ORIGINAL ARTICLE

Subdivisions of haplogroups U and C encompass mitochondrial DNA lineages of Eneolithic–Early Bronze Age Kurgan populations of western North Pontic steppe

Alexey G Nikitin[1], Svetlana Ivanova[2], Dmytro Kiosak[3], Jessica Badgerow[1,4] and Jeff Pashnick[1,5]

Prehistoric Europe experienced a marked cultural and economic shift around 4000 years ago, when the established Neolithic agriculture-based economy was replaced by herding-pastoralist industry. In recent years new data about the genetic structure of human communities living during this transition period began to emerge. At the same time, the genetic identities of the Eneolithic and Early Bronze Age (EBA) inhabitants from a prehistoric cultural crossroad in western North Pontic steppe region remain understudied. This report presents results of the investigation of maternal genetic lineages of individuals buried in kurgans constructed during the Eneolithic–EBA transition in the western part of the North Pontic Region (NPR). Mitochondrial DNA (mtDNA) lineages from the interments belonging to the Eneolithic as well as the EBA cultures such as Yamna (Pit Grave), Catacomb and Babino (Mnogovalikovaya or KMK) were examined. In the 12 successfully haplotyped specimens, 75% of mtDNA lineages consisted of west Eurasian haplogroup U and its U4 and U5 sublineages. Furthermore, we identified a subgroup of east Eurasian haplogroup C in two representatives of the Yamna culture in one of the studied kurgans. Our results indicate the persistence of Mesolithic hunter–gatherer mtDNA lineages in western NPR through the EBA, as well as suggesting a mtDNA lineage continuum connecting the western NPR inhabitants of the Early Metal Ages to the North Pontic Neolithic population groups.

Journal of Human Genetics (2017) 62, 605–613; doi:10.1038/jhg.2017.12; published online 2 February 2017

INTRODUCTION

The arid grasslands of the northern shores of the Black and Caspian Seas were the kind of an ethnocultural 'cauldron' that had a prominent role in the shaping of the key historical processes in Eurasia for millennia. A vital part in this dynamics was played by the western North Pontic Steppe (western North Pontic Region (NPR); Figure 1), which, for thousands of years, served as a link between the Balkan–Carpathian region and a vast strip of the Black Sea–Caspian steppes. The interaction of various 'demographic currents' produced a particularly intense concentration of prehistoric monuments in western NPR, underlying the complexity of cultural developments in the region (Supplementary Figures S5–S10). During the Early Metal Ages (EMAs, 5th–2nd millennium BC), western NPR served as a boundary between sedentary cultures with highly developed agriculture and metallurgy such as the Cucuteni-Trypillia and Gumelniţa, on the one hand, and numerous tribes with the defining role of pastoralism and the considerable mobility of the population, on the other. The EMA parstoralic nomadic cultures from the circum-Pontic area are the prime candidates for the introduction of the Indo-European languages to west Eurasia.[1] They have been historically considered to be migrants or even invaders from the east.[2] However, neither attributes of war nor cultural attributes of eastern provenance are widespread in their burial mounds called kurgans.

The steppe pastoralists from the Srednij Stog cultural horizon (4750 − 4200 cal BC; Supplementary Figure S8) are the likely founders of the kurgan tradition in the NPR in the early Eneolithic, but distinctive markers on top of burials can be traced to the local Neolithic groups of the NPR.[3] During the Early Bronze Age (EBA), the Yamna (Pit Grave, 3200 − 2200 cal BC; Supplementary Figure S9) culture communities extended and expanded the original Eneolithic kurgan mounds in the NPR as well as developing their own distinct kurgan burial tradition. In western NPR, the way the Bugeac Yamna group was repurposing Eneolithic kurgans is seen by some researchers as an indication of local cultural Eneolithic–EBA continuum.[4] These same kurgans were also used by the Catacomb (2700 − 2000 cal BC; Supplementary Figure S10) culture communities and, later, by the Babino (KMK, or Mnogovalikovaya) culture toward the end of the EBA (2250 − 1750 cal BC). In many cases, kurgan reuse continued through the Bronze Age and beyond.

[1]Biology Department, Grand Valley State University, Allendale, MI, USA; [2]Institute of Archaeology, National Academy of Sciences of Ukraine, Odessa, Ukraine and [3]I.I. Mechnikov Odessa National University, Odessa, Ukraine
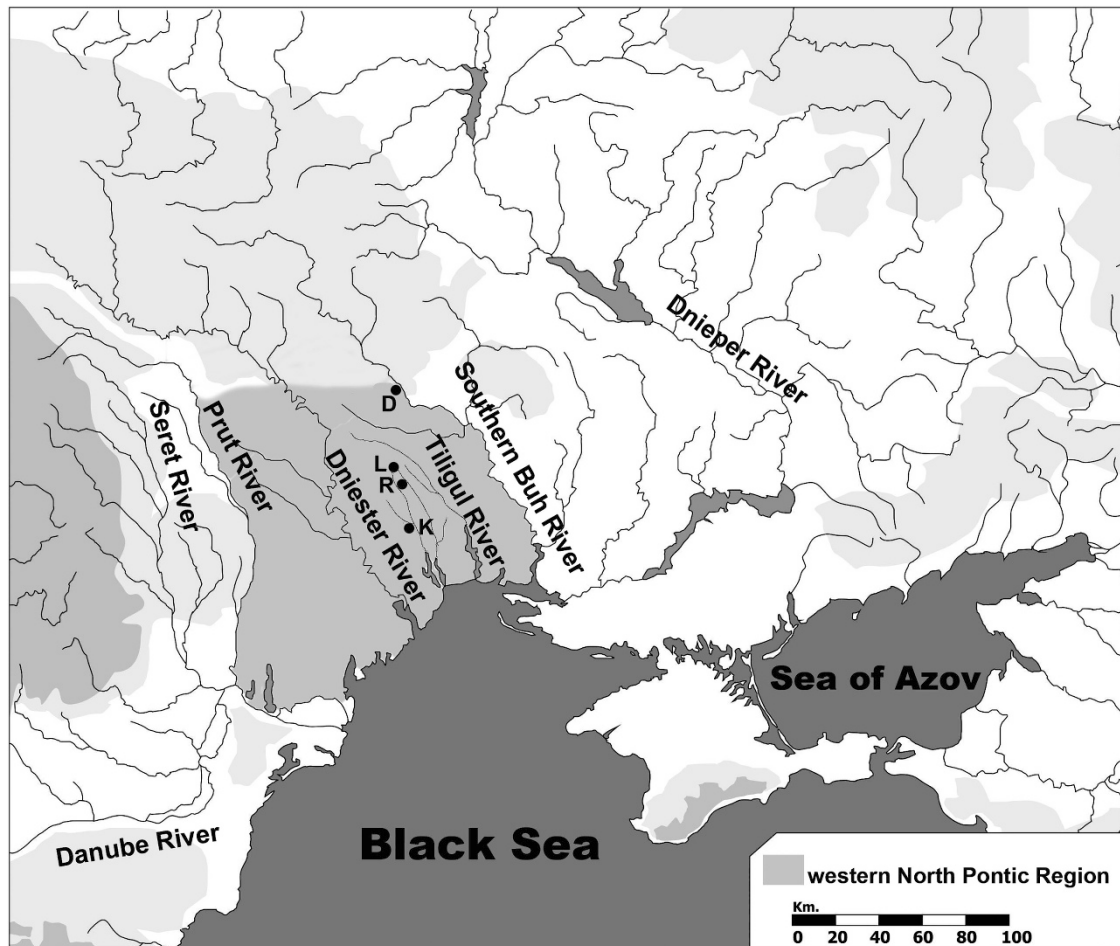[4]Current Address: Pine Rest Christian Mental Health Services, Grand Rapids, MI, USA.
[5]Current Address: Department of Plant Sciences and Plant Pathology, Montana State University, Bozeman, MT, USA.
Correspondence: Professor AG Nikitin, Biology Department, Grand Valley State University, 1 Campus Drive, Allendale, MI 49401, USA.
E-mail: nikitin@gvsu.edu

Figure 1 Western North Pontic region (NPR) with locations of kurgan sites sampled in the current study. D, Dubinovo; L, Liubasha; R, Revova; K, Katarzhino (1 and 2) kurgans.

