Interdisciplinary analyses of Bronze Age communities from Western Hungary reveal complex population histories

Dániel Gerber[1,2,3], Bea Szeifert[1,2,3], Orsolya Székely[1], Balázs Egyed[2], Balázs Gyuris[1,2,3], Julia I. Giblin[4], Anikó Horváth[5], László Palcsu[5], Kitti Köhler[6], Gabriella Kulcsár[6], Ágnes Kustár[7], Vajk Szeverényi[8], Szilvia Fábián[9], Balázs Gusztáv Mende[1], Mária Bondár[6], Eszter Ari[2,10,11,*], Viktória Kiss[6,*], Anna Szécsényi-Nagy[1*]

1) Institute of Archaeogenomics, Research Centre for the Humanities, Eötvös Loránd Research Network (ELKH); Tóth Kálmán utca 4., 1097 Budapest, Hungary 2) Department of Genetics, ELTE Eötvös Loránd University; Pázmány Péter sétány 1/C, 1117 Budapest, Hungary 3) Doctoral School of Biology, Institute of Biology, ELTE Eötvös Loránd University, Pázmány Péter sétány 1/C. 1117 Budapest, Hungary 4) Department of Sociology, Criminal Justice and Anthropology, Quinnipiac University; 275 Mount Carmel Avenue, Hamden, CT 06518, USA 5) Institute for Nuclear Research, ICER Centre; Bem tér 18/C, 4026 Debrecen, Hungary 6) Institute of Archaeology, Research Centre for the Humanities, Eötvös Loránd Research Network (ELKH); Tóth Kálmán utca 4., 1097 Budapest, Hungary 7) freelancer anthropologist, 1028 Budapest, Hungary 8) Déri Museum; Déri tér 1., 4026 Debrecen, Hungary 9) Hungarian National Museum; Múzeum krt. 14-16., 1088 Budapest, Hungary 10) HCEMM-BRC Metabolic Systems Biology Lab; Temesvári krt. 62., 6726 Szeged, Hungary 11) Synthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Centre, Eötvös Loránd Research Network (ELKH); Temesvári krt. 62, 6726 Szeged, Hungary
*These authors jointly supervised this work.

Correspondence to: Eszter Ari (arieszter@gmail.com), Viktória Kiss (kiss.viktoria@abtk.hu), Anna Szécsényi-Nagy (szecsenyi-nagy.anna@abtk.hu)

Keywords: mass grave, facial reconstruction, Bronze Age, Central Europe, hunter-gatherers, PAPline

Abstract

In this study we report 20 ancient shotgun genomes from present-day Western Hungary (3530 – 1620 cal BCE), mainly from previously understudied Baden, Somogyvár-Vinkovci, Kisapostag, and Encrusted Pottery archaeological cultures. Besides analysing archaeological, anthropological and genetic data, ^{14}C and strontium isotope measurements complemented reconstructing the dynamics of the communities discovered at the site Balatonkeresztúr. Our results indicate the appearance of an outstandingly high Mesolithic hunter-gatherer ancestry in the largest proportion (up to ~46%) among Kisapostag associated individuals, despite this component being thought to be highly diluted by the Early Bronze Age. We show that hunter-gatherer ancestry was likely derived from a previously unrecognised source in Eastern Europe that contributed mostly to prehistoric populations in Central Europe and the Baltic region. We revealed a patrilocal residence system and local female exogamy for this Kisapostag population that was also the genetic basis of the succeeding community of the Encrusted Pottery culture, represented by a mass grave that likely resulted from an epidemic. We also created a bioinformatic pipeline dedicated for archaeogenetic data processing. By developing and applying analytical methods for analysing genetic variants we found carriers of aneuploidy and inheritable genetic diseases. Furthermore, based on genetic and anthropological data, we present here the first female facial reconstruction from the Bronze Age Carpathian Basin.

Significance

Here we present a genomic time transect study from the Carpathian Basin (3530 – 1620 cal BCE), that sheds light on local and interregional population processes. We not only discovered long-distance mobility to provide detailed analysis of yet understudied Bronze Age communities, but we also recovered a previously hidden remnant hunter-gatherer genetic ancestry and its contribution to various populations in Eastern and

Central Europe. We integrated ^{14}C and strontium isotope measurements to the interdisciplinary interpretation of a site with 19 individuals analysed, where patrilocal social organisation and several health-related genetic traits were detected. Furthermore, we developed new methods and method standards for computational analyses of archaic DNA, implemented to our newly developed and freely available bioinformatic pipeline.

Introduction

A number of studies addressed population historical questions in Prehistoric Europe by recovering major events connected to the pre-Neolithic hunter-gatherers (HG)[1–3], their assimilation to early European farmers during the Neolithic era[2,4–6], and the appearance, expansion and admixture of steppe ancestry during the Eneolithic / Late Copper Age to the dawn of Early Bronze Age[4,7,8]. While some of these studies are essential for understanding the foundation of the European gene pool, studies are sparse in the literature that uncover regional interactions or social stratification via kinship[9–11]. Additionally, except for a few well-known markers in most archaic studies – *e.g.* basic pigmentation markers or lactose intolerance analysed large-scale in Mathieson et al. 2015[12] – no deeper analyses have been made e.g. on clinical variants. Our study aimed to make a transect analysis on a single site concerning understudied archaeological assemblies, as well as introducing the *PAPline* (Performing Archaeogenetic Pipeline, Supplementary Information section 6), a new bioinformatic pipeline for archaic DNA analysis. We analysed the archaeological finds from Balatonkeresztúr-Réti-dűlő site in Western Hungary, where - among others - Bronze Age assemblies were found during roadwork in 2003. Three Bronze Age archaeological horizons can be distinguished, from the Somogyvár-Vinkovci culture (~2500-2200 BCE), Kisapostag culture (~2200–1900 BCE) and to the Encrusted pottery culture (~1900–1450 BCE) that were named into Bk-I, II and III phases in this study, respectively (Table 1, Supplementary Information section 1, and Fig. S.1.2.1). In order to provide additional proxy to population ancestry of the region one further Late Copper Age individual from a multiple grave of the Baden culture (3600-2800 BCE) excavated at site Balatonlelle-Rádpuszta, ~30 km away from Balatonkeresztúr was added to our dataset.

