Maternal genetic origin of the Late and Final Neolithic human populations from present-day Poland

Short running title: Human mitochondrial genomes from Neolithic Poland

Anna Juras\(^1\), Edvard Ehler\(^2\), Maciej Chyleński\(^1\), Łukasz Pospieszny\(^3,4\), Anna Elżbieta Spinek\(^5\), Helena Malmström\(^6,7\), Maja Krzewińska\(^8,9\), Krzysztof Szostek\(^10\), Wojciech Pasterkiewicz\(^11\), Marek Florek\(^12\), Stanislaw Wilk\(^13,14\), Barbara Mnich\(^15\), Janusz Kruk\(^16\), Marzena Szmyt\(^17,18\), Sławomir Kozieł\(^5\), Anders Gótherström\(^8,9\), Mattias Jakobsson\(^6,7\), Miroslawa Dabert\(^19\)

\(^1\)Institute of Human Biology & Evolution, Faculty of Biology, Adam Mickiewicz University in Poznan, Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland
\(^2\)Laboratory of Genomics and Bioinformatics, Institute of Molecular Genetics of the ASCR, v. v. i., Vídeňská 1083, 142 20 Prague 4, Czech Republic
\(^3\)Department of Anthropology and Archaeology, University of Bristol, 43 Woodland Road, Bristol BS8 1UU, United Kingdom
\(^4\)Institute of Archaeology and Ethnology, Polish Academy of Sciences, Rubież 46, 61-612 Poznań, Poland
\(^5\)Department of Anthropology, Hirszfeld Institute of Immunology and Experimental Therapy, Polish Academy of Sciences, Podwale 75, 50-449 Wroclaw, Poland
\(^6\)Human Evolution, Department of Organismal Biology, Uppsala University, Norbyvägen 18C, SE-752 36 Uppsala, Sweden
\(^7\)Centre for Anthropological Research, University of Johannesburg, Auckland Park, 2006 Johannesburg South Africa
\(^8\)Archaeological Research Laboratory, Department of Archaeology and Classical Studies, Stockholm University, Lilla Frescativägen 7, SE-106 91 Stockholm, Sweden
\(^9\)Centre for Palaeogenetics, Svante Arrhenius väg 20C, SE-106 91 Stockholm, Sweden
\(^10\)Institute of Biological Sciences, Cardinal Stefan Wyszyński University in Warsaw, Wóycickiego 1/3, 01-938 Warszawa, Poland
\(^11\)Institute of Archaeology, University of Rzeszów, Moniuszki Street 10, 35-015 Rzeszów, Poland
\(^12\)Institute of Archaeology, Maria Curie-Skłodowska University, Marii Curie-Skłodowskiej sq. 4, 20-031 Lublin, Poland
\(^13\)Institute of Archaeology, Jagiellonian University, Gołębia 11, 31-007 Kraków, Poland
\(^14\)The Karkonosze Museum in Jelenia Góra, Matejki 28, 58-500 Jelenia Góra, Poland
\(^15\)Department of Anthropology, Institute of Zoology and Biomedical Research, Jagiellonian University in Kraków, Gronostajowa 9, 30-387 Kraków, Poland
\(^16\)Polish Academy of Sciences, Institute of Archaeology and Ethnology, Sławkowska str. 17, 31-016 Kraków, Poland
Objective
We aim to identify maternal genetic affinities between the Middle to Final Neolithic (3,850-2,300 BC) populations from present-day Poland and possible genetic influences from the Pontic steppe.

Materials and methods
We conducted ancient DNA studies from populations associated with Złota, Globular Amphora, Funnel Beaker, and Corded Ware cultures. We sequenced genomic libraries on Illumina platform to generate 86 complete ancient mitochondrial genomes. Some of the samples were enriched for mitochondrial DNA using hybridization capture.

Results
The maternal genetic composition found in Złota-associated individuals resembled that found in people associated with the Globular Amphora culture which indicates that both groups likely originated from the same maternal genetic background. Further, these two groups were closely related to the Funnel Beaker culture-associated population. None of these groups shared a close affinity to Corded Ware culture-associated people. Haplogroup U4 was present only in the Corded Ware culture group and absent in Złota group, Globular Amphora and Funnel Beaker cultures.

Discussion
The prevalence of mitochondrial haplogroups of Neolithic farmer origin identified in Early, Middle and Late Neolithic populations suggests a genetic continuity of these maternal lineages in the studied area. Although overlapping in time – and to some extent – in cultural expressions, none of the studied groups (Złota, Globular Amphora, Funnel Beaker), shared a close genetic affinity to Corded Ware culture-associated people, indicating a larger extent of cultural influence from the Pontic steppe than genetic exchange. The higher frequency of haplogroup U5b found in populations associated with Funnel Beaker, Globular Amphora and Złota cultures suggest a gradual maternal genetic influx from Mesolithic hunter-gatherers. Moreover, presence of haplogroup U4 in Corded Ware groups is most likely associated with the migrations from the Pontic steppe at the end of the Neolithic and supports the observed genetic distances.

INTRODUCTION
Ancient DNA (aDNA) studies have provided robust evidence for several key demographic events that shaped the European gene pool. As genomic data becomes available for an increasing number of prehistoric human groups it helps to refine our understanding of those major events. Moreover, genetic information starts matching the complexity found in archaeological records and enables to find answers to the most challenging questions about past societies and their evolution. However, there are still areas where available genomic data do not completely reflect the cultural diversity seen in the archaeological record.

