



## Supplementary Materials for

### **Ancient DNA Reveals Key Stages in the Formation of Central European Mitochondrial Genetic Diversity**

Guido Brandt,\* Wolfgang Haak,\* Christina J. Adler, Christina Roth, Anna Szécsényi-Nagy, Sarah Karimnia, Sabine Möller-Rieker, Harald Meller, Robert Ganslmeier, Susanne Friederich, Veit Dresely, Nicole Nicklisch, Joseph K. Pickrell, Frank Sirocko, David Reich, Alan Cooper, Kurt W. Alt, The Genographic Consortium

\*Corresponding author. E-mail: [brandtg@uni-mainz.de](mailto:brandtg@uni-mainz.de) (G.B.); [wolfgang.haak@adelaide.edu.au](mailto:wolfgang.haak@adelaide.edu.au) (W.H.)

Published 11 October 2013, *Science* **342**, 257 (2013)  
DOI: 10.1126/science.1241844

#### **This PDF file includes:**

Materials and Methods  
Supplementary Text  
Figs. S1 to S10  
References

**Other Supplementary Material for this manuscript includes the following:**  
available at [www.sciencemag.org/content/342/6155/257/suppl/DC1](http://www.sciencemag.org/content/342/6155/257/suppl/DC1)

Tables S1 to S17  
Movie S1

## Materials and Methods

### Archaeological background of the nine investigated Mittelbe-Saale cultures

The Mittelbe-Saale region (Fig. S1) is located in the south of Saxony-Anhalt, Germany, at the intersection of the Elbe and Saale rivers. This region spreads from Magdeburg in the North to Naumburg in the South and from the Harz Mountains in the West to Halle in the East. The fertile and highly productive loess plains of Mittelbe-Saale and its central location at the intersection of cultural axes along established trading routes and waterways resulted in a high density of archaeological distinguishable cultures in prehistoric times, either with regional or Pan-European distribution (Fig. S2) (4, 25-26).

In this study, human remains from nine subsequent archaeological cultures of Mittelbe-Saale were analyzed that covered 3,950 years of the Neolithic and EBA (5,500-1,550 cal BC) (Fig. S2), to test whether cultural changes in these millennia were accompanied by population changes. In the following, Neolithic and EBA cultures of Mittelbe-Saale are presented in chronological order describing archaeologically hypotheses about the development and interactions among cultures, their geographical distributions, as well as their characteristic features. The dates of each culture reflect their distribution in Mittelbe-Saale according to the chronology of Central Germany.

#### *Linear Pottery culture (LBK, 5,500-4,775 cal BC)*

The transition from a foraging to a farming lifestyle during the Early Neolithic in Central Europe is marked by the emergence of the LBK, which involved the introduction of domesticated animals and cultivation of plants, the establishment of long-houses in permanent settlements, the use of ceramics, novel stone tools, ritual practices and complex mortuary rites (1-3, 27). At its widest extent, the LBK spread from Western Hungary (Transdanubia) to the Paris Basin and covered the countries of today's Western Hungary, Slovakia, Austria, the Czech Republic, Poland, Germany, and Eastern France (Fig. S2). Archaeologists have reported an expansion of the LBK from Transdanubia to the westernmost corners of the Rhine valley (1-3, 27-30) within 150 years (31). The eponymous feature of the LBK was a characteristic pottery decoration technique consisting of impressed winding and angular lines. The homogenous unity of the LBK in Central Europe broke up into several smaller cultures with small-scale regional dispersal; the *Stroke Ornamented Pottery* culture in the eastern part, the *Lengyel* culture (~4,900-4,200 cal BC) in the south-eastern part, and the *Rössen* culture in the western part of the former LBK area (1-4, 25-26).

#### *Rössen culture (RSC, 4,625-4,250 cal BC)*

The RSC developed from the LBK substrate of Southwest Germany and spread over West and Central Germany, Belgium, and Northeastern France (Fig. S2) (4, 25). Characteristic pottery of the RSC consisted of footed bowls and beakers, which were

decorated with double incisions (4, 25-26). The end of the RSC was initiated by the emergence of cultural formations that are related to the *Lengyel* culture.

#### *Schöningen* group (SCG, 4,100-3,950 cal BC)

The distribution of the SCG was limited to the southern part of Mittelbe-Saale and bordering areas of Lower Saxony (Fig. S2). This culture was named after the eponymous sites of *Schöningen* in Lower Saxony. Characteristic vessels of the SCG included s-shaped bowls decorated with punctured bands, which show conjointly features with the RSC and *Lengyel* cultures (25). The SCG ended with the appearance of the *Baalberge* culture.

#### *Baalberge* culture (BAC, 3,950-3,400 cal BC)

The rise of the BAC in Mittelbe-Saale marks the transition to the Middle Neolithic, which was deeply linked to the emergence of the widespread *Funnel Beaker* culture complex (FBC) that arose about 4,100-2,650 cal BC in southern Scandinavia, Northern Germany, and Poland from post-LBK cultures (1-3, 25-26, 32). Several regional subgroups of the FBC can be distinguished, of which the BAC was the oldest group in Mittelbe-Saale. The BAC was mainly distributed in Central Germany, but settlements and burials were also found along the rivers Oder, Weser, and Elbe in Southwest Poland and in the Northern Czech Republic (Fig. S2). The archaeological record shows that the formation of the BAC was influenced by regional variants of the *Lengyel* culture (4, 25-26). Characteristic features of the BAC pottery included undecorated funnel-shaped vessels, jugs, amphorae, and cups. A particular feature of the BAC was the appearance of grave monuments, such as long barrows and trapezoid enclosures of stone that potentially indicated social differentiation within the BAC community (4, 25). The BAC ended with the emergence of the *Salzmünde* culture.

#### *Salzmünde* culture (SMC, 3,400-3100/3,025 cal BC)

The SMC was a regional variant of the FBC and was distributed in the southern part of Saxony-Anhalt (Fig. S2). Typical pottery shapes of the SMC included long-necked jugs with vertical linear patterns. Archaeological research has suggested that the SMC developed from BAC communities, based on the chronological sequence and common pottery characteristics of these cultures (4, 25). Special features of the SMC mortuary practice were burials with potsherd fillings (4, 25-26). The end of the SMC was initiated by the expansion of the *Bernburg* culture.

#### *Bernburg* culture (BEC, 3,100-2,650 cal BC)

The BEC was the latest representative of the FBC in Mittelbe-Saale and was distributed across Central Germany (Saxony-Anhalt, Thuringia, Saxony and Lower Saxony) (Fig. S2) (4, 25). Bulgy cups with broad handles characterized the ceramic style

of the BEC. The mortuary rites of the BEC reflected a Megalithic tradition. The dead were usually buried in collective burial chambers, constructed of planks and stones (25, 33). In Mittelbe-Saale the BEC communities were displaced by or arose in the subsequent *Corded Ware* culture.

#### *Corded Ware* culture (CWC, 2,800-2,200/2,050 cal BC)

The emergence of the CWC in Mittelbe-Saale marks the transition to the Late Neolithic. This period was characterized by two major archaeological phenomena with Pan-European distributions; the CWC in the East and the BBC in the West with an overlapping zone in Central Europe, encompassing Mittelbe-Saale. The CWC covered large areas of Northeast and Central Europe in the 3<sup>rd</sup> millennium BC extending from the Rhine in the west to the Volga in the east, and from the Danube in the south to Scandinavia and the Baltics in the north (Fig. S2) (2-4, 26, 34). Within this territory, the archaeological record has distinguished several local groups, which can be linked in a broader cultural context by their common characteristic features (e.g. ceramics, sex specific mortuary rites, and stone artifacts) (2-4, 26, 34). The CWC's vast distribution overlapped or bordered on various preceding cultures, including the FBC and Bronze Age Kurgan cultures of the Pontic-Caspian steppe (e.g. *Yamnaya* culture), suggesting various influences on the genesis of the CWC (2-4, 34-37) and an East European origin of the CWC in the region between the Weichsel and the Dnepr (26, 38). An eponymous feature of the CWC was the decoration of vessels by impressions of a twisted cord. During the CWC, the dead were buried in a flexed position with sex-specific differences. Males were predominantly buried on their right sides, females on their left, with both sexes facing south. As a consequence, men's heads were oriented to the west, while the women's heads were facing towards the east (2-4, 26, 34).

#### *Bell Beaker* culture (BBC, 2,500-2,200/2,050 cal BC)

The BBC represents a cultural phenomenon contrary to the CWC. The BBC was spread over vast areas of Western and Central Europe, the British Isles, and parts of North Africa. Similar to the CWC, several local BBC groups can be distinguished within this territory, which can be linked by characteristic features, such as beakers, archery sets, copper objects like daggers, and mortuary rites (2-3, 39). In Mittelbe-Saale, the BBC appeared ~300 years later than the CWC and both cultures coexisted in Central Europe for more than 300 years (Fig. S2) (2-3, 25, 40-41). Interestingly, the BBC and CWC showed not only contrasting geographical distributions but also opposing mortuary rites, suggesting the existence of two distinct 'counter' cultures (41-43). In the BBC, the dead were buried in a flexed position with sex-specific orientation. In contrast to the CWC, men were resting predominantly on their left with heads orientated to the north, while women were resting on their right with heads orientated to the south; and both sexes faced east (44). The most characteristic feature of the BBC was the eponymous ceramic

vessel, a bell shaped beaker. The spread of the BBC was closely linked to the emergence of metallurgy. As a consequence, metallic objects of copper and gold became common across a wide expanse of prehistoric Europe, leading to the manufacture of the first stylistic metal objects on the continent (2-3, 44). The oldest  $^{14}\text{C}$  dates for BBC finds are from the Tagus valley of the Portuguese Extremadura, suggesting an Iberian provenance of the Beaker package (2-3, 44). The BBC ended with the rise of Bronze Age societies in Europe.

#### *Unetice culture (UC, 2,200-1,550 cal BC)*

The emerging metallurgy was a characteristic feature of the Late Neolithic and the EBA. Although the technique to produce copper objects was introduced during the Late Neolithic period, the innovation to alloy tin and copper to bronze during the EBA led to deep economic and social changes in prehistoric Europe. Bronze metallurgies were developed in the 4<sup>th</sup> millennium BC in the Near East and the Caucasus reaching Central Europe ~1,000 years later. The UC was the first culture of the EBA in Mittelelbe-Saale and was named after the eponymous site of Únětice, in the Czech Republic. The distribution of the UC ranged from Thuringia, Saxony-Anhalt, and Saxony in Central Germany, Bohemia and Moravia in the Czech Republic, Silesia in Poland, Southwest Slovakia, and Lower Austria. Most of these regions overlapped with the dispersal of the CWC and BBC (Fig. S2). Thus, archaeologists have discussed the genesis of the UC from local CWC and BBC communities (4, 26, 40). Characteristic features of the UC included the ‘Unetice cup’ and metal objects, such as the bronze *Bohemian Ösenkopfnadel*, which was used for fixing clothes.

#### Archaeological sites

Genetic investigations were performed on 433 individuals that were selected, based on macroscopic preservation, out of a large collection of skeletal remains comprising ~810 individuals from 25 sites of Mittelelbe-Saale (Fig. S1), which is held at the State Museum of Prehistory of Saxony-Anhalt in Halle (Saale), Germany. Most samples were recovered from recent excavations, conducted between 2000 and 2010, which included the sites of Alberstedt, Benzingerode I, Benzingerode-Heimburg, Esperstedt, Eulau, Halle-Queis, Karsdorf, Osterwieck, Quedlinburg (VII, VIII, IX, XII, and XIV), Rothenschirnbach, and Salzmünde-Schiepzig. The remaining material was recovered from the sites of Derenburg-Meerenstieg II, Halberstadt-Sonntagsfeld, Leau 2, Naumburg, Oberwiederstedt (1, 2, 3, and 4), Plötzkau 3, and Röcken 2, which were excavated between 1991 and 1999. Many of the sites investigated were multiphase locations that were either settled several times or continuously during the Neolithic and EBA. These included Benzingerode-Heimburg, Esperstedt, Eulau, Karsdorf, Quedlinburg VII, Quedlinburg XII, and Salzmünde-Schiepzig. Excavations were conducted by the State Office for Heritage Management and Archaeology Saxony-Anhalt, Halle, Germany

(SHMA), which provided the samples, archaeological research, and the permission to conduct genetic analyses as part of joint third-party funded projects as outlined in the acknowledgements. Detailed information about the archaeological sites is given in Table S1.

### Dating

All individuals in this study were archaeologically dated and assigned to a prehistoric culture based on comprehensive archaeological information, including grave goods, mortuary practices, burial customs, radiocarbon dates, and an in-depth assessment of each site. The determination of precise dates and cultural assignments were particularly important for multiphase sites. In some instances, where absolute dates overlapped (e.g. in the Late Neolithic CWC and BBC), the attribution to a certain culture depended entirely on the archaeological context and therefore mis-assignments of single individuals in such a large dataset cannot be completely excluded. However, this risk was minimized by carefully considering all available archaeological information for each individual by experts, in addition to using radiocarbon dating (Table S2).

### Samples and post excavation history

Genetic investigations were conducted independently at ancient DNA (aDNA) facilities in both the Institute of Anthropology, University of Mainz, Germany (IoA) and the Australian Centre for Ancient DNA, University of Adelaide, Australia (ACAD). The samples were split between these institutions based on archaeological sites. The aim was to reproduce results by providing an independent chronological profile of the investigated cultures in both institutions (Table S1). In principle, both labs applied identical genetic typing methods. Hence, the description of the methodological outline is combined in the following sections. Additional or alternative steps to the described standard protocol are indicated, where appropriate.

Recently excavated sites were selected to reduce the impact of DNA degradation induced by long and inappropriate storage conditions (45). Since 2005/2006, the SHMA has taken samples routinely from all detected burials during excavations. All possible precautions were taken through the use of gloves, face masks, caps, long-sleeved clothes etc. to avoid contamination of the samples. Samples were subsequently stored at 4°C. The remaining material was sampled under DNA-free conditions prior to investigations by osteological and genetic researchers at the IoA and ACAD. Whenever possible, genetic samples were obtained from remains in their unwashed and untreated state after excavation. The genetic analyses were run in parallel with osteological investigations, which included identification and individualization of the skeletal remains. Therefore, we can state that each haplotype of this study represents a single individual.

For the genetic analyses, two to five well-preserved samples, from different anatomical regions were taken per individual. If available, teeth were preferred, otherwise long bones or the *pars petrosa* were chosen (Table S2).

#### Precautions against contaminations and ancient DNA work

In aDNA research of human remains there is always a risk of contaminating ancient material with modern human DNA, which can lead to false positive results (46). To minimize the influence of contamination with modern human DNA, the following precautions were taken:

- I The post and pre-PCR labs are located in separate buildings. Sample decontamination and preparation, DNA extraction and PCR set-up were carried out in the pre-PCR labs, while all cloning and sequencing work was conducted in post-PCR labs. The pre-PCR labs are physically isolated with no other molecular biology work in the same building.
- II All workers wore clean-room overalls with hood, laboratory designated shoes, face mask, face shield and several layers of gloves in the pre-PCR labs to exclude sample contamination by investigators.
- III All workers in the pre-PCR labs followed strict workflow protocols.
- IV All rooms, work benches, lab items and instruments were routinely cleaned with bleach and UV-irradiated over night.
- V All items entering the pre-PCR labs were thoroughly cleaned with bleach and subsequently irradiated with UV for at least 30 min.
- VI Sample preparation, DNA extraction, and PCR set-up were carried out in separate rooms in the pre-PCR labs.
- VII Every step, such as grinding, DNA extraction, and PCR set-up, was carried out in customized still-air boxes, which were cleaned with bleach before, between, and after use, in addition to UV irradiation over night.
- VIII Prior to use on ancient material, all chemicals for DNA extraction and PCR were tested extensively through the use of blank control reactions, to exclude the occurrence of chemical contaminations before using the reagents for analyses on ancient human material.
- IX Every grinding, DNA extraction, and PCR step was accompanied by several blank controls. At IoA, grinding controls were used, consisting of synthetic Hydroxyl-apatite to ensure the grinding vessels were being thoroughly cleaned and to detect potential cross-contamination during sample preparation. DNA extractions were accompanied by up to two controls at the IoA or a ratio 1:5 to 1:6 at ACAD. PCR reactions were carried out with at least two blanks with a ratio of approximate 1:5. DNA extraction and grinding controls were subsequently tested with every primer pair via PCR.

- X Positive blank controls were sequenced and the origin and influence of contaminations on relevant samples, DNA extracts or PCR products were assessed. PCR or extractions with multiple positive blank controls were discarded and repeated independently.
- XI All known persons having been involved in the investigation of this study were also typed to monitor potential sources of contamination.

#### Sample preparation and DNA extraction

Each sample was photographed and then decontaminated by initial UV-irradiation of major surfaces for 30 min. For genetic analyses, either whole teeth or 2 cm x 3 cm *compacta* of long bones were used. The bone samples were taken using a diamond drill (KaVO, Biberach Germany). The surface of the teeth and bone samples was removed by shot blasting with aluminum oxide abrasive (Harnisch & Rieth, Winterbach, Germany) at the IoA or by disposable Carborundum discs (Dremel, Australia) at ACAD. All samples were subsequently UV-irradiated for 30 min, ground to a fine powder by using a mixer mill (Retsch, Haan, Germany) at the IoA or a Mikro-Dismembrator (Sartorius, Göttingen, Germany) at ACAD and stored until use at 4°C. After each grinding step, the vessels (Retsch, Haan, Germany) were extensively washed with bleach and DNA free water and cleaned with Silicon dioxide (Roth, Karlsruhe, Germany) to avoid cross-contaminations. Digestion of teeth or bone powder was performed by incubating 0.2-0.5 g powder in an end volume of 3.33 ml lyses buffer consisting of 3 ml 0.5 M EDTA at pH 8.0 (Applied Biosystems/Ambion, Darmstadt, Germany), 15 mg N-lauryl sarcosine (Merck, Darmstadt, Germany) and 600 mg Proteinase K (Roche, Mannheim, Germany) on a rotary mixer over night at 37°C. DNA was extracted with phenol/chloroform/isoamyl alcohol (25:24:1, Roth, Germany). The supernatant was transferred to Amicon Ultra-15 filter units with 50 kDa cut off (Millipore Billerica, USA) and was concentrated and washed several times with DNA free water until the extract was clear. DNA extracts were stored at -20°C until amplification.

#### Amplification and reproduction

The genetic investigations of this study focused on mitochondrial DNA (mtDNA) variability. Therefore, informative positions of the mitochondrial control and coding region were analysed. This included the hyper variable segment I (HVS-I, nucleotide positions 16024-16383), the hyper variable segment II (HVS-II, nucleotide positions 73-386), and 22 diagnostic coding region SNPs (Tables S2, S3) as detailed in the following section.

- I HVS-I was sequenced using two or four overlapping primer pairs spanning either 356 bp (nucleotide positions 16046-16401) or 413 bp (nucleotide positions 15997-16409), respectively (Table S16). At IoA, a hierarchical approach was used to evaluate the DNA preservation of each sample and to adjust the process of



reproduction to the sample quality. Initially, DNA extracts were screened via HVS-I amplification and direct sequencing of two independent samples per individual. The state of DNA preservation was evaluated based on these sequences. Individuals were defined as better preserved if all PCR reactions with standard protocols resulted in consistent HVS-I profiles from both samples, containing a minority of ambiguous nucleotide positions. In these cases, HVS-I amplification was repeated at least three times from different extracts, producing 6-12 independent and overlapping amplicons (depending on primer pairs used). Additionally, selected PCR products that showed ambiguous nucleotide positions were cloned to monitor possible background contaminations and DNA damage. Individuals were classified as less well-preserved if the amplification success was limited and direct sequencing revealed the presence of numerous ambiguous nucleotide positions. For these individuals all PCR products (4-8) from the initial screening were cloned, and if necessary, additional amplifications were performed. PCR reactions were carried out as described previously (47) with the following adjusted cycling conditions: an initial denaturation at 94°C for 6 min, 40-50 cycles of 35 s at 94°C, 35 s at 56-58°C and 40 s at 72°C, followed by a final extension at 60°C for 30 min. Amplicons were purified either by using the Invisorb® Spin PCR Rapace Kit (Invitex GmbH, Berlin, Germany), according to the manufacturer's instructions, or by MultiScreen<sub>94</sub> PCR Plates (Millipore Billerica, USA) by eluting the PCR product in 180 µl of HPLC-water (Applied Biosystems, Darmstadt, Germany), transferring the mix to the MultiScreen<sub>94</sub> PCR Plates, vacuuming for 15 minutes, washing with 50 µl of HPLC-water, vacuuming for 5 min, and resuspending in 18 µl of water. At ACAD, HVS-I was amplified from at least two different samples, using four overlapping primer pairs that resulted in a minimum of 8 independent amplicons. PCR reactions and purification were carried out as described previously (17, 48). In both labs, individuals with inconsistent HVS-I results were either discarded or replicated using an independent third sample (if available) and applying a majority rule.

- II HVS-II sequences were collected for individuals from the same archaeological site, which had identical HVS-I haplotypes, in order to provide sub-haplogroup level information and to detect potential maternal kinship. This approach was only used at the IoA and applied on individuals with consistent HVS-I profiles. The HVS-II sequences were replicated by using two different samples per individual, which were amplified with four overlapping primer pairs that resulted in a minimum of 8 independent amplicons producing a sequence of 364 bp (nucleotide positions 34-397, Table S16) (21). In addition, selected PCR products with numerous ambiguous nucleotide positions in direct sequences were cloned.
- III 22 haplogroup-defining coding region SNPs of the mitochondrial genome were analysed to determine mtDNA lineages and provide an unambiguous classification of the mitochondrial haplogroup. At both labs, coding region SNPs were typed at least

once for each independent DNA extract using the GenoCoRe22 SNP multiplex assay (11).

### Cloning and sequencing

At IoA, PCR products were cloned and/or sequenced following standard protocols as described previously (47), with the following modifications: Clone PCR products were purified by adding 0.5 U of SAP and 2 U of ExoI (MBI Fermentas, St. Leon-Rot, Germany) directly to the PCR product, incubating at 37°C for 40 min and inactivating at 80°C for 15 min. Sequencing products were purified using the MultiScreen<sub>384</sub> SEQ Plates (Millipore Billerica, USA) by eluting in 10 µl of HPLC-water, transferring the mix to the MultiScreen<sub>384</sub> SEQ Plates, vacuuming for 5 min, washing with 50 µl of HPLC-water, vacuuming again for 15 min until the wells have dried, and resuspending in 20 µl Formamide (Applied Biosystems, Darmstadt, Germany). At ACAD, HVS-I amplicons were sequenced according to previous publications (17, 48).

Sequence analyses were performed using the programs SeqMan Pro™ and MegAlign™ from the DNASTAR software package version 9.04 (DNASTAR, Inc, Madison, USA) and Sequencher® version 4.7 (Gene Codes Corporation, Ann Harbor, Michigan, USA). Haplogroup and haplotype assignment was carried out using the software HaploGrep (49) (<http://haplogrep.uibk.ac.at/>) based on the mitochondrial haplogroup phylogeny of phylotree (50) (<http://www.phylotree.org/>, built 14, accessed 5 Apr 2012). Sequence polymorphisms were reported relative to the Reconstructed Sapiens Reference Sequence by using the software FASTmtDNA (RSRS, [www.mtcommunity.org](http://www.mtcommunity.org)) (51).

### Mittelbe-Saale data

Overall, 364 out of 433 analysed individuals of Mittelbe-Saale could be successfully replicated, yielding an amplification success rate of 84.1 % (Table S1). The analysis of significant coding region SNPs via the GenoCoRe22 SNP multiplex assay (11) provided reproducible profiles of 401 individuals (92.6%), highlighting the efficiency of this assay. All mtDNA SNP typing results were concordant with HVS-I sequence haplogroup assignments (Tables S2-S3).

The successfully replicated individuals (Table S2) were grouped according to affiliation with one of the nine investigated archaeological cultures (LBK=88, RSC=11, SCG=33, BAC=19, SMC=29, BEC=17, CWC=44, BBC=29 and UC=94). This collection includes 71 Mittelbe-Saale data that were published previously (LBK=34 (11, 12, 47), RSC=4 (12), SCG=2 (12), BAC=2 (12), SMC=2 (12), CWC=12 (12, 21), BBC=9 (12), and UC=6 (12)) (Tables S4). To increase the sample sizes of the investigated cultures, 14 additional LBK data from sites outside Mittelbe-Saale, such as Seehausen, Eilsleben, Flomborn, Schwetzingen, and Vaihingen in Germany and Asparn-Schletz in Austria (47) were integrated into the population genetic analyses (Table S4).

### Prehistoric comparative data

For comparative population genetic analyses in a prehistoric context, the datasets of Mittelbe-Saale cultures were combined with Mesolithic, Neolithic, and Bronze Age data from the literature. Prehistoric data were included in our analyses, when reported authentication criteria were transparent and described in sufficient detail to allow assessment. This included the use of separated post and pre-PCR labs, multiple independent extractions and amplifications, and partial or complete cloning. These data were pooled into populations according to their cultural assignment and geographic distribution as described in the original publications. Published data that could not be integrated unambiguously into one of these population pools by temporal and geographic affiliation or provide a sample size of less than eight individuals per culture or site (47 52-59), were omitted from the statistical analyses in order to generate a well-defined prehistoric dataset with sufficient population sizes. All used sequence data were defined or updated from the original publication to the mtDNA phylogeny of phylotree.com (50). The prehistoric dataset comprises the following eleven populations (Table S4):

- I A hunter-gatherer metapopulation from Central Europe (HGC, 29,356-2,250 cal BC) consisting of 16 individuals from the sites Dolni Vestonice in the Czech Republic (14), Bad Dürrenberg, Hohlenstein-Stadel, Hohler Fels and Oberkassel in Germany (10, 14), Drestwo and Dudka in Poland (10), Loschbour in Luxembourg (14), and Donkarnis, Kretuonas, and Spiginas in Lithuania (10).
- II A South European hunter-gatherer metapopulation including 13 individuals (HGS, 13,116-4,715 cal BC) from Toledo, Arapouco, Cabeço das Amoreiras, and Cabeço de Pez, in Portugal (60-61), and La Chora, La Pasiega, Erralla, Aizpea (22, 62), and La Braña (63) in Spain.
- III A East European hunter-gatherer metapopulation consisting of 14 individuals (HGE, 33,000-7,000 BP) from Kostenki (18), Chekalino, Lebyazhinka (10), Popovo, and Yuzhnyy Oleni Ostrov (17) in Russia.
- IV 19 individuals from the *Pitted Ware* culture (PWC, 3,300-2,150 cal BC) from Ajvide, Ire, and Fridtorp of Gotland in South Sweden (15-16), which has been regarded as a Neolithic hunter-gatherer population contemporaneous with the FBC.
- V The *Cardial/Epicardial* culture (CAR, 5,475-3,700 cal BC) of the Iberian Peninsula consisting of 18 individuals from Chaves, Sant Pau del Camp, Can Sadurni (23), and the Avellaner Cave of Catalonia in Spain (64-65).
- VI 17 individuals from the Portuguese Neolithic (NPO, 5,480-3,000 cal BC), which included the sites Algar do Bom Santo, Gruta do Caldeirão, and Perdigões (60-61).
- VII A Neolithic population comprising 43 samples from the sites Fuente Hoz, Marizulo, Los Cascajos, and Paternabidea of Basque country and Navarre in Spain (NBQ, 5,310-3,708 cal BC) (22, 62).

- VIII 10 individuals from the *Funnel Beaker* culture complex (FBC, 3,400-2,900 cal BC) from Fräsegråden in South Sweden (15-16) and Ostorf in North Germany (10) representing the Early Neolithic in southern Scandinavia.
- IX 29 individuals of the *Treilles* culture (TRE, 3,030-2,890 cal BC), retrieved from Treilles in Southeast France (24, 65).
- X A Bronze Age Kurgan culture sample set from South Siberia (BAS, 1,800-800 BC) including 11 individuals from the sites Solenoozernaïa I, Solenoozernaïa IV, Tatarka, Oust-Abakansty, and Bogratsky in the Krasnoyarsk region of Russia (19).
- XI 8 individuals of a Bronze Age population from Kazakhstan (BAK, 1,400-900 BC) recovered from the sites Ak-Mustafa, Izmaylovka, Oi-Zhaylau-III/Talapyt-II, and Vodokhranilische/Rybnyi Sakryl-III (20).

#### Present-day comparative data

In order to unravel affinities of the investigated Mittelbe-Saale cultures to present-day populations, we generated an extensive private database of 67,996 HVS-I sequences from contemporary populations of Europe, the Near East, Asia, and Africa, which were collected from the literature and pooled according to geography and/or ethnicity as described in the original publications. Pooling of populations, sequence generation from the haplotype data provided in the database, and - if applied - random selection of sequence data from population pools was carried out by using a database associated in-house program. All sequence data were defined or updated from the original publication to the mtDNA phylogeny of phylotree.com (50). The present-day population data were pooled into the following comparative datasets which were used for different statistical methods (see individual method sections):

- I A Central European metapopulation (CEM, n=500), representing the mtDNA variability of that region, was randomly drawn from a pool of 2,227 sequences from Austria, the Czech Republic, Germany, and Poland, which lie within the former geographic distribution area of the Early Neolithic LBK (Table S5).
- II A dataset of 73 populations from Europe, the Near East, Central Asia, North Asia, Southeast Asia, and North Africa was used for multidimensional scaling (MDS) and principal component analyses (PCA). This dataset was composed of an overall sample size of 50,688 sequences. Populations were pooled by country and ethnicity, with population sizes ranging from 116 to 3,014 individuals and an average sample size of 694 (Tables S10-S11).
- III Procrustes analyses and Ward clustering was carried out with 37,777 sequences of 56 modern-day populations from Europe, the Near East, Central Asia, and North Africa (removing North and Southeast Asian samples). Population sizes ranged from 116 to 2799 individuals with an average sample size of 675 (Table S12).
- IV To resolve the geographic distribution of mtDNA variability, genetic distance mapping was performed, using a dataset of 150 populations from Europe, the Near

East, North Africa, and Central Asia with a total of 44,799 sequences. For this dataset, administrative subdivisions (if applicable) were considered to increase the geographic resolution. The sample sizes ranged from 81 to 1,197 individuals with an average sample size of 299 individuals (Table S13).

### Haplogroup frequencies

In order to evaluate changes in haplogroup frequencies from the Mesolithic to present-day, we plotted the frequencies of 20 haplogroups (N1a, I, W, X, R, HV, V, H, T1, T2, J, U, U2, U3, U4, U5a, U5b, U8, K, and other) of the HGC, the nine Mittelbe-Saale cultures, and the CEM in chronological order (Fig. S3, Table S5).

In addition, we summarized the frequencies of particular haplogroups to components accordingly to their first appearance in Mittelbe-Saale and the characterization of cultural groups by the PCA (Fig. 1C). By using this approach, we were able to ascribe haplogroups to a particular culture/period and to distinguish mtDNA lineages and their proportion in each culture/population that I) were prevalent in the Mesolithic (the genetic substratum), II) likely have been brought in by incoming farmers, III) have emerged in later Neolithic phases, or IV) could not be assigned to one of the other three components without further comparative data or a higher phylogenetic resolution (Fig. 3, Table S5). We defined the haplogroups U, U4, U5 and U8 as characteristic hunter-gatherer; N1a, T2, K, J, HV, V, W, and X as Early/Middle Neolithic; I, U2, T1 and R as Late Neolithic lineages; and H, U3 as well as Asian and African lineages of the CEM as others. The haplogroups H and U3 were pooled with others to provide a conservative interpretation of the currently available palaeogenetic data, which is based on the following observations. U3 has been detected so far only from Neolithic LBK and SMC specimens, which results in a diffuse dissemination pattern through time (Fig. S3). The fact that U3 has been detected from one LBK individual suggests that this haplogroup was present since the Neolithic transition in Central Europe. However, it is known that pre-Neolithic hunter-gatherers show a high variability of U haplogroups including U, U2, U4, U5, and U8 (10, 14, 17-18, 63). This is also supported by population genetic studies based on the present-day U variability that have proposed an origin of this haplogroup in the Near East, which was then spread over Europe and West Asia by anatomical modern humans during the Upper Palaeolithic (~43,000 BC) (8). Thus, it seems plausible that haplogroup U3 was also present in Europe since Palaeolithic/Mesolithic times. However, due to the facts that mtDNA data from European hunter-gatherers are still incredible rare and U3 was not observed among Mesolithic specimens so far, we cannot argue either for a Mesolithic nor a Neolithic emergence of this haplogroup in Central Europe. In contrast, haplogroup H has been shown a high within group variation with more than 800 subgroups identified to date ([www.phylotree.com](http://www.phylotree.com)) (50). Furthermore, a recent aDNA study has shown that at different time periods distant H lineages can be observed in Central Europe, arguing for a complex population history of this haplogroup in Europe (12) including a potential origin

and expansion within Europe from a post-glacial refugium. However, most of the H variability is located outside the mitochondrial control region and thus, a precise dating of the emergence of H sub-haplogroups and an assignment to particular periods can be only addressed by a higher phylogenetic resolution.