Results

We shotgun sequenced genomes of 20 individuals with 0.008 to 0.17x coverage. We also sequenced reads of a capture set consisting 3000 nuclear SNPs (single nucleotide polymorphisms), and whole mitochondrial DNAs (mtDNAs) of all individuals. The shotgun and the capture sequenced samples ultimately resulted in an average ~104k SNPs/individuals using the 1240k SNP panel for genotype calling[12], see Materials and Methods and Supplementary Tables S4 and S7. We utilised STR analysis of the Y chromosome to recover paternal kinship patterns. Furthermore, we reconstructed the face of individual S13 (Bk-II), where all known biological and archaeological details were considered, see Supplementary Information section 4. The bioarchaeological analyses were completed with radiocarbon and strontium isotope analyses, the latter can be used to trace individual mobility.

Archaeological and anthropological evaluation of samples

Bk-I contained the remains of a single male individual having a very long (ultradolichocran) skull type which differentiates it from most individuals found at Bk-II and Bk-III that have a very short (brachycranic) skull type[13] (Table 1). In Bk-II and Bk-III male dominance (~78%) suggest distinctive funeral treatment for males and females. Bk-II is represented by 3 juveniles (16-19 years olds) and 7 adults (30+ years olds) distributed into two grave groups of A and B (Table 1, Supplementary Information Fig. S.1.2.1), and one child grave (individual S10) far from the others. Most of the burials contained no remaining grave goods except for small copper jewellery in S10 and S13. Radiocarbon dates place these inhumations to ca. 2200-1770 cal BCE, however, with Bayesian analysis using the OxCal software the timespan of the Bk-II burials can be reduced to ca. 2120-1900 cal BCE (95.4% CI), with two graves (individuals S10 and S11) possibly being slightly earlier (Supplementary Information section 1.8). The absence of children from the site is a common phenomenon that can be traced back to different preservation dynamics or burial practises to adults[14], while

the reason for the absence of young adults (~20-30 year olds) is unknown. Bk-III is represented by a single mass grave of 8 skeletal remains of all ages that turned out to be an unusual find in a period when the cremation practises and single inhumations were common, from ca. 1870-1620 cal BCE (95.4% CI). For details, see Supplementary Information section 1.

Uniparental genetics and kinship analyses

Both Bk-II and Bk-III show phylogeographic signals for their maternal and paternal lineages. Accordingly, Bk-II is mostly defined by mtDNA connections to the region of present-day Poland and its surroundings, whereas Bk-III has more diverse maternal composition, see Supplementary Information section 2.1. Male lineages in both Bk-II and Bk-III are mostly defined by Y chromosome haplogroup I2a-L1229 (Table 1), for which network analysis (Supplementary Information section 2.2) narrowed down regional affinities to the North European plain and shows continuity between these two horizons. Uniparental diversity makeup points to a patriarchal social structure similar to previously reported Bronze Age findings[9,11,15]. Results are highly similar to previous observations on Encrusted Pottery culture at the Jagodnjak site (Croatia)[10]. The kinship network of Bk-II follows the distribution of grave groups (Fig. 1) which were likely established along family relationships and chronology. Individuals buried in the Bk-III mass grave only show a few blood relations, like a half-brother and father-son and a dizygotic twin, the latter is the most archaic detection to date to our knowledge. However, Bk-III as an extended family group can not be excluded. For further details, see Supplementary Information section 2 and Supplementary Tables S1-S3.

Table 1 Summary of the investigated samples. MtDNA and ChrY denote mitochondrial haplogroup and Y chromosome haplogroup. In column "Kinship" 1st and 2nd mean the degree relations. For the feature, grave IDs and details on newly reported radiocarbon dates see the Supplementary Table S1.

Group	ID	Grave group	^{14}C (cal BCE) 2σ (95.4% CI)	Age	Sex	MtDNA	ChrY	Kinship
Baden	BAD002		3530-3370	8-9	M	K1a4a1	I-M170	None
Bk-I: Somogyvár - Vinkovci culture	S9		2560-2290	35-40	M	K1a3a	R1a-V2670	None
Bk-II: Kisapostag or Early Encrusted Pottery culture	S1	A	2120-1880	40+	M	V	I2a-L1229	2nd to S2 & S8
	S2	A	2120-1880	30-35	M	U5a2b1a	I2a-L1229	2nd to S1 & S8
	S4	B		17-19	M	H10a1	I2a-L1229	1st to S8
	S5	A		16-18	M	T1a4	I2a-L1229	1st to S6 & S11
	S6	A	2030-1770	17-18	M	T1a4	I2a-L1229	1st to S5 & S11
	S7	A	2120-1880	35-50	F	V		2nd to S8
	S8	B		30-40	M	T2b	I2a-L1229	1st to S4; 2nd to S1 & S2 & S7
	S10		2140-1940	7-8	M	K1a4a1g	I2a-L1229	None
	S11	B	2200-1980	34-43	M	T2b	I2a-L1229	1st to S5 & S6
	S13	B	2120-1890	35-45	F	J2b1		None
Bk-III: Transdanubian Encrusted Pottery culture	S14	Mass Grave B-938		7-8	F	H10a1		None
	S15			21-23	M	U4b1b1	I2a-L1229	2nd to S17
	S16		1890-1640	35-44	M	T2g2	I2a-L1229	None
	S17		1870-1540	26-35	M	U5b1b1+@16192	I2a-L1229	1st to S19; 2nd to S15
	S18			3-4	M	U4a2	R1b-Z2103	None
	S19			9-10	M	T2b	I2a-L1229	1st to S17
	S20			1.5-2	M	K1a+195	R1b-Z2103	1st to S21
	S21			1.5-2	F	K1a+195		1st to S20