One such example is the transition from Late to Final Neolithic (Eneolithic/Chalcolithic) 3,500-2,800 BC in East Central Europe. Several large-scale genomic studies have recently shown that this period was linked with a major migration from the Pontic-Caspian steppe associated with the Yamnaya culture (YAM), and resulting in the emergence of the Corded Ware culture (CWC) in Central Europe (Allentoft et al., 2015; Haak et al., 2015; Jones et al., 2017; Linderholm et al., 2020; Malmström et al., 2019; Mittnik et al., 2018; Saag et al., 2017). These results indicate that incoming steppe people, who carried the new genetic “steppe” component, met and admixed with local populations. However, as the genomic background of these resident Neolithic groups is less well known, especially on a finer-scale geographic level, identifying exactly which groups participated in this admixture – and when – remains to be resolved. Therefore, the south-eastern region of present-day Poland, which is in the direct proximity to the Pontic steppe, and where there is evidence of several cultural entities predating the introduction of YAM related ancestry, is of great importance. The most widespread of these cultures were the Funnel Beaker culture (FBC; also known as Trichterbecher or TRB; 3,850-2,850 BC (Włodarczak, 2006) and the Globular Amphora culture (GAC; ~3,300 – 2,700 BC) (Szmyt, 2017).

Ancient human DNA studies conducted on FBC-associated individuals, originated mostly from Scandinavia and present-day Germany, revealed their Neolithic farmer-related ancestry and, when compared with early Neolithic populations, a slightly higher degree of western Mesolithic hunter-gatherer genetic components (Brandt et al., 2013; Malmström et al., 2015; Skoglund et al., 2014). Late FBC partially overlapped in time, and coexisted with the following GAC (Nowak, 2017; Szmyt, 2017), but the genetic affinities of populations associated with GAC and FBC have not been fully addressed. The only studies of GAC-associated people from present-day Poland were conducted on a passage grave in Kierzkowo and double burial in Brześć Kujawski, as well as a collective burial in Koszyce, from North and South East Poland, respectively (Fernandes et al., 2018; Mathieson et al., 2018; Schroeder et al., 2019; Tassi et al., 2017). Individuals analysed in these burials did not carry steppe ancestry patterns and, instead, displayed higher contributions of Neolithic farmer ancestry and of the Mesolithic western hunter-gatherer genetic component, when compared to CWC-associated people (Schroeder et al., 2019). However, radiocarbon dates obtained for particular individuals from these two graves indicate that they might have predated or been contemporary with the YAM migrations, what plausibly explains the lack of steppe ancestry in their genetic makeup (Schroeder et
al., 2019; Tassi et al., 2017). Although there are indications that GAC-associated populations arose on a similar genetic background as preceding FBC-associated people (Schroeder et al., 2019), the possible genetic continuity between these two groups, especially in the region of Lesser Poland, has never been investigated.

The Złota group (ZLO) was less common, but also present in South East Poland during the Late and Final Neolithic. This group name refers to an archaeological unit defined by a specific type of burials known only from a limited number of sites, and named after two cemeteries discovered in the village of Złota in the 1920s (Krzak, 1976). Funeral rites, grave equipment, and radiocarbon dates made it possible to date the usage of both necropolises to the first half of the 3rd millennium BC (~2,900 BC - 2,600/2,500 BC). ZLO is considered a local phenomenon at the confluence of CWC, GAC and Baden culture (BAD) (~3,650 – 2,900 BC) groups (Furholt, 2008; Włodarczak, 2017). In Lesser Poland the number of BAD burials is very small, and no genetic data is available for them. The existence of ZLO coincides with the expansion of the Pontic steppe herders westward and the emergence of the CWC-associated populations. The genetic influences of steppe-related groups on both ZLO and GAC have, however, not been fully explored. So far, only 6 ZLO-associated individuals were analysed (Schroeder et al., 2019). Nevertheless, the question remains whether the population associated with ZLO, represented by a much larger number of ancient individuals from different archaeological sites and showing cultural influences not only from GAC, CWC, and BAD but also from the YAM (Włodarczak, 2014) shared the genetic composition with GAC-associated groups or resembled CWC genetic makeup.

Our aim is to identify maternal genetic affinities between the Middle to Final Neolithic populations, and to investigate the possible genetic influences from the Pontic steppe. Therefore, in order to bridge the gap in genetic data from East Central Europe, in particular from present-day Poland, we performed analyses of 86 ancient mitochondrial genomes from individuals associated with ZLO and local groups of GAC, CWC and FBC.

MATERIALS AND METHODS

Bone materials

We sampled teeth and/or petrous bones from 125 ancient individuals recovered from archaeological sites associated with ZLO (N=53), GAC (N=21), CWC (N=8) and FBC (N=43). The analyzed individuals came from Lesser Poland with the exception of 6 GAC-associated samples from the Nakonowo archaeological site, located in the Kuyavia region in north-central Poland. The geographic locations of the analyzed populations are shown in Figure 1 were plotted using QGIS 2.12.2 (QGIS Development Team, 2015). The detailed description of archaeological sites and individuals is presented in Table S1 in Supporting Information S2.

