

ARTICLE

Received 20 Sep 2012 | Accepted 27 Feb 2013 | Published 16 Apr 2013

DOI: 10.1038/ncomms2656

# Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans

Paul Brotherton<sup>1,2,\*</sup>, Wolfgang Haak<sup>1,\*</sup>, Jennifer Templeton<sup>1</sup>, Guido Brandt<sup>3</sup>, Julien Soubrier<sup>1</sup>, Christina Jane Adler<sup>1,†</sup>, Stephen M. Richards<sup>1</sup>, Clio Der Sarkissian<sup>1,†</sup>, Robert Ganslmeier<sup>4</sup>, Susanne Friederich<sup>4</sup>, Veit Dresely<sup>4</sup>, Mannis van Oven<sup>5</sup>, Rosalie Kenyon<sup>6</sup>, Mark B. Van der Hoek<sup>6</sup>, Jonas Korfach<sup>7</sup>, Khai Luong<sup>7</sup>, Simon Y.W. Ho<sup>8</sup>, Lluís Quintana-Murci<sup>9</sup>, Doron M. Behar<sup>10</sup>, Harald Meller<sup>4</sup>, Kurt W. Alt<sup>3</sup>, Alan Cooper<sup>1</sup> & The Genographic Consortium<sup>‡</sup>

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.

<sup>1</sup>The Australian Centre for Ancient DNA, School of Earth and Environmental Sciences, University of Adelaide, Adelaide, South Australia 5005, Australia. <sup>2</sup>Archaeogenetics Research Group, School of Applied Sciences, University of Huddersfield, Huddersfield HD1 3DH, UK. <sup>3</sup>Institute of Anthropology, Colonel-Kleinmann Weg 2, Johannes Gutenberg University Mainz, D-55128 Mainz, Germany. <sup>4</sup>State Office for Heritage Management and Archaeology Saxony-Anhalt/State Museum for Prehistory Halle, Richard-Wagner-Straße 9, D-06114 Halle/Saale, Germany. <sup>5</sup>Department of Forensic Molecular Biology, Erasmus MC, University Medical Centre Rotterdam, 3000 CA Rotterdam, The Netherlands. <sup>6</sup>SA Pathology, SA Health, Adelaide, South Australia 5000, Australia. <sup>7</sup>Pacific Biosciences, Menlo Park, California 94025, USA. <sup>8</sup>School of Biological Sciences, The University of Sydney, Sydney, New South Wales 2006, Australia. <sup>9</sup>Institut Pasteur, Unit of Evolutionary Genetics, 75015 Paris, France. <sup>10</sup>Rambam Medical Centre, 31096 Haifa, Israel. \* These authors contributed equally to this work. † Present address: Institute of Dental Research, Westmead Centre for Oral Health, The University of Sydney, Sydney, New South Wales 2145, Australia (C.J.A.); Centre for Geogenetics, Natural History Museum of Denmark, 1350 Copenhagen, Denmark (C.D.S.). Correspondence and requests for materials should be addressed to P.B. (email: p.m.brotherton@hud.ac.uk) or to W.H. (email: wolfgang.haak@adelaide.edu.au). ‡A full list of authors for the Genographic Consortium and their affiliations appears at the end of the paper.

**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 (ENE), through to the socially stratified chiefdoms of the Early Bronze Age<sup>1–3</sup>. The exact nature and genetic context of the transformative changes that took place over these four millennia remain unclear<sup>4,5</sup>, although current genetic patterns of mitochondrial DNA (mtDNA) haplogroup (hg) distribution suggest a complex series of events in European prehistory<sup>4–9</sup> and hint at multiple inputs from outside Central Europe<sup>4,10,11</sup>.

Phylogeographic studies suggest that mt hg H arrived in Europe from the Near East before the Last Glacial Maximum (22,000 BP), and survived in glacial refugia in Southwest Europe before undergoing a post-glacial re-expansion<sup>4,12</sup>. 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 Caucasus<sup>10</sup>. However, it remains uncertain when and how H became the dominant European hg. Traditional approaches (including ancient DNA studies) have been unable to resolve either the phylogeny or phylogeographic distribution of H sub-haplogroups (sub-hgs)<sup>6</sup>, however, they have generally relied on sequencing only 300–400 bp of the mt D-loop or control region<sup>10,13</sup>. A number of studies based on complete 16.6 kb human mt genomes have revealed a complex evolutionary history for hg H (for example, refs 12,14–18, with phylogenetic analyses recognizing 87 H sub-hgs<sup>19</sup>). These complete mt genomes revealed that 71% of hg H polymorphic diversity is located outside the D-loop, in the coding region<sup>20</sup> 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. Owing 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 >3,500 years of the Central European Neolithic period (Table 1, Supplementary Table S1), from the first farmers of the ENE 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 (LNE) 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, because of 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

