Supplementary Information to

Bronze Age communities from Western Hungary reveal complex population histories

Dániel Gerber, Bea Szeifert, Orsolya Székely, Balázs Egyed, Balázs Gyuris, Julia I. Giblin, Anikó Horváth, László Palcsu, Kitti Köhler, Gabriella Kulcsár, Ágnes Kustár, Vajk Szeverényi, Szilvia Fábián, Balázs Gusztáv Mende, Mária Bondár, Eszter Ari, Viktória Kiss, Anna Szécsényi-Nagy

Table of content

1) Archaeological, isotope and anthropological data

by Mária Bondár, Szilvia Fábián, Viktória Kiss, Kitti Köhler, Gabriella Kulcsár, Ágnes Kustár

1.1) Summary of the archaeological context and preliminary genetic analyses of BAD002

Skull cults were prevalent among Mesolithic and Neolithic groups in Europe and the Middle East as well. In Late Copper Age Baden culture this practice was rare, only a few exceptions are known, including the context of Feature 415 where a complete skeleton of a child was found, referred as BAD002 in this study (Figure S.1.1.1) from Balatonlelle-Rádpuszta site [1], approximately 30 km away from Balatonkeresztúr-Réti-dűlő site. BAD002 corresponds to a 7-8 years old male individual surrounded by four skulls and one fragmented skull of children and a young adult male. Two of them were presumably close relatives of BAD002 according to mtDNA results. The skulls lack mandibles and were secondary used placing them into the burial of BAD002, contemporaneously with the funeral act. Since previous findings of skull cult all over from Europe appear it does not necessarily point out his western origin, while copper and black jet beads [2], as well as mitochondrial haplogroup U5b3 from two skulls in Feature 415 does [3]. Preliminary results revealed that BAD002 had an outstandingly high HG component compared to known Neolithic and Copper Age groups from the Carpathian Basin, therefore we included it in this study.

Fig. S.1.1.1

Balatonlelle-Rádpuszta, the complete skeleton of BAD002 in Feature 415 surrounded by pottery and skulls of children and an adult individual. According to Bondár & Szécsényi-Nagy 2020[3], two of the skulls possess the same mitochondrial haplogroup as BAD002, suggesting that at least some of the skulls belong to close relatives of this young male individual. Remains are dated to 3530-3370 cal BCE (95.4% CI).

1.2) Overall description of Balatonkeresztúr site

Balatonkeresztúr-Réti-dűlő site (Fig. S.1.2.1) is located at 3 km from the southern shore of Lake Balaton. In 2003–2004 an area of 45,000 m^2 was investigated preceding the construction of the M7 motorway. As a result, the traces of occupations associated with nine archaeological cultures spreading over eight chronological periods were identified by the analysis of 2,976 archaeological features [4,5]: Middle and Late Copper Age (Balaton-Lasinja, Furchenstich, Boleráz and Baden cultures), Early Bronze Age (Somogyvár-Vinkovci and Kisapostag/earliest Encrusted Pottery cultures), Middle Bronze Age (Transdanubian Encrusted Pottery culture), Late Iron Age (La Téne D), Migration period (Langobards), Árpádian Age (12-13th century) and Late Middle Ages (13-15th century[6]). Human remains were discovered from settlement features: one pit (Somogyvár-Vinkovci) contained bones of one individual, while another pit (Transdanubian Encrusted Pottery) contained the remains of 8 human bodies. East from the settlement of the Kisapostag culture a small cemetery with 11 skeletal burials was discovered. Most of the burials contained no grave goods except for two with small bronze ornaments [6].

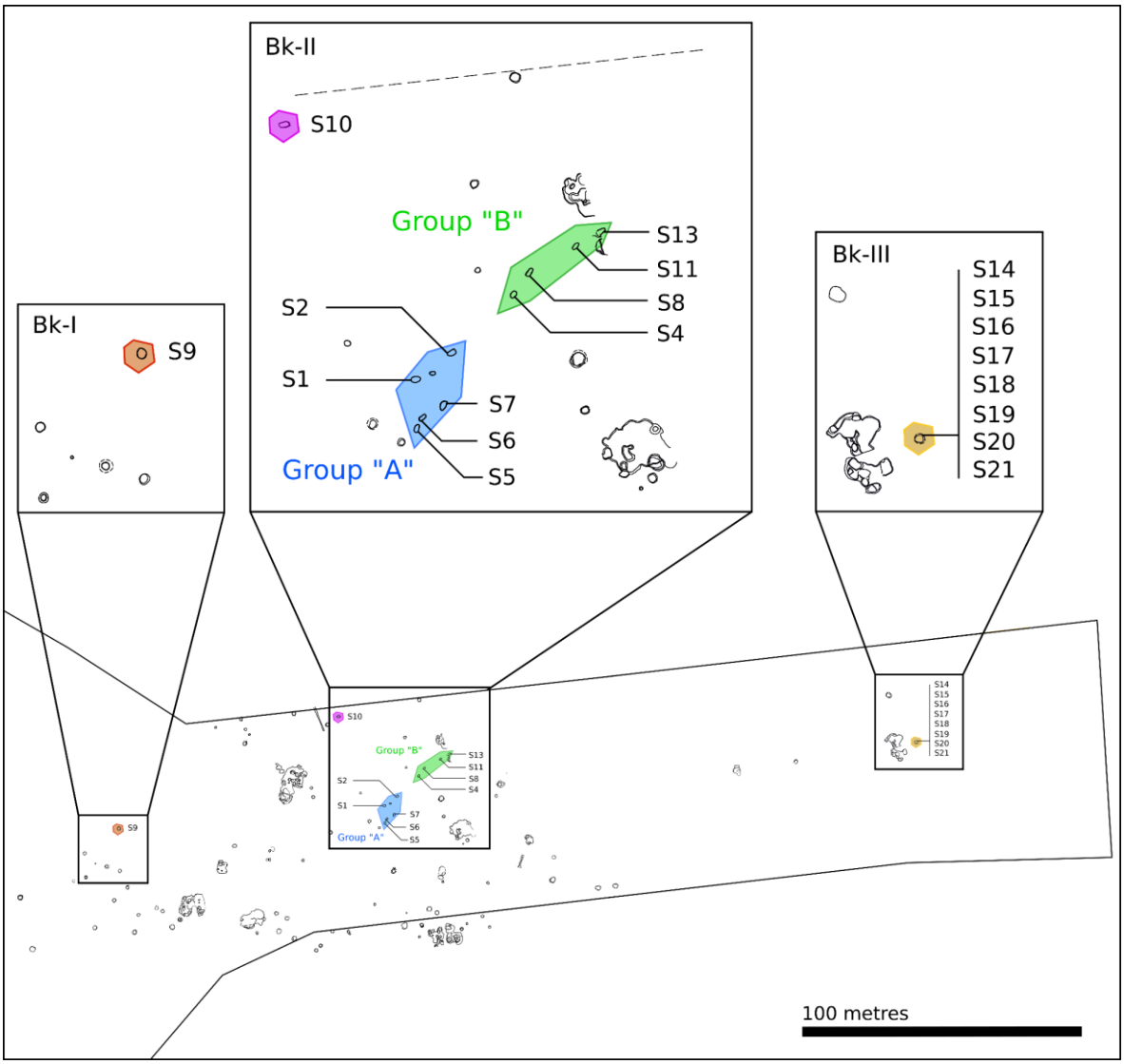


Fig. S.1.2.1

Map of Balatonkeresztúr-Réti-dűlő site, where graves and grave groups are showed in separate enlarged boxes. The only child grave is situated near to the edge of the excavation site for possible reasons explained in the 'Discussion' section in the main text of the manuscript. The mass grave of Bk-III is also situated far from the core of Bronze Age objects (uncoloured) at the lower point of the site.

1.3) Overall description of the studied era

The first thousand years of the Bronze Age in the Carpathian Basin – between 2500-1500 BCE – is an important period in connection with the spread of bronze metallurgy, new pottery styles, development of new settlements and emerging social inequalities in Central Europe. While at the first part of the period under study during the Early Bronze Age in Hungary (2500-2000/1900 BCE; parallel with the onset and older phase of the Central European Early Bronze Age, Reinecke Br A0 and A1) we can find arsenical bronze axes and daggers, as well as ornaments made of pure copper, after 2000/1900 BCE tin bronze weapons and tools appear in all spheres of everyday life [7–9]. Two different burial practises, cremation and inhumation can be observed during the Early Bronze Age [10,11]. After the beginning of the Middle Bronze Age in western Hungary around 2000/1900 BCE (parallel with the beginning of Central

European developed Early Bronze Age, Reinecke Br A2) cremation burials were applied nearly exclusively, see *e.g.* the cemetery excavated at Bonyhád-Biogas factory, with only 6.5% inhumations of all burials that can be dated to the earliest period of the cemetery [12]. From this point of view the physical anthropological analyses, as well as stable isotope and ancient DNA study of the few known inhumations and special settlement pit burials are very important in the discovery of physical anthropological and genetic makeup, as well in understanding of social processes of the Early, Middle and Late Bronze Ages of the region.

1.4) Somogyvár-Vinkovci culture

In the first part of the Early Bronze Age in western Hungary hilltop/hillfort settlements as well as smaller and larger villages of this culture can be distinguished. Only a very small number of burials are known from this period: urn cremations and several inhumations. Beside normal burials a few inhumations from settlement pits were also discovered [10]. At Balatonkeresztúr-Réti-dűlő only one burial (individual S9) can be associated with this archaeological culture according to its ^{14}C date (3929±30 BP, 2560-2290 cal BCE, 95.4% CI). Remains of this individual (Fig. S.1.4.1) were discovered in a settlement pit, lying in a special position, on his back and with the knees and head on the left side, oriented with N/NE-S/SW [13]. Here, only pottery shards and animal bones were found, probably belonging to the waste pit. Bones were preserved in worse condition than other human remains discovered in the group of burials, suggesting bad life conditions. The skeletal remains of this individual consist of a moderately preserved skull and well preserved bone material. Determined as a male of 35-40 years with a calculated height of 154.8 cm, his anthropological features, mainly his skull shape (ultradolichokran curvoccipital) are resembling more to the preceding Copper Age population [14–16]. No particular pathological condition could be observed, except some cribra orbitalia (porotic lesions on the upper wall of the eye-orbit), which may resulted from the increased erythrocyte number that signalise a wide variety of pathological conditions from malnutrition to malaria [17].

Fig. S.1.4.1

Grave of individual S9 surrounded by animal remains, such as cattle metatarsus under his skull. His burial position shows resemblance to people of high steppe ancestry[13], that which this individual possesses too. Photo by Szilvia Fábián.

1.5) Kisapostag/Earliest Encrusted Pottery culture

The two-tiered settlement structure mentioned above, with scattered dwellings at smaller and larger villages, and sometimes hilltop/hillfort sites and settlements surrounded by ditches is characteristic for the communities of the Kisapostag/earliest Encrusted Pottery. Open settlements, like the ones discovered in Kaposvár and Balatonkeresztúr, consist of a few houses, storage and refuse pits [18,19]. Metal finds usually made of fahlore copper, rich in arsenic, antimony and silver impurities; tin bronzes can be observed in a few cases only at the end of the Early Bronze Age [19,20]. Before the 1990s, the practice of cremation burials placed in urns was solely associated with the Kisapostag/earliest Transdanubian Encrusted Pottery culture. In the last decades beside urn burials and scattered cremations, around 90 inhumation burials of this culture are documented[11]. Anthropological analyses, however, are available in a few cases only without detailed bioarchaeological testing [11,12,18]. Eleven inhumation burials can be associated with this culture according to archaeological features and ^{14}C dates (Supplementary Table S1) in Balatonkeresztúr site. The graves, with bodies placed in flexed position, laying on their side, were found east from the settlement of the same period in two groups. Group A (individuals S1, S2, S3, S5, S6, S7, N=6) and Group B (individuals S4, S8, S11, S13, N=4) and a single inhumation (individual S10) can be observed, from where individual S3 was not sampled due to the complete destruction of the poorly preserved remains during excavation. Group A consists of three juveniles and three adults, five males and a female (individual S7), while in Group B we find one juvenile male, and three adults, two males and one female (individual S13). One inhumation, found a bit further from the two burial groups, belongs to a young male individual. Three graves (individuals S10, S12 (cremated), S13) contained metal ornaments as grave goods. Many samples are only preserved as scattered and heavily destroyed remains, and most skeletons show slight or modest distortion. The overall pathological makeup of the population does not show extremities and the average adult biological age is high, listed below along with their archaeological artefacts in detailed descriptions:

Fig. S.1.5.1

Grave of individual S1, Group A. The body was placed on the right side in a flexed position, oriented with NE-SW. One piece of a flint was found in the grave pit. The remains consist of fragmented skeletal bones and skull, ^{14}C dated back to 3611 ± 31 BP (2120-1880 cal BCE, 95.4% CI), and were belonging to a male of over 40 years (determination of age at death based on the teeth abrasion: level 5). Sexualisation index: - 29 teeth remained, three upper and one lower molar were decayed (caries); the lower right first molar is ante mortem missing, indicated by the healed alveolus. Other pathological conditions were not revealed. The skull has a brachycranic (planoccipital?) shape, which feature is almost ubiquitous to this group. Photo by Szilvia Fábián.

Fig. S.1.5.2

Disturbed grave of individual S2, Group A. The remains and endogenous DNA are poorly preserved. The body was placed on the left side in a flexed position, oriented with NE-SW. Only several non-diagnostic pieces of Bronze Age pottery were found in the burial pit. The remains consist of fragmented bones, with absolute terms the grave is dated to 3609 ± 32 BP (2120-1880 cal BCE, 95.4% CI), and belongs to a male of 30-35 years (determination of age at death based on the teeth abrasion: level 3). Sexualisation index: -0,50. No molar decay was determined on the remaining 12 teeth, his skull type is brachycranic (planoccipital?). A rare anatomical variation of perforated fossa olecrani can be observed on the left humerus. Photo by Szilvia Fábián.

Fig. S.1.5.3

Grave of individual S4, Group B. The skeletal remains and the DNA were preserved poorly. The deceased was placed on its right side in a flexed position, oriented slightly NE-SW. Human remains only consist of a small amount of fragmentary bone material and teeth. The remains belong to a 17-19-year-old male individual (determination of age at death based on muscular adhesion surfaces, and teeth abrasion: level 2) with 16 remaining teeth and, with no visible pathological condition. One boar incisor and fragmentary pottery were recovered from the grave. ^{14}C *dating was not performed (the remains can be dated by the genetically determined father, individual S8, as it is shown in the 'Results' section in the main text). Photo by Szilvia Fábián.*

Fig. S.1.5.4

Grave of individual S5, Group A. Moderately preserved remains of a 16-18 years old male individual in a flexed position, oriented to N-S. Only several non-diagnostic Bronze Age pottery shards were found in the grave. Determination of age at death based on muscular adhesion surfaces, and teeth abrasion: level 2-3) with 32 remaining teeth. Skull type is brachycranic (planoccipital?), besides some severely decayed molars (left upper first molar: caries) no other pathological conditions can be observed. Some fragmentary pottery was also recovered from the grave, ^{14}C dating was not performed (the remains can be dated by the genetically determined father, individual S11, as it is shown in the 'Results' section in the main text of the manuscript). Photo by Szilvia Fábián.

Fig. S.1.5.5

Grave of individual S6,Group B. The skeletal remains and the DNA were preserved poorly. The body was placed on the left side in a heavily flexed position, oriented with NE-SW. The moderately preserved bone fragments were ^{14}C dated to 3571 ± 31 BP (2030-1770 cal BCE, 95.4% CI), and belonged to a 17-18 year-old male individual (sexualisation index: +0,71; determination of age at death based on muscular adhesion surfaces, and teeth abrasion: level 2). 24 remaining teeth; besides one decayed molar (right lower first molar: caries) no other visible pathological condition was revealed on the skeletal remains. Photo by Szilvia Fábián.

Fig. S.1.5.6

Grave of individual S7, Group A. Moderately preserved remains of a 35-50 years old female individual (sexualisation index: -0,18; determination of age at death based on sutures ossification, and teeth abrasion), ^{14}C dated to 3611 ± 32 BP (2120-1880 cal BCE, 95.4% CI). The skull was of brachycranic type with a long facial structure. Beside 28 remaining teeth (with no caries), some are ante mortem missing (right lower second premolar and second molar, and left lower second premolar). Extended porotic lesions were found in the skull (os frontale et parietale), in addition with some enthesopathy, i.e. bone spikes caused by heavy labour on the right calcaneus (left is not examined because of its fragmentary condition). No grave goods were recovered from the grave besides some non-diagnostic Bronze Age pottery sherds. Photo by Szilvia Fábián.

Fig. S.1.5.7

Grave of individual S8, Group B. The body was placed on the left side in a flexed position, oriented with NE-SW. The poorly preserved bones belonged to a 30-40 years old male (sexualisation index: +0,80; determination of age at death based on abrasion of teeth: level 4, sutures ossification and the surface of facies auricularis) with a calculated height of 168.2 cm. Just like individual S2 the body had an anatomical variation of perforated fossa olecrani on the right humerus, likely as a result of blood relationship (individuals S2 and S8 were 2nd degree relatives). Only four teeth remained, showing no signs of decay (caries). On the diaphysis of his right tibia periostitis is observed, while both calcanei show signs of enthesopathy. Photo by Szilvia Fábián.

Fig. S.1.5.8

Grave of individual S10, the only child grave from Bk-II. Moderately preserved remains of a 7-8 years old male individual (determination of age at death based on teeth, and size of long bones), ^{14}C dated to 3661 ± 30 BP (2140-1940 cal BCE, 95.4% CI), placed in a flexed position, oriented to NE-SW. The bones do not show any sign of pathological conditions. Copper or bronze bracelet on the arm shows higher social status for this individual. Photo by Szilvia Fábián.

Fig. S.1.5.9

Grave of individual S11, Group B. Moderately preserved remains of a 34-43 years old male (sexualisation index: +1,16, determination of age at death based on the trajectory of proximal epiphysis of the humerus and femur and abrasion of the remaining 26 teeth: level 3), ^{14}C dated to 3705 ± 30 BP (2200-1980 cal BCE, 95.4% CI), placed in a flexed position, oriented to NE-SW. The dolichocranic skull was highly deformed under soil pressure. The bones are distorted, similarly to many other individuals from this archaeological culture. The calculated height of the body is 171.9 cm, and the individual has had a severe walking condition for his left leg owing to an early life injury or a developmental disorder. The grave contained only some non-diagnostic Bronze Age pottery sherds. Photo by Szilvia Fábián.

Fig. S.1.5.10

Grave of individual S13, Group B. Well preserved remains of a 35-45 years old female (sexualisation index: -0,81, determination of age at death is based on sutures ossification, the surface of facies symphyses and teeth abrasion: level 4-5), ^{14}C dated to 3618 ± 30 BP (2120-1890 cal BCE, 95.4% CI). The body was placed on the left side in a flexed position, oriented with N-S. On the other hand, the remains had a strange body placement compared to other individuals, implying special treatment for this individual. The skull is of brachycranic type, the calculated height of the body is 158.7 cm, and the individual had a rare anatomical variation of extra sutural bones and maxillary prognathy. 29 teeth remained, with caries observable only on lower left second molar; lower left third molar had fallen out ante mortem. The only pathological condition was one missing and one decayed tooth. The remains of the individual had several fragments of metal beads made of copper sheet at the left side of the head along with a small burnt animal bone fragment, as well traces of green patina on the right side of the skull. The remains and skull preserved the best among all individuals in the site enabling facial reconstruction, grave reconstruction and detailed anthropological description for this individual by using genetic data as well, see in section '4) Facial reconstruction'. Photo by Szilvia Fábián.

1.6) Transdanubian Encrusted Pottery culture

This archaeological culture, distributed in a large area of Western Hungary, is characterised by cremation burials with a large number of ornamentations highlighted by white encrustations on the dark (brown or black) surface of the pottery. The origin of this pottery was connected to the encrusted decoration of the Vučedol culture [21,22]. However, later studies emphasised the continuous development from the Kisapostag/Earliest Encrusted Pottery tradition [12,19,23]. Cremation burials of the older period can be dated between the 19th and 17th centuries BCE, while younger periods between the 17th and the 15th centuries BCE [24]. Rescue excavations of the last decades resulted in the discovery of the large villages of this culture. These data suggest both smaller and larger, permanently occupied settlements with a relatively small population that repeatedly shifted their location. Besides single-layered, hilltop and (fewer) fortified settlements were also established. Local production of metallurgy is proven by a small number of moulds found in these settlements, as well as the bronze jewellery depositions of the Tolnanémedi hoard horizon [25,26]. Urn burials and scattered cremation burials both occur, while

inhumation is hardly ever documented [19]. However, these irregular settlement burials and several inhumations in ‘urnfields' of the culture are important for the study of the anthropological and genetic makeup of the Middle Bronze Age population in western Hungary. During 2003 a round pit was excavated in Balatonkeresztúr site in which eight human skeletons of three adults and five children were found without evidence of grave goods (Figure S.1.6.1). The feature was identified as a household refuse pit utilised for mortuary purposes dating to the Late Copper Age or the Early Bronze Age [27,28]. ^{14}C samples from the pit measured by the VERA AMS laboratory in Vienna, Austria yielded a date of 3455±35 BP (1890-1640 cal BCE, 95.4% CI) and from HEKAL AMS laboratory in Debrecen, Hungary yielded a date of 3387±30 BP (1870-1540 cal BCE, 95.4% CI) which falls to the period of the Hungarian MBA. This multiple pit-burial, however, could be considered as an important funerary assemblage in the light of the recently heightened interest in irregular prehistoric burials [29,30]. The physical anthropological examinations of the Balatonkeresztúr remains prove the presence of the *planoccipital brachycranic taurid* type crania [28], a physical trait which could be linked to the Bell Beaker group [31]. This outcome is significant in the context of cremation tradition practised widely by Transdanubian populations of the period. Prior to the examination, virtually nothing was known about the physical anthropological makeup of these communities. Individuals S14 to S21 belong to this group: circles with solid lines represent first degree, dashed lines second degree relationships (Figure S.1.6.2).

Fig. S.1.6.1

Mass Grave B-938 of Bk-III. Photo by Szilvia Fábián.

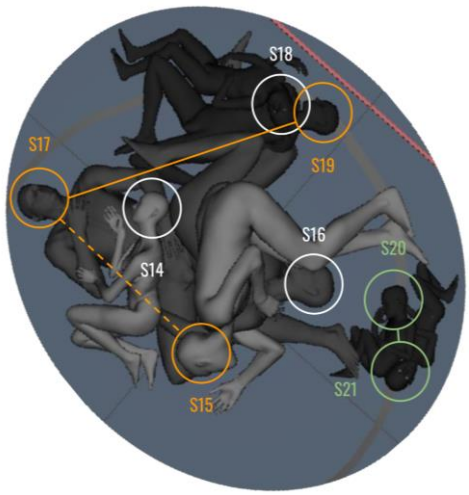


Fig. S.1.6.2

Digital reconstruction by Zsolt Réti of individuals from the mass grave of Bk-III. Connected individuals are related (solid line for first degree, dashed line for second degree).

1.7) Anthropological and paleopathological analyses

By Kitti Köhler, Balázs Gusztáv Mende

The condition of the studied anthropological material is medium, or sometimes poor. In the case of most of the skulls, the facial part was reconstructed only in a few cases, while the postcranial remains are relatively well preserved. The finds are hosted in Rippl-Rónai Museum, Kaposvár. The estimation of age-at-death and the sex determination were based on the methods commonly used in physical anthropology [32–39]. The paleopathological lesions were investigated macroscopically [40–42]. Bk-I EBA pit burial and Bk-II inhumation burials are first published here, while the physical anthropological analyses

of S14 to S21 individuals from Bk-III pit burial were already published[28], and examinations are here completed by a detailed archaeological interpretation and results of isotope and aDNA testing.

1.8) Radiocarbon dates of the Balatonkeresztúr site

by Vajk Szeverényi

Radiocarbon dating was performed for all Bronze Age horizons of the site. Where ^{14}C data are missing, DNA based kinship supports the corresponding dating. Overall 10 samples have been radiocarbon dated from the site, one sample was taken from the earliest Bronze Age burial connected to the period of the Somogyvár-Vinkovci type material (S9), six samples from Early Bronze Age graves with Kisapostag type material (four from Grave Group A (S1, S2, S6 and S7) and two from Grave Group B (S11 and S13), while two individuals were sampled from the Middle Bronze Age mass grave (S16 and S17).

All radiocarbon measurements were carried out at the HEKAL AMS C-14 facility of the Institute for Nuclear Research, Debrecen [43,44], except for individual S16, which was measured at the VERA AMS facility in Vienna. The dates were calibrated with the 'OxCal' v4.4 software [45] using the IntCal20 Northern Hemisphere radiocarbon calibration curve [46], see the individually calibrated dates at Supplementary Table S1. Individual S9 is dated to ca. 2560-2290 cal BCE (all calibrated dates are presented at two-sigma range). The dates from the graves of the BK-II phase range between 2200 and 1770 cal BCE, while BK-III dates fall between 1890 and 1540 cal BCE.

Bayesian analysis using the 'OxCal' software was implemented to attempt to obtain more precise age estimates. For the first model, independently identified typological phases were introduced as prior information. As a result of the analysis (Figure S.1.7.1), the ranges of the Somogyvár and Encrusted Pottery dates changed somewhat, moving the more probable part of the date ranges towards the Kisapostag dates. The latter did not change considerably as the result of the Bayesian analysis.