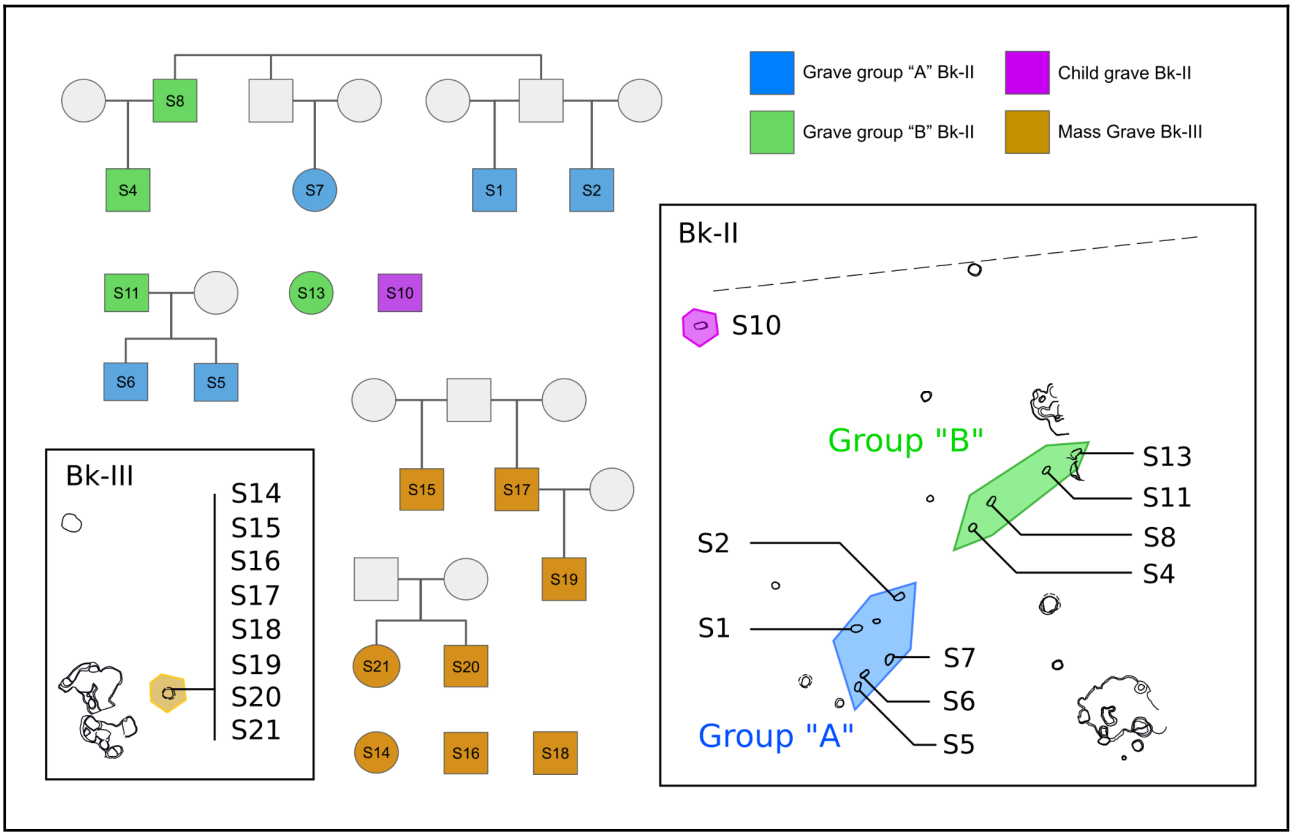

Fig. 1 Kinship network at Balatonkeresztúr site based on the biological age of the individuals, and the results of the uniparental and and the READ/MPMR analyses. Blue colour represents the Bk-II grave group "A" that consists of descendant individuals, while the green coloured individuals – found in Bk-II grave group "B"– are mostly the ancestors, which suggests a kinship or chronological based geographical distribution of graves. In the mass grave Bk-III (orange, east to Bk-II graves, for full site map see Supplementary Information Fig. S.1.2.1) only a partial kinship network can be observed. Specimens buried in Bk-II S10 (purple) and S13 (green), and Bk-III S14, S16 and S18 (orange) do not have 1st and 2nd degree relatives in the uncovered graves.

Genetic disorders and pigmentation

Investigating genetic disorders in archaic datasets is potentially valuable for history of health and medicine, and also highlights the overall genetic health of past populations. Inherited genetic disorders, if accompanied with severe phenotypic anomalies, could also explain unusual burial practises, as it was described in cases of dwarfism[16]. For detailed results of this topic, see Supplementary Information section 3.

Aneuploidies

The abnormal number of chromosomes result in a few well known diseases which we tested thoroughly. We found one individual, S10, the only child burial in Bk-II, with XYY gonosomal genotype, described as Jacob's syndrome. Although this syndrome remains in most cases silent as it is relatively frequent (~0.1%) in today's populations, it occasionally comes with a wide scale of symptoms[17], which may be linked to its separate burial, but due to poor bone preservation for S10 no further assessments could be made.

Mitochondrial DNA diseases

We examined the clinical significances of the polymorphisms that can be found in the mtDNA by using the *mitopathotool* software on the *AmtDB* database[18], and found that individual S1 (40+ years old) from Bk-II had one of the defining mutations (T14484C) of Leber's hereditary optic neuropathy (LHON) causing complete

vision loss in ~10% of females and in ~50% of males between 20-40 years of age, rarely accompanying with other neuropathies[19].

Pigmentation

Pigmentation patterns highly differ between horizons, as Bk-I mostly possess variants for light pigmentation, blue eyes and blonde hair, while Bk-II is more similar to populations of Neolithic Europe (Fig. 2), although some variants for lighter pigmentation exist within this group too. Members of Bk-III on the other hand show a wide range from dark to light tones and even the presence of variants for red hair (Supplementary Table S5, Supplementary Information section 3.2.1).

Fig. 2 Reconstruction of individual S13. Her mouth was partly open due to maxillary prognathia and her burial position differs from the others by her unusual arm position. She likely had higher social status for the rare copper beads she had around her head. Her genomic makeup and pigmentation pattern blends well to other Bk-II individuals, and while she did not have any blood relatives at the site up to second degree, according to strontium isotope data she lived in the region, suggesting her origin from a nearby community of the same population. For more information, see Supplementary Information sections 1 and 4.

Nuclear variants with clinical significance

We also examined the nuclear genomes to find regions with clinical significance. Since a complete panel for determining disease susceptibility only exists in commercial DNA kits, and detailed description for the 1240k SNP set is not available, we created our own SNP calling panel (included in *PAPline*) focusing on various conditions including amyotrophic lateral sclerosis, Alzheimer disease, autism, Crohn's disease, diabetes, lactose intolerance, tumor markers, mental disorders, Parkinson disease, schizophrenia and ulcerative colitis. For this study we used 3,874 clinically significant SNPs, which were marked as "pathogenic" or "likely pathogenic" in the ClinVar database[20], by ignoring deletion, duplication and copy number variants, as well as SNPs with questionable ("reported", "conflicting reports", *etc.*) contribution to diseases. From this set we found ~2,200 SNPs which covered at least one individual, out of these 27 positions showed clinically relevant substitutions. However, test runs on database data resulted in numerous positive hits for pathogenic variants most likely related to DNA damage, which highlights the unreliability of low coverage SNP data for variant identification (Supplementary Information section 3.2.2). To overcome this issue, we considered

positions with more than 0.99 genotype likelihood (GL) values calculated using ANGSD v0.931[21] (Supplementary Table S6) or when skeletal features supported results. A variant for Lig4 syndrome (rs104894421[22], GL=0.999) in individual S15 was detected, and some of his skeletal features (*e.g.* congenital hip dysplasia) show possible onset of symptoms (Supplementary Information section 3.2.2.1). In another case the physical manifestation of hereditary spastic paraplegia is likely for S11 and S6, father and son but the genotype likelihood is lower (0.67; see Supplementary Information section 3.2.2.3). Interestingly, a tumor marker on the BRCA2 gene (rs80358920, GL=0.999) in individual S9 is nowadays only prevalent in Asian populations[22]. For further discussions, see Supplementary Information section 3.2.