**Table 1 | Summary of genotyping data against the Reconstructed Sapiens Reference Sequence (RSRS).**

Culture/age	Individual	Hg*	Hg H sequence variants compared with RSRS
LBK (5450–4775 BC)	HAL36	H23	<b>C1021T</b>
	HAL11	H	T16093C, G16129A!
	HAL32	H26	<b>T1152C</b>
	HAL39	H1e	<b>G3010A, G5460A</b>
	DEB9	H88	<b>A8596G</b>
	DEB21	H1j	<b>G3010A, T4733C</b>
	KAR6a	H1bz	G1719A, <b>G3010A</b> , C14380T
	KAR11b	H	T152C!
	KAR16a	H46b	<b>C2772T</b> , A11893G
Rössen (4625–4475/4250 BC)	OSH2	H89	A6932G, C8068T, T12696C
	OSH3	H1	<b>G3010A</b>
	OSH1	H16	<b>T152C!, C10394T</b>
	OSH7	H5b	<b>C456T, G5471A, T16304C</b> , C16519T
Schöningen (4100–3950 BC)	SALZ18a	H10i	C13503T, <b>T14470a, T16093C</b>
	SALZ21b	H1e7	T1766C, <b>G3010A, G5460A</b>
Baalberge (3950–3400 BC)	ESP30	H1e1a5	<b>G3010A, G5460A</b> , (C5960T), <b>A8512G</b> , G8865A, <b>C14902T</b> , <b>A4793G, C15409T</b> , G16388A
	HQU4	H7d5	
Salzmünde (3400–3100/3025 BC)	SALZ57a	H3	T152C!, <b>T6776C</b>
	SALZ77a	H3	<b>T6776C</b>
Corded Ware (2800–2200/2050 BC)	ESP15	H6a1a	<b>T239C, G3915A, A4727G, G9380A, T11253C, T16362C, A16482G</b> , C16519T
	BZH6	H1_TBD	<b>G3010A</b> , A8149G, A9377G, T9467C, A13671G, T14319C, <b>T16189C!</b>
Bell Beaker (2500–2200/2050 BC)	BZH4	H1e7	<b>G3010A, G5460A</b> , A15220G, A15401G, A16293G
	ROT6	H5a3	<b>C456T, G513A, T4336C, G15884A, T16304C</b> , C16519T
	ALB1	H3b	<b>A2581G, T6776C</b>
	ROT1	H3aa2	C4577T, <b>T6776C, C16256T</b>
	ROT2	H5a3	<b>C456T, G513A, T4336C, G15884A, T16304C</b> , C16519T
	QUEXII1	H4a1	<b>C3992T, A4024G, T5004C, G9123A, C14365T, A14582G</b> , C16519T
	QUEXII2	H4a1	<b>C3992T, A4024G, T5004C, G9123A, C14365T, A14582G</b> , C16519T
Unetice (2200–1575 BC)	QLB26a	H1	<b>G3010A</b>
	QUEXII3	H13a1a2c	<b>C2259T, A4745G, G9025A, A13542G, C13680T, C14872T</b> , C16519T
	QLB28b	H1	<b>G3010A</b>
	BZH1	H11a	<b>T195C!, T961G, T8448C, (G13759A), A16293G, T16311C!</b> , C16519T
Unetice (2200–1575 BC)	BZH8	H2a1a3	<b>G951A, G1438A, G4769A, C6173T, T13095C</b> , A16240t, <b>C16354T</b> , C16519T
	BZH14	H82a	T195C!, <b>A16220G</b>
	EUL41a	H4a1a1a5	<b>A73G!, C3992T, A4024G, T5004C, G8269A, G9123A, A10044G</b> , C13545T, <b>C14365T, A14582G</b> , C16519T
	EUL57B	H3	T152C!, <b>T6776C</b>
Unetice (2200–1575 BC)	QUEVII4	H7h	<b>A4793G, G16213A</b>
Nuragic Bronze Age (1624 BC)	—	H1aw1	<b>G3010A, A8701G!</b> , C15912T
Iron Age (500 BC)	—	H90	C5435T, T8911C, T10237C, T15109C

Abbreviation: SNP, single-nucleotide polymorphism.

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

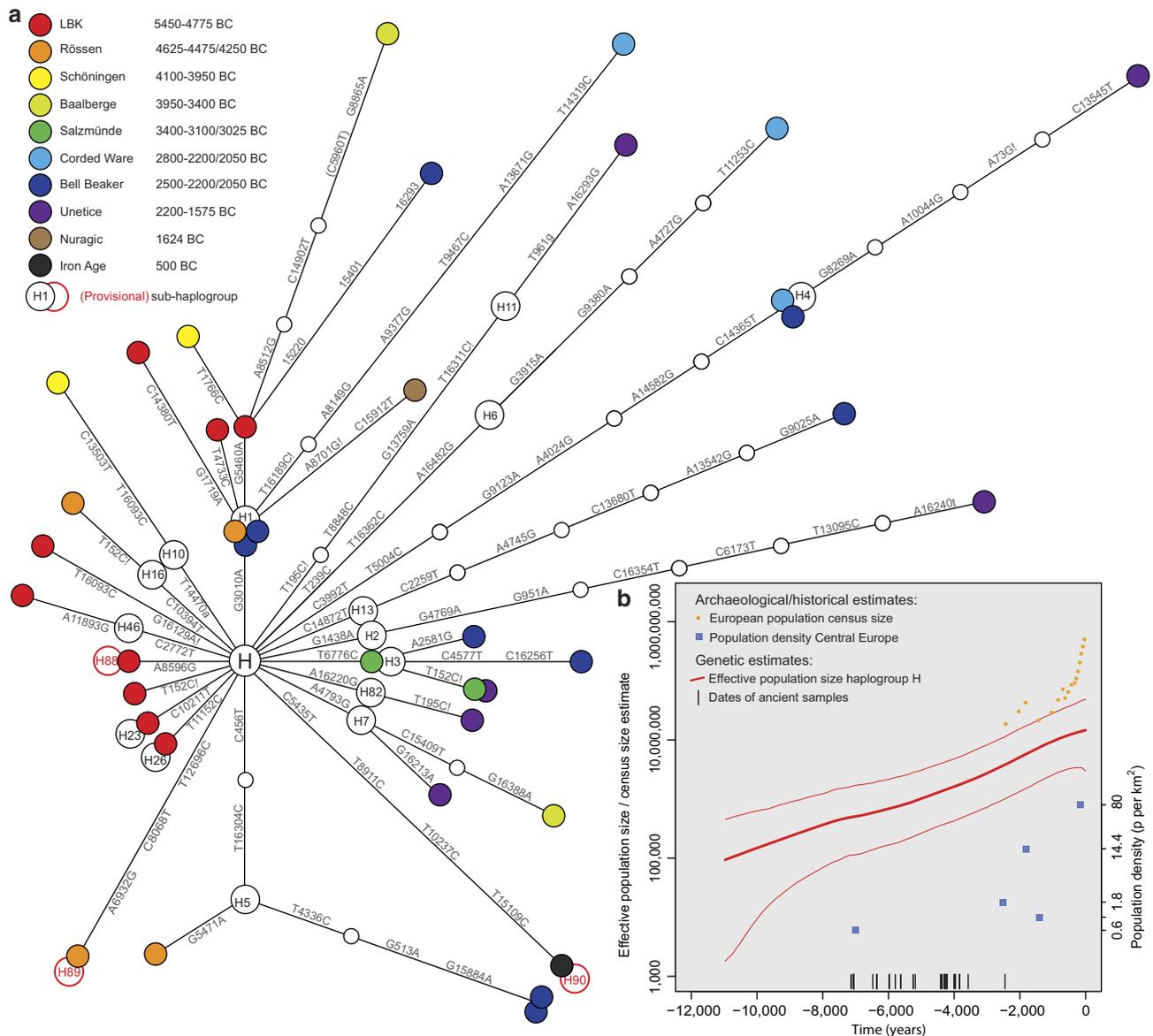
\*Haplogroup H designations based on the <http://www.phylotree.org/mtDNA/tree/Build14> (5 April 2012)<sup>19,20</sup>.