Currently, it remains unclear whether the burial sites utilization in the NPR steppe was linked to an ancestral connection among the successive populations of steppe dwellers or it was a mere repurposing of the existing anthropomorphic landscape symbolism by the incoming outsider pastoralist groups. Kurgan utilization ties directly into the question of the origins of the pastoralist EMA cultures in the NPR. Many eastern European archeologists and anthropologists consider the cultures of the Eneolithic–EBA in the NPR to have autochthonous origins, albeit influenced to some degree by cultures from the east as well as from the west. The extent of these influences as well as the origins of the Eneolithic–EBA cultures in the NPR remain poorly understood, largely because such questions cannot be adequately resolved at the archeological level of analysis.

Archeogenetics can provide reference points and benchmarks for cultural and historical reconstructions. Genetic links among kurgan interments can help refine cultural interaction and succession concepts, thus producing a more comprehensive understanding of cultural dynamics in prehistory. The long-term interests of our research group are to understand population dynamics in southeast Europe at the time of the shift from sedentary agriculture to pastoralist economy and to trace the genetic and cultural transformations that accompanied this transition period. Our research method is based on the quantification of sociocultural changes in the key geographic areas where such transformations were most pronounced and the examination of the genetic variation that accompanied such changes.

If different culture groups contributing to the development of the post-Neolithic population landscape in the Pontic steppe had retained traceable genetic associations with their respective progenitor populations, then the tracing of differential genetic determinants of culture-specific interments in the steppe kurgans would allow us to discern the ultimate geographic origins of the first kurgan builders. From that, we can infer cultural associations and build a comprehensive picture of the transformations that had a profound effect on the development of post-Neolithic Europe. Focusing on individual life histories of the representatives of prehistoric populations within a defined geographic area may provide a better approach to quantifying population history compared with a wide sampling of transregional genetic variation that disregards local genetic specificity, thus subjecting the data interpretation to unnecessary generalization.

In this report, we present the maternal genetic lineage composition of kurgan builders in western NPR during the Eneolithic–EBA from a subset of geographically linked stratified kurgans, using the method of low-resolution mitochondrial DNA (mtDNA) PCR-SNP (single-nucleotide polymorphism-PCR). While the recently developed sophisticated DNA sequencing techniques like the next-generation DNA sequencing technology greatly expand the resolution of ancient DNA (aDNA) analysis, some basic questions about the connections among various population groups can still be answered by a low-resolution mtDNA screening of representative samples, especially when dealing with samples with potentially low DNA yield.

MATERIALS AND METHODS

Specimen description

In 2004, in a result of rescue excavations associated with the expansion of the Odessa-Kyiv highway in the Odessa region of Ukraine, a series of kurgans representing a wide temporal range of burials from the Eneolithic to the Iron Age have been investigated. Osteological material for the current study was obtained from the Dubinovo (D) kurgan 1, Liubasha (L) kurgan 2, Katarzhyno (K) kurgans 1 and 2 and Revova (R) kurgan 3[5] (Figure 1). The D kurgan was located in the southern part of the forest-step zone, whereas the rest were found in the steppe zone of western NPR.

The kurgans have individual peculiarities setting them apart from other kurgans in the region. The K1 kurgan is one of the five tallest EMA kurgans studied in western NPR (6.45 m in height). Four of the five kurgans were erected on top of Eneolithic structures. The D, K, R and D main Eneolithic burials featured elements of megalithic architecture in the burial chamber construction. Both Eneolithic burials in K1 and K2 were most likely produced by the people who succeeded the Stednij Stog culture complex. Eneolithic main burials R3.19a–b in the R kurgan are chronologically the oldest human remains found in a western NPR kurgan. The disarticulated partial remains (in layers likely to imitate a sitting position) of R3.19b were placed in the center of an ancient sanctuary, which likely already contained another burial (R3.19a). The remains of the latter burial are represented by a patellae found at a distance from the R3.19b bone 'package'. The L kurgan is the only kurgan in the current selection with the main burial belonging to the Yamna culture. Elsewhere in western NPR kurgans with the main Yamna burial comprise 80% of all kurgans. The interments from the D kurgan contained material culture artifacts not commonly found in Bronze Age kurgans, namely pottery of non-local origin (Supplementary Table S1).

Ancient DNA extraction and analysis

Samples from 16 individuals were selected for DNA analysis (Table 1 and Supplementary Table S1). The selection was primarily based on sample availability as well as the breadth of culture representation and the overall bone preservation based on visual morphology analysis. From the K kurgan, only the Eneolithic interments were available for the current genetic analysis. Only Catacomb culture burials were available from the D kurgan. The R and L kurgans provide the most comprehensive selection of chronological dates and associated cultures. Overall, three Eneolithic, five Yamna (the Bugeac subgroup of the Yamna cultural horizon), five Catacomb (including two belonging to the Ingul subgroup of the Catacomb horizon) and three Babino individuals were selected.

Ancient DNA extraction and analysis were performed as described in Nikitin *et al.*[6] DNA quantification was performed on a BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA). In all but two cases, two to three independent DNA extractions were performed on each specimen. For specimens L19 and R3.16, a single extraction was performed because of the limited initial amount of tissue sample.

The low-resolution PCR-SNP analysis was undertaken in a dedicated aDNA facility at Grand Valley State University (Allendale, MI, USA). All stages of the aDNA extraction and analysis were spatially separated. All aDNA extractions, PCR and cloning ligation reactions were performed under a laminate flow hood with internal UVC light source. Bone samples were cleaned by sanding off ~ 1 mm of the surface to remove surface contamination and then irradiated on all sides by UV light.

About 1 g of bone tissue was removed per extraction using a Dremel tool (Mount Prospect, IL, USA) and powdered using a sterilized porcelain mortar and pestle. The powder was washed three times with EDTA (pH 8.0) followed by three rinses with sterile water (pH 7.5). DNA was extracted using a QIAGEN DNA Investigator Kit (Qiagen, Valencia, CA, USA) protocol for bone extractions. Extracted DNA was eluted in 20–25 µl of sterile water and stored at − 20 °C.

Four overlapping primer pairs were used to amplify hypervariable region 1 (HVR-1) of the mtDNA control region using previously reported primers.[6] Diagnostic mtDNA coding region sites for haplogroups H (position 7028) and U (position 12308) were amplified using previously published primers.[7] Negative controls were used to detect the presence of contamination and positive controls, set up in isolation from aDNA, were used to establish effective PCR chemistry. Amplification was carried out using a Fast-Cycling PCR Kit

(Qiagen) as directed in the kit protocol following the conditions optimized for fragments in the 10–100 copy range, except PCR cycles were kept at 49 rounds. Each coding and control region segment was amplified up to four times per extraction or until two independent amplification products were obtained. Successful amplifications were cleaned using a QIAGEN MinElute Kit (Qiagen) and eluted into 10 µl of sterile water.