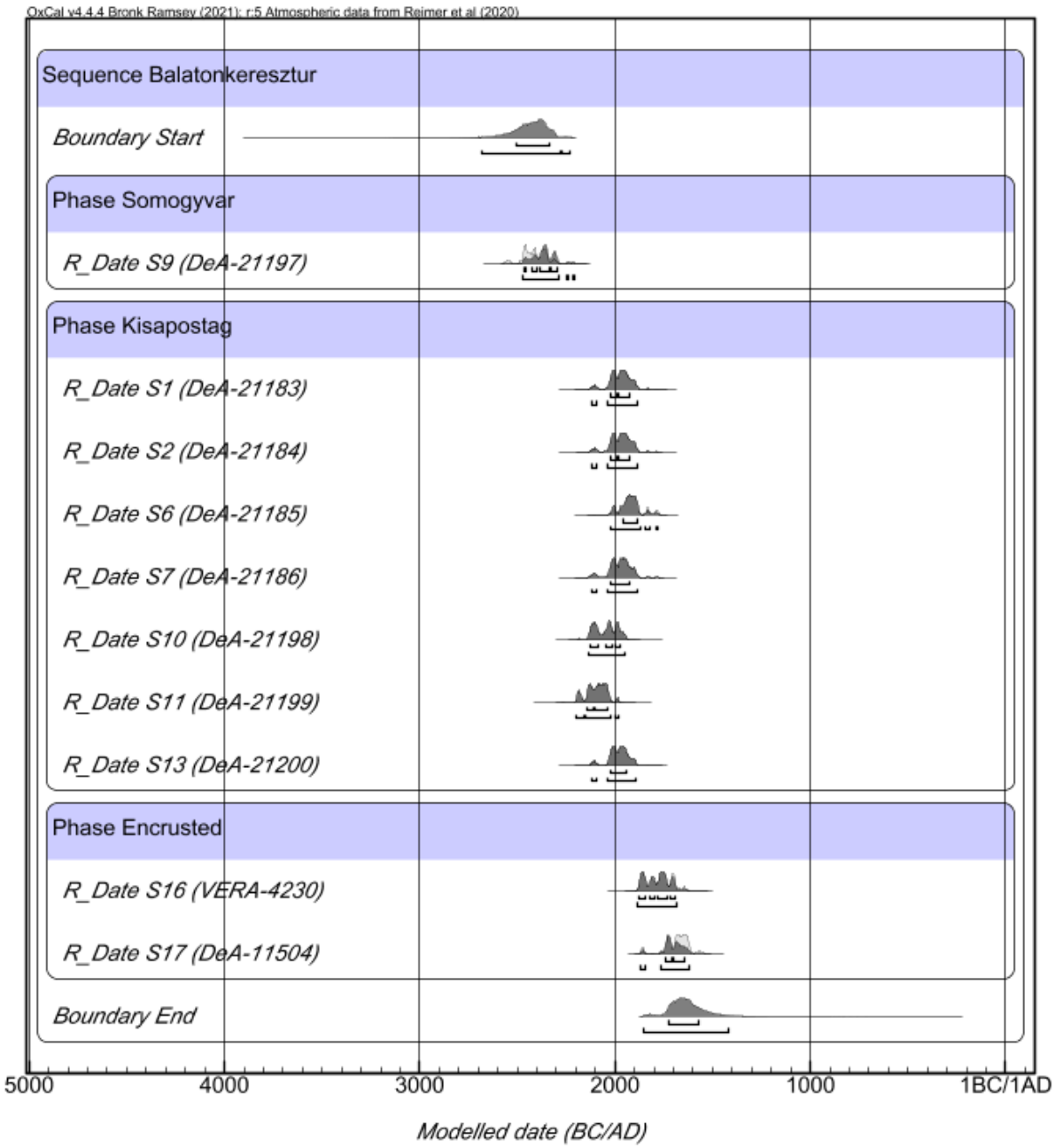


Fig. S.1.7.1

Modelled radiocarbon dates of samples from Balatonkeresztúr site, a multiplot for all three Bronze Age period (BK-I-III) radiocarbon series.

Only the Kisapostag dates were included in the second model as contexts belonging to a single phase. Grave Groups A and B show no archaeological signs of chronological difference, thus were treated as belonging to the same phase. As a result (Fig. S.1.7.2), the timespan of the use of the BK-II cemetery could be reduced to ca. 2120-1900 cal BCE.

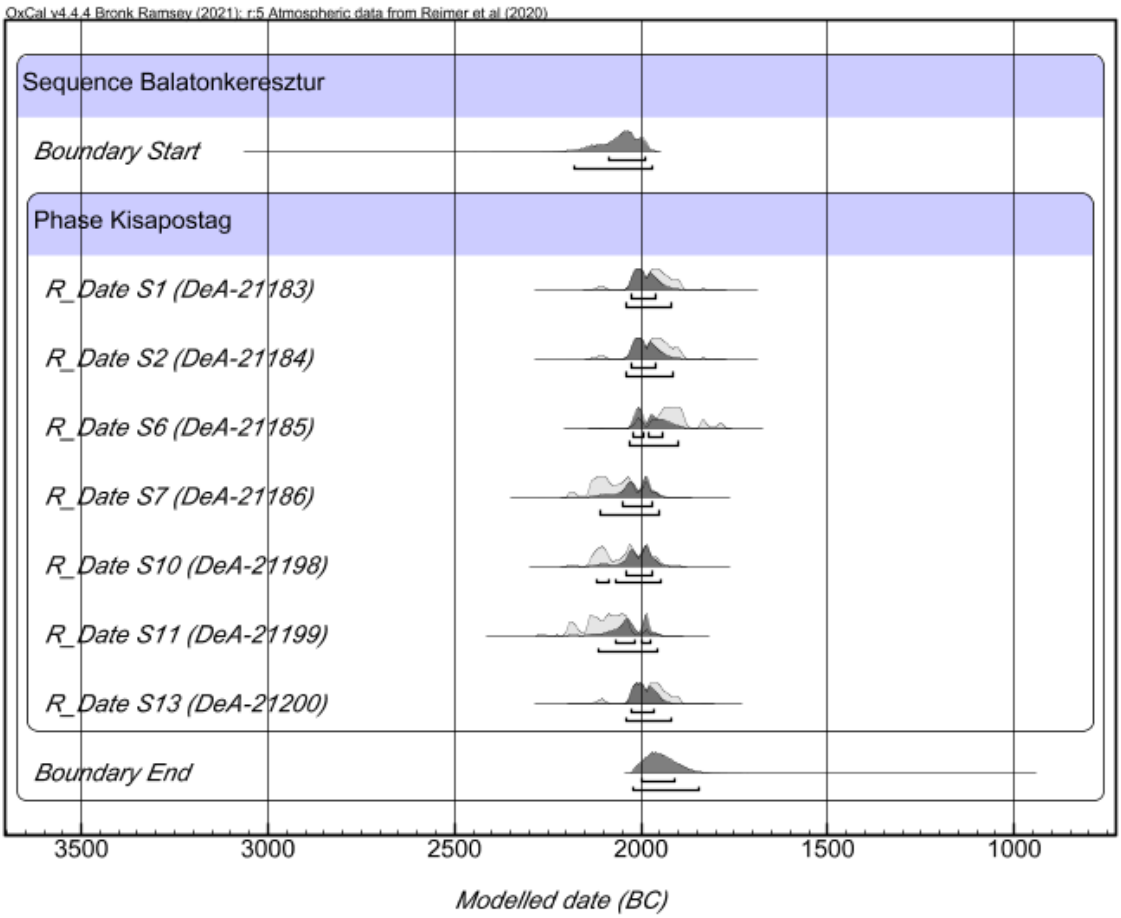


Fig. S.1.7.2

Modelled radiocarbon dates of samples from the Balatonkeresztúr site, a multiplot for the Kisapostag period (BK-II) radiocarbon series.

Finally, since the Middle Bronze Age samples derive from a mass grave in a pit, where the context and the full articulation of the skeletons indicate that the individuals deceased at the same time or very close to each other in time, the two dates were combined to date a single event. As a result (Fig. S.1.7.3), the combined date for the mass grave is 1870-1620 cal BCE (95,4%), where the timespan between 1770 and 1620 cal BCE has a significantly greater probability (93,1%).

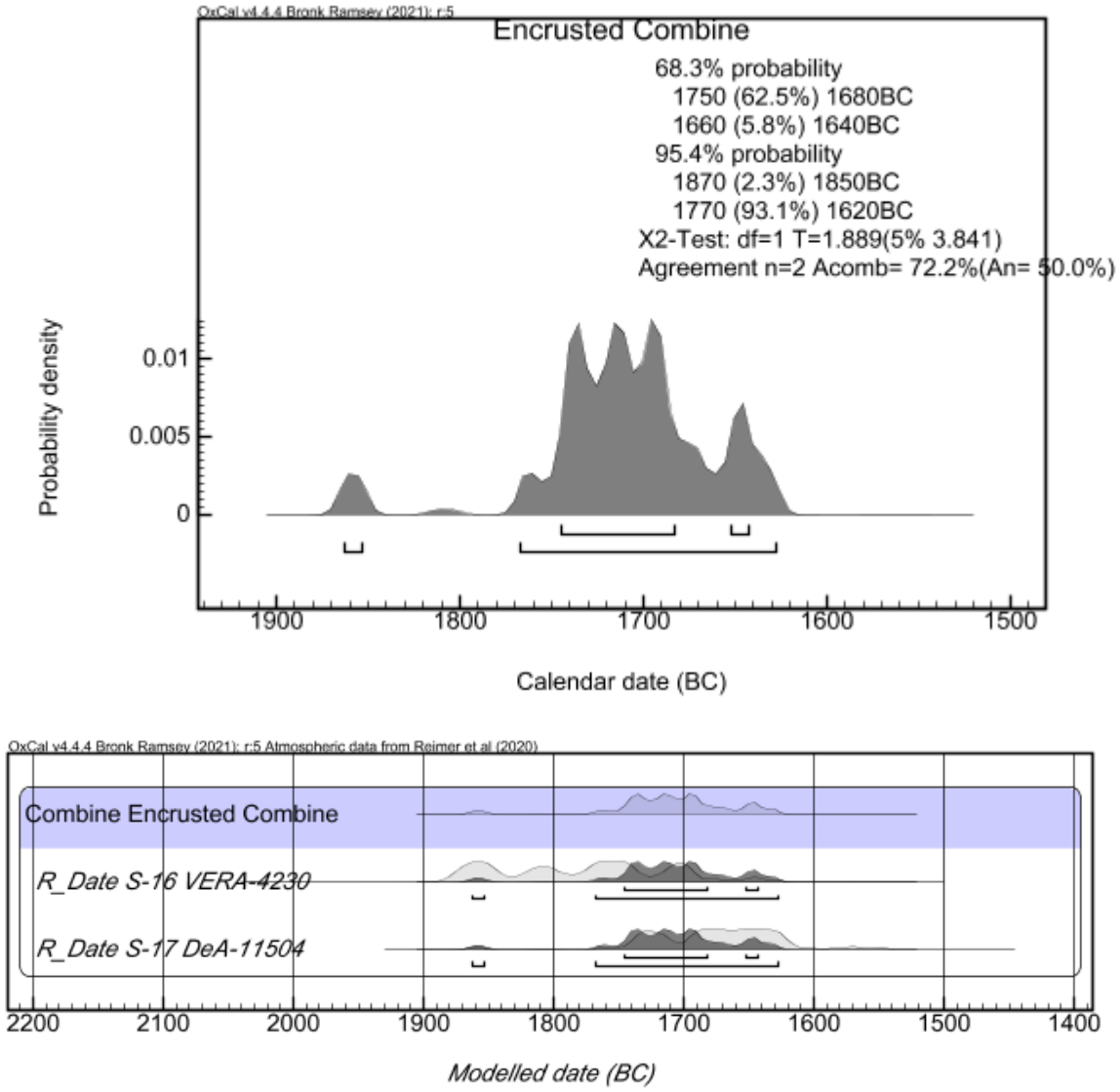


Fig. S.1.7.3

Combined radiocarbon date of samples, from the Encrusted Pottery period (BK-III) mass grave from the Balatonkeresztúr site.

1.9) Sr isotope data of Balatonkeresztúr site

by Julia I. Giblin, Anikó Horváth, László Palcsu

Samples from individuals S1-S13 were analysed for strontium ($^{87}Sr/^{86}Sr$ ratio) by the ICER laboratory, Atomki, Debrecen (Hungary), where the bone and enamel samples were physically pre-cleaned using a Dremel 300 drill equipped with a diamond tipped tool. Any remaining dentin on the teeth or cancellous bone on the cortical bone were removed using a diamond tipped Dremel tool attachment. Bones were repeatedly cleaned of dust by ultrasonication in milliQ water, after which they were rinsed in a 1 M acetic acid solution to remove diagenetic carbonate. Then, the samples were rinsed 3 times in milliQ water and dried overnight at 60 °C. 5-20 mg of enamel and 20 mg of bone samples, respectively, were weighed and placed into cleaned, labelled PFA beakers. Enamel samples were digested twice in 1ml of 14M HNO_3, bone sample was digested in 1 ml of 14M HNO_3 and 100 µl of $HClO_4$ and evaporated to

dryness on hotplate at 120 °C. When the sample was fully demineralized, it was taken up in 8M HNO_3 and loaded into the pre-cleaned Sr-specific ion exchange resin column to separate Sr from the other matrix elements. Ultrapure water (UPW) (Merck, 18 MΩ·cm) and doubly distilled 14 M nitric acid were used for sample digestion, resin cleaning, and elution. Crown ether-based Sr-Spec Resin (100–150 µm particle size) from Triskem International, France was pre-cleaned with 8M HNO3 and UPW prior to use. Sr was isolated from the matrix components and from Rb to avoid an isobaric overlap of $^{87}Sr^+$ and $^{87}Rb^+$. Additional blank and standard solutions were included to verify the blank and accuracy of the chemical preparation. All sample preparation occurred in a Class 1000 cleanroom. Standard solution was prepared from the NBS987 $SrCO_3$. Strontium isotope ratios were measured using a Neptune Plus MC-ICPMS (multi-collector inductively-coupled-plasma mass spectrometer,Thermo Scientific), equipped with an Aridus-3 (CETAC) desolvating system at the Isotope Climatology and Environmental Research Centre (ICER) in Debrecen, Hungary. The $^{87}Sr/^{86}Sr$ ratio was corrected for instrumental mass discrimination using $^{88}Sr/^{86}Sr = 8.375209$, as well as by applying an interference correction for $^{87}Rb^+$ and $^{86}Kr^+$ with $^{85}Rb^+$ and $^{83}Kr^+$, respectively. All values were normalised to the reported value of 0.710240 for NIST SRM 987.

Dental enamel samples analysed by J. Giblin (S14-20) were isolated using a diamond drill bit attached to a Dremel tool and then chemically pretreated at the Center for Anthropological Research (CAR) at Quinnipiac University in a 2% bleach solution (NaOCl) to remove organic contaminants (13 hours), followed by a 0.1 M acetic acid (CH_3CO_2H) leach for 4 hours to remove non-structural carbonates. Samples were then freeze dried. Approximately five milligrams of each pretreated sample were weighed and digested in a Picotrace class ten clean room at the Yale Metal Geochemistry Center (Department of Geology and Geophysics at Yale University) by using in-house distilled ultra-pure acids. Each sample was dissolved in 1 ml 6.2 M hydrochloric acid (HCl) in 5 ml acid-cleaned Teflon beakers and put on a hot plate at 100°C overnight to digest. Two drops of pure hydrogen peroxide (H_2O_2) were added to dissolve potential organic matter and samples were then evaporated dry. Dried samples were then dissolved in 1 ml 2 M nitric acid (HNO_3). A split of this sample solution was purified for $^{87}Sr/^{86}Sr$ analysis using an ESI PrepFast- MC- Sr system. The resulting elution of strontium was evaporated on a hot plate and then raised in 1 ml weak nitric acid (5% HNO_3 v/v) for analysis on a Thermo Neptune MC- ICP- MS using NIST SRM 987 as a bracketing standard (average $^{87}Sr/^{86}Sr$ ratio of 0.71034±0.00001, 2SD). The data were corrected for mass bias using $^{88}Sr/^{86}Sr = 8.375209$.

The radiogenic strontium isotope data ($^{87}Sr/^{86}Sr$) from all of the Balatonkeresztúr individuals are displayed in Figure 4 of the main text and SI Table 1. They were consistent with “local” values measured from plants, water and soil from the region to the south of Lake Balaton [47]. The local estimate in this case was calculated by taking the mean value of the published plant and water samples from Alt et al. 2014[47] plus or minus two standard deviations. There were no significant differences between males and females or subadults versus adults. The samples from the Transdanubian Encrusted Pottery culture (Pit: B-938) were slightly less radiogenic and more variable than the earlier time periods. This is particularly apparent for samples from three individuals (S15, S16, and S17) where the first and third molars were sampled. While both values could have come from the region south of Lake Balaton, they exhibit some spread, indicating movement within the region during early adolescence.

SI Table 1

Radiogenic isotope data for Balatonkeresztúr site samples

Laboratory ID	Laboratory	Burial/Specimen	Tooth Sampled	$^{87}Sr/^{86}Sr$
I/2584/2	ICER Laboratory, Atomki	Balatonkersztúr, S1	Second molar	0.709843
I/2584/3		Balatonkersztúr, S2	Second molar	0.709790
I/2584/4		Balatonkersztúr, S4	Second molar	0.709797
I/2584/5		Balatonkersztúr, S5	Second molar	0.709896
I/2584/6		Balatonkersztúr, S6	First molar	0.709778
I/2584/7		Balatonkersztúr, S7	First molar	0.709969
I/2584/8		Balatonkersztúr, S8	Second molar	0.709802

Laboratory ID	Laboratory	Burial/Specimen	Tooth Sampled	$^{87}Sr/^{86}Sr$
I/2584/9		Balatonkersztúr, S9	Third molar	0.709692
I/2584/10		Balatonkersztúr, S10	Second molar	0.709716
I/2584/11		Balatonkersztúr, S11	Second molar	0.709763
I/2584/12		Balatonkersztúr, S13	Second molar	0.709757
CAR0995	Quinnipiac University, Yale Metal Geochemistry Center	Balatonkersztúr, S-14 (Pit B-938)	First molar	0.709413
CAR0991		Balatonkersztúr, S-15 (Pit B-938)	First molar	0.70956
CAR1016		Balatonkersztúr, S-15 (Pit B-938)	Third molar	0.70872
CAR0992		Balatonkersztúr, S-16 (Pit B-938)	First molar	0.70948
CAR1015		Balatonkersztúr, S-16 (Pit B-938)	Third molar	0.70928
CAR0993		Balatonkersztúr, S-17 (Pit B-938)	First molar	0.70953
CAR1014		Balatonkersztúr, S-17 (Pit B-938)	Third molar	0.70895
CAR0994		Balatonkersztúr, S-18 (Pit B-938)	First molar	0.70959
CAR0990		Balatonkersztúr, S-19 (Pit B-938)	First molar	0.70961
CAR0996		Balatonkersztúr, S-20 (Pit B-938)	First molar	0.709432

2) Uniparental genetics and kinship

by Eszter Ari, Dániel Gerber, Bea Szeifert, Orsolya Székely

2.1) The mitochondrial DNA haplogroups and their phylogenetic evaluation

Studying maternal lineages in the era of whole genome studies still has its potential for kinship assessments and preliminary evaluation of population connections. Here we retrieved mitochondrial DNA (mtDNA) haplogroups for all 20 individuals (Supplementary Table S1), and inferred phylogenies for each subgroup by using all available mtDNA sequences from modern and archaic databases (ENA and NCBI), see below. We used the Bayes method for inferring phylogenies of subhaplogroups (implemented in the 'MrBayes' v3.2.6 software [48]). First, FASTA files were created from the BAM files by a custom pipeline implemented in the 'PAPline' using the 'Samtools' v1.6[49,50] and the 'ANGSD' v0.931 software[51] and a custom 'R' v3.4.4 script. The threshold for thecoverage were setted to 2x< and 50%< (majority rule) allele occurrence. Sequence alignments using the 'MAFFT' v7.271 software [52] were created using all publicly available recent (http://www.ianlogan.co.uk/) and archaic (AADR[53]) whole mtDNA sequences for each macrohaplogroup phylogeny, restricting the analysis to ~1500 samples per inferred subtree. Alignments were checked using the 'SeaView' software [54] by eye and ambiguous alignment regions as well low coverage (>30% of N sites) samples were discarded. We used the '4by4' nucleotide model, substitution type of 2, site variation model of invgamma [+I +G], 4 chain, 1 million generations, disabled parsimony model and 0,25 burn-in fraction for the phylogenetic inferences by the 'MrBayes' software The results are shown on Figs. S.2.1.1 - S.2.1.10. We also restricted visualisation of subtrees where posterior probability (indicated with node numbers on the Bayesian trees) was higher than 0.6, other phylogenies were considered unreliable or uninformative.

The individual BAD002 branches together with Bell Beaker culture (BBC) associated individuals from Spain in line with the Western European affinities of his genomic makeup (Figure S.2.1.1). Both in Bk-II and Bk-III maternal lineages various affinities appear, such as connections to e.g. Balkan (haplogroup J2b1, individual S13), Scandinavia (haplogroup T2b, individual S19), Central Europe (haplogroup K1a4a1g, individual S10) or British isles (haplogroup T1a4, individuals S5 and S6). Differences in geographical affinities clearly represent both changes and overlaps in genomic ancestry between Bk-II to Bk-III. One individual of the U4b1b1 subhaplogroup (ind. S15 from Bk-III) has direct maternal

connections to Iron Gates HG groups [55], which we tested using the 'Beast' v1.10.4 software [56] for divergence dates (Figure S.2.1.10). The maternal lineage of the individual S15 originated from the Mesolithic, thus its connection to HG groups is rather too old to firm inferences, while it shows a clear migration pattern from the Iron Gates to the Baltics. The mtDNA provides weak signal for population continuity at sequence level, only H10a1 subhaplogroup shows direct phylogenetic connection between Bk-II and Bk-III, but considering the mitogenome sequence diversity (calculated by the R package 'Popgenome' [57]) within periods (Bk-II *d*=0.978; Bk-III *d*=0.964), this is less surprising.

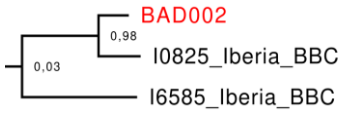


Fig. S.2.1.1

Haplogroup K1a4a1 subtree, calculated by the MrBayes software.

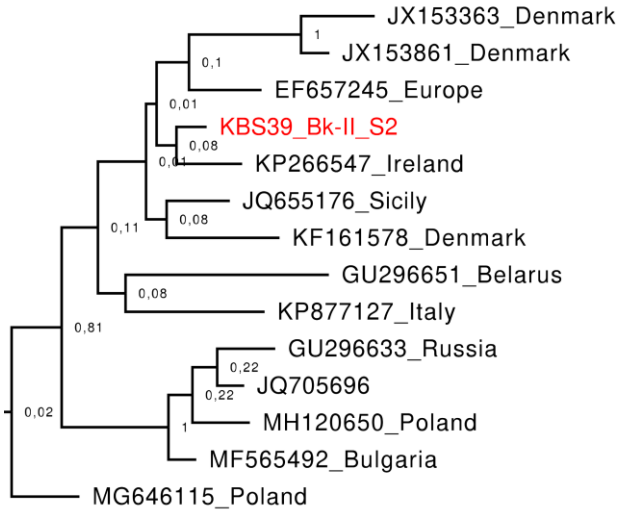


Fig. S.2.1.2

Haplogroup U5a2b1a subtree, calculated by the MrBayes software.

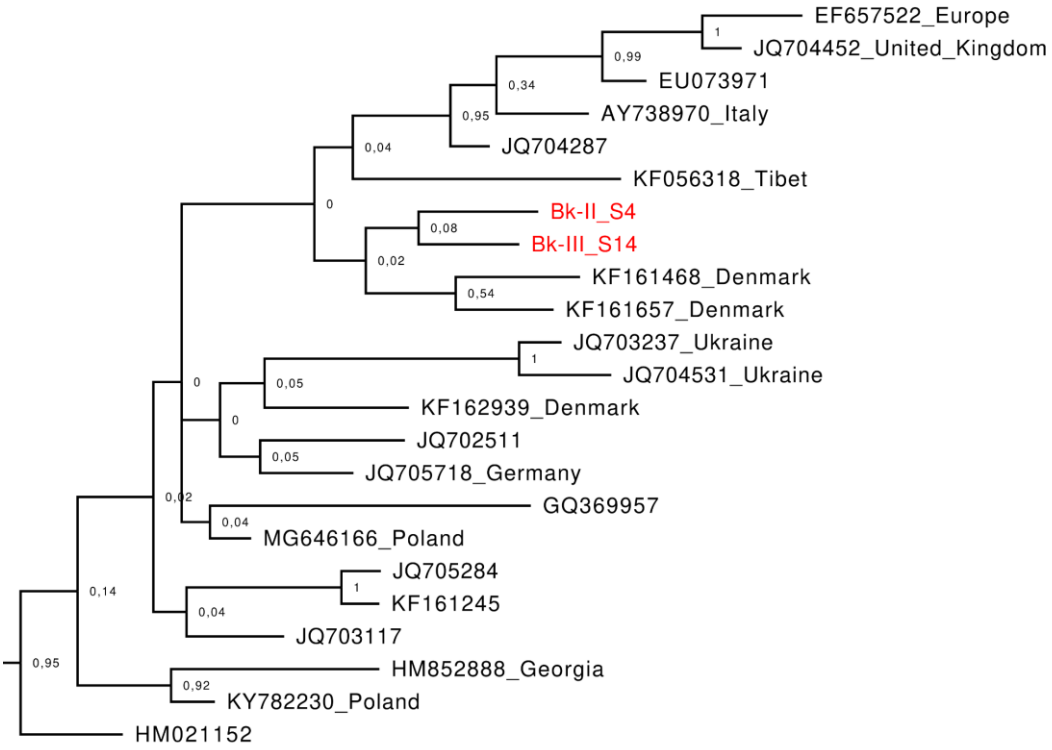


Fig. S.2.1.3

Haplogroup H10a1 subtree, calculated by the MrBayes software.

