplementary Table S7).

Multivariate analysis of variance

We performed a non-parametric multivariate analysis of variance (NP-MANOVA) to test whether the temporal grouping of ancient individuals according to archaeological time periods are statistically supported. The NP-MANOVA was performed on a Raup-Crick distance matrix, which was produced from the presence/absence of the 15 hg H sub-clades used in the PCA. Calculations were performed in PAST version 2.09 with 10,000 permutations per test and post-hoc Bonferroni correction to account for multiple

comparisons and small sample sizes⁵³. We also tested whether the clustering pattern between the ancient and present-day populations observed in the PCA was significantly supported.

Bayesian skyride analyses and mutation rate calculation

The data set comprised 37 newly sequenced, non-related, ancient mt genomes, five sets of randomly chosen, distinct, present-day hg H mt genomes from Phylotree (<http://www.phylotree.org>, mtDNA tree Build 12 (20th July 2011)) and 420 newly available hg H sequences¹⁷. The Sequences were manually aligned to the revised Cambridge Reference Sequence (rCRS: AC_000021)⁵⁴ using the program SeaView⁵⁵. The alignment was partitioned into 4 subsets, representing the D-loop, the protein-coding regions (1st+2nd codon positions and 3rd codon position) and a concatenation of tRNA and RNA genes. Insertions at nps 309.1C(C), 315.1C, 523-524d (aka 522-523d), 16182C, 16183C, 16193.1C(C) were not considered for phylogenetic reconstruction and position 16519 was removed from the D-loop subset²⁰. The best substitution models were selected using ModelGenerator 0.85⁵⁶ by comparison of Bayesian Information Criterion scores: HKY+G for D-loop, TN+G for protein-coding regions and HKY for RNA genes. Considering the short evolutionary timescale being studied (intra-haplogroup), models including a proportion of invariant sites were excluded. A Bayesian skyride analysis⁵⁷ was performed using the phylogenetic software BEAST 1.6.1⁵⁸ and calibrated using radiocarbon dates from 18 of the ancient individuals and mean archaeological dates for the remaining individuals. This allowed us to achieve a broad temporal coverage for ~3500 years of the Neolithic period in Central Europe (5500–2000 BC) and to generate the most precise demographic reconstruction of hg H. Results were replicated using independent sets of 100 (1×), 200 (3×), and 300 (1×) mt genomes. A strict molecular clock was used, allowing for a distinct rate in each subset of the alignment. Additional analysis using an uncorrelated log normal relaxed clock to account for potential rate variations could not reject the strict clock assumption. Convergence was checked by sampling from two independent Markov chains. Each MCMC analysis was run for 100,000,000 steps and samples from the two chains were combined, after discarding the first 10% of samples as burn-in. All parameters showed sufficient sampling, indicated by effective sample sizes above 200. Tracer 1.5 was used to produce the skyride plot (Figure 1b)⁵⁹.

We carried out a ‘date randomisation test’³⁸, to test whether the signal from the radiocarbon dates associated with the ancient sequences was sufficient to calibrate the hg H phylogeny⁶⁰. This test randomises all dates associated with the sequences (including present-day ones) and replicates of the phylogenetic analysis as described above. If the structure and spread of the ancient sequences in the tree were sufficient to calibrate the analysis, the inferred mean rate of the randomised analysis should be significantly different from the rate calculated using the correct association date/sequence. In other words, the 95% HPD of the randomised analysis should not overlap with the mean rate estimated without randomisation. The comparison of estimated rates from the main analysis and from ten replicates with randomised dates presented in Supplementary Fig. S4 confirms the presence of sufficient signal to calibrate the tree provided by dates from the 37 ancient samples.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are indebted to Matt Kaplan and Ryan Spriggs at Arizona Research Laboratories, Division of Biotechnology, University of Arizona Genetics Core Facility, <http://uagc.arl.arizona.edu/>, Tyson Clark, Michael Brown, Kristi

Spittle, and Matthew Boitano (Pacific Biosciences) for sequencing work, Jeremy Timmis for help with DNA sonication protocols, and Robin Skeates, and Hubert Steiner for additional samples and contextual information. We thank the Australian Research Council (grant LP0882622), the Deutsche Forschungsgemeinschaft (AI 287/7-1 and Me 3245/1-1) and National Geographic's Genographic Project for funding. M.v.O. was supported in part by the Netherlands Forensic Institute (NFI) and a grant from the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) within the framework of the Forensic Genomics Consortium Netherlands (FGCN).

