

Supplementary information for:

Tracing the genetic origin of Europe's first farmers reveals insights into their social organization

Anna Szécsényi-Nagy^{a,b}, Guido Brandt^a, Victoria Keerl^a, János Jakucs^b, Wolfgang Haak^c, Sabine Möller-Rieker^a, Kitti Köhler^b, Balázs Gusztáv Mende^b, Marc Fecher^a, Krisztián Oross^b, Tibor Marton^b, Anett Osztás^b, Viktória Kiss^b, György Pálfi^d, Erika Molnár^d, Katalin Sebők^e, András Czene^f, Tibor Paluch^g, Mario Šlaus^h, Mario Novakⁱ, Nives Pećina-Šlaus^j, Brigitta Ószk^k, Vanda Voicsek^k, Krisztina Somogyi^e, Gábor Tóth^l, Bernd Kromer^m, Eszter Bánffy^{b,n}, Kurt W. Alt^{a,o,p,q}

^a Institute of Anthropology, Johannes Gutenberg University of Mainz, D-55128 Mainz, Germany,

^b Archaeological Institute, Research Centre for the Humanities, Hungarian Academy of Sciences, H-1014 Budapest, Hungary,

^c Australian Centre for Ancient DNA, School of Earth and Environmental Sciences, AUS-5005 University of Adelaide, Australia,

^d Department of Biological Anthropology, University of Szeged, H-6720 Szeged, Hungary,

^e Institute of Archaeological Sciences, University of ELTE, H-1088 Budapest, Hungary,

^f Directorate of Pest County Museums, H-2000 Szentendre, Hungary,

^g Móra Ferenc Museum, H-6720 Szeged, Hungary,

^h Anthropological center, Croatian Academy of Sciences and Arts, 10000 Zagreb, Croatia,

ⁱ School of Archaeology, University College Dublin, 4 Dublin, Ireland,

^j Department of Biology, School of Medicine, University of Zagreb, 10000 Zagreb, Croatia,

^k National Heritage Protection Centre of the Hungarian National Museum, Department Pécs, 7621 Pécs, Hungary

^l Biology Department, University of West Hungary, H-9700 Szombathely, Hungary,

^m Curt-Engelhorn-Centre for Archaeometry, D-68159 Mannheim, Germany,

ⁿ German Archaeological Institute, Roman-Germanic Commission, D-60325 Frankfurt am Main, Germany,

^o State Office for Heritage Management and Archaeology Saxony-Anhalt and Heritage Museum, D-06114 Halle, Germany,

^p Institute for Prehistory and Archaeological Science, University of Basel, 4003 Basel, Switzerland,

^q Danube Private University, A-3500 Krems, Austria.

Index of the supporting text file

Material and Methods.....	4
Archaeological sites and dating.....	4
Samples and sampling.....	4
Ancient DNA work and authenticity of the results.....	5
Analyses of mitochondrial DNA.....	6
Analyses of Y chromosomal DNA.....	7
Comparative data for population genetic analyses.....	8
Haplotype diversity.....	10
Fisher’s exact test.....	10
Genetic distances.....	11
Test of population continuity (TPC).....	11
Principal component analysis (PCA).....	11
Multidimensional scaling (MDS).....	13
Analysis of molecular variance (AMOVA).....	13
Ancestral shared haplotype analysis (ASHA).....	14
Genetic distance maps.....	14
References.....	16
Supplementary figures.....	20
Supplementary Datasets.....	26

Material and Methods

Archaeological sites and dating

For the genetic investigations, we selected one Mesolithic site from south Croatia, six STA sites from western Hungary, three STA sites from northern Croatia, and eight Hungarian sites of the LBKT (Figure 1, Dataset S1). Our aim was to cover the major distribution area of both cultures and sample the most important Early and Middle Neolithic sites in western Hungary (1, 2). The LBKT site of Harta-Gátórház geographically does not belong to today's Transdanubia, but rather to the Danube-Tisza Interfluve, although culturally this site was connected to the Transdanubian distribution of the LBK.

Since graveyards are absent in the STA and LBKT, all investigated samples are from settlement burials. Most of the studied sites were inhabited throughout several archaeological periods (Dataset S1). Each analyzed individual was dated and assigned to a particular archaeological culture by characteristic grave goods, archaeological context and stratigraphic position (Dataset S1-S2). In 26 instances, where characteristic grave goods were absent or the archaeological context was insufficient for unambiguous attribution of the individuals to a certain culture, we dated the human skeletal remains by radiocarbon analyses (Dataset S2).

Samples and sampling

The skeletons from western Hungary were uncovered during the last decade in the course of rescue excavations preceding motorway constructions or other infrastructure projects, while the sites from northern Croatia were excavated between 1977 and 1999. Samples from three STA and two LBKT sites (Dataset S2) were obtained in their untreated and unwashed state after

the excavation. To monitor potential sources of contamination, we took swab samples from all genetic and anthropological investigators as well as most of the archaeological investigators who came into contact with the skeletons or the samples (Dataset S17). The sampling was carried out by A.S-N., J.J., M.F., and V.Ke. with the anthropological assistance of BG.M. and K.K.. Human remains of multiple burials were individualized by anthropological experts as listed in Dataset S2. To avoid contamination of the skeletal remains with modern DNA, the samples were taken with all possible precautions through the use of gloves, face masks and disposable oversleeves. All instruments and materials used were cleaned extensively with bleach before, between, and after sampling. We took two to five samples per individual, from different skeletal elements. Whenever possible, teeth were favored for aDNA analyses, otherwise pieces of long bone compacta or the *pars petrosa ossis temporalis* were taken, which were sawn out using a cleaned diamond drill. The samples were then directly transferred to the Institute of Anthropology at the University of Mainz and stored at -20 °C.