For each haplogroup and component the 95% confidence interval was calculated by performing a conservative non-parametric bootstrapping with 10,000 replicates in R 2.13.1 (The R Foundation for Statistical Computing 2011, <http://www.r-project.org/>) (66). For each culture, we calculated the frequency of haplogroups and components in each of 10,000 bootstrap datasets generated by sampling with replacement from the observed data (all bootstrap datasets were the same size as the original dataset). The 2.5th and 97.5th percentiles of the distribution of calculated frequencies provided estimates of the lower and upper 95% confidence interval limits for each empirical frequency (Figs. 3, S3, Table S5).

### Ward clustering

To unravel groups with similar haplogroup composition, we performed cluster analysis with the Mittelbe-Saale cultures and different sets of comparative data based on the haplogroup frequencies of the corresponding datasets.

We used a chronological dataset consisting of the HGC, the Mittelbe-Saale cultures, and the CEM and performed the Cluster analysis based on frequencies of 20 haplogroups (N1a, I, W, X, R, HV, V, H, T1, T2, J, U, U2, U3, U4, U5a, U5b, U8, K, and other) (Fig. 2A, Table S5).

In order to identify cluster of prehistoric populations with similar haplogroup composition the frequencies of 22 haplogroups of the Mittelbe-Saale cultures and eleven ancient comparative data were used, that were observed from the ancient data (C, Z, N\*, N1a, I, W, X, R, HV, V, H, T1, T2, J, U, U2, U3, U4, U5a, U5b, U8, and K) (Fig. 1D, Table S9).

Cluster analyses of Mittelbe-Saale cultures and 56 present-day populations were carried out based on 23 haplogroup frequencies (N1a, I, I1, W, X, HV, HV0/V, H, H5, T1, T2, J, U, U2, U3, U4, U5a, U5b, U8, K, African haplogroups (L), Asian haplogroups (A, B, C, D, E, F, G, Q, Y, and Z), and all remaining haplogroups (other)) to classify mitochondrial variability of the investigated samples into the clustering of present-day populations (Figs. S6A-I, Table S12).

All analyses were conducted with Ward's hierarchical clustering method and Manhattan distances and were visualized as dendrogram by using the *hclust* function in R.2.13.1. Significance of each cluster was evaluated by 10,000 bootstrap replicates using the *pvclust* function in R.2.13.1.

### Genetic distances and Mantel test

To reveal similarities or differences between the HGC, the Mittelbe-Saale cultures and the Central European Meta-population (CEM) on the sequence haplotype level we calculated genetic distances ( $F_{st}$ ) based on HVS-I data (nucleotide positions 16059-16400) (Fig. 2B, Table S6). We used a resampling strategy to evaluate whether different sample sizes of the CEM yield similar  $F_{st}$ -values to the Mittelbe-Saale cultures, which would exclude biases by largely fluctuating sample sizes. Therefore, we generated ten additional CEMs each with 100 or 50 randomly selected individuals out of a pool of 2,227 sequences and included these samples into the analysis.  $F_{st}$ -values were computed in Arlequin 3.5.1 (<http://cmpg.unibe.ch/software/arlequin35/>) (67) by using the Tamura & Nei substitution model (68) and an associated gamma value of 0.300, which were evaluated as most suitable for the data by the Akaike and Bayesian information criterion (AIC and BIC) in jModelTest 0.1.1 (<http://darwin.uvigo.es/software/modeltest.html>) (69). Significant variations in  $F_{st}$ -values were tested by performing 10,000 permutations in Arlequin. P-values were adjusted post hoc to correct for multiple comparisons and minimize the probability of type I errors (false positives). As we have a large number of comparisons ( $n=55$ ), the widely used Bonferroni correction can be considered too conservative (adjusting the significance level to  $\alpha=0.00090$  in our case), causing a reduction in power to detect type II errors or false negatives (70-71). We therefore chose the Benjamini-Hochberg method, an alternative multiple testing procedure, which uses the concept of the false discovery rate (72) and is considered intermediate between performing no correction and the highly conservative Bonferroni correction. P-values were adjusted using the function *p.adjust* in R (Fig. 2B, Table S6).

In addition, genetic distances between the Early/Middle Neolithic cultures (LBK, RSC, SCG, BAC, SMC) and Late Neolithic/Early Bronze Age cultures (BEC, CWC, BBC, UC) were calculated accordingly to the settings mentioned above (Fig. 2C, Table S6), but with an adjusted gamma-value of 0.242. All genetic distances were visualized by a level-plot that was generated using the function *levelplot* implemented in the *lattice* package of R 2.13.1.

We performed a Mantel-test to examine whether the increasing genetic distances correlate with the temporal distance between the ancient cultures. A negative correlation would be the expected effect of genetic drift on small populations over time. We used the genetic distance matrix from above and a temporal matrix based on the terminal dates of each archaeological culture (Table S1), assuming a maximum effect of genetic drift at the end of each culture. The Mantel-test was computed using the function *mantel* in R with 10,000 permutations and selecting Pearson's correlation method (73).

### Analysis of molecular variance (AMOVA)

In order to analyze genetic variation between the Mittelbe-Saale cultures as well as between the Mittelbe-Saale cultures and the CEM on different population genetic levels

(among groups, among populations within groups, and within populations) AMOVA was performed on HVS-I sequences (nucleotide positions 16059-16400).

The nine investigated Mittelbe-Saale cultures were arranged into 289 different combinations consisting of two, three or four groups with varying numbers of populations within each group and AMOVA was conducted for each constellation (Table S7). According to the supported differentiation of Mittelbe-Saale cultures into two groups by Ward clustering, we tested at first all possible two group arrangements (no. 1-246) to determine the best culture combination with the highest ‘among groups’ and the lowest ‘within groups’ variance. Adjacent, this arrangement was split into all possible three (no. 247-259) or four (no. 260-289) group combinations.

In order to identify, which ancient Mittelbe-Saale cultures were most similar to the present-day metapopulation the cultures and the CEM were arranged into 37 different combinations and AMOVA was computed for each constellation (Table S14). The AMOVA of the Mittelbe-Saale cultures supported a differentiation into an Early/Middle Neolithic and Late Neolithic/Early Bronze Age group (Table S7). This constellation was used as prior in the analyses with the CEM. Consequently, all possible two (no 1-2) or three (3-37) group arrangements with the CEM were tested by combining one or more cultures of the Early/Middle Neolithic or Late Neolithic/Early Bronze Age group with the CEM. The best combination of one or more Mittelbe-Saale cultures with the CEM was determined by the highest ‘among groups’ and the lowest ‘within groups’ variance.

All AMOVA were carried out with the standard AMOVA function implemented in Arlequin 3.5.1 using the Tamura & Nei (68) substitution model and a gamma value of 0.298 (without CEM) or 0.242 respectively (with CEM), which were evaluated by jModelTest 0.1.1 as the most suitable settings for the data used in the analyses. Significant variations in  $F_{st}$ -values were tested by performing 10,000 permutations.

#### Test of population continuity (TPC)

We consider the following problem: we have sampled mtDNA from individuals at the same geographical location from two time points, and wish to test a simple model of ‘population continuity’. That is, we want to test if the allele frequencies at the later time point are consistent with the allele frequencies at the early time point and some amount of genetic drift. The information comes from the fact that mtDNA has many alleles - while genetic drift results in changes in allele frequency and decreased genetic diversity, it cannot cause the appearance of new alleles. Our approach is to fit a model of genetic drift to the observed data, and then to evaluate how well this model fits the data. If it does not fit, we conclude that there has most likely been migration from an outside source between the two time points. An alternative interpretation of a poorly fitting model is that natural selection influenced the allele frequencies between the two time points; this seems



unlikely given other lines of evidence (for example, selection seems unlikely to generate the shifts in affinities to modern day populations seen in Figs. S4A-I-S7A-I).

We first describe our model of genetic drift, and then how we evaluate its fit. The basic model is an extension of the Gaussian approximation to genetic drift (74-75) to multi-allelic systems. We ignore mutation and selection, and wish to model changes in allele frequency over time in a population of arbitrary demography.

Call the earlier population in time  $A$  and the later population  $B$ , and let them be separated by  $t$  generations. Consider a multi-allelic locus in  $A$ . The  $L$  alleles (i.e. mtDNA haplogroups) at this locus have frequencies  $\vec{f}^A = [f_1^A, f_2^A, f_3^A, \dots, f_L^A]$ . We now want to model the allele frequencies in  $B$ . Let the effective population size of the population be  $N_e$ . After a single generation, the allele frequencies in this intermediate population (call it  $A'$ ) are a multinomial sample from the allele frequencies in  $A$ . As in the standard bi-allelic case, the expected allele frequencies remain unchanged, but the variance depends on  $N_e$ . Since now there are multiple alleles, we also model the covariance in frequencies between alleles  $i$  and  $j$ :

$$\begin{aligned} E[\vec{f}^{A'}] &= \vec{f}^A \\ \text{Var}(f_i^{A'}) &= \frac{f_i^A(1 - f_i^A)}{N_e} \\ \text{Cov}(f_i^{A'}, f_j^{A'}) &= \frac{-f_i^A f_j^A}{N_e}. \end{aligned}$$

These are simply properties of multinomial sampling. Now assume that the number of generations between  $A$  and  $B$ ,  $t$ , is small compared to  $N_e$ , such that the amount of genetic drift in the population is relatively small. We introduce a parameter  $c \approx \frac{t}{N_e}$ . If the allele frequencies were not constrained to be between zero and one, we could use a simple multivariate normal approximation for the allele frequencies in  $B$ . Instead, following Coop and colleagues (76), we introduce surrogate allele frequency variables  $\vec{\theta}$  that are not constrained to be between zero and one, and model these as:

$$\vec{\theta} \sim \text{MVN}(\vec{f}^A, c\mathbf{W}),$$

where

$$\mathbf{W}_{ij} = \begin{cases} f_i^A(1 - f_i^A) & \text{if } i = j \\ -f_i^A f_j^A & \text{otherwise.} \end{cases}$$

Now,  $\vec{f}^B$  is a simple transformation of  $\vec{\theta}$ :

$$f_i^B = g(\theta_i) = \begin{cases} 0 & \text{if } \theta_i < 0 \\ \theta_i & 0 < \theta_i < 1 \\ 1 & \theta_i > 1. \end{cases}$$

Note that because  $\sum_i f_i^B$  must be one,  $\mathbf{W}$  as written is singular. In practice we only model  $L - 1$  alleles as above, and then set  $f_L^B = 1 - \sum_{i=1}^{L-1} f_i^B$ . Now assume we have sampled  $n_A$  alleles from population  $A$  and  $n_B$  alleles from population  $B$ , and have observed counts  $\vec{c}^A = [c_1^A, c_2^A, \dots]$  and  $\vec{c}^B = [c_1^B, c_2^B, \dots]$ . The counts are multinomial draws from  $\vec{f}^A$  and  $\vec{f}^B$ . To complete the model, we place a prior Dirichlet distribution on  $\vec{f}^A$  and a prior Gaussian distribution on  $\ln(c)$  (see discussion below). Putting everything together, the full hierarchical model is:

$$\begin{aligned} \vec{f}^A &\sim \text{Dirichlet}(\alpha) \\ \ln(c) &\sim N(\ln(\mu), \sigma^2) \\ \vec{\theta} &\sim \text{MVN}(\vec{f}^A, c\mathbf{W}) \\ \vec{f}^B &= g(\vec{\theta}) \\ \vec{c}^A &\sim \text{Multinomial}(n_A, \vec{f}^A) \\ \vec{c}^B &\sim \text{Multinomial}(n_B, \vec{f}^B) \end{aligned}$$

The  $c$  parameter controls our prior belief about the amount of genetic drift that has occurred between  $A$  and  $B$ . After experimenting with using an exponential distribution for our prior on  $c$ , we settled on a prior of the form shown above. For all applications, we set  $\sigma^2 = 1$ . This provides a relatively wide range of drift, while avoiding extremely small amounts of drift. For example, consider two populations separated by 200 generations (or about 6,000 years). If we set the mean amount of genetic drift as 0.02 (which corresponds to  $N_e = 10,000$ ), this prior places 95% of its weight on the range [1,400-71,000]. There is thus a substantial amount of prior probability on small values of  $N_e$ , which seems like a desirable property if we would like to be conservative in rejecting continuity. We evaluated this prior using simulations (see below).

We would like to have a measure for how well the above model fits observed data from two populations. We will use two test statistics. First, consider the standard  $\chi^2$  statistic used to measure the difference in distributions underlying  $\vec{c}^A$  and  $\vec{c}^B$ :

$$\chi^2 = \sum_{i \in (A,B)} \sum_{j=0}^L \frac{(c_j^i - E[c_j^i])^2}{E[c_j^i]},$$

where

$$E[c_j^i] = \frac{(c_j^A + c_j^B)n^i}{n^A + n^B}.$$

We want to know: under the model of drift, what is the probability of seeing a  $\chi^2$  statistic as extreme as that observed in the real data? (Note that in the absence of drift, this could be calculated from the  $\chi^2$  distribution). We will use the concept of the posterior predictive p-value (77-78). Imagine repeating the sampling of mtDNA haplotypes from A and B. Call these hypothetical resamples  $\vec{c}^{A'}$  and  $\vec{c}^{B'}$ , and the statistic computed from these samples  $\chi^2'$ . We want to estimate:

$$p_{\chi^2} = P(\chi^2' \geq \chi^2 | \vec{c}^A, \vec{c}^B).$$

In the context of our model, we can write this as:

$$p_{\chi^2} = \int I_{\chi^2' \geq \chi^2} P(\vec{c}^{A'}, \vec{c}^{B'} | c, \vec{\theta}, \vec{f}^A) P(c, \vec{\theta}, \vec{f}^A | \vec{c}^B, \vec{c}^A) d\vec{c}^{A'} d\vec{c}^{B'} dc d\vec{\theta} d\vec{f}^A,$$

where  $I$  is the indicator function. To evaluate this integral, we simulate  $\vec{c}^{A'}$  and  $\vec{c}^{B'}$  from the posterior distribution of the model parameters, and count the number of times the simulated  $\chi^2$  statistic is larger than that observed in the real data. To simulate from the posterior, we use the MCMC algorithm described below.

The other test statistic we use captures the intuition that genetic drift acts to decrease haplotype diversity. Thus, if the model is not a good fit, it will tend to overestimate the haplotype diversity in A. This test statistic is simply the number of observed haplogroups in A:  $Y = \sum_{i=1}^L I(c_i^A > 0)$ , where  $I$  is the indicator function. We can compute a similar posterior predictive p-value for this statistic. Again, we denote this statistic computed on a hypothetical sample  $Y'$ , and:

$$p_Y = \int I_{Y' \leq Y} P(\vec{c}^{A'} | c, \vec{\theta}, \vec{f}^A) P(c, \vec{\theta}, \vec{f}^A | \vec{c}^B, \vec{c}^A) d\vec{c}^{A'} d\vec{c}^{B'} dc d\vec{\theta} d\vec{f}^A.$$

To have a single p-value that captures the model fit, we combine  $p_{\chi^2}$  and  $p_Y$  using Fisher's method. This p-value as a measure of model fit is likely to be extremely conservative, since we have used the data twice: once to fit the model and again to evaluate this fit (78).

We want to simulate from the joint posterior distribution  $P(c, \vec{\theta}, \vec{f}^A | \vec{c}^B, \vec{c}^A)$ . We do this by MCMC. We begin by setting all the values in both  $\vec{\theta}$  and  $\vec{f}^A$  to  $1/L$  and  $c$  to 1. In

all cases we set  $\alpha$  (the Dirichlet prior on  $\vec{f}^A$ ) to an  $L$ -vector of 0.5. The prior on  $c$  depends on the pair of populations, described in the section on application. We then do the following updates:

- I Update  $\vec{f}^A$ : We propose a new vector of allele frequencies in  $A$  as a mixture of a Dirichlet distribution centered on  $\vec{f}^A$  with a tuning parameter  $K$  and the prior:

$$P(\vec{f}^{A'}|\vec{f}^A) = 0.99\text{Dirichlet}(K\vec{f}^A) + 0.01\text{Dirichlet}(\alpha).$$

- II Update  $c$ : We propose a new value of  $\ln(c')$  centered on  $\ln(c)$  (the logarithm ensures non-negativity) as  $\ln(c')|\ln(c) \sim N(\ln(c), \sigma^2)$ , where in all applications  $\sigma^2 = 0.4$ .
- III Update  $\vec{\theta}$ : We propose a new value of  $\vec{\theta}$  by adding an independent bit of noise to each entry, so  $\theta'_i \sim N(\theta_i, \sigma^2)$ , where  $\sigma^2$  is small, on the order of  $1 \times 10^{-5}$ . In practice the entire vector is updated at once. If the allele frequencies implied by  $\vec{\theta}$  sum to greater than one, the proposal is rejected. Otherwise all acceptance probabilities are by Metropolis-Hastings.

To simulate counts from the posterior, after a sufficient burn-in period, we sample values of  $\vec{f}^A$  and  $\vec{f}^B$  and simulate  $n^A$  and  $n^B$  alleles from each.

We applied the above procedure to all pairs of populations in the transect data. We set the prior on  $c$  based on the number of generations between the two populations, such that the mean prior on the exponential was  $\frac{t}{10,000}$ , where  $t$  is the difference in time depths of the two populations. For populations that are extremely close in time (less than 30 generations apart), we set the prior to be  $\frac{30}{10,000}$ , so as to allow some amount of genetic drift. For each pair of populations, we tuned the transition probabilities in the Markov chain to have acceptance probabilities around 10-50% (this meant smaller jumps for more closely related populations). We then ran the Markov chain for 50,000 iterations, sampling every 10 states after a burn-in of 2,000 iterations.

To test the performance of this model, we performed simulations of two scenarios: one 'null' scenario of two populations where one is the direct descendant of the other, and one 'alternative' scenario of complete population replacement. To generate the null data we did the following:

- I Draw  $\vec{f}^A$ , a vector of 20 allele frequencies, from a Dirichlet distribution with parameter  $\alpha = [0.5, 0.5 \dots 0.5]$ .
- II Draw 40 haplotypes from  $\vec{f}^A$  by multinomial sampling.
- III Draw  $\theta$  from a multinomial distribution centered on  $\vec{f}^A$  with genetic drift set to  $c$ .
- IV Convert  $\theta$  to  $\vec{f}^B$  by setting all negative values to zero and all value over one to one.

V Draw 40 haplotypes from  $\vec{f}^B$  by multinomial sampling.

This is a simulation of a situation where  $B$  is a direct descendant of  $A$ . To generate the data under the alternative of population replacement, we generated counts of an additional 40 haplotypes as done for population  $B$  above, with the same values for  $\vec{f}^A$ . Call this population  $C$ . In this simulation,  $C$  is *not* a direct descendant of  $B$ , so running our method on  $B$  and  $C$  is what would be expected in the situation of complete population replacement.

For each scenario, we ran 200 simulations with  $c = 0.02$ . We then ran our method on each simulation, and generated posterior predictive p-values as described. We were most interested in using the null simulations to test the calibration of the p-values and their robustness to mis-specification of the prior on  $c$ . We thus ran our method on the null simulations three times, with  $\mu$  (the mean of the prior on  $c$ ) set to 0.02, 0.01, and 0.002. Since the true value of  $c$  in the simulations is 0.02, the second and third runs correspond to situations where the prior mean is below the true amount of genetic drift, either slightly or by an order of magnitude. Shown in Fig. S9 are the distributions of p-values from the simulations in each situation. When the prior mean is near the true value of drift, the distribution of p-values is as expected. It is only when our prior mean is an order of magnitude below the true value that there is a slight excess of small p-values in the null simulations. Even in this case, however, this excess is small (2.5% of the p-values are less than 0.01, as compared to the expected 1%). By contrast, in the simulations of population replacement, the shift towards small p-values is much clearer (Fig. S9).

#### Multidimensional scaling (MDS)

In order to unravel genetic affinities of the investigated samples to present-day populations on the sequence level, genetic distances between the Mittelbe-Saale cultures and 73 present-day Eurasian populations were computed based on HVS-I sequences (nucleotide positions 16059-16400) (Table S10). Genetic distances were calculated in Arlequin 3.5.1 using the Tamura & Nei substitution model (68) and a gamma value of 0.385, which were supported as best settings for the data used in these analyses by jModelTest 0.1.1. Slatkin  $F_{st}$ -values (79) were used for MDS and visualized in a three dimensional space (Figs. S4A-I) using the *scatterplot3d* function of the corresponding R package.

#### Principal component analysis (PCA)

PCA was carried out in order to characterize each Mittelbe-Saale culture based on their haplogroup composition and to indicate their affiliations within the mtDNA diversity of eleven Mesolithic, Neolithic, and Bronze Age populations as well as 73 present-day populations from Europe, the Near East, Asia, and North Africa.

For the PCA with the prehistoric comparative data haplogroups were divided into the following 22 groups, which were observed from the ancient mtDNA data: C, Z, N\*, N1a, I, W, X, R, HV, V, H, T1, T2, J, U, U2, U3, U4, U5a, U5b, U8, and K (Fig. 1C, Table S9).

The PCA with the present-day comparative data were conducted with the following 23 haplogroup frequencies, which cover the most frequent haplogroups of Eurasian populations: N1a, I, I1, W, X, HV, HV0/V, H, H5, T1, T2, J, U, U2, U3, U4, U5a, U5b, U8, K, African haplogroups (L), Asian haplogroups (A, B, C, D, E, F, G, Q, Y, and Z), and all remaining haplogroups (other) (Figs. S5A-I, Table S11).

All analyses were performed using the *prcomp* function for categorical PCA implemented in the R 2.13.1 package and plotted in a two (prehistoric comparative data) or three dimensional space (present-day comparative data), displaying the first and second or the first three principal component respectively. Three dimensional PCA was plotted using the *scatterplot3d* function of the corresponding R package.

#### Procrustes analysis

Procrustes analyses were conducted to visualize the geographic transformation of the Mittelbe-Saale cultures according to their affinities to present-day populations. Procrustes transformation was carried out with 23 haplogroup frequencies (N1a, I, I1, W, X, HV, HV0/V, H, H5, T1, T2, J, U, U2, U3, U4, U5a, U5b, U8, K, African haplogroups (L), Asian haplogroups (A, B, C, D, E, F, G, Q, Y, and Z), and all remaining haplogroups (other)) and geographic coordinates of the Mittelbe-Saale cultures and 56 present-day populations (Table S12). Haplogroup frequencies were used for an initial PCA and the PCA scores of the first and second principal component were rotated to the best fit with the geographic coordinates of the populations used in the analyses. PCA was carried out according to the settings described in detail above. For each population pool used in these analyses, a point of reference was defined, which best described the available geographic information (Table S12). Procrustes transformation was performed and tested on significance by 100,000 permutations using the functions *procrustes* and *protest* of the *vegan* package of R 2.13.1 and subsequently plotted on a geographic map (Figs. S6A-I).

#### Genetic distance maps

To resolve the geographic distribution of mtDNA variability, genetic distances between the Mittelbe-Saale cultures and 150 present-day populations were plotted on a geographic map. Genetic distances were computed using a high-resolution frequency dataset of 107 sub-haplogroups, which can be distinguished from their HVS-I sequences.  $F_{st}$ -values between each culture and present-day populations were computed in Arlequin 3.5.1 using the implemented F-statistic for frequency based distance computations, combined with longitudes and latitudes, and interpolated by the Kriging method implemented in ArcGis version 10.0 (Arcmap, Environmental Systems Research Institute

(Esri) Inc, Redlands, USA, Figs. S7A-I). For each population pool used in this analysis, a point of reference was defined which best described the available geographic information (Table S13).

#### Coalescent simulations (BayeSSC)

The software program BayeSSC (80) was used to determine the most likely demographic scenario, which could explain the mtDNA structure of the investigated nine Neolithic and EBA cultures and their relationship to the present-day CEM. As our genetic analyses indicated that not all cultures can be treated as discrete populations in our models we decided not to choose a separate deme for each culture, and instead incorporated them in chronological order in one or two larger deme(s) with exponential growth. Our main aim was therefore to examine the observed genetic differentiation between Early/Middle and Late Neolithic/EBA cultures could be explained by genetic drift over time (in an expanding Central European population) or by subsequent population genetic events (i.e. influx, migration, replacement after the initial Neolithic transition). We focused on the transition between the Early/Middle and Late Neolithic period as the Meso-Neolithic transition has been addressed in previous studies (10-11). The demographic models or scenarios that were tested generally followed archaeological hypotheses, especially concerning the proposed interaction and likely origin/grouping of cultures, which are largely based on the described affinities in the material culture (see archaeological background above). BayeSSC was performed in accordance with previous publications (10-11, 15, 23, 81-82), with the following parameters of sequence evolution for all simulations used: a generation time of 25 years, a fixed mutation rate of  $7.5 \times 10^{-6}$  substitutions per site per generation (83), a transition/transversion ratio of 0.9841 (84), and a gamma distribution of rates with shape parameter of 0.205 (theta) and 10 (kappa) (83). A uniform distribution was applied to estimate the effective population size ( $N_e$ ) for an Upper Palaeolithic source population (100-5,000) 1800 generations ago, an initial Neolithic population (100-50,000) 400 generations ago, and the present-day Central European deme (100-10,000,000) at time 0 (today). The maximum value of the prior distribution of the Central European demes  $N_e$  represents 7% ( $1/14^{\text{th}}$ ) of the present-day census size of approximately 144 million people, which includes the countries Germany, Poland, Austria, the Czech and Slovak Republics based on the July 1, 2009 estimate by the United Nations Department of Economic and Social Affairs ([http://www.un.org/esa/population/publications/wpp2008/wpp2008\\_text\\_tables.pdf](http://www.un.org/esa/population/publications/wpp2008/wpp2008_text_tables.pdf)).

In order to test whether the signals of population differentiation between the Early/Middle and the Late Neolithic/EBA could be explained by drift alone (null hypothesis  $H_0$ ), a scenario of genetic continuity in a single, panmictic, Central European population was assumed (incorporating the nine Mittelbe-Saale cultures in chronological order), and the  $N_e$  was allowed to grow exponentially since Paleolithic times ( $H_{0a}$ ; 1,800 generations ago) or alternatively since the Neolithic ( $H_{0b}$ ; 400

generations ago). Growth rates were drawn from a uniform prior distribution based on the natural logarithm function of the respective deme sizes. In the alternative scenarios H1 and H2 we explored the importance of the two Neolithic periods by splitting the present-day deme into two pseudo-demes and considered them as separate statistic groups with a prior distribution of 100-5,000,000 individuals, assuming that present-day mitochondrial diversity resulted from two structured Neolithic populations in Central Europe, which had experienced exponential growth since the Neolithic and to which each of the nine Neolithic/EBA cultures was assigned. In models summarized under H1, we assumed continuity between the present-day CEM and Early and Middle Neolithic cultures (LBK, RSC, SCG, BAC, SMC) as deme 0 and tested the proportion of migrants (10-50 %) in Mittelbe-Saale during the Late Neolithic and EBA (BEC, CWC, BBC, UC) from Neolithic deme 1, starting 190 generations ago with an exponential growth for both demes since 320 generations ago, a convergence of both demes 400 generations ago followed by a constant population size since 1,800 generations ago. In models summarized under H2 we tested the opposite scenario by assuming a continuous deme 0 including CEM and Late Neolithic/EBA cultures and various proportions of Early/Middle Neolithic migrants (10-50 %) from deme 1 starting at 320 generations ago. We maximized the chances of gene flow between the two Early/Middle Neolithic and Late Neolithic/EBA derived demes in post Neolithic times by applying a migration matrix ( $m = 0.02$ ) between time zero and the time of the population split before merging both into one single Upper Paleolithic deme 1,800 generations ago. In total, twelve variants of three main demographic scenarios were tested and details of the alternative models are illustrated in Fig. S8.

All models were simulated using 250,000 genealogies in BayeSSC (<http://www.stanford.edu/group/hadlylab/ssc/index.html>). Summary statistics were calculated on the observed and simulated data using a customized executable to overcome the differences in underlying parameter calculations (SCStat.exe, C. Anderson, personal communication). To condition the runs and compare simulated and observed data with summary statistics, two within-population (Tajima's D and haplotype diversity) and three between-population parameters ( $F_{st}$ , average pairwise distances and number of private alleles) were chosen, which best reflected population differentiation between the nine Neolithic/EBA cultures. As the HVS-I region contains many recurrent mutations that can produce statistical noise, a 22 base pair tag resulting from the GenoCoRe22 results to the HVS-I sequences was concatenated to obtain more 'realistic' haplogroup based estimates for the summary statistics. Sequence tags for modern-day population data were inferred from the haplogroup status of each respective sequence in accordance with the current mtDNA phylogeny (50). Simulated and observed values were compared by an Approximate Bayesian Computation (ABC) framework (85) in R 2.14.1 using available scripts (<http://www.stanford.edu/group/hadlylab/ssc/index.html>). We retained 1% of the simulations with smallest Euclidian distance to observed population statistics to construct



posterior distributions of population parameters. The *mle* function was used to pick the ‘maximum credible’ version of each model. The prior distribution of these model versions was replaced with MLE values from the posterior distributions and each model was run for 1,000 genealogies. Goodness of fit for the different models tested was compared using Akaike Information Criteria (AICs) (86) and Akaike’s weights  $\omega$  (87-88). Results of the simulations and ABC are summarized in Table S15.

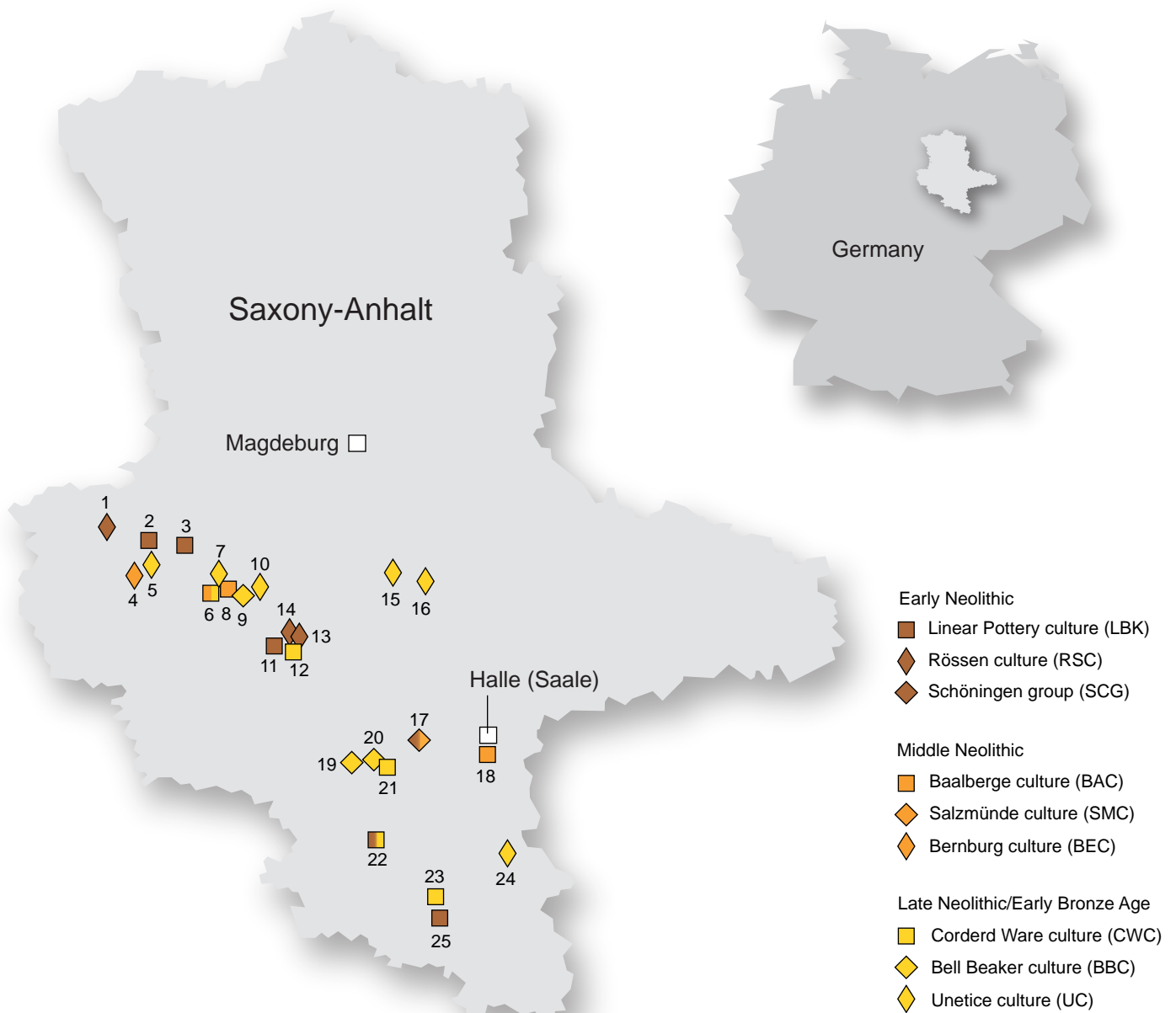
### Authenticity of aDNA results

Because the risk of modern human DNA contamination in prehistoric human remains is highly probable in archaeogenetic investigation, numerous criteria for endogenous aDNA authenticity have been established (46, 89-91). Even if all recommended criteria were considered and strict scientific protocols were applied, contamination of a skeleton cannot be ruled out entirely (90). Therefore ancient human DNA must be evaluated by completing several independent experiments in a logical framework that ensure the assessment of authenticity probability (91). Accordingly, we discuss the following criteria that were applied to our data to compile a chain of evidence and which provide confidence that the results of this study are authentic and derived from endogenous DNA.