References

1. Whittle, AWR.; Cummings, V. Going over: the mesolithic-neolithic transition in North-West Europe. Oxford University Press; Oxford: 2007. p. 632
2. Sherratt, A. Plough and pastoralism: aspects of the secondary products revolution. In: Hodder, I.; Isaac, G.; Hammond, N., editors. *Patterns of the Past: Studies in honour of David Clarke*. Cambridge University Press; Cambridge: 1981. p. 261-305.
3. Bogucki, PI.; Crabtree, PJ. *Ancient Europe 8000 B.C.--A.D. 1000: Encyclopedia of the Barbarian World*. Charles Scribner's Sons; 2004. p. 1221
4. Soares P, et al. The archaeogenetics of Europe. *Curr Biol*. 2010; 20:R174–83. [PubMed: 20178764]
5. Pinhasi R, Thomas MG, Hofreiter M, Currat M, Burger J. The genetic history of Europeans. *Trends in genetics: TIG*. 2012
6. Bramanti B, et al. Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science*. 2009; 326:137–40. [PubMed: 19729620]
7. Haak W, et al. Ancient DNA from European Early Neolithic Farmers Reveals Their Near Eastern Affinities. *PLoS Biology*. 2010; 8:e1000536. [PubMed: 21085689]
8. Gamba C, et al. Ancient DNA from an Early Neolithic Iberian population supports a pioneer colonization by first farmers. *Molecular Ecology*. 2012; 21:45–56. [PubMed: 22117930]
9. Lacan M, et al. Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108:9788–91. [PubMed: 21628562]
10. Richards M, et al. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet*. 2000; 67:1251–76. [PubMed: 11032788]
11. Skoglund P, et al. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science*. 2012; 336:466–9. [PubMed: 22539720]
12. Pereira L, et al. High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res*. 2005; 15:19–24. [PubMed: 15632086]
13. Torroni A, Achilli A, Macaulay V, Richards M, Bandelt HJ. Harvesting the fruit of the human mtDNA tree. *Trends Genet*. 2006; 22:339–45. [PubMed: 16678300]
14. Roostalu U, et al. Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: The near eastern and Caucasian perspective. *Molecular Biology and Evolution*. 2007; 24:436–448. [PubMed: 17099056]
15. Loogväli EL, et al. Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Molecular Biology and Evolution*. 2004; 21:2012–2021. [PubMed: 15254257]
16. Achilli A, et al. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *American Journal of Human Genetics*. 2004; 75:910–8. [PubMed: 15382008]
17. Behar DM, et al. The Basque Paradigm: Genetic Evidence of a Maternal Continuity in the Franco-Cantabrian Region since Pre-Neolithic Times. *American Journal of Human Genetics*. 2012; 90:486–493. [PubMed: 22365151]
18. Alvarez-Iglesias V, et al. New population and phylogenetic features of the internal variation within mitochondrial DNA macro-haplogroup R0. *PLoS ONE*. 2009; 4:e5112. [PubMed: 19340307]
19. Behar DM, et al. A "Copernican" Reassessment of the Human Mitochondrial DNA Tree from its Root. *American Journal of Human Genetics*. 2012; 90:675–684. [PubMed: 22482806]
20. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*. 2009; 30:E386–94. [PubMed: 18853457]