Ancient DNA work and authenticity of the results

The ancient DNA work was carried out in specialized facilities of the Institute of Anthropology, Bioarchaeometry Group, at the Johannes Gutenberg University of Mainz following well-established protocols to prevent and minimize contamination with modern DNA (3–6). These facilities are composed of pre-PCR labs for sample preparation, DNA extraction, and PCR set-up and post-PCR labs for amplification, sequencing, and cloning. The precautions against contamination, the sample preparation and DNA-extraction followed our standard protocols as described previously (6) with the following modifications: during the grinding step, every tenth sample was a grinding blank, consisting of DNA free hydroxyl apatite powder. For the

extraction, 0.2-1 g bone or tooth powder was used. 8-22 samples were processed at once, with one or two extraction blanks and a grinding blank at each extraction event.

In previous publications, we have discussed several criteria compiling a chain of evidence for the authentication of our ancient data (3–6), which we have applied analogously in this study.

Analyses of mitochondrial DNA

Mitochondrial DNA diversity was investigated by the analyses of multiple independent and informative loci of the mitochondrial genome, including the HVS-I and II of the control region and 22 haplogroup-defining SNPs of the coding region (Dataset S3-S4). Depending on the state of DNA preservation of each sample, HVS-I was amplified using one of three different primer systems consisting of two, four or six overlapping primer pairs with decreasing amplicon length (Dataset S16). These primer systems produced contiguous HVS-I sequences of 356 (np 16046-16401), 413 (np 15997-16409), and 383 base pairs (np 16019-16401), respectively. HVS-I sequences were replicated by at least three independent amplifications from a minimum of two samples per DNA extracts, producing 6-18 independent and overlapping amplicons (depending on the primer system used). In addition, selected PCR products with ambiguous nucleotide positions were cloned, and an average of 5 clones per amplicon were sequenced to monitor possible background contaminations and DNA damage. Poorly preserved samples with numerous ambiguous nucleotide positions were cloned entirely. Individuals with inconsistent HVS-I results were either discarded or replicated using an independent third sample. HVS-II sequences were obtained from individuals from the same archaeological site that showed consistent and identical HVS-I motifs in order to detect potential maternal kinship. HVS-II sequences were amplified at least twice from two extracts by using four overlapping primer

pairs (Dataset S16) that produced a contiguous sequence of 364 bp (np 34-397). Amplicons with numerous ambiguous nucleotide positions were additionally cloned. HVS-I and HVS-II were amplified, purified, sequenced and cloned according to our standard protocols as described previously (3, 6). Coding region information was achieved using the GenoCoRe22 SNP multiplex assay (5), which was amplified once per DNA extract.

Mitochondrial sequence polymorphisms were reported relative to the revised Cambridge Reference Sequence (rCRS) (7) as well as the Reconstructed Sapiens Reference Sequence (RSRS, www.mtdnacommunity.org) (8). Haplogroup determination was carried out according to the mtDNA phylogeny of PhyloTree build 14, accessed 05 April 2012 (www.phylotree.com) (9).

Analyses of Y chromosomal DNA

Y chromosome diversity was obtained through the analyses of 33 haplogroup-defining SNPs of the NRY, which enabled us to distinguish the most frequent Eurasian haplogroups (Dataset S5). Most of these informative SNPs (25) were retrieved using the GenoY25 SNP multiplex assay (5). In addition, we designed singleplex PCR for 8 additional SNP markers to increase the sub-haplogroup resolution of particular branches (I, G and F*) of the Y chromosome phylogeny (Dataset S16), which were detected in our ancient samples. Multiplex assays were replicated at least 4 times, twice from each extract. The singleplex PCRs were performed once from each DNA extract. Due to the incomplete morphological sex data, all individuals with reproduced mtDNA results were tested for NRY markers. The GenoY25 SNP multiplex was carried out according to the protocol published previously (5) with the following modifications: PCR was set up in a volume of 16 μ l consisting of 1 \times Buffer Gold, 8 mM MgCl₂ (Applied Biosystems), 0.7 mM dNTPs (Qiagen), \leq 0.2 μ M of each primer, 13.4 μ g BSA (Roche), 1.25 U of Amplitaq Gold

Polymerase (Applied Biosystems), and 4 µl target DNA. Thermocycling conditions consisted of an initial denaturation at 95°C for 6 min, 37–45 cycles of 10 sec at 95°C, 1 min at 60°C, and 30 sec at 65°C, followed by a final extension at 65°C for 6 min. PCR products were purified by incubating 2.5 µl PCR product with 1.5 U FastAP, 0.6 U ExoI, 1 µl 10x FastAP buffer (Fermentas, Thermo Fischer Scientific) and 1 µl HPLC water at 37°C for 10 min, followed by heat inactivation at 75°C for 5 min. SBE reaction was performed with 30 cycles, and the amplicons were purified with 1U FastAP (Fermentas) using cycling conditions consistent to the purification of the PCR products. The singleplex PCR were carried out according to the protocol used for mtDNA. Amplicons were purified with ExoI and FastAP and subsequently sequenced in line with the protocols used for mtDNA.

Y chromosomal haplogroups were determined using the phylogeny of the International Society of Genetic Genealogy (2014), Y-DNA Haplogroup Tree 2014, Version: 9.52, Date: 5 April 2014 (<http://www.isogg.org/tree/>).

Comparative data for population genetic analyses

For comparative analyses of the two investigated cultures, we used prehistoric and present-day mtDNA and NRY data from published sources.

The mtDNA data of the STA and LBKT samples were compared with 487 published prehistoric data across Europe, which were pooled into 14 groups according to cultural, chronological, and geographic aspects (Dataset S6). These groups comprise two hunter-gatherer populations from Central/North Europe (10–12) and southwestern Europe (13–15), the LBK from Central Europe (3, 5, 6, 16), a temporal succession of four cultures from the 5th/4th millennium BC of central Germany (6), four cultures from the 3rd/2nd millennium BC of central Germany (4, 6, 17), and

three southwest European populations from the 6th-4th millennia BC of Portugal (13), Basque County and Navarra (14), and Catalonia (18, 19) representing the Neolithic of the Iberian Peninsula. Previously published mtDNA data from the Carpathian Basin (20) were omitted from the population genetic analysis since serious doubts have been raised concerning the accurate dating and cultural assignment of some of these samples (21).