Whole genome composition and genetic ancestry

Balatonkeresztúr site samples

According to the principal component analysis (PCA) based on 590k nuclear SNPs (Fig. 3 a.) Bk-I is clearly separated from Bk-II and Bk-III, where Bk-II has a strong shift towards HG samples[23] overlapping with only a fraction of known archaic samples[23] and Bk-III. Admixture and qpAdm analyses for assessing genetic components (Supplementary Tables S9, S12-16; Supplementary Information sections 5.2 and 5.5.2) revealed ~17% HG, ~40% European farmer, and ~43% steppe ancestry for Bk-I, similar to average Bronze Age Europeans. Bk-I is most likely derived from a single source that is genetically related to a Poland Southeast Bell Beaker culture (BBC) associated population (p = 0.784) in line with archaeological observations[24]. Bk-II comprises a unique makeup of ~42% HG, ~41% European farmer, and ~17% steppe ancestries. qpAdm analysis revealed most plausible sources of Neolithic Sweden Funnel Beaker culture (~32±8%), Poland Southeast BBC (~41±6%) and an extra HG (~29±3%) ancestry of yet unknown origin. Despite the lower supported models Bk-I can not be excluded as an ancestry component for Bk-II, while the affinity of Sweden Funnel Beaker culture associated population likely reflects a more closely related group, such as population related to the Poland Globular Amphora culture, see Supplementary Information section 5.5.2.2 and Supplementary Tables S12 and S15. Bk-III shows a shift in ancestry composition (~29% HG, ~46% European farmer, ~25% steppe). qpAdm analyses revealed that the main ancestry component for Bk-III is Bk-II (~53±5%), while “dilution” of Bk-II to Bk-III is mostly driven by population events that are yet to be uncovered.

Genetic outliers from previous studies

Many samples were defined as genetic outliers in their genetic context by previous studies from Bronze Age Europe. We selected such outlier individuals with high HG ancestry components to assess whether they are related to Bk-II. Selection was based on previous observations and also by using Dixon’s *Q*-test[25] at 90% confidence interval on HG component’s upper deviation using results of the Admixture analysis. First, we ran *f4*-statistics in the form of f4(W=test outlier, X=corresponding population, Y=Bk-II, Z=Yoruba)[26]. This test resulted in positive values for some outliers (W) meaning that these are genetically closer to Bk-II (Y) than its presumed population (X). However, *Z*-scores are low in many cases, and false positives may appear solely by high HG component, not by true relationship (see Supplementary Table S10 and Supplementary Information section 5.3). To check true relationship between Bk-II and groups/samples with high HG ancestry, we performed an outgroup *f3*-statistics in the form of f3(X=Bk-II, Y=test HG-s, Z=Yoruba) for all relevant archaic populations and outliers[23] (Supplementary Table S11) that resulted a table of allele frequency based distances between test pairs Xs and Ys. Euclidean clustering based on the results of f3-statistics revealed that a number of samples and even three populations from the Baltic (Fig. 3 b, Supplementary Information section 5.4) from AADR[23] form a cluster with Bk-II and Bk-III, suggesting actual genetic connection via a common HG ancestry source.

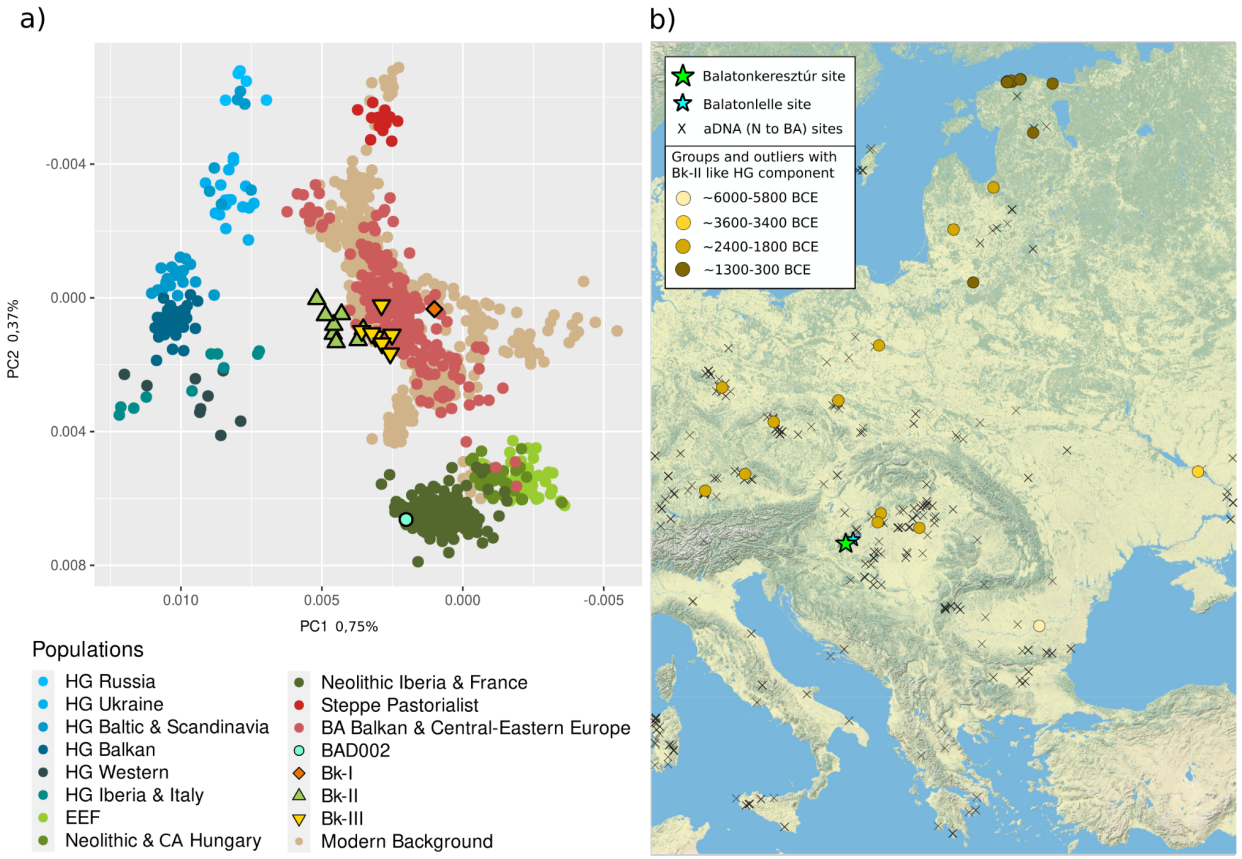

Fig. 3 a) Principal Component Analysis based on 590k SNPs calculated by the smartpca software. Bk-II clearly separated from any known archaic Central-Eastern European populations. b) Highlighted genetic outliers from previous studies (Supplementary Information section 5) that show a similar hunter-gatherer (HG) ancestry component to Bk-II among ancient datasets. The origin of the HG component likely lies between the Carpathian mountains and the Dnieper River or the Black Sea, from where it expanded further North and from where the Bk-II population most likely originated from. Outliers signalise a westward migration route of this component detected in Bk-II.