PCR amplifications were cloned by ligation into QIAGEN pDrive vectors using a QIAGEN PCR Cloning Kit (Qiagen). Transformed cells were grown on sterile LB-Amp agar plates and incubated at 37 °C for 16 to 20 h. Cells containing the PCR insert were selected by blue-white differentiation, replated and incubated again at 37 °C for 20–26 h. Subcultured cells were eluted into 250 µl of sterile water using a sterile loop. Clone DNA amplification was performed by using 1 µl of resuspended cells as template with SP6 and T7 universal primers to amplify the entire fragment within the plasmid cloning site. After an initial 5 min at 95 °C to lyse cells, 29 PCR cycles were as follows: 94 °C for 30 s, 42 °C for 45 s and 72 °C for 90 s with one elongation step of 72 °C for 5 min at the end of the 29 cycles. To verify an insertion of the desired PCR fragment into the plasmid vector, PCR products were visualized on a 2.5% agarose gel using a Low Molecular Weight DNA Ladder (New England Biolabs, Ipswich, MA, USA).

Sanger DNA sequencing analysis was performed at the Annis Water Research Institute at Grand Valley State University. Sequencing reactions were carried out on 96-well plates using BigDye Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) for 45 rounds. Sequencing reactions were cleaned before sequencing using a standard Sephadex protocol. Samples were run on an ABI 3130 × 1 Genetic Analyzer (Applied Biosystems) with a 50-cm capillary array.

DNA sequence analysis was accomplished using the tools from NCBI BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) through alignment with the revised Cambridge Reference Sequence (rCRS) of mtDNA (GenBank Accession No. NC 012920) to determine SNP differentiation. SNP variations were referenced with the phylogenetic tree of global human mtDNA variation, based on both coding and control region polymorphisms, at phylotree.org to determine haplogroup assignment. All chromatograms were thoroughly inspected using the 4Peaks program (A. Griekspoor and Tom Groothuis, mekentosj.com) and ambiguous base assignments were manually called by researchers.

Authenticity criteria

Ancient DNA studies are notoriously prone to contamination with modern DNA, usually coming from the people handing the remains, including archeologists, anthropologists and DNA researchers. There have been a variety of techniques suggested to minimize such contamination including amplifying small DNA fragments, cloning of PCR products and maintaining the strict conditions of sterility.[8] We followed all of these precautions including others such as determining the mtDNA polymorphism pattern of the researchers who had most direct contact with the specimens (Table 1). While this determination provides control over the most likely source of modern DNA contamination, it does not exclude the possibility of contamination by other exogenous sources that came into contact with the specimens between excavation and DNA extraction.

To insure the authenticity of aDNA results, multiple criteria were used, including DNA fragment length quantification based on Bioanalyzer data (aDNA is expected to be highly fragmented); two to three temporally separated DNA extractions from the same specimen (for all but two specimens); fragment size-dependent amplification success frequency (shorter fragments should be preferentially amplified in an aDNA/contamination mix); multiple PCR amplifications of the same region for each specimen (to achieve at least three successful PCR amplifications for each amplified region); short length of amplified fragments (128–164 bp); overlapping fragment amplification; SNP match in overlapping fragments; and the cloning of amplified DNA fragments. The molecular behavior of amplified fragments was examined. The sequences considered to be genuine aDNA displayed signs of deamination damage as well as displaying increased chimerization. Furthermore, SNP patterns from genuine aDNA were expected to make phylogenetic sense.

Personnel involvement

A1 (Table 1) handled the specimens during excavation and examination, using reasonable precautions. LP1 handled the specimens during cleanup, DNA

Table 1 Cultural associations, radiocarbon dating, nucleotide variation and mtDNA haplogroup assignment of the specimens presented in the report along with mtDNA haplotyping data for the personnel involved in the processing of the samples

Specimen ID, skeletal part used in DNA extraction	Number of clones sequenced	Archeological culture	C14, uncalibrated BP (Lab Code)/cal BC, 2δ	Coding and HVR-1 nucleotide variation	Haplogroup
D1.8, right ulna	39	Catacomb (EBA)	3940±70 (Ki-11200)/2621–2205	7028T, 12308G, 16192T, 16256T, 16270T, 16294T, 16296T	U5a1a1h
D1.10, tibia fragment	15	Catacomb (EBA)	Not dated	7028T, 12308G, 16270T	U5
D1.11, ulna	34	Catacomb (Ingul) (EBA)	3720±70 (Ki-11202)/2341–1921	7028T, 12308G	U
D1.12, 1st premolar	14	Catacomb (EBA)	3900±80 (Ki-11203)/2580–2140	rCRS	Contamination
K1.10, 5th right metatarsal	47	Eneolithic (Post-Stog)	4950±70 (Ki-9523)/3944–3638	7028T	Not determined
K2.1, mandibular fragment	25	Eneolithic (Post-Stog)	4270±90 (Ki-17331)/3311–2578	12308G, 16241G, 16270T	U5
L8, right humerus	24	Yamna (EBA)	3990±70 (Ki-11177)/2855–2291	16218T, 16223T, 16288C, 16298C, 16305T, 16327T, 16357C	C4a2
L9, 1st premolar	0	Babino (EBA)	3520±80 (Ki-11173)/2120–1636	No amplification	Not determined
L11, ulna fragment	53	Babino (MBA)	3230±70 (Ki-17332)/1687–1386	7028T, 16067T, 16192T	HV1
L15, skull fragment	21	Yamna (EBA)	3740±70 (Ki-11201)/2433–1945	16218T, 16223T, 16298C, 16327T, 16357C	C4a2
L19, 1st molar	28	Yamna (EBA)	4030±60 (Ki-11249)/2864–2350	7028T, 12308G, 16270T	U5
R3.7, long bone fragment	25	Yamna (EBA)	3910±60 (Ki-11058)/2568–2206	7028T, 12308G, 16192T, 16256T, 16270T	U5a
R3.13, long bone fragment	39	Catacomb (Ingul) (EBA)	3940±60 (Ki-11172)/2581–2209	7028T, 12308G, 16270T	U5
R3.14, mandibular fragment	0	Babino (EBA)	3590±70 (Ki-11175)/2137–1753	No amplification	Not determined
R3.16, premolar	31	Yamna (EBA)	4135±60 (Ki-11059)/2886–2503	7028T, 12308G, 16311C	U
R3.19a, right patellae	52	Eneolithic (Stog/Post-Stog)	5450±80 (Ki-1117)/4454–4054	7028T, 12308G, 16356C	U4
Personnel					
A1 (Archeologist)				rCRS	H
LP1 (pre-PCR)				16304C	H5a
LP2 (post-PCR)				7028T, 16126C, 16163G, 16186T, 16189C, 16194T	T1
LP3 (post-PCR)				7028T, 12308G, 16134T, 16356C	U4a1
LP4 (post-PCR)				7028T, 12308G, 16224C, 16311C	K

Abbreviations: EBA, Early Bronze Age; HVR-1, hypervariable region 1; LP, ancient DNA lab personnel; MBA, Middle Bronze Age; mtDNA, mitochondrial DNA; rCRS, revised Cambridge Reference Sequence.
Radiocarbon dates are from Ivanova et al.,[5] except for specimen K2.1, which appears for the first time in this report. Coding and HVR-1 region nucleotide variation was scored against rCRS and haplogroup assignment was determined according to the classification in van Oven and Kayser[13] and phylotree.org (mtDNA tree Build 17).

extraction and all DNA amplifications via PCR, which are the steps where most of the contamination was expected to occur. LP2, LP3 and LP4 handled the cloning of successful PCR products and the clone analysis steps and have not been in direct contact with the specimens nor have been involved in any pre-PCR manipulations at any point during the study.