37 Mittelbe-Saale individuals, as well as two samples from Italy (Supplementary Table S1), previously assigned to hg H by simplex and multiplex PCR<sup>7</sup>. 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 v2.0 (ref. 21) (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<sup>22,23</sup>. Six of the 39 target-enriched libraries were also analysed via a single-molecule, real-time (SMRT<sup>24</sup>) 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 H<sup>20</sup>,

confirming that a single human was typed in each case (Table 1). The mt 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  $0.000421 \pm 0.000225$ ), 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 ~3,500-year time transect (Fig. 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



**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)<sup>28</sup>.

branches. This temporal relationship provides further support for the authenticity of the ancient mt genomes.

Network analysis (Fig. 1) reveals pronounced differences in the composition of sub-hgs between the ENE cultures (LBK, Rössen, Schöningen), and those of the Mid Neolithic (MNE)/LNE to Early Bronze Age (Baalberge, Salzmünde, Corded Ware, Bell Beaker, Unetice). ENE (and in particular LBK) mt genomes are either rare today (H16, H23 and H26), extinct or have not yet been observed in present-day populations (H46b, H88 and H89). In sharp contrast, most of the later H sub-hgs are more common in present-day European populations (for example, hg H3, H4, H6, H7, H11 and H13)<sup>12,14–16</sup>. Of the 39 haplotypes detected, only three (within the common, basal, sub-hg H1) were shared between ENE and MNE/LNE cultures. As the observed gene diversity is high, we might expect the number of shared haplotypes within and between cultures to be low. However, as 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 ENE versus LNE (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 ( $F_{ST}=0.08722$ ;  $P=0.00386 + / - 0.0006$ ) between the ENE 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 and BBC) and 37 present-day Western Eurasian populations (Supplementary Table S8). PCA of the frequencies of the 15 most common H sub-hgs 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 2: (i) Iberia in the west; (ii) the Caucasus, the Near East and Anatolia; and (iii) Central and Eastern Europe from the Urals to France (Fig. 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 (Fig. 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 coordinates 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