In order to identify affinities of our prehistoric sample sets in the maternal and paternal gene pool of present-day Eurasian and African populations, we gathered 67,996 mitochondrial HVS-I sequences and 49,516 NRY SNP profiles from the literature. We generated different mtDNA and NRY datasets, which were used for PCA and genetic distance maps and we pooled the modern-day data into different populations according to geography or ethnicity, as described in the original publications.

For PCA with mtDNA data, the present-day samples were grouped into 73 populations. This dataset was composed of 50,688 sequences with an average sample size of 694 samples per population (Dataset S12). Mitochondrial genetic distance maps were generated from HVS-I sequences of 130 modern-day populations. Whenever possible, the administrative subdivisions of a country were considered in order to increase the phylogeographic resolution. In this dataset, we only included population data with a minimum sequence range of np 16068-16365 to exclude biases by varying sequence ranges. Each population is represented by a maximum of 140 randomly selected individuals, which resulted in a total amount of 17,074 sequences used in the analysis (Dataset S13).

For population genetic analysis of NRY data, we combined our results with three published LBK data (5) to enlarge the prehistoric dataset up to twelve individuals. Present-day population data were only considered when the Y chromosome sub-haplogroups I, G, F, K, and R1a were differentiated, the selection of which includes the most frequent haplogroups observed from

the prehistoric data. PCA was carried out with 24,464 samples from 80 present-day populations with an average sample size of 305 individuals per population (Dataset S14). The Y chromosome genetic distance map consists of 100 modern-day populations with 215 samples per population on average, using of 21,478 individuals (Dataset S15) from the dataset.

Haplotype diversity

Haplotype diversity (22) of the Central/North European hunter-gatherers, the two Carpathian Basin cultures, and the Central European LBK was computed in DnaSP Version 5.10.01 (23), using HVS-I sequences (np 16056-16400).

Fisher's exact test

We used the mtDNA haplogroup frequencies in order to identify significant variation between the haplogroup composition of the STA, LBKT, and LBK (Table 1) using Fisher's exact test (24). In addition, we also included available data from the Central/North European hunter-gatherers into the mtDNA analysis. Overall, 16 mtDNA (H, HV, V, J, K, N1a, T1, T2, U, U2, U3, U4, U5a, U5b, W, and X) haplogroups were distinguished. The Fisher test was carried out in R 3.0.2 (The R Foundation for Statistical Computing 2011, <http://www.r-project.org/>) by using the implemented *fisher.test* function. Significant variation in haplogroup compositions was assessed by 10,000 permutations.

Genetic distances

F_{st} values were computed in Arlequin 3.5.1 (25) based on HVS-I sequences (np 16056-16400) of the Central/North European hunter-gatherers, the two Carpathian Basin cultures, and the Central European LBK (Table 1). We used the Tamura & Nei substitution model (26) and an associated gamma value of 0.177, which were inferred from the software FindModel based on PAML likelihoods (www.hcv.lanl.gov/content/sequence/findmodel/findmodel.html) and tested significant variations in F_{st} -values by 10,000 permutations. The p values were adjusted post hoc to correct for multiple comparisons with the Benjamin and Hochberg method, using the function *p.adjust* in R 3.0.2.

Test of population continuity (TPC)

We performed tests of population continuity as described by Brandt and his colleagues (6) using the absolute haplogroup frequencies of the hunter-gatherer Central/North, STA, LBKT and LBK datasets (Dataset S7). In order to apply conservative parameters, i.e. maximizing the chances of genetic drift, we used the terminal dates of the Mesolithic in the Carpathian Basin (6000 cal BC) and of each Neolithic culture's timespan to define the difference in time between populations in n generations of 25 years. We also ran each of the pairwise tests of all possible group combinations with three different effective population sizes ($N_e=500$, 5000 and 30000) (Dataset S11). The TPC script is available at <https://github.com/joepickrell/tpc>.

Principal component analysis (PCA)

PCA was carried out based on mtDNA and NRY haplogroup frequencies of prehistoric and modern-day populations. On the prehistoric level, the mtDNA haplogroup composition of the

STA and LBKT samples were compared to 14 ancient cultural entities. In this analysis, we considered 22 mtDNA haplogroups (H, H5, HV, HV0, V, I, J, K, N, N1a, R, T1, T2, U, U2, U3, U4, U5a, U5b, U8, W, and X), which were observed from the ancient samples (Dataset S7, Figure 2). In order to exclude biases induced by potential maternal kinship within the prehistoric datasets, which could have led to an overestimation of haplogroup frequencies and genetic affinities, we included a reduced dataset (marked with the symbol *) in the analysis. Redundant haplotypes with identical HVS-I and II sequences from the same site, were counted once.

Mitochondrial haplogroup frequencies of the two Neolithic Carpathian Basin cultures and 73 populations were used for PCA with present-day comparative data. The following 21 haplogroups were differentiated that cover the most frequent haplogroups of modern-day Eurasian populations: H, HV, HV0/V, I, J, K, N1a, T1, T2, U, U2, U3, U4, U5a, U5b, U8, W, and X, African haplogroups (L), Asian haplogroups (A, B, C, D, E, F, G, Q, Y, and Z) and other (all remaining haplogroups) (Dataset S12, Figure S2).

The Y chromosomal data of the STA, LBKT, and LBK were pooled and compared to 80 modern-day populations. Y chromosome haplogroup frequencies were condensed to 13 groups (AB, E, DHC, F, G, I, J, KTS, L, NO, N, PR, and R1a) (Dataset S14, Figure 5), based on the Y chromosome phylogeny and phylogeography. This haplogroup classification was conditioned by the varying resolution of the published comparative data, which were in many cases insufficiently resolved to distinguish further subgroups such as I1 and I2 or G1 and G2.

All PCA were carried out using the *prcomp* function for categorical PCA, implemented in the R 3.0.2 package and plotted in a two-dimensional (prehistoric culture PCA) or three-dimensional space (present-day population PCA), displaying the first two or three principal components, respectively.