Origin of high HG ancestry component in BK-II and the Kisapostag population

f4-statistics in the form of f4(W=Sweden HG, X=Serbia IronGates HG, Y=Bk-II, Z=Yoruba) revealed that while Pitted Ware culture associated individuals have similarly high HG levels to Bk-II from Scandinavian HG-s, in Bk-II this type of HG component is significantly negative meaning no or minimal shared ancestry (Supplementary Information section 5.3). Results of outgroup in *f3*-statistics in the form of f3(X=Bk-II, Y=HG-s, Z=Yoruba) showed that HGs fit the best to Bk-II are geographically widely distributed (Supplementary Table S11), thus we conclude that the true source for this component is yet to be described. We can take into consideration the outgroup *f3*-statistics results, chronology, timing of HG admixture according to Freilich 2021[10] between ~3400-2400 BCE, qpAdm results and the geographical distribution of groups and outliers of similar HG makeup. These suggest a dated migration pattern for this undescribed population with dominant HG genetic ancestry from what is present-day Bulgaria to the Baltic through the Eastern borders of the Carpathians (Fig. 3 b). Ancestors of Bk-II likely branched off from this migration route and started to move towards West, by at least around ~2500 BCE, subsequently intermixing with various groups. Interestingly, the phylogeography of mitochondrial haplogroup U4b1b1 (individual S15) perfectly fits this scenario. Notably, strontium ($^{87}Sr/^{86}Sr$ ratio) isotope results from molars (Fig. 4, Supplementary Information section 1.9) indicate that every individual of this study was raised and lived at least close proximity to Balatonkeresztúr site in their childhood and early adolescence, suggesting that Bk-II group does not represent the first generation of the Kisapostag culture associated population in Transdanubia (Western

Hungary). Correspondingly, the high number of supported models in qpAdm results could be the result of subsequent local admixtures.

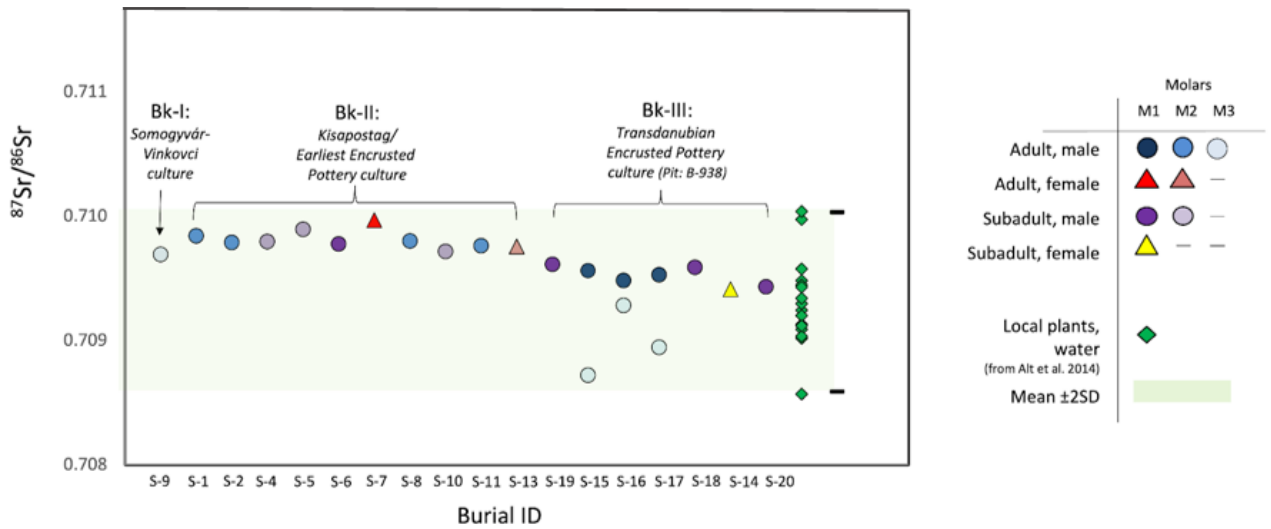

Fig. 4. Sr isotope data from the Balatonkeresztúr site. Samples were taken from dental enamel (first, second and third molars) to evaluate whether individuals were born in the area, or grew up in a geologically distinct region. All of the samples are consistent with previously published plant and water $^{87}Sr/^{86}Sr$ ratio (green diamonds) data collected from the southern portion of Lake Balaton[14]. For further data see Supplementary Information section 1.9.

A Late Copper Age outlier individual from Balatonlelle site

We included a Late Copper Age individual from Balatonlelle site to this study for its high HG genomic ancestry component. Mitogenome of BAD002 (K1a4a1) shows affinity to Iberian BBC associated individuals (Supplementary Information Fig. S.2.1.1), and his Y chromosomal haplogroup belongs to I-M170. On genomic PCA this sample groups with Iberian and French Neolithic individuals, due to higher HG component (~34%) than known Neolithic and Copper Age populations in the Carpathian Basin have, and due to the lack of steppe related ancestry. qpAdm estimates pointed to a source of Neolithic communities from present-day (Northwestern) France (87±8%)[27] with a minor extra HG component (13±8%) with p=0.948 (Supplementary Information section 5.5, Supplementary Tables S12 and S13). Pigmentation pattern of BAD002 shows resemblance to average Neolithic Europeans. The foreign cultural traits of the boy's jewellery is in line with his outlier genetic composition in the study region[28]. Notably, further tests (outgroup *f3*-statistics, qpAdm) excluded contribution nor any connection of BAD002 to Bk-II (Supplementary Tables S11, S12 and S13, Supplementary Information section 5). Therefore we conclude that this individual testifies large-scale migration in the Copper Age, providing research questions for future studies.