The samples investigated herein were collected by us or our collaborators according to the cultural and heritage law. Given the age of samples and lack of cultural continuity in the region,
descendant communities are no longer in existence. In case of samples coming from museums or anthropological collections located at the universities we obtained official permissions from the head of the respective institution to collect teeth and/or petrous parts of temporal bone and conduct DNA studies. The museums and universities are bound by the heritage law to follow specific guidelines for acquisition and storage of skeletal human remains. Furthermore, due to recent developments in molecular extraction methods, we minimized the invasiveness of the selected protocols in order to preserve the ancient material for future investigations by using the smallest amount of bone material (<200 mg).

**DNA extraction, library preparation and sequencing**

The pre-PCR lab work procedures leading to high-throughput sequencing were performed in a sterile laboratory dedicated to aDNA analyses at the Institute of Human Biology & Evolution, Adam Mickiewicz University in Poznań (AMU), Poland. DNA extraction was performed using the method presented by (Yang et al., 1998) with slight modifications as in (Malmström et al., 2007) and purified using MinElute PCR Purification Kit (Qiagen, Germany). We built blunt-end genomic libraries following (Meyer & Kircher, 2010) with minor modifications as in (Günther et al., 2015) and purified one blunt-end library from each extract using AMPure® XP Reagents (Agencourt-Beckman Coulter, US) following manufacturer's instructions. DNA fragment length distribution and concentrations of the libraries were estimated using the TapeStation High Sensitivity D1000 screen tapes (Agilent Technologies, US). The single-indexed genomic libraries were screened on Illumina HiSeq X Ten (150bp paired-end, each library on 1/20 lane) or Illumina HiSeq2500 (125bp, paired-end, each library 1/10 lane) at the SNP & SEQ technology platform in Uppsala, Sweden.

**Mitochondrial DNA capture enrichment and sequencing**

We performed mitochondrial DNA (mtDNA) enrichment by capture hybridization on 46 samples which yielded less than 4.9-fold average mtDNA coverage in the NGS screening (table S1). We carried out two rounds of mtDNA enrichment using home-made RNA bait library following (Carpenter et al., 2013) with minor modifications as in (Juras et al., 2018). Three of the enriched libraries (Table S1 in Supporting Information S2) were converted into Ion Torrent libraries by PCR with indexed fusion primers following (Juras et al., 2017) and sequenced on the Ion 318 chip using the Ion Torrent PGM system (Thermo Fisher Scientific) at the Molecular Biology Techniques Laboratory, AMU, Poland. The remaining 43 enriched libraries were sequenced on Illumina HiSeq2500 (125bp, pair end, each library 1/10 lane) at the National Genomics Infrastructure (NGI) in Stockholm, Sweden.

**Bioinformatics analyses**
Preliminary pipeline computations of sequenced data were processed using a custom analytical pipeline (Günther et al., 2015) and the computational infrastructure provided by the Swedish National Infrastructure for Computing (SNIC) through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) (Lampa et al., 2013). Mitochondrial DNA data from capture hybridization were processed using the resources provided by The Polish Grid Infrastructure (PL-Grid). Merging of the read pairs and removing of adapters were carried out following (Meyer & Kircher, 2010). We used BWA software package version 0.7.8 (Li & Durbin, 2009) with the non-default parameters -l 16500 -n 0.01 -o 2 -t 2, to map merged reads as single-end reads against the revised Cambridge Reference Sequence (rCRS) (Anderson et al., 1981; R. M. Andrews et al., 1999) (GenBank: NC_012920). We used PL-Grid computational infrastructure to demultiplex DNA sequences generated by the PGM Ion Torrent system using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). The long (−M 110), short (−m 35), and low-quality sequences (−q 20) were removed using Cutadapt v.1.8.1(Martin, 2011). The filtered reads were analyzed with FastQC v 0.11.3 (Andrews, 2012) following the options described previously by (Chyleński et al., 2017). The sequences were mapped against the rCRS using TMAP v3.4.1 (Merriman et al., 2012). We used FilterUniqueSAMCons.py to collapse duplicate sequence reads with identical start and end coordinates for both PGM and Illumina sequence data following (Meyer & Kircher, 2010). Consensus sequences were built using ANGSD v0.910 (Korneliussen et al., 2014) and we accepted only positions with mapping scores of 30, a minimum base quality of 20, and covered by at least 3 reads as in (Chyleński et al., 2017). Mitochondrial haplogroups were determined for each individual using the PhyloTree phylogenetic tree build 17 (van Oven et al., 2014) and Mitomaster (Lott et al., 2013).

**Ancient DNA authenticity**

We utilized mapDamage v2.0.5 (Jónsson et al., 2013) to analyze misincorporation patterns which involved fragment length distributions and the deaminations characteristic for aDNA such as C to T transitions accumulated at 5′ ends and corresponding G to A transitions at 3′ ends of DNA fragments. The levels of mtDNA contaminations were calculated using contamMix (Fu et al., 2013).

**Population genetic studies**

For statistical analyses we used comparative population sets from the ancient human mt genomes database (AmtDB) (Ehler et al., 2019) and from the literature (Table S2 and S3 in Supporting Information S2). To better understand the local maternal genetic variability of FBC-associated populations, we divided it into two geographic groups, according to their geographic origin. The first north-western group (FBCnw) consisted of Scandinavian and western European individuals and the second east-central group (FBCce) encompassed individuals from present-day Poland. GAC
comprised the individuals from both central-north and south-eastern Poland and Ukraine. CWC included samples from this study and other Central European CWC from the literature.