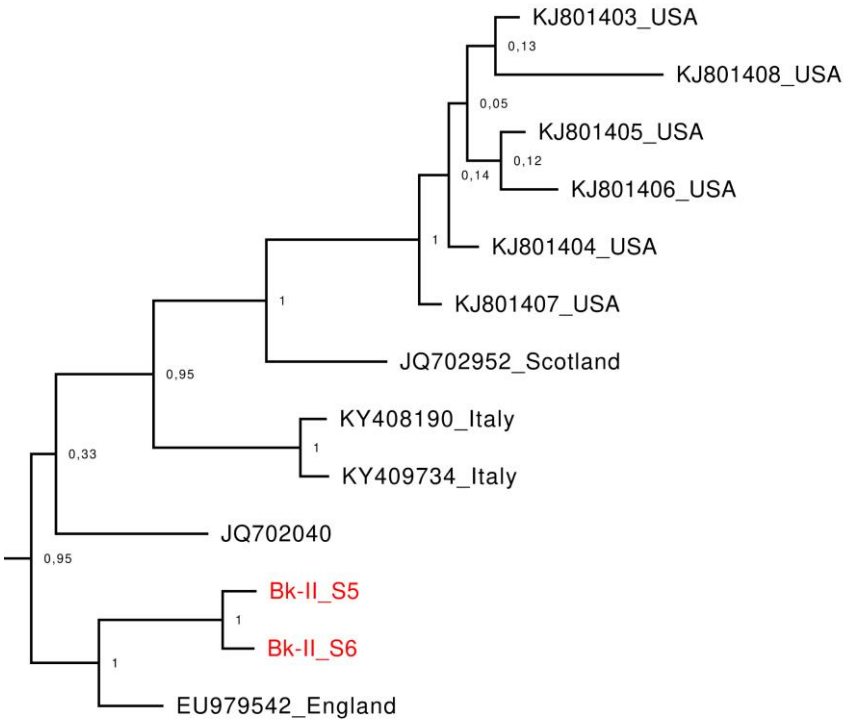


Fig. S.2.1.4

Haplogroup T1a4 subtree, calculated by the MrBayes software.

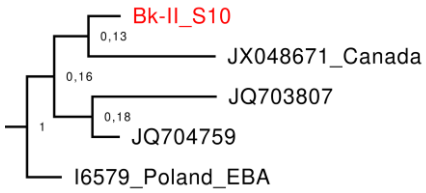


Fig. S.2.1.5

Haplogroup K1a4a1g subtree, calculated by the MrBayes software.

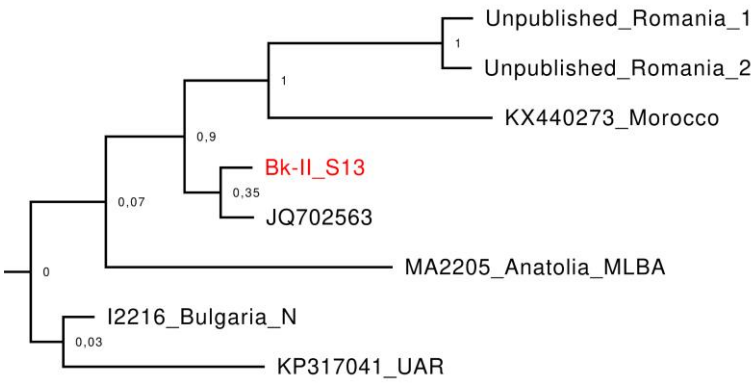


Fig. S.2.1.6

Haplogroup J2b1 subtree, calculated by the MrBayes software.

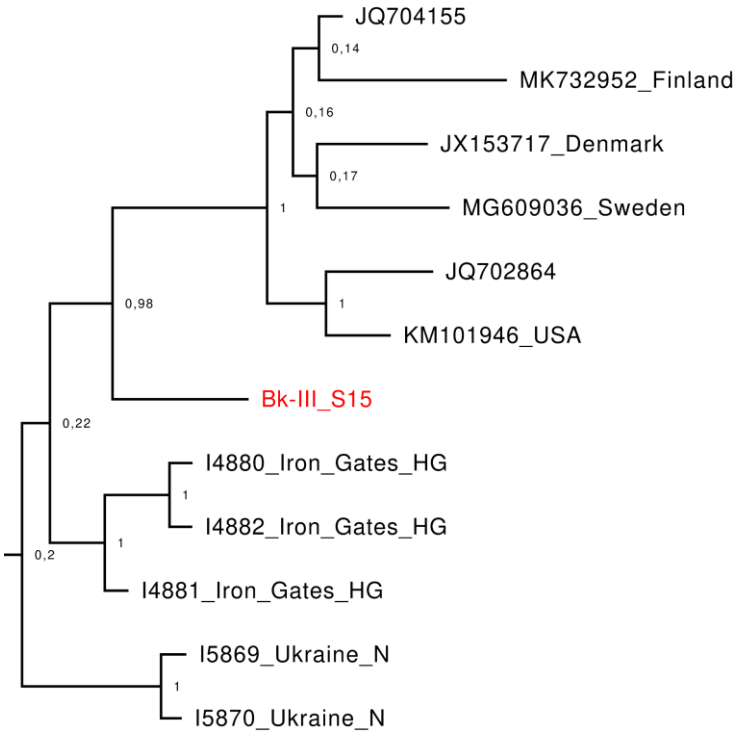


Fig. S.2.1.7

Haplogroup U4b1b1 subtree, calculated by the MrBayes software.

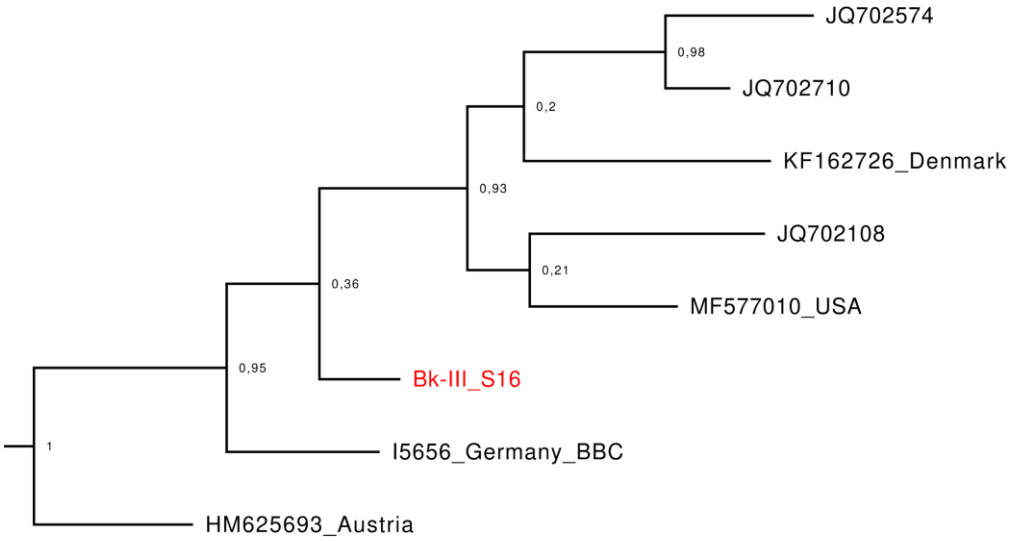


Fig. S.2.1.8

Haplogroup T2g2 subtree, calculated by the MrBayes software.

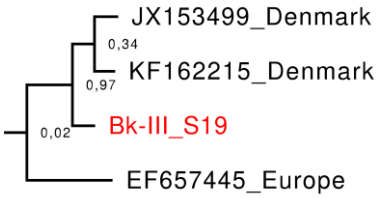


Fig. S.2.1.9

Haplogroup T2b subtree, calculated by the MrBayes software.

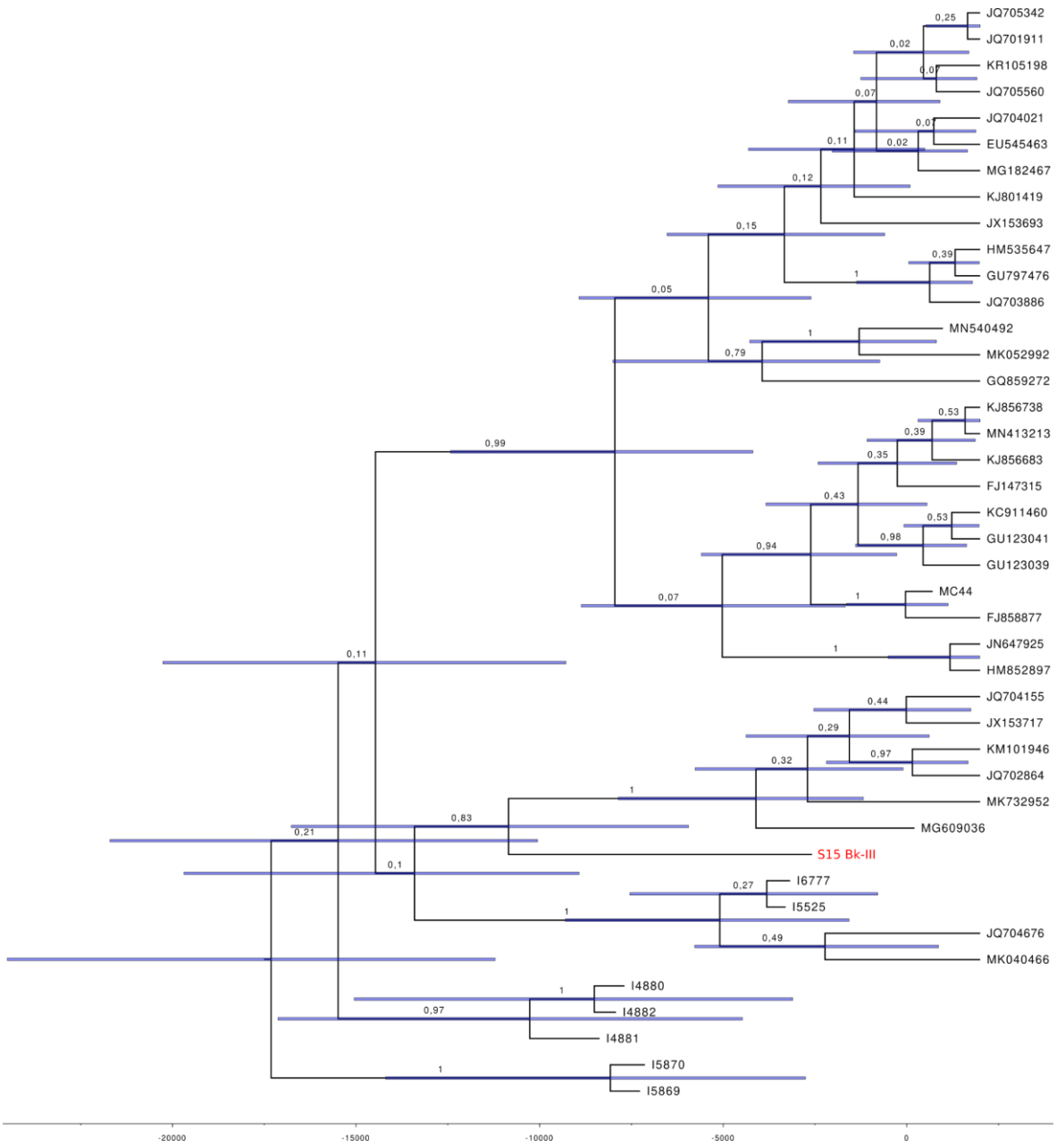


Fig. S.2.1.10

Haplogroup U4b1b1 subtree, calculated by the BEAST software, individual S15 diverged from the rest between ~6000-17000 BCE.

2.2) Y chromosomal haplogroups and STR network analyses

For the Y chromosomal haplogroup determination we used a combination of shotgun, capture and 17 short tandem repeat (STR) markers from AmpFISTR® Yfiler® PCR Amplification Kit (Applied Biosystems) analysed in 'NEVGEN' (https://nevgen.org) (Supplementary Tables S1 and S3). Fine detailed classification was made by SNP set from capture and shotgun sequencing by using the Yleaf v1 software [58] with a custom set of SNP list from ISOGG 14.255 [59]. We also made the Y STR network by using the 'Network' v10.1.0.0 and the 'Network publisher' v2.1.2.5 [60,61] software, and a custom R script that automatically pre-selected samples from available database data that are at either maximum 2 or 3 steps from our samples. We maximised 2 steps in cases where STR marker number is below 17,

as these may include too many otherwise distantly related individuals to the network. Individual S9 from Bk-I has the R1a-Z280>>V2670 Y chromosome subgroup, which according to STR network analysis of modern samples is identical with a number of individuals from Northeastern Europe. Although the poor diversity pattern of this Y subgroup makes the results conditional. Contrary to maternal lineages, the paternal makeup for Bk-II shows high homogeneity, only a few STR marker differences can be observed, although a haplotype diversity test similar to mtDNA can not be performed due to the scarcity of data (Table 1 in the main text, Figures S.2.2.1/2/3). High Y homogeneity within and between our groups shows Bk-II successiveness and population continuity in Bk-III, supported by qpAdm results as well (section 5.5.2.3) BK-III). It also suggests a strong patriarchal social network and – in combination with mtDNA data – likely a relatively smaller founder population. All lineages from BK-II and the most lineages from Bk-III can be attested to I2a-M223>>L1229>>(Z2054), which subgroup was already present in Mesolithic/Neolithic France [62]. I2a-L1229 is sparse in database but group I2a-M223 were present in Megalithic cultures from the British Isles to today's Czechia [63–65], reflecting STR network analysis really well, and is on a par with the observations of [62]. The qpAdm analyses (section 5.5.2.2) Bk-II) revealed that a major component of Bk-II is a Funnel Beaker culture (FBC) or related population, which groups frequently carried I2a-M223, pointing to a likely source for paternal lineages in Bk-II. Jagodnjak group (Encrusted Pottery) from Croatia carried G2a2-Z31430 that also can be linked to these groups [66], which strengthens this theory. A two-step neighbour lineage from Early Mediaeval Hungary signalises subsequent, probably limited, survival for these paternal lineages in Central Europe [67]. R1b-Z2103 is represented by two individuals in Bk-III, and it appears in contemporaneous populations such as in BBC period samples from Hungary [68] or a Vučedol culture associated individual from Croatia [55]. 16 marker STR network analysis shows more southeastern oriented connections for Bk-III R1b-Z2103 lineages, while the 17 STR marker network is more similar to I2a network distribution making results controversial (Figures S.2.2.1-3). Similarly to R1a, R1b is a relatively novel and widely distributed paternal lineage in the region which weakens its phylogeographic signal. Taking into consideration both previous appearances of this lineage in the region with the results of Network analysis, we can conclude that either it was present in Bk-II yet unsampled or came with Middle Bronze Age admixing groups, and at this level both scenarios are plausible.

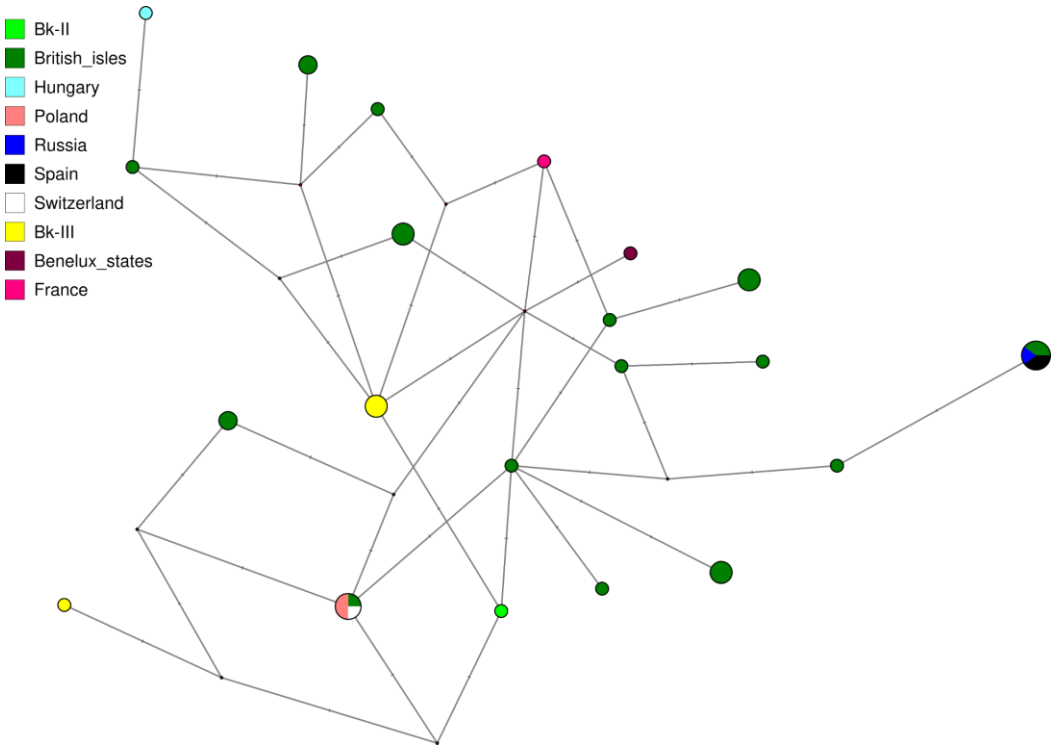


Fig. S.2.2.1

Median Joining Network with 12 STR markers where sample subselection was maximised to two step distance among I2a-L1229 haplotypes. Affinities to the British Isles are apparent.

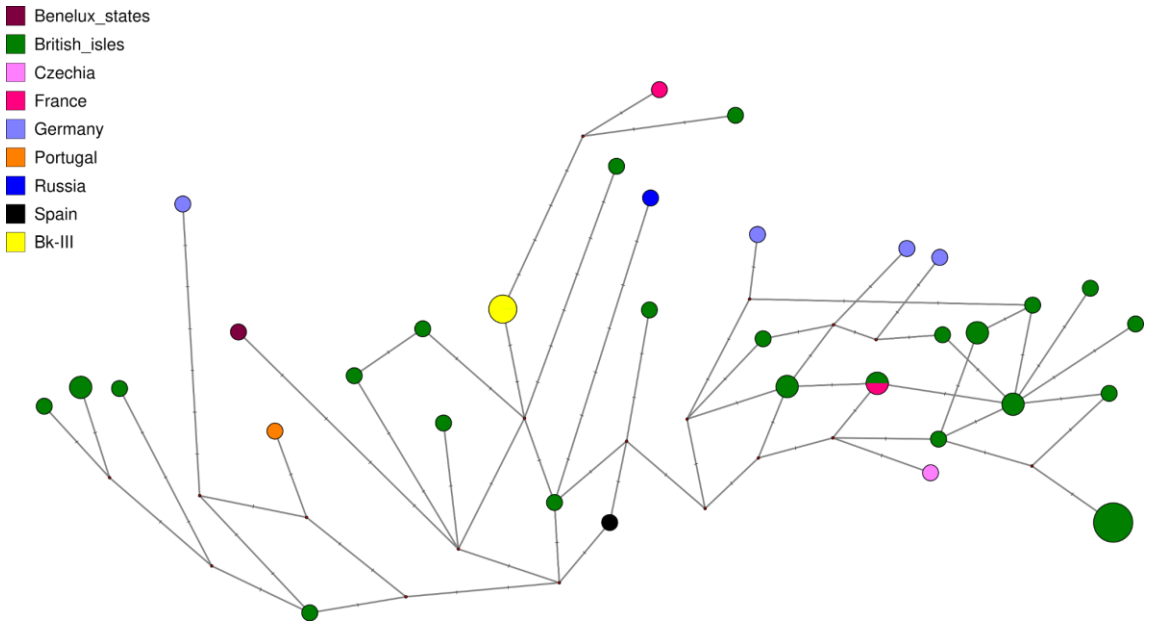


Fig. S.2.2.2

Median Joining Network with 17 STR markers where sample subselection was maximised to three step distance among I2a-L1229 haplotypes. Population affinities are highly similar seen on 12 STR marker network.

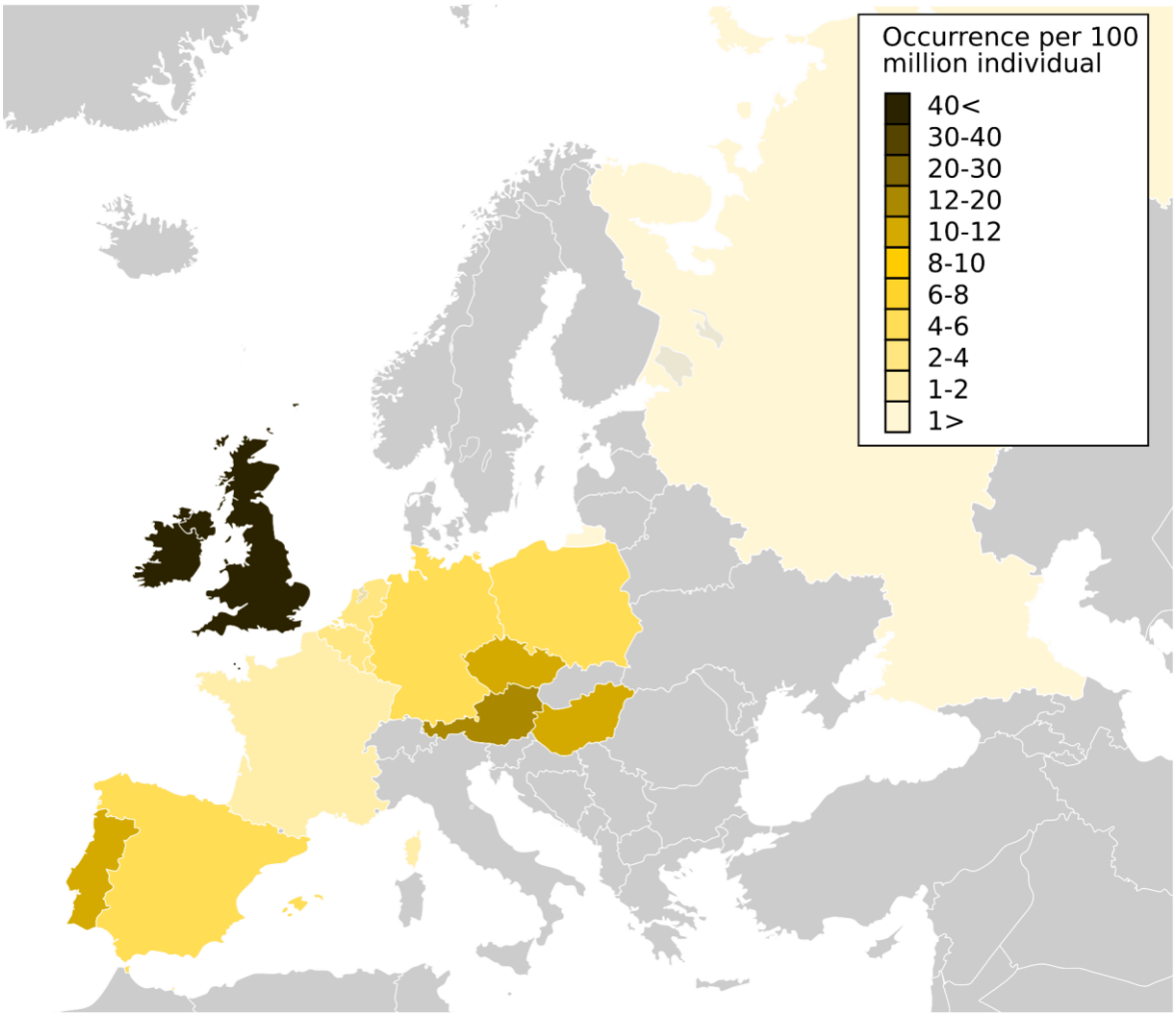


Fig. S.2.2.3

I2a-L1229 Median Joining Network results from Fig. S.2.2.1 and Fig. S.2.2.2 are combined, from which we calculated sample number occurrences per 100 million individuals per country. This figure shows these proportional occurrences.