Statistical analysis

The principal component analysis (PCA) was carried out on mtDNA haplogroup frequencies of 39 prehistoric Eurasian populations listed in Supplementary Table S2, following the mtDNA haplogroup subdivisions identified for each population in the corresponding publication source (Supplementary Tables S2 and S3). The PCA analysis was performed using the R Statistics Package v.3.1 (The R Foundation for Statistical Computing, Vienna, Austria) with graphical output generated with the component information for each population obtained from the R Statisitcs Package output. F_{ST} and exact test analyses of population differentiation were performed using Arlequin v.3.5.[9] In F_{ST} and exact test analyses, testing panmixia ($F_{ST}=0$) was the null hypothesis. PCA is an Eigen vector-based, multivariate and nonparametric test that shows the variance of the individual data points in an analysis by grouping them depending on the amount of variation explained by the vectors. F_{ST} is the allele fixation index and also measures the extent of genetic differentiation among populations.[10,11] The exact test of population differentiation was run to correct for the small and irregular sample size of ancient population data through 10 000 permutations using a Markov chain method.[12]

Radiocarbon analysis

Radiocarbon analysis was conducted at the Kyiv Radiocarbon Dating Laboratory (Ki) at the Institute of Environmental Geochemistry in Kyiv, Ukraine.

RESULTS

Of the 16 ancient specimens processed in the course of the study, 14 produced DNA amplification and sequencing data. Specimens L9 and R3.14, both representatives of the Babino culture, failed to yield amplifiable DNA. The SNP pattern in specimen D1.12 showed no deviation from rCRS and did not produce a deamination pattern expected in genuine aDNA, suggesting contamination. DNA sequences have been deposited in GenBank (http://www.ncbi.nlm.nih.gov/genbank/) under accession numbers KY073325—KY073335.

The amplification and sequencing of specimens D1.11 and R3.16 produced a recurrent transition at nucleotide 12308 in the coding region in three different extractions of D1.11 and one extraction of R3.16, designating these to belong to the U clade. However, the HVR-1 region amplification, cloning and sequencing failed to produce sufficient amplification to further refine the haplogroup assignment in these specimens. Specimen R3.16 produced a transition at the nucleotide position 16 311 in a single amplification and six subsequent clones of the corresponding HVR-1 fragment. Although the transition
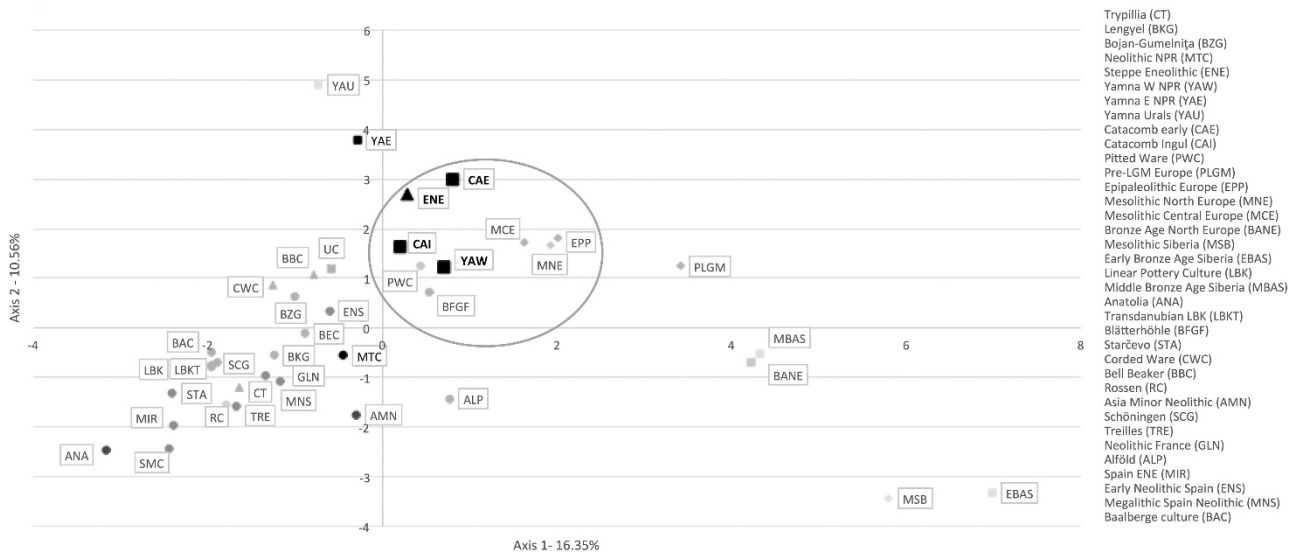
Figure 2 Principal component analysis of mitochondrial DNA (mtDNA) haplogroup frequencies of 39 Eurasian populations from the Paleolithic, Epipaleolithic, Mesolithic (rhombs), Neolithic (circles), Eneolithic (triangles) and Bronze Age (squares) cultures of Eurasia including the western North Pontic Region (NPR) Eneolithic (part of the Steppe Eneolithic, ENE) and Early Bronze Age specimens (western NPR Yamna, YAW, early Catacomb, CAE, and Ingul Catacomb, CAI) from this study. Symbols for the cultures from the Ponto-Caspian area are shaded in black, cultures from Asia Minor and Anatolia are 75% shaded, western and southern Europe—50% shaded, central Europe—35% shaded, northern Europe—25% shaded, east Eurasia—15% shaded. The Early Metal Age NPR-European Stone Age cluster is circled on the graph. Corresponding geographic locations, population sizes and the sources of data are given in the Supplementary Tables S2 and S3. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

at 16311 is frequent in the U clade, in the absence of other diagnostic polymorphisms in the HVR-1 region, it is not possible to further refine the sublineage assignment for R3.16 based on the available data. All R3.16 post-PCR manipulations were handled by Lab Personnel 3 (Table 1).

Specimen K1.10 produced a recurring thymine at the nucleotide position 7028. The diagnostic for haplogroup U region failed to amplify in specimen K1.10. The sequencing of 45 clones of the HVR-1 region amplicons displayed deamination pattern characteristic of genuine aDNA, but no polymorphic sites deviating from rCRS were detected. Thus, we were unable to determine the haplogroup lineage for K1.0.

Specimens L8 and L15 shared the 16223T–16298C–16327T–16357C HVR-1 motif characteristic for the C4a2 sublineage of haplogroup C, according to the classification in van Oven and Kayser[13] based on the SNP pattern in the HVR-1 region. Both specimens featured a transition at nucleotide position 16218 not previously reported for the C4a2 subclade. In addition, L8 featured additional transitions at 16288 and 16305, the former also being diagnostic for haplogroup C5. The limited amount of bone material precluded the analysis of the coding region of mtDNA in these two specimens. In specimen L8, the 16218 transition has occurred in five out of five clones derived from two independent amplifications of HVR-1 in the 16164–16277 region, which also contained the 16223 transition occurring in every clone of that segment. In specimen L15, of the six clones from two independent amplifications of the same region, the 16218 transition was observed in four out of six (derived from both amplifications) and all clones also featured the 16223 transition. While the presence of artifacts in aDNA work can never be fully ruled out, the occurrence of the 16218 transition in multiple independent amplifications in both specimens speaks for this transition to be endogenous for the aDNA in these two specimens, rather than being an amplification artifact.

Specimen L11, the only representative of the Babino group that produced amplifiable aDNA in our selection, belonged to haplogroup HV1 with the 16067–16192T SNP motif at HVR-1.

Specimen R3.19a carried a 16356 HVR-1 transition along with the 12308 coding region transition in three independent extractions, placing it into haplogroup U4. Nineteen clones of four PCR amplicons bearing the nucleotide position 16356 were sequenced, all bearing the transition at 16356. Lab Personnel 3 and 4 (Table 1) were involved in post-PCR manipulations.