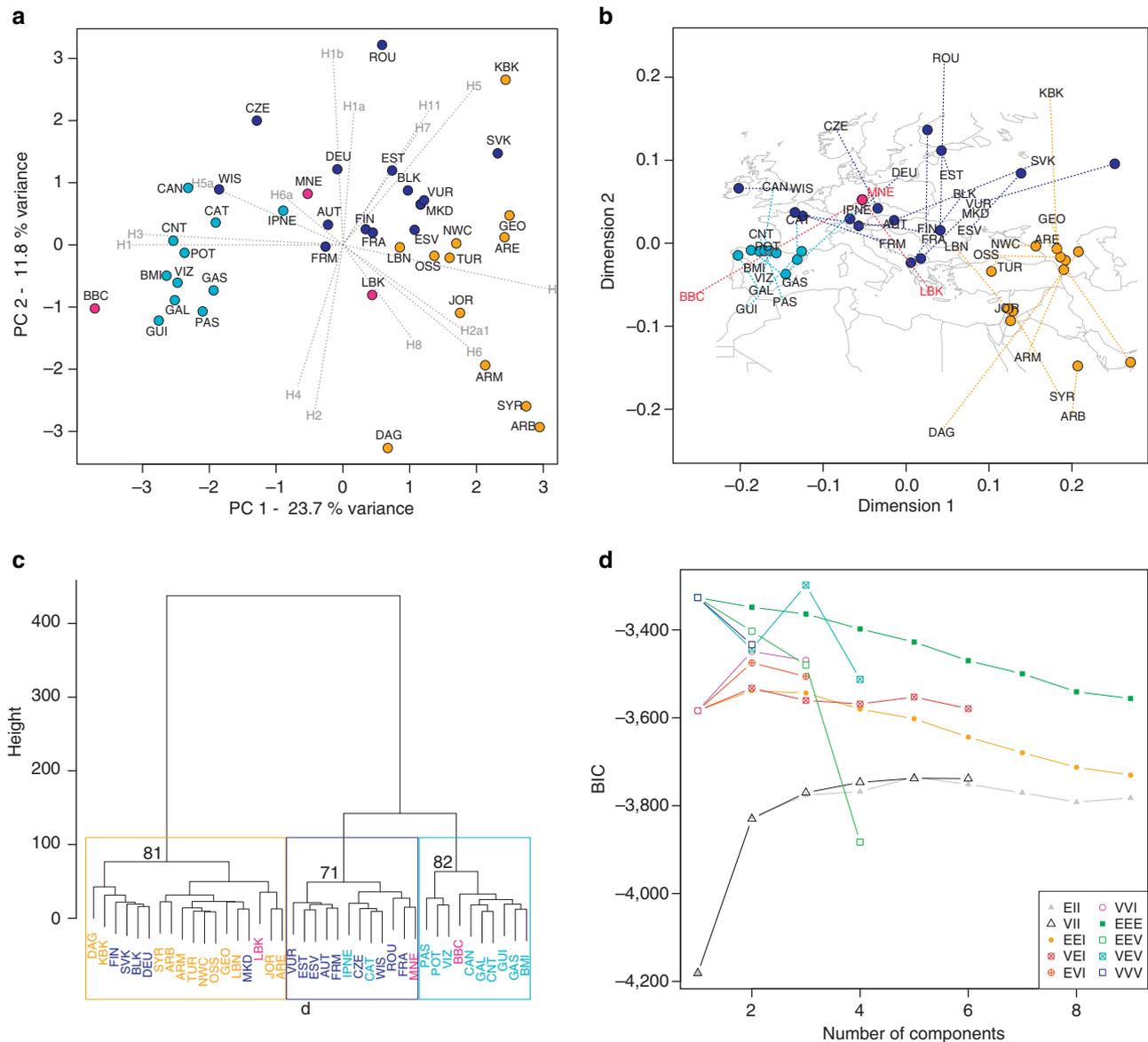
**Table 2 | Population pairwise and linearised Slatkin's  $F_{ST}$  and NP-MANOVA tests.**

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

Abbreviations: BBC, Bell Beaker culture; MNE, Mid Neolithic; NP-MANOVA, non-parametric multivariate analysis of variance; PCA, principal component analysis. Neolithic samples pooled in different time periods: (a) four time periods; (b) three time 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).

described above (Fig. 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 noted<sup>7</sup>. In contrast, individuals from the successor series of regional post-LBK (and MNE) Rössen, Schöningen, Baalberge and Salzmünde cultures (ca. 4625–3025 BC, MNE;  $n=10$ ) cluster with present-day Central European populations (Fig. 2). Mitochondrial genomes from BBC individuals in Mittelbe-Saale (BBC;  $n=7$ ) display close genetic affinities to present-day Iberian populations (Fig. 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 refugium<sup>13</sup> and which have also been reported from ancient Neolithic sites from France and Spain<sup>8,25</sup>. Other LNE samples add further to the genetic complexity. Individuals from the 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



**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 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 data set (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 Bayes Information Criterion (BIC); VEV = multivariate mixture model (ellipsoidal, equal shape)).

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-hgs in present-day populations<sup>13</sup>. 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 hg in Europe, as archaeogenetic and palaeodemographic reconstructions have very large uncertainties<sup>4,26</sup>. 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 (Fig. 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 ~3,500 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 size<sup>27</sup> and population density estimates from archaeological sites<sup>28</sup> in the Late Holocene, but also provides detailed estimates for prehistoric times for which data points remain very scarce (Fig. 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 years<sup>29</sup>. The rate calibrated by the Neolithic and Bronze Age sequences is  $2.4 \times 10^{-8}$  substitutions per site per year ( $1.7\text{--}3.2 \times 10^{-8}$ ; 95% high posterior density) for the entire mt genome, which is  $1.45 \times$  (44.5%) higher than current estimates based on the traditional human/chimp split (for example,  $1.66 \times 10^{-8}$  for the entire mt genome<sup>30</sup> and  $1.26 \times 10^{-8}$  for the coding region<sup>31</sup>). Consequently, the calibrated 'Neolithic' rate infers a considerably younger coalescence date for hg H (10.9–19.1 kya) than those previously reported (19.2–21.4 kya for HVSI<sup>10</sup>, 15.7–22.5 kya for the mt coding region<sup>31</sup> or 14.7–22.6 kya when corrected for purifying selection<sup>30</sup>).