21. Hartmann A, et al. Validation of Microarray-Based Resequencing of 93 Worldwide Mitochondrial Genomes. *Human Mutation*. 2009; 30:115–122. [PubMed: 18623076]
22. Maricic T, Whitten M, Paabo S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE*. 2010; 5:e14004. [PubMed: 21103372]
23. Margulies M, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005; 437:376–80. [PubMed: 16056220]
24. Korlach J, et al. Real-time DNA sequencing from single polymerase molecules. *Methods in enzymology*. 2010; 472:431–55. [PubMed: 20580975]
25. Lacan M, et al. Ancient DNA suggests the leading role played by men in the Neolithic dissemination. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108:18255–9. [PubMed: 22042855]
26. Fu Q, Rudan P, Pääbo S, Krause J. Complete Mitochondrial Genomes Reveal Neolithic Expansion into Europe. *PLoS ONE*. 2012; 7:e32473. doi: 10.1371/journal.pone.0032473. [PubMed: 22427842]
27. Livi-Bacci, M. *A Concise History of World Population*. Blackwell Publishing; Malden, Oxford, Carlton: 2007. p. 279
28. Zimmermann A, Hilpert J, Wendt KP. Estimations of population density for selected periods between the Neolithic and AD 1800. *Human Biology*. 2009; 81:357–80. [PubMed: 19943751]
29. Ho SY, Shapiro B, Phillips MJ, Cooper A, Drummond AJ. Evidence for time dependency of molecular rate estimates. *Systematic biology*. 2007; 56:515–22. [PubMed: 17562475]
30. Soares P, et al. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet*. 2009; 84:740–59. [PubMed: 19500773]
31. Mishmar D, et al. Natural selection shaped regional mtDNA variation in humans. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100:171–6. [PubMed: 12509511]
32. Keller A, et al. New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nature communications*. 2012; 3:698.
33. Sanchez-Quinto F, et al. Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. *Current biology: CB*. 2012; 22:1494–9. [PubMed: 22748318]
34. Behar DM, et al. The Basque paradigm: genetic evidence of a maternal continuity in the Franco-Cantabrian region since pre-Neolithic times. *American Journal of Human Genetics*. 2012; 90:486–93. [PubMed: 22365151]
35. Gunnarsdottir ED, et al. Larger mitochondrial DNA than Y-chromosome differences between matrilineal and patrilineal groups from Sumatra. *Nature communications*. 2011; 2:228.
36. Brandstätter A, et al. Dissection of mitochondrial superhaplogroup H using coding region SNPs. *Electrophoresis*. 2006; 27:2541–50. [PubMed: 16721903]
37. Heyd V. Families, prestige goods, warriors & complex societies: Beaker groups of the 3rd millennium cal BC along the upper & middle Danube. *Proceedings of the Prehistoric Society*. 2007; 73:327–379.
38. Shennan S, Edinborough K. Prehistoric population history: from the late glacial to the late neolithic in central and northern Europe. *Journal of Archaeological Science*. 2007; 34:1339–1345.
39. Cunliffe, B.; Koch, JT., editors. *Celtic from the West: Alternative Perspectives from Archaeology, Genetics, Language and Literature*. Oxbow Books; Oxford: 2010. p. 384
40. Koch, JT. *Tartessian. Celtic in the South-west at the Dawn of History*. Aberystwyth; 2009.
41. Renfrew, C. *Archaeology and language: the puzzle of Indo-European origins*. Vol. XIV. Cape; London: 1988. 346 S. Ill., Kt.
42. Nowak M. Transformations in East-Central Europe from 6000 to 3000 BC: local vs. foreign patterns. *Documenta Praehistorica*. 2006; XXXIII:143–158.
43. Malmström H, et al. Ancient DNA reveals lack of continuity between neolithic hunter-gatherers and contemporary Scandinavians. *Current biology: CB*. 2009; 19:1758–62. [PubMed: 19781941]
44. Sampietro ML, et al. Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proceedings. Biological sciences / The Royal Society*. 2007; 274:2161–7. [PubMed: 17609193]

45. Hervella M, et al. Ancient DNA from hunter-gatherer and farmer groups from Northern Spain supports a random dispersion model for the Neolithic expansion into Europe. *PLoS ONE*. 2012; 7:e34417. [PubMed: 22563371]
46. Haak W, et al. Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science*. 2005; 310:1016–8. [PubMed: 16284177]
47. Patel M, Sive H. PCR-based subtractive cDNA cloning. *Current protocols in molecular biology*. 2001; Chapter 25(Unit 25B 2)
48. Tao SC, Gao HF, Cao F, Ma XM, Cheng J. Blocking oligo--a novel approach for improving chip-based DNA hybridization efficiency. *Molecular and cellular probes*. 2003; 17:197–202. [PubMed: 12944123]
49. Drummond, AJ., et al. Geneious v5.4. 2011. Available from <http://www.geneious.com/>
50. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2010.
51. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*. 2010; 10:564–7. [PubMed: 21565059]
52. Posada D. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*. 2008; 25:1253–6. [PubMed: 18397919]
53. Hammer O, Harper DAT, Ryan PD. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica*. 2001; 4
54. Andrews RM, et al. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat.Genet*. 1999; 23:147. [PubMed: 10508508]
55. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*. 2010; 27:221–4. [PubMed: 19854763]
56. Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McLnerney JO. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC evolutionary biology*. 2006; 6:29. [PubMed: 16563161]
57. Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution*. 2008; 25:1459–71. [PubMed: 18408232]
58. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*. 2007; 7:214. [PubMed: 17996036]
59. Rambaut, A.; Drummond, A. Tracer. v1.4.2007. Available from <http://beast.bio.ed.ac.uk/software/tracer>
60. Ho SY, et al. Bayesian estimation of substitution rates from ancient DNA sequences with low information content. *Systematic biology*. 2011; 60:366–75. [PubMed: 21296909]