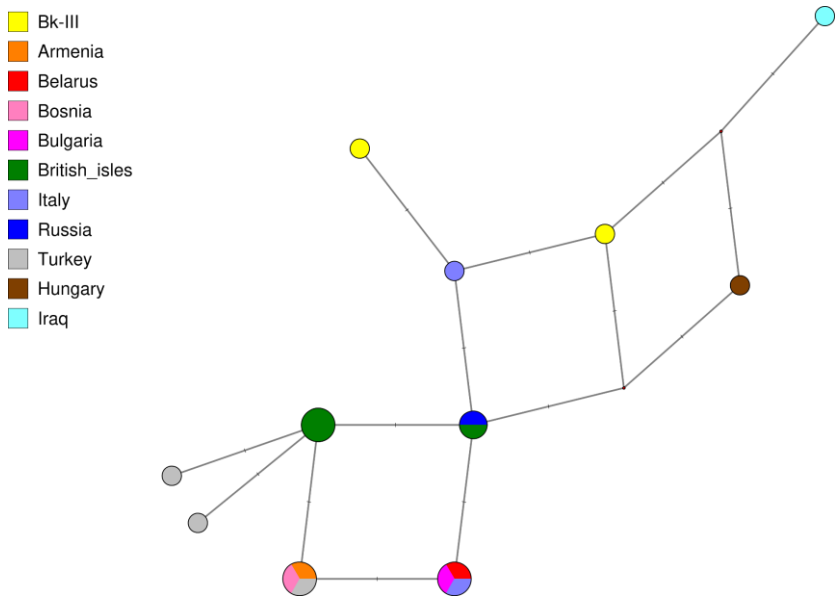


Fig. S.2.2.4

Median Joining Network with 16 STR markers where sample subselection was maximised to two step distance among R1b-Z2103 haplotypes. More prominent southern affinities appear in line with modern and archaic geographical distribution of the lineage.

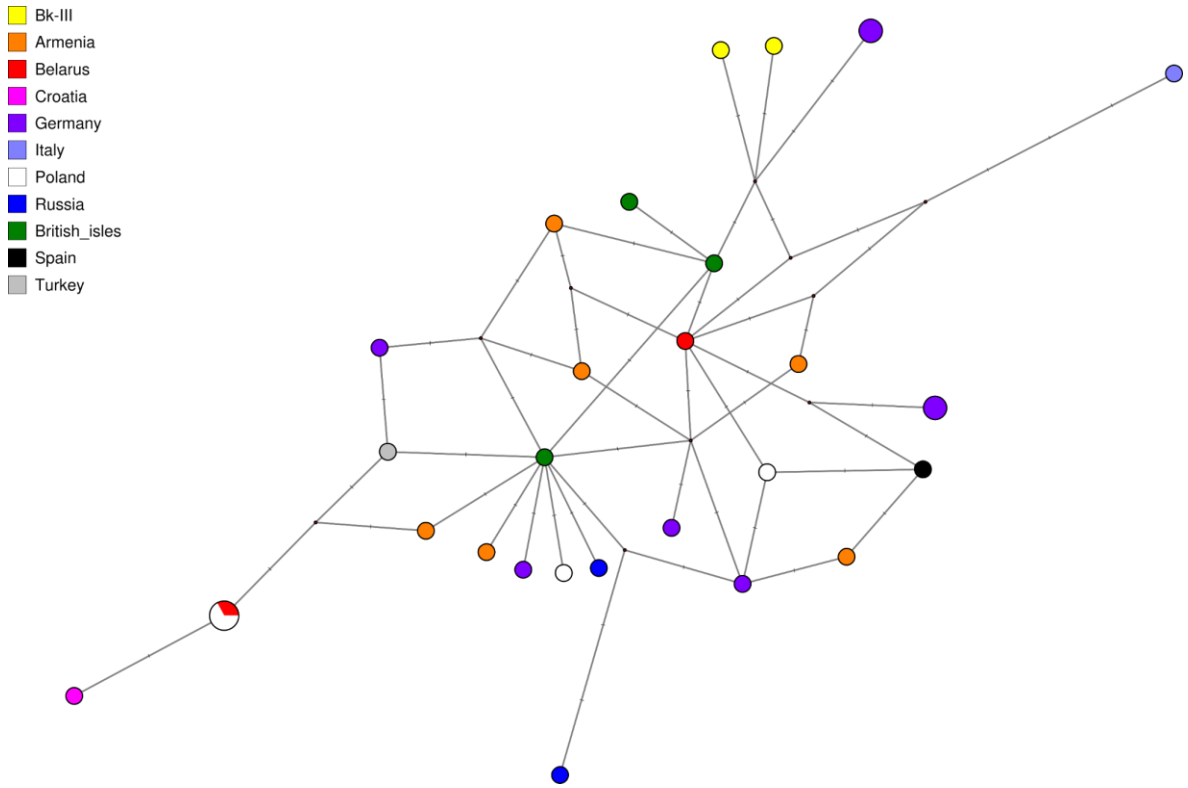


Fig. S.2.2.5

Median Joining Network with 17 STR markers where sample subselection was maximised to three step distance among R1b-Z2103 haplotypes. Network shows more northern affinities for Bk-III, contradicting the 16 STR marker network.

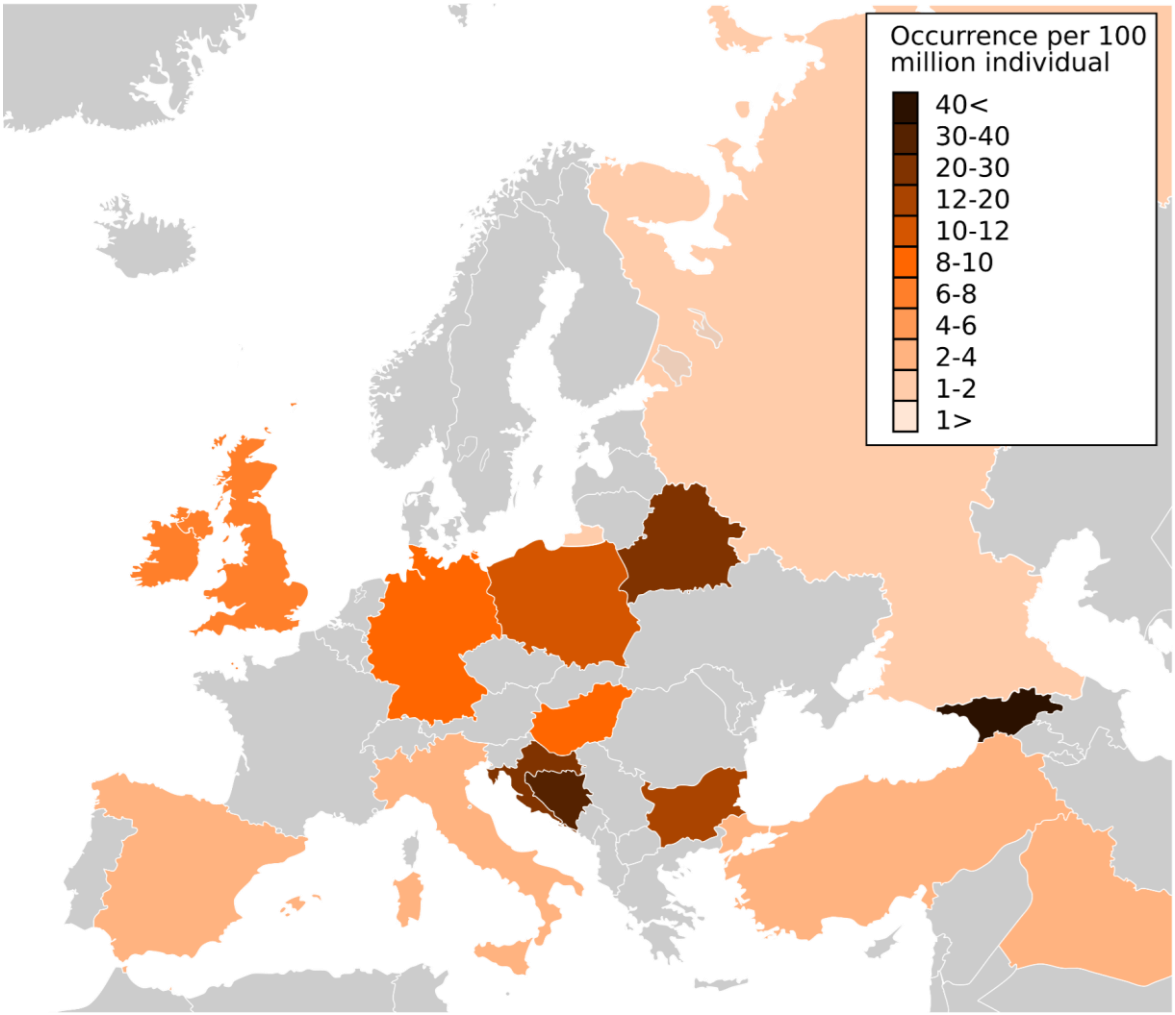


Fig. S.2.2.6

R1b-Z2103 Median Joining Network results from Fig. S.2.2.4 and Flg. S.2.2.5 are combined, from which we calculated sample number occurrences per 100 million individuals per country. This figure shows these proportional occurrences.

2.3) Kinship analysis

Kinship analyses are important to reveal social structures in prehistory that is a period before written records or human documentation, however, several recent archaeological studies emphasised that prehistoric social kinship relations are not necessarily determined by biogenetic links in all cases [69,70]. Pairwise genetic kinship assessments in this study were made by two independent methods due to low coverage genomes: identity by descent (IBD) estimated by the 'READ' software[71] and pairwise mismatch rate (PMR)[72]. The latter was calculated by the pairwise matches divided by the shared alleles, from where the smallest value (*min*) from the arbitrary population (here all samples from Balatonkeresztúr site) was extracted and the remaining value was multiplied by *1/(1-min)*, as

theoretically this would result in a ~40-50<% similarity rate for first degree, and ~30-40% similarity rate for second degree relationships (we named this method “modified pairwise match rate”, MPMR). For each, we call the 1240k dataset, as it consists of a comparable set of SNPs for human variation. We kept only those SNPs that have at least 100,000 base (100k) distance from each other, also setting the minor allele frequency (MAF) to 0.05 within Balatonkeresztúr site, which means that in our dataset at least one individual has to have a minor allele at each position. We also added RISE479 (discovered at Érd, from MBA period, Hungary) [72] to the pool, as this individual turned out to be highly similar to Bk-II and Bk-III, also giving sufficient background for IBD estimates. Supplementary Table S2 shows the results of the IBD and the MPMR [73] patterns. Where pairwise MPMR values above 0.4 indicate first degree, between 0.3-0.4 second degree relationship, however, in practice, values close to 0.3 may suggest third degree rather than second degree relationships, when we compare these with the results of the READ software. Since MPMR aimed to support or disprove READ results in cases when fewer than 1000 SNPs overlap between two samples, RISE479 was excluded from MPMR analyses since 1) it turned out to be unrelated to any of our samples according to READ 2) RISE479 is one of the best covered archaic individual in AADR database 3) thus it would highly distort MPMR results.

As expected based on the chronological gap, individual S9 from Bk-I does not show genetic relationship to any individual from Bk-II nor Bk-III up to the second degree (uncle-nephew, half-sibling, grandparent-grandchild). Within Bk-II a number of first (parent-offspring, siblings) and second degree relatives were found. The structure of the grave groupings also suggest apparent social links: ascendants, *i.e.* the earliest members of the group, also the genetically most distant individuals to others represent themselves in Grave Group B within Bk-II, while their children, nieces and nephews in Grave Group A. Bk-III shows extensive kinship relations, however, it shall not be considered as one blood related family burial due to the structure of kinship network. Additionally, possible further connections may remain hidden due to low genomic coverages. It is notable that even at this small population size we have found two pairs of half-siblings (ind. S1-S2 and S15-S17), which along with the low Y chromosomal diversity could either signalise polygamy or second marriage due to higher female mortality, although the latter seems more plausible in the light of previous findings [74]. Male dominance strongly influenced burial customs of the site both in Bk-II and Bk-III hinting at a special social structure compared to known Bronze Age groups, *e.g.* BA cemeteries from Germany [69] or from Serbia [75], where genetically confirmed male-female ratios are approximately even.

3) Phenotype assessment

by Dániel Gerber

3.1) Genetic sex and aneuploidy

Our new method that is implemented in the 'PAPline' pipeline to determe genetic sex and the presence of any form of aneuploidies at low coverage data. First we obtained coverage data for all chromosomes for (1) the individuals in our dataset, (2) a selected ancient shotgun genome database [72], and (3) a yet unpublished mediaeval and modern day dataset studied in our group by using the 'bedtools' v2.27.1 software [76] Then we analysed the results in a custom R script. We used *Z*-score adjusted coverages – named as *ZAC* – for each chromosome by the following calculation: $ZAC_i = \frac{chr_{ij} - min_i}{max_i - min_i}$, where *chr* is the coverage of the *i*-th chromosome of the *j*-th sample, *min* is the minimum coverage of the *i*-th chromosome across all samples, and *max* is the maximum coverage of the *i*-th chromosome across all samples. By providing the ratio of the Y and X chromosomes of these *AC*s, the presence of the Y chromosome can be confidently detected even at really low coverages (Figure S.3.1.1). ZAC values are also eligible to test aneuploidy, as coverages correlate with ploidy. This method shrinks coverages to the same scale (except Y chromosome needs to be further divided by 2 as normally it exists in one or zero copy), from which we can easily infer ploidy even as low as 0.01x genomic coverage (Figure S.3.1.2).

Accordingly, individual S10 possesses an extra copy of Y chromosome, an aneuploidy known as Jacob's syndrome (Figure S.3.1.3). Notably, data type (shotgun *vs.* capture), mapping method and post filter processes may significantly alter output if one compares different datasets to each other, thus we recommend to analyse similarly processed data with this method.

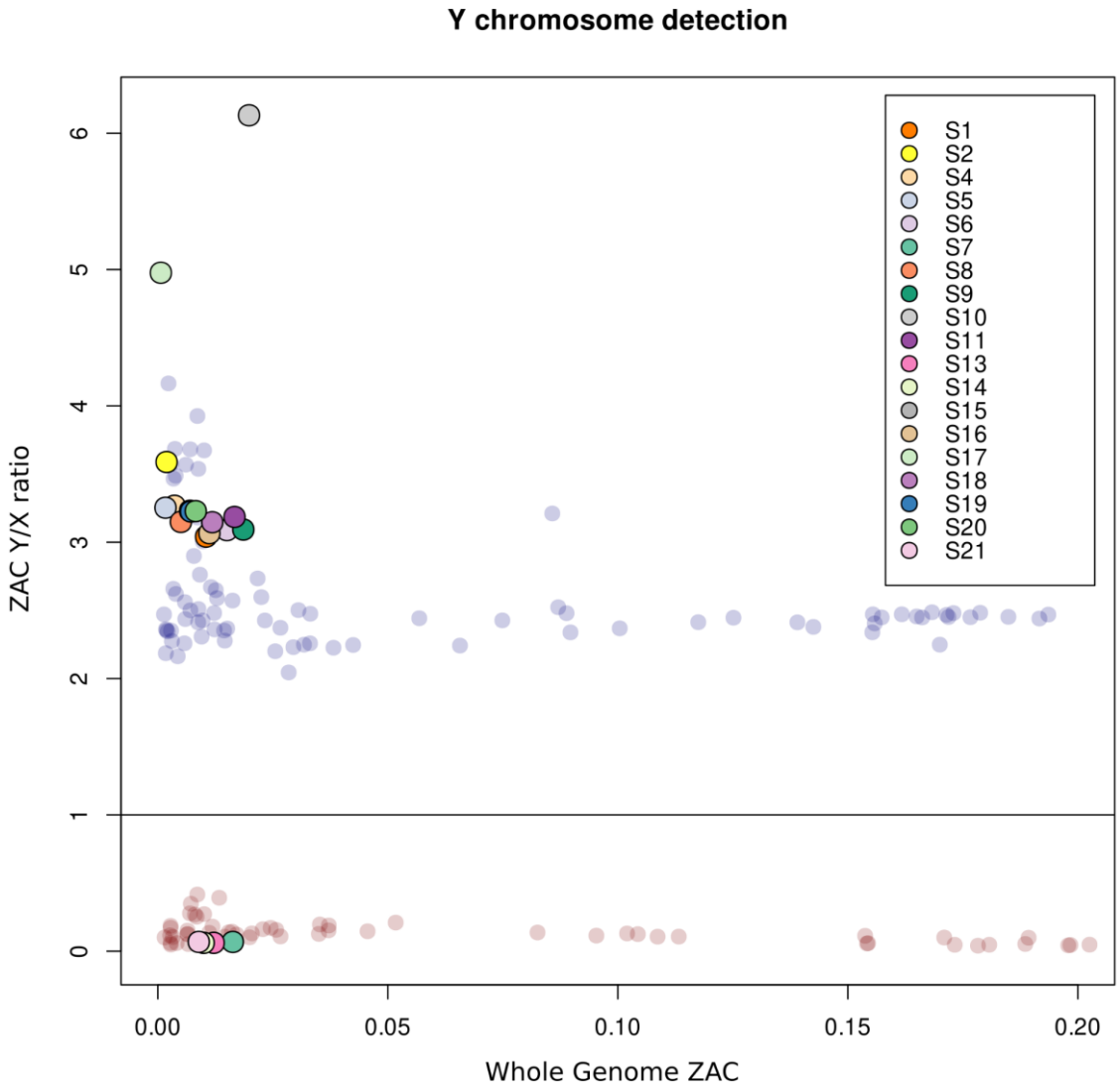


Fig. S.3.1.1

Visualisation of ZAC X and Y chromosome ratios showing the robustness of Y chromosome detection even when analysing genomes with extremely low coverage. Blue dots represent detectable Y chromosomes (males), red dots non-detectable Y chromosomes (females). The extremely high value of individual S10 is due to Jacob's syndrome, discussed further in Fig. S.3.1.3 and the main text.

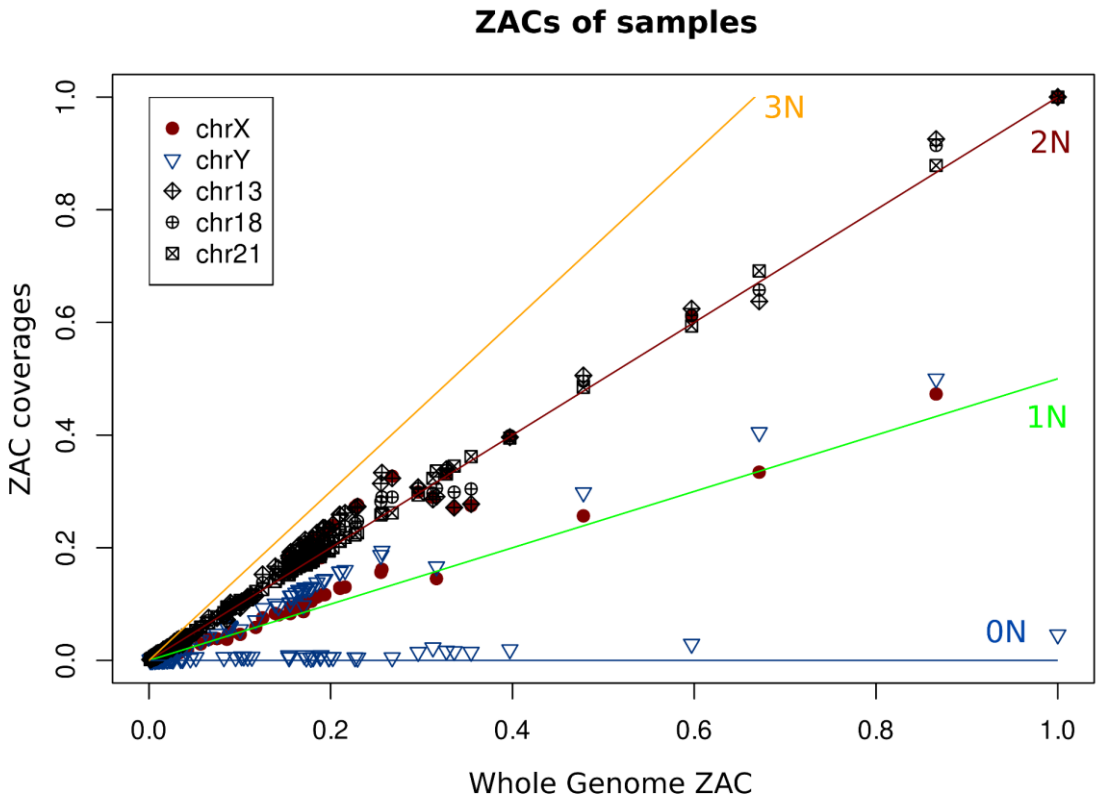


Fig. S.3.1.2

ZACs of our and database data. Lines represent ploidy (blue: zero copy, green: 1x copy, red: 2x copy and yellow: 3x copy), symbols listed in legend represent ZAC value of a certain chromosome of a certain individual at a certain genomic coverage. We selected chromosomes 13, 18, 21, X and Y to image as these tend to vary in copy number for natural and/or pathological reasons. Clear distribution between 0-2 copies of a certain chromosome can be observed, however, due to different sequencing approaches (capture vs. shotgun) and read mapping methods, the standard deviation from statistical mean is high across samples, thus the actual application of this method is recommended for similarly pre-processed datasets.

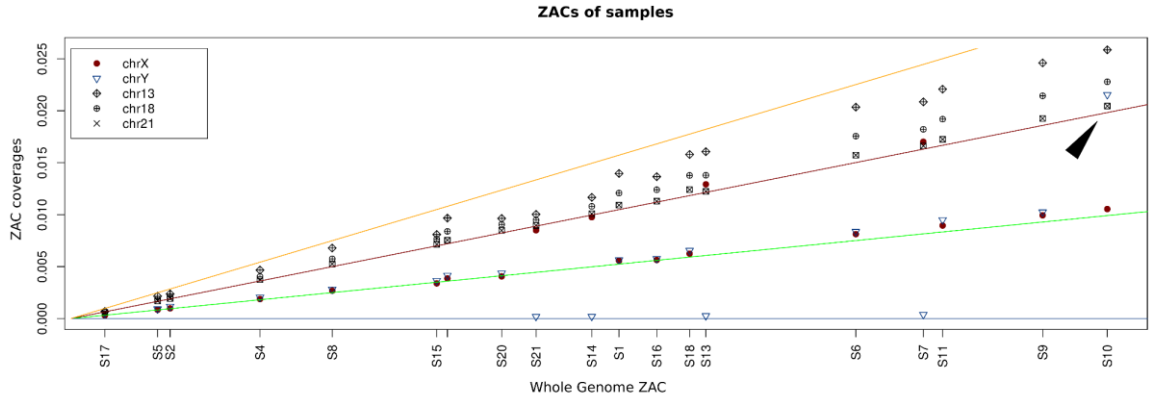


Fig. S.3.1.3

ZACs of our data subselected from Fig. S.3.1.2 where coverages were renamed to the correspondent sample ID. Besides clear differences between biological male and female individuals, S10 shows 2n copy number for Y chromosome (highlighted with a black arrow) indicative for Jacob's syndrome.

3.2) Variant discovery

3.2.1) SNPs of pigmentation

For pigmentation assessment, we used a custom set of phenotypically relevant SNPs listed in the HIrisPlex-S system [77], which we extended with variants from the SNPedia database [78], for results, see Supplementary Table S5. SNPs were called with the following filters: trimmed read ends (2 bp), BaseQuality>30, and MappingQuality>30. According to the available data, BAD002, despite providing weak signals for pigmentation, likely blends to the average Neolithic European variation. S9 from Bk-I had European light, almost pale skin that probably had no freckles and was likely a bit less sensitive to sunburn. His hair was dark blonde and his eyes were likely blue, which traits are in accordance with previous studies of people with steppe origin [79]. Bk-II shows almost uniform makeup, even with sparse data only a couple of variants show heterozygosity within the population. Accordingly, the overall pigmentation was dark, skin colour was probably darker than today's average European, however, light and blue eye colouration occurred, thus most individuals probably had lighter hazel or green eyes, in addition of other SNPs associated with light pigmentation, freckles and blonde hair, which may have played a role in their appearance. The individuals in the Bk-II population were probably more sensitive to sunburn and had low tan response. Surprisingly, Bk-III shows a high variability of pigmentation patterns. The colouration of the preceding population appears, but more and well pronounced lighter pigmentation patterns can be observed. Two individuals (S14 and S17) from the mass grave probably had reddish blonde hair, green eyes and pale skin, maybe even freckles, likely as a result of admixture to populations of various origin.

3.2.2) SNPs of clinical significance

We created a custom set of variants obtained from the ClinVar database [80], which we tested on our data. 27 SNPs show malignant substitutions in some of our samples, which finally restricted to only four trustable positions, in which cases the genotype likelihood (GL) values (calculated using the 'ANGSD' v0.931 software[51]) were above 0.99. Notably, we considered the majority of variant detection as *false* positives likely caused by DNA damage since most of them were transitions, and these transitions are extremely rare in today's population. On the other hand, a couple of variants with low GL values have high prevalence both in today's and in the investigated ancient populations or involved in skeletal

development. In this section we discuss diseases where trustable or otherwise interesting SNPs were recovered. The final estimates were summarised in Supplementary Tables S1 and S6.