The rest of the specimens in our sample selection (D1.8, D1.10, K2.1, L19, R3.7 and R3.13) carried the diagnostic for the U clade coding region transition at position 12308 and HVR-1 SNP pattern identifying them as members of the U5 subclade, although specimens D1.10, L19, K2.1 and R3.13 lacked one of the two diagnostic for the U5 polymorphisms in the control region at the nucleotide position 16192. The 16270 transition for specimens D1.10, L19, K2.1 and R3.13 was recorded for the majority of clones from independent PCR amplifications of the 16164–16277 and 16266–16385 regions both containing the 16270 polymorphism, from independent DNA extractions. Specimen K2.1 produced a recurrent transition at nucleotide position 16241 in 9 out of 11 clones of the corresponding HVR-1 segment from four different amplifications from two independent extractions. Specimens D1.8 and R3.7 were designated as members of the U5a subclade based on the nucleotide polymorphism pattern at HVR-1.

To investigate the genetic relationship between the EMA population groups in the NPR as well as the relatedness of the NPR populations to other population groups of Eurasia, PCA and the genetic distance analysis were performed on mtDNA haplogroup frequencies from this report combined with those available from literature and unpublished data (Figure 2) spanning Eurasian populations from the Paleolithic to the Bronze Age. For the purpose of these analyses, the Yamna cultural complex was split into western NPR, eastern NPR and Ural groups

610

(Supplementary Table S2). The western NPR Yamna group included the Yamna specimens west of the Southern Buh River delta, whereas eastern NPR encompassed the remaining Yamna specimens from the Ponto-Caspian steppe studied to date. The Ural group included specimens from the Caspian forest-steppe zone west of the Southern Ural Mountains (Supplementary Tables S2 and S3). The Catacomb cultural complex was divided into early Catacomb and Ingul Catacomb groups, to follow the currently accepted chronological division within the Catacomb culture group.[14]

When mtDNA haplogroup frequencies were considered in the space of principal components, component 1 accounted for 16.35% of the variance and component 2 comprised 10.56% of the variance (Figure 2). On the PCA graph, the steppe Eneolithic, western NPR Yamna as well as the early and Ingul Catacomb formed a cluster (circled on the graph) that also included the Epipaleolithic as well as the Mesolithic North and Central European, Pitted Ware and Blätterhöhle groups.

The genetic distance analysis (Supplementary Table S4) indicated a close genetic relationship among the steppe populations from the Eneolithic to the EBA in the NPR. For western NPR Yamna, the lowest population differentiation score was obtained with the early Catacomb group ($F_{ST} = -0.00436$, $P > 0.05$) followed by Bernburg, a regional subgroup of the Funnel Beaker/*Trichterbecker* culture from Germany ($F_{ST} = 0.01379$ $P > 0.05$) and eastern NPR Yamna ($F_{ST} = 0.01844$, $P > 0.05$). The NPR Eneolithic was closest to Ingul Catacomb ($F_{ST} = -0.01653$, $P > 0.05$) and Bernburg ($F_{ST} = -0.00286$, $P > 0.05$). Besides the Eneolithic NPR, Ingul Catacomb appeared closest to the Alföld group of the Eastern Linear Pottery Culture from Hungary ($F_{ST} = 0.00125$, $P > 0.05$) and the early Catacomb group ($F_{ST} = 0.00992$, $P > 0.05$).

DISCUSSION

Ancient DNA research is prone to errors, in part, based on the nature of the material and, in part, on the methods used. Although the low-resolution SNP-PCR method may have a potentially higher error rate than the whole-genome capture analysis using next-generation DNA sequencing technology, it can deliver results on the samples where other methods fail, particularly on those samples coming from geographic areas with potentially poor DNA preservation conditions. In this report, we made every effort to insure the authenticity of the obtained results, by following the customary clean lab procedures as well as best practices for working with aDNA (please refer to the Materials and methods section). Additional steps to minimize non-authentic findings included reducing the personnel conducing the pre-PCR work to a single person, working on a single specimen at a time and cloning all PCR amplicons. Although all these efforts may substantially reduce the chance of sample contamination as well as minimizing the possibility of elevating sequencing artifacts to the status of genuine aDNA, chances of random research artifacts affecting the outcome of aDNA findings may still persist. At the same time, it is also worth pointing out that the overall sequence quality can also be affected by such factors as disease (either at an individual level or at a level affecting a specific lineage), which is pertinent to modern and ancient humans alike and which may result in deviations from the lineage motifs that have already been described. Finally, unique undescribed nucleotide sequence polymorphisms may still exist in extant and ancient human populations, particularly in regions that have been systematically understudied.

The analysis of mtDNA lineages presented in this study provides an insight into the population history of the Eneolithic inhabitants of western NPR and their EBA successors. Our results suggest that the

composition of the maternal genetic lineages in western NPR steppe appear to stay relatively homogeneous from the Eneolithic through the EBA, represented by west and east Eurasian mtDNA lineages, the former belonging exclusively to lineages of haplogroup U in our sample selection. We have not observed any culture-specific segregation of the mtDNA lineages in our study, suggesting that the burial tradition (such as catacomb versus pit grave) does not necessarily imply population subdivision, at least from the maternal genetics point of view. The similarity of the composition of the Eneolithic and EBA mtDNA packages and their close genetic relationship evidenced from the PCA and genetic distance analyses based on mtDNA haplogroup frequencies from our study combined with previously published data suggests local Eneolithic–EBA continuity of mtDNA lineages in western NPR and in the Ponto-Caspian steppe overall. HVR-1 sequence matches (identical in all but one nucleotide position) to the R3.7 Yamna specimen from this report have been identified in the Yamna specimens from eastern NPR.[14,15] Exact sequence matches to the HVR-1 motif in the Eneolithic R3.19 specimen from our study have been identified in specimens belonging to the Catacomb culture from central and eastern NPR.[14] This overall biological continuity is further supported by craniological studies of the series of Ponto-Caspian skulls from the Mesolithic to the Middle Bronze Age.[16] At the same time, exact HVR-1 sequence matches to R.19 and D1.8 HVR-1 (in all but two nucleotide positions for the latter) have been found in Bell Beakers from Denmark[17] as well as the Bronze Age Andronovo culture from southwestern Siberia (exact sequence HVR-1 matches with R3.19 and R3.7)[18] and the Karakol Bronze Age population from the Altai region (for R3.7),[19] potentially suggesting other transregional associations.

A distinct inclusion of lineages of east Eurasian origin belonging to mtDNA haplogroup C was observed in the Yamna group of western NPR. The two C-bearing Yamna specimens from the L kurgan appear to possess a similar, albeit different at two nucleotide positions, SNP motif at HVR-1 in the C4a2 lineage of haplogroup C. Specimen L8, the chronologically older of the two C4a2-carrying specimens in our study, is more derived than L15, suggesting, on the one hand, an *ex situ* diversification of C4a2 lineages before their arrival in western NPR. On the other hand, it is equally likely that these lineages had diversified on the local substrate of the Neolithic builders of Mariupol-type cemeteries of extended burials of the Dnieper–Donets Culture Complex from the Dnieper Rapids in the central part of the NPR, specimens from which featured representatives of the C haplogroup.[6] In fact, both EBA lineages of C presented in this study could be viewed as derived from the lineage of the Dnieper–Donets Culture Complex specimen Ya34 from the Mariupol-type cemetery at Yasinovatka[6] via the transition at 16218. Sequence matches to the HVR-1 nucleotide motif of Ya34 have been identified in modern Siberian populations from the circum-Baikal area[20] as well as Sherpa populations from Tibetan highlands.[21] The lack of coding region polymorphism for the C lineages in the NPR limits our understanding of their phylogenetic relationships.