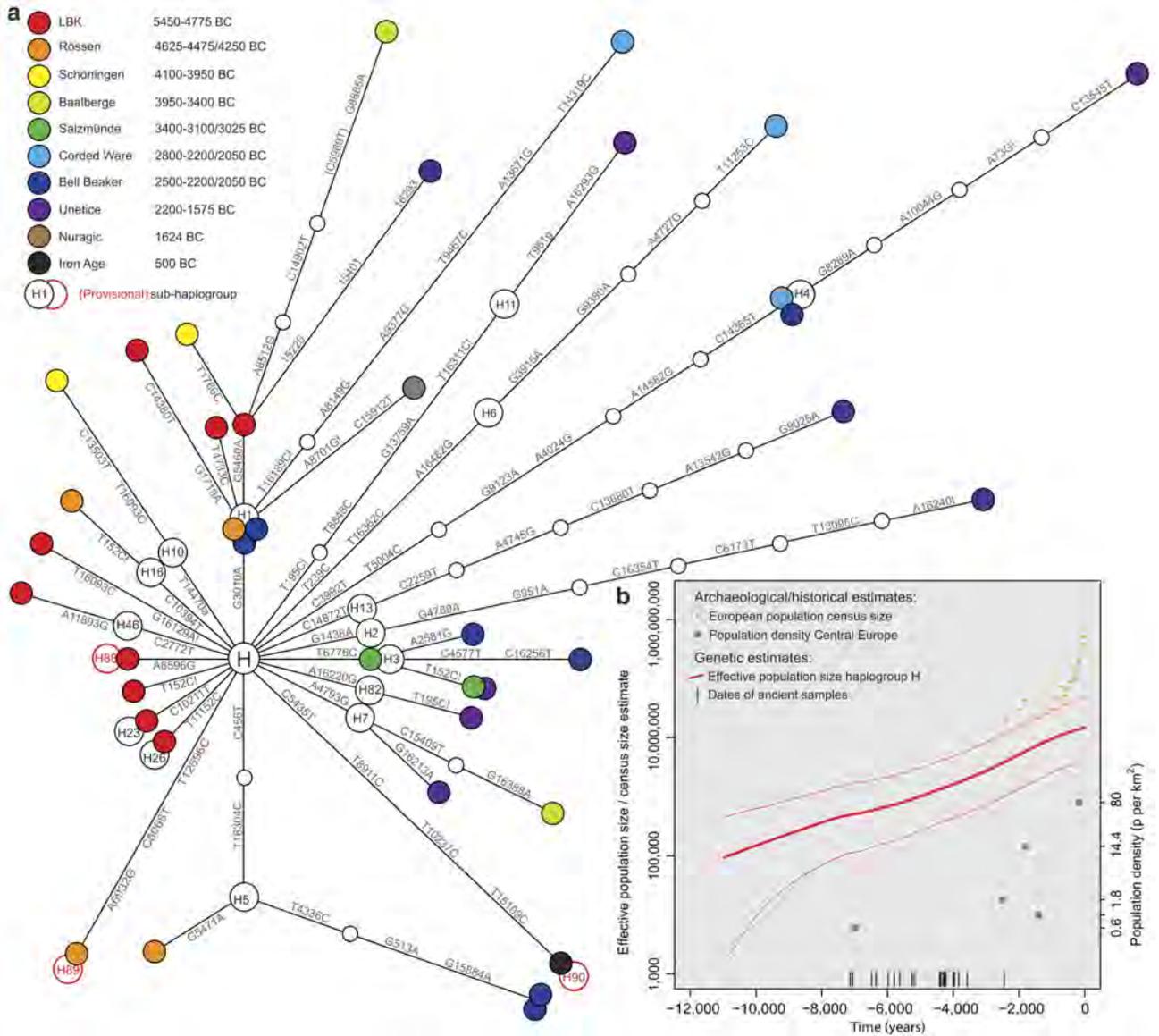


Figure 1. Mitochondrial haplogroup H sequence evolution

(a) Phylogenetic network of 39 prehistoric mitochondrial genomes sorted into two temporal groupings: Early Neolithic (left) and Mid-to-Late Neolithic (right). Node colours represent archaeological cultures. (b) A Bayesian skyride plot of 200 representative present-day and 39 ancient hg H mt genomes (the thick red line denotes the posterior median, thinner flanking lines denote the 95% credibility interval; note the logarithmic scale of the y-axis). Prehistoric samples (18 radiocarbon and 21 mean archaeological dates) served as internal calibration points (black bars). For comparison, census size estimates for the European population are shown as orange dots. Population density estimates from the archaeological record for key periods in Central Europe are plotted as blue squares in chronological order: LBK, Iron Age, Roman period, Merovingian, and Pre-industrial modern times (y-axis on the right)²⁸.