- I Various precautions against contaminations were taken into account (see individual method sections on sample preparation, extraction and PCR).
- II The samples were taken under DNA-free conditions. In the case of more recent excavations, the DNA samples were taken *in situ*. Furthermore, most of the samples originated from excavations which were not older than 10-15 years. This reduces the degree of DNA degradation, which can lead to a significant increase of background contaminations. Whenever possible, samples were not washed, treated, or examined before taking DNA samples.
- III All HVS-I profiles were replicated and reconstructed from at least two independent extracts obtained from two anatomically distant regions. The HVS-I was amplified by numerous overlapping amplicons. At IoA, ambiguous sequences from PCR products were cloned, and 4-8 clones per PCR were sequenced.
- IV The overlapping amplicons enable to identify gapless, contiguous sequences, which phylogenetically correspond to the current mtDNA phylogeny. Importantly, overlapping regions of the HVS-I contain numerous phylogenetically informative positions, which resulted in SNP calling redundancy (at least 4x).
- V The authenticity of the individuals depends on several independent amplifications targeting different loci of the mitochondrial genome (HVS-I, HVS-II and 22 coding region SNPs). In all cases, the HVS-I sequence support the haplogroup assignment by the GenoCoRe22 SNP multiplex assay. Furthermore, the phylogenetic classification of the HVS-II profiles is consistent with the HVS-I and 22 coding region SNP designations.

- VI Characteristic *post-mortem* DNA damage-derived substitutions (mostly C>T and G>A transition) were observed in PCR product sequences. Generally, these substitutions are neither reproducible nor consistent and can be distinguished from genuine SNP calls.
- VII Only few researchers were involved in the genetic analyses (GB, WH, CJA, CR, ASN, SK). Most of the samples were analysed by GB and WH. All other genetic investigators were restricted to single sites and cultures. HVS-I and in some cases additional HVS-II profiles or complete mitochondrial genomes (12) of all genetic investigators and people involved in sampling and excavation were sequenced and compared to the aDNA results in order to monitor potential contamination. Overall, only two researchers show a mitochondrial profile matching samples they worked on (researcher no. 5 and 10) (Table S17). Both researchers show the ‘modal’ haplogroup H haplotype A16129G T16187C C16189T T16223C G16230A T16278C C16311T, which matches the rCRS on the HVS-I. Furthermore, in case of researcher no. 10 the HVS-II sequence is also available (G73A C146T C152T C195T A247G 309.1C 309.2C 315.1C). These two researchers were involved in the genetic investigation of some individuals of the SMC (no. 5) and the sampling of the sites of Karsdorf, Quedlinburg and Oberwiederstedt II (no. 10), respectively. Considering the available HVS-I and HVS-II data and the restricted access of both researchers to particular skeletal material, only the individuals SALZ116, KAR16, KAR 29, QLB26, QLB28, QUEXII1, QUEXII2, QUEXII3 and OBW2 match their mitochondrial profile. However, this modal HVS-I H haplotype is very common in present-day populations of Europe and also in our full ancient dataset. Thus, in the light of all precautions against contaminations during the sampling and the genetic investigations and the reproduction of the results through multiple independent extractions and amplification of different mitochondrial loci, the probability of contaminating several skeletons is highly unlikely. All other researchers can be excluded as potential contamination by different HVS-I, HVS-II, and complete mitochondrial genomes (12) and/or they did not have access to the particular samples.
- VIII Overall, the favorable climatic and soil conditions in Mittelbe-Saale resulted in exceptional DNA preservation, as indicated by the high amplification success rates of reproducible HVS-I sequences (84.1%) and GenoCoRe22 SNP profiles (92.6%). Most of the investigated archaeological sites show a consistent amplification success rate (70-100%) with very few exceptions, such as Osterwieck (0%), Halle-Queis (20%), and Eulau (61.8%).
- IX In the chronological dataset, fluctuations in haplotype and haplogroup frequencies were observed, which show chronological trends and – when assigned to particular cultures – can be interpreted as characteristic haplogroup compositions.

- X The haplogroup composition of Mittelbe-Saale cultures observed genetic affinities to modern-day populations of different regions that are in accordance to previous publications and/or can be explained by archaeological hypothesis.
- XI The most striking argument for the authenticity of the presented data is the independent reproduction of the overall results in two different laboratories (IoA and ACAD). Both labs generated highly similar chronological datasets from different sites, which independently supported population dynamic events in the later Neolithic.



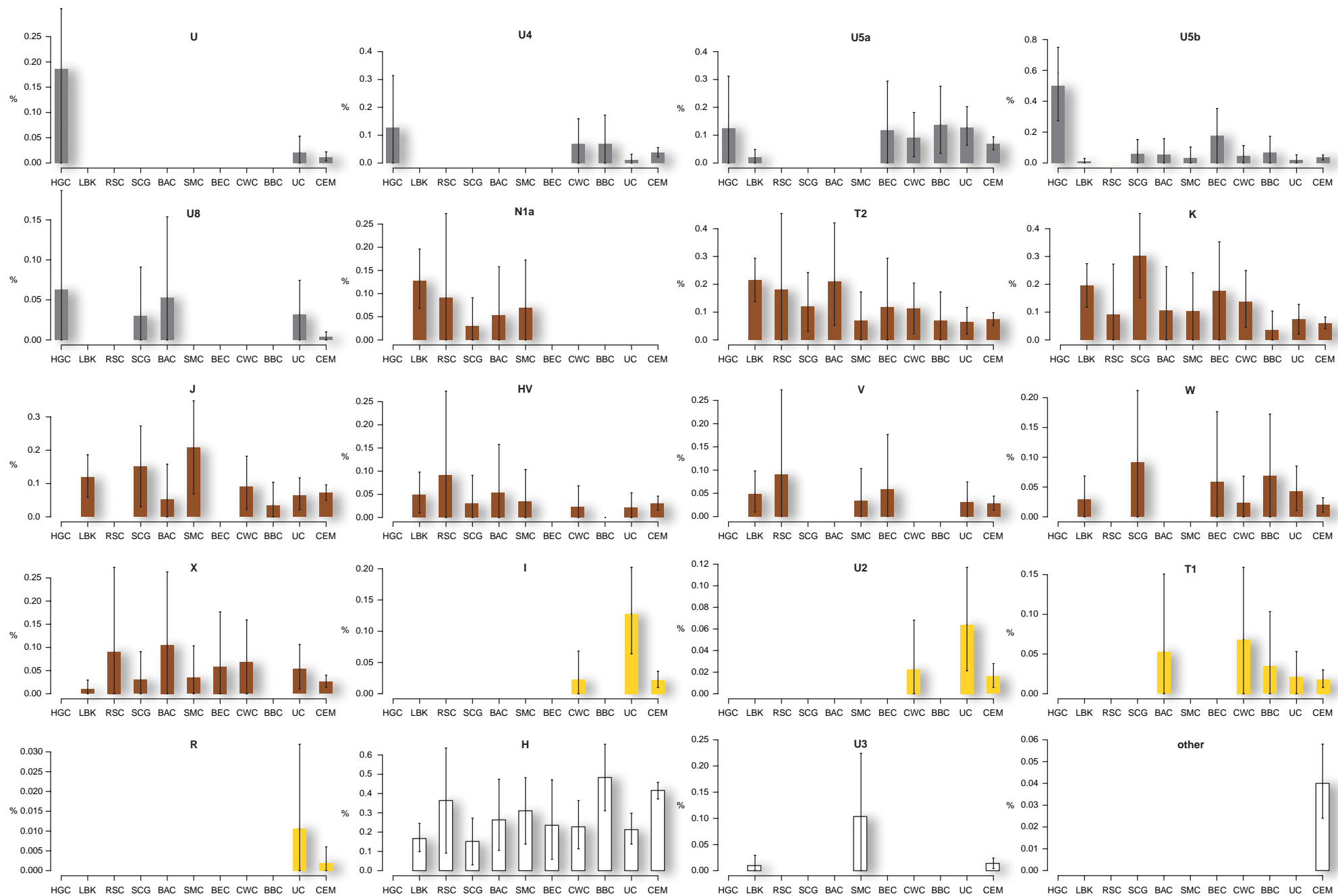
**Fig. S1. Map of investigated sites in the Mittelbe-Saale region**

The Mittelbe-Saale region is located in southern Saxony-Anhalt, Germany. This study includes 433 individuals from 25 sites (1-25) ascribed to nine distinct archaeological cultures. Osterwieck (1), Derenburg-Meerenstieg II (2), Halberstadt-Sonntagsfeld (3), Benzingeroode I (4), Benzingeroode-Heimburg (5), Quedlinburg VII (6), Quedlinburg VIII (7), Quedlinburg IX (8), Quedlinburg XII (9), Quedlinburg XIV (10), Oberwiederstedt (11), Oberwiederstedt 2 (12), Oberwiederstedt 3 (13), Oberwiederstedt 4 (14), Plötzkau 3 (15), Leau 2 (16), Salzmünde-Schiepzig (17), Halle-Queis (18), Rothenschirmbach (19), Alberstedt (20), Esperstedt (21), Karsdorf (22), Eulau (23), Röcken 2 (24), Naumburg (25). The symbols and color shadings of the archaeological cultures indicate their association to Neolithic/Early Bronze Age phases. Multiphase sites are indicated by several shadings. Further information of the cultures and sites investigated are listed in Table S1 and S2.



**Fig. S2. Chronology and distribution of prehistoric cultures in the Mittelbe-Saale region**

Maps depict the geographic distribution of Neolithic and Early Bronze Age cultures in Mittelbe-Saale across Europe or Central Germany in chronological order. Eponymous or characteristic ceramic vessels are shown for each culture and dates refer to the occurrence and duration in Mittelbe-Saale.

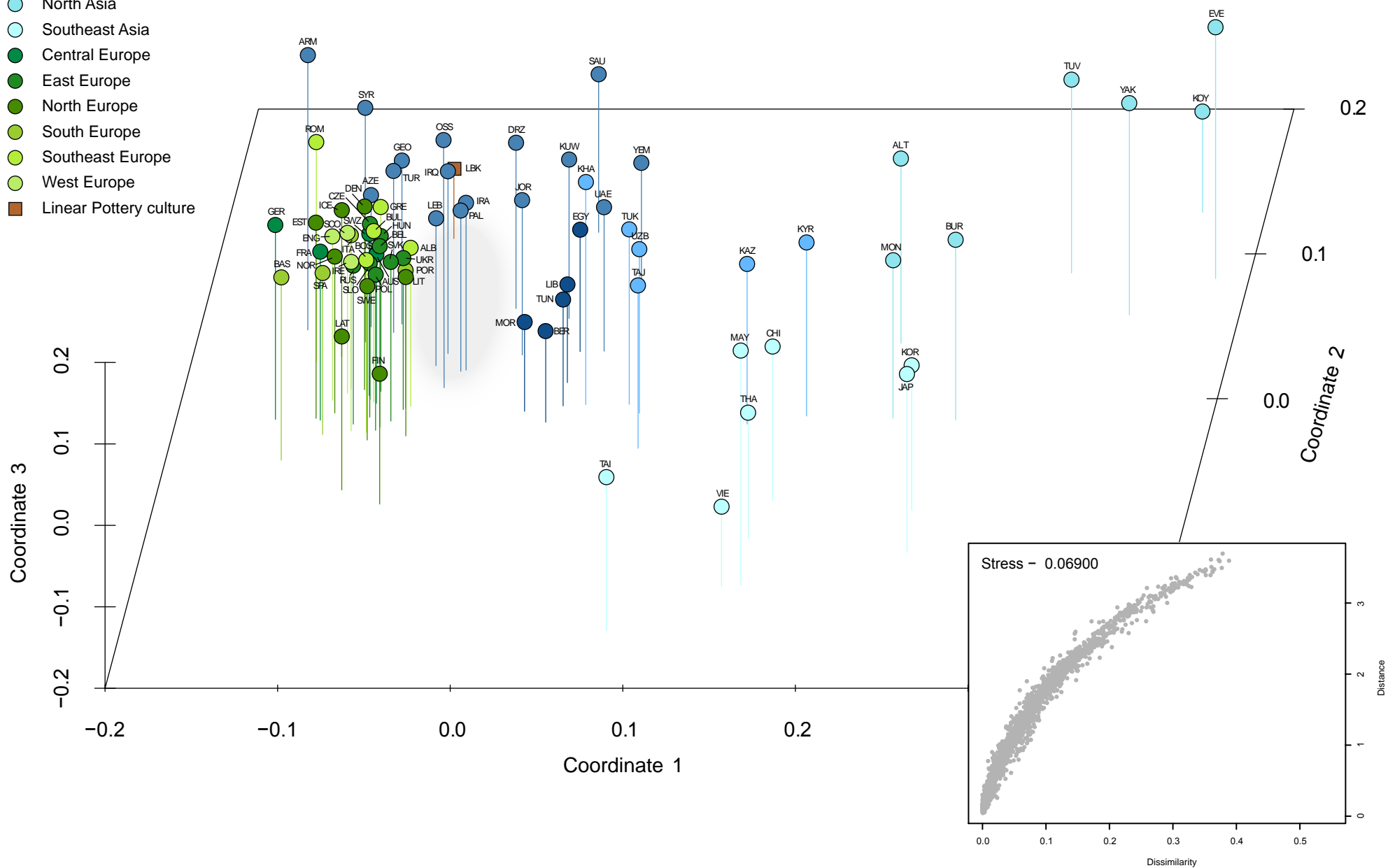


**Fig. S3. Chronological profiles of haplogroup frequencies**

Haplogroup frequencies of 20 sub-groups of hunter-gatherers from Central Europe (HGC), the Mittelbe-Saale cultures (LBK, RSC, SCG, BAC, SMC, BEC, CWC, BEC, UC), and a Central European metapopulation (CEM) were plotted in chronological order. The shading denotes characteristic hunter-gatherer (grey), Early/Middle Neolithic (brown), and Late Neolithic/Early Bronze Age (yellow) haplogroups as identified by PCA. Haplogroups that could not assign to one of these periods were defined as other (white). Error bars of haplogroup frequencies indicate 95% confidence intervals of 10,000 bootstrap replicates. See Tables S2, S4, and S5 for further details.

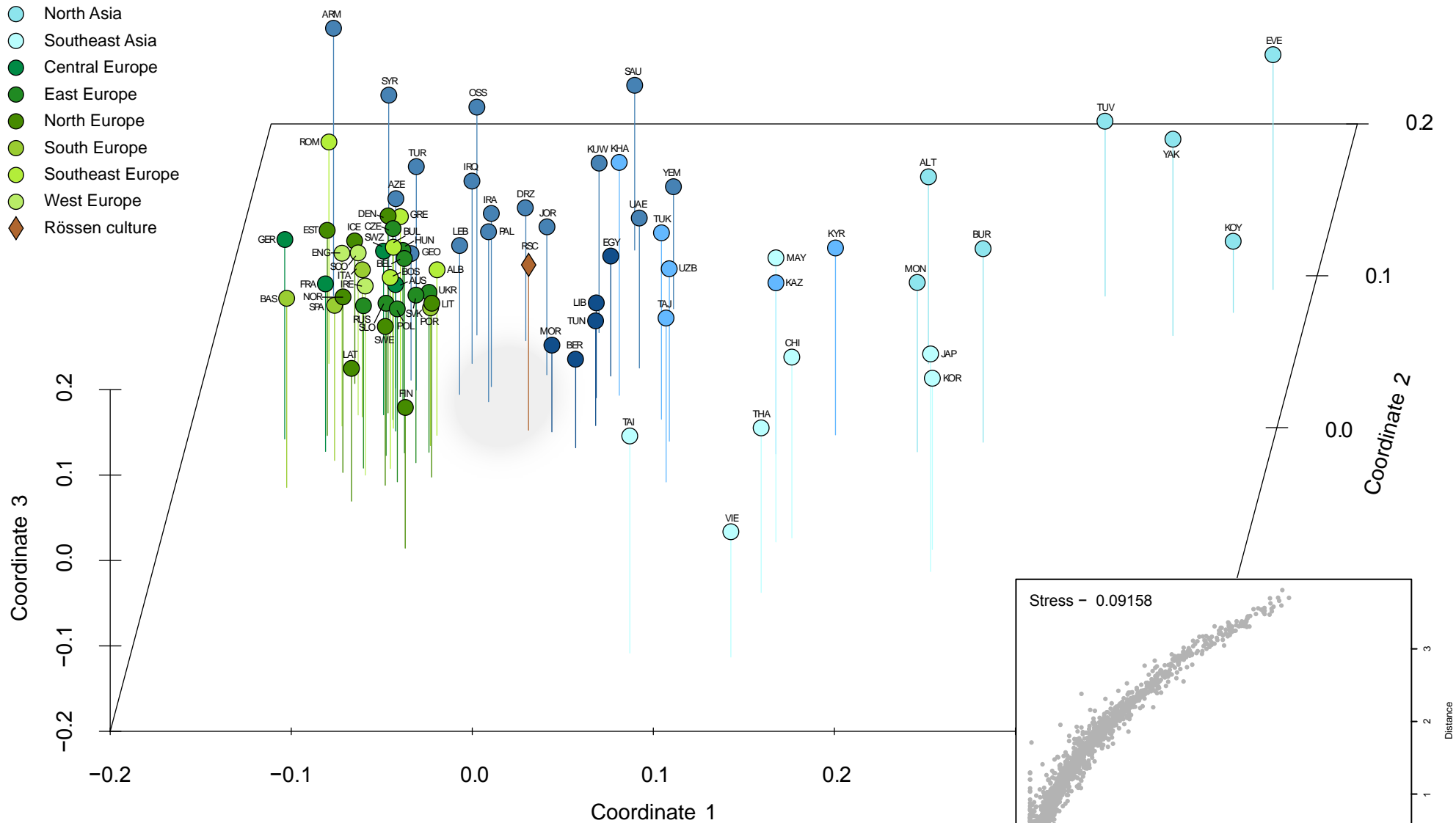
**A**

- North Africa
- Southwest Asia
- Central Asia
- North Asia
- Southeast Asia
- Central Europe
- East Europe
- North Europe
- South Europe
- Southeast Europe
- West Europe
- Linear Pottery culture



**B**

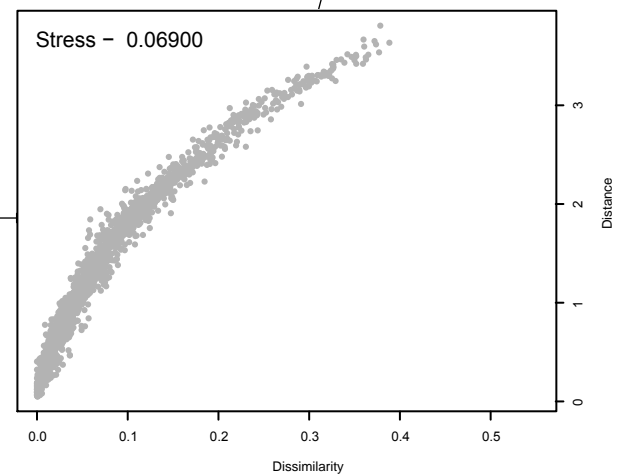
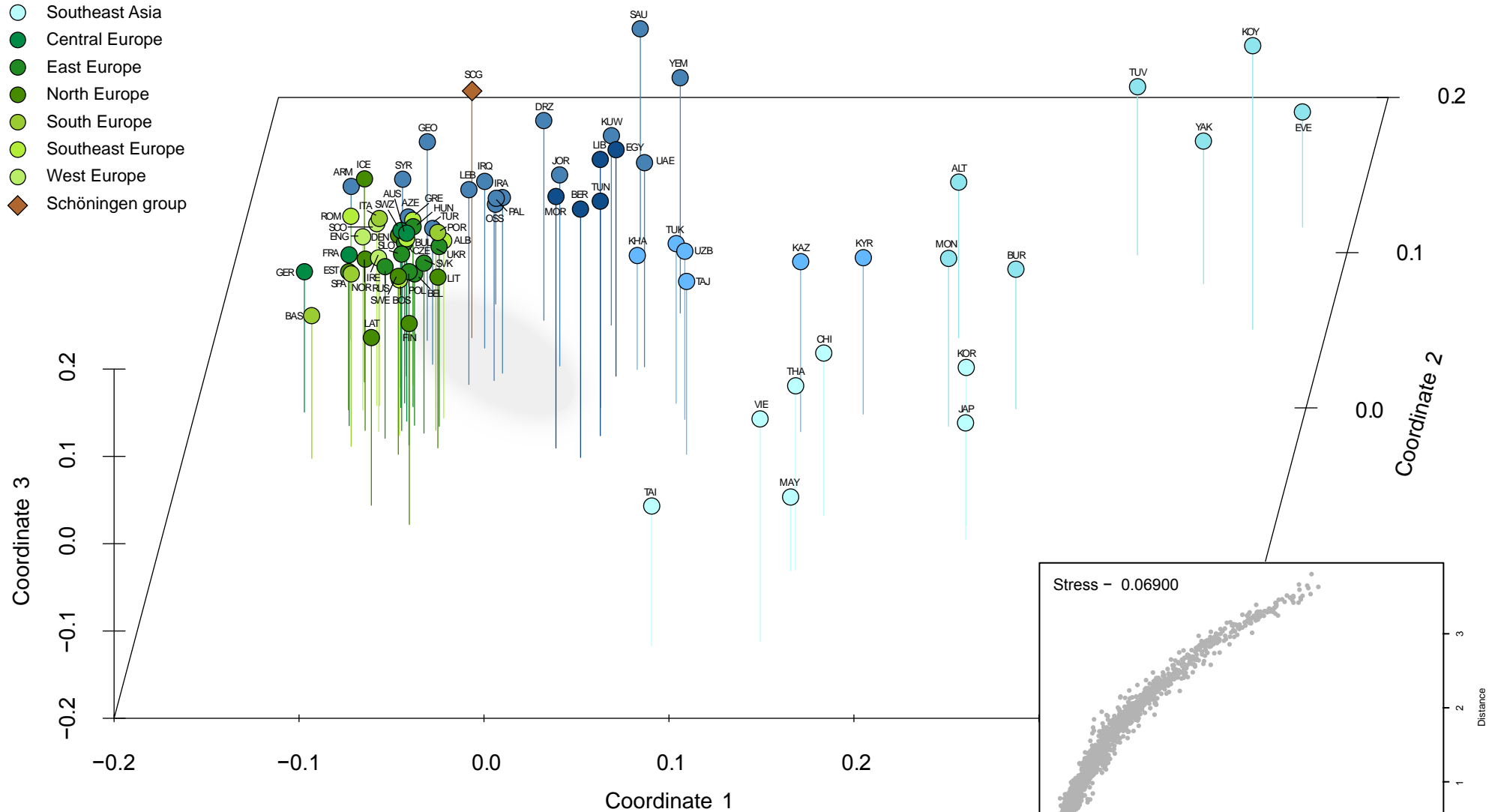
- North Africa
- Southwest Asia
- Central Asia
- North Asia
- Southeast Asia
- Central Europe
- East Europe
- North Europe
- South Europe
- Southeast Europe
- West Europe
- Rössen culture





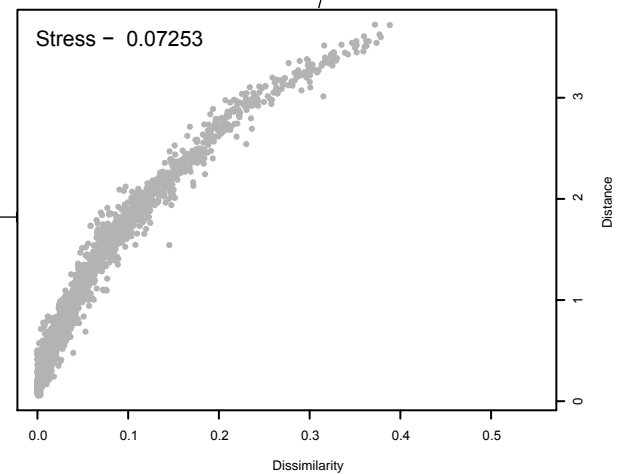
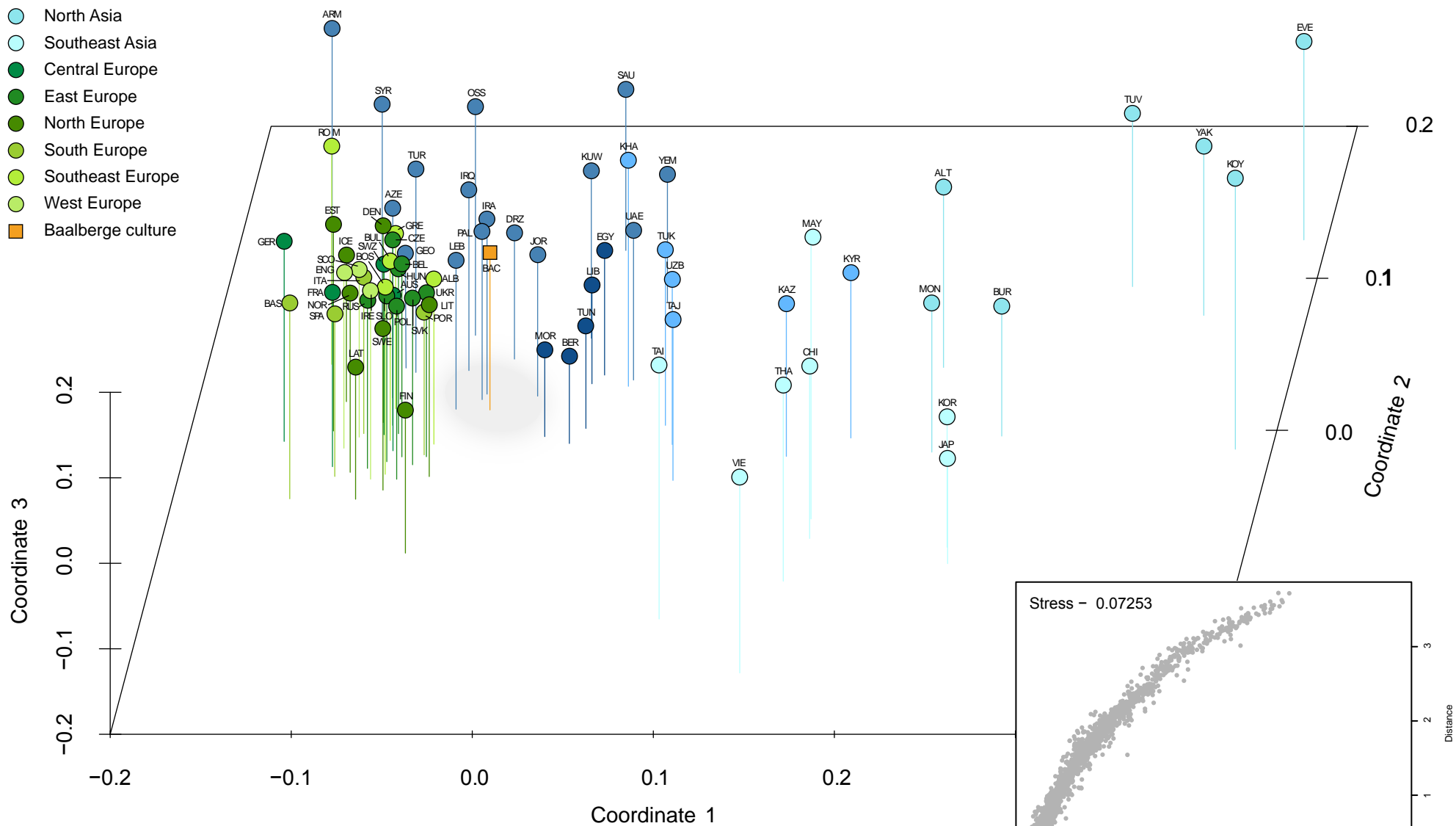
C

- North Africa
- Southwest Asia
- Central Asia
- North Asia
- Southeast Asia
- Central Europe
- East Europe
- North Europe
- South Europe
- Southeast Europe
- West Europe
- ◆ Schöningen group



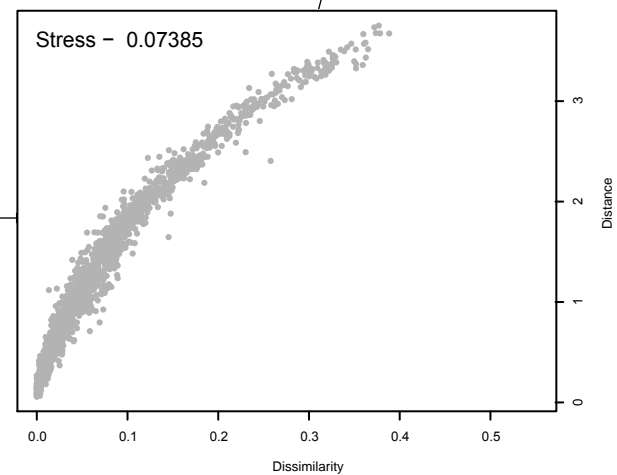
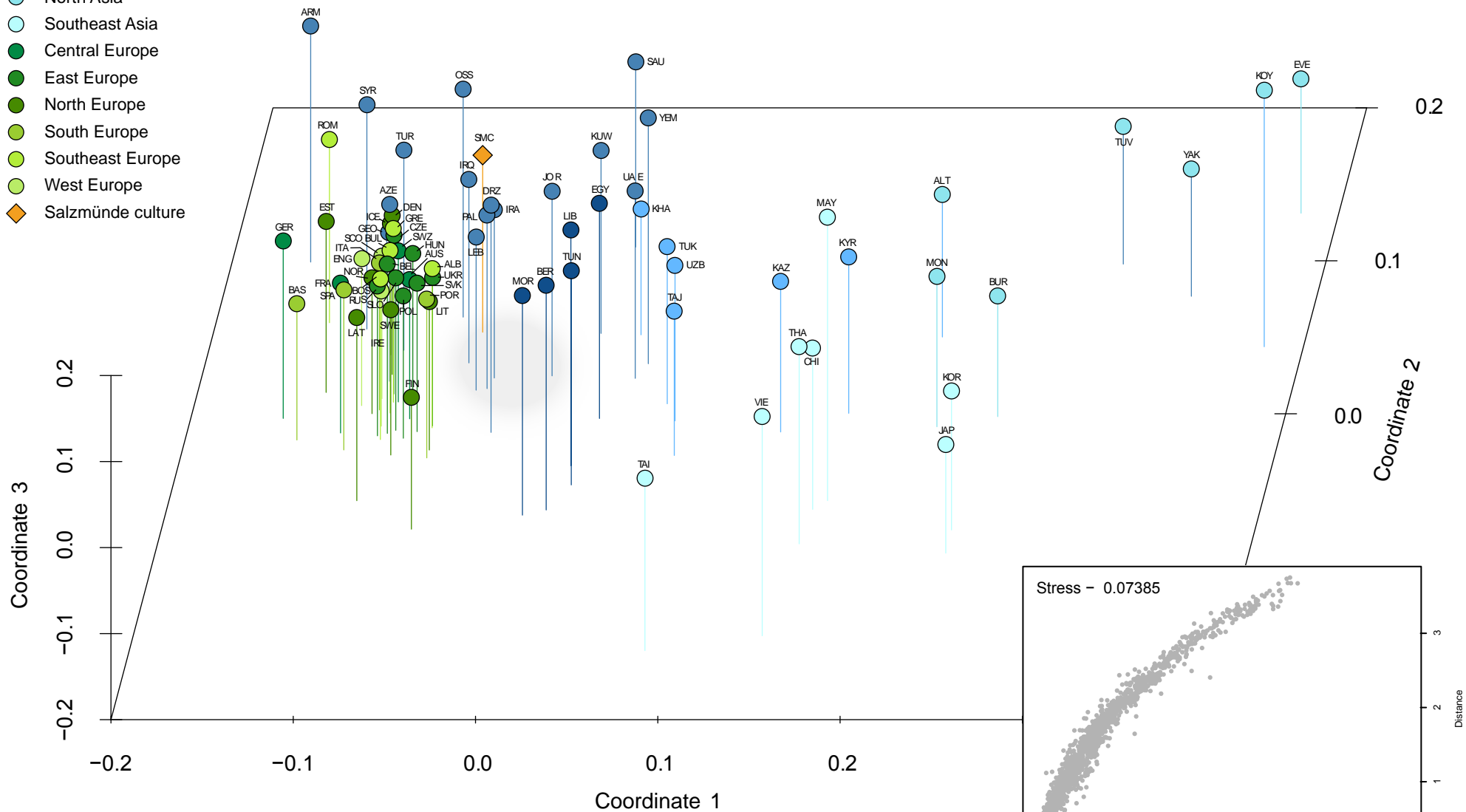
D

- North Africa
- Southwest Asia
- Central Asia
- North Asia
- Southeast Asia
- Central Europe
- East Europe
- North Europe
- South Europe
- Southeast Europe
- West Europe
- Baalberge culture