Anthropological data support a link between NPR's Dnieper–Donets Culture Complex and EBA populations, as well as connecting them with northern Europe, including the Mesolithic population of the South Olenij Ostrov.[6,22] The latter was featured as a part of the Mesolithic North European population in our PCA and genetic distance analyses, both indicating a mtDNA haplogroup frequency similarity between the NPR EBA and Mesolithic North European populations (Figure 2 and Supplementary Table S4). Representatives of east Eurasian mtDNA haplogroups such as C and D have been found on Olenij Ostrov.[23]

Mitochondrial lineages of U matching those obtained in the present study are widespread in modern and ancient European populations. In modern populations, exact HVR-1 sequence matches to specimens D1.10, L19 and R3.13 have been found in Italy,[24] Denmark[25] and Slovakia.[26] Specimens D1.10, L19 and R3.13 along with K2.1 lack one of the two diagnostic polymorphisms for U5 at the nucleotide position 16192 in the HVR-1 region, which is not uncommon, although it is usually associated with a gain of transitions at nucleotide positions 16256 and 16399 in the U5a1 sublineage. The reason for the incompletes of the HVR-1 SNP pattern in these four specimens remains unclear, but may relate to sample preservation or be a result of sequencing artifacts, especially for the 16192 polymorphic site, located within a C stretch, which may result in DNA polymerase 'slippage' at that location.[14] At the same time, a stand-alone 16270 transition in the HVR-1 region may reflect genuine HVR-1 polymorphism pattern in the U5 mtDNA lineage, as the finds of exact matches with specimens from a Neolithic hunter–gatherer group from Germany[27] as well the above-mentioned modern specimens indicate.

Exact sequence matches to the HVR-1 motif from R3.7 and R3.19 from the present study have been identified in modern populations from Spain,[28] Serbia,[29] Slovakia, Belarus, Russia, Poland,[26,30] Ukraine[31] and Denmark,[25] as well as in Neolithic populations from Hungary[32] and Germany,[33] and Mesolithic populations from Germany,[27,34] Lithuania and Poland,[27] northwest Russia (Olenij Ostrov)[23] and Sweden.[35] The high percentage of the mitochondrial lineages of U (U4, U5/U5a) in the specimens in the present study, HVR-1 sequence matches between the samples from the current report and European Mesolithic specimens, the dominance of the U4/U5 lineages in European Mesolithic populations[36] and the clustering of the EMA groups from the NPR and European Mesolithic populations on the PCA plot (Figure 2) as well as their mtDNA haplogroup frequency similarities in the genetic distance analysis (Supplementary Table S4) make a strong case for the maternal genetic connection between the inhabitants of the NPR, western NPR in particular, in the EMA and Mesolithic hunter–gatherer populations of Europe. At the same time, in the absence of the nuclear genome data and taking into account the small sample size of the currently available specimens, it is not possible to claim a direct genetic connection between western NPR and Mesolithic Europe at this point. The western NPR counterparts of the Mesolithic north and central European hunter–gatherers were such groups as the Mesolithic Kukrek and Grebenyky cultures (Supplementary Figures S5 and S6), which are the most likely candidates for the source of the hunter–gatherer lineages in western NPR, although since the genetic data on these populations is not yet available, a direct link between the Mesolithic and EMA inhabitants of western NPR cannot be made at the present time.

Although the U4/U5/U5a lineages of haplogroup U have been a major component of the mtDNA lineage variety in west Eurasian Mesolithic populations, they have also been found in ancient and modern populations of east Eurasia. Sequence matches to the HVR-1 motifs of R3.7 and R3.19 have been identified in an Eneolithic Ust-Tartas population from the Baikal region of west Siberia[37] as well as in modern Siberian Buryats.[26]

The U5-bearing K2.1 Eneolithic burial likely belongs to the Post-Stog culture group, which arrived in western NPR during Trypillia BII. Post-Stog, along with Cernavodă/Khadzhyder and Trypillia, likely contributed to the formation of the Usatovo culture. According to archeology, the Post-Stog group also likely participated in the founding of the Yamna culture.[5] Specimens with identical HVR-1 sequences in all but one nucleotide position to the

U5a-carrying specimen R3.7 have been identified[14] in late Eneolithic populations of the Dereivka culture from central NPR, belonging to the Srednij Stog cultural horizon (Supplementary Figure S8). In the mtDNA analysis of three Yamna specimens from the Usatovo culture site at Mayaki in western NPR,[14] two of the three belonged to haplogroups U and U5. The former carried a transition at nucleotide position 16311, also found in the U-carrying R3.16, and the latter displayed the 16256–16270 SNP motif, also found in the U5a-carrying specimen R3.7. Both R3.7 and R3.16 are from the Yamna culture. At the same time, two Eneolithic specimens from Mayaki, belonging to the Usatovo culture, yielded mtDNA haplogroups T2b and X2b.[14] Although the T2 and X mtDNA lineages have been identified in Yamna populations from the NPR and the Volga–Ural region,[14,38,39] these are typically associated with European Neolithic farming groups.[40] As Usatovo is considered to have formed, in part, on the foundation of the Eneolithic Trypillian farming culture,[41] it would be expected for its representatives to carry mtDNA haplogroups associated with the 'farming mtDNA package'. The presence of the T2b lineage in a Trypillian population from the Carpathian piedmont[42] confirms this assumption. At the same time, the persistence of mtDNA lineages of U/U5 in the Eneolithic samples from western NPR (our study and Wilde et al.[14]) points at the hunter–gatherer roots of the Eneolithic pastoralists of the Srednij Stog/Post-Stog populations in western North Pontic steppe. Furthermore, the Eneolithic–EBA continuum of U/U5 in western NPR lays support to the involvement of Post-Stog in the formation of Yamna at the genetic level. At the same time, the absence of U/U5 lineages in Usatovo so far suggests the lack of maternal genetic continuity between Post-Stog and Usatovo.

Of the four Yamna and three Catacomb specimens in our sampling, one Yamna and two Catacomb specimens presented variations of essentially the same U5 mtDNA lineage. The relationship between the Yamna and Catacomb cultures is a subject of a considerable debate among archeologists. While the Catacomb is generally considered to be a regional successor of Yamna in the NPR, recent radiocarbon data argue for the existence of an overlap between the two culture groups, particularly in western NPR.[43] Although the similarity of mtDNA haplogroup lineages does not clarify the question about the existence of an overlap between these cultures, our results suggest a shared maternal genetic pool. The contemporaneous D1.8 Catacomb and the R3.7 Yamna specimens both belonged to the U5a haplogroup, but the D1.8 lineage is more derived. The closest genetic matches to the D1.8 Catacomb specimen from our selection were found in a specimen from the Comb-Marked Pottery or Yelshanskaya culture in the forest-steppe zone of the middle Volga region (16192–16256–16270–16294 HVR-1 sequence motif) dated to the period of the Mesolithic–Neolithic transition[27] as well as in a late Eneolithic specimen from the Black Sea coast from Bulgaria (16114a–16192–16256–16270–16294 HVR-1 sequence motif).[14] Thus, it is likely that the source population harboring the 16192–16256–16270–16294 HVR-1 SNP pattern is autochthonous to the Ponto-Caspian region. At the same time, it is worth noting that an identical to the Yelshanskaya specimen in the HVR-1 motif sequence has been identified in a representative of the Blätterhöhle population from Germany dated to the early Mesolithic.[34] The Blätterhöhle population, as well as specimens from a Neolithic German hunter–gatherer site at Ostorf[27] also carried representatives with the exact HVR-1 sequence matches to the D1.10, L19 and R3.13 specimens from the current report, making the phylogeographic ancestral associations of the EBA populations from western NPR potentially more complex.