Multidimensional scaling (MDS)

HVS-I sequences (np 16056-16400) of the two Carpathian Basin cultures and 14 prehistoric populations were used for genetic distance computation in Arlequin 3.5.1 (25), with the same substitution model as at the genetic distance calculation. Analogous to the PCA, we also integrated the reduced datasets (*) to exclude biases by potential maternal kinship. The F statistic was calculated based on 10,000 permutations. MDS was applied on the matrix of linearized Slatkin F_{st} values (27) and visualized in a two-dimensional space using the *metaMDS* function based on Euclidean distances implemented in the *vegan* library of R 3.0.2. (Dataset S8, Figure S1).

Analysis of molecular variance (AMOVA)

We arranged the HVS-I sequences (np 16056-16400) of the STA and LBKT together with nine archaeological cultures from Central Europe, ranging from the LBK to the Early Bronze Age into varying groups. According to our previous publication (6), we pooled the Central European cultures into 6th-4th millennia BC and 3rd/2nd millennium BC groups and subsequently transferred one or more cultures to the STA and LBKT until each group was subsumed with our samples. Overall we tested 82 different arrangements to find out the best combination indicated by the greatest among-groups and the least within-group variance (Dataset S9). Variance, F_{st} , and significant values (p) were computed with the standard AMOVA function implemented in Arlequin 3.5.1 (25) by using the Tamura & Nei substitution model (26) and a gamma value of 0.177. F_{st} values were tested on significance by 10,000 permutations.

Ancestral shared haplotype analysis (ASHA)

We used shared haplotype analysis (25) and modified this approach by accounting for the temporal succession of cultures in order to ascribe mtDNA haplotypes to particular cultures or time periods, and to identify the amount of ancestral lineages in each culture. Therefore, hunter-gatherers from Central/North Europe, the two investigated Carpathian Basin cultures, and nine cultures from Central Europe ranging from the LBK to the Early Bronze Age were placed into a chronological order. Each lineage within a given cultural dataset was traced back to its earliest match in the chronology and regarded as ancestral lineage that arose in this culture for the first time (Dataset S10, Figure 3). This approach enabled us to estimate the amount of mtDNA lineages that were prevalent in Central European cultures since i) the Mesolithic, ii) the STA, iii) the LBKT, iv) the LBK, or v) that emerged in Central Europe in later periods.

Genetic distance maps

Genetic distance maps were generated from mtDNA and NRY data. HVS-I sequences (np 16068-16365) of the two investigated Carpathian Basin cultures were compared to 130 present-day Eurasian populations (Dataset S13, Figure 4ab). Genetic distances were calculated in Arlequin 3.5.1 (25) using the Tamura and Nei substitution model (26) and a gamma value of 0.177.

The Y chromosomal data of the STA, LBKT, and LBK were pooled and compared to 100 modern-day populations (Dataset S15, Figure S3). Haplogroup frequencies were differentiated into 16

groups (AB, D, E, F, G, H, C, I, J, KTS, L, NO, N, PR, R1, and R1a) and pairwise F_{st} values were computed in Arlequin 3.5.1 using the conventional F statistic.

Mitochondrial and Y chromosome genetic distances between the cultural datasets and modern-day populations were combined with longitudes and latitudes according to the sampling information in the literature and interpolated with the Natural Neighbor method implemented in ArcGis version 10.0 (Arcmap, Environmental Systems Research Institute [Esri] Inc, Redlands, USA).

References

1. Kalicz N (2010) in *Die Neolithisierung Mitteleuropas. The Spread of Neolithic to Central Europe. International Conference, 24-26th June 2005. Mainz RGZM*, eds Gronenborn D, Petrasch J (RGZM, Mainz), pp 235–254. Band 4.
2. Oross K, Marton T (2012) Neolithic burials of the Linearbandkeramik settlement at Balatonszárszó and their European context. *Acta Archaeol* 63:257–299.
3. Haak W et al. (2005) Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* 310:1016–8.
4. Haak W et al. (2008) 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* 105:18226–31.
5. Haak W et al. (2010) Ancient DNA from European early Neolithic farmers reveals their Near Eastern affinities. *PLoS Biol* 8:e1000536.
6. Brandt G et al. (2013) Ancient DNA reveals key stages in the formation of Central European mitochondrial genetic diversity. *Science* 342:257–261.
7. Andrews RM et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23:147.
8. Behar DM et al. (2012) A “Copernican” reassessment of the human mitochondrial DNA tree from its root. *Am J Hum Genet* 90:675–84.

9. Van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30:E386–94.
10. Bramanti B et al. (2009) Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* 326:137–40.
11. Bollongino R et al. (2013) 2000 years of parallel societies in Stone Age Central Europe. *Science* 342:479–481.
12. Fu Q et al. (2013) A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr Biol* 23:1–7.
13. Chandler H, Sykes B, Zilhão J (2005) in *Actas dell III Congreso del Neolítico en la Península Ibérica, Santander, Monografías del Instituto internacional de Investigaciones Prehistóricas de Cantabria 1*, eds Arias P, Ontañón R, García-Moncó C, pp 781–786.
14. Hervella M et al. (2012) 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.
15. Sánchez-Quinto F et al. (2012) Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. *Curr Biol* 22:1494–9.
16. Bramanti B (2008) Ancient DNA: genetic analysis of aDNA from sixteen skeletons of the Vedrovice. *Anthropol* 46:153–160.
17. Lee EJ et al. (2012) Emerging genetic patterns of the European Neolithic: perspectives from a late Neolithic Bell Beaker burial site in Germany. *Am J Phys Anthropol* 148:571–9.

18. Lacan M et al. (2011) Ancient DNA suggests the leading role played by men in the Neolithic dissemination. *Proc Natl Acad Sci U S A* 108:18255–9.
19. Gamba C et al. (2012) Ancient DNA from an Early Neolithic Iberian population supports a pioneer colonization by first farmers. *Mol Ecol* 21:45–56.
20. Guba Z et al. (2011) HVS-I polymorphism screening of ancient human mitochondrial DNA provides evidence for N9a discontinuity and East Asian haplogroups in the Neolithic Hungary. *J Hum Genet* 56:784–96.
21. Bánffy E, Brandt G, Alt KW (2012) “Early Neolithic” graves of the Carpathian Basin are in fact 6000 years younger—appeal for real interdisciplinarity between archaeology and ancient DNA research. *J Hum Genet* 57:467–9; author reply 470–1.
22. Nei M (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York).
23. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–2.
24. Fisher RA (1922) On the interpretation of χ^2 from contingency tables, and the calculation of P. *J R Stat Soc* 85:87–94.
25. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10:564–7.
26. Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10:512–26.

27. Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–62.

Supplementary figures

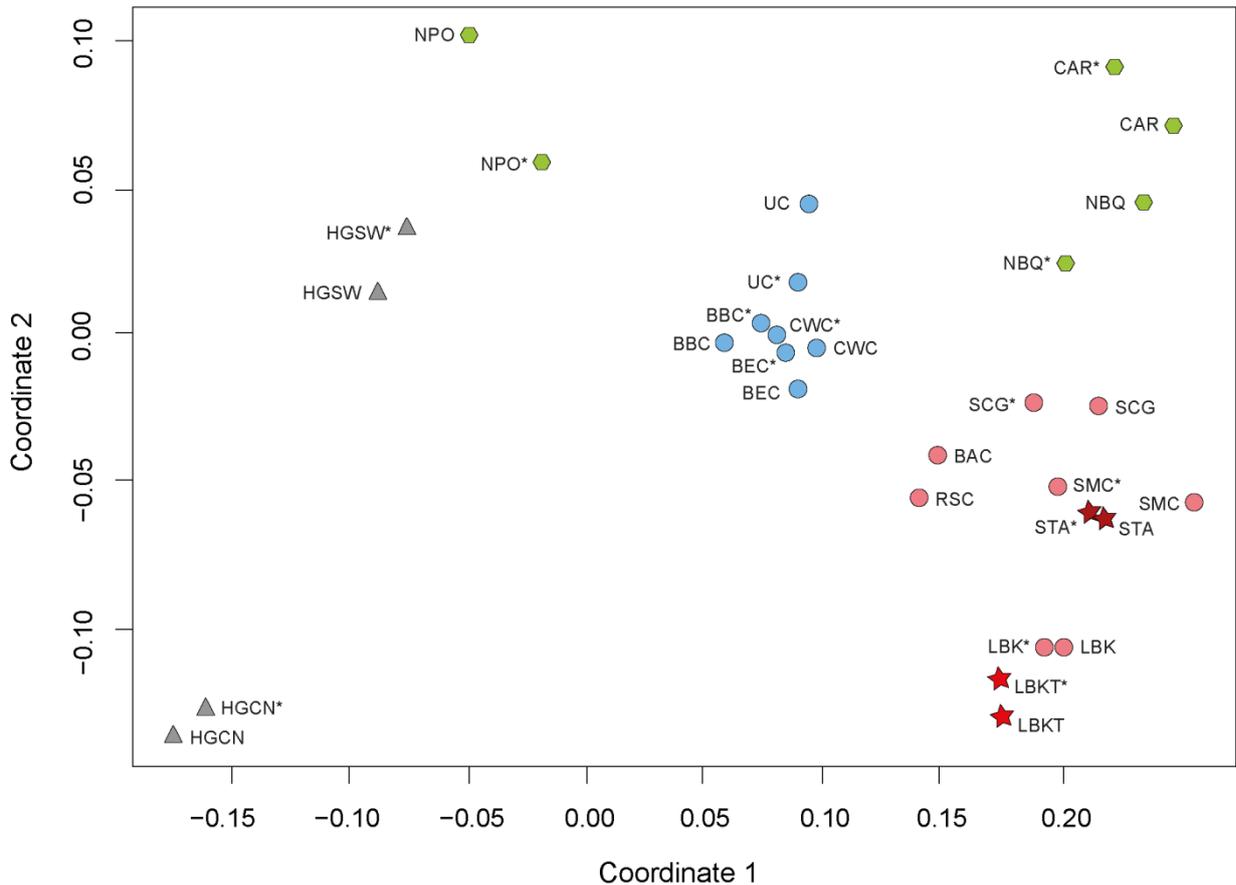


Figure S1. Multidimensional scaling of 16 prehistoric cultures.

Genetic distances (F_{st}) were computed between the STA, LBKT and 14 prehistoric cultures, and visualized by multidimensional scaling, with a stress value of 0.10639. Color shadings and symbols denote populations of different periods or European regions according to figure 2. The reduced version of each dataset is marked by an asterisk (*). Detailed information about the comparative data and F_{st} values are listed in Dataset S8.

Culture abbreviations: hunter-gatherers in Central and North Europe (HGCN), hunter-gatherers in Southwestern Europe (HGSW), Starčevo culture (STA), LBK in Transdanubia (LBKT), LBK in Central Europe (LBK), Rössen culture (RSC), Schöningen group (SCG), Baalberge culture (BAC), Salzmünde culture (SMC), Bernburg culture (BEC), Corded Ware culture (CWC), Bell Beaker culture (BBC), Únětice culture (UC), Cardial and Epicardial culture (CAR), Neolithic Basque Country and Navarre (NBQ), Neolithic Portugal (NPO).