Mitochondrial haplogroup frequencies were calculated for each population (Table S2 in Supporting Information S2) and used for the Principal Component Analysis (PCA), Shared Haplogroup Distance (SHD) and t-distributed Stochastic Neighbor Embedding (t-SNE). The results of these analyses were visualized with Matplotlib v. 3.0.3 Python package (Hunter, 2007). To trace the genetic affinities between ancient populations, the PCA was calculated in Python 3.6. The results obtained for the first five principal components from the PCA were used as input data for k-means clustering. All k-means variants were presented in the supplementary electronic material. We computed SHD and calculated a matrix of pairwise population distances and its p-values using Python 3.6. The significance of the pairwise SHD distances (p-values) were assessed using the Arlequin software as proposed by (Excoffier & Lischer, 2010) where null hypothesis of no differences between populations is tested by permuting haplotypes between the populations and p-value being the proportion of permutations with SHD distance higher than the observed value. Obtained SHD pairwise matrix was utilized to compute the multidimensional scaling (MDS) using Python Seikit-learn v. 0.20.3 package (Pedregosa et al., 2011). Mitochondrial haplogroup frequencies were used also to compute t-SNE, a modern and highly efficient method for visualizing multidimensional data (Maaten & Hinton, 2008). This analysis was performed using the methods previously described (Juras et al., 2020).

Complete mt genome sequences were used to calculate $F_{ST}$ genetic distance and in haplotype sharing analysis. These pairwise genetic distances, as well as p-values based on 10,000 permutations and linearization (Slatkin, 1995), were computed in Arlequin 3.5 (Excoffier & Lischer, 2010). For this particular analysis we used the overall 960 complete mt genomes from the literature and amtDB (Ehler et al., 2019) (Table S3 in Supporting Information S2). $F_{ST}$ matrix and p-values were used as input values for both the molecular variance (AMOVA) and MDS analyses. We run AMOVA analysis to investigate the within and between group variability using Arlequin 3.5 software package (Excoffier & Lischer, 2010) and tested the position of GAC, ZLO, FBC and CWC-associated populations in regards of three population groups: Group 1 (hunter-gatherers); Group 2 (central European farmers); Group 3 (steppe populations). Shared haplotypes between populations were identified using Arlequin 3.5 (Excoffier & Lischer, 2010). Those instances were then counted for each population pair ($p_1, p_2$) and normalized by sample size of both population in the pair ($\text{shared haplotypes count} / (n_{p1} + n_{p2})$). This measure can have values higher than 1 in case of sharing common frequent haplotypes between the populations. The value for shared/same haplotypes in one population was set to 0. Vizualization of the shared haplotypes using the chord diagrams was created in Python 3.6 and the chord library (v. 1.0.1) (https://pypi.org/project/chord/).
U4 haplogroup network

We have constructed a network of U4 haplotypes using 78 ancient U4-assigned individuals with complete mtDNA available from this study and the literature (Table S7 in Supporting Information S2). Mitochondrial genomes were aligned using MUSCLE v3.8.31 (Edgar, 2004) and the alignments was manually corrected where it was necessary (i.e. stretches of mononucleotides with insertions/deletions). In total, 59 variable bi-allelic positions (SNPs) were input for network construction. We have used Networks (version 10.1.0, https://www.fluxus-engineering.com/) for the calculations. Firstly, we have utilized the star-contraction procedure (delta = 5), followed by reduced median and median joining computations. The last step was the maximum parsimony calculation to simplify the resulting network (Bandelt et al., 1995, 1999; Forster et al., 2001; Polzin & Daneshmand, 2003).

Radiocarbon dating

We report new radiocarbon dates for 15 individuals representing ZLO (N=6), GAC (N=8) and CWC (N=1), dated at the Poznań Radiocarbon Laboratory, Poznań, Poland. Additional AMS and conventional dates for 8 individuals were collected from literature (supplementary electronic material, table S1). All 14C dates were calibrated with OxCal 4.4 (Bronk Ramsey, 2009) using the IntCal 20 curve (Reimer et al., 2020).

RESULTS

Sequencing results and contamination controls

We retrieved mt genomes for 86 of the 125 analyzed individuals associated with ZLO (N=28), GAC (N=18), FBC (N=32) and CWC (N=8). Among them, 40 mt genomes were obtained from Illumina shotgun screening and 46 from capture hybridization with the average coverage ranging from ca. 5- to 287-fold, and ca. 4- to 1118-fold, respectively. Calculated levels of contamination revealed that the sequence data carried >90% authentic DNA sequences (Table S1 in Supporting Information S2). In addition, the nucleotide misincorporation patterns reported for each sample indicated damages at the 5’ and 3’ ends characteristic for aDNA (Figure S9 in Supporting Information S1). Mitochondrial genomes generated in this study have been deposited in GenBank under accession numbers MT588211-MT588296.