## Discussion

Despite recent successes in sequencing portions of nuclear genomes from Meso- and Neolithic samples<sup>11,32,33</sup>, mtDNA remains the most widely studied and best described marker in population genetics. Although its interpretation is limited to the matrilineal genetic history<sup>4,13,19</sup>, this can be an important socio-cultural and demographic signal additional to that gained from autosomal loci<sup>34,35</sup>. 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 data set (Fig. 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. Although a temporal pattern could be expected in an expanding population with stable/increasing hg H frequencies (Fig. 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 H<sup>14,16</sup>. Under an assumption of genetic continuity, we would expect MNE/LNE and Bronze Age individuals to be on the same sub-hg branches as ENE individuals. Instead, ENE mt genomes are generally either rare today<sup>19</sup> or have not yet been observed in present-day populations, possibly owing 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 ENE made a marginal contribution to LNE 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 ENE and the MNE/LNE in Central Europe (Fig. 1; Table 1), again suggesting that ENE lineages were largely superseded during the MNE/LNE (~4100–2200 BC) in a previously unrecognised major genetic transition. This pronounced genetic changeover between ENE and MNE/LNE 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)<sup>6,7</sup>. When compared with hg H diversity in present-day Central Europe<sup>14,15,18,36</sup>, the network in Fig. 1 suggests that much of the present-day diversity can be attributed to the incorporation of new lineages in the MNE/LNE and emerging Bronze Age (from 2200 BC). The LNE in particular is known to have been a period of profound cultural and economic change<sup>37</sup>, 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<sup>7</sup>.

Our observation that individuals from the successor series of regional post-LBK and MNE cultures (Rössen, Schöningen, Baalberge and Salzmünde) cluster with present-day Central European populations could be explained by a loss of lineages from the ENE LBK period during a short phase of population decline in the centuries after 5000 BC (as proposed in some archaeological models)<sup>38</sup>. However, our results suggest that mtDNA H sub-hg diversity established during the MNE 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 MNE cultures outside Mittelbe-Saale, which (perhaps in concert with LBK population decline) could have resulted in a replacement of most ENE 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 LNE 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 (Fig. 2) is striking and throws fresh light on long-disputed archaeological models<sup>3</sup>. We suggest these data indicate a considerable genetic influx from the West during the LNE. 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 Europe<sup>39</sup>. This hypothesis suggests that early members of the Celtic language family (for example, Tartessian)<sup>40</sup> 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 East<sup>41</sup>.

Other LNE population movements appear to have added further genetic complexity, as exemplified by the 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<sup>37,42</sup>, marks a consolidation of social and cultural systems in Mittelbe-Saale that were established during the LNE by the two pan-European Bell Beaker and CWCs. The Unetice culture appears contemporaneously with the last Neolithic horizon (~2200 BC) in areas where elements of both the Bell Beaker and CWCs 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 before the Holocene (12 kya)<sup>13</sup>. Our data complement 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 day<sup>26</sup>. 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<sup>12,16</sup> (as indicated by our population skyride 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 (Fig. 1b).

Support for this position comes from data suggesting that hg H was virtually absent among Central and Northern European hunter-gatherers<sup>6,43</sup> and formed only 19% in LBK individuals, most likely introduced from Southeast Europe and/or the Near East<sup>7</sup>. In our updated data set 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, MNE/LNE cultures with cultural associations to the North and Northeast, such as the Bernburg and CWCs, 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<sup>11,43</sup>. 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 France<sup>9</sup>; at ~25% from Iberian (Epi-)Cardial Neolithic samples<sup>8,25</sup>; at 36% from a Neolithic site in Catalonia<sup>44</sup>; and at 44% from Neolithic sites from the Basque Country and Navarre<sup>45</sup>. Importantly, a recent study on Iberian hunter-gatherers revealed the presence of hg H there in Mesolithic times<sup>45</sup>. 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 today<sup>16</sup>. Having reached significant levels, and assuming a generally higher rate of population growth in southern and western Europe in post-Neolithic times<sup>27</sup>, these Neolithic processes appear to have been the major factor in hg H becoming the predominant European mtDNA hg.

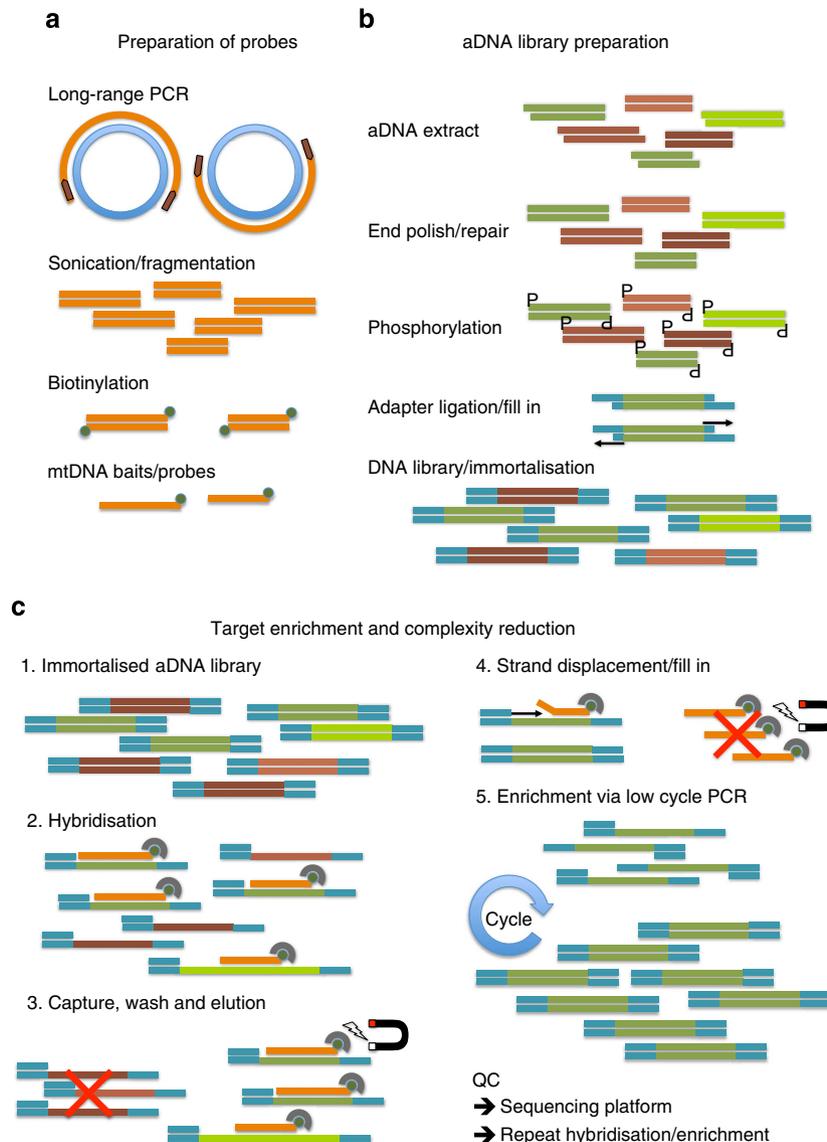
Overall, our results suggest that the broad foundations of the Central European mtDNA pool, here approximated via hg H, were formed during the Neolithic rather than the post-glacial period. ENE hg H mt lineages brought in from the Near East by Central Europe's first farmers do not appear to have contributed significantly to present-day Central Europe's hg H diversity, instead being largely superseded during the MNE and LNE (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 MNE, 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, LNE 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<sup>7,46</sup>. Multiplex SNP typing of 22 hg informative SNPs (GenoCoRe22) was carried out using a SNaPshot-based protocol as described previously<sup>7</sup>.