3.2.2.1) Lig4 syndrome

The individual S15 carried a variant (rs104894421, GL>0.99) in heterozygous form which potentially causes Lig4 syndrome when carried in homozygous form. Therefore, in theory the genotype of individual S15 did not result in the actual onset of symptoms. On the other hand, in the Jagodnjak group the very same variant appears in individual JAG93 (1x coverage), while it is also present in individual JAG82 but on another locus (rs104894420) of the same gene. While it is merely speculative since individual S15 is not covered in the latter locus, there is a possibility that these two SNPs were both present in the Encrusted pottery culture associated population, and individual S15 inherited two different locus both causing frameshift mutation resulting in an actual onset of symptoms, supported by some of his skeletal features, such as elongated face and hip dysplasia [81].

3.2.2.2) Diabetes susceptibility

Individuals BAD002, S1, S9, S10 and S11 carried variants (rs11196205, rs12255372, rs237025) for diabetes susceptibility, which despite having low GL-s (~0.67) in each individual, we can assume that at least a few of these SNPs were likely present, as these are also highly prevalent in today's European populations (~30-50%) [82].

3.2.2.3) Hereditary spastic paraplegia

rs121434442 is an autosomal dominant SNP in the KIF5A gene that causes hereditary spastic paraplegia, a neurological disorder causing progressive muscle stiffness in the limbs, taking onset between 2 to 40 years of age. Complex form causes a high variety and number of symptoms, while the condition itself ranges between low-key to serious disability. Individual S6 carried the malignant variant at this locus (GL=0.67), which could have resulted in the onset of the disease. Interestingly, S11, the father of individual S6, has a lower limb condition that could be associated with years-long onset of symptoms [83].

3.2.2.4) VCP gene malignant substitutions

Diseases associated with the mutations of VCP gene are autosomal recessive and their homozygous form cause a number of symptoms, such as amyotrophic lateral sclerosis, Paget disease and inclusion body myopathy [84]. Individual S21 (infant) carried heterozygous variants in both rs387906789 and rs121909330 locuses (GL>0.99 in both cases), which most definitely did not cause any symptoms, considering her age as well. Individual S16 could also be a carrier at a different locus (rs121909335, GL=0.67), but his skeleton does not show any kind of malformations.

3.2.2.5) Obesity

Individual S10 have a substitution at locus rs4994 related to obesity. While as a transition it is highly unreliable (GL=0.67), ~7% of today's Europeans possess this allele [82], thus we concluded that the actual presence of the variant as well the onset of the symptoms can not be excluded.

3.2.2.6) 3-Methylglutaconic aciduria type 1

Individual S10 have a substitution (GL=0.67) at locus rs200030276 related to 3-Methylglutaconic aciduria type 1. This rare disease is an inborn error in leucine metabolism causing mainly neurological and physiological issues where the symptoms are taking onset at early childhood [85]. The actual presence of the variant is highly questionable, however, it can not be excluded and may be related to the early death of this individual.

3.2.2.7) Dandy-Walker syndrome

The Dandy-Walker syndrome is marked by cerebral malformation, and rarely accompanied by macrocephaly [86]. Individual S20 carries a SNP associated with this disease (rs576405108, GL=0.67), however, many even non-genetic factors contribute to the onset of symptoms. The SNP is a C>T transition of 1x coverage, but being an infant, S20 had no time to develop symptoms, thus the possibility of him being a carrier can not be excluded.

4) Facial reconstruction

by Ágnes Kustár

The method of facial reconstruction helps us to take insight into the appearance of past individuals, which is also frequently used today by forensic identification. The skull of individual S13 is well preserved, except the right zygomatic and temporal bones were slightly damaged. According to previous evaluation by Kitti Köhler, the skeletal remains suggest a gracile and short (~154 cm) physique without any traces of perimortem injuries. Secondary analysis by Ágnes Kustár confirmed these results by an addition of no traces for - at least intensive - childbirth, and two small and shallow perpendicular cuts on one of the ribs that could be a perimortem damage, although neither the surrounding area nor the placing of the cuts suggest a life-threatening injury.

The physical appearance of the skull reflects the past facial features. Absolute size and form is small and gracile, overall highly feminine in characteristics. The shape of the skull is high and short, the forehead is narrow and bulging. The nape is curved, muscular joints (*linea nuchae superior et suprema*) are prominent, the external occipital protuberance is well developed, and while the mastoid is small the features indicate strong neck muscles. The nasal cavity is moderately wide (*mesorrhine*), its lower margin is sharp (*anthropine*), both indicate moderately wide nasal wings. The nasal root is shallow and the nasal bridge is narrow. Distal parts of the nasal bones are broken and were reconstructed with resin. The anterior nasal spine is moderately developed that slightly turns upward. Each feature combined with moderately protruding nasal bones from the facial plane suggest a moderately prominent cartilaginous nasal structure (*nasus externus*). The orbit is high (*hypsikonch*), its shape is rounded, and the upper margin is slightly reclined. The distance between orbits is relatively small. The zygomatic bones are smooth, short and slim, the canine fossa is shallow. Due to the prominent protrusion of both jawbones the upper and lower incisors are highly protruding. Accordingly, the lips may have been moderately full and also protruding. The mandible is small and short, and is moderately thick in its body. The *ramus mandibularis* is short, the bone joint is small, but the jaw corner is almost rectangular with moderately developed muscular joints. The chin bone is sharp and slightly protruding.

For facial reconstruction a detailed copy of the skull was made by rapid prototyping technique. First, the skull was scanned by Computer Tomography at Huniko Egészségügyi Szolgáltató Kft. (Hungary), then a resin cast was made by selective laser sintering at Varinex Zrt. (Hungary). Finally, a plaster cast was made after filling the gaps in the skull copy with beeswax (Figure S.4.1). During the process the plaster cast was used as a base where soft tissues were built back for achieving facial features. The reconstruction was performed by well defined artistic and anatomical methodology [87–89]. Plasticine clay was used to form muscles on the skull [90,91]. The thickness of the muscles were estimated by the roughness of the bone surfaces [92] in 45 measuring points (Figure S.4.2). Markers of tissue thickness were placed on the plaster skull to measure points, lengths of the marker spikes were determined according to average values listed in SI Table 2. Long needles were used to mark regions and borders on the skull which became covered during the process to keep monitoring important morphological points (Figure S.4.3). Eyeballs made of resin were measured to be 25 mm in diameter fitting eye sockets. The cartilaginous nasal septum (*septum nasi cartilagineum*) was made of hardened beeswax to preserve shape, while the nasal spine and nasal wings were made of plasticine. The shape and

dimensions of the nasal structure was estimated based on the shape of the nasal bones, the proportions of the nasal cavity and the direction of the anterior nasal spine [93]. Shape, size and length of mimicry muscles were estimated according to the imprints of muscular joints on bone surface, anatomical regularities and individual characteristics. In the final artistic phase harmonisation of the features was performed (Figures S.4.4).

SI Table 2

Grades according to bone relief: 1: Very gracile, smooth, 2: Less gracile, a little rough, 3: Rough, 4: Robust, very rough

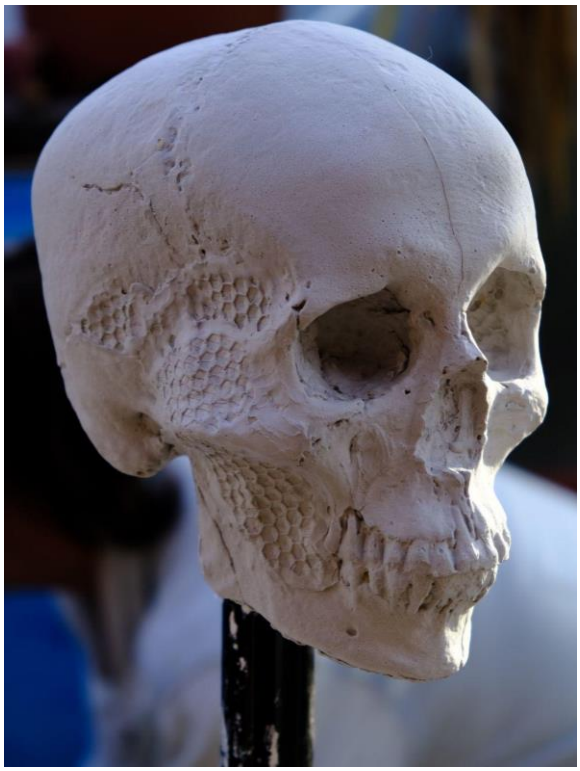
Soft tissue thickness data on the face of individual S13		
Measuring point	**Degree**	**Thickness (mm)**
Bregma (b)	1	4
Metopion (m)	1	4
Glabella (g)	1	5
Nasion (n)	1	4
Rhinion (rhi)	1	2
Philtrum (ph)	1	7
Labiomentale (lab)	1	7
Pogonion (pog)	1	8
Gnathion (gn)	1	7
Arcus sup.medialis (acm)	1	7
Arcus sup.lateralis (acl)	1	4
Ectoconchion (ek)	1	3
Orbitale (or)	1	3
Dacryon (da)	1	2
Lacrimale (la)	1	2
Lat.apertura pir. (lat.ap)	1	2
Alare (al)	1	3
Subspinale lat. (ss lat)	1	9
Caput mandibulae (cap)	1	3
Gonion (go)	2	4
Zygion (zyg)	1	2
Facies zygomaticus (fac.zyg)	1	4
Zygomaxillare (zm)	1	3
Proc.mastoideus (mast)	1	3
Lambda (l)	2	5
Opisthocranion (op)	2	5
Subnasale (sn) (H11)*	NA	13
Labrale superius (ls)(H12)*	NA	11
Labrale inferius (li)(H13)*	NA	12
Mid mandibular border (mmb)(H28)*	NA	11.5
Euryon (eu)(H29)*	NA	5.5

**H11-H29 measurements according to Helmer [92] (30-39 years old women)*

The reconstructed face reflects the shape of the skull, the head is wide and short, the forehead is narrow and convex, the *glabella* and eyebrows are less protruding and is slightly arched. The face itself is moderately wide and it narrows towards the chin, while her neck is relatively robust. The nasal root is moderately deep, the nasal bridge is also moderately protruding and narrow. The tip of the nose is sharp and is in a slight upward position. In the frontal view the nasal root is narrow, the nasal bridge and nasal wings are moderately wide. According to the small protuberances in the margins of the orbit as indicators for the joint points of the eyelid hanging fibres the eye openings are completely horizontal. The eyes are slightly sitting close to each other, the eyelid fold is moderate. The mouth is moderately wide and thick, the incisors are protruding and the protruding lips are not closing entirely. The jawline is not prominent, the chin is slightly protruding, the ears cannot be confidently reconstructed, thus they

were adjusted in harmony to other facial features, especially to the nose and eyebrows. Pictures were taken of the clay model in frontal (Figure S.4.5) and profile (Figure S.4.6) view.

Nutrition information can not be obtained from osteological features, thus we suggested average condition in facial reconstruction. Individual S13 died between 35-45 years of age, hence we created mimicry wrinkles specific to 40-60 years olds. Colouration was created according to genetic results (Supplementary Table S5, section '3.2.1) SNPs of pigmentation') by Zsuzsa Herceg (Figure S.4.7). The overall pigmentation was transitional with lighter but creole skin tone and hazel eyes with blondish-brown hair accompanied with the presence of freckles. Hairstyle and hair structure was created according to genetic and artistic evidence, as Encrusted pottery figurines show straight, braided hair (Figure S.4.6 [12]). Additionally, digital grave reconstruction was made by Fanni Gerber according to burial position, suspected clothing and jewelry (Fig. S.4.7). Finally, an attempt for digital sculpting of the finished reconstruction was made by the newly developed Unreal Engine's Metahuman Creator [94] (Figure S.4.8), however, despite the life-like results, some important facial characteristics, such as maxillary prognathy, could not be implemented to this software, making this reconstruction unreliable. On the other hand, by creating this digital version we showed the endless future opportunities for such reconstructions.



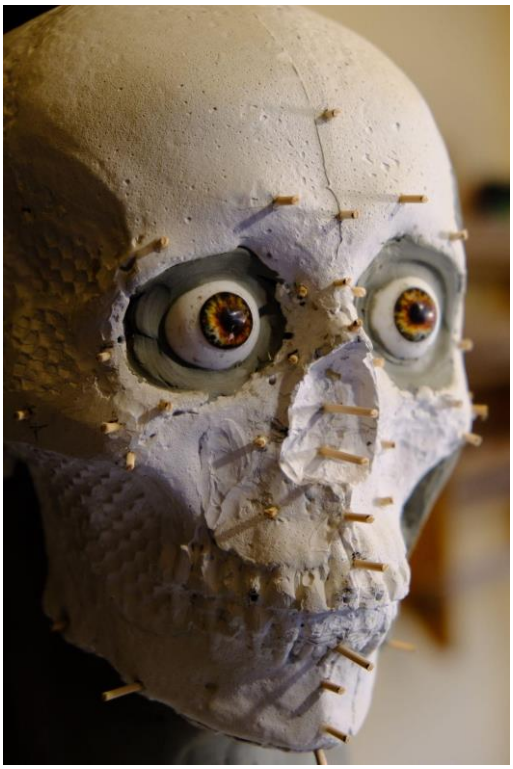


Fig. S.4.1

Plaster cast of the skull by Ágnes Kustár. Photo by Dániel Gerber.

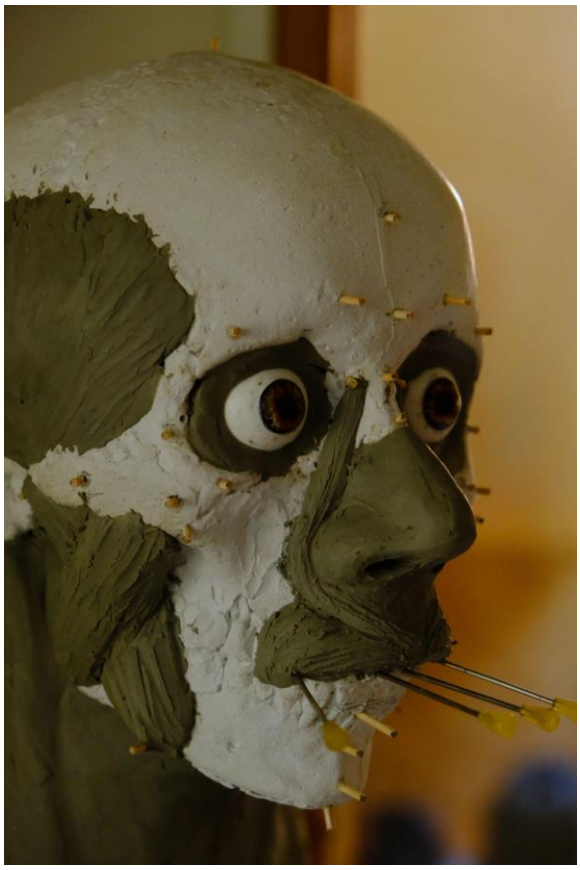


Fig. S.4.2

Pins placed to measuring points and trimmed according to calculated tissue thickness (SI Table 2) by Ágnes Kustár. Photo by Dániel Gerber.

Fig. S.4.3

During the tissue rebuilding process long needles were used to keep monitoring the important measurement points by Ágnes Kustár. Photo by Dániel Gerber.

Fig. S.4.4

Half of the skull is finished, the other half was kept halfway ready to monitor accuracy by Ágnes Kustár. Photo by Dániel Gerber.

Fig. S.4.5

Frontal view of the final clay sculpture by Ágnes Kustár. Photo by Dániel Gerber.

Fig. S.4.6

Profile view of the final clay sculpture by Ágnes Kustár. Photo by Dániel Gerber.

Fig. S.4.7

Finished and coloured reconstruction of individual S13. Bronze bead was pinned to her hair as similar jewellery was found next to her skull by Zsuzsanna Herceg. Photo by Dániel Gerber.

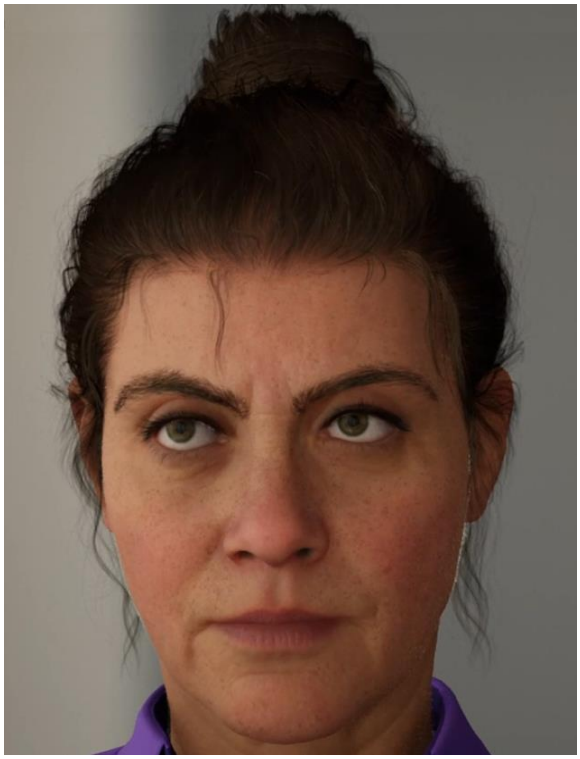


Fig. S.4.8

Digital reconstruction made by Dániel Gerber with Unreal Engine's Metahuman Creator. While it is highly realistic, due to the lack of distinguishing features, such as maxillary prognathy, and raw shaping of facial features it remains unreliable.

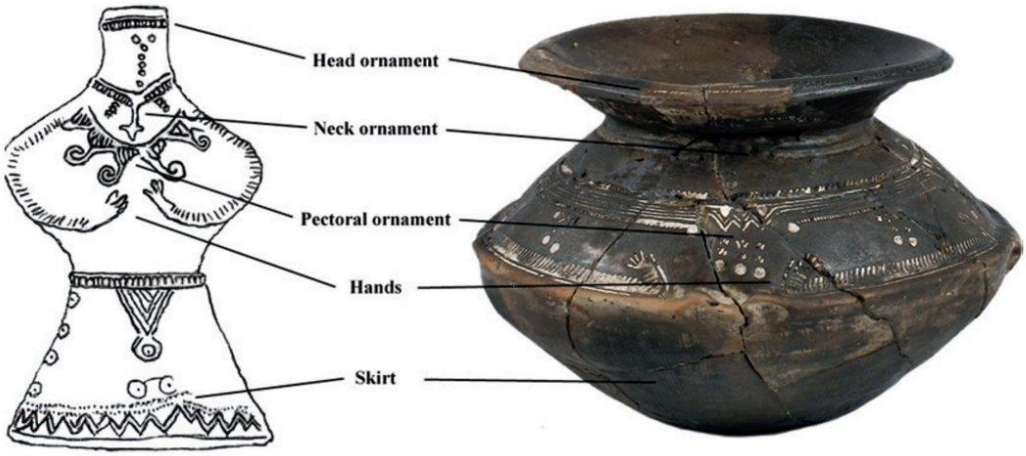


Fig. S.4.9

Reconstruction of wear based on pottery ornaments [12].

Fig. S.4.10

Digital grave reconstruction of individual S13 by Fanni Gerber according to available evidence of wear and artefacts in her burial.

5) Whole genome analyses

by Dániel Gerber, Balázs Gyuris, Anna Szécsényi-Nagy

5.1) PCA

We used the 'smartpca' software [95] and calculated principal components (PCs) on a set of present-day European, Near Eastern and Caucasian populations (here using their database[53] name: Abkhasian, Adygei, Albanian, Armenian, Assyrian, Balkar, Basque, BedouinA, BedouinB, Belarusian, Bulgarian, Canary_Islander, Chechen, Chuvash, Croatian, Cypriot, Czech, Druze, English, Estonian, Finnish, French, Georgian, German, Greek, Hungarian, Icelandic, Iranian, Irish, Irish_Ulster, Italian_North, Italian_South, Jew_Ashkenazi, Jew_Iranian, Jew_Iraqi, Jew_Libyan, Jew_Moroccan, Jew_Tunisian, Jew_Turkish, Jew_Yemenite, Jordanian, Kumyk, Lebanese, Lebanese_Christian, Lebanese_Muslim, Lezgin, Lithuanian, Maltese, Mordovian, North_Ossetian, Norwegian, Orcadian, Palestinian, Polish, Romanian, Russian, Sardinian, Saudi, Scottish, Shetlandic, Sicilian, Sorb, Spanish, Spanish_North, Syrian, Turkish, Ukrainian), using 590k SNPs in the calculation with an option of shrinkmode=YES, lsq=YES, and the PCAplot.r script from the 'PAPline' for plotting.

5.2) Admixture

We made supervised admixture analysis (Figure S.5.2.1) by using the 'ADMIXTURE' software [96] with a fixed set of individuals as sources listed in the Supplementary Table S8 and results seen in Supplementary Table S9. We selected these sources based on uniparental makeup and known sources of European genepool of HG[55,62,97–103], early farmer[55,62,66,99,104] and steppe[55,99,105] ancestries. We further divided HG source to Western (WHG) and Eastern (EHG) components for their high genetic dissimilarity

based on Rivollat et al. 2020[62], but we excluded the GoyetQ2 type component because of their low coverage representative genomes in the database.

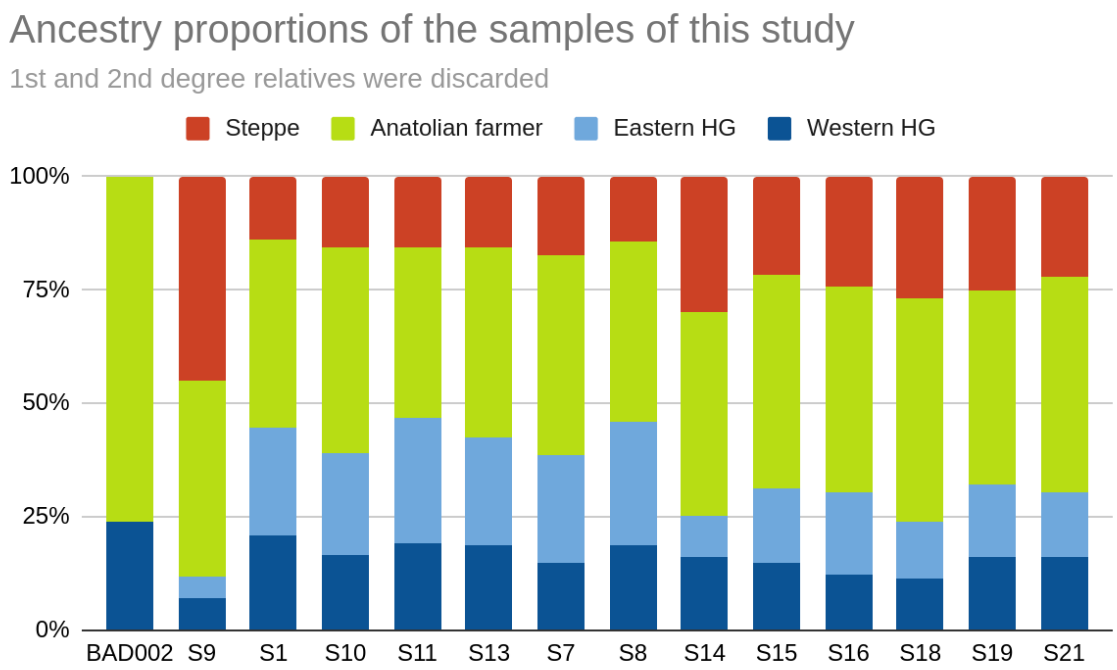


Fig. S.5.2.1

Ancestry proportions of BAD002 and samples from Balatonkeresztúr site. In the transition between Bk-II and Bk-III a clear shift in ancestry proportions can be observed where the HG components decrease and the steppe component increases. Bk-I shows high steppe ancestry, while BAD002 lacks this component.

5.3) *f4*-statistics

We used *R* packages 'admixr' [106] and 'tidyverse' [107] and the Admixtools v7.0 [108] to calculate *f4*-statistics in the form of f4(W=test HG, X=Serbia Iron Gates, Y=target, Z=Yoruba). In theory, if we consider Serbia Iron Gates as sources for most HG ancestries in East-Central European populations, this test could reveal additional HG admixture components on top. Indeed, despite the high standard deviation results for BAD002 suggest additional sources of HG from Western Europe, confidently pointing to Loschbour and Villabruna HG, meaning highly WHG characterised admixture (Figure S.5.3.1). Bk-I surprisingly show rather similar makeup, although the HG levels (and coverage) for S9 are so low that it could bias the results (Figure S.5.3.2). Bk-II (Figure S.5.3.3) and Bk-III (Figure S.5.3.4) show a rather similar makeup with possible candidates of Ireland Mesolithic [109], Italy Mesolithic [97], BDB001 from Germany [62], Mesolithic HG and EMN Narva culture from Lithuania [98], Körös period HG (Hungary) [66,99], Brześć Kujawski Group (BKG) HG outlier (N22) from Poland [110], Iron Gates HG outlier (I5409) from Serbia [55], Loschbour HG [100] and a HG (I1875) from Croatia [55] This highly diverse pattern points to a special, probably yet unsampled HG source somewhere in East-Central Europe most dissimilar to WHG, Iberia HG, Scandinavian HG and EHG (Eastern HG from Ukraine and Russia).