Mitochondrial haplogroup diversity and distribution of U4 haplogroup

Among all ancient individuals, we identified mt lineages prevalent in present day Eurasian populations (H, J, K, T, U, W, X, I, V, R, HV). In general, GAC, ZLO, and FBC-associated populations displayed higher frequencies of Neolithic farmer origin mt haplogroups i.e., H (32.2–43.2%) and J (7.7-19.3%) and high frequency of U5b (7.7%-15.4%) commonly associated with western Mesolithic hunter-gatherers (Brandt et al., 2013; Jones et al., 2015). In contrast to this, CWC-
associated individuals exhibited the highest frequencies of mt haplogroup H (25.7%), U4 (20%) and U5a (11.4%), the latter two are frequently connected with either eastern or northern Mesolithic and sub-Neolithic hunter-gatherers (Jones et al., 2017; Malmström et al., 2015; Skoglund et al., 2014). This is supported by the major difference seen in haplotype sharing analysis, in which CWC has notably higher proportion of maternal hunter-gatherer ancestry (where U, including U4, is predominant) than GAC or ZLO (Figure S16 in Supporting Information S1 and Table S8 in Supporting Information S2).

Due to the absence of the haplogroup U4 in GAC, ZLO and FBC-associated populations, but at the same time its high frequency in the CWC complex, we employed the network analysis to get an insight into haplogroup U4 variability in ancient populations (Figures 2A and 2B; Table S6 in Supporting Information S2). We identified three main sources of mt U4 haplotypes in our dataset that comprised the Neolithic steppe populations, Mesolithic hunter-gatherers (and their remnants from Baltic/Scandinavia region), and Bronze Age (and later) samples from Central Europe. The major proportion of U4 haplotypes of possible steppe origin was found in almost every branch of the graph. This is true also for the Mesolithic haplotype U4 background represented in our dataset by hunter-gatherers from Balkans and north-eastern Europe. Most of the mt lineages from the steppe and the Baltic region (Lithuania, Estonia, Latvia, and Sweden) formed mixed branches of subhaplogroups U4a2, U4b1b1 and U4b1a2. It should be noted that the central, ancestral U4 node, represented by 22 nearly identical sequences, was not detected in the Baltic region. It consisted of samples from the steppe region, Central Europe, Balkans and Pannonia. Two other major nodes that included 5 and 7 individuals that belonged to subhaplogroups U4b1a1a1 and U4b1b, respectively, represented two groups of closely related haplotypes without direct influence from the steppe. The latter lineage was clearly a remnant of haplogroup U4 variability from pre-Neolithic Europe, where the majority of samples were coming from Mesolithic Balkans. The subhaplogroup U4b1a1a1 node displayed Late Neolithic, Bronze Age and Middle Age (Longobards) samples from present-day Poland, Germany, and Hungary. Haplotypes of U4 identified in western areas of Europe such as Great Britain were always found both in the west of Europe and in the Pontic steppe region or eastern Europe (Table S6 in Supporting Information S2).

Genetic affinities of Złota to other Middle to Final Neolithic human groups

To trace the genetic affinities of ZLO to other ancient human groups, we used mt haplogroup frequencies and visualized them in PCA space (Figure 3), t-SNE (Figure 4), and SHD based MDS (Figure 5). All results indicated that the ZLO-associated population was maternally related to GAC and FBC-associated groups, and separated from CWC. In the first PCA plot, which accounted for 50.38% of the total variance, the ZLO was grouped in direct proximity to GAC, but also close to FBCce and FBCnw as well as other Neolithic populations such as Late Danubian Neolithic (LDN) and Salzmünde (SMC) (Figure 3). The combination of PCA with k-means clustering (with $k=9$ as the best
representation of the data, at the average silhouette 0.29) (Figure 3, Figure S10 in Supporting Information S1) showed a clear division of populations into 9 groups representing people of Neolithic farmer origin (purple, yellow and pink), hunter-gatherers (dark green, light green and light blue), steppe origin cluster (orange), Asian origin clusters (red and dark blue). Correspondingly, in the t-SNE plot combined with the \(k\)-means clustering (with the \(k\) value of 6 as the best representation of the data, at the average silhouette of 0.59), ZLO was grouped close to GAC in a blue cluster further encompassing also FBC, SMC, and Middle Neolithic Germany (MNG) (Figure 4 and S11 in Supporting Information S1). Other clusters in the t-SNE plot represented hunter-gatherers, but also Andronovo (AND) and Bronze Age Siberians (BASI) (light green), Neolithic farmers (purple and light blue), and groups associated with the Pontic steppe (dark blue). The last cluster (yellow) comprised mostly populations with high amount of steppe ancestry associated with CWC, Bell Beakers complex (BBC), Únětice culture (UNC), Scythians from Moldova and Ukraine (SCU), Catacomb culture (CAT), Trzciniec cultural circle (TCC) and Mierzanowice culture (MCC). Moreover, SHD based MDS plot (Figure 5) combined with the \(k\)-means clustering (with the \(k\) value of 6 as the best representation of the data, at the average silhouette of 0.39) (Figure S12 in Supporting Information S1 and Table S5 in Supporting Information S2) resembled the PCA and t-SNE results and grouped ZLO in direct proximity to GAC and FBC-associated populations. SHD values indicated ZLO-associated groups to be genetically closer to GAC and FBCce communities (0.247 and 0.305, respectively, \(p>0.005\)) (Table S5 in Supporting Information S2 and Figure S13 in Supporting Information S1). We found a statistically significant genetic differences between ZLO and YAM-associated populations (0.529, \(p<0.005\)) as well as GAC and YAM-associated people (0.567, \(p<0.005\)). Complete mt DNA sequences were used to calculate pairwise \(F_{ST}\) values and to visualize them on MDS (stress value = 0.156) (Figure S14 in Supporting Information S1). Obtained results also showed ZLO-associated population to be maternally closely related to GAC (\(F_{ST} = 0.00, p>0.05\)) and FBCce (\(F_{ST} = 0.00, p>0.05\)), and distant to CWC group (\(F_{ST} = 0.005, p>0.05\)) and YAM (\(F_{ST} = 0.05, p>0.05\)) (Table S4 in Supporting Information S2 and Figure S15 in Supporting Information S1). This pattern is also supported by haplotype sharing (Figure S16 in Supporting Information S1, Table S8 in Supporting Information S2). Moreover, ZLO and GAC-associated people were closer to Neolithic farmers from Balkans (NEBA) (\(F_{ST} = 0.00, p>0.05\)), and at the same time displayed the greatest maternal genetic distances to hunter-gatherers (Figure S15 in Supporting Information S1 and Table S4 in Supporting Information S2). In addition, AMOVA indicated the best variability distribution when the ZLO individuals were grouped together with GAC, FBC, CWC, and Neolithic farmer populations, while the two other groups were clustered into steppe origin people (YAM, SCU and SRU) and hunter-gatherers (HGW, HGN, HGS). The results of AMOVA clustering were driven by maternal Neolithic Farmer and not steppe ancestry. This is evident by clustering CWC, which we know has steppe ancestry, with Neolithic populations. This distribution displayed the highest variability among
groups (7.78%) and the lowest among populations within groups (1.12%) (Table S6 in Supporting Information S2).