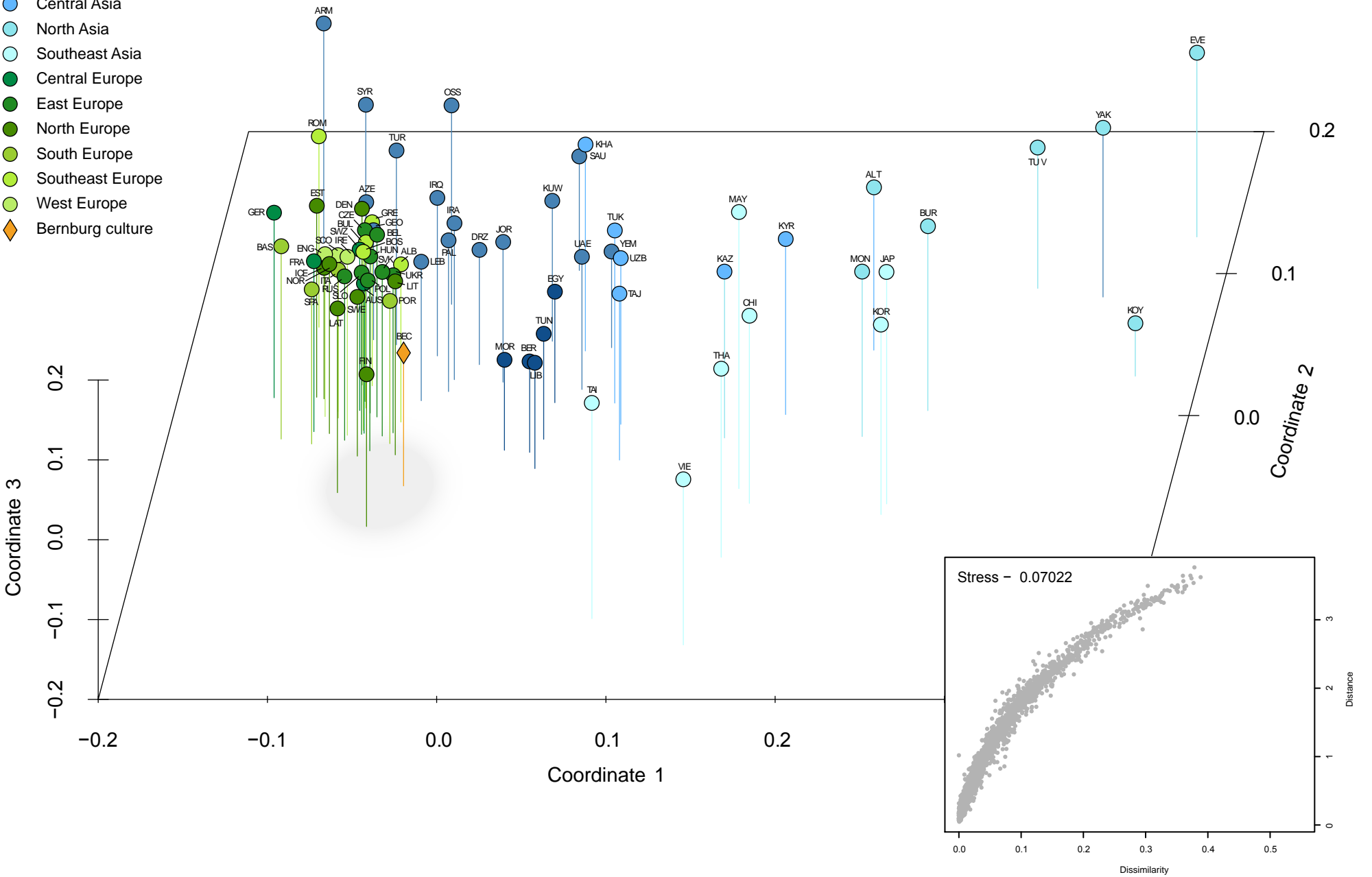
E

- North Africa
- Southwest Asia
- Central Asia
- North Asia
- Southeast Asia
- Central Europe
- East Europe
- North Europe
- South Europe
- Southeast Europe
- West Europe
- ◆ Salzmünde culture



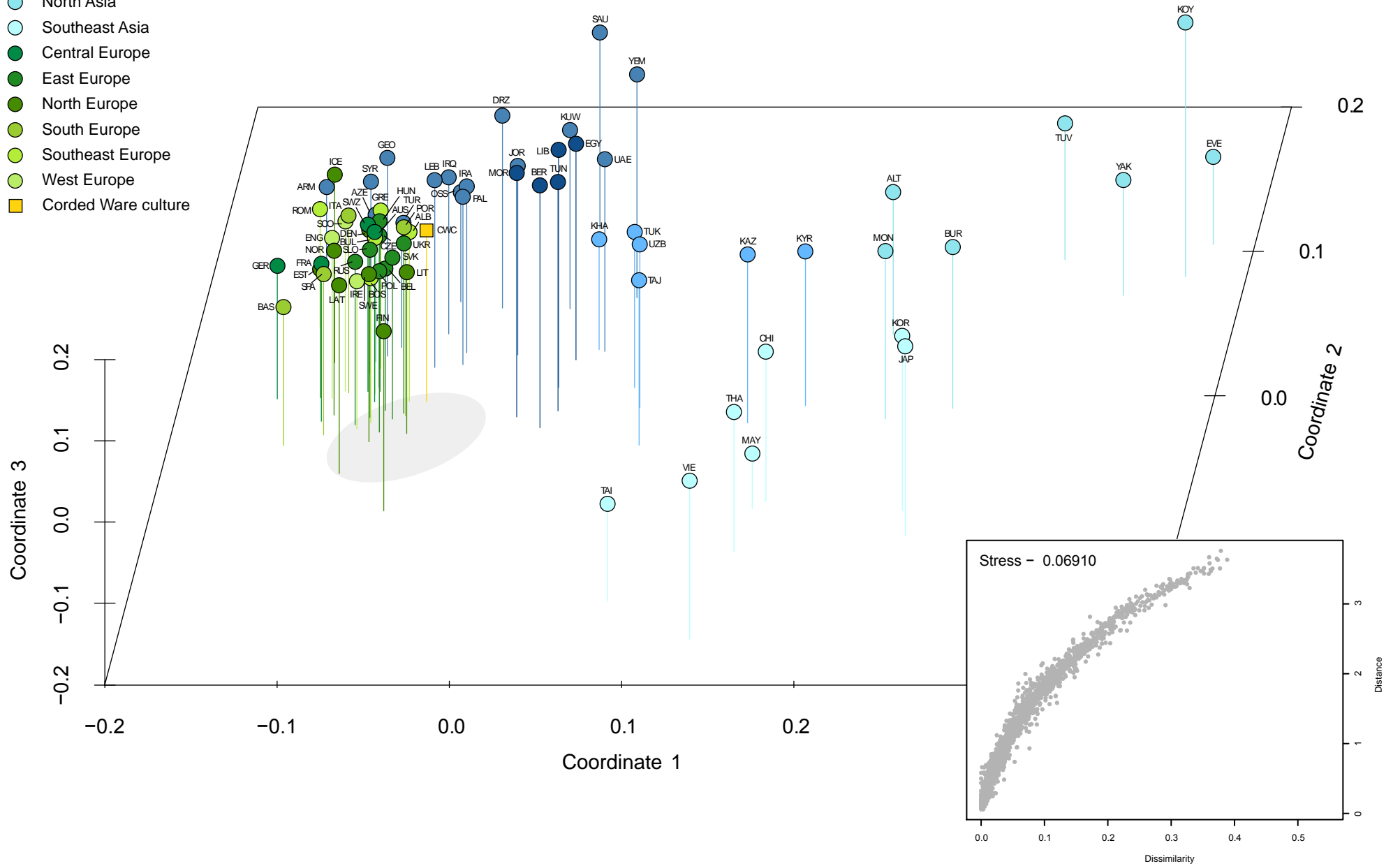
**F**

- North Africa
- Southwest Asia
- Central Asia
- North Asia
- Southeast Asia
- Central Europe
- East Europe
- North Europe
- South Europe
- Southeast Europe
- West Europe
- Bernburg culture



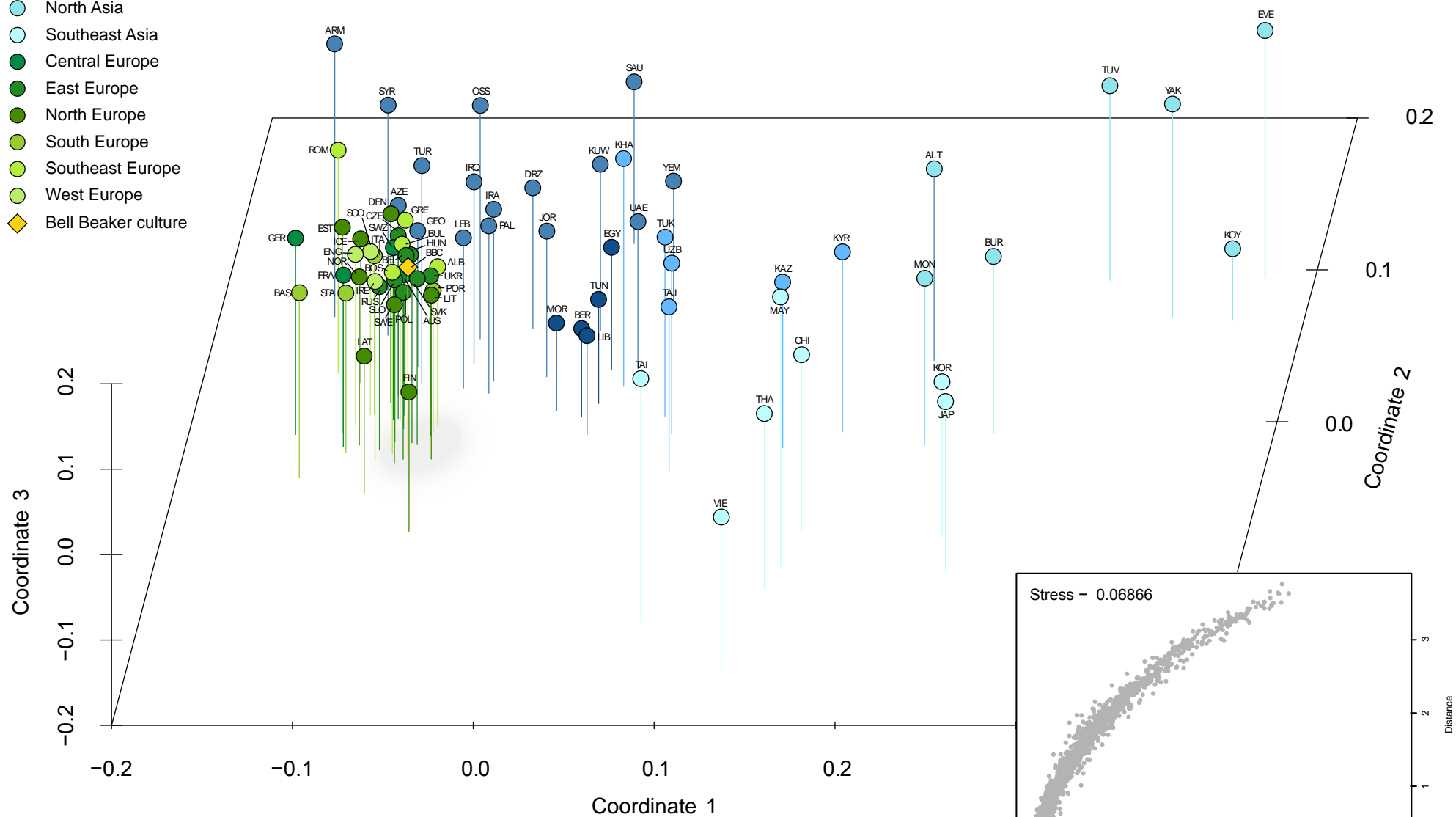
**G**

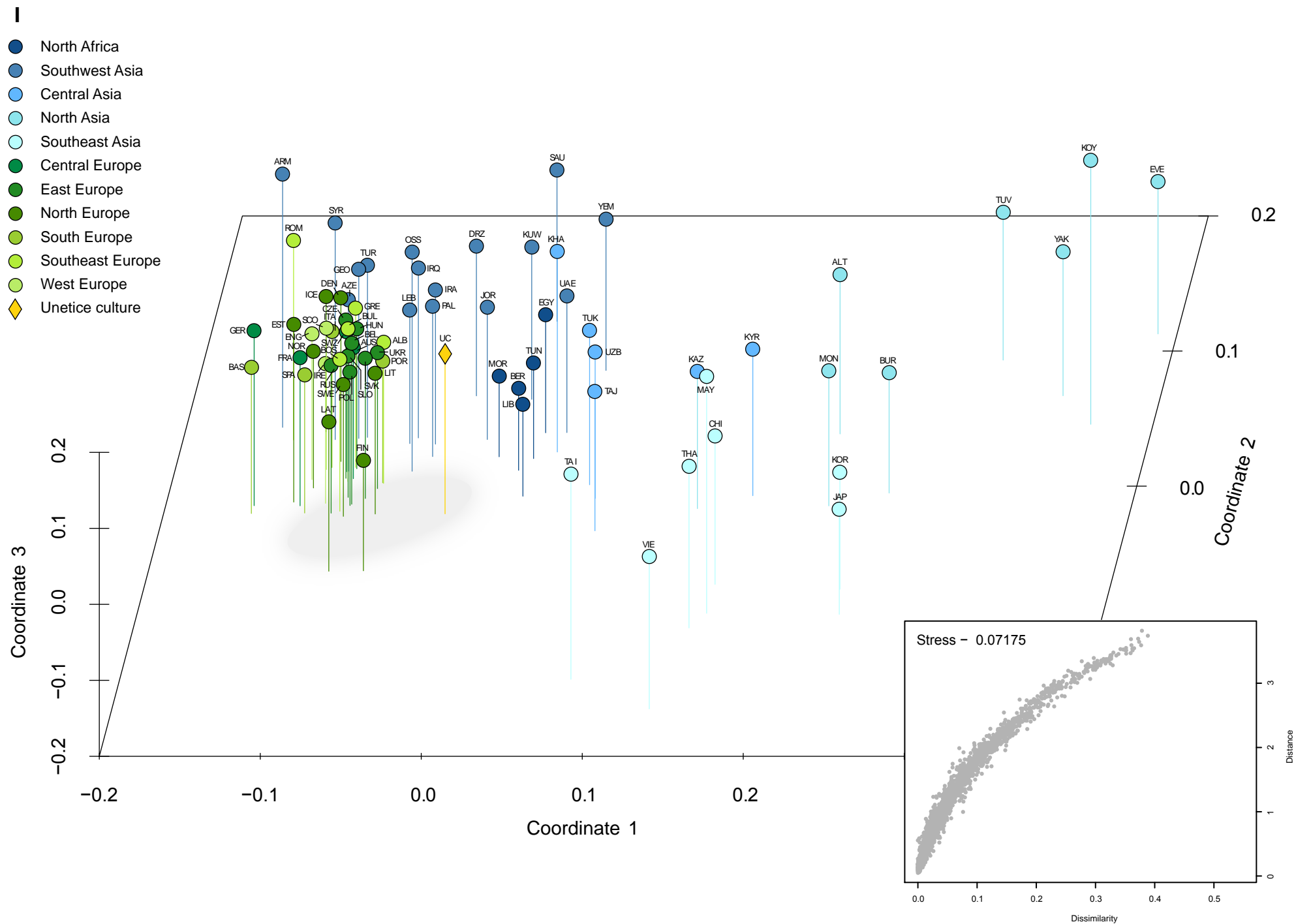
- North Africa
- Southwest Asia
- Central Asia
- North Asia
- Southeast Asia
- Central Europe
- East Europe
- North Europe
- South Europe
- Southeast Europe
- West Europe
- Corded Ware culture



H

- North Africa
- Southwest Asia
- Central Asia
- North Asia
- Southeast Asia
- Central Europe
- East Europe
- North Europe
- South Europe
- Southeast Europe
- West Europe
- ◆ Bell Beaker culture





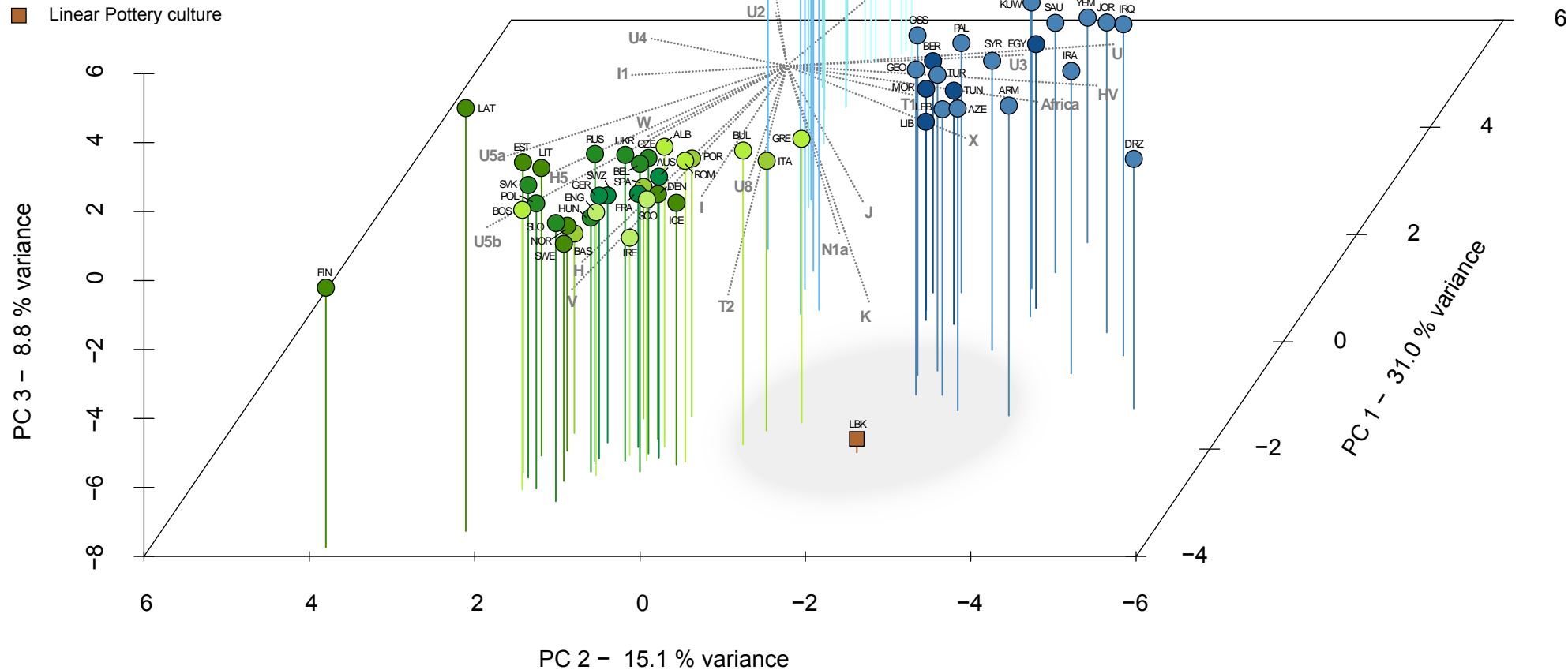
**Fig. S4A-I. Multidimensional scaling plots of prehistoric Mittelbe-Saale cultures and 73 present-day populations**

Genetic distances between Mittelbe-Saale cultures and 73 present-day populations of Europe, Asia, and North Africa were computed based on HVS-I sequences and visualized in a three dimensional space by multidimensional scaling (MDS) including Shepard diagram and associated stress value: A: Linear Pottery culture (LBK), B: Rössen culture (RSC), C: Schöningen group (SCG), D: Baalberge culture (BAC), E: Salzmünde culture (SMC), F: Bernburg culture (BEC), G: Corded Ware culture (CWC), H: Bell Beaker culture (BBC), and I: Unetice culture (UC). Color shading of data points indicates populations from different regions in Europe, Asia and Africa. Interestingly, the distribution of modern-day populations is consistent with geography, providing confidence about the plotting of ancient cultures within the present-day mitochondrial diversity. Grey circles display the closest present-day populations to the ancient sample sets. Population information, abbreviations, and  $F_{ST}$ -values are listed in Table S10.



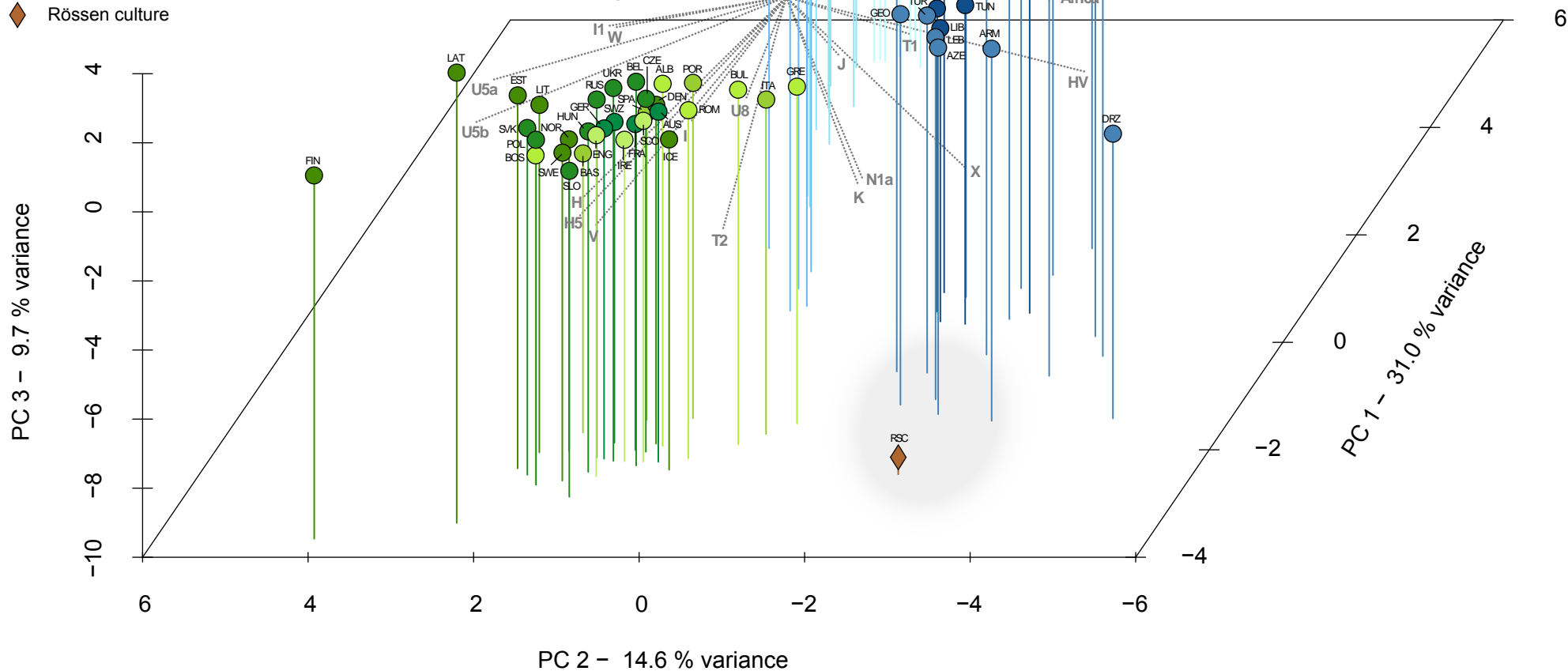
**A**

- North Africa
- Southwest Asia
- Central Asia
- North Asia
- Southeast Asia
- Central Europe
- East Europe
- North Europe
- South Europe
- Southeast Europe
- West Europe
- Linear Pottery culture



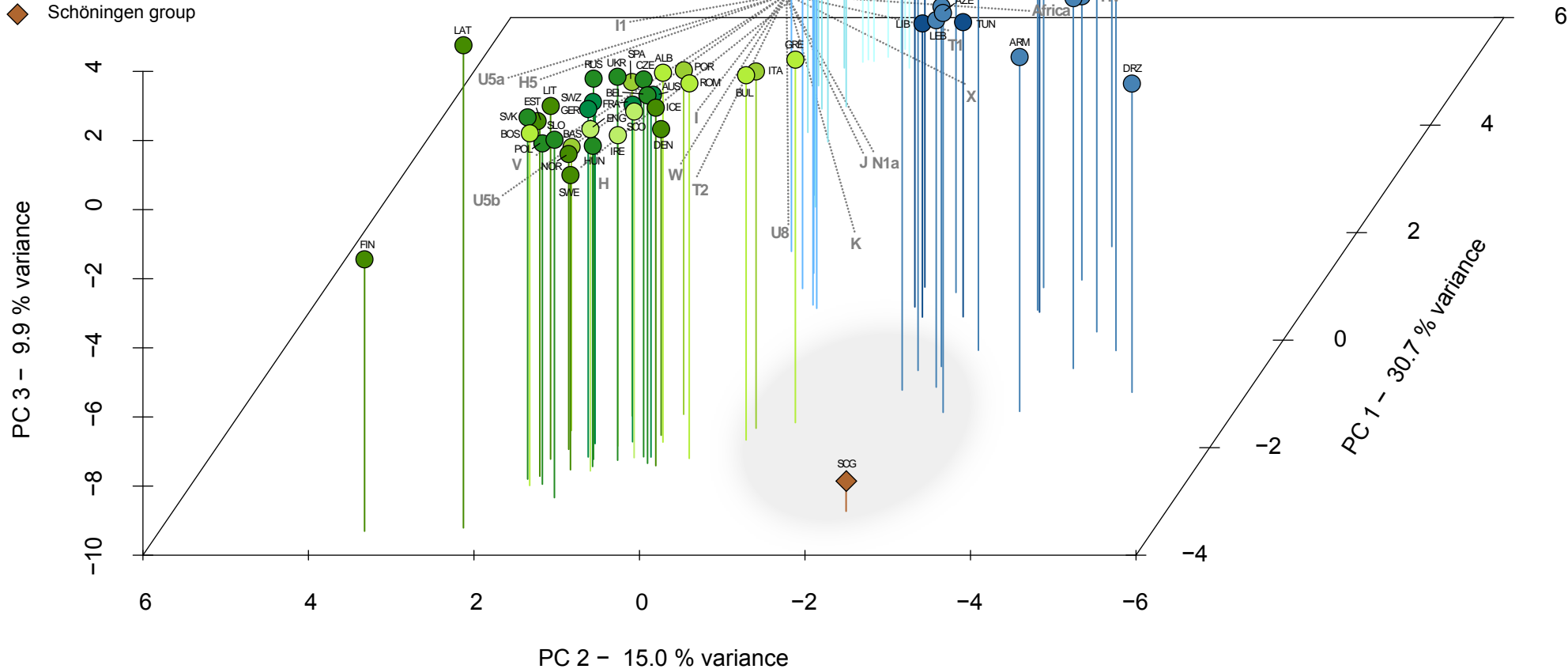
**B**

- North Africa
- Southwest Asia
- Central Asia
- North Asia
- Southeast Asia
- Central Europe
- East Europe
- North Europe
- South Europe
- Southeast Europe
- West Europe
- ◆ Rössen culture



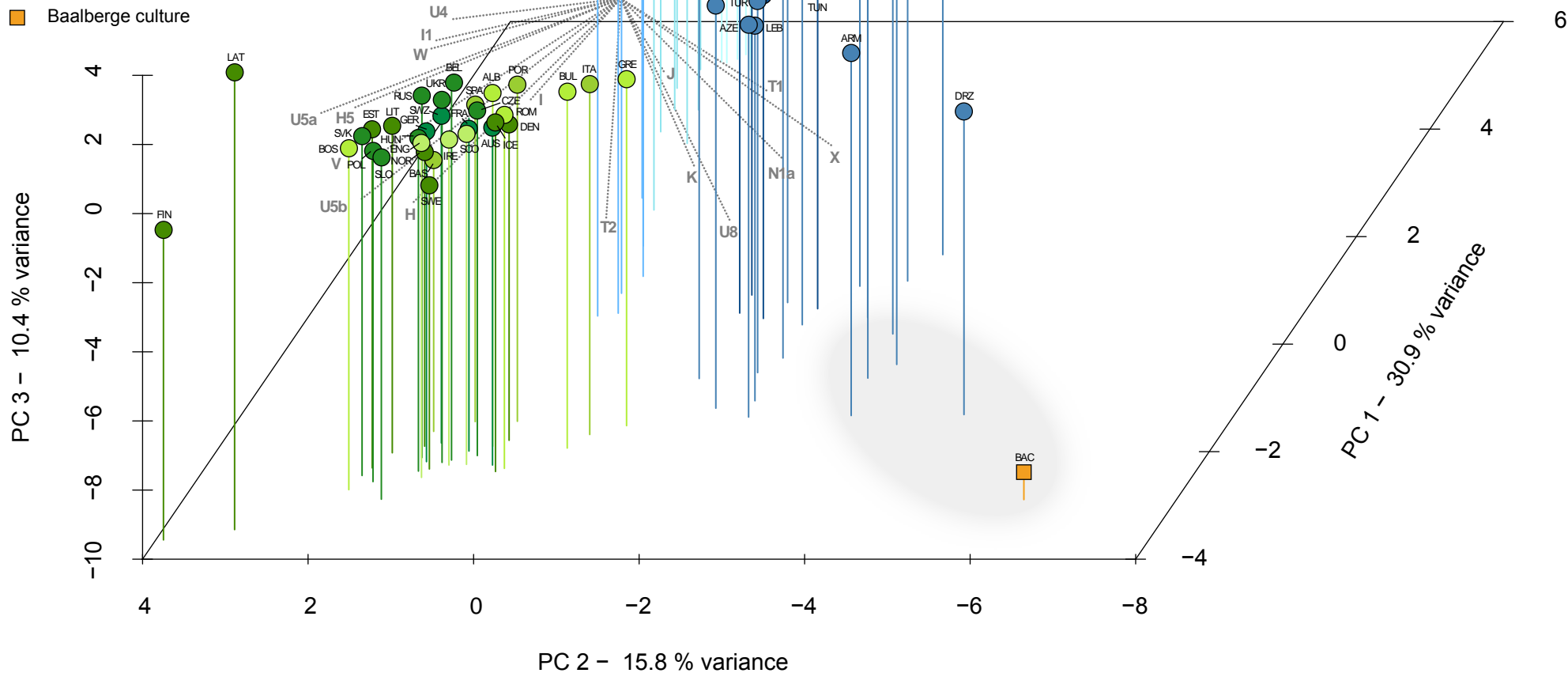
C

- North Africa
- Southwest Asia
- Central Asia
- North Asia
- Southeast Asia
- Central Europe
- East Europe
- North Europe
- South Europe
- Southeast Europe
- West Europe
- ◆ Schöningen group



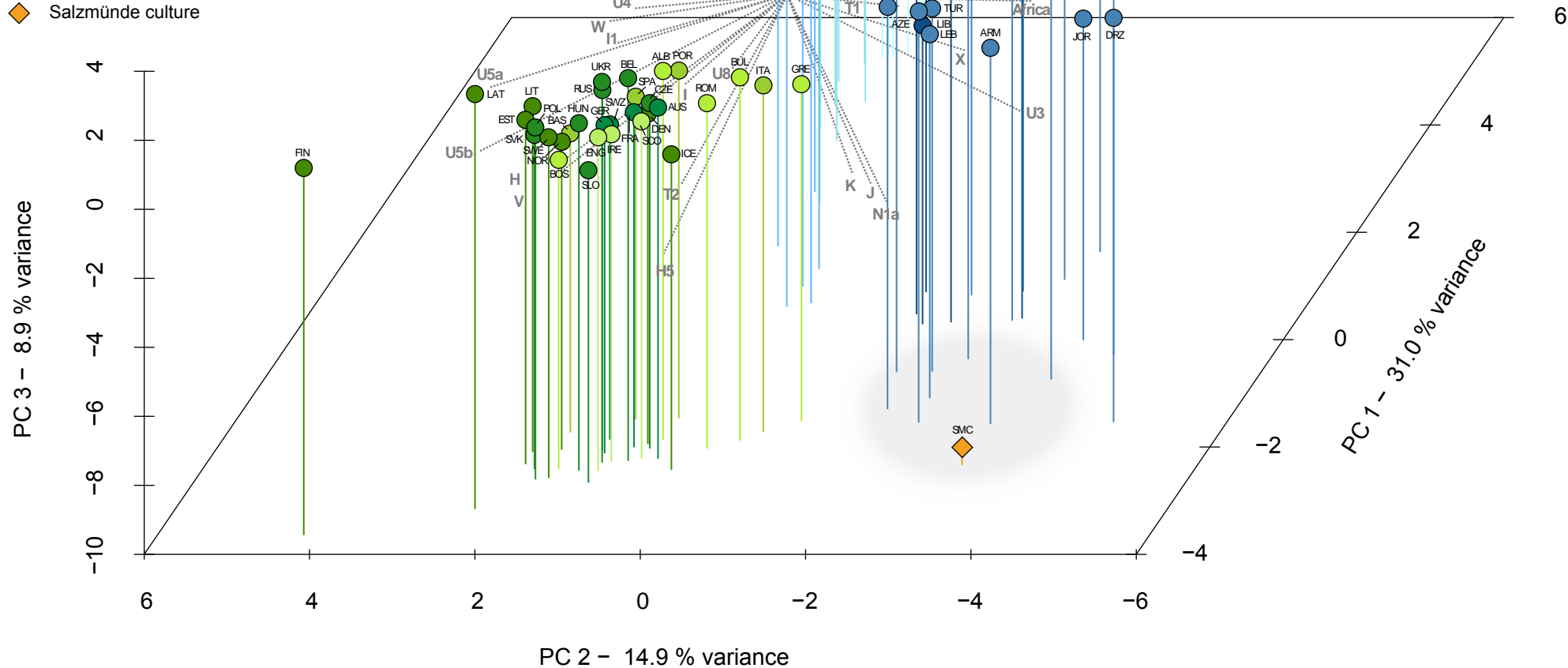
D

- North Africa
- Southwest Asia
- Central Asia
- North Asia
- Southeast Asia
- Central Europe
- East Europe
- North Europe
- South Europe
- Southeast Europe
- West Europe
- Baalberge culture