PAPline

We introduce our newly developed, freely available bioinformatic pipeline, named *PAPline* (Performing Archaeogenetic Pipeline), written in *linux bash, R,* and *python v3.8.10* programming languages. One can use this software primarily to analyse next generation sequencing data of archaeogenomic samples, supplemented by non-pipeline scripts. The main distinguishing feature of the *PAPline* compared to the EAGER[29] and the Paleomix[30] pipelines is its user friendliness, as its installation process is less complicated and it provides a graphical interface with almost complete automatisation with practical details. For detailed description visit the github page or see Supplementary Information section 6.

Discussion

The Carpathian Basin was inhabited by the Baden cultures' population at the end of the Copper Age, and their genetic composition was represented by an early farmer and a slightly increased HG genetic component, compared to the previous Neolithic populations of the region[6]. Here we demonstrated that in the early phase of this culture, a group of Western European origin appeared in Transdanubia, diversifying what we knew about the region's Late Copper Age substrate up to now.

The Carpathian Basin experienced the influx of steppe-related genetic ancestry at the dawn of the Bronze Age[5,8], and this transformation was already detectable at Balatonkeresztúr-Réti-dűlő site as well, where we could examine multiple populations. The earliest Bronze Age horizon Bk-I (representative of the Somogyvár-Vinkovci culture) shows similarities to Poland Southeast BBC associated population with high steppe ancestry that was replaced by the Kisapostag culture associated Bk-II around the 23-22th century BCE, while at least some sort of genetic ancestry of Bk-I in this population can not be excluded. According to our results, the Bk-II population had outstandingly high HG genetic ancestry levels, compared to other Bronze Age groups of the region, which can be traced back to today's Ukraine, Belarus, Moldavia or Romania, pointing to a long standing previously unsampled population with dominant HG ancestry. Calculated admixture dates[10] suggest the presence of a genetically pure or at least highly HG specific population in Eastern Europe as late as the end of the Copper Age. Part of this group subsequently admixed with populations of mainly steppe (likely Poland Southeast BBC) and early farmer (most likely a Globular Amphora culture related) ancestry during their westward migration on a Northern route, leaving genetic traces in Corded Ware culture, BBC, and other Bronze Age populations. The paternal lineage of BK-II was likely linked to the farmer component, as I2a-M223 (upstream to I2a-L1229) was a frequent paternal lineage among Globular Amphora culture and related populations. Looking for the possible source areas of the Kisapostag culture, a number of archaeological theories need to be considered. The pottery decoration technique originated either from Corded Ware in the Middle Dnieper region (Ukraine), epi-Corded Ware groups (northern Carpathians), e.g. Chłopice-Veselé or Nitra groups (Slovakia), the latter two is also supported by inhumation practises[31–36]. However, connections with the Litzenkeramik or Guntramsdorf-Drassburg group (eastern Austria, Slovenia, western Croatia) were also raised[37,38]. Pottery forms were connected to local development of communities with eastern (Makó–Kosihy–Čaka) or southern (Somogyvár–Vinkovci) origins, too[39]. BBC influence was also mentioned based on connections of pottery and craniometry data (so called *Glockenbecher* or brachycran skull type[40–42]). The results of this study fit best with the Middle Dnieper area origin of BK-II, especially when we consider individual I4110 from Dereivka I (Ukraine Eneolithic) as one of the earliest representatives of their genomic makeup.

Strontium isotope ($^{87}Sr/^{86}Sr$ ratio) data, representing through nutrition the bioavailable Sr in the area where people lived in a certain age interval, shows local values for both sexes in both Bk-II and Bk-III. These results push back the timing of their arrival a few generations, meaning that local or southern impact of cultural traits and maybe even genetic admixtures likely occurred during this short period as well, which also could explain the culture's archaeological heterogeneity.

The population of Bk-III was the direct descendant of Bk-II, forming not just cultural (Encrusted pottery) but also genetic continuity for at least ~500 years, even if the radiocarbon sequences allow a few decades of hiatus at the studied site. Continuous female-biased admixture with various groups occurred during this period according to our and previous genetic[10] and archaeological[31,43] evidence, diluting the BK-II genetic ancestry.

In both periods, the homogeneity of paternal lineages suggest a similar social organisation described in[9,10] of a patrilocal residence system. However, strontium isotope data shows local values for both sexes, which along with similar genomic makeup of females and males suggest exogamy most probably between villages of the same population. Two pairs of half-sibling graves in the two periods may indicate polygamy, although remarriage for high female mortality is more plausible. Notably, almost none of the uniparental markers are identical even at the haplogroup level with individuals from the Croatian Encrusted Pottery culture Jagodnjak

site, despite high similarities in cultural traits, social structure and genomic composition of the communities. This points to clan-like or patriarchal superfamily structure of Kisapostag and Encrusted Pottery groups. The relatively limited presence of female and children burials in both Bk-II and Bk-III periods may suggest distinctive treatment or another (here undiscovered) burial group for women and children at the same site. Although, in other Bronze Age cemeteries, *e.g.* Ordacsehi and Bonyhád in Hungary, males, females and children were buried close to each other, suggesting high variance of burial practises[34,35,44].

While low genomic coverage did not allow fine SNP recovery, we did find evidence for malignant variants within all of our tested groups, and undoubtedly showed the presence of LHON and Jacob's syndrome within Bk-II. Additionally, the disease panel we created can be extended and used in future studies, providing insight into past population health qualities.

Considering the unstructured age and kinship distribution in the mass grave Bk-III compared to Bk-II, the coetaneous death of eight people at least, the absence of traumatic or ritual events on bones, and non-cremated nature of the burial all signals a sudden tragic event in the Encrusted pottery period, most likely an epidemic, as first suggested based on the anthropological analyses[45]. Interestingly, comparative strontium isotope analyses on the first and third molar of the individuals in the BK-III mass grave indicate that subadult males - including a severely disabled individual (S15) with hip dysplasia - left their community for a while and then returned to their birthplace prior to their death, raising further questions for future studies on prehistoric lifeways and social organisations.

Materials and Methods

Isotope analyses

Radiocarbon dating was performed at the HEKAL AMS C-14 facility of the Institute for Nuclear Research in Debrecen, Hungary (see Supplementary Information section 1.8). $^{87}Sr/^{86}Sr$ isotope measurements were performed in the ICER Centre, Institute for Nuclear Research Debrecen, Hungary and at Quinnipiac and Yale University, Connecticut, USA (see Supplementary Information section 1.9).