**Maternal kinship patterns**

We have identified 70 distinct mt haplotypes among the 86 analyzed individuals associated with ZLO, GAC, FBC and CWC. Potential maternal kin relatives were found in GAC, FBC and ZLO-associated groups. The highest number of plausible maternal kinship was found in GAC and FBC-associated populations. In two collective graves from Sadowie (GAC), we identified two groups of potential maternal relatives. These graves consisted of six and three individuals that belonged to haplogroups I2 and H3v, respectively (Table S1 in Supporting Information S2). Similarly, in Nakonowo (GAC), we found two individuals out of six, which belonged to subhaplogroup J1c3. Among the FBC-associated populations, we identified possible maternally related individuals in a collective grave (no. XIII) from Bronocice. For this grave, we retrieved DNA from 16 skeletons belonging to children and young adults, among which we found 3 children assigned to haplogroup W3, 2 children and 1 young adult reported to belong to haplogroup H, 2 children assigned to haplogroup J1c and 1 child and 1 young adult identified to belong to H, but displaying different SNPs profile than the previous 2 individuals also assigned to H. Although ZLO comprised the largest sample set, we found only 3 potentially maternally related adults from Złota archaeological site that were assigned to haplogroup H40.

**Radiocarbon dating**

The results of AMS $^{14}$C-dating can be found in Table S1 in Supporting Information S2, Figure S17 in Supporting Information S1. The calibrated dates obtained for ZLO-associated individuals ranged from 2,919 to 2,497 BC and from 2,911 to 2,465 BC for GAC, while for CWC-associated individuals, the date ranged from 2,462 to 2,209 BC.

**DISCUSSION**

From 3,850 BC to 2,300 BC (Middle-Final Neolithic) the southern region of present-day Poland was settled by populations associated with a number of archaeological cultures, such as FBC, GAC, BAD, ZLO and CWC. These people often coexisted in close proximity to each other and partially overlapped in time. Moreover, the archaeological record shows bilateral exchange of goods and possibly ideas between contemporary groups (Włodarczak, 2014, 2017). However, mtDNA data obtained in this work indicate that only ZLO, GAC, and FBC-associated populations were maternally related and that they had no maternal genetic connection to CWC-associated groups.

In this study we used a larger number of ZLO-associated individuals (28 samples) compared to what was available before (6 samples) (Schroeder et al., 2019) and showed close genetic affinities between GAC and ZLO-associated people. Our results suggest that the latter could have been
genetically a part of GAC population or that both originated from the same maternal genetic background. This is consistent with previous data (Schroeder et al., 2019) that showed ZLO-associated individuals to be genetically closely related to local GAC individuals. These genetic similarities between GAC and ZLO-associated people correspond with the links observed in their material culture and funerary practices, especially in a form of collective burials, sometimes with fragmented and partially burned skeletons or bearing traces of post-deposition manipulations. Therefore, archaeologists often define ZLO as a part of the GAC tradition (Włodarczak, 2014, 2017). However, ZLO also show material culture similarities to CWC (e.g. stone shaft-hole axes and copper, bone and shell ornaments) and to the Pontic steppe tradition (e.g. niche construction of graves, use of ochre) (Włodarczak, 2014, 2017). This, on the other hand, is not supported by our data, as no maternal genetic relations were found between these two groups. Therefore, the genetic distances of ZLO to CWC and steppe pastoralists, such as YAM and other associated groups, indicate that the connections between these groups were likely limited to cultural exchange. Notably, our results also did not show any maternal genetic links between GAC-associated people and the Pontic steppe herders. This is especially interesting since most of the samples used here predated or overlapped YAM-associated expansion from the steppe that influenced the formation of the CWC. Although, the oldest CWC graves are dated to 2,900 – 2,700 BC (Allentoft et al., 2015; Malmström et al., 2019), the steppe migration probably took place shortly before that time, however, the precise dates are still a subject of a debate (Furholt, 2014; Kristiansen et al., 2017; Racimo et al., 2020). Thus, more detailed, regional studies are needed to fully resolve the timing and nature of migrations from the steppe, including regional modes of interaction. Nevertheless, the genetic distances between GAC and YAM-associated populations found in this study are concordant with the results of nuclear genomic studies that showed a lack of steppe ancestry in GAC-associated individuals (Fernandes et al., 2018; Mathieson et al., 2018; Schroeder et al., 2019; Tassi et al., 2017). This also stays in agreement with current archaeological evidence for indigenous Central European development of GAC, not supporting previous hypothesis of its steppe origins (Jarosz et al., 2016; Nowaczyk et al., 2017; Szmyt, 2010).