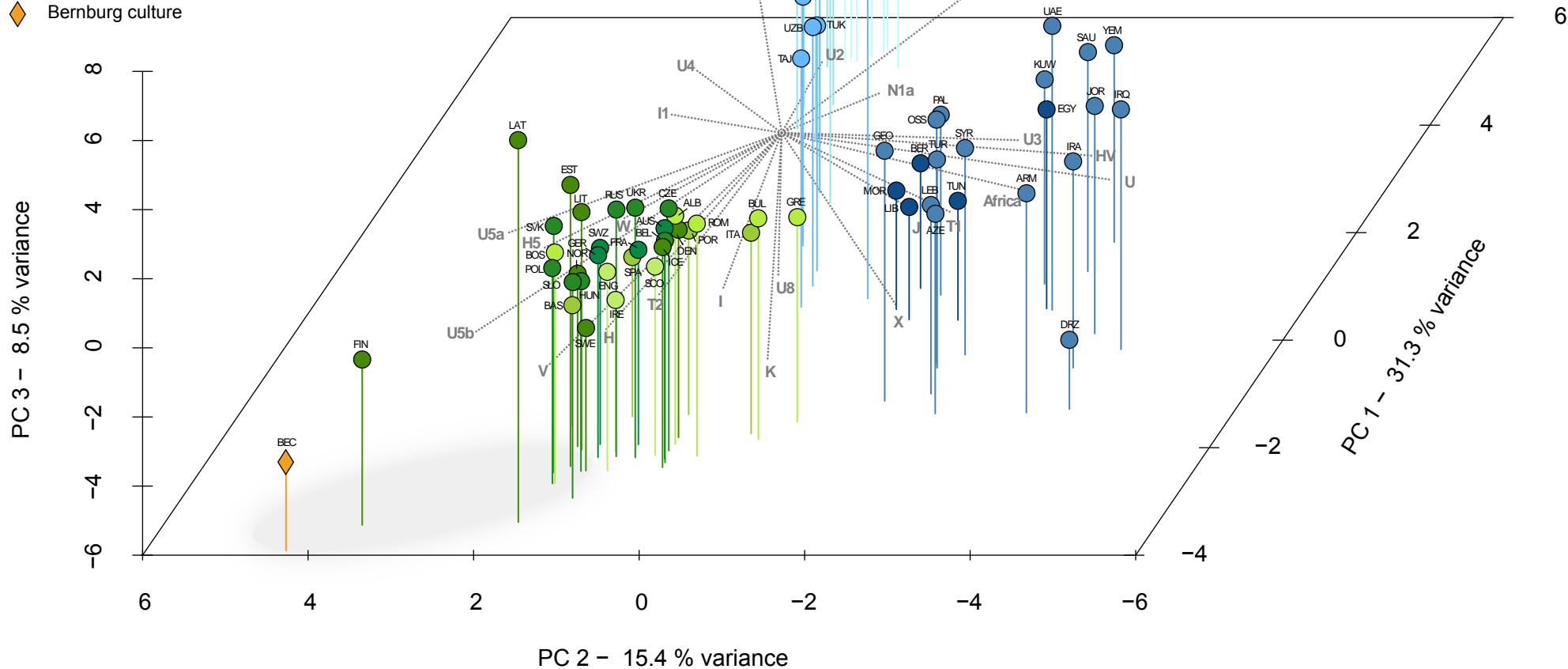
E

- North Africa
- Southwest Asia
- Central Asia
- North Asia
- Southeast Asia
- Central Europe
- East Europe
- North Europe
- South Europe
- Southeast Europe
- West Europe
- ◆ Salzmünde culture



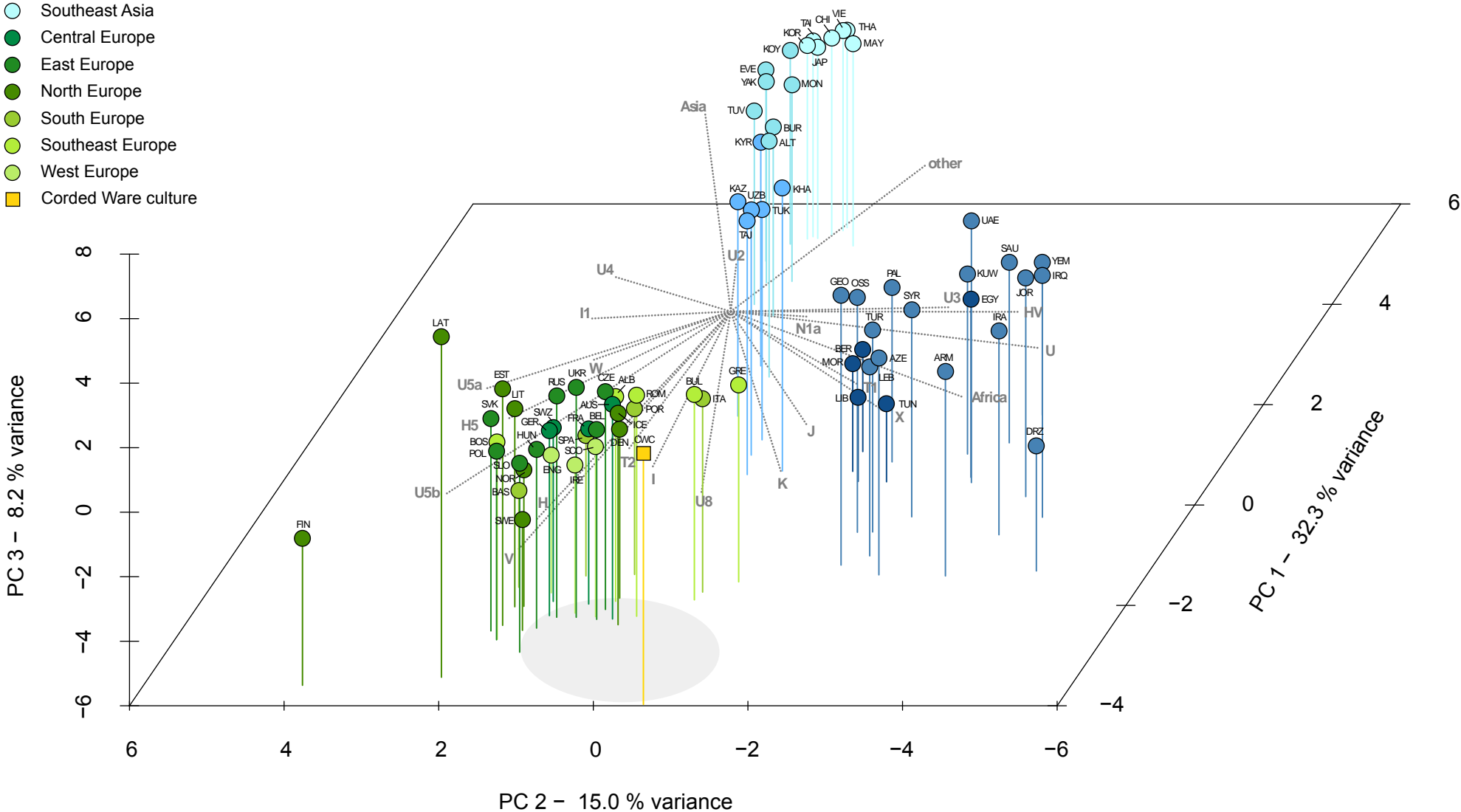
**F**

- North Africa
- Southwest Asia
- Central Asia
- North Asia
- Southeast Asia
- Central Europe
- East Europe
- North Europe
- South Europe
- Southeast Europe
- West Europe
- ◆ Bernburg culture



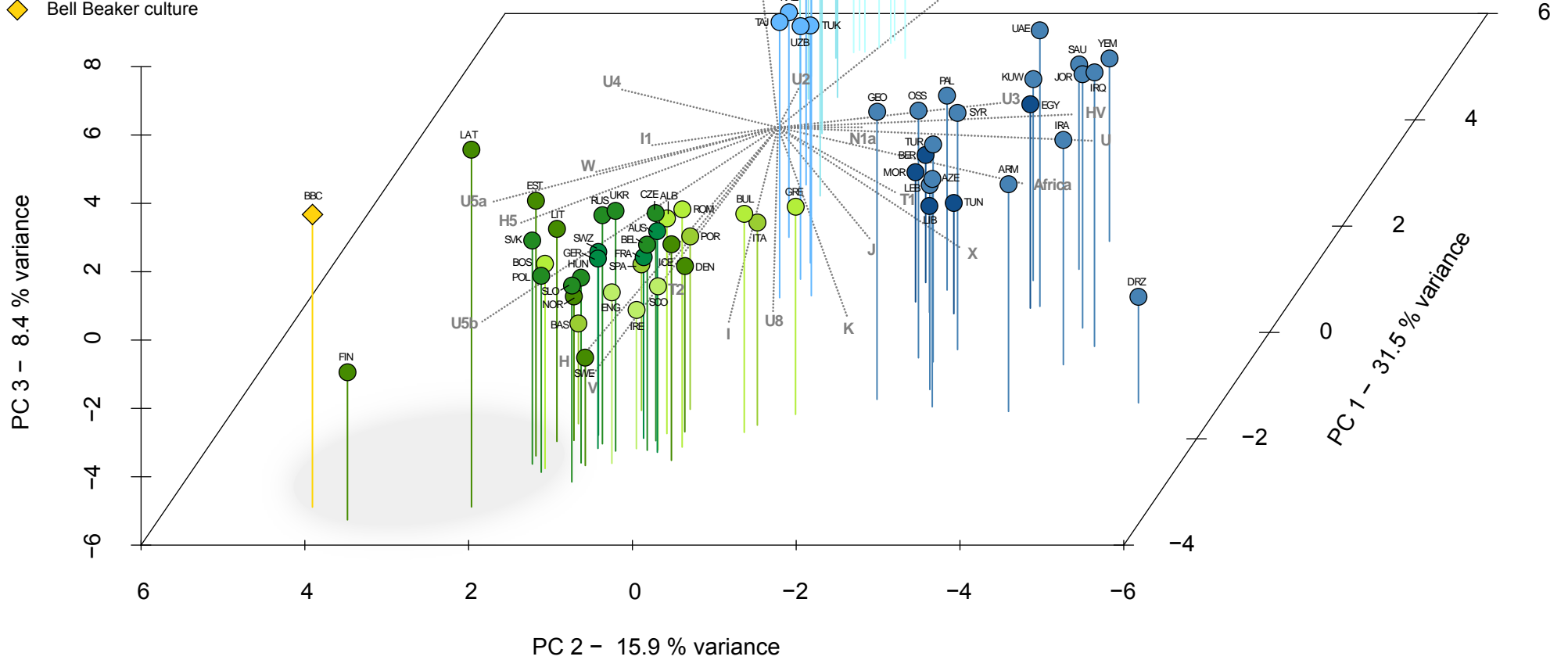
**G**

- North Africa
- Southwest Asia
- Central Asia
- North Asia
- Southeast Asia
- Central Europe
- East Europe
- North Europe
- South Europe
- Southeast Europe
- West Europe
- Corded Ware culture

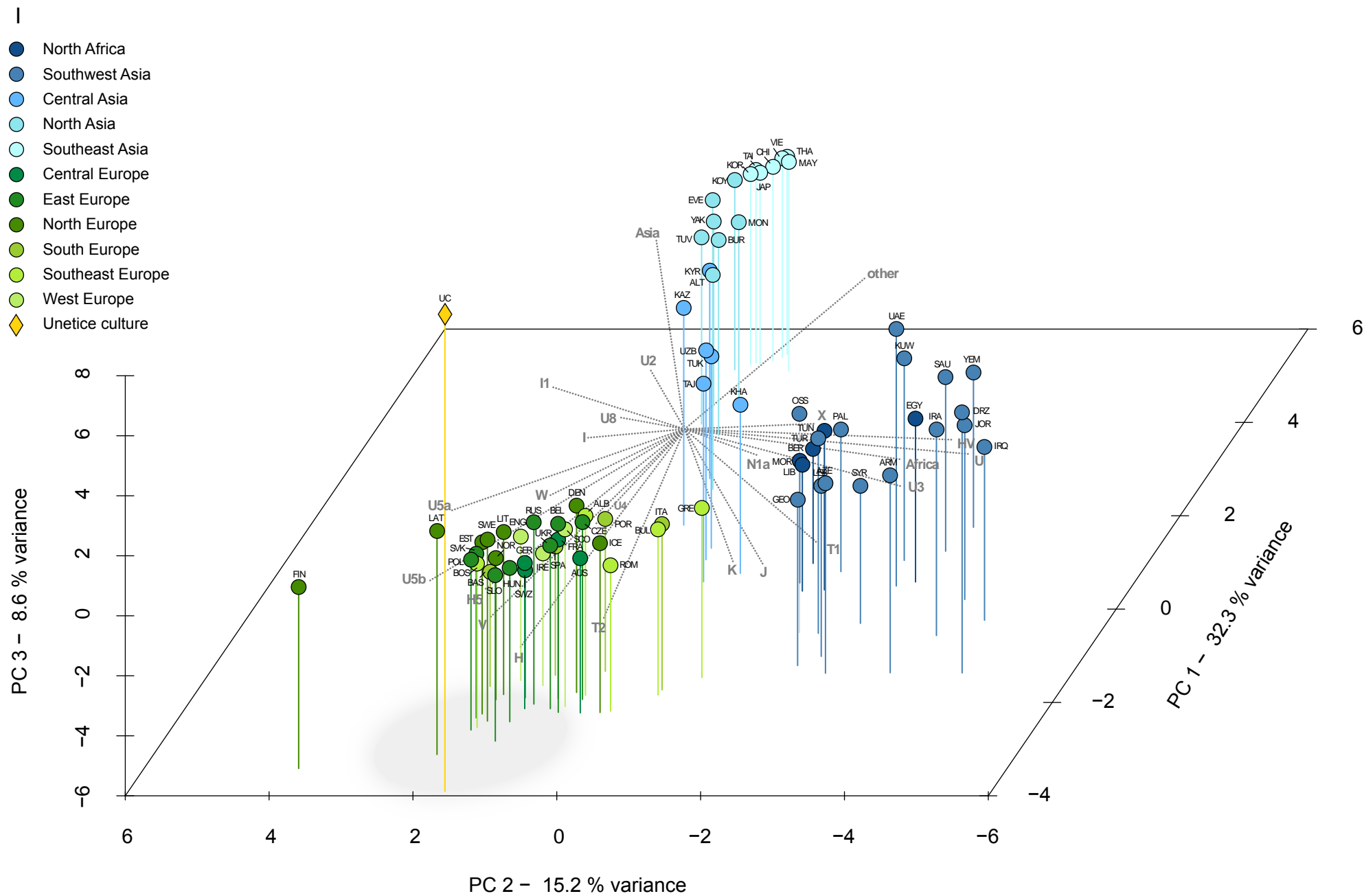


H

- North Africa
- Southwest Asia
- Central Asia
- North Asia
- Southeast Asia
- Central Europe
- East Europe
- North Europe
- South Europe
- Southeast Europe
- West Europe
- Bell Beaker culture



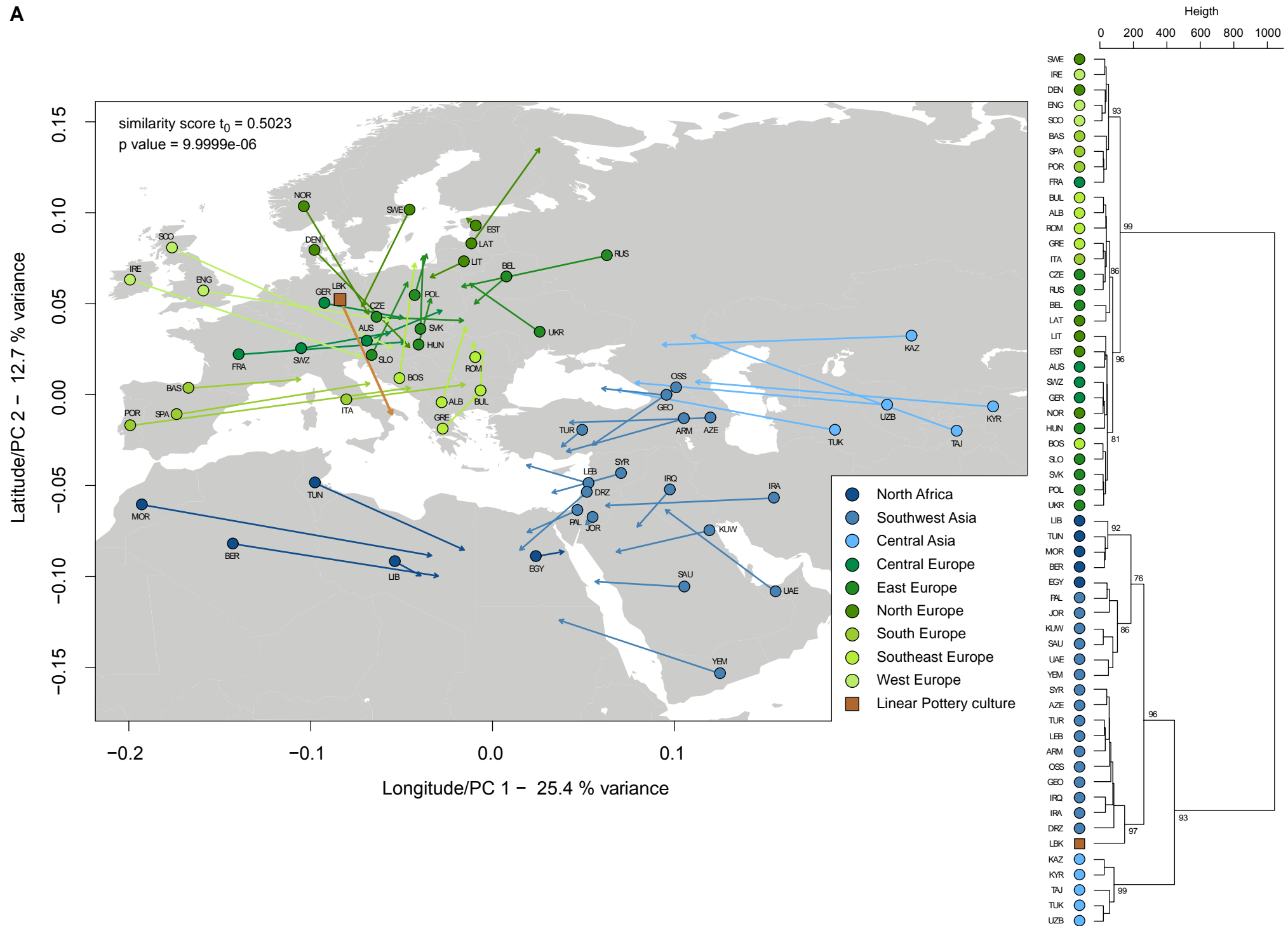




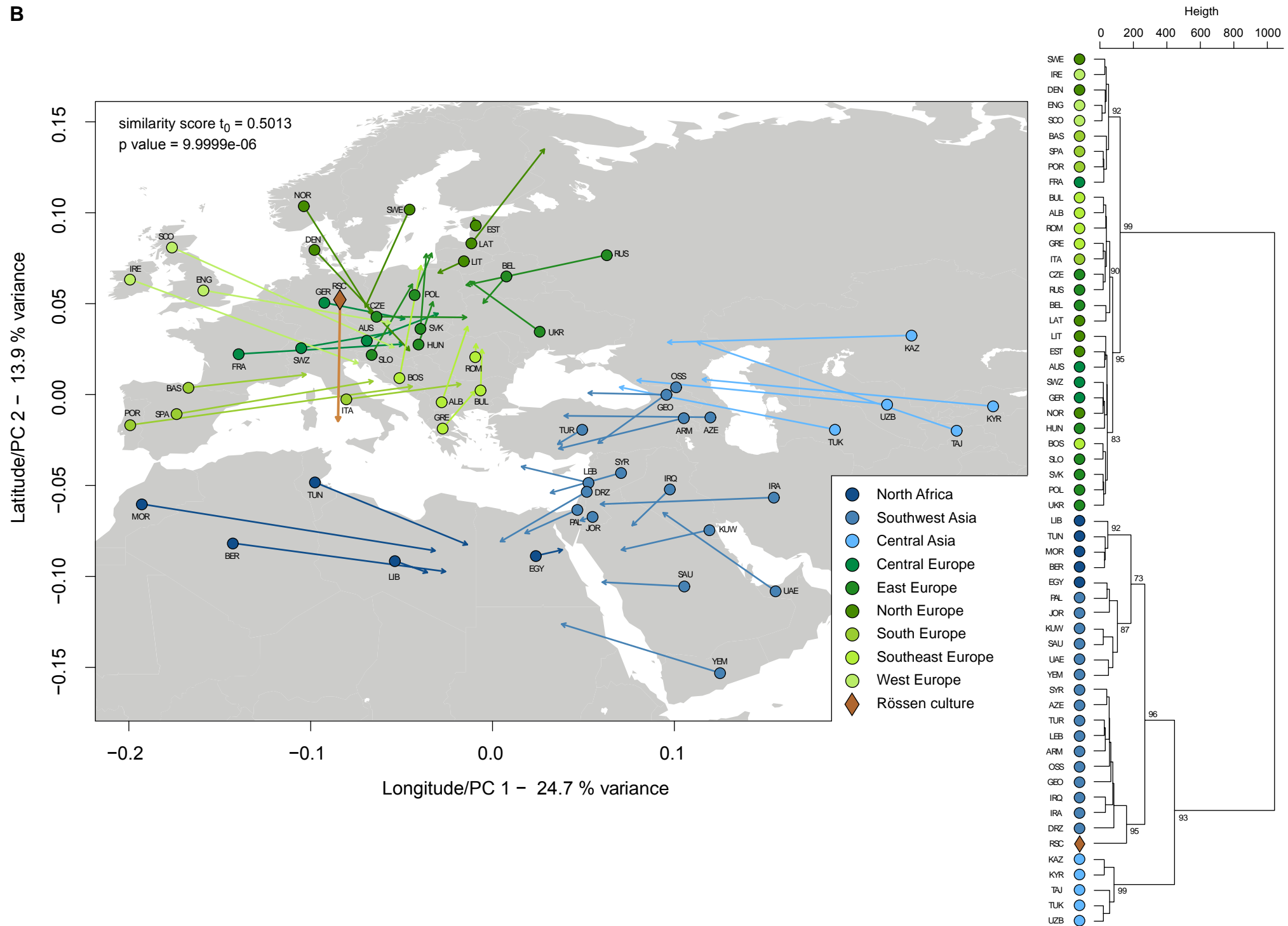
### **Figure S5A-I. Principal component analysis plots of prehistoric Mittelbe-Saale cultures and 73 present-day populations**

Principal component analysis (PCA) based on 23 mitochondrial haplogroup frequencies of the Mittelbe-Saale cultures and 73 present-day populations of Europe, Asia, and North Africa was plotted in a three dimensional space displaying the first three principal components: A: Linear Pottery culture (LBK), B: Rössen culture (RSC), C: Schöningen group (SCG), D: Baalberge culture (BAC), E: Salzmünde culture (SMC), F: Bernburg culture (BEC), G: Corded Ware culture (CWC), H: Bell Beaker culture (BBC), and I: Unetice culture (UC). The contribution of each haplogroup is superimposed as grey component loading vector. Color shadings of data points indicate populations from different regions in Europe, Asia, and Africa. The first three principal components account for 54.9% (LBK), 55.3% (RSC), 55.6% (SCG), 57.1% (BAC), 54.8% (SMC), 55.2% (BEC), 55.5% (CWC), 55.8% (BBC), and 56.1% (UC) of the total genetic variation. In general, the first principal component clearly separates European, Near Eastern, and North African populations from Central, North, and Southeast Asia. The second principal component separates the Near Eastern and North African population from Europeans and the third principal component shows a separation of the ancient cultures from all present-day populations, which indicates their unique haplogroup compositions. Grey circles display the closest present-day populations to the ancient sample sets. Population information, abbreviations, and haplogroup frequencies are listed in Table S11.