Ancient DNA laboratory work

Petrous bones and teeth were taken from skulls for genetic investigation (Supplementary Table S1). Laboratory work was performed in a dedicated ancient DNA laboratory facility (Institute of Archaeogenomics, Research Centre for the Humanities, Eötvös Loránd Research Network, Budapest, Hungary). Each step was carried out in separate rooms under sterile conditions, during work protective clothing was used. Irradiated UV-C light, DNA-ExitusPlus™ (AppliChem) and/or bleach were applied for cleaning after and between work stages, and also, blank controls were utilised at all times.

Sample surfaces were cleaned by sandblasting and mechanically milled to powder. DNA extraction was performed according to Dabney et al. 2013[46] with minor changes according to Lipson et al. 2017[6]. DNA extraction success was verified by PCR using mtDNA primer pairs (F16209-R06348; F16045-R06240). Half-UDG treated libraries were used according to Rohland et al. 2015[47] with minor changes. Unique double internal barcode combinations were used for each library (Supplementary Table S1). Libraries were amplified with TwistAmp Basic (Twist DX Ltd) and purified with AMPure XP beads (Agilent). Then, concentration measurements were taken on Qubit 2.0 fluorometer, fragment sizes were checked on Agilent 4200 TapeStation System (Agilent High Sensitivity D1000 ScreenTape Assay).

Hybridisation capture method for mtDNA and 3k nuclear SNP was used besides whole genome shotgun, as described by Haak et al. 2015, Lipson et al. 2017 and Csáky et al. 2020[4,6,48]. Bait production was based on Fu et al. 2016[1] and N. Rohland's personal communication, the oligos as a pool was ordered from CustomArray Inc. Both for shotgun and capture libraries, universal iP5 and unique iP7 indexes were used.

Sequencing was done on Illumina MiSeq and NovaSeq platforms with custom setup and 150, 200 and 300 cycles, respectively.

Additionally, we investigated Y chromosome STR profiles (17 markers) with AmpFISTR® Yfiler® PCR Amplification Kit (Applied Biosystems), having one blank and one positive control at each reaction preparation. The workflow followed the recommended protocol except the PCR cycles were increased from 30 to 34 and reactions were halved in volume. Two repeats were done where at least 4 markers yielded results. Data analyses were carried out in GeneMapper® ID Software v3.2.1 (Applied Biosystems), results are summarised in Supplementary Table S3.

Bioinformatic analyses

Illumina sequencing reads were processed by the *PAPline*, for details, see Supplementary Information section 6. We used the GRCH37.p13 reference sequence for calling pseudohaploid genomes. For kinship inferences we applied the *READ* software[49] and a custom script (named *MPMR*, see Supplementary Information section 2.3 and Supplementary Table S2). MtDNA analyses included phylogenetic analyses using the *MrBayes* v3.2.6[50] and the *BEAST* v1.10.4[51] softwares and diversity tests using the *Popgenome*[52] *R* package, see Supplementary Information section 2.1. For Y chromosome haplogroup determination the *Yleaf* v1[53] software was applied. For network analysis of STR data we used *Network* v10.1.0.0 and *Network publisher* v2.1.2.5[54,55], see Supplementary Information section 2.2. Due to low genomic coverages (<10,000 SNPs) we discarded individuals S2, S5 and S17 from the population genetic analyses. PCA was made by the Eigensoft smartpca software[56] using the Human Origins Panel SNP set[26], for other analyses the 1240k array[12] was used for SNP call, for results, see Supplementary Table S7. Individuals S4, S5, S6 and S20 were discarded from further tests for them being first degree relatives of other samples. For investigating ancestry estimates we used supervised admixture analysis calculated by the *ADMIXTURE* v1.3.0 software[57]. The results were visualised by custom *R* scripts. *f*-statistics and qpAdm were performed using the *admixr* v0.9.1[58] and the *admixtools* v2.0.0[26] *R* packages.

Data availability

All other study data are included in the article and/or Supplementary Information and tables. Sequencing data are deposited in the European Nucleotide Archive (ENA) and are available to download under accession number PRJEB49524. Our new software 'PAPline' is freely available at https://www.github.com/gerberd-workshop/papline.

Acknowledgement

This study was funded by the Hungarian Academy of Sciences through the Momentum Mobility research project (LP2015-3/2015). This paper and A.S-N. was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. E.A. 's work was supported by the National Research, Development and Innovation Office, Hungary (NKFIH) grant (PD-19/131839). M.B. 's work and the analyses of the BAD002 sample was supported by a NKFIH grant under project code K-18/128413. We would like to greatly thank restorer Zsuzsanna Herceg and digital artist Fanni Gerber for the artwork of individual S13. We are grateful for the radiocarbon dates to István Major and Mihály Molnár at the HEKAL AMS C^{14} facility of the Institute for Nuclear Research in Debrecen, Hungary.

Author contribution

D.G, V.K., A.Sz-N. conceived and designed the experiments. D.G. processed the sequencing data, created the PAPline, and performed downstream bioinformatic analyses. B.Sz. did all molecular laboratory work. O.Sz. created the mtDNA database for phylogenetic analyses. B.E. obtained Y chromosome STR data.

B.Gy. and E.A. optimised genetic analyses. J.I.G., A.H., L.P. performed Sr isotope analysis. G.K., Sz.F., V. K. and M.B. evaluated the archaeological context. B.G.M., Á.K. and K.K. did the anthropological examination of the remains. Á.K. made the facial reconstruction. V.Sz. performed radiocarbon calibrations and modelling. B.G.M. sampled the remains. E.A., V.K. and A.Sz-N. jointly supervised the research and wrote the paper with D.G. All authors provide critical feedback for this study and contribute to the final manuscript.

Ethics declarations

The authors declare that they had requested and got permission for the destructive bioarchaeological analyses of the archaeological material in the study from the stakeholders, excavator and processor archaeologists.

Competing interests

The authors declare no competing interests.

References

1. Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
2. Mittnik, A. *et al.* The genetic prehistory of the Baltic Sea region. *Nat. Commun.* **9**, 442 (2018).
3. 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).
4. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
5. Mathieson, I. *et al.* The genomic history of southeastern Europe. *Nature* **555**, 197–203 (2018).
6. Lipson, M. *et al.* Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* **551**, 368–372 (2017).
7. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
8. Olalde, I. *et al.* The Beaker phenomenon and the genomic transformation of northwest

Europe. *Nature* **555**, 190–196 (2018).