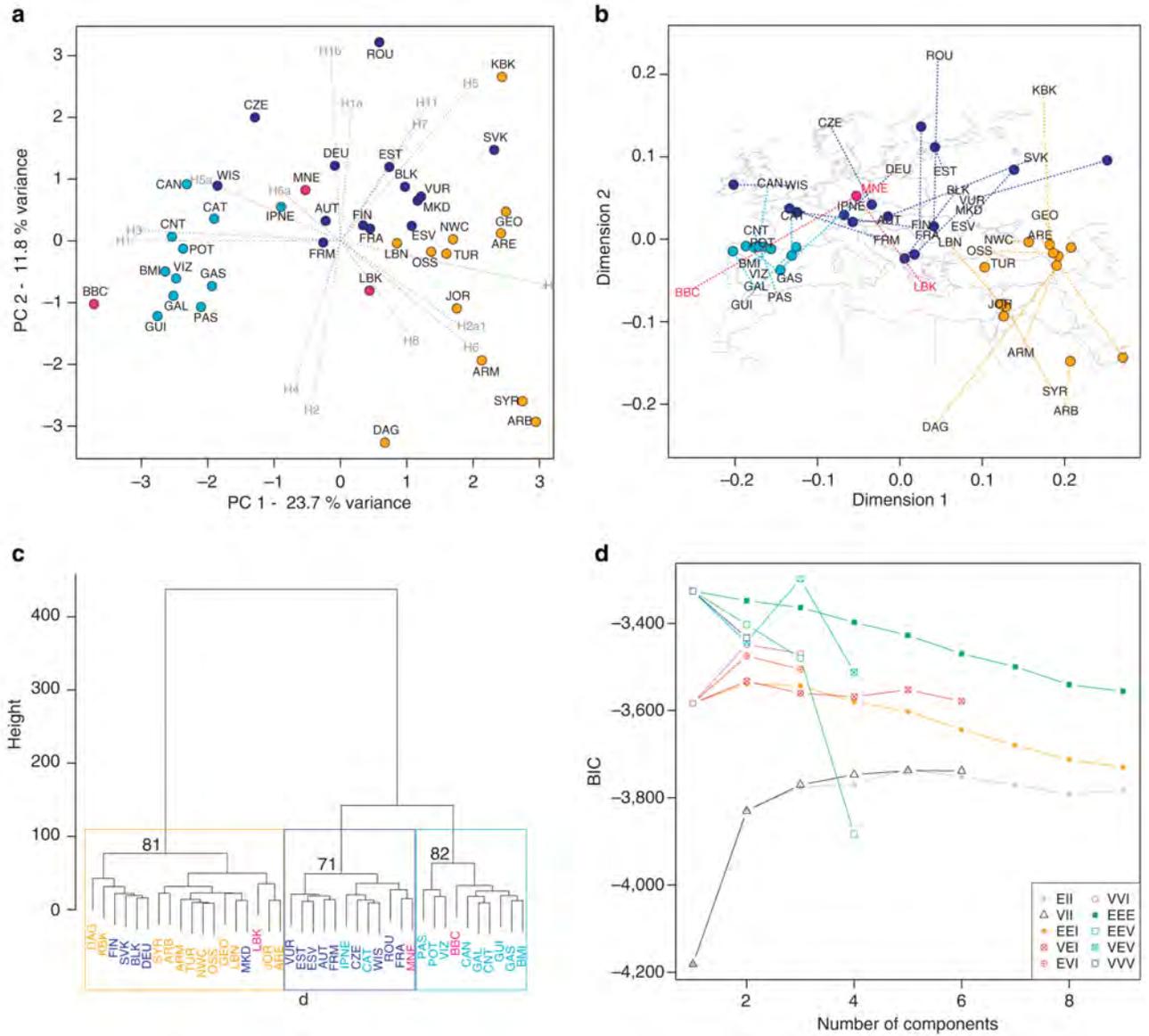


Figure 2. Population affinities of select Neolithic cultures

(a) PCA biplot based on the frequencies of 15 hg H sub-haplogroups (component loadings) from 37 present-day Western Eurasian and three ancient populations (light blue: Western Europe; dark blue: Central and Eastern Europe; orange; Near East, Caucasus, and Anatolia; and pink: ancient samples). Populations are abbreviated as follows: GAL, Galicia; CNT, Cantabria; CAT, Catalonia; GAS, Galicia/Asturia; CAN, Cantabria2; POT, Potes; PAS, Pasiegos; VIZ, Vizcaya; GUI, Guipuzcoa; BMI, Basques; IPNE, Iberian Peninsula Northeast; TUR, Turkey; ARM, Armenia; GEO, Georgia; NWC, Northwest Caucasus; DAG, Dagestan; OSS, Ossetia; SYR, Syria; LBN, Lebanon; JOR, Jordan; ARB, Arabian Peninsula; ARE, Arabian Peninsula2; KBK, Karachay-Balkaria; MKD, Macedonia; VUR, Volga-Ural region; FIN, Finland; EST, Estonia; ESV, Eastern Slavs; SVK, Slovakia; FRA, France; BLK, Balkans; DEU, Germany; AUT, Austria; ROU, Romania; FRM, France Normandy; WIS, Western Isles; CZE, Czech Republic; LBK, Linear pottery culture; BBC, Bell Beaker culture; MNE, Middle Neolithic. (b) Procrustes analyses of geographic coordinates and PCA scores of the same dataset (similarity score $t_0 = 0.733$, $p < 10^{-6}$, 100

000 permutations). **(c)** Ward clustering dendrogram of the three ancient groups and present-day populations (colour code as above and p values in % of approximately unbiased bootstrapping for the following three main clusters). **(d)** Results of the model-based test to identify the number of clusters by the model with the highest support (highest BIC; VEV= multivariate mixture model (ellipsoidal, equal shape)).