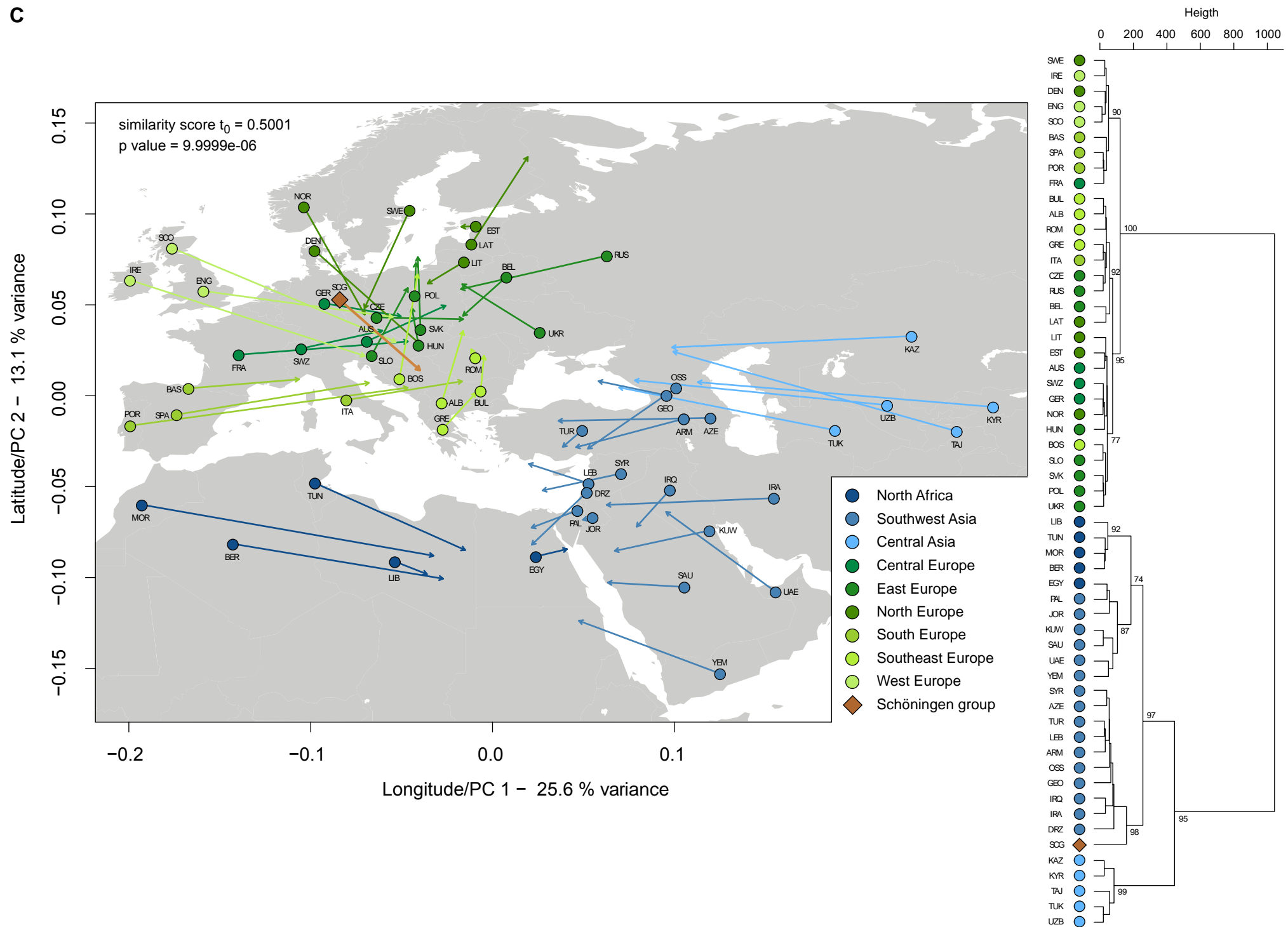
A



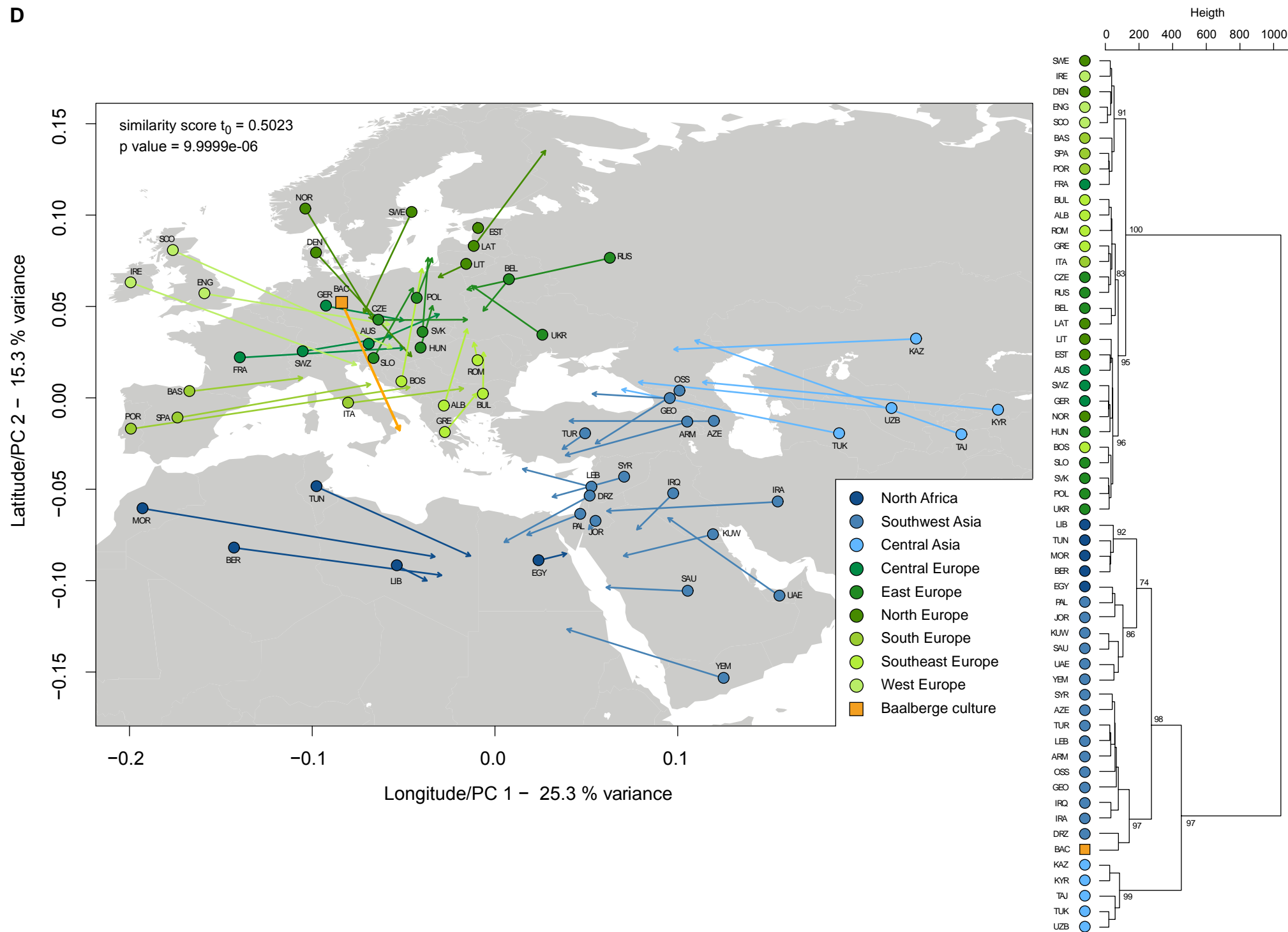
B



C

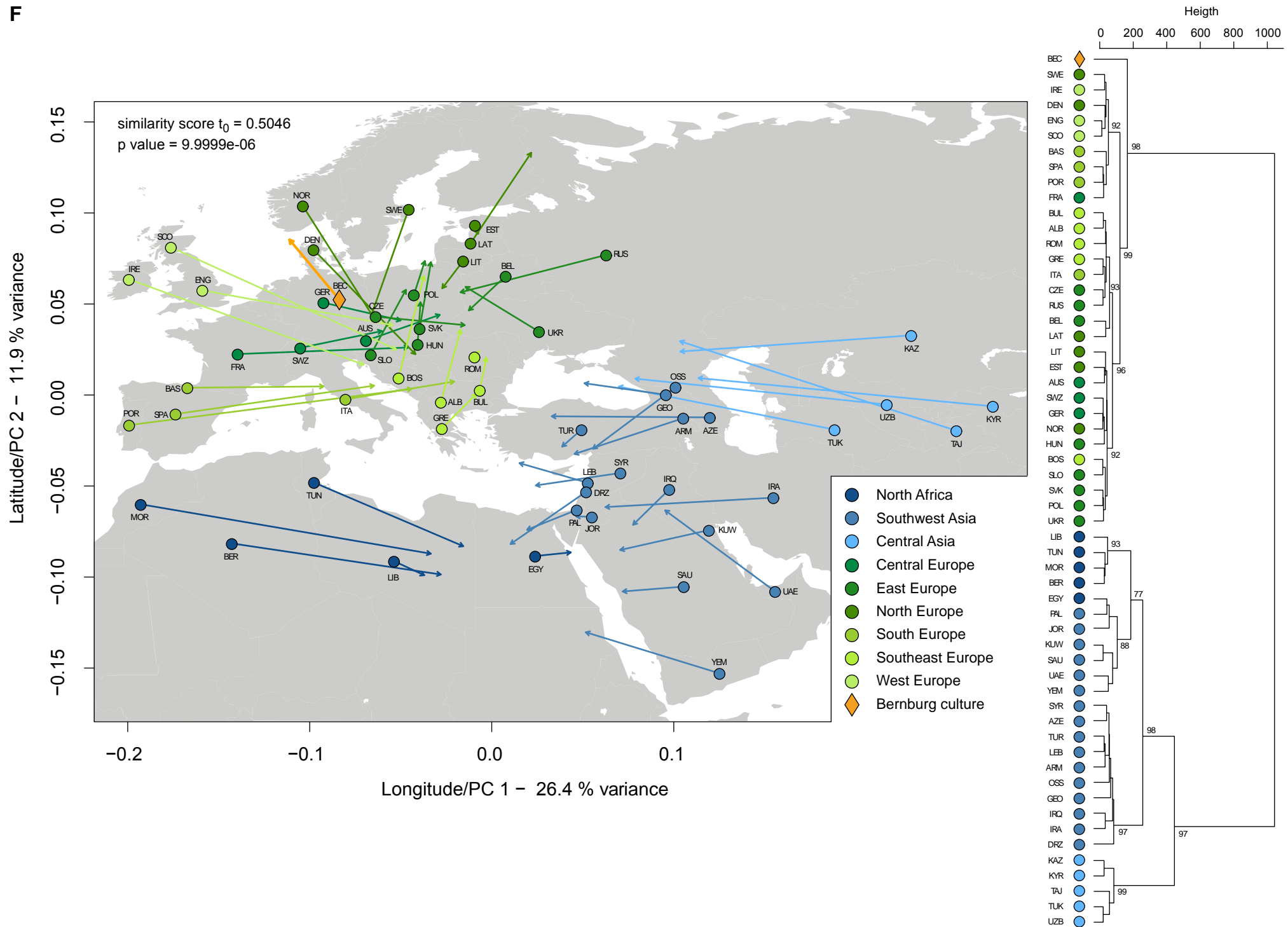


D



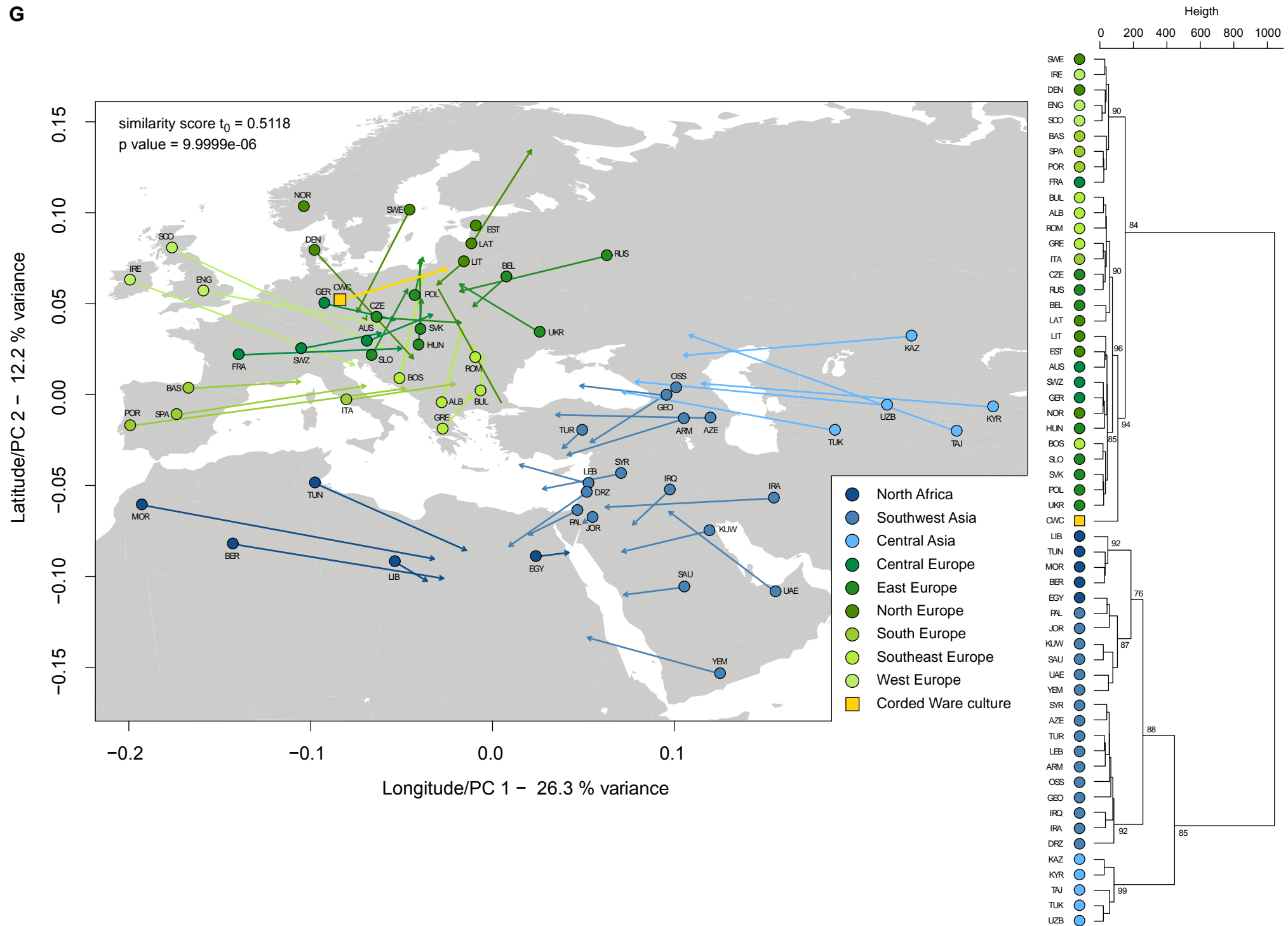


Latitude/PC 2 - 11.9 % variance

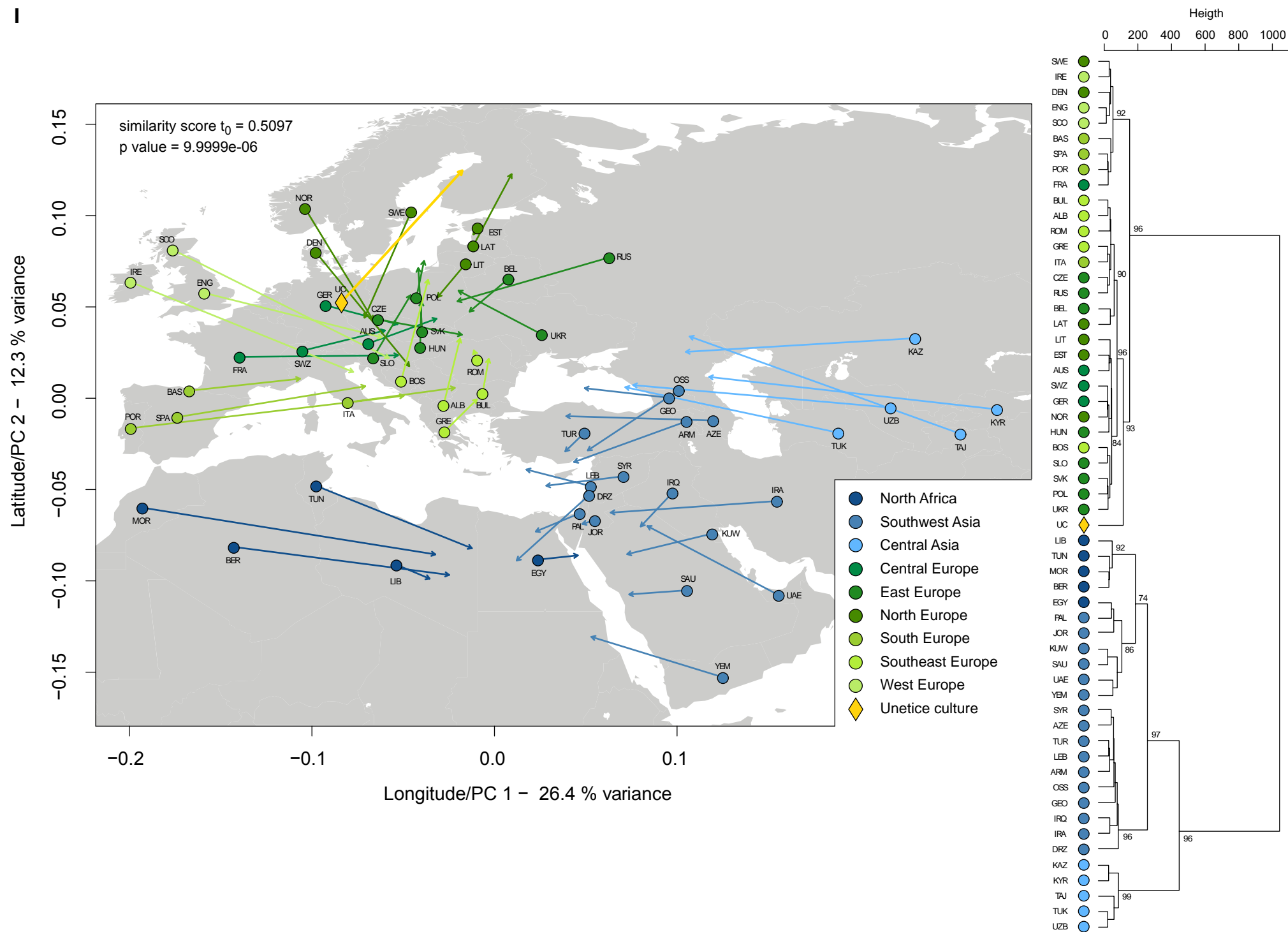




G

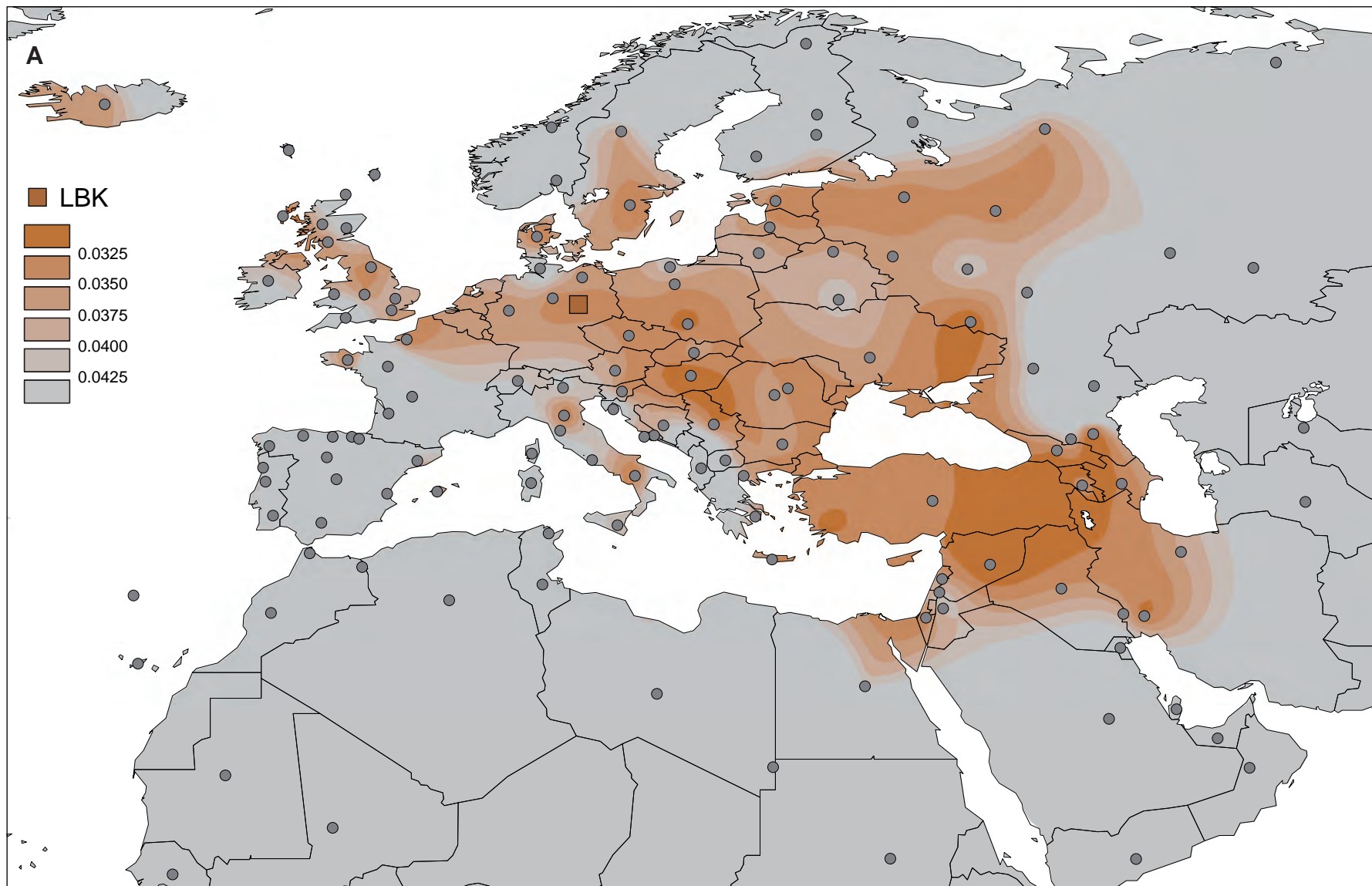


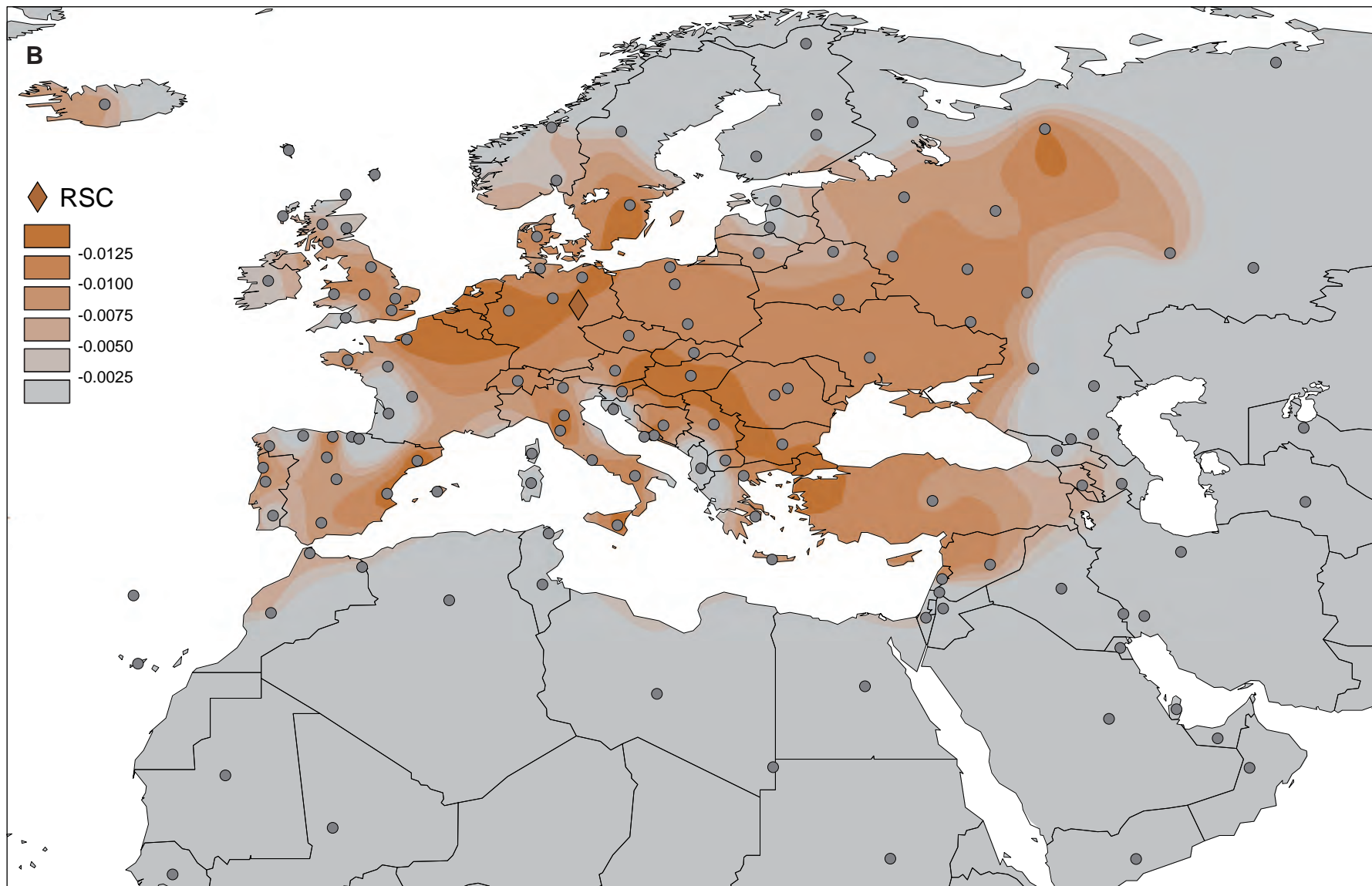




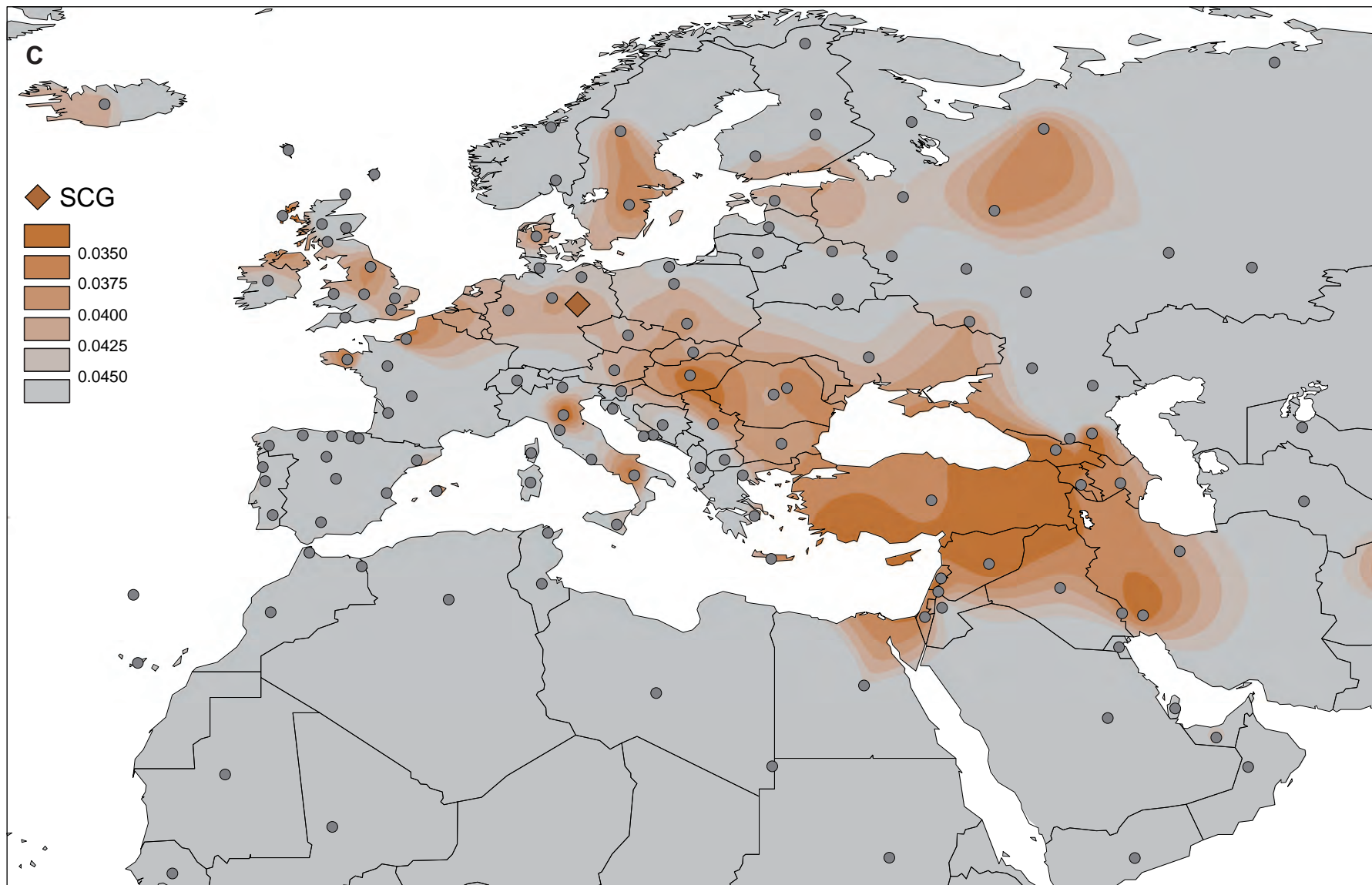
**Fig. S6A-I. Procrustes analyses and Ward Clustering of prehistoric Mittelbe-Saale cultures and 56 present-day populations**

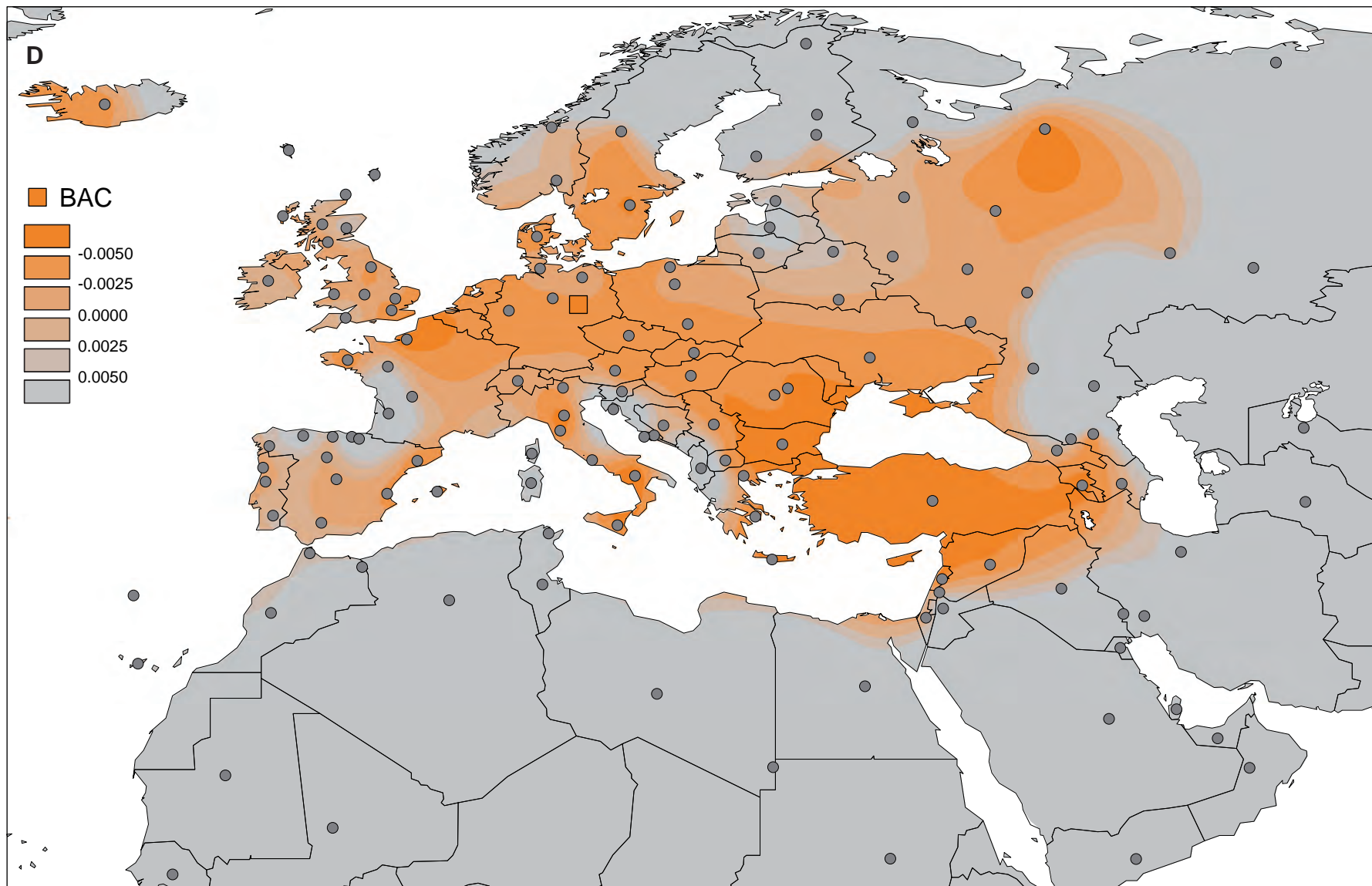
Procrustes analyses (left) and Ward clustering (right) were carried out based on 23 haplogroup frequencies of the Mittelbe-Saale cultures and 56 present-day populations of Europe, Asia, and North Africa: A: Linear Pottery culture (LBK), B: Rössen culture (RSC), C: Schöningen group (SCG), D: Baalberge culture (BAC), E: Salzmünde culture (SMC), F: Bernburg culture (BEC), G: Corded Ware culture (CWC), H: Bell Beaker culture (BBC), and I: Unetice culture (UC). For Procrustes analyses the haplogroup frequencies were used for an initial PCA and PCA scores of the first and second principal component were rotated to the best fit with the geographic coordinates of the populations. Procrustes analyses were tested on significance by 100,000 permutations. Color shadings of data points denote the geographic location of populations from different regions in Europe, Asia, and Africa. The vectors indicate the rotation of data points according to their mtDNA affinities inferred from the PCA. The first two components of the PCA account for 38.1% (LBK), 38.6% (RSC), 38.7% (SCG), 40.6% (BAC), 37.9% (SMC), 38.3% (BEC), 37.5% (CWC), 39.6% (BBC), and 38.7% (UC) of the total genetic variation. The same dataset was used for Ward clustering dendrograms. Color code of data points is consistent to Procrustes analyses. p-values of the main clusters are stated in percent of reproduced cluster based on 10,000 bootstrap replicates. Ward clustering distinguished two main clusters of European and Asian/African populations that can be further differentiated largely into West Europe, Central/East Europe, North Africa, the Arabian Peninsula, the Near East/Anatolia/Caucasus, and Central Asia. Population information, abbreviations, and haplogroup frequencies are listed in Table S12.