We note that the prevalence of mt haplogroups of Neolithic farmer origin identified in ZLO and GAC-associated people resembles those found in FBC communities in Central Europe (FBCce) and preceding later Early Neolithic Farmers (LDN), which suggests a genetic continuity of maternal lineages between Early and Middle/Late Neolithic in the region of present-day Poland. We found that mt U5b haplotypes, previously mostly associated with western Mesolithic hunter-gatherers, were prevalent in central European FBC communities (second most common lineage) and in the GAC-associated population (third most common lineage in the present study). These results suggest gene flow between Mesolithic hunter-gatherers and Neolithic farmers, which is in line with previous observations in later Early Neolithic LDN-associated populations from the same region (Chyleński et al., 2017; Fernandes et al., 2018). Ancient mt genomes support the idea that Early Neolithic populations gradually evolved with an influx of genes from Mesolithic hunter-gatherers, and gave rise
to the Middle and Late Neolithic populations associated with FBC and GAC. This gradual increase of hunter-farmer admixture seems to be a general pattern observed not only in the area of present-day Poland but also in other parts of Central and Western Europe (González-Forbes et al., 2017; Haak et al., 2015; Lipson et al., 2017; Mathieson et al., 2015; Rivollat et al., 2020). For example, nuclear genomic studies has shown that LDN, FBC and GAC-associated people from north-central Poland had similar genetic compositions and consisted mostly of Neolithic farmer and western Mesolithic hunter-gatherer ancestries, without influx of any steppe ancestry (Fernandes et al., 2018; Mathieson et al., 2018; Schroeder et al., 2019; Tassi et al., 2017). A similar pattern is also seen in Scandinavia (Malmström et al., 2015; Skoglund et al., 2014), where the first farmers in the area (FBCnw), display similar maternal genetic compositions as the FBCce in present-day Poland.

Similarities between GAC and FBC-associated populations were displayed not only in their mt haplogroup composition but also in their reconstructed maternal kinship patterns. Shared haplotypes between individuals, and thus, plausible maternal relations, were identified in GAC-associated populations from this (Sadowie and Nakonowo) and previous studies (Koszyce and Kierzkowo) (Mathieson et al., 2018; Schroeder et al., 2019; Tassi et al., 2017). This suggests that burying of biologically related individuals might have been a funerary custom typical for GAC in East Central Europe. However, the sudden deaths of individuals discovered in Koszyce, likely caused by inter- or intragroup conflict (Mathieson et al., 2018; Schroeder et al., 2019; Tassi et al., 2017), did not occur in Nakonowo (as evidenced by distinct radiocarbon ages of buried individuals) (Szmyt, 2002) and most probably also not in Sadowie. We also found possible maternal relatives in an FBC-associated population, among individuals interred in a collective grave from Bronocice (grave no. XIII). However, this type of collective burials are highly unusual in FBC societies in Lesser Poland, as most burials are typically single graves (Kruk & Milisauskas, 2018). Therefore, we cannot clearly state if these possible kinship patterns were generally prevalent in FBC-associated burial rites. To get a clearer picture, detailed studies, preferably using nuclear DNA on a larger number of FBC-associated individuals, are needed. Nevertheless, biological kinship patterns have also been found in other Neolithic farmer related collective graves which were spread across north-western Europe (Cassidy et al., 2020; Sánchez-Quinto et al., 2019). Due to the low number of samples coming from varied and scattered archaeological sites associated with CWC, we did not identify any potential maternal kin relatives in this group.

Our results support the general idea that CWC-associated people were genetically much closer to populations with an origin from the Pontic steppe than other Middle and Late Neolithic populations from the area of present-day Poland. This is in accordance with previous mtDNA data (Brandt et al., 2013; Juras et al., 2018) and nuclear genomic studies (Allentoft et al., 2015; Haak et al., 2015; Jones et al., 2017; Linderholm et al., 2020; Malmström et al., 2019; Mittnik et al., 2018; Saag et al., 2017) indicating that CWC emerged after migration of Pontic steppe people toward Central Europe at the end of the Neolithic. Interestingly, our findings demonstrate that the second most frequent mt lineage
in CWC from the area of present-day Poland was haplogroup U4, which was absent in other Middle and Late Neolithic populations from this study. Furthermore, the network analysis of haplogroup U4 revealed that this lineage was present in Mesolithic hunter-gatherers from Eastern Europe, the Baltic region and the Balkans. It is also found among Neolithic individuals from the Baltic region and Pontic steppe, but only in groups exhibiting Mesolithic hunter-gatherer genetic origins. The re-emergence of U4 haplogroup at the end of the Neolithic in the Central and Western Europe seems to be directly connected with the introduction of steppe ancestry and could possibly be used as its indicator on the maternal side, analogous to Y-chromosome R1a/b on the paternal side. Moreover, this conclusion supports our previous study (Jurasc et al., 2018) in which we indicated a contribution of both females and males to the formation of populations associated with eastern CWC and puts the findings of Goldberg et al. (2017) into perspective.