**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 (Fig. 3). Following every step, DNA was purified using MinElute spin columns (Qiagen) as per the



**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 adaptors, 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).

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 mt probe) was previously described<sup>47</sup> and the overall scheme is outlined in Fig. 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 molecules<sup>48</sup>. 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 mt 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 mt Resequencing 2.0 Arrays for 17 h 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 Geneious<sup>49</sup> (Supplementary Fig. S1,S2, Supplementary Tables S2,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 (Fig. 1a) was constructed manually using the most up-to-date version of the mt phylogenetic tree (PhyloTree.org, mtDNA tree Build 14) as a scaffold on which to place the observed hg H lineages<sup>19,20</sup>. 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 reconstruction<sup>20</sup>.

**Procrustes-based PCA and Ward Clustering.** PCA was used to describe and visualise the maternal genetic relationships among the Neolithic cultures investigated, as well as to 37 present-day European and Near Eastern populations (Fig. 2a). PCA was performed on the frequency of H sub-hgs taken from the literature (Supplementary Table S8). To minimise statistical noise caused by rare sub-hgs and to allow for data compatibility across published studies, we considered only the following 15 most common H sub-hgs 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 2.11.1 (ref. 50) 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 MNE) 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) followed by Ward hierarchical clustering of sub-hg H frequencies using the packages `mclust`, `pvcust` (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 3.5 (ref. 51). We used `jMODELTEST 0.1.1`<sup>52</sup> 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 Bayes Information criterion), we subsequently used the Tamura and 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 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-hgs 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 sizes<sup>53</sup>. 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 sequences<sup>17</sup>. The Sequences were manually aligned to the revised Cambridge Reference Sequence (rCRS: AC\_000001)<sup>54</sup> using the program `SeaView`<sup>55</sup>. The alignment was partitioned into four subsets, representing the D-loop, the protein-coding regions (1st + 2nd codon positions and 3rd codon position) and a concatenation of tRNA and rRNA 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 subset<sup>20</sup>. The best substitution models were selected using `ModelGenerator 0.85` (ref. 56). By comparison of Bayesian Information Criterion scores: HKY + G for D-loop, TN + G for protein-coding regions and HKY for rRNA genes. Considering the short evolutionary timescale being studied (intra-hg), models including a proportion of invariant sites were excluded. A Bayesian skyride analysis<sup>57</sup> was performed using the phylogenetic software `BEAST 1.6.1` (ref. 58). 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 (Fig. 1b)<sup>59</sup>.

We carried out a 'date randomisation test', to test whether the signal from the radiocarbon dates associated with the ancient sequences was sufficient to calibrate the hg H phylogeny<sup>60</sup>. 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 10 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.