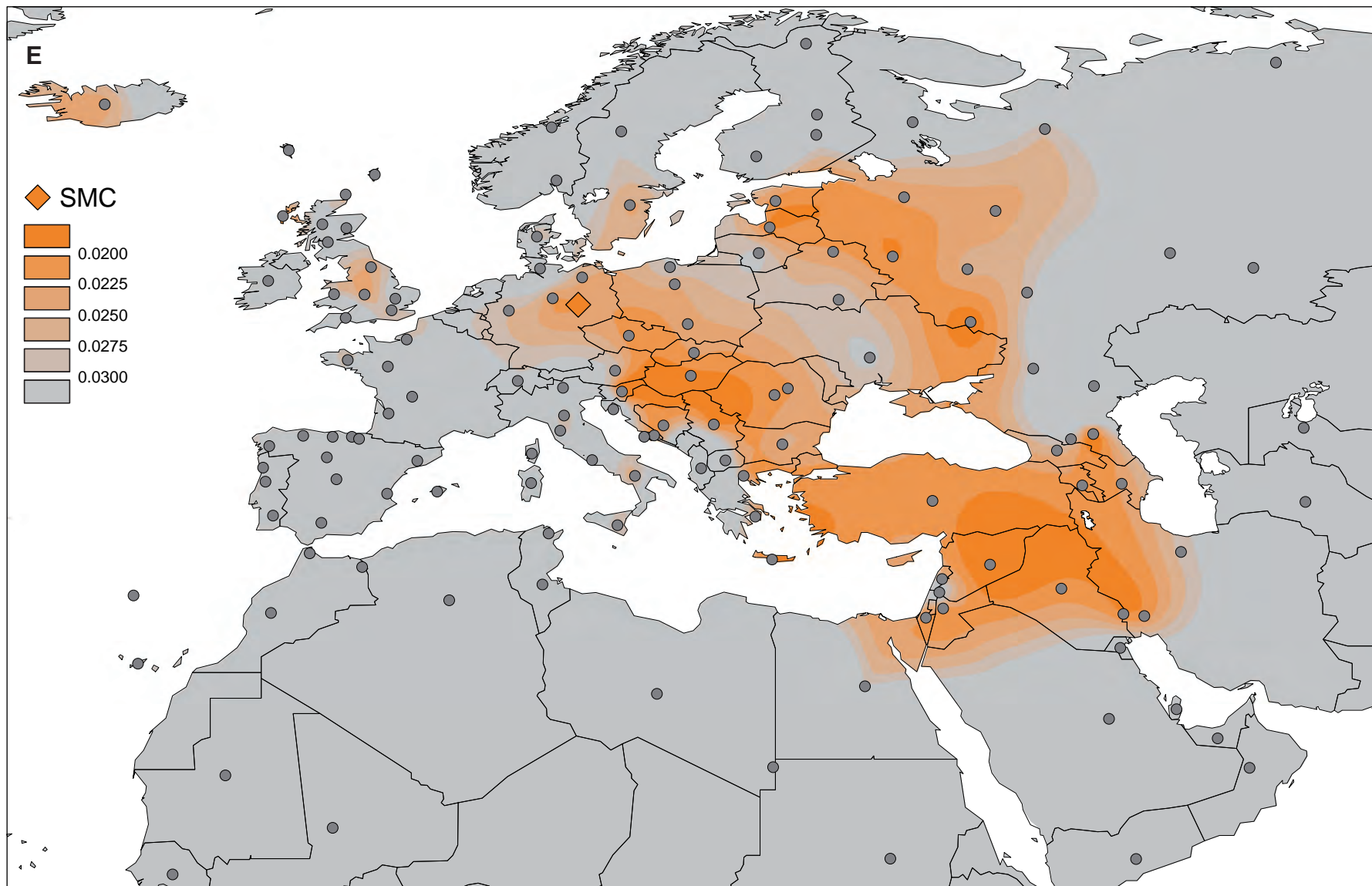


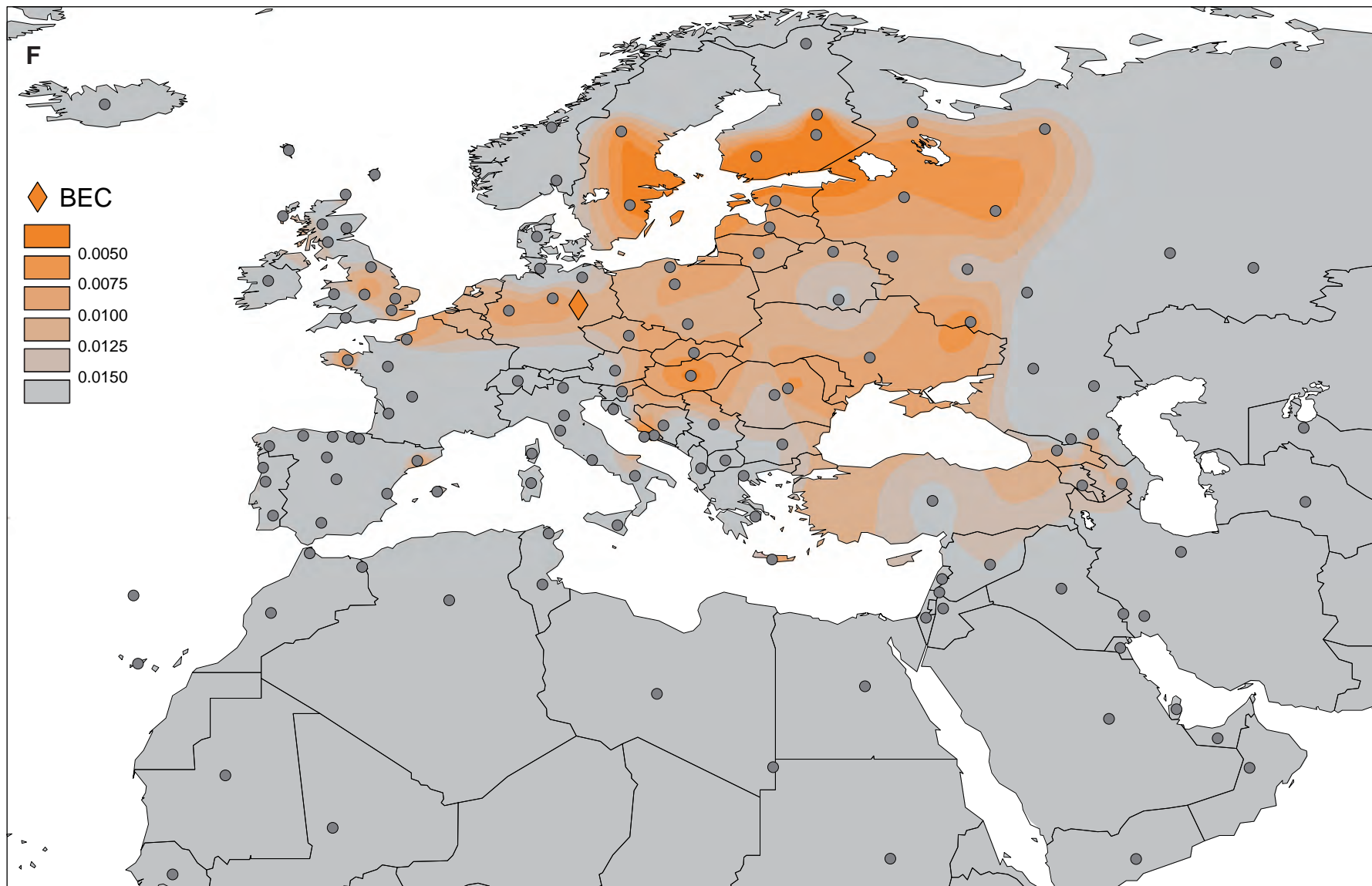


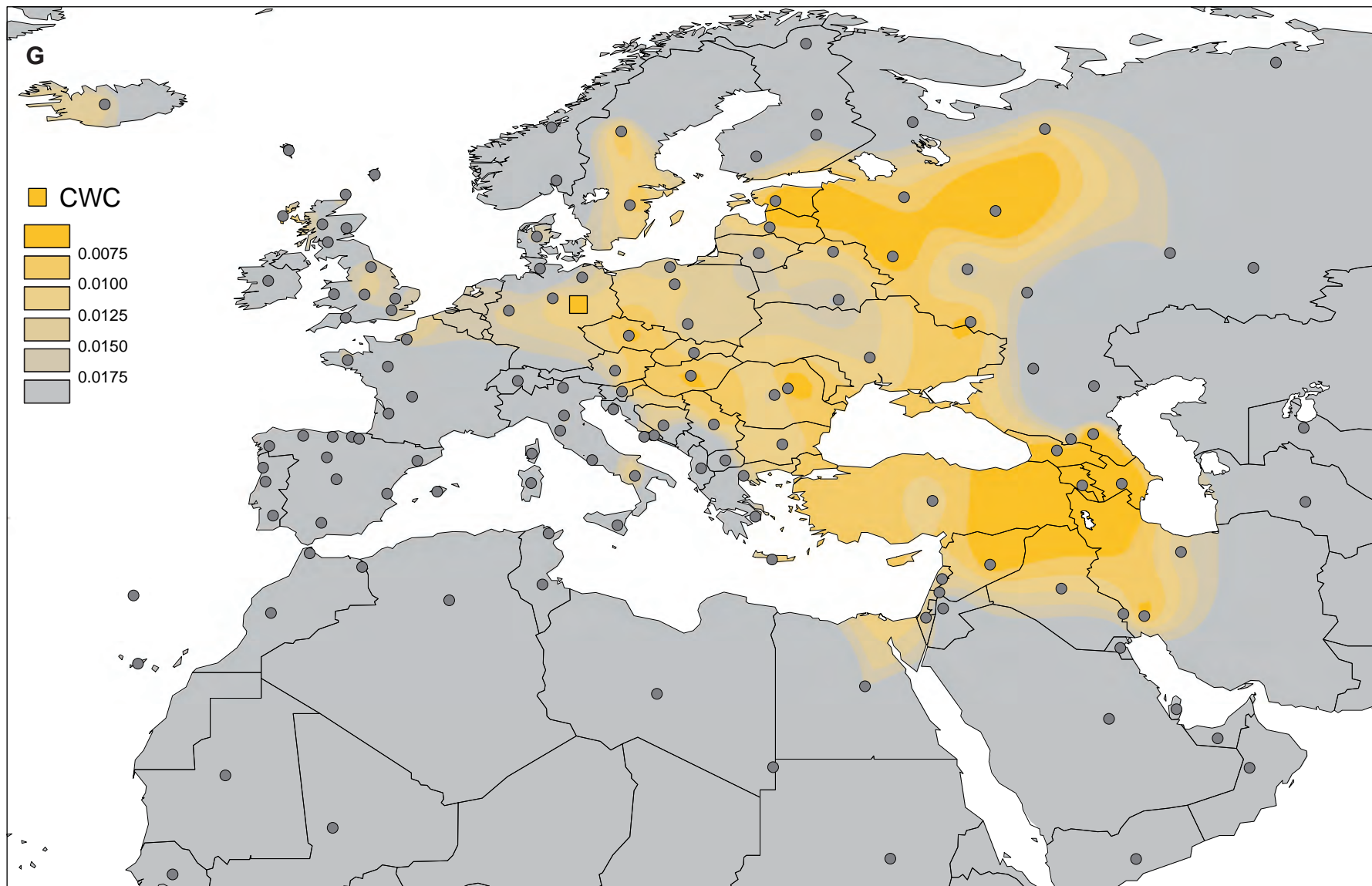


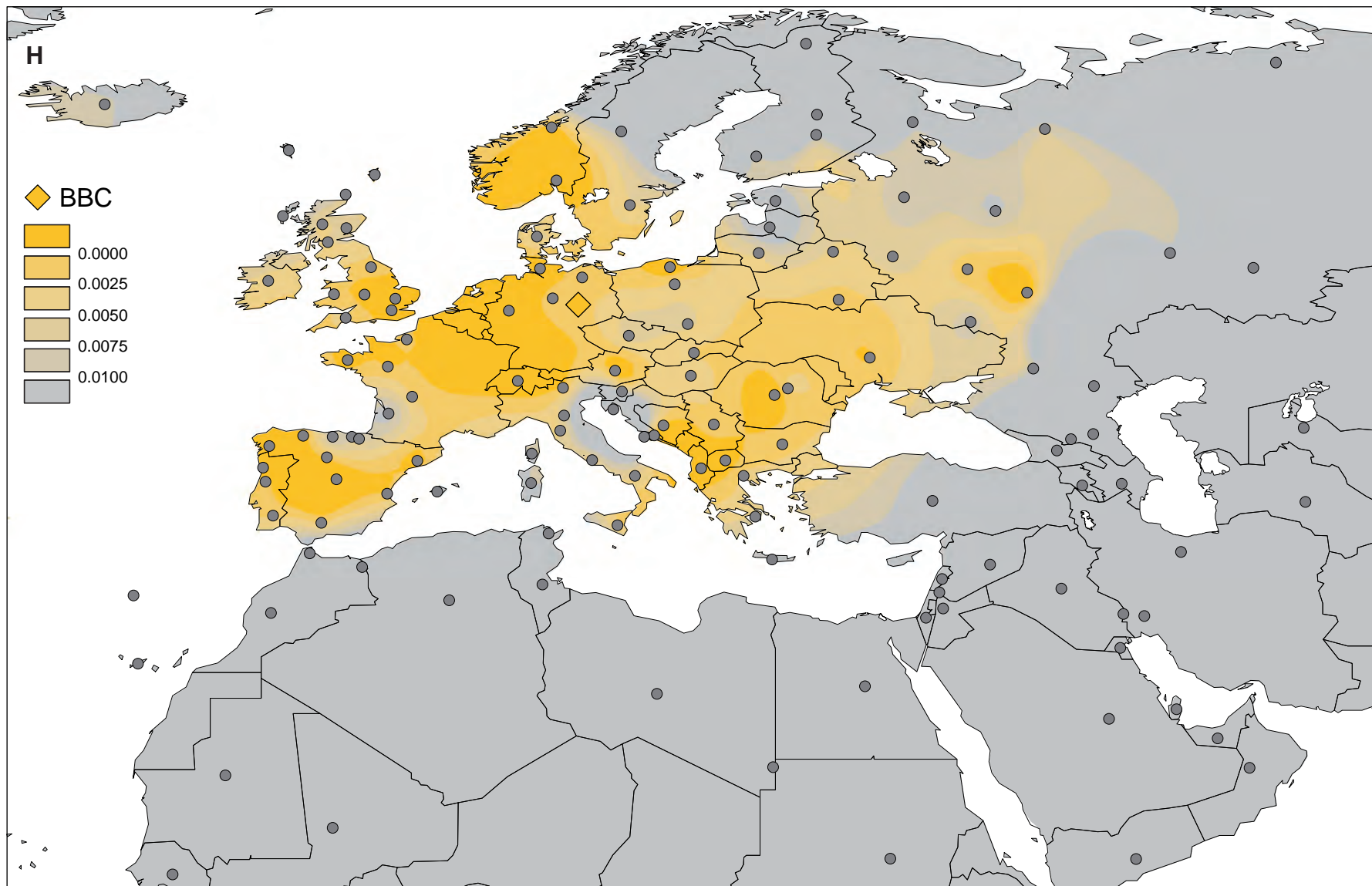




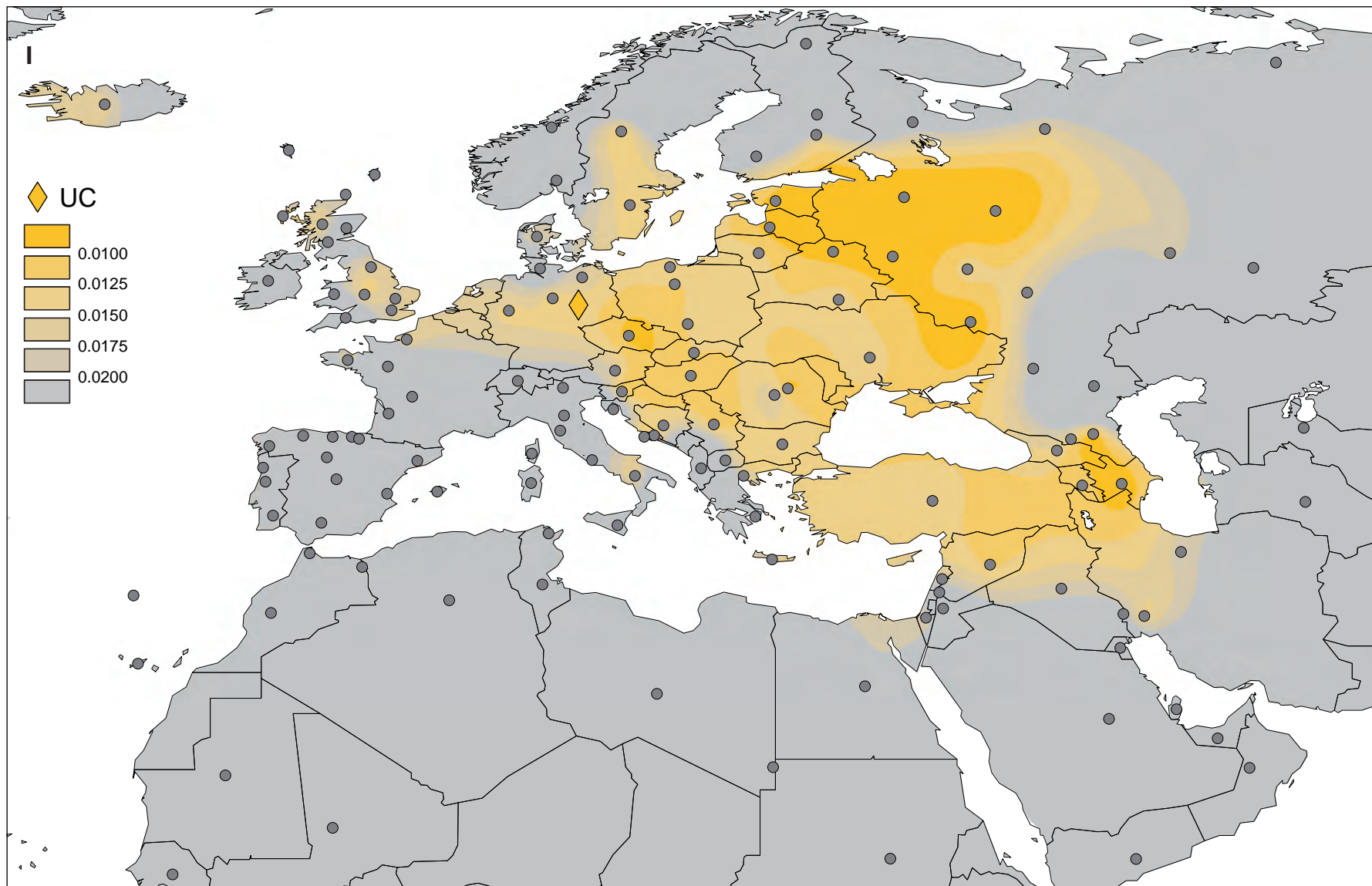






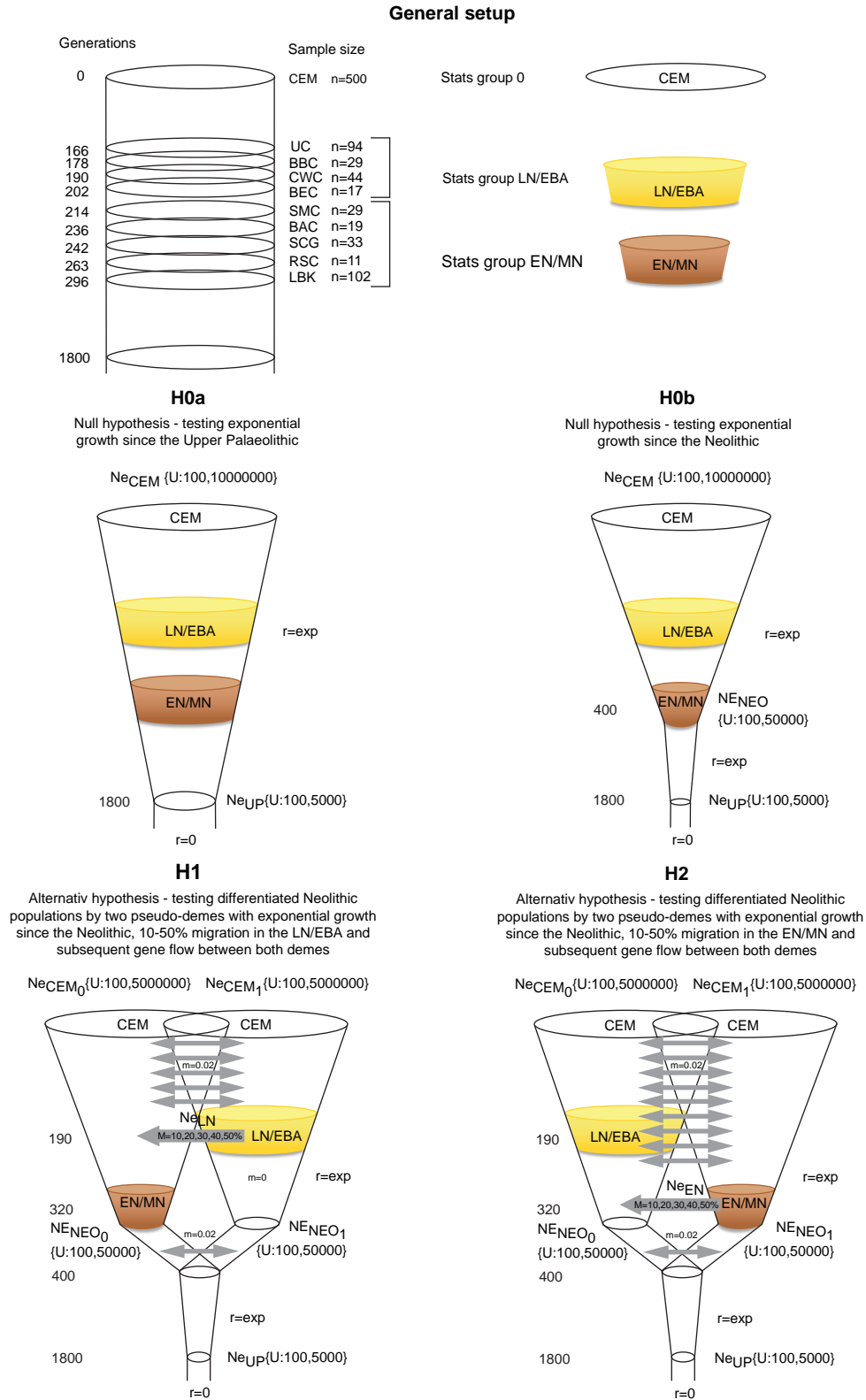






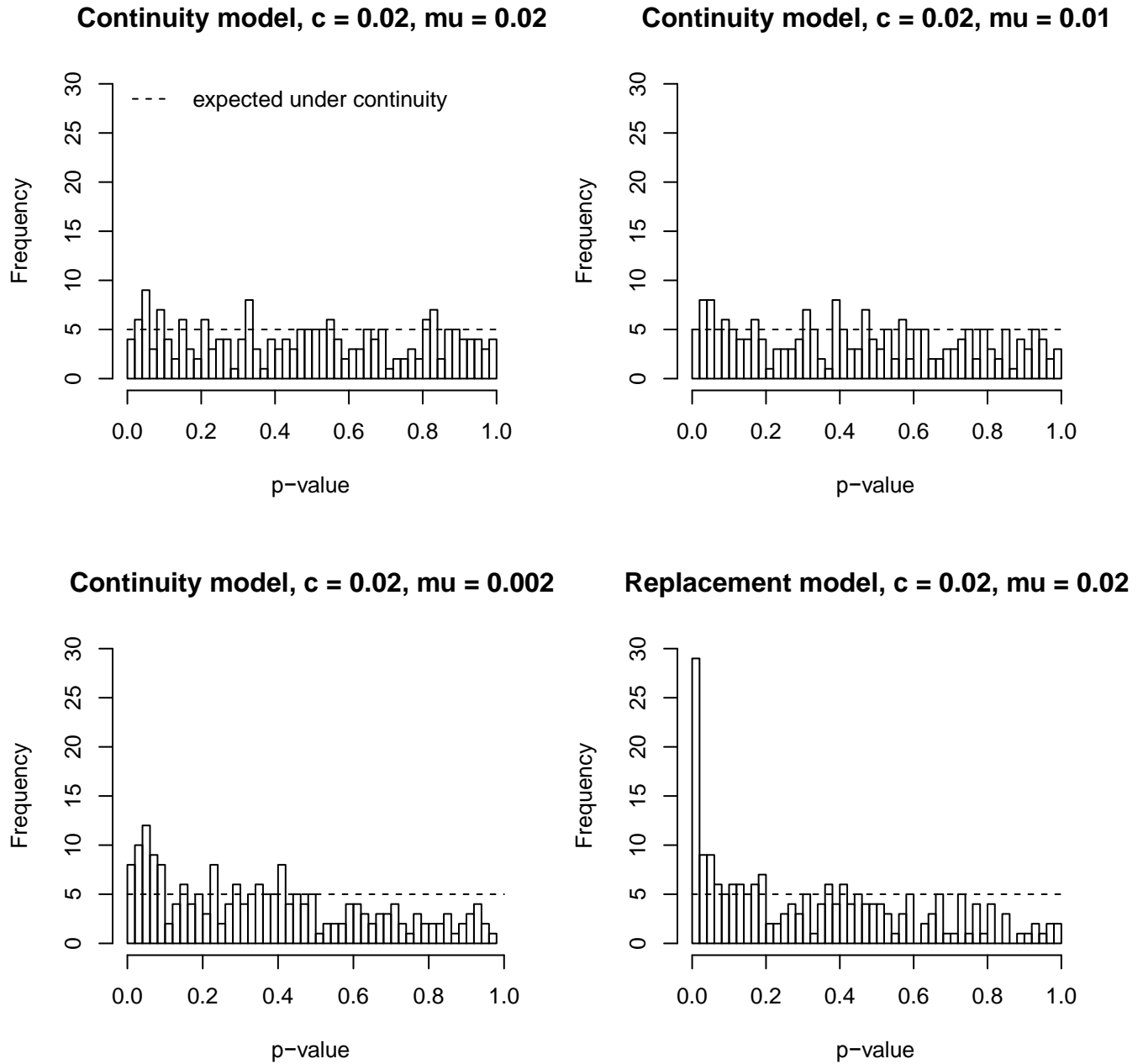
**Fig. S7A-I. Genetic distance maps of prehistoric Mittelbe-Saale cultures and 150 present-day populations**

Genetic distances of Mittelbe-Saale cultures and 150 present-day populations of Western Eurasia and North Africa were computed based on mitochondrial haplogroup frequencies and visualized by genetic distance maps (GDM): A: Linear Pottery culture (LBK), B: Rössen culture (RSC), C: Schöningen group (SCG), D: Baalberge culture (BAC), E: Salzmünde culture (SMC), F: Bernburg culture (BEC), G: Corded Ware culture (CWC), H: Bell Beaker culture (BBC), and I: Unetice culture (UC). Frequencies of 107 subhaplogroups, as assigned by distinct HVS-I motifs were considered in the analysis in order to increase the resolution on the haplogroup level. Grey dots denote the location of present-day populations. Color shadings indicate the degree of similarity or dissimilarity of the prehistoric cultures to the 150 modern populations. Short distances (i.e. greatest similarity) to present-day populations are highlighted in brown (Early Neolithic: LBK, RSC, SCG), orange (Middle Neolithic: BAC, SMC, BEC), or yellow (Late Neolithic: CWC, BBC; and Early Bronze Age: UC). Population information, abbreviations and  $F_{st}$ -values are listed in Table S13.



**Fig. S8. Demographic models tested by Bayesian Serial Simcoal**

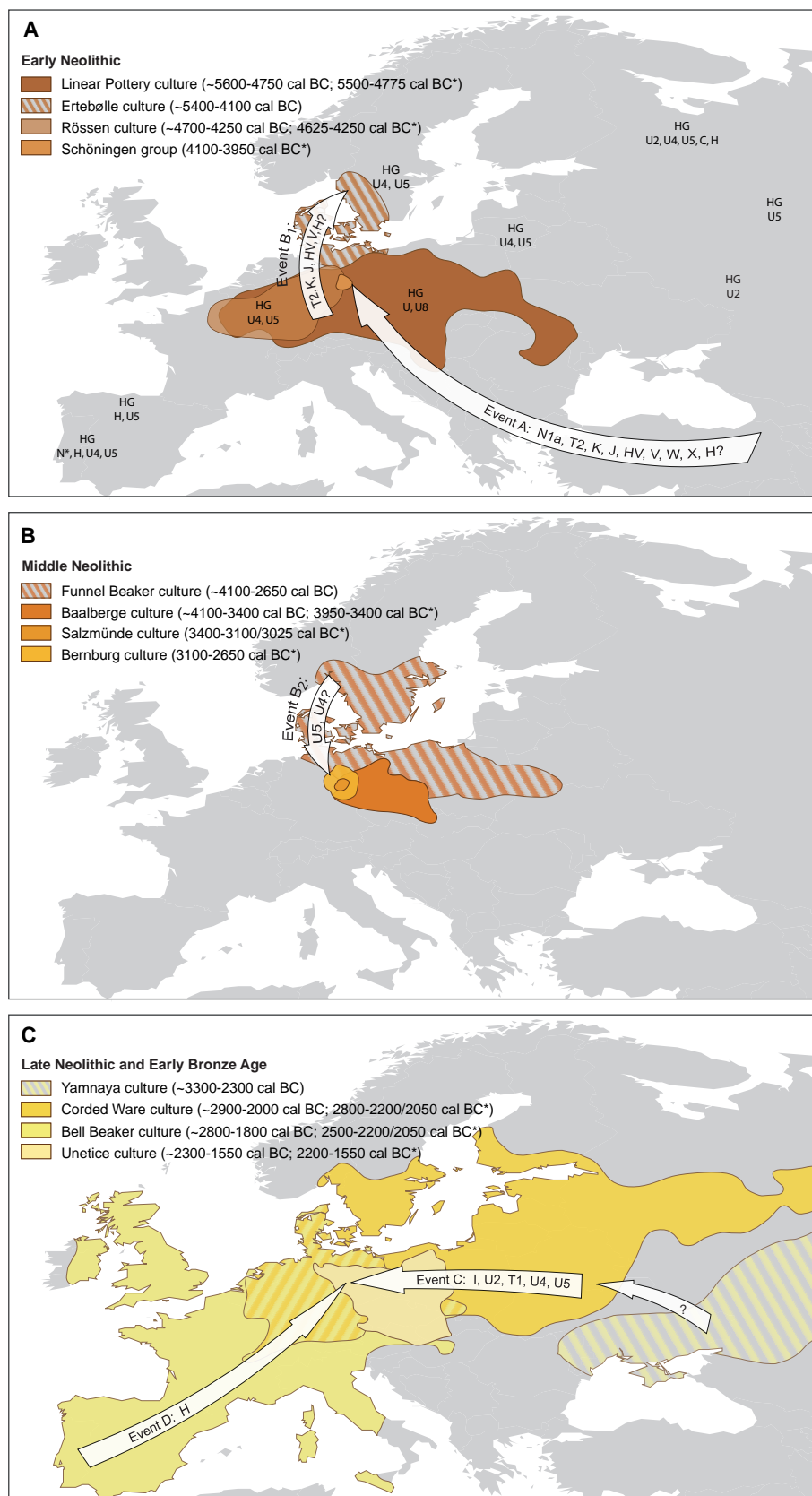
The Mittelbe-Saale cultures and the Central European metapopulation (CEM) were assigned to three statistic groups (group 0 = present-day metapopulation (CEM), group 1 = Late Neolithic/Early Bronze Age (LN/EBA: BEC, CWC, BBC, UC), group 2 = Early/Middle Neolithic (EN/MN: LBK, RSC, SCG, BAC, SMC)). Calibrated BC dates for each culture (Table S2) were converted into generations ago. Population divergence times are given as generation times: the Palaeolithic was set to 1,800 and the beginning of the Neolithic period to 400 generations ago. Overall, 12 variants of three main demographic models were tested. Models H1 and H2 were simulated with different migration rates of 10%, 20%, 30%, 40%, and 50%. All models were simulated by 250,000 generations. For abbreviations see Table S2 and S5. The simulation results are listed in Table S15.



**Fig. S9. P-values from simulated data**

We simulated data under two situations: one of population continuity, and one of population replacement. We then ran our method with different setting for  $\mu$ , the prior mean of  $c$ . Shown are histograms of p-values from 200 simulations in each situation. Each histogram is labeled by the simulated model, the true value of  $c$ , and the prior mean of  $c$ . The dotted line shows the expected distribution of p-values if the true model is population continuity.





**Fig. S10. Summary of population dynamic events during the Neolithic period in Europe**

The summary maps illustrate the population dynamic events (A-D) during the three phases of the Neolithic period until the Early Bronze Age: Early Neolithic (A), Middle Neolithic (B), and Late Neolithic/Early Bronze Age (C). Different shadings and patterns denote the geographic distributions of the nine Neolithic cultures under study (plain) and other contextually important cultures (striped). The dates for each culture concern their distribution across Europe and in Mittelelbe-Saale according to the chronology of Central Germany (\*). White arrows display hypothesised geographic expansions and their associated haplogroups according to the events A-D. The mtDNA variability of indigenous hunter-gatherer populations (HG) is mentioned in the respective regions (A).

**Table S1. Summary of archaeological sites and sample sizes**

The Mittelbe-Saale sites are listed in chronological and alphabetical order. Multiphase sites are listed numerous times. Basic information of geographic location (county of Saxony-Anhalt), year of excavation, periods discovered (NEO = Neolithic, BZA = Bronze Age, IA = Iron Age, RIA = Roman Iron Age, RIA = Roman Iron Age, MP = Migration Period, MA = Middle Ages), Neolithic cultures discovered, archaeological context of the investigated individuals (single/multiple burials, graveyard, settlement pits, burial chamber) and sample sizes are listed. Sample sizes of all excavated, analysed, and successfully typed individuals, and the lab in which DNA work was carried out (IoA= Institute of Anthropology, University of Mainz, Germany; ACAD=Australian Centre for ancient DNA, University of Adelaide, Australia) are presented separately per site and culture.

**Table S2. Summary of samples, 14C dates, and genetic results**

Sequence polymorphisms were compared to the Reconstructed Sapiens Reference Sequence (RSRS, [www.mtcommunity.org](http://www.mtcommunity.org)). Haplogroups were determined according to phylotree built 14, accessed 05 Apr 2012 ([www.phylotree.com](http://www.phylotree.com)). Generally, all control region sequences and SNP profiles were replicated from at least two independent extracts with the exception of the HVS-II sequence of individual KAR 20 that could not be replicated by a second sample (red letters). In cases when SNP profiles were derived from only one extract, we report the extraction number (A or B). Individuals with insufficient DNA preservation and unreliable HVS-I sequences or SNP profiles are reported as n.d. (not determined). Individuals with coding region information represent parts of the Brotherton et al. in 2013 study. Identical HVS-I, and if provided HVS-II sequences, of individuals originated from the same site and culture are highlighted by different shadings indicating potential kinship. Detailed results of the GenoCoRe22 SNP assay (Table S3) were summarized to a consensus profile of replicated SNPs and reported in forward direction (L-strand). Additional single SNPs to determine haplogroup H and U were used at the site Benzingerode and Salzmünde without applying the GenoCoRe22 SNP assay. Tooth notations followed the FDI World Dental Federation if determinable, otherwise the abbreviations M (molar), P (premolar), C (caninus), and I (incisivus) were used. In case of ambiguity, all possible teeth were listed and separated by slashes. Bone samples were named by anatomical notation followed by r (right) or l (left) if determined. Question marks in culture assignment indicate burials without or with uncharacteristic grave goods or burial rites. However, the respective date is based on an inclusive assessment of the site and the archaeological context, which supports the culture assignment.

**Table S3. Results of the GenoCoRe22 SNP multiplex assay**

SNP profiles were compared to the Reconstructed Sapiens Reference Sequence (RSRS, [www.mtcommunity.org](http://www.mtcommunity.org)). SNPs were detected in forward (L-strand) or reverse orientation (H-strand, underlined SNPs). Generally, individuals were typed from two independent extracts. Multiplex reactions that failed or were not determined results (n.d.) are mentioned in the haplogroup row. Dashes indicate either allelic dropout or a relative fluorescent unit (rfu) value below the threshold of 50. Color shadings of alleles indicate

derived SNPs (green), ambiguous SNP results, (violet) derived and ancestral SNPs with low rfu values less than 100 (blue and light red), and SNPs that were not detected but required for a particular haplogroup (red).

**Table S4. Published mitochondrial DNA data**

Published prehistoric mitochondrial DNA data were used for comparative analyses with the Mittelbe-Saale cultures. Prehistoric data were considered, when the studies have been accounted for authentication criteria, such as separated post and pre-PCR labs, multiple independent extractions and amplifications, and partial or complete cloning. These data were grouped according to their cultural assignment and/or geographical distribution. Groups with sample sizes of up to eight individuals were used for comparative analyses, while the remaining available data were considered in the discussion of the population genetic results. Haplotype sequences, sequence range, haplogroup assignment and dates were taken from the original publications. Whenever possible, published BP dates were converted to cal BC dates (cal BC dates in brackets) either by considering associated archaeological literature (indicated by a varying number of \* that refers to the corresponding publication) or by using the IntCal 09 curve of the program OxCal 4.1 (<https://c14.arch.ox.ac.uk/OxCal/OxCal.html>) (without \*). Sequence polymorphisms were converted relative to the Reconstructed Sapiens Reference Sequence (RSRS, [www.mtdnacommunity.org](http://www.mtdnacommunity.org)) and haplogroup assignment was confirmed or updated to the mitochondrial phylogeny of phylotree built 14, accessed 05 Apr 2012 ([www.phylotree.com](http://www.phylotree.com)). Partial data from different sites of Mittelbe-Saale were published previously. These data were included in the corresponding cultural datasets (Table S2).

**Table S5. Haplogroup and component frequencies of Mittelbe-Saale cultures and the hunter-gatherer/present-day metapopulation of Central Europe**

The haplogroups of the Mittelbe-Saale cultures, the hunter-gatherer metapopulation of Central Europe (HGC), and the present-day metapopulation of Central Europe (CEM) were differentiated into 20 groups and used for Ward clustering. In addition, characteristic haplogroups of the hunter-gatherers, Early/Middle Neolithic and Late/Neolithic/EBA cultures were summarized to respective components accordingly to the differentiation in the PCA. Haplogroups that could not be assigned unambiguously to one of the three components were defined as other. The 95% confidence intervals of each haplogroup and component are listed in the lower table. Culture information is presented in Table S2 and S4.

**Table S6. Details and results of genetic distance computation between Mittelbe-Saale cultures and the hunter-gatherer/present-day metapopulation of Central Europe**

Pairwise  $F_{st}$  (top) and significance values (bottom) were computed based on HVS-I sequences between the Mittelbe-Saale cultures, the hunter-gatherer metapopulation of Central Europe (HGC), and the present-day metapopulation of Central Europe (CEM) as well as between the Early/Middle Neolithic (EN/MN) and Late Neolithic/Early Bronze Age (LN/EBA) cultures of Mittelbe-Saale. The CEM was resampled by generating ten additional dataset each with 100 or 50 randomly selected individuals (CEM-100-I-V and

CEM-50-I-V) to exclude biases by fluctuating sample sizes. The p-values (regular) were adjusted post hoc to correct for multiple comparisons by using the Benjamini-Hochberg method (italicized). The shading indicates the significance of pairwise  $F_{st}$ -values. Light green represents significant ( $p \leq 0.05$ ) and dark green highly significant values ( $p \leq 0.001$ ). For abbreviations and population details see Table S2 and S4.