Elsewhere in the Ponto-Caspian steppe, mtDNA lineages of the Catacomb populations appear to be more diversified than in western

612

NPR, but, overall, 52% are representatives of haplogroup U and its sublineages (U4 and U5).[14] Noteworthy, over half of the Catacomb U lineages belong to haplogroup U4 and its representatives are about evenly distributed between the early and Ingul Catacomb groups. In contrast, in all of the Yamna specimens from the Ponto-Caspian steppe, representatives of U4 comprise ~7% of the total mtDNA lineage variety.[14,38,39] These ratios suggest a stronger influence of the bearers of haplogroup U4 on the mtDNA lineage pool of the Catacomb group. Alternatively, this is a reflection of genetic drift. At the same time, considering the small sample size of the Catacomb representatives studied to date (25 individuals excluding those from the current study), these values can be artifacts of limited sample selection.

In the Eneolithic Ponto-Caspian region, mtDNA haplogroup U4 has been found in western NPR (R3.19a, this study) as well as at the opposite end of the Ponto-Caspian steppe in the Volga–Ural region.[39] The R19a specimen likely belongs to Srednij Stog/Post-Stog groups, whereas the Volga–Ural specimen is a representative of the contemporary of the Srednij Stog Khvalynsk culture complex,[44] thus pointing to the autochthonous origins of U4 carriers in the Ponto-Caspian steppe as well as providing a genetic link between the east and west proto-kurgan nomadic groups in the region.

A single mtDNA lineage, HV1, potentially associated with farming Neolithic expansion in Europe, has been detected in our study in the L11 specimen from the Babino culture, belonging to the Middle Bronze Age. An exact sequence match to the HVR-1 sequence of L11 was identified in a modern specimen from Sudan.[45] As L11 is the only Babino specimen for which mtDNA lineage is available to date, we do not have enough data to discuss its significance in the context of maternal lineages of post-EBA Pontic steppe at this time.

The genetic distance analysis of mtDNA lineage frequencies in the Ponto-Caspian steppe suggests that at the level of mitochondrial lineage frequencies in the current state of resolution, the NPR Eneolithic, western NPR Yamna as well as the early and Ingul Catacomb population groups are derived essentially from the same maternal genetic pool. This observation is corroborated by the next-generation DNA sequencing technology analysis of whole ancient genomes from the Ponto-Caspian region. A study of Eneolithic and EBA genomes from the Volga–Ural and eastern NPR steppes demonstrated a close association of representatives of these groups from both steppe regions at the whole-genome level.[39] The study also showed that both groups gravitate towards East European Hunter Gatherers rather than European Neolithic and Eneolithic farmers. Additional analysis of these samples in a wider genetic context of prehistoric Eurasia further confirmed that the ancestry of the EBA steppe groups is almost 60% East European Hunter Gatherer,[46] fitting well with the data from the current study. The other significant genetic component in the EBA steppe ancestry appeared to hail from the Zargos mountain area in northwestern Iran.[46] The extent of the influence of this or other transregional genetic determinants on the genetic pool of Eneolithic–EBA populations in western NPR remains an open question, which should be addressed with additional sampling from western NPR and the use of the next-generation DNA sequencing technology.

Close genetic associations at the mtDNA level between the steppe Eneolithic, western NPR Yamna and the central European Bernburg group of the Funnel Beaker/*Trichterbecker* culture complex are difficult to evaluate using the currently available data. Presently, there is little archeological evidence connecting the above-mentioned cultures beyond the fact that both Funnel Beakers and western NPR Eneolithic groups employed megalithic architecture in the construction of their

burials.[5,47] Overall, small sample sizes of the NPR steppe populations and many other populations involved in the comparison analysis of mtDNA haplogroup frequencies in this report impede further examination of genetic links beyond the immediate phylogeographic relationships and do not allow further clarifications of the intercontinental connections of the steppe cultures of western NPR at this point.

The maternal genetic relationship among the interred in each kurgan from which specimens in our study came from is difficult to evaluate because of a small sample size and the deficiencies in nucleotide sequencing resolution. While it overall appears that the maternal genetic lineages of the prehistoric inhabitants of western NPR stem from the genetic pool of European Mesolithic hunter-gatherers, a more inclusive sample as well as complete mtDNA profiles, Y-chromosome and nuclear genome data are needed to comprehensively evaluate genetic relationships within and among various prehistoric population groups in the region. A broader sample selection should ideally include multiple individuals from region-specific well-stratified group of kurgans. Furthermore, the aDNA evaluation should be correlated with the burial context of a particular archeological culture, to avoid misinterpretation and misrepresentation of the results of the aDNA analysis.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1   Piazza, A., Rendine, S., Minch, E., Menozzi, P., Mountain, J. & Cavalli-Sforza, L. L. Genetics and the origin of European languages. *Proc. Natl Acad. Sci. USA* **92**, 5836–5840 (1995).
2   Gimbutas, M. A. in *American School of Prehistoric Research, Bulletin No. 20* (ed. Hencken, H.) (Peabody Museum, Harvard University, Camridge, MA, USA, 1956).
3   Rassamakin, Y. Y. in *Ancestral Landscapes TMO 58* (ed. Muller-Celka, S.) 293–306 (Maison de l'Orient et la Méditerranée, Lyon, France, 2011).
4   Petrenko, V. G. Kurgan epokhi paleometalla na poberezh'e Khadzhubejskogo limana. *MASP* **11**, 303–368 (2010).
5   Ivanova, S. V., Petrenko, V. G. & Betchinnikova, N. E. *Kurgans of Ancient Herdsmen from the Southern Buh and Dniester interfluve* (KP OGT, Odessa, Ukraine, 2005).
6   Nikitin, A. G., Newton, J. R. & Potekhina, I. D. Mitochondrial haplogroup C in ancient mitochondrial DNA from Ukraine extends the presence of East Eurasian genetic lineages in Neolithic Central and Eastern Europe. *J. Hum. Genet.* **57**, 610–612 (2012).
7   Santos, C., Montiel, R., Anglés, N., Lima, M., Francalacci, P., Malgosa, A. *et al.* Determination of human caucasian mitochondrial DNA haplogroups by means of a hierarchical approach. *Hum. Biol.* **76**, 431–453 (2004).
8   Cooper, A. & Poinar, H. N. Ancient DNA: do it right or not at all. *Science* **289**, 1139 (2000).
9   Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567 (2010).
10 Nei, M. Analysis of gene diversity in subdivided populations. *Proc. Natl Acad. Sci. USA* **70**, 3321–3323 (1973).
11 Wright, S. *Evolution and the Genetics of Populations* Vol. 4. *Variability Within and Among Natural Populations* (University of Chicago Press, Chicago, IL, USA, 1978).
12 Raymond, M. & Rousset, F. An exact test for population differentiation. *Evolution* **49**, 1280–1283 (1995).
13 van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–E394 (2009).

14 Wilde, S., Timpson, A., Kirsanow, K., Kaiser, E., Kayser, M., Unterländer, M. *et al.* Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5000 y. *Proc. Natl Acad. Sci. USA* **111**, 4832–4837 (2014).

15 Allentoft, M. E., Sikiora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).

16 Turkina, I. S. & Kazarnitsky, A. A. in *Cultures of the Steppe Zone of Eurasia and their Interaction with Ancient Civilizations* (ed. Alyoshkin, V. A.) 87–92 (Russian Academy of Science, Institute for the History of Material Culture, St Petersburg, Russian Federation, 2012).