## References

- Whittle, A. W. R. & Cummings, V. *Going over: The Mesolithic-Neolithic Transition in North-West Europe* 632 (Oxford University Press, Oxford, 2007).
- Sherratt, A. Plough and pastoralism: aspects of the secondary products revolution. *Patterns of the Past: Studies in honour of David Clarke*. In: Hodder, I., Isaac, G. & Hammond, N. (eds). 261–305 (Cambridge University Press, Cambridge, 1981).
- Bogucki, P. I. & Crabtree, P. J. *Ancient Europe 8000 B.C.-A.D. 1000: Encyclopedia of the Barbarian World*, 1221p (Charles Scribner's Sons, 2004).
- Soares, P. *et al.* The archaeogenetics of Europe. *Curr. Biol.* **20**, R174–R183 (2010).
- Pinhasi, R., Thomas, M. G., Hofreiter, M., Currat, M. & Burger, J. The genetic history of Europeans. *Trends genet.* **28**, 496–505 (2012).
- Bramanti, B. *et al.* Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* **326**, 137–140 (2009).
- Haak, W. *et al.* Ancient DNA from European Early Neolithic Farmers Reveals Their Near Eastern Affinities. *PLoS Biol.* **8**, e1000536 (2010).
- Gamba, C. *et al.* Ancient DNA from an Early Neolithic Iberian population supports a pioneer colonization by first farmers. *Mol. Ecol.* **21**, 45–56 (2012).
- Lacan, M. *et al.* Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. *Proc. Natl Acad. Sci. USA* **108**, 9788–9791 (2011).
- Richards, M. *et al.* Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* **67**, 1251–1276 (2000).
- Skoglund, P. *et al.* Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336**, 466–469 (2012).
- Pereira, L. *et al.* High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res.* **15**, 19–24 (2005).
- Torroni, A., Achilli, A., Macaulay, V., Richards, M. & Bandelt, H. J. Harvesting the fruit of the human mtDNA tree. *Trends Genet.* **22**, 339–345 (2006).
- 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. *Mol. Biol. Evol.* **24**, 436–448 (2007).
- Loogväli, E. L. *et al.* Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol. Biol. Evol.* **21**, 2012–2021 (2004).
- 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. *Am. J. Hum. Genet.* **75**, 910–918 (2004).
- Behar, D. M. *et al.* The Basque paradigm: genetic evidence of a maternal continuity in the Franco-Cantabrian region since Pre-Neolithic times. *Am. J. Hum. Genet.* **90**, 486–493 (2012).
- Alvarez-Iglesias, V. *et al.* New population and phylogenetic features of the internal variation within mitochondrial DNA macro-haplogroup R0. *PLoS ONE* **4**, e5112 (2009).
- Behar, D. M. *et al.* A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012).
- van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–E394 (2009).
- Hartmann, A. *et al.* Validation of microarray-based resequencing of 93 worldwide mitochondrial genomes. *Hum. Mutat.* **30**, 115–122 (2009).
- Maricic, T., Whitten, M. & Paabo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* **5**, e14004 (2010).

23. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
24. Korlach, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Methods Enzymol.* **472**, 431–455 (2010).
25. Lacan, M. *et al.* Ancient DNA suggests the leading role played by men in the Neolithic dissemination. *Proc. Natl Acad. Sci. USA* **108**, 18255–18259 (2011).
26. Fu, Q., Rudan, P., Pääbo, S. & Krause, J. Complete mitochondrial genomes reveal Neolithic Expansion into Europe. *PLoS ONE* **7**, e32473 (2012).
27. Livi-Bacci, M. *A Concise History of World Population 279* (Blackwell Publishing, Malden, Oxford, Carlton, 2007).
28. Zimmermann, A., Hilpert, J. & Wendt, K. P. Estimations of population density for selected periods between the Neolithic and AD 1800. *Hum. Biol.* **81**, 357–380 (2009).
29. Ho, S. Y., Shapiro, B., Phillips, M. J., Cooper, A. & Drummond, A. J. Evidence for time dependency of molecular rate estimates. *Syst. Biol.* **56**, 515–522 (2007).
30. Soares, P. *et al.* Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* **84**, 740–759 (2009).
31. Mishmar, D. *et al.* Natural selection shaped regional mtDNA variation in humans. *Proc. Natl Acad. Sci. USA* **100**, 171–176 (2003).
32. Keller, A. *et al.* New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat. Commun.* **3**, 698 (2012).
33. Sanchez-Quinto, F. *et al.* Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. *Curr. Biol.* **22**, 1494–1499 (2012).
34. Behar, D. M. *et al.* The Basque paradigm: genetic evidence of a maternal continuity in the Franco-Cantabrian region since pre-Neolithic times. *Am. J. Hum. Genet.* **90**, 486–493 (2012).
35. Gunnarsdottir, E. D. *et al.* Larger mitochondrial DNA than Y-chromosome differences between matrilineal and patrilineal groups from Sumatra. *Nat. Commun.* **2**, 228 (2011).
36. Brandstätter, A. *et al.* Dissection of mitochondrial superhaplogroup H using coding region SNPs. *Electrophoresis* **27**, 2541–2550 (2006).
37. Heyd, V. Families, prestige goods, warriors & complex societies: Beaker groups of the 3rd millennium cal BC along the upper & middle Danube. *Proc. Prehist. Soc.* **73**, 327–379 (2007).
38. Shennan, S. & Edinborough, K. Prehistoric population history: from the late glacial to the late neolithic in central and northern Europe. *J. Archaeol. Sci.* **34**, 1339–1345 (2007).
39. Cunliffe, B. & Koch, J. T. (eds) *Celtic from the West: Alternative Perspectives from Archaeology, Genetics, Language and Literature*, 384 (Oxbow Books, Oxford, 2010).
40. Koch, J. T. *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*. XIV, 346 S. Ill., Kt(Cape, London, 1988).
42. Nowak, M. Transformations in East-Central Europe from 6000 to 3000 BC: local vs. foreign patterns. *Documenta Praehistorica XXXIII*, 143–158 (2006).
43. Malmström, H. *et al.* Ancient DNA reveals lack of continuity between neolithic hunter-gatherers and contemporary Scandinavians. *Curr. Biol.* **19**, 1758–1762 (2009).
44. Sampietro, M. L. *et al.* Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proc. Biol. Sci./Royal Soc.* **274**, 2161–2167 (2007).
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* **7**, e34417 (2012).
46. Haak, W. *et al.* Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* **310**, 1016–1018 (2005).
47. Patel, M. & Sive, H. PCR-based subtractive cDNA cloning. *Curr. Protocols Mol. Biol.* Chapter 25, Unit 25B, 2 (2001).
48. Tao, S. C., Gao, H. F., Cao, F., Ma, X. M. & Cheng, J. Blocking oligo—a novel approach for improving chip-based DNA hybridization efficiency. *Mol. Cell. Probes* **17**, 197–202 (2003).
49. Drummond, A. J. *et al.* Geneious v5.4. Available from <http://www.geneious.com/> (2011).
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, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resources* **10**, 564–567 (2010).
52. Posada, D. jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253–1256 (2008).
53. Hammer, O., Harper, D. A. T. & Ryan, P. D. PAST: paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* **4** (2001).
54. Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147 (1999).
55. Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224 (2010).
56. Keane, T. M., Creevey, C. J., Pentony, M. M., Naughton, T. J. & McLnerney, J. O. 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 Evol. Biol.* **6**, 29 (2006).
57. Minin, V. N., Bloomquist, E. W. & Suchard, M. A. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol. Biol. Evol.* **25**, 1459–1471 (2008).
58. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
59. Rambaut, A. & Drummond, A. Tracer v1.4. Available from <http://beast.bio.ed.ac.uk/Tracer> (2007).
60. Ho, S. Y. *et al.* Bayesian estimation of substitution rates from ancient DNA sequences with low information content. *Syst. Biol.* **60**, 366–375 (2011).