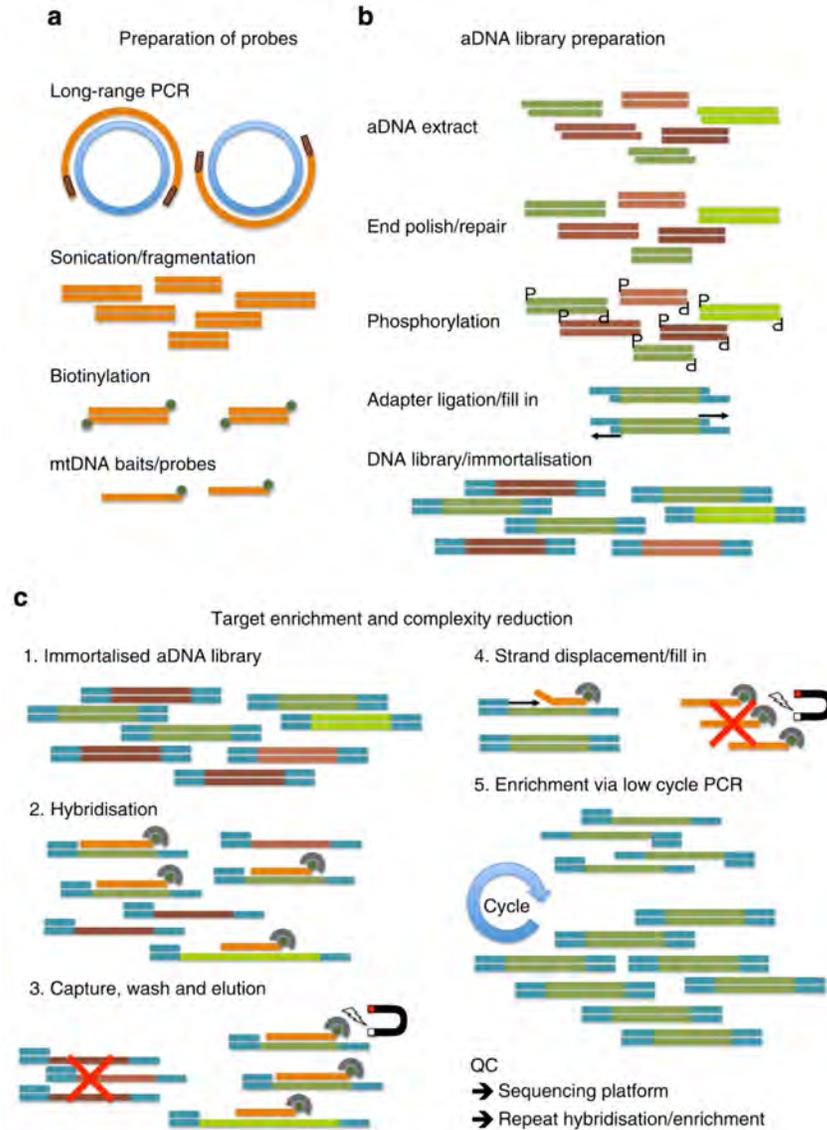


Figure 3. Schematic representation of experimental steps

(a) Probe DNA was prepared by amplifying a complete mitochondrial genome in two overlapping fragments by long-range PCR, followed by DNA fragmentation and biotinylation to form mtDNA ‘baits’ for targeted hybridisation. (b) Ancient DNA was enzymatically blunt-ended and phosphorylated, ligated to custom library adapters, followed by polymerase ‘fill-in’ to create ‘immortalised’ double-stranded DNA libraries. (c) Hybridisation-based DNA-capture using biotinylated probe bound to Streptavidin magnetic beads; following stringency washes, captured library constructs enriched in mtDNA sequences are eluted from the beads/probe via a novel polymerase strand-displacement reaction followed by PCR library reamplification. These steps can be carried out iteratively to maximise mtDNA content in enriched libraries (see Supplementary Methods for full details).

Table 1
Summary of genotyping data against the Reconstructed Sapiens Reference Sequence

Sub-haplogroup defining SNPs are shown in bold and ‘private’ / as-yet-unknown sequence variants in regular print.

Culture/Age	Individual	Hg ^I	Hg H sequence variants compared to RSRS
LBK (5450-4775 BC)	HAL36	H23	C10211T
	HAL11	H	T16093C, G16129A!
	HAL32	H26	T11152C
	HAL39	H1e	G3010A, G5460A
	DEB9	H88	A8596G
	DEB21	H1j	G3010A, T4733C
	KAR6a	H1bz	G1719A, G3010A , C14380T