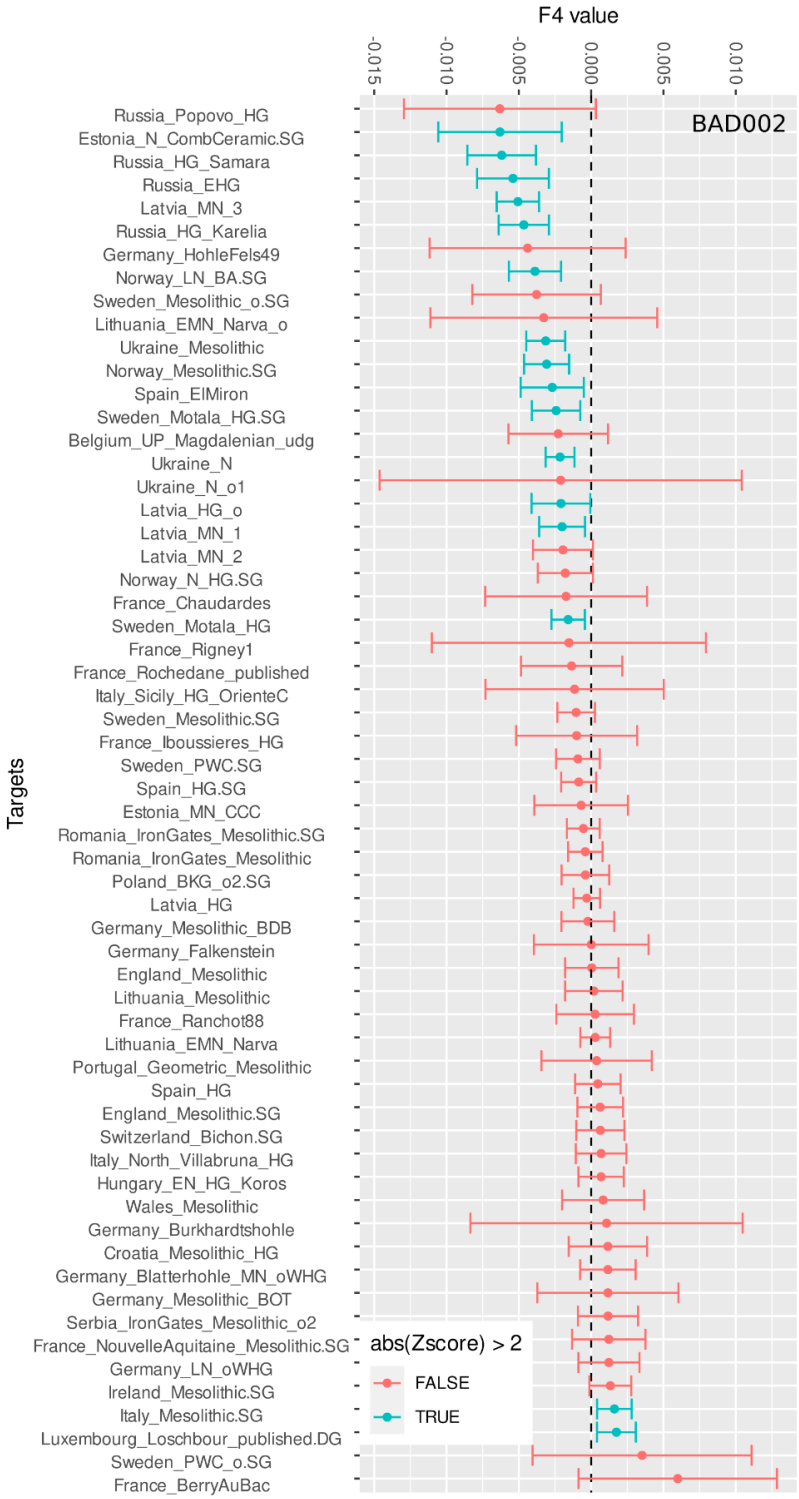


Fig. S.5.3.1

The result of the f4-statistics of the BAD002 individual as target – in the form of (W=test HG, X=Serbia Iron Gates HG, Y=target, Z=Yoruba), where the HG component shows the most resemblance to Western HG-s.

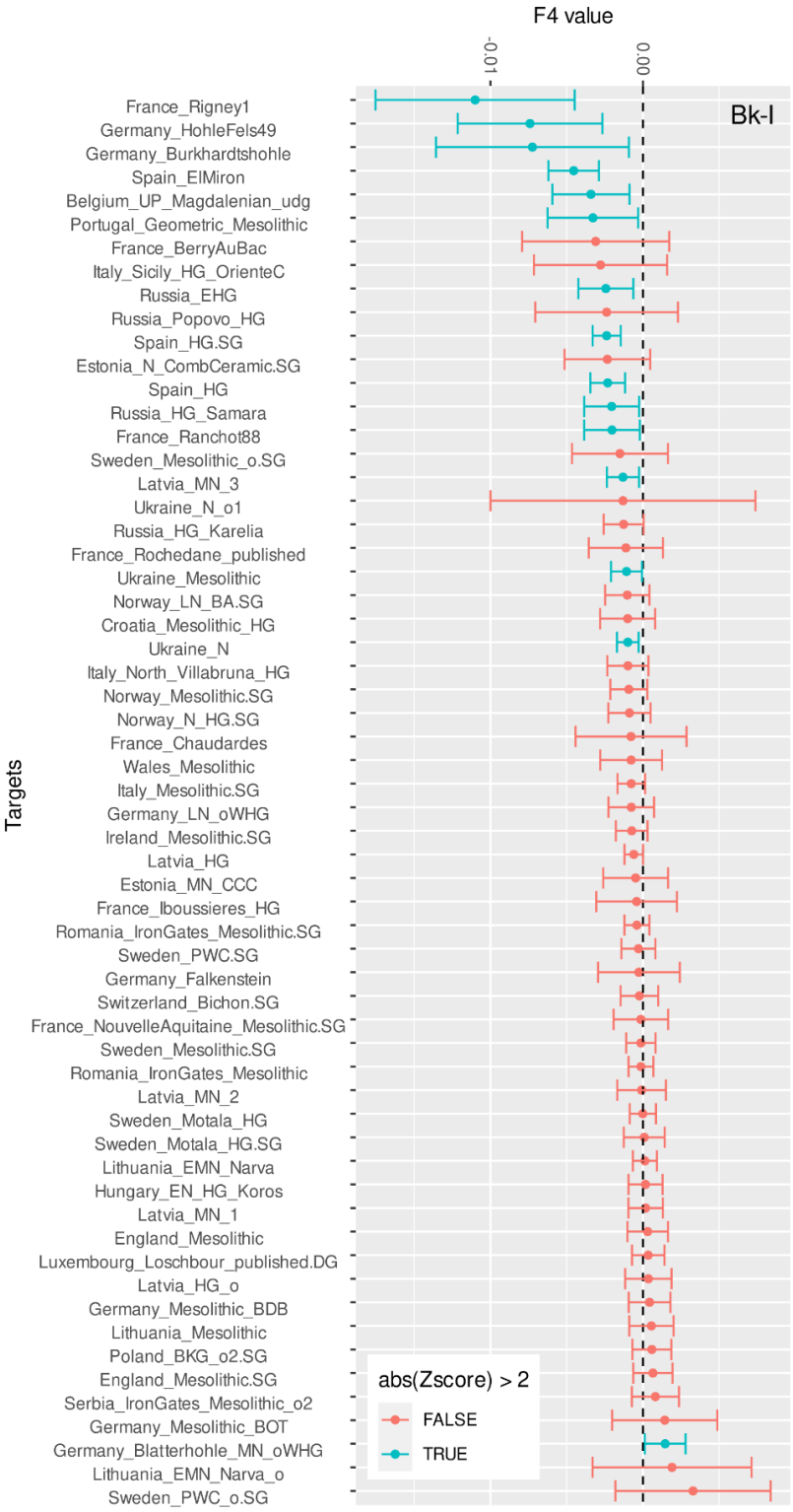


Fig. S.5.3.2

The results of the f4-statistics of the S9 (BK-I) individual as target – in the form of (W=test HG, X=Serbia Iron Gates HG, Y=target, Z=Yoruba). Since the HG component is low for this individual, the results tend to be a bit controversial.

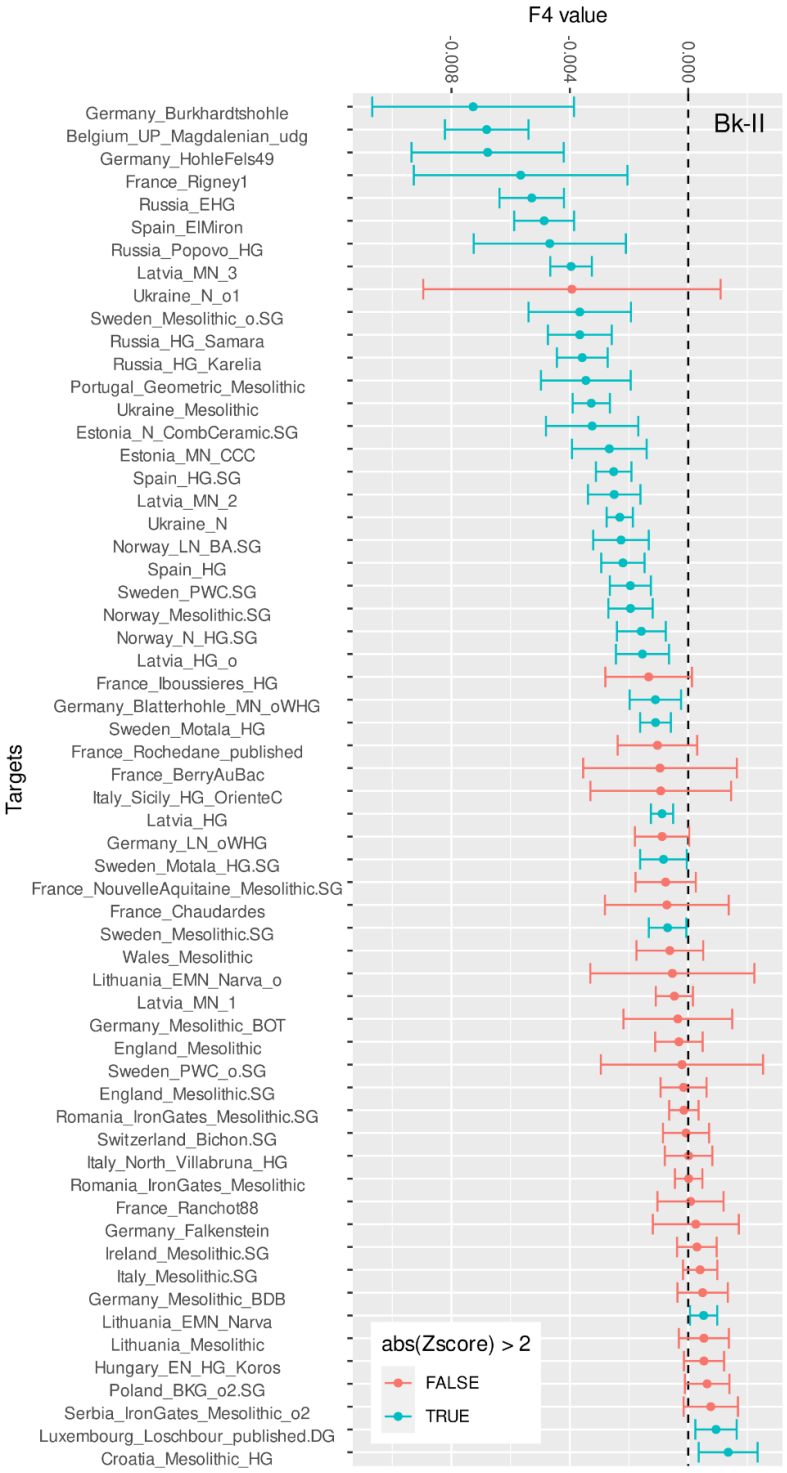


Fig. S.5.3.3

The results of the f4-statistics of the Bk-II individuals as targets – in the form of (W=test HG, X=Serbia Iron Gates HG, Y=target, Z=Yoruba). Best affinities of HG components besides Loschbour tend to point to regions of today's Croatia, Serbia, Poland, Hungary and Lithuania. Loschbour and Irish HG affinity is likely due to the Neolithic component of an FBC related population which groups knowingly had higher, western type HG component [64,65,98].


F4 value
-0.010
-0.005
0.000
Bk-III
Targets
Belgium_UP_Magdalenian_udg
Germany_Burkhardtshohle
Germany_HohleFels49
France_Rigney1
Spain_ElMiron
Estonia_N_CombCeramic.SG
Russia_EHG
Latvia_MN_3
France_Chaudardes
France_Iboussieres_HG
Ukraine_N_o1
Russia_HG_Karelia
Estonia_MN_CCC
Portugal_Geometric_Mesolithic
Ukraine_Mesolithic
Russia_HG_Samara
Spain_HG.SG
Norway_Mesolithic.SG
Spain_HG
Russia_Popovo_HG
Norway_LN_BA.SG
Sweden_Mesolithic_o.SG
Latvia_MN_2
Norway_N_HG.SG
Ukraine_N
Sweden_PWC.SG
Sweden_Motala_HG.SG
Sweden_Mesolithic.SG
Italy_Sicily_HG_OrienteC
Germany_Blatterhohle_MN_oWHG
Latvia_HG_o
Sweden_Motala_HG
Latvia_HG
Germany_Mesolithic_BOT
Switzerland_Bichon.SG
France_Rochedane_published
Romania_IronGates_Mesolithic.SG
France_NouvelleAquitaine_Mesolithic.SG
Italy_North_Villabruna_HG
Germany_Mesolithic_BDB
Germany_LN_oWHG
Sweden_PWC_o.SG
Ireland_Mesolithic.SG
England_Mesolithic
Romania_IronGates_Mesolithic
Latvia_MN_1
England_Mesolithic.SG
Hungary_EN_HG_Koros
France_Ranchot88
France_BerryAuBac
Germany_Falkenstein
Lithuania_EMN_Narva
Wales_Mesolithic
Poland_BKG_o2.SG
Italy_Mesolithic.SG
Luxembourg_Loschbour_published.DG
Lithuania_Mesolithic
Serbia_IronGates_Mesolithic_o2
Croatia_Mesolithic_HG
Lithuania_EMN_Narva_o
abs(Zscore) > 2
FALSE
TRUE

Fig. S.5.3.4

The results of the f4-statistics of the Bk-III individuals as targets – in the form of (W=test HG, X=Serbia Iron Gates HG, Y=target, Z=Yoruba). Due to decreasing HG component the affinities to certain HG groups are less prominent, while in characteristics, highly resembles to Bk-II.

5.4) *f3*-statistics

The outgroup *f3*-statistics and clustering analyses were based on Patterson et al. 2012[108] and made by Admixtools v7.0 and using *R* packages 'admixr', 'tidyverse', 'ggplot2' [107], 'plotly' [111], 'heatmaply' [112], 'RColorBrewer' [113]. In theory the concept is simple: we are interested in relative divergence times between pairs populations, and want to know in which approximate order they split off from each other, *i.e.* check whether the genome or a certain genomic component is closely or distantly related to a selected group. To address this problem, we could use the *f3*-statistics by fixing the outgroup as Yoruba, and calculating pairwise *f3*-statistics between all HG and HG admixed groups. As a result, pairwise distances reveal candidates as a source for admixture analyses. The results are similar to the results of the *f4*-statistics in all of the investigated groups if we use only HG-s as potential source populations. Euclidean distance based clustering [114] grouped populations together having most similar patterns in HG composition (Figure S.5.4.1). It is notable that the similarities of Loschbour to Bk-II and Bk-III could be the result of admixing Neolithic groups (see in section 5.5) qpAdm) containing large portions from this type of HG ancestry, rather than the direct similarity to this unsampled HG source.

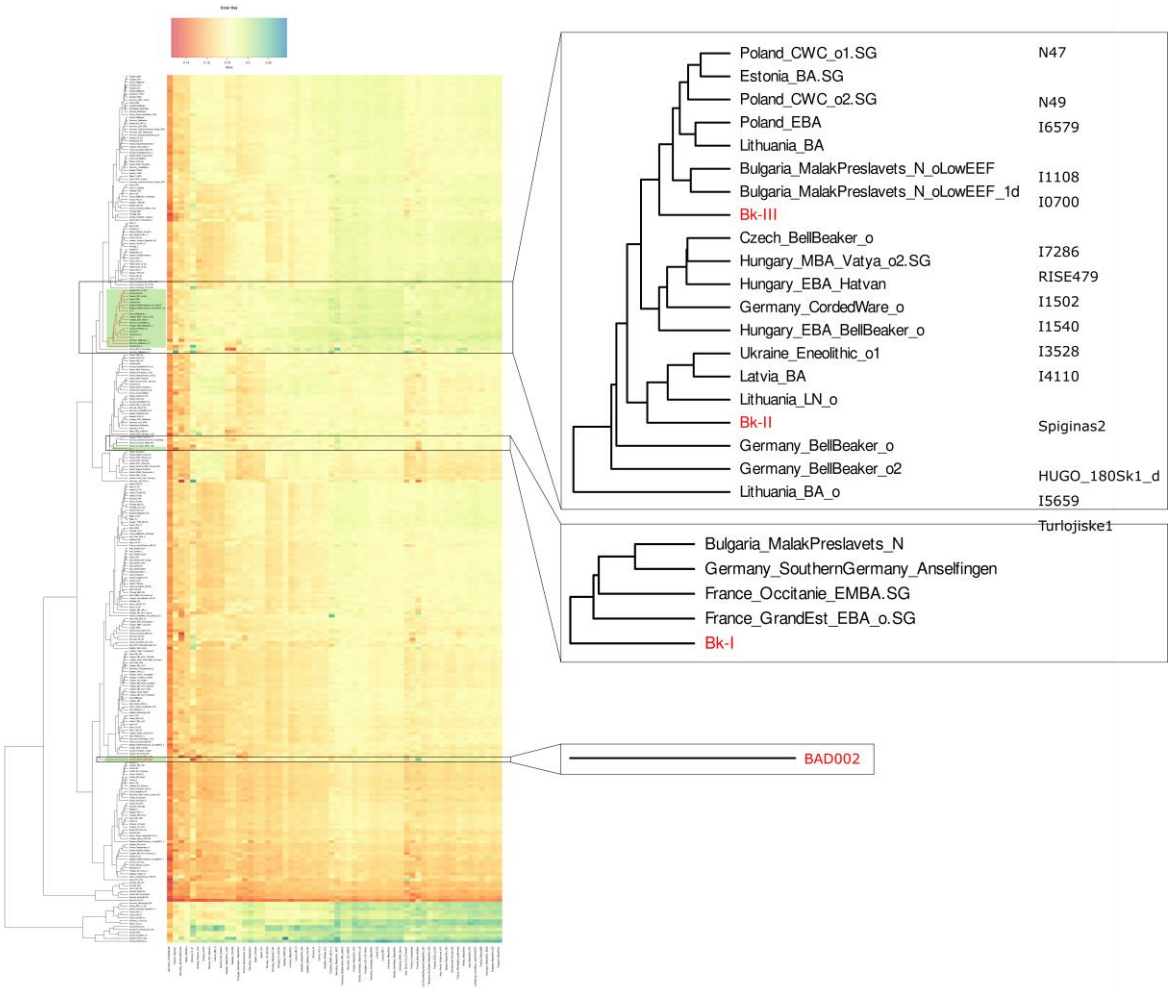


Fig. S.5.4.1

Euclidean clustering of outgroup f3-statistics results. The X axis represents HG individuals and populations and the Y axis all tested individuals and populations. The individual BAD002 is completely dissimilar to any known test-populations from the Y axis, likely due to an extra HG component that did not contribute to any sampled Neolithic population from Europe. Bk-I (individual S9) resembles most to Northern Bronze Age groups in terms of HG affinity, while Bk-II and Bk-III are creating their unique cluster consisting of a number of outliers and three Baltic Bronze Age populations.

5.5) qpAdm

qpAdm calculations were performed for each of our groups for which we used Admixtools v7.0 and R packages 'admixtools' [108] and 'tibble' [107]. Potential sources of ancestry were selected based primarily on principal component analysis (PCA), ADMIXTURE and uniparental analyses results to avoid unnecessary model testing and reduce computational time. For outgroup we chose 'Mbuti.DG', 'Russia_Ust_Ishim.DG', 'Russia_MA1_HG.SG', 'Ethiopia_4500BP_published.SG', 'Russia_Kostenki14', 'Iran_GanjDareh_N', 'Georgia_Kotias.SG', 'Russia_North_Caucasus', 'Luxembourg_Loschbour.DG', 'Russia_EHG', 'Israel_Natufian_published', 'Russia_Afanasievo', 'Spain_HG', 'Russia_Steppe_Maikop', and 'Germany_BellBeaker' (only for BAD002). We created a custom R script (provided in the 'PAPline') for running all possible combinations of single, dual and trial ancestry sources. Since qpAdm results tend to give high *false* positive rates in case of low coverage,

we restricted analyses to source populations that have at least 300,000 SNPs in combination (*e.g.* either a single sample above 300k SNPs or a population of 3 samples with at least 100k SNPs each, *etc.*), we made one exception by adding Bk-I to model selection for Bk-II and Bk-III. Our script automatically rejects combinations where negative components appear or standard errors reach a limit while it provides a list of best combination candidates (*i.e.* models that were not rejected in an ascending order of *p*-values). We considered the model the most plausible when standard deviation was low (we set a threshold of below 0.1 and/or lower than the corresponding component's weight), has a decent (above 0.05) *p*-value, and is in line with the results of *f3*/*f4*-statistics and uniparental makeup. Multiple possible sources emerged especially in the case of Bk-II and Bk-III (Figure S.5.5.1), here we applied different strategies to reject or accept models, discussed below.

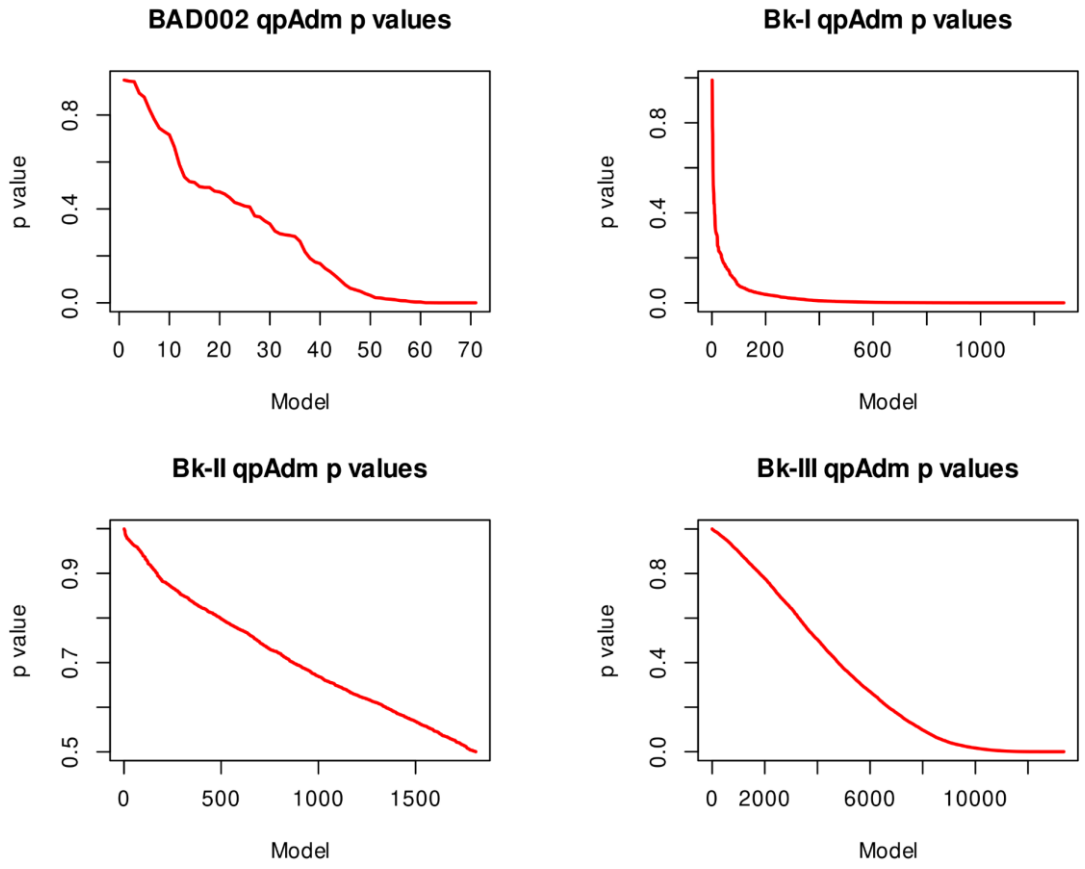


Fig. S.5.5.1

qpAdm models p-values in ascendant order for all of our studied groups, where each element in the X axis represents a particular model. Bk-I has the most ideal slope having a few really high p-value models for ancestry, whereas Bk-III has the less ideal slope with more than 10k possible models.