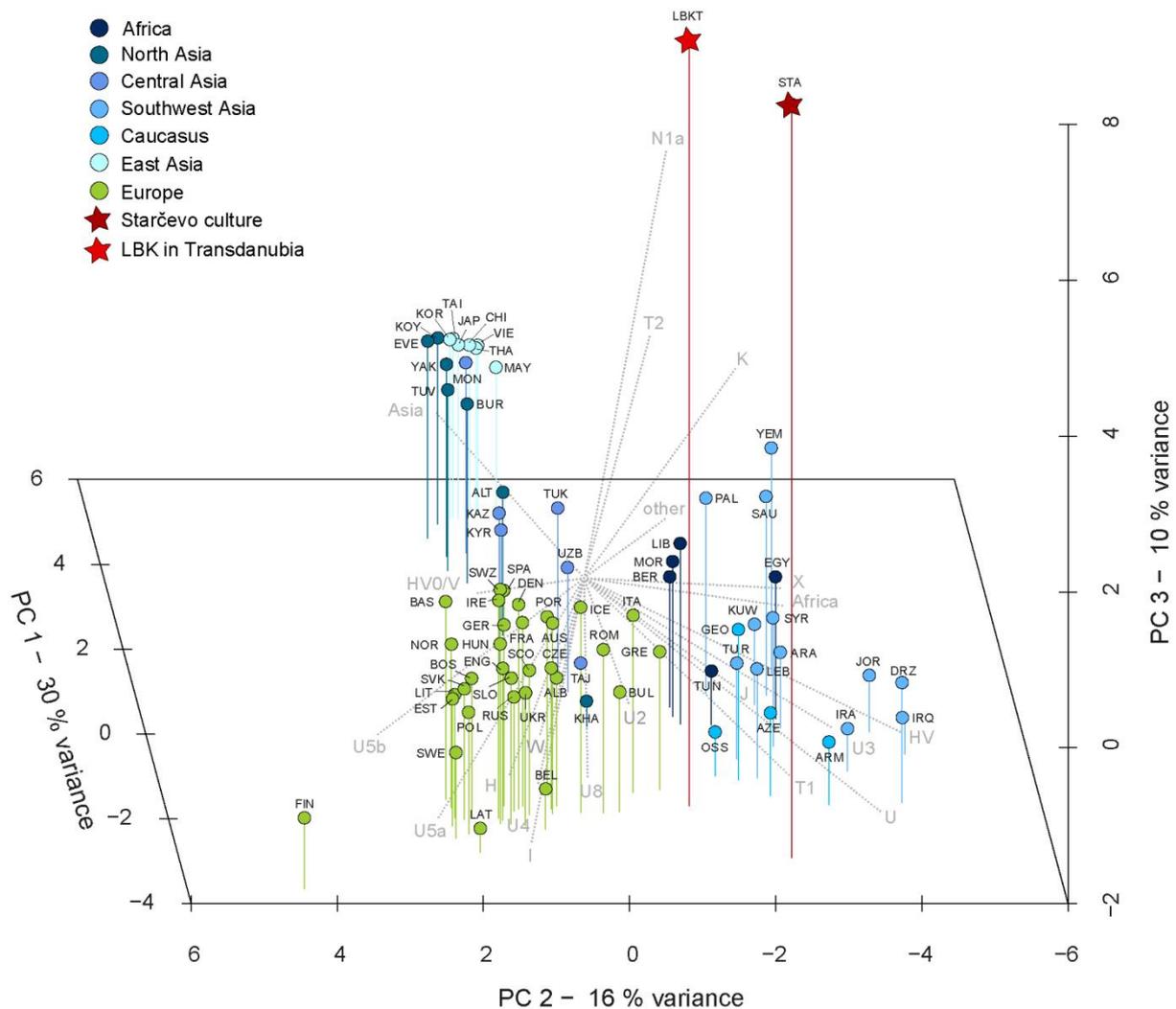


Figure S2. PCA comparing the Starčevo, LBKT and modern-day mtDNA data.

PCA of the STA, LBKT and 73 present-day populations of Eurasia, and North Africa plotted in a three dimensional space. Colors indicate populations from different Eurasian and African regions. The contribution of each haplogroup is superimposed as grey component loading vector. The first three principal components of the PCA display 56% of the total genetic variation. Population information and haplogroup frequencies are listed in Dataset S12.

Population codes: Linearbandkeramik culture in Transdanubia (LBKT), Starčevo culture (STA), Albanians, Macedonians (ALB), Altai (ALT), Arabs in UAE, Oman, Qatar (ARA), Armenians (ARM), Austrians (AUS), Azeris (AZE), Basques (BAS), Belarusians (BEL), Berber (BER), Bosnians, Croatians, Serbians (BOS), British

(ENG), Bulgarians (BUL), Buryats (BUR), Chinese, Tibetan (CHI), Czechs (CZE), Danes (DEN), Druze (DRZ), Egyptians (EGY), Estonians (EST), Evenks (EVE), Finns (FIN), French (FRA), Georgians (GEO), Germans (GER), Greeks (GRE), Hungarians (HUN), Icelanders (ICE), Iranians (IRA), Iraqi (IRQ), Irish (IRE), Italians (ITA), Japanese (JAP), Jordanians (JOR), Kazakhs (KAZ), Koreans (KOR), Koryaks (KOY), Kuwaiti (KUW), Kyrgyz (KYR), Latvians (LAT), Lebanese (LEB), Libyans (LIB), Lithuanians (LIT), Malays (MAY), Khants, Mansi (KHA), Mongolians (MON), Moroccans (MOR), Norwegians (NOR), Ossetians (OSS), Palestinians (PAL), Poles (POL), Portuguese (POR), Romanians (ROM), Russians (RUS), Saudi Arabians (SAU), Scots (SCO), Slovaks (SVK), Slovenians (SLO), Spaniards (SPA), Swedes (SWE), Swiss (SWZ), Syrians (SYR), Taiwanese (TAI), Tajiks (TAJ), Thai (THA), Tunisians (TUN), Turkmen (TUK), Turks (TUR), Tuvians (TUV), Ukrainians (UKR), Uzbeks (UZB), Vietnamese (VIE), Yakuts, Yukaghir (YAK), Yemenis (YEM).