	KAR11b	H	T152C!
	KAR16a	H46b	C2772T , A11893G
Rössen (4625-4475/4250 BC)	OSH2	H89	A6932G, C8068T, T12696C
	OSH3	H1	G3010A
	OSH1	H16	T152C! , C10394T
	OSH7	H5b	C456T, G5471A, T16304C , C16519T
Schöningen (4100-3950 BC)	SALZ18a	H10i	C13503T, T14470a, T16093C
	SALZ21b	H1e7	T1766C, G3010A, G5460A
Baalberge (3950-3400 BC)	ESP30	H1e1a5	G3010A, G5460A , (C5960T), A8512G , G8865A, C14902T
	HQU4	H7d5	A4793G, C15409T , G16388A
Salzmünde (3400-3100/3025 BC)	SALZ57a	H3	T152C!, T6776C
	SALZ77a	H3	T6776C
Corded Ware (2800-2200//2050 BC)	ESP15	H6a1a	T239C, G3915A, A4727G, G9380A, T11253C, T16362C, A16482G , C16519T
	BZH6	H1_TBD	G3010A , A8149G, A9377G, T9467C, A13671G, T14319C, T16189C!
Bell Beaker (2500-2200/2050 BC)	BZH4	H1e7	G3010A, G5460A , A15220G, A15401G, A16293G
	ROT6	H5a3	C456T, G513A, T4336C, G15884A, T16304C , C16519T
	ALB1	H3b	A2581G, T6776C
	ROT1	H3ao2	C4577T, T6776C, C16256T
	ROT2	H5a3	C456T, G513A, T4336C, G15884A, T16304C , C16519T
	QUEXIII	H4a1	C3992T, A4024G, T5004C, G9123A, C14365T, A14582G , C16519T
	QUEXII2	H4a1	C3992T, A4024G, T5004C, G9123A, C14365T, A14582G , C16519T
	QLB26a	H1	G3010A
	QUEXII3	H13a1a2c	C2259T, A4745G, G9025A, A13542G, C13680T, C14872T , C16519T
QLB28b	H1	G3010A	
Unetice (2200-1575 BC)	BZH1	H11a	T195C! , T961g, T8448C, (G13759A) , A16293G, T16311C! , C16519T
	BZH8	H2a1a3	G951A, G1438A, G4769A, C6173T, T13095C, A16240t, C16354T , C16519T
	BZH14	H82a	T195C!, A16220G