**Table S7. Details and results of AMOVA with prehistoric Mittelbe-Saale cultures**

AMOVA was performed based on HVS-I sequences of the Mittelbe-Saale cultures. Cultures were arranged into 289 different combinations consisting of two, three or four groups with varying numbers of populations within each group and AMOVA was conducted for each constellation. At first, all possible two group arrangements were tested to determine the best culture combination with the highest ‘among groups’ and the lowest ‘within groups’ variance (no. 246). Adjacent, this arrangement was split into all possible three (no. 247-259) or four (no. 260-289) group combinations. Groups are separated by + and cultures within one group are reported in brackets separated by underscores ( ). The table provides the ‘among groups’, ‘within groups’, and ‘within populations’ variance,  $F_{st}$ , and significant values (p) of each constellation tested. Color shadings indicate the best (dark green) and second best arrangements (light green). For abbreviations and population details see Table S2.

**Table S8. Details and results of population continuity test between Mittelbe-Saale cultures and the hunter-gatherer/present-day metapopulation of Central Europe**

Test of population continuity was performed based on haplogroup frequencies of the Mittelbe-Saale cultures, the hunter-gatherer metapopulation of Central Europe (HGC), and the present-day metapopulation of Central Europe (CEM, Table S5)). The p-value depicts the probability that changes in haplogroup frequencies between two populations cannot be explained by genetic drift alone. Light green represents significant ( $p \leq 0.05$ ) and dark green highly significant values ( $p \leq 0.001$ ). For abbreviations and population details see Table S2 and S4.

**Table S9. Details and haplogroup frequencies of 20 Mesolithic, Neolithic, and (Early) Bronze Age cultures used in principal component analysis and Ward clustering**

Relative haplogroup frequencies of the Mittelbe-Saale cultures and eleven previously published Mesolithic, Neolithic and (Early) Bronze Age cultures were used for principal component analysis and Ward clustering. Haplogroups were differentiated into 22 groups. For population details see Table S2 and S4.

**Table S10. Details and  $F_{st}$ -values of prehistoric Mittelbe-Saale cultures and 73 present-day populations used for multidimensional scaling**

Genetic distances based on HVS-I sequences were computed between the Mittelbe-Saale cultures and 73 present-day populations of Europe, the Near East, North Africa, and Asia. The table provides pairwise  $F_{st}$  (regular) and Slatkin  $F_{st}$ -values (italicized). Slatkin  $F_{st}$ -values were used for multidimensional scaling.

**Table S11. Details and haplogroup frequencies of prehistoric Mittelbe-Saale cultures and 73 present-day populations used in principal component analysis**

Relative haplogroup frequencies of the Mittelbe-Saale cultures and 73 present-day populations of Europe, the Near East, North Africa, and Asia were used for the principal component analysis. Haplogroups were differentiated into 23 sub-groups.

**Table S12. Details and haplogroup frequencies of prehistoric Mittelbe-Saale cultures and 56 present-day populations used in Procrustes analyses and Ward clustering**

Relative haplogroup frequencies of the Mittelbe-Saale cultures and 56 present-day populations of Europe, the Near East, North Africa, and Asia were used for Procrustes analyses and Ward clustering. Haplogroups were differentiated into 23 groups. For Procrustes analyses an initial PCA was carried out and PCA scores of the first and second principal component were rotated against the geographic coordinates (latitudes and longitude) which were defined for each population by reference points in Google Earth.

**Table S13. Details and  $F_{st}$ -values of prehistoric Mittelbe-Saale cultures and 150 present-day populations used for genetic distance maps**

Pairwise  $F_{st}$ -values based on high-resolution haplogroup frequencies were computed between the Mittelbe-Saale cultures and 150 present-day populations of Europe, the Near East, and North Africa. Geographic coordinates (latitudes and longitude) for each population were defined by reference points in Google Earth.

**Table S14. Details and results of AMOVA with prehistoric Mittelbe-Saale cultures and the Central European metapopulation**

AMOVA was carried out based on HVS-I sequences of the Mittelbe-Saale cultures and a Central European metapopulation consisting of 500 individual from Austria, Germany, the Czech Republic, and Poland (CEM). Cultures and the metapopulation were arranged into 37 different combinations and AMOVA was conducted for each constellation. AMOVA and Ward clustering supported the differentiation of the Mittelbe-Saale cultures into two groups of Early/Middle Neolithic and Late Neolithic/Early Bronze Age cultures (Table S7). This differentiation was used as prior and all possible two (no 1-2) or three (3-37) group arrangements with the CEM were tested by combining one or more cultures of the Early/Middle Neolithic or Late Neolithic/Early Bronze Age group with the CEM. The best combination of one or more Mittelbe-Saale cultures with the CEM was determined by the highest ‘among groups’ and the lowest ‘within groups’ variance (no. 2). Groups are separated by + and cultures/populations within one group are reported in brackets separated by underscores (\_). The table provides the ‘among groups’, ‘within groups’ and ‘within populations’ variance,  $F_{st}$ , and significant values (p) of each constellation tested. Color shadings indicate the best (dark green) and second best arrangements (light green). For abbreviations see Table S2 and S5.

**Table S15. Details and results of the Bayesian Serial Simcoal simulation, ABC and AIC goodness-of-fit estimates**

The Mittelbe-Saale cultures and the Central European metapopulation (CEM) were assigned to three statistic groups (group 0 = present-day metapopulation (CEM), group 1

= Late Neolithic/Early Bronze Age (LN/EBA: BEC, CWC, BBC, UC), group 2 = Early/Middle Neolithic (EN/MN: LBK, RSC, SCG, BAC, SMC)). Calibrated BC dates for each culture (Table S1) were converted into generations ago (LBK = 296, RSC = 263, SCG = 242, BAC = 236, SMC = 214, BEC = 202, CWC = 190, BBC = 178, UC = 166). Population divergence times are given as generation times: the Palaeolithic was set to 1,800 and the beginning of the Neolithic period to 400 generations ago. Overall, 12 variants of three main demographic models were tested. Models H1 and H2 were simulated with different migration rates of 10%, 20%, 30%, 40%, and 50%. All models were simulated using 250,000 genealogies.

#### **Table S16. Sequences of overlapping HVS-I and HVS-II primer pairs**

Primers are given in 5' to 3' direction. Primer names indicate forward (L-strand (L)) and reverse primer orientation (H-strand (H)) and denote the 3' end of the primer sequence for all HVS-I primers or the first nucleotide of the amplicon in case of the HVS-II primers

#### **Table S17. HVS-I and HVS-II sequences of researchers**

All researchers involved in genetic and anthropological analyses as well as all people who were involved in the sampling were typed. Archaeologists were typed if they were known to have been involved in a particular excavation. However, not all archaeologists were available, especially in case of older excavations. Sequence polymorphisms were compared to the Reconstructed Sapiens Reference Sequence (RSRS, [www.mtcommunity.org](http://www.mtcommunity.org)) and haplogroups were assigned according to the mitochondrial phylogeny of phylotree built 14, accessed 05 Apr 2012 ([www.phylotree.com](http://www.phylotree.com)).

#### **Movie S1.**

To best illustrate the dynamics of the genetic landscape in Neolithic Central Europe we animated the genetic distance maps, haplogroup frequencies as well as haplotype diversity through time. The timeline covers 4,500 years of prehistory from the late Mesolithic (~6,000 cal BC) to the end of the Early Bronze Age (2,200 cal BC). The timing is proportional to the time elapsed, i.e. the duration of each cultural period. Events A, B<sub>1</sub>, B<sub>2</sub>, C and D mark the genetic changes described in the main text, which are also visible in the alternating genetic affinities on the genetic distance maps (darker colors indicate a greater similarity with the respective Neolithic culture). White arrows summarize the underlying vectors in the form of substantially increasing/decreasing and/or newly arriving haplogroups as observed in the bar graphs at the bottom. Colored symbols on the genetic distance maps indicate the sampling location of the respective data and black dotted lines denote the distribution area of each Mittelelbe-Saale culture. The movie is prepared for the Flash player.

#### **Supplementary Notes**

##### Author Contributions

G.B., W.H., C.A., A.S-N., S.K., and C.R. performed the palaeogenetic analyses; N.N., G.B., and W.H. collected the samples; G.B., C.R., A.S-N., and S.K. collected reference data for the population genetic analyses; S.M-R, C.R., A.S-N. cloned PCR

products; G.B., W.H., J.P., D.R., and C.A. performed the biostatistic analyses; N.N. conducted the anthropological analyses including individualization; R.G., S.F., V.D., and H.M. excavated the sites, provided samples and archaeological context; K.W.A., H.M., and W.H. designed the study; G.B., W.H., A.C., and K.W.A. wrote the paper. All Authors discussed the results and commented the manuscript.

#### Members of The Genographic Consortium

Clio Der Sarkissian, Australian Centre for Ancient DNA, Adelaide, Australia; 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 Hernanz & 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.

## References and Notes

1. A. W. R. Whittle, V. Cummings, Eds., *Going Over: The Mesolithic-Neolithic Transition in North-West Europe* (Oxford Univ. Press, Oxford, 2007).
2. P. Bogucki, P. J. Crabtree, Eds., *Ancient Europe 8000 B.C. - A.D. 1000. Encyclopaedia of the Barbarian World* (Scribner, New York, 2004).
3. B. Cunliffe, Ed., *Europe Between the Oceans: Themes and Variations: 9000BC – AD 1000* (Yale Univ. Press, New Haven, CT, 2008).
4. H. Behrens, *Die Jungsteinzeit im Mittelelbe-Saale-Gebiet* (Veröffentlichungen des Landesmuseums für Vorgeschichte in Halle 27, Berlin, 1973).
5. R. Pinhasi, M. G. Thomas, M. Hofreiter, M. Currat, J. Burger, The genetic history of Europeans. *Trends Genet.* **28**, 496–505 (2012). [Medline](#)  
[doi:10.1016/j.tig.2012.06.006](https://doi.org/10.1016/j.tig.2012.06.006)
6. P. Soares, A. Achilli, O. Semino, W. Davies, V. Macaulay, H. J. Bandelt, A. Torroni, M. B. Richards, The archaeogenetics of Europe. *Curr. Biol.* **20**, R174–R183 (2010). [Medline](#) [doi:10.1016/j.cub.2009.11.054](https://doi.org/10.1016/j.cub.2009.11.054)
7. A. J. Ammerman, L. L. Cavalli-Sforza, *The Neolithic Transition and the Genetics of Populations in Europe* (Princeton Univ. Press, Princeton, NJ, 1984).
8. M. Richards, V. Macaulay, E. Hickey, E. Vega, B. Sykes, V. Guida, C. Rengo, D. Sellitto, F. Cruciani, T. Kivisild, R. Villems, M. Thomas, S. Rychkov, O. Rychkov, Y. Rychkov, M. Gölge, D. Dimitrov, E. Hill, D. Bradley, V. Romano, F. Calì, G. Vona, A. Demaine, S. Papiha, C. Triantaphyllidis, G. Stefanescu, J. Hatina, M. Belledi, A. Di Rienzo, A. Novelletto, A. Oppenheim, S. Nørby, N. Al-Zaheri, S. Santachiara-Benerecetti, R. Scozari, A. Torroni, H. J. Bandelt, Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* **67**, 1251–1276 (2000). [Medline](#)
9. A. Achilli, C. Rengo, C. Magri, V. Battaglia, A. Olivieri, R. Scozzari, F. Cruciani, M. Zeviani, E. Briem, V. Carelli, P. Moral, J. M. Dugoujon, U. Roostalu, E. L. Loogväli, T. Kivisild, H. J. Bandelt, M. Richards, R. Villems, A. S. Santachiara-Benerecetti, O. Semino, A. Torroni, 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). [Medline](#) [doi:10.1086/425590](https://doi.org/10.1086/425590)
10. B. Bramanti, M. G. Thomas, W. Haak, M. Unterlaender, P. Jores, K. Tambets, I. Antanaitis-Jacobs, M. N. Haidle, R. Jankauskas, C. J. Kind, F. Lueth, T. Terberger, J. Hiller, S. Matsumura, P. Forster, J. Burger, Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* **326**, 137–140 (2009); [10.1126/science.1176869](https://doi.org/10.1126/science.1176869). [Medline](#)  
[doi:10.1126/science.1176869](https://doi.org/10.1126/science.1176869)
11. W. Haak, O. Balanovsky, J. J. Sanchez, S. Koshel, V. Zaporozhchenko, C. J. Adler, C. S. Der Sarkissian, G. Brandt, C. Schwarz, N. Nicklisch, V. Dresely, B. Fritsch, E. Balanovska, R. Villems, H. Meller, K. W. Alt, A. Cooper, Ancient DNA from



- European early neolithic farmers reveals their near eastern affinities. *PLoS Biol.* **8**, e1000536 (2010). 10.1371/journal.pbio.1000536 [Medline](#)  
[doi:10.1371/journal.pbio.1000536](#)
12. P. Brotherton, W. Haak, J. Templeton, G. Brandt, J. Soubrier, C. Jane Adler, S. M. Richards, C. D. Sarkissian, R. Ganslmeier, S. Friederich, V. Dresely, M. van Oven, R. Kenyon, M. B. Van der Hoek, J. Korlach, K. Luong, S. Y. Ho, L. Quintana-Murci, D. M. Behar, H. Meller, K. W. Alt, A. Cooper, S. Adhikarla, A. K. Ganesh Prasad, R. Pitchappan, A. Varatharajan Santhakumari, E. Balanovska, O. Balanovsky, J. Bertranpetit, D. Comas, B. Martínez-Cruz, M. Melé, A. C. Clarke, E. A. Matisoo-Smith, M. C. Dulik, J. B. Gaieski, A. C. Owings, T. G. Schurr, M. G. Vilar, A. Hobbs, H. Soodyall, A. Javed, L. Parida, D. E. Platt, A. K. Royyuru, L. Jin, S. Li, M. E. Kaplan, N. C. Merchant, R. John Mitchell, C. Renfrew, D. R. Lacerda, F. R. Santos, D. F. Soria Hernanz, R. Spencer Wells, P. Swamikrishnan, C. Tyler-Smith, P. Paulo Vieira, J. S. Ziegler, Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nat. Commun* **4**, 1764 (2013). 10.1038/ncomms2656 [Medline](#)  
[doi:10.1038/ncomms2656](#)
13. Materials and methods are available as supplementary materials on *Science* Online.
14. Q. Fu, A. Mittnik, P. L. Johnson, K. Bos, M. Lari, R. Bollongino, C. Sun, L. Giemsch, R. Schmitz, J. Burger, A. M. Ronchitelli, F. Martini, R. G. Cremonesi, J. Svoboda, P. Bauer, D. Caramelli, S. Castellano, D. Reich, S. Pääbo, J. Krause, A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013). 10.1016/j.cub.2013.02.044 [Medline](#)  
[doi:10.1016/j.cub.2013.02.044](#)
15. H. Malmström, M. T. Gilbert, M. G. Thomas, M. Brandström, J. Storå, P. Molnar, P. K. Andersen, C. Bendixen, G. Holmlund, A. Götherström, E. Willerslev, Ancient DNA reveals lack of continuity between neolithic hunter-gatherers and contemporary Scandinavians. *Curr. Biol.* **19**, 1758–1762 (2009). [Medline](#)  
[doi:10.1016/j.cub.2009.09.017](#)
16. P. Skoglund, H. Malmström, M. Raghavan, J. Storå, P. Hall, E. Willerslev, M. T. Gilbert, A. Götherström, M. Jakobsson, Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336**, 466–469 (2012). [Medline](#)  
[doi:10.1126/science.1216304](#)
17. C. Der Sarkissian, O. Balanovsky, G. Brandt, V. Khartanovich, A. Buzhilova, S. Koshel, V. Zaporozhchenko, D. Gronenborn, V. Moiseyev, E. Kolpakov, V. Shumkin, K. W. Alt, E. Balanovska, A. Cooper, W. Haak, Ancient DNA reveals prehistoric gene-flow from siberia in the complex human population history of North East Europe. *PLoS Genet.* **9**, e1003296 (2013). [Medline](#)  
[doi:10.1371/journal.pgen.1003296](#)
18. J. Krause, A. W. Briggs, M. Kircher, T. Maricic, N. Zwyns, A. Derevianko, S. Pääbo, A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr. Biol.* **20**, 231–236 (2010). [Medline](#) [doi:10.1016/j.cub.2009.11.068](#)

19. C. Keyser, C. Bouakaze, E. Crubézy, V. G. Nikolaev, D. Montagnon, T. Reis, B. Ludes, Ancient DNA provides new insights into the history of south Siberian Kurgan people. *Hum. Genet.* **126**, 395–410 (2009). [Medline doi:10.1007/s00439-009-0683-0](#)
20. C. Lalueza-Fox, M. L. Sampietro, M. T. Gilbert, L. Castri, F. Facchini, D. Pettener, J. Bertranpetit, Unravelling migrations in the steppe: Mitochondrial DNA sequences from ancient central Asians. *Proc. Biol. Sci.* **271**, 941–947 (2004). [Medline doi:10.1098/rspb.2004.2698](#)
21. W. Haak, G. Brandt, H. N. de Jong, C. Meyer, R. Ganslmeier, V. Heyd, C. Hawkesworth, A. W. G. Pike, H. Meller, K. W. Alt, Ancient DNA, Strontium isotopes, and osteological analyses shed light on social and kinship organization of the Later Stone Age. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18226–18231 (2008). [Medline doi:10.1073/pnas.0807592105](#)
22. M. Hervella, N. Izagirre, S. Alonso, R. Fregel, A. Alonso, V. M. Cabrera, C. de la Rúa, 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). 10.1371/journal.pone.0034417 [Medline doi:10.1371/journal.pone.0034417](#)
23. C. Gamba, E. Fernández, M. Tirado, M. F. Deguilloux, M. H. Pemonge, P. Utrilla, M. Edo, M. Molist, R. Rasteiro, L. Chikhi, E. Arroyo-Pardo, Ancient DNA from an Early Neolithic Iberian population supports a pioneer colonization by first farmers. *Mol. Ecol.* **21**, 45–56 (2012). [Medline doi:10.1111/j.1365-294X.2011.05361.x](#)
24. M. Lacan, C. Keyser, F. X. Ricaut, N. Brucato, F. Duranthon, J. Guilaine, E. Crubézy, B. Ludes, Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9788–9791 (2011). [Medline doi:10.1073/pnas.1100723108](#)
25. H. J. Beier, R. Einicke, Eds., *Das Neolithikum im Mittelbe-Saale-Gebiet und in der Altmark. Eine Übersicht und ein Abriß zum Stand der Forschung* (Beier and Beran, Wilkau-Hasslau, Germany, 1994).
26. J. Preuss, Ed., *Das Neolithikum in Mitteleuropa* (Beier and Beran, Weißbach, Germany, 1998).
27. T. D. Price, Ed., *Europe's First Farmers* (Cambridge Univ. Press, Cambridge, 2000).
28. E. Bánffy, The Late Starčevo and the earliest Linear Pottery groups in Western Transdanubia. *Doc. Praehist.* **27**, 173–185 (2000).
29. D. Gronenborn, A variation on a basic theme: The transition to farming in southern central Europe. *J. World Prehist.* **13**, 123–210 (1999). [doi:10.1023/A:1022374312372](#)
30. D. Gronenborn, Migration, acculturation and culture change in western Temperate Eurasia, 6500–5000 cal BC. *Doc. Praehist.* **30**, 79–91 (2003).

31. P. Dolukhanov, A. Shukurov, D. Gronenborn, D. Sokoloff, V. Timofeev, G. Zaitseva, The chronology of Neolithic dispersal in Central and Eastern Europe. *J. Archaeol. Sci.* **32**, 1441–1458 (2005). [doi:10.1016/j.jas.2005.03.021](https://doi.org/10.1016/j.jas.2005.03.021)
32. P. Rowley-Conwy, Westward Ho! The Spread of agriculture from Central Europe to the Atlantic. *Curr. Anthropol.* **52**, S431–S451 (2011). [doi:10.1086/658368](https://doi.org/10.1086/658368)
33. B. Berthold *et al.*, *Die Totenhütte von Benzingerode*, Archäologie in Sachsen-Anhalt 7 [Landesamt für Denkmalpflege und Archäologie Sachsen-Anhalt, Halle (Saale), Germany, 2008].
34. M. Buchvaldek, C. Strahm, Eds., *Die kontinentaleuropäischen Gruppen der Kultur mit Schnurkeramik* (Univerzita Karlova, Prague, Czech Republic, 1992).
35. M. Gimbutas, in *Indo-European and Indo-Europeans*, G. Cardona, H. M. Hoenigswald, A. M. Seen, Eds. (Univ. of Pennsylvania Press, Philadelphia, 1970).
36. J. P. Mallory, *In Search of the Indo-Europeans* (Thames and Hudson, London, 1989).
37. D. W. Anthony, *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World* (Princeton Univ. Press, Princeton, NJ, 2007).
38. M. Furholt, *Die absolutchronologische Datierung der Schnurkeramik in Mitteleuropa und Südsandinavien* (Habelt, Bonn, 2003).
39. M. Van der Linden, What linked the Bell Beakers in third millennium BC Europe? *Antiquity* **81**, 343–352 (2007).
40. J. Czebreszuk, M. Szmyt, Eds., *The Northeast Frontier of Bell Beaker* (British Archaeological Reports, Oxford, 2003).
41. M. Vander Linden, For equalities are plural: Reassessing the social in Europe during the third millennium BC. *World Archaeol.* **39**, 177–193 (2007). [doi:10.1080/00438240701249678](https://doi.org/10.1080/00438240701249678)
42. V. Heyd, Families, treasures, warriors, and complex societies: Beaker groups and the third millennium BC along the Upper and Middle Danube. *Proc. Prehist. Soc.* **73**, 321–370 (2007).
43. R. Harrison, V. Heyd, The transformation of Europe in the third millennium BC: The example of “Le Petit Chasseur I & III” (Sion, Valais, Switzerland). *Prähist. Z.* **82**, 129–214 (2007).
44. F. Nicolis, Ed., *Bell Beakers Today: Pottery, People, Culture, Symbols in Prehistoric Europe: Proceedings of the International Colloquium, Riva del Garda (Trento, Italy), 11-16 May 1998* (Provincia autonoma di Trento, Servizio beniculturali, Ufficio beni archeologici, Trento, Italia, 2001).
45. M. Pruvost, R. Schwarz, V. B. Correia, S. Champlot, S. Braguier, N. Morel, Y. Fernandez-Jalvo, T. Grange, E. M. Geigl, Freshly excavated fossil bones are best for amplification of ancient DNA. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 739–744 (2007). [Medline doi:10.1073/pnas.0610257104](https://doi.org/10.1073/pnas.0610257104)

46. M. Hofreiter, D. Serre, H. N. Poinar, M. Kuch, S. Pääbo, Ancient DNA. *Nat. Rev. Genet.* **2**, 353–359 (2001). [Medline doi:10.1038/35072071](#)
47. W. Haak, P. Forster, B. Bramanti, S. Matsumura, G. Brandt, M. Tänzer, R. Villems, C. Renfrew, D. Gronenborn, K. W. Alt, J. Burger, Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* **310**, 1016–1018 (2005). [Medline](#)
48. C. J. Adler, W. Haak, D. Donlon, A. Cooper, Survival and recovery of DNA from ancient teeth and bones. *J. Archaeol. Sci.* **38**, 956–964 (2011). [doi:10.1016/j.jas.2010.11.010](#)
49. A. Kloss-Brandstätter, D. Pacher, S. Schönherr, H. Weissensteiner, R. Binna, G. Specht, F. Kronenberg, HaploGrep: A fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum. Mutat.* **32**, 25–32 (2011). [Medline doi:10.1002/humu.21382](#)
50. M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–E394 (2009). [Medline doi:10.1002/humu.20921](#)
51. D. M. Behar, M. van Oven, S. Rosset, M. Metspalu, E. L. Loogväli, N. M. Silva, T. Kivisild, A. Torroni, R. Villems, A “Copernican” reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012). [Medline doi:10.1016/j.ajhg.2012.03.002](#)
52. D. Caramelli, C. Lalueza-Fox, C. Vernesi, M. Lari, A. Casoli, F. Mallegni, B. Chiarelli, I. Dupanloup, J. Bertranpetit, G. Barbujani, G. Bertorelle, Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6593–6597 (2003). [Medline doi:10.1073/pnas.1130343100](#)
53. B. Bramanti, D. N. A. Ancient, Genetic analysis of aDNA from sixteen skeletons of the Vedrovice collection. *Anthropologie* **46**, 153–160 (2008).
54. A. G. Nikitin, J. R. Newton, I. D. Potekhina, Mitochondrial haplogroup C in ancient mitochondrial DNA from Ukraine extends the presence of East Eurasian genetic lineages in Neolithic Central and Eastern Europe. *J. Hum. Genet.* **57**, 610–612 (2012). [Medline doi:10.1038/jhg.2012.69](#)
55. M. F. Deguilloux, L. Soler, M. H. Pemonge, C. Scarre, R. Joussaume, L. Laporte, News from the west: Ancient DNA from a French megalithic burial chamber. *Am. J. Phys. Anthropol.* **144**, 108–118 (2011). [Medline doi:10.1002/ajpa.21376](#)
56. G. Di Benedetto, I. S. Nasidze, M. Stenico, L. Nigro, M. Krings, M. Lanzinger, L. Vigilant, M. Stoneking, S. Pääbo, G. Barbujani, Mitochondrial DNA sequences in prehistoric human remains from the Alps. *Eur. J. Hum. Genet.* **8**, 669–677 (2000). [Medline doi:10.1038/sj.ejhg.5200514](#)
57. L. Ermini, C. Olivieri, E. Rizzi, G. Corti, R. Bonnal, P. Soares, S. Luciani, I. Marota, G. De Bellis, M. B. Richards, F. Rollo, Complete mitochondrial genome sequence of the Tyrolean Iceman. *Curr. Biol.* **18**, 1687–1693 (2008). [Medline doi:10.1016/j.cub.2008.09.028](#)

58. L. Melchior, N. Lynnerup, H. R. Siegismund, T. Kivisild, J. Dissing, Genetic diversity among ancient Nordic populations. *PLoS ONE* **5**, e11898 (2010). 10.1371/journal.pone.0011898 [Medline](#) [doi:10.1371/journal.pone.0011898](#)
59. E. J. Lee, C. Makarewicz, R. Renneberg, M. Harder, B. Krause-Kyora, S. Müller, S. Ostritz, L. Fehren-Schmitz, S. Schreiber, J. Müller, N. von Wurmb-Schwark, A. Nebel, Emerging genetic patterns of the European Neolithic: Perspectives from a late Neolithic Bell Beaker burial site in Germany. *Am. J. Phys. Anthropol.* **148**, 571–579 (2012). [Medline](#) [doi:10.1002/ajpa.22074](#)
60. H. Chandler, B. Sykes, J. Zilhão, in *Actas dell III Congreso del Neolítico en la Península Ibérica, Monografías del Instituto internacional de Investigaciones Prehistóricas de Cantabria*, P. Arias, R. Ontañón, C. García-Moncó, Eds. (Univ. of Cantabria, Santander, Spain, 2005).
61. H. Chandler, thesis, University of Oxford, Oxford, UK (2003).
62. M. Hervella, thesis, University of the Basque Country, Leioa, Spain (2010).
63. F. Sánchez-Quinto, H. Schroeder, O. Ramirez, M. C. Avila-Arcos, M. Pybus, I. Olalde, A. M. Velazquez, M. E. Marcos, J. M. Encinas, J. Bertranpetit, L. Orlando, M. T. Gilbert, C. Lalueza-Fox, Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. *Curr. Biol.* **22**, 1494–1499 (2012). [Medline](#) [doi:10.1016/j.cub.2012.06.005](#)
64. M. Lacan, C. Keyser, F. X. Ricaut, N. Brucato, J. Tarrús, A. Bosch, J. Guilaine, E. Crubézy, B. Ludes, Ancient DNA suggests the leading role played by men in the Neolithic dissemination. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18255–18259 (2011). [Medline](#) [doi:10.1073/pnas.1113061108](#)
65. M. Lacan, thesis, University of Toulouse, Toulouse, France (2011).
66. B. Efron, R. Tibshirani, *An Introduction to the Bootstrap* (Chapman and Hall, London, 1993).
67. L. Excoffier, H. E. L. Lischer, Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567 (2010). [Medline](#) [doi:10.1111/j.1755-0998.2010.02847.x](#)
68. K. Tamura, M. Nei, Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**, 512–526 (1993). [Medline](#)
69. D. Posada, jModelTest: Phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253–1256 (2008). [Medline](#) [doi:10.1093/molbev/msn083](#)
70. O. J. Dunn, Multiple comparisons among means. *J. Am. Stat. Assoc.* **56**, 52–64 (1961). [doi:10.1080/01621459.1961.10482090](#)
71. S. R. Narum, Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conserv. Genet.* **7**, 783–787 (2006). [doi:10.1007/s10592-005-9056-y](#)
72. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc, B* **57**, 289–300 (1995).

73. P. Legendre, L. Legendre, *Numerical Ecology*, 2nd English Edition (Elsevier, Amsterdam, 1998).
74. L. L. Cavalli-Sforza, A. W. Edwards, Phylogenetic analysis. Models and estimation procedures. *Am. J. Hum. Genet.* **19**, 233–257 (1967). [Medline](#)
75. G. Nicholson, A. V. Smith, F. Jonsson, O. Gustafsson, K. Stefansson, P. Donnelly, Assessing population differentiation and isolation from single-nucleotide polymorphism data. *J. R. Stat. Soc., B* **64**, 695–715 (2002). [doi:10.1111/1467-9868.00357](#)
76. G. Coop, D. Witonsky, A. Di Rienzo, J. K. Pritchard, Using environmental correlations to identify loci underlying local adaptation. *Genetics* **185**, 1411–1423 (2010). [Medline](#) [doi:10.1534/genetics.110.114819](#)
77. A. Gelman *et al.*, *Bayesian Data Analysis* (Chapman and Hall/CRC, London, 2004).
78. X.-L. Meng, Posterior predictive p-values. *Ann. Stat.* **22**, 1142–1160 (1994). [doi:10.1214/aos/1176325622](#)
79. M. Slatkin, A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457–462 (1995). [Medline](#)
80. C. N. Anderson, U. Ramakrishnan, Y. L. Chan, E. A. Hadly, Serial SimCoal: A population genetics model for data from multiple populations and points in time. *Bioinformatics* **21**, 1733–1734 (2005). [Medline](#) [doi:10.1093/bioinformatics/bti154](#)
81. E. M. Belle, A. Benazzo, S. Ghirotto, V. Colonna, G. Barbujani, Comparing models on the genealogical relationships among Neandertal, Cro-Magnoid and modern Europeans by serial coalescent simulations. *Heredity* **102**, 218–225 (2009). [Medline](#) [doi:10.1038/hdy.2008.103](#)
82. S. Ghirotto, S. Mona, A. Benazzo, F. Paparazzo, D. Caramelli, G. Barbujani, Inferring genealogical processes from patterns of Bronze-Age and modern DNA variation in Sardinia. *Mol. Biol. Evol.* **27**, 875–886 (2010). [Medline](#) [doi:10.1093/molbev/msp292](#)
83. S. Y. Ho, P. Endicott, The crucial role of calibration in molecular date estimates for the peopling of the Americas. *Am. J. Hum. Genet.* **83**, 142–146 (2008). [Medline](#) [doi:10.1016/j.ajhg.2008.06.014](#)
84. M. Kimura, A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120 (1980). [Medline](#) [doi:10.1007/BF01731581](#)
85. M. A. Beaumont, W. Zhang, D. J. Balding, Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025–2035 (2002). [Medline](#)
86. H. Akaike, A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**, 716–723 (1974). [doi:10.1109/TAC.1974.1100705](#)
87. K. P. Burnham, D. R. Anderson, Eds., *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer, New York, ed. 2, 2002).



88. D. Posada, T. R. Buckley, Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Syst. Biol.* **53**, 793–808 (2004). [Medline](#) [doi:10.1080/10635150490522304](https://doi.org/10.1080/10635150490522304)
89. E. Willerslev, A. Cooper, Ancient DNA. *Proc. Biol. Sci.* **272**, 3–16 (2005). [Medline](#) [doi:10.1098/rspb.2004.2813](https://doi.org/10.1098/rspb.2004.2813)
90. S. Pääbo, H. Poinar, D. Serre, V. Jaenicke-Despres, J. Hebler, N. Rohland, M. Kuch, J. Krause, L. Vigilant, M. Hofreiter, Genetic analyses from ancient DNA. *Annu. Rev. Genet.* **38**, 645–679 (2004). [Medline](#) [doi:10.1146/annurev.genet.37.110801.143214](https://doi.org/10.1146/annurev.genet.37.110801.143214)
91. M. T. Gilbert, H. J. Bandelt, M. Hofreiter, I. Barnes, Assessing ancient DNA studies. *Trends Ecol. Evol.* **20**, 541–544 (2005). [Medline](#) [doi:10.1016/j.tree.2005.07.005](https://doi.org/10.1016/j.tree.2005.07.005)