9. Schroeder, H. *et al.* Unraveling ancestry, kinship, and violence in a Late Neolithic mass grave. *Proc. Natl. Acad. Sci.* **116**, 10705–10710 (2019).
10. Freilich, S. *et al.* Reconstructing genetic histories and social organisation in Neolithic and Bronze Age Croatia. *Sci. Rep.* **11**, 16729 (2021).
11. 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).
12. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
13. Köhler, K. Anthropological examination of the Late Copper Age human remains. in *The Prehistoric Settlement at BalatonőszödTemetői-dűlő. The Middle Copper Age, Late Copper Age and EarlyBronze Age Occupation* 269–292 (Archaeolingua, 2014).
14. 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).
15. Mittnik, A. *et al.* Kinship-based social inequality in Bronze Age Europe. *Science* **366**, 731–734 (2019).
16. Slon, V., Nagar, Y., Kuperman, T. & Hershkovitz, I. A Case of Dwarfism from the Byzantine City Rehovot-in-the-Negev, Israel: Dwarfism and Christianity. *Int. J. Osteoarchaeol.* **23**, 573–589 (2013).
17. Bardsley, M. Z. *et al.* 47,XYY Syndrome: Clinical Phenotype and Timing of Ascertainment. *J. Pediatr.* **163**, 1085–1094 (2013).
18. Ehler, E. *et al.* AmtDB: a database of ancient human mitochondrial genomes. *Nucleic Acids Res.* **47**, D29–D32 (2019).
19. Tońska, K., Kodroń, A. & Bartnik, E. Genotype–phenotype correlations in Leber

hereditary optic neuropathy. *Biochim. Biophys. Acta BBA - Bioenerg.* **1797**, 1119–1123 (2010).

20. Landrum, M. J. *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).
21. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).
22. dbSNP. *NCBI* https://www.ncbi.nlm.nih.gov/snp/docs/RefSNP_about/ (2021).
23. Reich, D. David Reich Lab: Ancient DNA, Biology, and Disease. https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data (2021).
24. Buchvaldek, M. Zu den Beziehungen zwischen Böhmen, Mähren und dem Karpatenbecken in der Frühbronzezeit. *Bp. Régiségei* **36**, 211–221 (2002).
25. Dean, R. B. & Dixon, W. J. Simplified Statistics for Small Numbers of Observations. *Anal. Chem.* **23**, 636–638 (1951).
26. Patterson, N. *et al.* Ancient Admixture in Human History. *Genetics* **192**, 1065–1093 (2012).
27. Brunel, S. *et al.* Ancient genomes from present-day France unveil 7,000 years of its demographic history. *Proc. Natl. Acad. Sci.* **117**, 12791–12798 (2020).
28. Bondár, M. & Szécsényi-Nagy, A. Skull cult in the Late Copper Age. *Ziridava* **34**, 91–104 (2020).
29. Peltzer, A. *et al.* EAGER: efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).
30. 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).

31. Bóna, I. Die mittlere Bronzezeit Ungarns und ihre südöstlichen Beziehungen. *Archaeol. Hung.* **49**, 73–76, 229–230 (1975).

32. 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).

33. Bándi, G. Die „Kisapostag-Problematik". Die Kultur der transdanubischen inkrustierten Keramik. in *Kulturen der Frühbronzezeit des Karpatenbeckens und Nordbalkans.* 257–281 (1984).

34. Szabó, G. A Dunántúli mészbetétes edények népe kultúrájának kialakulása és belső időrendje a Bonyhádon feltárt temetőrészlet tükrében. in *Wosinsky Mór Megyei Múzeum Évkönyve* vol. 32 101–128 (2010).

35. Hajdu, T. A bronzkori Dunántúli mészbetétes edények népe kultúrájának bonyhádi temetője feltárása és az embertani leletek vizsgálata során alkalmazott módszerek tanulságai. in *Wosinsky Mór Megyei Múzeum Évkönyve* vol. 32 129–140 (2010).

36. 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).

37. Bóna, I. Bronzezeitliche Tell-Kulturen in Ungarn. in *Bronzezeit in Ungarn. Forschungen in Tell-Siedlungen an Donau und Theiss.* 9–42 (Frankfurt am Main, 1992).

38. Črešnar, M. New research on the Urnfield period of Eastern Slovenia. A case study of Rogoza near Maribor. *Arheol. Vestn.* **61**, 114 (2010).

39. 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).

40. Mozsolics, A. A kisapostagi korabronzkori urnatemető (Der frühbronzezeitliche Urnenfriedhof von Kisapostag). *Archaeol. Hung.* **26**, (1941).

41. K. Zoffmann, Z. Az M7-es autópálya nyomvonalán előkerült őskori embertani leletek rövid áttekintése. in *Gördülő idő. Régészeti feltárások az M7-es autópálya Somogy megyei szakaszán Zamárdi és Ordacsehi között – Rolling time. Excavationson the M7 motorway in County Somogy between Zamárdi and Ordacsehi.* 309–313 (2007).

42. K. Zoffmann, Z. A bronzkori Gáta-Wieselburg kultúra Nagycenk-Laposi rét lelőhelyen feltárt temetkezéseinek embertani vizsgálata. The anthropologic study of the burials unearthed at the Nagycenk-Laposi rét site of the Bronze Age Gáta-Wieselburg Culture. in *Múzeumi Közlemények* vol. 2 9–34 (2008).

43. Vicze, M. Bronze Age Cemetery at Dunaújváros-Duna-dűlő. *Diss. Pannonicae* **4**, 34–36, 52–53 (2011).

44. Somogyi, K. A kisapostagi kultúra birituális temetője Ordacsehi-Cserefölön – Das birituelle Gräberfeld der Kisapostag-Kultur on Ordacsehi-Csereföld. in *Őskoros Kutatók III. Összejövetelének konferenciakötete* vol. ΜΩΜΩΣ 349–381 (2004).

45. Köhler K. Őskori tömegsír embertani leletei Balatonkeresztúrról. **17**, 8 (2006).

46. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci.* **110**, 15758–15763 (2013).

47. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20130624 (2015).

48. Csáky, V. *et al.* Early medieval genetic data from Ural region evaluated in the light of archaeological evidence of ancient Hungarians. *Sci. Rep.* **10**, 19137 (2020).

49. Monroy Kuhn, J. M., Jakobsson, M. & Günther, T. Estimating genetic kin relationships

in prehistoric populations. *PLOS ONE* **13**, e0195491 (2018).

50. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
51. Suchard, M. A. *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, (2018).
52. 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).
53. 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).
54. Bandelt, H. J., Forster, P. & Rohl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).
55. Network Software. *Fluxus-Engineering* https://www.fluxus-engineering.com/sharepub.htm#a1 (2008).
56. Patterson, N. *et al. EIGENSOFT*. (2017).
57. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
58. Petr, M., Vernot, B. & Kelso, J. admixr—R package for reproducible analyses using ADMIXTOOLS. *Bioinformatics* **35**, 3194–3195 (2019).