17 Melchior, L., Lynnerup, N., Siegismund, H. R., Kivisild, T. & Dissing, J. Genetic diversity among ancient Nordic populations. *PLoS ONE* **5**, e11898 (2010).

18 Keyser, C., Bouakaze, C., Crubézy, E., Nikolaev, V. G., Montagnon, D., Reis, T. *et al.* Ancient DNA provides new insights into the history of south Siberian Kurgan people. *Hum. Genet.* **126**, 395–410 (2009).

19 Chikisheva, T. A., Gubina, M. A., Kulikov, I. V., Karafet, T. M., Voevoda, M. I., Romaschenko, A. G. *et al.* A paleogenetic study of the prehistoric populations of the Altai. *Archaeol. Ethnol. Anthropol. Eurasia* **32**, 130–142 (2007).

20 Shimada, M. K., Kim, C.-G, Takahashi, A., Spitsyn, V. A., Ikeo, K. & Gojobori, T. *et al.* in *DNA Polymorphism* Vol. 10 (ed. Ishidu, H.) 151–155 (Toyo Shoten, Tokyo, Japan, 2002).

21 Kang, L., Zheng, H.-X., Chen, F., Yan, S., Liu, K., Qin, Z. *et al.* mtDNA lineage expansions in Sherpa population suggest adaptive evolution in Tibetan Highlands. *Mol. Biol. Evol.* **30**, 2579–2587 (2013).

22 Potekhina, I. D. *Naselenie Ukrainy v epokhi Neolita i rannego Eneolita po antropologicheskim dannym* (National Academy of Sciences of Ukraine, Kiev, Ukraine, 1999).

23 Der Sarkissian, C. *Mitochondrial DNA in Ancient Human Populations of Europe* (PhD thesis, Univ. Adelaide, 2011).

24 Achilli, A., Rengo, C., Battaglia, V., Pala, M., Olivieri, A., Fornarino, S. *et al.* Saami and Berbers—An unexpected mitochondrial DNA link. *Am. J. Hum. Genet.* **76**, 883–886 (2005).

25 Raule, N., Sevini, F., Li, S., Barbieri, A., Tallaro, F., Lomartire, L. *et al.* The co-occurrence of mtDNA mutations on different oxidative phosphorylation subunits, not detected by haplogroup analysis, affects human longevity and is population specific. *Aging Cell* **13**, 401–407 (2014).

26 Malyarchuk, B., Derenko, M., Grzybowski, T., Perkova, M., Rogalla, U., Vanecek, T. *et al.* The peopling of Europe from the mitochondrial haplogroup U5 perspective. *PLoS ONE* **5**, e10285 (2010).

27 Bramanti, B., Thomas, M. G., Haak, W., Unterlaender, M., Jores, P., Tambets, K. *et al.* Genetic discontinuity between local hunter–gatherers and Central Europe's first farmers. *Science* **326**, 137–140 (2009).

28 Cardoso, S., Valverde, L., Alfonso-Sánchez, M. A., Palencia-Madrid, L., Elcoroaristizabal, X., Algorta, J. *et al.* The expanded mtDNA phylogeny of the Franco-Cantabrian region upholds the pre-Neolithic genetic substrate of Basques. *PLoS ONE* **8**, e67835 (2013).

29 Davidovic, S., Malyarchuk, B., Aleksic, J. M., Derenko, M., Topalovic, V., Litvinov, A. *et al.* Mitochondrial DNA perspective of Serbian genetic diversity. *Am. J. Phys. Anthropol.* **156**, 449–465 (2015).

30 Malyarchuk, B., Grzybowski, T., Derenko, M., Perkova, M., Vanecek, T., Lazur, J. *et al.* Mitochondrial DNA phylogeny in Eastern and Western Slavs. *Mol. Biol. Evol.* **25**, 1651–1658 (2008).

31 Derenko, M., Malyarchuk, B., Denisova, G., Perkova, M., Litvinov, A., Grzybowski, T. *et al.* Western Eurasian ancestry in modern Siberians based on mitogenomic data. *BMC Evol. Biol.* **14**, 217 (2014).

32 Szécsényi-Nagy, A., Brandt, G., Haak, W., Keerl, V., Jakucs, J., Möller-Rieker, S. *et al.* Tracing the genetic origin of Europe's first farmers reveals insights into their social organization. *Proc. Biol. Sci.* **282**, 20150339 (2015).

33 Keller, M., Rott, A., Hoke, N., Schwarzberg, H., Regner-Kamlah, B., Harbeck, M. *et al.* United in death-related by blood? Genetic and archeometric analyses of skeletal remains from the neolithic earthwork bruchsal-aue. *Am. J. Phys. Anthropol.* **157**, 458–471 (2015).

34 Bollongino, R., Nehlich, O., Richards, M. P., Orschiedt, J., Thomas, M. G., Sell, C. *et al.* 2000 Years of parallel societies in Stone Age Central Europe. *Science* **342**, 479–481 (2013).

35 Malmström, H., Linderholm, A., Skoglund, P., Storå, J., Sjödin, P., Gilbert, M. T. *et al.* Ancient mitochondrial DNA from the northern fringe of the Neolithic farming expansion in Europe sheds light on the dispersion process. *Philos. Trans. R. Soc. Lond. Ser. B* **370**, 20130373 (2015).

36 Brandt, G., Szécsényi-Nagy, A., Roth, C., Alt, K. W. & Haak, W. Human paleogenetics of Europe—the known knowns and the known unknowns. *J. Hum. Evol.* **79**, 73–92 (2015).

37 Molodin, V. I., Pilipenko, A. S., Romaschenko, A. G., Zhuravlev, A. A., Trapezov, R. O., Chikisheva, T. A. *et al.* in *Population Dynamics in Prehistory and Early History* (eds Kaiser, E., Burger, J. & Schier, W.) 93–112 (De Gruyter, Berlin, Germany, 2012).

38 Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).

39 Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).

40 Brandt, G., Haak, W., Adler, C. J., Roth, C., Szécsényi-Nagy, A., Karimnia, S. *et al.* Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. *Science* **342**, 257–261 (2013).

41 Patokova, E. F., Petrenko, V. G., Burdo, N. B. & Polischuk, L. Y. *Pamyatniki Tripol'skoj kul'tury v Severo-Zapadnom Prichernomorje* (Naukova Dumka, Kiev, Ukraine, 1989).

42 Nikitin, A. G., Sokhatsky, M. P., Kovaliukh, M. M. & Videiko, M. Y. Comprehensive site chronology and ancient mitochondrial DNA analysis from Verteba Cave—a Trypillian culture site of Eneolithic Ukraine. *Interdiscip. Archaeol. Nat. Sci. Archaeol.* **1**, 9–18 (2010).

43 Pustovalov, S. Z. Analiz radiokarbonnykh dat iz pogrebenij yamnoj i katakombnoj obschnostej, opublikovannykh v Baltic-Pontic Studies, No7 (1999). *Vita Antiq.* **5–6**, 45–59 (2003).

44 Mallory, J. P. & Adams, D. Q. *Encyclopedia of Indo-European Culture* (Fitzroy Dearborn Publishers, Chicago, IL, USA, 1997).

45 Musilová, E., Fernandes, V., Silva, N. M., Soares, P., Alshamali, F., Harich, N. *et al.* Population history of the Red Sea-genetic exchanges between the Arabian Peninsula and East Africa signaled in the mitochondrial DNA HV1 haplogroup. *Am. J. Phys. Anthropol.* **145**, 592–598 (2011).

46 Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D. C., Rohland, N., Mallick, S. *et al.* Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016).

47 Müller, J. *Megaliths and Funnel Beakers: Societies in Change 4100–2700 BC* (Joh. Enschedé Amsterdam BV, Amsterdam, The Netherlands, 2011).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)