5.5.1) BAD002

We selected a set of possible candidates from Hungary (Balaton-Lasinja, Lengyel and Baden cultures [66,99], as these are the most proximate to BAD002 geographically and chronologically), Western Europe (France and Iberia according to previous results and Britain for proxy) as well Villabruna and El Mirón cave WHG-s [101]. Resulting models for the Neolithic components from the present-day Hungary and British Isles were rejected. On the other hand, models fit really well with Neolithic groups from modern day France and Spain with and without an additional WHG component with minor differences in *p*-value. Finally, after the higher drop of *p*-values under 0.9, we considered the best models the top 3, despite

48 models being statistically significant (above *p*=0.05), see Table S13. Accordingly, either Hauts de France Neolithic or France Occitanie Neolithic represented as the main source with or without an addition of El Mirón or Villabruna HG. Applying the outgroup exclusion strategy, *i.e.* adding one possible source to outgroup and run the test for another candidate, revealed that France Occitanie group is the most plausible source for BAD002. However, since *f3*- and *f4*-statistics exclude El Mirón as possible HG source, and the results of *f3*-statistics suggest that the main HG is Ranchot88 from France, we conclude that the most likely scenario is indeed the one with the best *p*-value in the original run: Hauts De France (~87±8%) + Villabruna HG (~13±8%) with *p*=0.949, see Supplementary Table S12.

5.5.2) Balatonkeresztúr site

For all horizons we selected a number of sources from all over Neolithic to Bronze Age Europe consisting of an overall set of 93 potential sources in ~148k combinations. For Bk-II we also added Bk-I and for Bk-III we added both Bk-I and Bk-II in order to see whether gene flow between horizons occurred.

5.5.2.1) Bk-I

Too many best fitting models exist above *p*=0.05 for Bk-I, thus we excluded these below *p*=0.1 and examined the remaining 86. Considering the high drop of p values after the first two best models, by applying outgroup exclusion Poland Southeast BBC [115] turned out to be the best candidate (*p*=0.78, *p*=0.712 when Hungary MBA Vatya [72] is among the outgroups) for modelling Bk-I, see Supplementary Tables S12 and S14.

5.5.2.2) Bk-II

Too many best fitting models exist for Bk-II without significant drop of *p*-values, thus we restricted the search for the most possible model by excluding all models below *p*=0.95, obtaining 81 possible scenarios. This data, on the other hand, can be filtered further by leveraging the trend between models: each one is a three way model consisting of a Neolithic, a highly steppe related and a HG component, latter highly supporting the introgression of an isolated group with remnant HG genetic composition or a group with extremely high HG component compared to Early Bronze Age populations. Analysing further this data, we marked different HG groups (Romania IronGates HG, Latvia HG, Lithuania Narva culture, Sweden Motala HG, Sweden Mesolithic, Romania oHG, Ukraine Neolithic) as the same for the unknown exact HG source discussed further at section 5.2 and 5.3, then deleted duplicates keeping models with the best *p*-values. 38 models of an overall 26 sources remained (Figure S.5.5.2.2.1, Table S15), with differing frequencies. Here, we applied outgroup exclusion on the Neolithic component, as this one is the most restricted to a few candidates among all, and as a result, Sweden FBC [64] turned out to be the most likely source. However, all *p*-values remain above 0.6 in this test, except for Germany Middle Neolithic Blätterhöhle, which likely got high *p*-value models (*p*=0.987) for their outstandingly high HG component. Globular Amphora culture (GAC), FBC, Brześć Kujawski Group (BKG) and Megalithic cultures are archaeologically and genetically related to each other, thus as a final conclusion on this component, we suggest that while Sweden FBC is the best fitting model, Poland GAC as the most frequent source fits best chronologically, archaeologically and geographically to Bk-II, while a previously unsampled but genetically highly similar group can not be excluded as source. On the other hand, for steppe-related sources this strategy provided controversial results, likely due to the lower diversity within this component. Thus we considered the most frequent as well as the highest *p*-value model as the most plausible for modelling Bk-II (HG: ~29±3%, Poland GAC: ~31±8%, Poland BBC: ~40±6%), for results, see Supplementary Table S12. Since this model presents Poland Southeast BBC as the third component, we can not confidently exclude the possibility of gene flow between Bk-I and Bk-II, while at best only a model of *p*=0.75 supported this scenario, which, again, could be the result of low coverages. Finally, possible models for HG + FBC/GAC + BBC/Bk-I are listed in Supplementary Table S15.

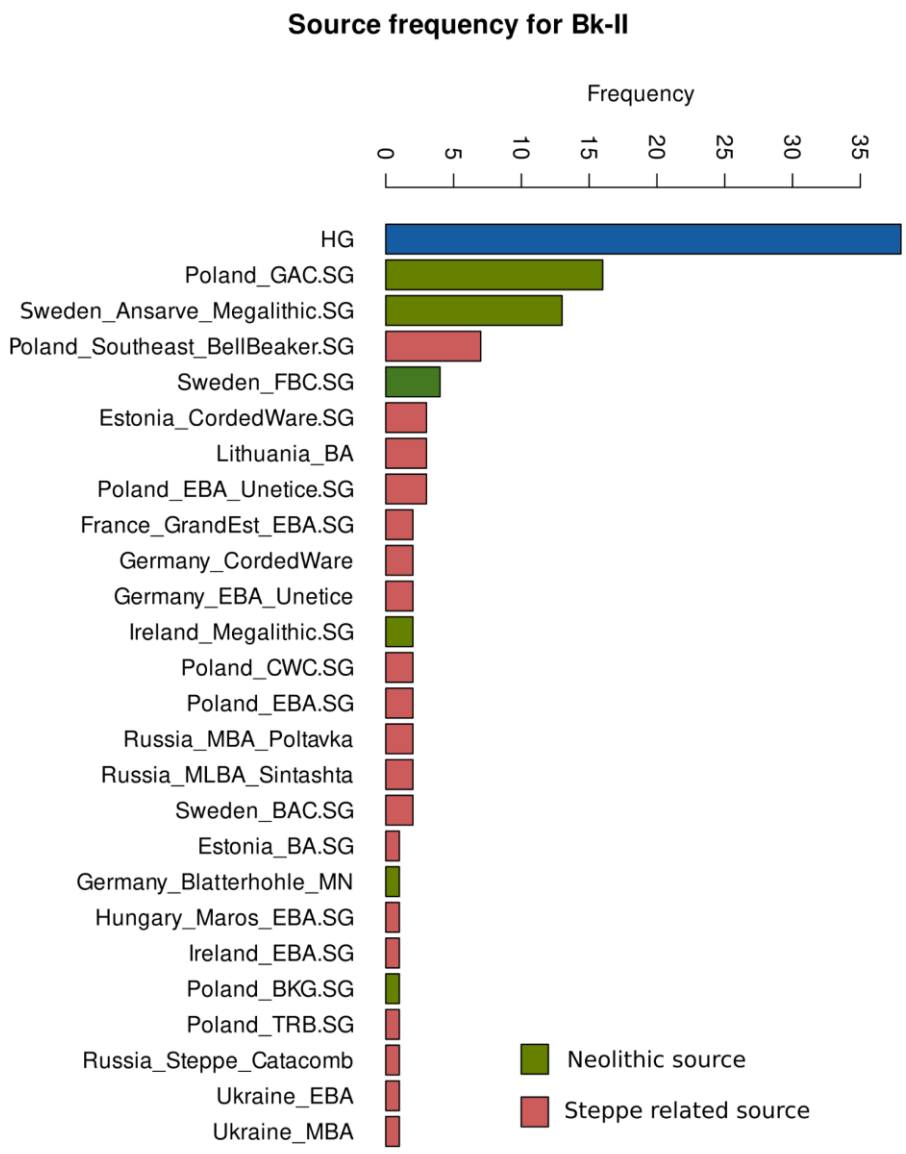


Fig. S.5.5.2.2.1

Source frequency in qpAdm model series for Bk-II, where a merged (yet undefined) HG population is present in each.

5.5.2.3) Bk-III

There is no significant difference between the first 100 best models (all above p=0.99), thus a similar method was applied to see the possible sources for Bk-III. According to source frequency results, Bk-II appears in half of the cases, which we consider a true source for previous results, thus we excluded models where this population did not appear. Bk-II represents the main source for Bk-III in each case (average weight is 53±5%), however, finding the other half of the sources is really challenging, which issue was also present in case of the Jagodnjak group. According to previous results, steppe ancestry increased in Bk-III, which, in line with the observations of Freilich *et al.* (2021) [116], likely suggest local, eastern or northern origin (Figure S.5.5.3.1). This assumption is also supported by the most frequently appearing regions (Poland 19 cases, Sweden and Italy 15 cases, Hungary 14 cases) among the 50 best models next to Bk-II (Table S16), which are also frequent among mtDNA parallels. As a final

conclusion at this point we can only suspect that the “dilution” of Bk-II to Bk-III is mostly driven by population histories that are yet to be uncovered.

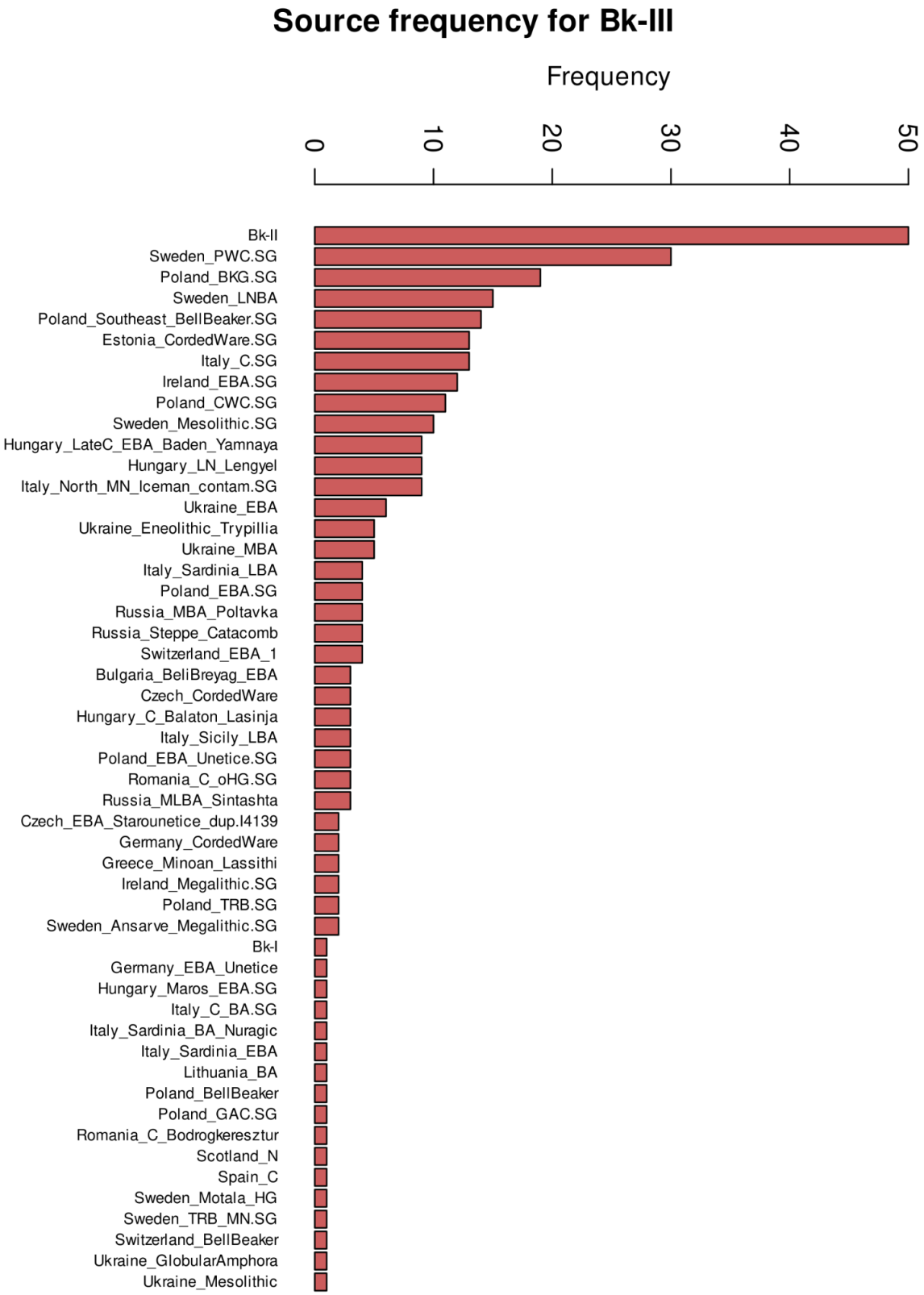


Fig. S.5.5.3.1

Source frequency in acceptable qpAdm models for Bk-III, where Bk-II is present in half of the cases.

6) 'PAPline' description and the bioinformatic analyses performed in this study

by Dániel Gerber, Eszter Ari

6.1) The overall description of the 'PAPline'

'PAPline' is an abbreviation for "Performing Archaeogenetic Pipeline", it can be downloaded from https://www.github.com/gerberd-workshop/papline and used freely under Debian based Linux distributions. This pipeline was written in *bash* and is supplemented by scripts written in *R* and *python* v3.8.10 programming languages. The program aims to analyse raw sequencing data and perform read alignment, apply various data filtering methods, and finally provide summary tables about the run with additional haplogroup assessments, genetic sex determinations, aneuploidies, *etc*. The concept of 'PAPline' is highly similar to the 'Paleomix' [117] and the 'EAGER' [118], however, we aimed to create a program that needs only a few adjustments before actual application, and is also user friendly. The software and R packages that are needed to be installed and added to the path to use the 'PAPline' are listed in SI Table 3. Regular updates are planned for 'PAPline' in the future by releasing new versions every second year.

SI Table 3

Name	Available from	Ubuntu repository
ANGSD	http://www.popgen.dk/angsd/index.php/ANGSD	
bamtools	https://github.com/pezmaster31/bamtools	✓
bamUtil	https://github.com/statgen/bamUtil	
bedtools	https://bedtools.readthedocs.io/en/latest/	✓
Bowtie2	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml	✓
BWA	https://github.com/lh3/bwa	✓
ContamMix	contact Philip L.F. Johnson plfj@umd.edu	
cutadapt	https://github.com/marcelm/cutadapt	✓
FastQC	https://github.com/s-andrews/FastQC	✓
Haplogrep2	https://github.com/seppinho/haplogrep-cmd	
Mafft	https://mafft.cbrc.jp/alignment/software/	✓
MapDamage	https://ginolhac.github.io/mapDamage/	✓
plink1.9	https://www.cog-genomics.org/plink2	✓
preseq	https://github.com/smithlabcode/preseq	
samtools	https://github.com/samtools/samtools	✓
SeqPrep	https://github.com/jstjohn/SeqPrep	✓
SequenceTools	https://github.com/stschiff/sequenceTools	
vcftools	https://vcftools.github.io/index.html	✓
yLeaf2	https://github.com/genid/Yleaf	
admixr	R package at CRAN	
admixtools	R package at https://github.com/uqrmaie1/admixtools	
dplyr	R package at CRAN	✓
ggplot2	R package at CRAN	✓
ggrepel	R package at CRAN	✓
seqinr	R package at CRAN	✓
stringr	R package at CRAN	✓

Name	Available from	Ubuntu repository
tibble	R package at CRAN	✓

The following parameters are needed to run the PAPline:

- parameter file
- input directory
- output directory
- sample list file

The parameter file contains information about the run parameters, it can be created by using text editor or the graphical interface of the PAPline. Input and output directories are straightforwardly can be used, while the sample list file needs special attention. This latter file is a *csv* file of sample names and inner barcodes (when absent, leave this field empty), where each line represents one sample and the corresponding inner barcode sequence. For an example see SI Table 4.

SI Table 4

SampleID	P5 barcode	P7 barcode
S001	AGTCTAG	AGATCGA
S002	GGCCTAG	TACAGTA
S003	CCAGATA	TTACAGA

However, **the header must be excluded from the actual csv file** and the first line has to be the first sample! After installing the 'PAPline' an example sample list file (named as list.csv) is provided in the folder /xmp. No varying columns are allowed in the *csv* file to avoid mixing samples with one, two or without barcodes.

6.2) Options

PAPline is available in the CLI version with a graphical interface panel (Figure S.6.2.1) for creating the parameter file.

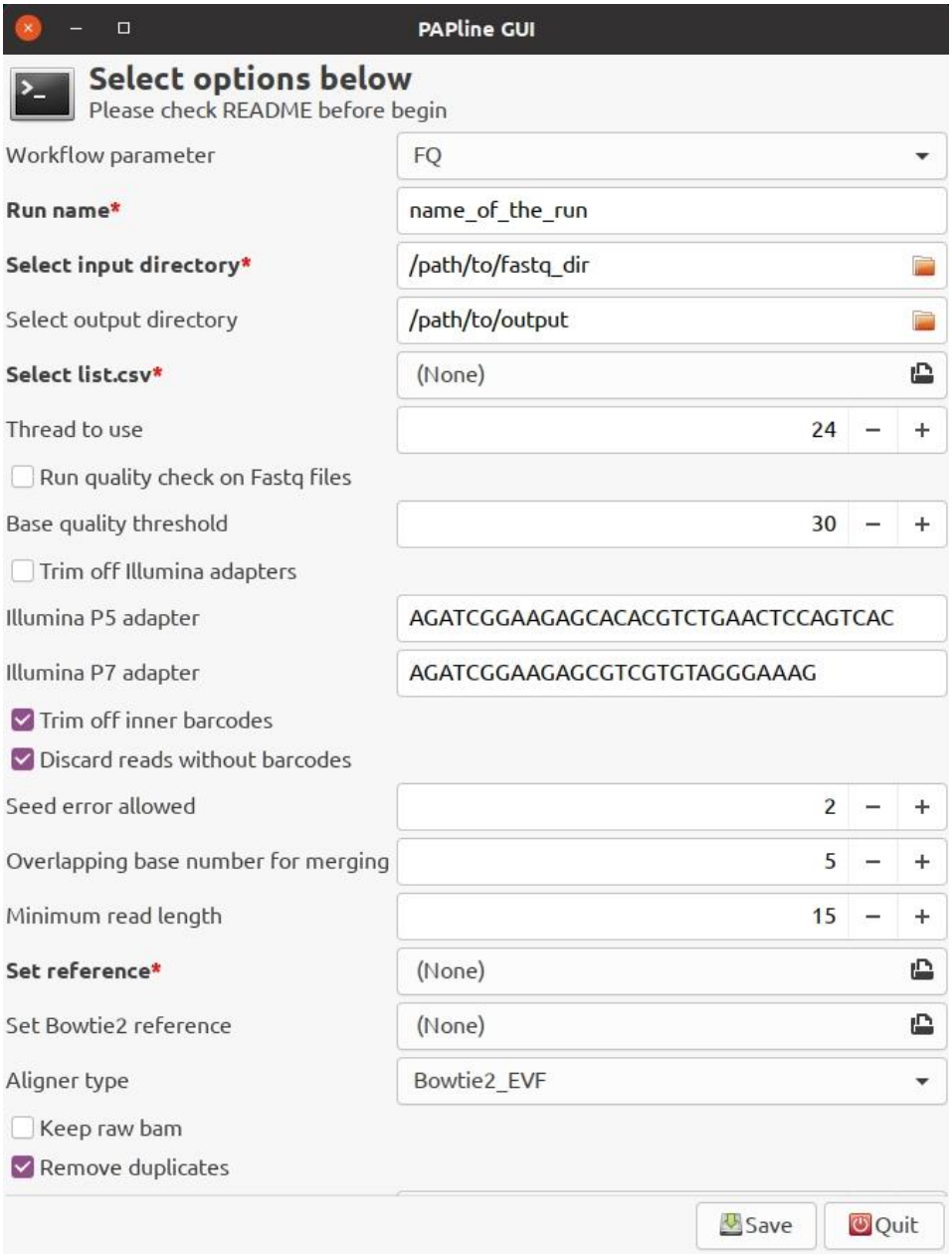


Fig. S.6.2.1

The graphical interface panel to create the parameter file in PAPline, which can be scrolled down for further parameters.

When starting a PAPline session from the Terminal by typing `papline` the following options are available:

```
-a     Auto mode
-h     Help
-g     GUI panel
```

The “auto mode” expects a parameter file, for which an example is available under the /xmp folder in the installed version of the ‘PAPline’, and besides the GUI panel it can be created manually. The parameter file may contain options listed in SI Table 5.

SI Table 5

*Parameters marked with a * are essential. Multiple choice means as it is described in the “Description”; STR means string; PATH means the absolute or relative path to the file; FILE means the name of the file; INT means an integer number; TRUE/FALSE mean TRUE or FALSE logical type of operator.*

Name	Type	Description
*`workf`	Multiple choice	Setting up the main workflow: `FQ` (*fastq*) starts the full pipeline and expects a directory where either gzipped or unzipped *fastq* files are present. Paired end (PE) read files must contain “R1” and “R2” in the file names and the file extensions must be either “fq”, “fq.gz”, “fastq” or “fastq.gz”. Single end (SE) read files are expected with no “R1” or ”R2” present in filenames. Also, `FQ` option allows more than 2 *fastq* files per sample, *i.e.* if a sequencing run resulted in *e.g.* 3 pairs of *fastq* files per sample then ‘PAPline’ can process these as a single sample, but correct formatting is needed for this, see section “6.6) PAPline file formatting requirements”. `BAM` expects *bam* files. `PAPLINE` indicates the usage of ‘PAPline’ library structure, *i.e.* if a run was started and subfolders are created but for some reason the run was interrupted or need to be rerunned, ‘PAPline’ continues from the last interrupted step instead of deleting all existing libraries, however, fails when expected files to continue with are missing or were being relocated. It uses a “*papli_log*” file saved to the “*indir*” folder, further described below, that lists out finished steps, and can be maintained manually.
*`runam`	STR	The name of the run, if `workf=FQ/BAM` a folder with the same name will be created at the “*endir*” path, or the “*indir”* path in case the “*endir”* is missing.
*`indir`	PATH	Path to the input folder containing *fastq* or *bam* files or the ‘PAPline’ file structure, according to `workf`.
`endir`	PATH	Path to the output folder, where directory will be created under a name provided in `runam` when `workf=FQ/BAM`. Existing directories under the same name will be deleted and re-created, except when the `PAPLINE` option is `TRUE`, in which case the `endir` option will be ignored.
*`listf`	FILE	The name of the comma separated (*csv*) file containing samples and corresponding barcodes, without header. Example is given in the /xmp folder. If no barcodes are given, the ‘PAPline’ automatically skips the `btrim` and the `bdisc` options. When the `workf=BAM` no other columns are considered besides the first with the sample names. No varying columns are allowed in the *csv* to avoid mixing samples with one, two or without barcodes. ‘PAPline’ can only process a single data type (PE (paired end), SE (single end), barcoded, or non-barcoded) in a single run.
`thred`	INT	Number of threads used by the ‘PAPline’. The default value is 24.
`fstqc`	TRUE/FALSE	When `fstgc=TRUE` (default option) the ‘PAPline’ will perform quality checks on each *fastq* file by using the ‘FastQC’ software.
`bqthr`	INT	Base quality (BQ) threshold (PHRED score, Illumina based value range 0-62) to be considered in analyses, default is 30, if 0, no BQ check or filtering will be applied during analyses.
`atrim`	TRUE/FALSE	Argument to set if Illumina adapters from the end of the sequencing reads should be trimmed off by the ‘PAPline’. The default value is `FALSE`.
`p5ada`	STR	Defines the P5 adaptor sequence when `workf=FQ`.
`p7ada`	STR	Defines the P7 adaptor sequence when `workf=FQ`.
`btrim`	TRUE/FALSE	When `workf=FQ` and `btrim=TRUE` (default option) the ‘PAPline’ trims off inner barcodes provided in the *csv file*.
`bdisc`	TRUE/FALSE	When `workf=FQ` and `bdisc=TRUE` (default option) the ‘PAPline’ discards reads lacking a barcode sequence.