## Acknowledgements

We are indebted to Matt Kaplan and Ryan Spriggs at Arizona Research Laboratories, Division of Biotechnology, University of Arizona Genetics Core Facility, <http://uagc.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 (Al 287/7-1 and Me 3245/1-1) and National Geographic's Geographic 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).

## 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, radio-carbon 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., J.K. and members of the Geographic Consortium. All authors discussed the paper and gave comments.

## Additional information

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

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors claim no competing financial interests associated with this paper.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Brotherton, P. *et al.* Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nat. Commun.* **4**:1719 doi: 10.1038/ncomms2656 (2013).

Syama Adhikarla<sup>11</sup>, Arun Kumar Ganesh Prasad<sup>11</sup>, Ramasamy Pitchappan<sup>11</sup>, Arun Varatharajan Santhakumari<sup>11</sup>, Elena Balanovska<sup>12</sup>, Oleg Balanovsky<sup>12</sup>, Jaime Bertranpetit<sup>13</sup>, David Comas<sup>13</sup>, Begoña Martínez-Cruz<sup>13</sup>, Marta Melé<sup>13</sup>, Andrew C. Clarke<sup>14</sup>, Elizabeth A. Matisoo-Smith<sup>14</sup>, Matthew C. Dulik<sup>15</sup>, Jill B. Gaieski<sup>15</sup>,

Amanda C. Owings<sup>15</sup>, Theodore G. Schurr<sup>15</sup>, Miguel G. Vilar<sup>15</sup>, Angela Hobbs<sup>16</sup>, Himla Soodyall<sup>16</sup>, Asif Javed<sup>17</sup>, Laxmi Parida<sup>17</sup>, Daniel E. Platt<sup>17</sup>, Ajay K. Royyuru<sup>17</sup>, Li Jin<sup>18</sup>, Shilin Li<sup>18</sup>, Matthew E. Kaplan<sup>19</sup>, Nirav C. Merchant<sup>19</sup>, R. John Mitchell<sup>20</sup>, Colin Renfrew<sup>21</sup>, Daniela R. Lacerda<sup>22</sup>, Fabrício R. Santos<sup>22</sup>, David F. Soria Hernanz<sup>23</sup>, R. Spencer Wells<sup>23</sup>, Pandikumar Swamikrishnan<sup>24</sup>, Chris Tyler-Smith<sup>25</sup>, Pedro Paulo Vieira<sup>26</sup> & Janet S. Ziegler<sup>27</sup>

<sup>11</sup> The Genographic Laboratory, School of Biological Sciences, Madurai Kamaraj University, Madurai 625 021, Tamil Nadu, India. <sup>12</sup> Research Centre for Medical Genetics, Russian Academy of Medical Sciences, 115478 Moscow, Russia. <sup>13</sup> Institut de Biologia Evolutiva (CSIC-UPF), Departament de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, 08003 Barcelona, Spain. <sup>14</sup> Department of Anatomy, University of Otago, Dunedin 9054, New Zealand. <sup>15</sup> Department of Anthropology, University of Pennsylvania, Philadelphia, Pennsylvania, 19104-6398, USA. <sup>16</sup> National Health Laboratory Service, Sandringham 2131, Johannesburg, South Africa. <sup>17</sup> IBM, Yorktown Heights, New York 10598, USA. <sup>18</sup> School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai 200433, People's Republic of China. <sup>19</sup> Arizona Research Laboratories, University of Arizona, Tucson, Arizona 85721, USA. <sup>20</sup> Department of Genetics, School of Molecular Sciences, La Trobe University, Melbourne, Victoria 3086, Australia. <sup>21</sup> McDonald Institute for Archaeological Research, University of Cambridge, Cambridge, CB2 3ER, UK. <sup>22</sup> Departamento de Biologia Geral, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, CEP 31270-901, Brazil. <sup>23</sup> National Geographic Society, Washington, District of Columbia 20036-4688, USA. <sup>24</sup> IBM, Somers, New York 10589, USA. <sup>25</sup> Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK. <sup>26</sup> Universidade Federal do Rio de Janeiro, Rio de Janeiro, CEP 21941-901, Brazil. <sup>27</sup> Applied Biosystems, Foster City, California 94494, USA