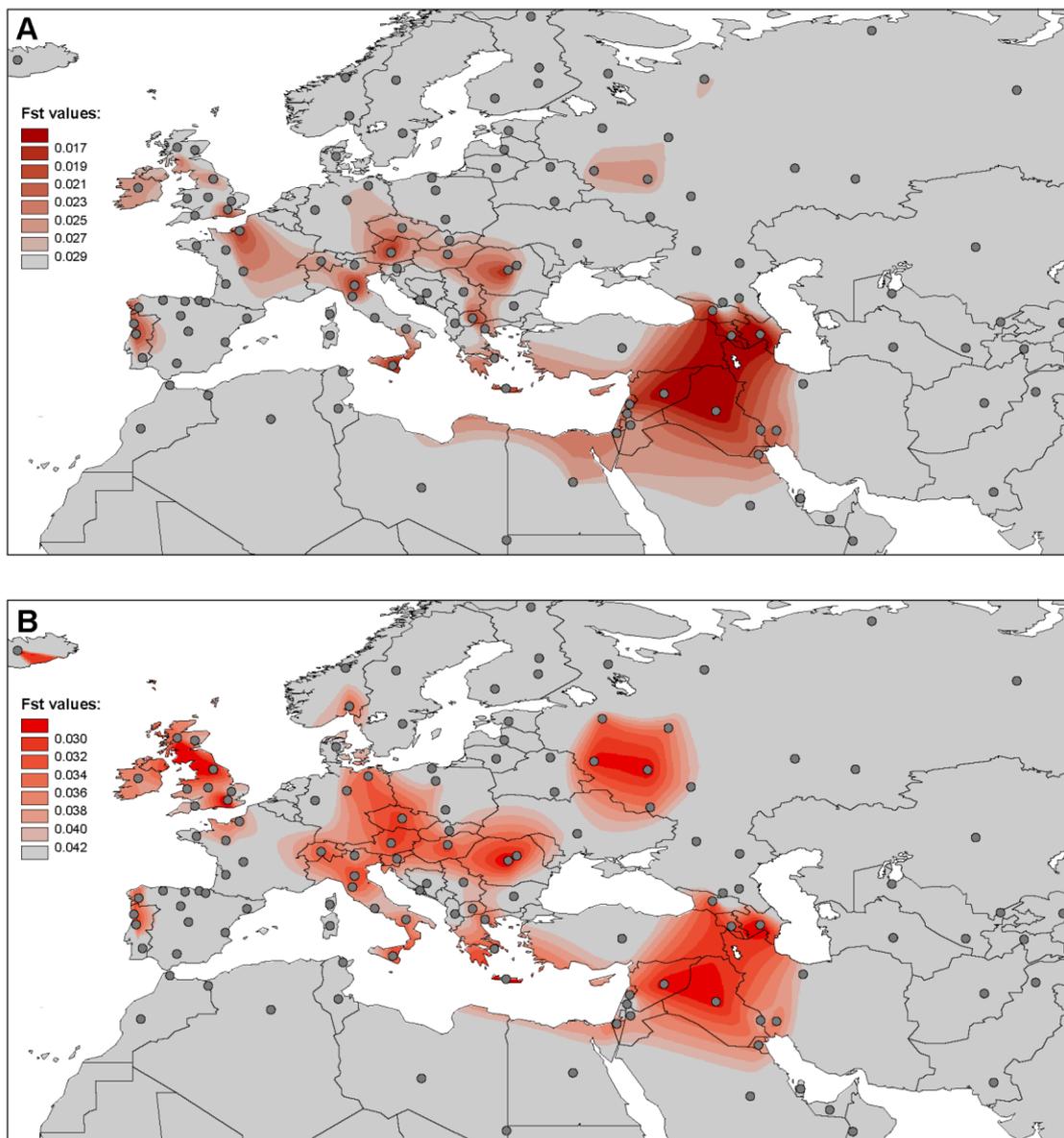


Figure S3a-b. MtDNA genetic distance maps of the STA, LBKT and 130 present-day populations.

Genetic distances (F_{st}) between the STA (A) and LBKT (B) and 130 present-day populations of Eurasia and North Africa were computed based on HVS-I sequences and visualized on a geographic map. Grey dots denote the location of modern populations. Color shadings indicate the degree of similarity or dissimilarity of the Neolithic cultures to these populations. Short distances and great similarities are marked by dark red (STA) and red (LBKT) areas. F_{st} values are scaled by an interval range of 0.002. F_{st} values higher than 0.029 (STA), 0.042 (LBKT) were not differentiated (grey areas). Population information and F_{st} values are listed in Dataset S13.