Name	Type	Description
seeds	INT	Sets maximum differences in seed for the atrim and the btrim options, as well as the 'BWA ALN' mapping method. The default value is 2.
overl	INT	When workf=FQ and overl>0, paired end (PE) reads will be merged to single fragments and saved to a *fastq* file to create an input for the read mapping. The overlINT parameter gives the number of basepairs to overlap between read pairs. The default value is 5.
minle	INT	When workf=FQ minle will define the minimum read length of reads to be considered. Shorter reads will be discarded from further analyses. The default value is 15.
*refer	FILE	The name of the reference *fasta* file, which has to be indexed with 'Samtools' and 'BWA'.
befer	FILE	Filename of the 'Bowtie2' indexed reference file. When using the 'Bowtie2' reference in the graphical interface mode, one has to select one of the indexed reference genome files, and 'PAPline' will automatically trim off the unnecessary tags. When the parameter file is being created manually, the path still has to point to an existing file, *i.e.* the 'Bowtie2' tag must remain or the 'PAPline' will provide an error message regarding the 'Bowtie2' indexing error.
align	Multiple choice	When workf=FQ the alignment method can be selected using this argument. The available methods are: BWA_Aln, BWA_Mem, Bowtie2_EVF (end-to-end very fast), Bowtie2_EF (end-to-end fast), Bowtie2_ES (end-to-end sensitive), Bowtie2_EVS (end-to-end very sensitive), Bowtie2_LVF (local very fast), Bowtie2_LF (local fast), Bowtie2_LS (local sensitive), and Bowtie2_LVS (local very sensitive). The default method is Bowtie2_EVF. The parameter setted to runam will be added as read groups to 'Bowtie2' mappings.
krbam	TRUE/FALSE	When krbam=TRUE the 'PAPline' will keep the initial raw *BAM* files after various filters were applied. The default is FALSE.
rmdup	TRUE/FALSE	Remove PCR duplicates from the *BAM* file. The default value is TRUE.
mapqt	INT	Mapping quality (MQ) threshold (PHRED score, Illumina based value range 0-62) to be considered in analyses, default is 30, if 0, no MQ check or filtering will be applied during analyses.
mapda	TRUE/FALSE	Run the 'MapDamage' software to check damage patterns. The default value is TRUE.
cntmx	TRUE/FALSE	Run the 'ContaMix' software to assess contamination levels. The default value is FALSE. (It is really slow.)
ctrim	INT	The number of base pairs to be trimmed from both ends of the reads, if ctrim=0 this step will be skipped. The default value is 2.
mtfas	Multiple choice	Defines the mode for base calling method to create a mitochondrial DNA *fasta* file. The available options are: Permissive: At positions where the coverage is equal to the value given by the mtcov argument a base is only called when all reads contain the same base. Below this given threshold no bases are called, above this given threshold the majority rule will be applied. This option is default. Majority rule: The majority rule is applied to call a base, minimum coverage is given by mtcov argument. Random: Calls the base from a randomly selected read if the position is covered at least by the number given in the mtcov argument. None: No *fasta* file will be created.
mtcov	INT	Coverage threshold for calling a base when creating the mitochondrialDNA *fasta* file. The default value is 2.
cotrv	TRUE/FALSE	When cotrv=TRUE the 'PAPline' calls only transversions for nuclear genome even if the *BED* file contains transitions as well, and considers only transversions when assigning the Y chromosome haplogroup. The default value is FALSE.
psudo	TRUE/FALSE	When psudo=TRUE the 'PAPline' calls pseudohaploid genome. The default value is TRUE.
disis	INT	Indicates the variant quality for calling SNPs of clinical significance. It is similarly calculated as the BQ or MQ, *i.e.* bqthr=10 equals to disis=10. Also, only heterozygous calls are made when disis>0 (the default value is 30). disis=0 indicates skipping this step.
pigme	INT	Indicates the variant quality for calling SNP-s of pigmentation. It is similarly calculated as the BQ or MQ, *i.e.* bqthr=10 equals to pigme=10. Also, only heterozygous calls are made when pigme>0 (the default value is 4). pigme=0 indicates skipping this step.
glqua	INT	Defines the genotype likelihood quality. It is similarly calculated as 'pigme' quality. The default value is 1. glqua=0 indicates skipping this step, only when psudo=FALSE.
eigen	TRUE/FALSE	When eigen=TRUE the 'PAPline' calls the genotypes and transforms data into *EIGENSTRAT* format. The default value is FALSE.
plink	TRUE/FALSE	When plink=TRUE the 'PAPline' calls genotypes and transforms data into *PLINK* format. The default value is FALSE.

The graphical interface version opens a panel where all of the options listed in SI Table 5 can be adjusted and saved.

While a number of software have to be pre-installed before using the 'PAPline', the pipeline runs properly if the installation fails for an unused software.

6.3) Library structure of the PAPline

The `endir` option creates a working directory. Under this working directory subdirectories of samples will be created. Besides, 'PAPline' places the final report file and the individual result files here. The final report file contains all the information most frequently needed for an ancient NGS data analysis, *e.g.* number of reads entered the pipeline, number of reads mapped, average genomic coverage, mitochondrial DNA haplogroup, *etc*. This report file is formatted in a supplementary-ready format, thus exhaustive manual labour can be saved. Also, the *papli_log* file is placed here which contains information about the run, *i.e.* each line represent run step tags marked as true or false, *e.g.*:

```
fstqc_run=T
bqtrm_run=T
atrim_run=F
...
```

This *papli_log* file enables the user to take full control over workflow steps, when the `papli=TRUE` . It switches on and off workflow steps, *i.e.* all steps set to be *F* will be processed, whereas steps set to be *T* or missing from list will be skipped. However, when `workf=PAPLINE` *fastqc_run* and *bqtrim_run* logs files will be ignored because the input for these are raw BAM files from outside of the 'PAPline' library structure.

In the sample folders a number of files and other sub-folders can be found, each one contains various files produced by steps of the workflow, *e.g. ALNFILES* folder contains the finalised BAM file(s), *etc*. Files in the sample folder consist of a number of files, *e.g.* results of 'MapDamage' run, or the results of coverage calculations.

6.4) Running the PAPline

Finally, when everything is set, 'PAPline' can be run by using the following command:

```
./papline -a /path/to/parameterfile
```

The running time depends on the hardware, the size of the input files, and the workflow settings. For example calculating the results for this study took a couple of hours on a 24 threaded Ryzen CPU and 36 Gb RAM PC when applying the settings described in the section "6.5) PAPline options used in this study". If the parameter file contains discrepancies, the program shuts down and asks for revision.

Eigenstrat and PLINK formatted BED and SNP files can be found under the */references* folder. The pre-formatted BED files used for the pigmentation and disease essays are located in the same folder. These files can be modified if needed. Besides the basic default pre-formatted BED files contain the dbSNP IDs, the positions of the hg19 chromosomes, the reference and alternate allele, as well as the phenotype of the variants, which data table can be extended or modified at will by new data, except headers.

6.5) The PAPline options used in this study

We ran the 'PAPline' to analyse data of this study by using the following options and steps (most of them are the default parameters):

- end trim of reads by BQ threshold of 30 ('cutadapt')
- barcode trimming allowing 1 mismatch and discarding read pairs not having any of the barcode sequences, thus excluding DNA fragments that possibly contaminated the sample post-sterile laboratory ('cutadapt')
- merging paired end (PE) reads to single fragments by 5 base pair overlap ('SeqPrep-master')
- mapping with the 'Bowtie2' EVF following the recommendations of Poullet & Orlando (2020) [119]
- duplication remove ('samtools')
- 'MapDamage' and 'ContaMix' were applied
- hard clipping of read ends (2 base pair)
- Creation of the *fasta* file from reads of the mitochondrial DNA with the `mtfas='Permissive` option by using 2x minimum coverage, detailed in section "2.1) The mitochondrial DNA haplogroups and their phylogenetic evaluation"
- obtaining EIGENSTRAT and PLINK formats by using `mapqt=30` and `bqthr=30`.
- pigmentation (detailed in section "3.2.1) SNPs of pigmentation"), disease (detailed in section "3.2.2) SNPs of clinical significance") and Y chromosome haplogroup (detailed in section "2.2) Y chromosomal haplogroups and STR network analyses") assessment, as well genetic sex and aneuploidy determination (detailed in section "3.1) Genetic sex and aneuploidy").

6.6) The file formatting requirements of PAPline

Since 'PAPline' offers multiple input processing, certain requirements are needed when naming the input files. First, when analysing paired end (PE) reads, each *fastq* file must contain r1/R1 and r2/R2 in their names to distinguish p5 and p7 pairs. Second, when more than two pairs of *fastq* files exist per sample, the correct filename should look like the following:

```
SAMPLEID_someinfo_otherinfo_PAIRID_someinfo_R1/2_someinfo.fastq
```

Where the "SAMPLEID" is the ID of the sample, "someinfo" and "otherinfo" can be anything while "PAIRID" also can be anything but these have to be unique for the pair of fastq files. When the filenames are correct the 'PAPline' can do the pairing of the files. Parts of the file name, like "SAMPLEID", "someinfo", *etc.* must be separated by single underscores "_", since "PAPline" only considers underscore separated parts of the filename for pairing.

7) References

1. Bondár M. Késő rézkori különleges temetkezések Balatonlelléről (Somogy megye). *Kaposvári Rippl-Rónai Múz. Közleményei* 89–108 (2020) doi:10.26080/krrmkozl.2020.7.89.
2. Bondár M. *et al.* KÜLÖNLEGES „GAGÁT" GYÖNGY EGY KÜLÖNLEGES KÉSŐ RÉZKORI SÍRBÓL. *Archeometriai Műh.* **18**, 143–156 (2021).
3. Bondár, M. & Szécsényi-Nagy, A. Skull cult in the Late Copper Age. *Ziridava* **34**, 91–104 (2020).
4. Honti S. *et al.* Régészeti kutatások az M7-es autópálya Somogy megyei szakaszán és a 67-es úton (2004–2005) Elõzetes jelentés IV. **17**, 64 (2006).
5. Fábián, S. & Serlegi, G. Settlement and environment in the Late Copper Age along the southern shore of Lake Balaton in Hungary. in *Reimagining Regional Analyses: The Archaeology of Spatial and Social Dynamics* 199–231 (Cambridge Scholars Publishing, 2009).
6. Honti S. *et al.* A tervezett M7-esautópálya Somogy megyei szakaszának megelőző régészeti feltárása (2002–2003) Előzetes jelentés III. **16**, 3–70 (2004).
7. Dani, J. The Significance of Metallurgy at the Beginning of the Third Millennium BC in the Carpathian Basin. in *Transitions to the Bronze Age. Interregional Interaction and Socio-Cultural Change in the Third Millennium BC Carpathian Basin and Neighbouring regions* 203–231 (Archaeolingua, 2013).
8. Fischl, K. P., Kiss, V. & Szeverényi, V. Transformations in the Carpathian Basin around 16oo B. C. in *Kultureller Umbruch im Schatten des Thera-Ausbruchs?(16oo – Cultural change in the shadow of theThera-Eruption?).* vols 14–16 355–371 (Tagungen des Landesmuseums für Vorgeschichte Halle, 2013).
9. Dani, J., Fischl, K. P., Kulcsár, G., Szeverényi, V. & Kiss, V. Visible and invisible inequality in Early and Middle Bronze Age Hungary. in *Arm und Reich – Zur Ressourcenverteilung in prähistorischen Gesellschaften. Rich and Poor – Competing for resources in prehistoric societies.* vol. 1 219–242 (2016).
10. Kulcsár, G. *The beginnings of the Bronze Age in the Carpathian Basin: the Makó-Kosihy-Čaka and the Somogyvár-Vinkovci cultures in Hungary*. (Archaeolingua, 2009).
11. Kiss, V. The Bronze Age burial from Balatonakali revisited. in *Objects, Ideas and Travelers. Contacts between the Balkans, the Aegean and Western Anatolia during the Bronze Age and Early Iron Age. Conference to the Memory of Alexandru Vulpe.* 553–568 (Deutsche Nationalbibliothek, 2020).
12. Hajdu, T., György-Toronyi, A., Pap, I., Rosendahl, W. & Szabó, G. The chronology and meaning of the Transdanubian encrusted pottery decoration. *Praehistorische Z.* **91**, 353–368 (2016).
13. Preda-Balanica, B., Frinculeasa, A. & Heyd, V. The Yamnaya Impact North of the Lower Danube - A Tale of Newcomers and Locals. *Bull. Société Préhistorique Francaise* **117**, 85–101 (2020).
14. Köhler, K. The Anthropological Remains from the Budakalász Cemetery. in *The Copper Age Cemetery of Budakalász* 303–364 (Pytheas, 2009).
15. Köhler, K. A harang alakú edények népe Szigetszentmiklós-Felső Ürge hegyi dűlő lelőhelyének embertani vizsgálati eredményei (Anthropological examination of the Bell Beaker cemetery at Szigetszentmiklós-Felső-Ürge-hegyi dűlő). *Anthr. Közlemények* **52**, 55–76 (2011).
16. Köhler, K. *et al.* Possible cases of leprosy from the Late Copper Age (3780-3650 cal BC) in Hungary. *PLOS ONE* **12**, e0185966 (2017).
17. Marciniak, S. *et al.* An integrative skeletal and paleogenomic analysis of prehistoric stature variation suggests relatively reduced health for early European farmers. *BioRxiv Prepr.* (2021) doi:https://doi.org/10.1101/2021.03.31.437881.
18. Somogyi, K. A kisapostagi kultúra birituális temetője Ordacsehi-Csereföldön – Das birituelle Gräberfeld der Kisapostag-Kultur on Ordacsehi-Csereföld. in *Őskoros Kutatók III. Összejövetelének konferenciakötete* vol. ΜΩΜΩΣ 349–381 (2004).
19. Kiss, V. *Middle Bronze Age encrusted pottery in western Hungary*. (Archaeolingua, 2012).
20. Kiss, V. Recent data on chronology, distribution, and connections of Kisapostag, Transdanubian Encrusted Pottery and Litzenkeramik. in *KEĎ BRONZ VYSTRIEDAL MEĎ* 27–38 (Archaeologica

Slovaca Monographiae, 2015).
21. Bóna, I. Geschichte der frühen und mittleren Bronzezeit in Ungarn und im mittleren Donauraum. in *Annales Universitatis Scientiarum Budapestinensis de Rolando Eötvös nominatae III-IV* vol. 3 3–22 (ELTE, 1961).
22. Bóna, I. Die mittlere Bronzezeit Ungarns und ihre südöstlichen Beziehungen. *Archaeol. Hung.* **49**, 73–76, 229–230 (1975).
23. Torma, I. A balatonakali bronzkori sír (Das bronzezeitliche Grab inBalatonakali). *Veszprémi Megyei Múzeumok Közleményei* **13**, 15–24 (1978).
24. Kiss, V. *et al.* Chronology of the Early and Middle Bronze Age in Hungary: New results. in *Reinecke´s Heritage. Terminology, Chronology and Identity in Central Europe, 2300 and 1600 BC.* vol. 2 173–197 (2019).
25. Kiss, V. The Life Cycle of Middle Bronze Age Bronze Artefacts from the Western Part of the Carpathian Basin. in *Metals and Societies. Studies in honour of Barbara S. Ottaway.* vol. 169 328–335 (2009).
26. Kiss, V. Transformations of Metal Supply during the Bronze Age in the Carpathian Basin. *Hung. Hist. Rev.* **9**, 315–330 (2020).
27. Fábián S. Õskori tömegsír Balatonkeresztúr-Réti-dûlõ lelõhelyrõl. **17**, 79–88 (2006).
28. Köhler K. Õskori tömegsír embertani leletei Balatonkeresztúrról. **17**, 8 (2006).
29. Müller-Scheeßel, N. *‚Irreguläre' Bestattungen in der Urgeschichte: Norm, Ritual, Strafe …? / Akten der Internationalen Tagung in Frankfurt a. M. vom 3. bis 5. Februar 2012.* vol. 19 (Deutsche Nationalbibliothek, 2013).
30. Earle, T., Kulcsár, G., Kiss, V., Serlegi, G. & Szeverényi, V. RECENT RESULTS FROM THE BRONZE AGE RESEARCHINTO THE BENTA VALLEY. *Hung. Archaeol. E-J.* **2014**, (2014).
31. K. Zoffmann, Z. Anthropological sketch of the prehistoric population of the Carpathian Basin. *Acta Biol. Szeged.* **44**, 75–79 (2000).
32. Acsádi, G. & Nemeskéri, J. *History of human life span and mortality*. (Akadémiai Kiadó, 1970).
33. Stloukal, M. & Hanáková, H. Die Lange der Langsknochen altslawischer Bevölkerungen unter besonderer Berücksichtigung von Wachstumsfragen. *Homo* **29**, 53–69 (1978).
34. Fazekas, I. G. & Kósa, F. *Forensic fetal osteology*. (Akadémiai Kiadó, 1978).
35. Ferembach, D., Schwidetzky, I. & Stloukal, M. Empfehlungen für die Alters- und Geschlechtsdiagnose am Skelett. *Homo* **30**, 1–32 (1979).
36. Işcan, M., Loth, S. & Wright, R. Age estimation from the rib by phase analysis: white males. *J. Forensic Sci.* **29**, 1094–104 (1984).
37. Işcan, M., Loth, S. & Wright, R. Age estimation from the rib by phase analysis: white females. *J. Forensic Sci.* **30**, 853–63 (1985).
38. Ubelaker, D. H. *Human skeletal remains, excavation, analysis, interpretation*. (Taraxacum, 1989).
39. Bernert, Z., Évinger, S. & Hajdu, T. New data on the biological age estimation of children using bone measurements based on historical populations from the Carpathian Basin. *Ann. Hist.-Nat. MUSEI Natl. Hung.* **99**, 199–206 (2007).
40. Manchester, K. *The archaeology of disease*. (University of Bradford Press, 1983).
41. Aufderheide, A. C. & Rodríguez-Martín, C. *The Cambridge encyclopedia of human paleopathology*. (Cambridge University Press, 1998).
42. Ortner, D. *Identification of Pathological Conditions in Human Skeletal Remains*. (London-New York Academic Press, 2003).
43. Molnár, M. *et al.* EnvironMICADAS: A Mini ^{14}C AMS with Enhanced Gas Ion Source Interface in the Hertelendi Laboratory of Environmental Studies (HEKAL), Hungary. *Radiocarbon* **55**, 338–344 (2013).
44. Molnár, M. *et al.* Status Report of the New AMS ^{14}C Sample Preparation Lab of the Hertelendi Laboratory of Environmental Studies (Debrecen, Hungary). *Radiocarbon* **55**, 665–676 (2013).
45. Bronk Ramsey, C. Bayesian Analysis of Radiocarbon Dates. *Radiocarbon* **51**, 337–360 (2009).
46. Reimer, P. J. *et al.* The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP). *Radiocarbon* **62**, 725–757 (2020).

47. Alt, K. W. *et al.* Lombards on the Move – An Integrative Study of the Migration Period Cemetery at Szólád, Hungary. *PLoS ONE* **9**, e110793 (2014).
48. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
49. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
50. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
51. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).
52. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
53. Reich, D. AADR - Allen Ancient DNA Resource. https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data (2021).
54. Gouy, M., Guindon, S. & Gascuel, O. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol. Biol. Evol.* **27**, 221–224 (2010).
55. Mathieson, I. *et al.* The genomic history of southeastern Europe. *Nature* **555**, 197–203 (2018).
56. Suchard, M. A. *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, (2018).
57. Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E. & Lercher, M. J. PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses in R. *Mol. Biol. Evol.* **31**, 1929–1936 (2014).
58. Ralf, A., Montiel González, D., Zhong, K. & Kayser, M. Yleaf: Software for Human Y-Chromosomal Haplogroup Inference from Next-Generation Sequencing Data. *Mol. Biol. Evol.* **35**, 1291–1294 (2018).
59. Y-DNA Haplogroup Tree 2019-2020. *International Society of Genetic Genealogy* https://isogg.org/tree/ (2019).
60. Bandelt, H. J., Forster, P. & Rohl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).
61. Network Software. *Fluxus-Engineering* https://www.fluxus-engineering.com/sharepub.htm#a1 (2008).
62. Rivollat, M. *et al.* Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers. *Sci. Adv.* **6**, eaaz5344 (2020).
63. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
64. Malmström, H. *et al.* The genomic ancestry of the Scandinavian Battle Axe Culture people and their relation to the broader Corded Ware horizon. *Proc. R. Soc. B Biol. Sci.* **286**, 20191528 (2019).
65. Sánchez-Quinto, F. *et al.* Megalithic tombs in western and northern Neolithic Europe were linked to a kindred society. *Proc. Natl. Acad. Sci.* **116**, 9469–9474 (2019).
66. Lipson, M. *et al.* Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* **551**, 368–372 (2017).
67. Amorim, C. E. G. *et al.* Understanding 6th-century barbarian social organization and migration through paleogenomics. *Nat. Commun.* **9**, 3547 (2018).
68. Olalde, I. *et al.* The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**, 190–196 (2018).
69. Mittnik, A. *et al.* Kinship-based social inequality in Bronze Age Europe. *Science* **366**, 731–734 (2019).
70. Brück, J. & Frieman, C. J. Making kin: The archaeology and genetics of human relationships. *TATuP* **30**, 47–52 (2021).
71. Monroy Kuhn, J. M., Jakobsson, M. & Günther, T. Estimating genetic kin relationships in

prehistoric populations. *PLOS ONE* **13**, e0195491 (2018).
72. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
73. Wang, J. Estimating pairwise relatedness in a small sample of individuals. *Heredity* **119**, 302–313 (2017).
74. Rebay-Salisbury, K. Bronze Age Beginnings: The Conceptualization of Motherhood in Prehistoric Europe. in *Motherhood in Antiquity* (eds. Cooper, D. & Phelan, C.) 169–196 (Springer International Publishing, 2017). doi:10.1007/978-3-319-48902-5_9.
75. Žegarac, A. *et al.* Ancient genomes provide insights into family structure and the heredity of social status in the early Bronze Age of southeastern Europe. *Sci. Rep.* **11**, 10072 (2021).
76. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
77. Chaitanya, L. *et al.* The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. *Forensic Sci. Int. Genet.* **35**, 123–135 (2018).
78. Cariaso, M. & Lennon, G. SNPedia: a wiki supporting personal genome annotation, interpretation and analysis. *Nucleic Acids Res.* **40**, D1308–D1312 (2012).
79. Reich, D. *Who we are and how we got there: Ancient DNA and the new science of the human past*. (Oxford University Press, 2018).
80. Landrum, M. J. *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).
81. Altmann, T. & Gennery, A. R. DNA ligase IV syndrome; a review. *Orphanet J. Rare Dis.* **11**, 137 (2016).
82. ClinVar. *NCBI* https://www.ncbi.nlm.nih.gov/clinvar/ (2021).
83. Shribman, S., Reid, E., Crosby, A. H., Houlden, H. & Warner, T. T. Hereditary spastic paraplegia: from diagnosis to emerging therapeutic approaches. *Lancet Neurol.* **18**, 1136–1146 (2019).
84. Koppers, M. *et al.* VCP mutations in familial and sporadic amyotrophic lateral sclerosis. *Neurobiol. Aging* **33**, 837.e7-837.e13 (2012).
85. IJlst, L. *et al.* 3-Methylglutaconic Aciduria Type I Is Caused by Mutations in AUH. *Am. J. Hum. Genet.* **71**, 1463–1466 (2002).
86. Dandy-Walker syndrome. *MedlinePlus Genetics* https://medlineplus.gov/genetics/condition/dandy-walker-malformation/ (2021).
87. Gerasimov, M. M. Osnovy vosstanovleniya litsa po cherepu. [Bases of the craniofacial reconstruction.]. in 185 pp. (Nauka, 1949).
88. Prag, J. & Neave, R. Making faces: Using forensic and archaeological evidence. in 256 pp. (1997).
89. Taylor, K. T. *Forensic art and illustration*. (CRC Press, 2001).
90. Sjøvold, T. Árpás anatomical method for face reconstruction. in 203–204 (1981).
91. Kustár, Á. & Skultéty, G. A benepusztai honfoglaláskori férfi arcrekonstrukciója [Craniofacial reconstruction of a Hungarian conqueror male from the site of Benepuszta]. *Savaria* **22**, 179–190 (1996).
92. Röhrer-Ertl, O. & Helmer, R. Zu Stand und Möglichkeiten der Erneut modifizierten Kollmann–Methode. (Gesichtsrekonstruktion aufgrund des Schädels.). in vol. 3 369–373 (1984).
93. Rynn, C., Wilkinson, C. M. & Peters, H. L. Prediction of nasal morphology from the skull. *Forensic Sci. Med. Pathol.* **6**, 20–34 (2010).
94. *Metahuman Creator*. (2021).
95. Patterson, N., Price, A. L. & Reich, D. Population Structure and Eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
96. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
97. Antonio, M. L. *et al.* Ancient Rome: A genetic crossroads of Europe and the Mediterranean. *Science* **366**, 708–714 (2019).
98. Mittnik, A. *et al.* The genetic prehistory of the Baltic Sea region. *Nat. Commun.* **9**, 442 (2018).
99. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**,

499–503 (2015).
100. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
101. Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
102. Günther, T. *et al.* Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLOS Biol.* **16**, e2003703 (2018).
103. Brace, S. *et al.* Ancient genomes indicate population replacement in Early Neolithic Britain. *Nat. Ecol. Evol.* **3**, 765–771 (2019).
104. Feldman, M. *et al.* Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia. *Nat. Commun.* **10**, 1218 (2019).
105. Narasimhan, V. M. *et al.* The formation of human populations in South and Central Asia. *Science* **365**, eaat7487 (2019).
106. Petr, M., Vernot, B. & Kelso, J. admixr—R package for reproducible analyses using ADMIXTOOLS. *Bioinformatics* **35**, 3194–3195 (2019).
107. *Tidyverse packages*. (2021).
108. Patterson, N. *et al.* Ancient Admixture in Human History. *Genetics* **192**, 1065–1093 (2012).
109. Cassidy, L. M. *et al.* A dynastic elite in monumental Neolithic society. *Nature* **582**, 384–388 (2020).
110. Fernandes, D. M. *et al.* A genomic Neolithic time transect of hunter-farmer admixture in central Poland. *Sci. Rep.* **8**, 14879 (2018).
111. *Plotly graphing packages*. (2021).
112. *Heatmaply package*. (2021).
113. *R Color Brewer's palettes*. (2021).
114. Galili, T., O'Callaghan, A., Sidi, J. & Sievert, C. heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics* **34**, 1600–1602 (2018).
115. Linderholm, A. *et al.* Corded Ware cultural complexity uncovered using genomic and isotopic analysis from south-eastern Poland. *Sci. Rep.* **10**, 6885 (2020).
116. Freilich, S. *et al.* Reconstructing genetic histories and social organisation in Neolithic and Bronze Age Croatia. *Sci. Rep.* **11**, 16729 (2021).
117. Schubert, M. *et al.* Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat. Protoc.* **9**, 1056–1082 (2014).
118. Peltzer, A. *et al.* EAGER: efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).
119. Poullet, M. & Orlando, L. Assessing DNA Sequence Alignment Methods for Characterizing Ancient Genomes and Methylomes. *Front. Ecol. Evol.* **8**, 105 (2020).