Culture/Age	Individual	Hg ^I	Hg H sequence variants compared to RSRS
	EUL41a	H4a1a1a5	A73G! , C3992T , A4024G , T5004C , G8269A , G9123A , A10044G , C13545T , C14365T , A14582G , C16519T
	EUL57B	H3	T152C! , T6776C
	QUEVIII4	H7h	A4793G , G16213A
Nuragic Bronze Age (1624 BC)	-	H1aw1	G3010A , A8701G! , C15912T
Iron Age (500 BC)	-	H90	C5435T , T8911C , T10237C , T15109C

^IHaplogroup H designations based on the <http://www.phylotree.org> mtDNA tree Build 14 (5 April 2012)^{19,20}.

Table 2
Population pairwise and linearised Slatkin's F_{ST} s and NP-MANOVA tests

Neolithic samples pooled in different time periods: a) four time periods; b) three times periods; c) two time periods, and d) from cultures grouped with geographic regions. For NP-MANOVA, Bonferroni corrected values are given in bold print and areas shaded grey indicate significant values ($p < 0.05$). Slatkin F_{ST} s are italicised (upper diagonal) and significant pairwise distances are given in bold print (lower diagonal).

a) NP-MANOVA Four time periods ($p=0.0696$)				
	Early Neolithic	Middle Neolithic	Late Neolithic	Bronze Age
Early Neolithic (13)	0	0.1262	0.024	0.0574
Middle Neolithic (6)	0.7572	0	0.8575	0.7782
Late Neolithic (9)	0.144	1	0	0.742
Bronze Age (6)	0.3444	1	1	0
F_{ST} Four time periods F_{ST}				
	Early Neolithic	Middle Neolithic	Late Neolithic	Bronze Age
Early Neolithic (13)	0	<i>0</i>	<i>0.03179</i>	<i>0.09555</i>
Middle Neolithic (6)	0.01135	0	<i>0</i>	<i>0.02299</i>
Late Neolithic (9)	0.02247	-0.01165	0	<i>0.01148</i>
Bronze Age (6)	0.08722	0.03081	-0.02250	0
b) NP-MANOVA LBK, BBC and pooled intermediate Neolithic (MNE) as used in PCA ($p=0.2355$)				
	LBK	MNE	BBC	
LBK (9)	0	0.2084	0.0916	
MNE (10)	0.6252	0	0.8025	
BBC (7)	0.2748	1	0	
F_{ST} LBK, BBC, and pooled intermediate Neolithic (MNE) as used in PCA				
	LBK	MNE	BBC	
LBK (9)	0	<i>0</i>	<i>0.03369</i>	
MNE (10)	-0.02587	0	<i>0</i>	
BBC (7)	0.03260	-0.00704	0	
c) NP-MANOVA Two time periods ($p=0.0072$)				
	Early Neolithic	Late Neolithic		
Early Neolithic (13)	0	0.0109		
Late Neolithic (16)	0.0109	0		
F_{ST} Two time periods				
	Early Neolithic	Late Neolithic		
Early Neolithic (13)	0	<i>0.01459</i>		
Late Neolithic (16)	0.01438	0		
d) NP-MANOVA Cultures grouped with geographic regions as in Figure 2 ($p < 0.0001$)				
	Iberia	Near East	Mainland Europe	
Iberia	0	0	0.0001	
Near East	0	0	0.0004	
Mainland Europe	0.0003	0.0012	0	