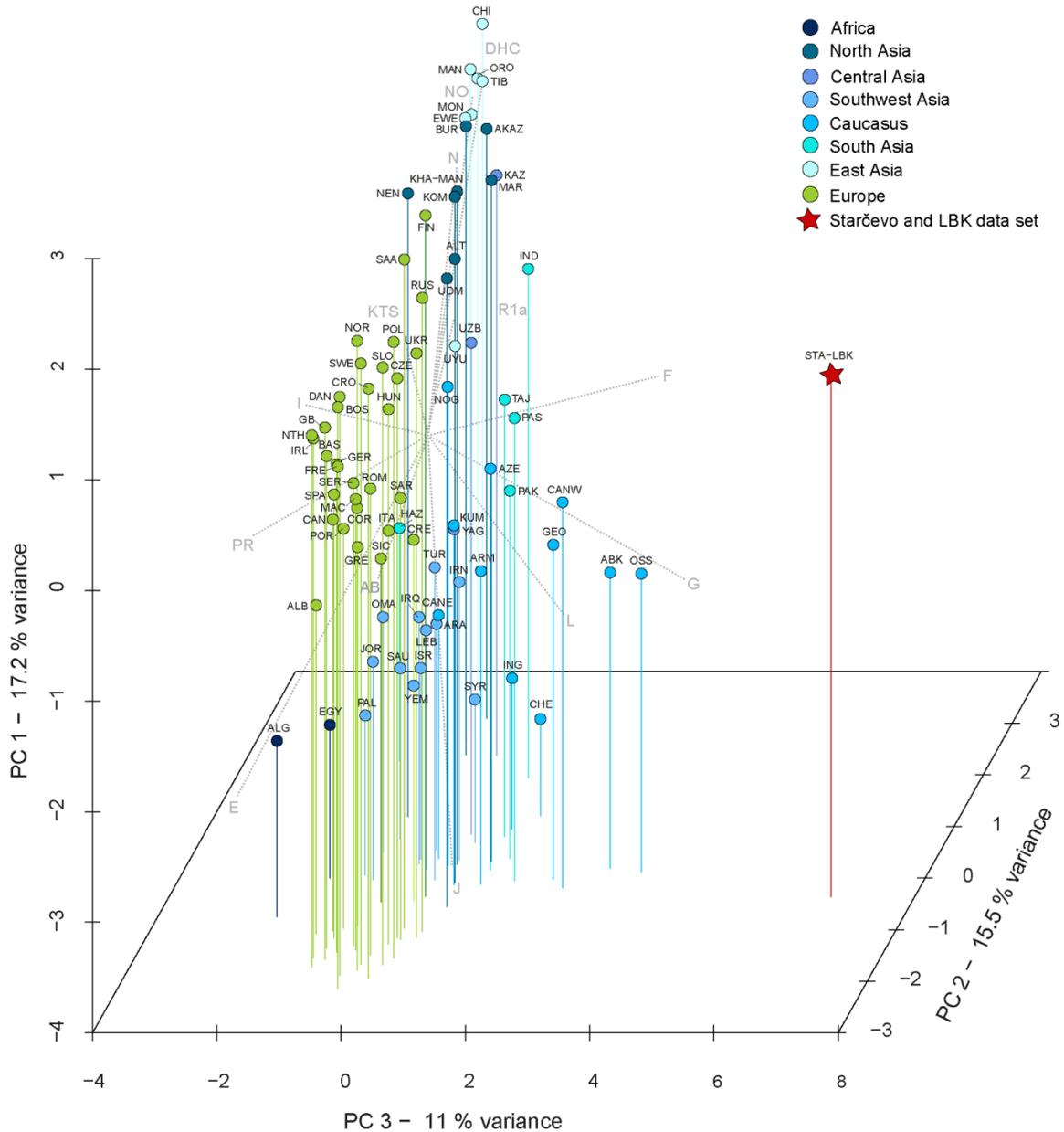


Figure S4. PCA from Y chromosomal data of the STA-LBK sample set and 80 modern-day populations.

The PCA, based on the frequencies of 13 Y chromosomal haplogroup in the STA-LBK sample set and 80 present-day populations from Eurasia and North Africa, were plotted in a three dimensional space. Colors of data points indicate populations from different Eurasian and African regions. The contribution of each haplogroup is superimposed as grey component loading vector. The first three principal

components of the PCA display 43.7% of the total genetic variation. Population information and haplogroup frequencies are listed in Dataset S14.

Population codes: Starčevo and Linearbandkeramik culture (STA-LBK), Abkhazians (ABK), Albanians (ALB), Algerians (ALG), Altai (ALT), Altai Kazakhs (AKAZ), Arabs in UAE, Qatar, Kuwait (ARA), Armenians (ARM), Azeri (AZE), Basques (BAS), Bosnians (BOS), British (GB), Buryats (BUR), north-east Caucasus (CANE), north-west Caucasus (CANW), Chechens (CHE), Chinese (CHI), Corsicans (COR), Crete (CRE), Croatians (CRO), Czech-Slovakians (CZE), Danish (DAN), Dutch (NTH), Egyptians (EGY), Ewenki (EWE), Finns (FIN), French (FRE), Georgians (GEO), Germans (GER), Greek (GRE), Hazara (HAZ), Hungarians (HUN), Indian (IND), Ingush (ING), Iranians (IRN), Iraqis (IRQ), Irish (IRL), Israeli (ISR), Italians (ITA), Jordanians (JOR), Kazakhs (KAZ), Komi (KOM), Kumyks (KUM), Lebanese (LEB), Macedonians (MAC), Manchu (MAN), Mansi & Khanti (KHA-MAN), Mari (MAR), Mongolians (MON), Nenets (NEN), Nogays & Kara Nogays (NOG), Norwegians (NOR), Omani (OMA), Oroqen (ORO), south-north Ossetians (OSS), Pakistani (PAK), Palestinian (PAL), Pashtun (PAS), Poles (POL), Portuguese (POR), Romanians (ROM), Russians (RUS), Sami (SAA), Sardinians (SAR), Saudi Arabians (SAU), Serbians (SER), Sicilians (SIC), Slovenians (SLO), Spaniards (SPA), Spaniards in Canary Islands (CAN), Swedes (SWE), Syrians (SYR), Tajik (TAJ), Tibetans (TIB), Turks (TUR), Udmurts (UDM), Ukrainians (UKR), Uyghurs (UYU), Uzbeks (UZB), Yagnobi (YAG), Yemeni (YEM).

Supplementary Datasets

- Dataset S1: Basic information of the archaeological sites.
- Dataset S2: Anthropological data and ¹⁴C dating of the sampled skeletons.
- Dataset S3: Summary of mtDNA and Y chromosome results from western Hungary and Croatia.
- Dataset S4: Detailed results of the GenoCoRe22 SNP multiplex assay.
- Dataset S5: Detailed results of the GenoY25 SNP Multiplex and singleplex PCR analyses.
- Dataset S6: Summary of published prehistoric mtDNA data used for population genetic analyses.
- Dataset S7: Mt haplogroup frequencies of 16 hunter-gatherer, Neolithic, and early Bronze Age cultures, used for the PCA.
- Dataset S8: Fst values, its p values and Slatkin matrix of 16 prehistoric cultures.
- Dataset S9: Results of AMOVA.
- Dataset S10: Results of ASHA of the Central/North European hunter-gatherers, Neolithic and Early Bronze Age cultures.
- Dataset S11: Results of TPC with STA, LBKT, LBK and the hunter-gatherer metapopulation of Central and North Europe.
- Dataset S12: Population information and mtDNA haplogroup frequencies used for PCA with Neolithic Starčevo, LBKT cultures and 73 present-day populations.
- Dataset S13: Population information and Fst values used for mapping the genetic distances between the Carpathian Basin cultures and 130 present-day populations.
- Dataset S14: NRY haplogroup frequencies of 80 modern populations and the combined Starčevo-LBK dataset.
- Dataset S15: Information of 100 modern-day populations and their Fst values, counted from the STA-LBK dataset, used for the Y chromosomal genetic distance map.
- Dataset S16: List of primers, used for the mtDNA and NRY singleplex amplifications.
- Dataset S17: Genetic profiles of the coworkers.