Conclusions

Ancient mt genomes suggest that Middle, Late and Final Neolithic ZLO, GAC and FBC-associated populations from present-day Poland were maternally related. Moreover they were genetically distant to CWC-associated people, even in the case of GAC and ZLO groups post-dating the introduction of steppe ancestry. The close affinity between the ZLO-associated group and the GAC people further indicates that these populations possibly originated from the same maternal genetic background. However, to better understand their genetic and cultural relation, both future high resolution genomic and integrative studies are needed. Moreover, the prevalence of mt haplogroups of Neolithic farmer origin identified in ZLO and GAC-associated people resembles the haplogroup composition found in FBCce and preceding LDN, which points to a genetic continuity of maternal lineages between Early and Middle to Final Neolithic in the region of present-day Poland. Additionally, the higher frequency of haplogroup U5b found in FBC, GAC and preceding LDN-associated populations suggest gradual maternal genetic influx of Mesolithic hunter-gatherer ancestry. Furthermore, CWC-associated people, including the CWC from Lesser Poland, had much closer maternal genetic affinities with steppe origin populations than to other Late and Middle Neolithic groups that were in their close proximity. In fact, the second most frequent mt lineage in CWC from the area of present-day Poland was haplogroup U4, which is absent in the other Middle to Final Neolithic populations from this study. Distribution of haplogroup U4 is most likely associated with the migrations from the Pontic steppe in the end of Neolithic.

DATA ACCESSIBILITY

The mitochondrial genome data for 86 individuals are available at the GenBank under accession numbers MT588211-MT588296.

AUTHOR CONTRIBUTIONS
A.J. and E.E. conceived and designed the study. A.J, M.C. and A.S. performed the experiments. A.J.,
E.E., and M.C processed and analyzed the data. A.J., Ł.P., H.M., M.K., A.G. and M.J. discussed
results and reviewed the manuscript. A.S., K.S., W.P., M.F., S.W., B.M., J.K., M.S., S.K. contributed
samples and provided input about archaeological and anthropological context. A.J., M.D. and M.J.
coordinated the study. A.J wrote the manuscript with input from all authors.

COMPETING INTERESTS
The authors declare they have no conflicting interests with the work herein.

FUNDING
A.J. was supported by a grant project awarded by the Polish National Science Center [grant no.
2017/01/X/NZ8/01472]. A.J. and E.E were supported by the mobility grant project awarded by
NAWA [PPN/BCZ/2019/1/00010] and MSMT [8J20PL063]. E.E was supported by ELIXIR CZ
research infrastructure project [MSMT Grant No: LM2018131] including access to computing and
storage facilities. Ł.P was supported by the European Union’s Marie Skłodowska-Curie Individual
Fellowship [H2020-MSCA-IF-2017, grant no. 798894, ISOPATH] at the University of Bristol, and by
the Polish National Science Centre [grant no. 2014/15/D/HS3/01304]. M.J. was supported by Knut
and Alice Wallenberg foundation and H.M. by the Swedish Research Council [grant no. 2017-02503].
M.S. was supported in radiocarbon dating of Nakonowo samples by a grant of the AMU Faculty of
Archaeology. Particular computations were conducted with the support of PL-Grid Infrastructure,
UPPMAX and Scilife SNP and Seq platform.

ACKNOWLEDGEMENTS
We are thankful to Paweł Sobczyk from the Museum of the Kujawy and Dobrzyń Land in Włocławek,
Poland for providing samples from Nakonowo. We are also grateful to Marcin Przybyła and Paweł
Micyk for providing samples from Czaple Wielkie.

References
Allentoft, M. E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard,
P. B., Schroeder, H., Ahlström, T., Vinner, L., Malaspina, A.-S., Margaryan, A., Higham, T.,
Chivall, D., Lynnerup, N., Harvig, L., Baron, J., Casa, P. D., Dąbrowski, P., … Willerslev, E.
https://doi.org/10.1038/nature14507
Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I.
C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R., & Young,


Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., Fu, Q., Mittnik, A., Bänffy, E., Economou, C., Francken, M.,


FIGURE LEGENDS

Figure 1. (A) Localization of the sampled Neolithic archaeological sites and ranges of the studied cultures - Funnel Beaker culture (FBC), Globular Amphora culture (GAC), Zlota group (ZLO), Yamnaya culture (YAM). Numbers indicate archaeological sites: (1) Czaple Wielkie; (2) Koniusza; (3) Bronocice; (4) Książnice Wielkie; (5) Książnice; (6) Ilża -Chwałowski Trakt; (7) Sadowie; (8) Malice Kościelne; (9) Święcica; (10) Samborzec; (11) Wilczyce; (12) Złota; (13) Sandomierz; (14) Garbów Stary; (15) Nakonowo. (B) Chronology of analyzed individuals.

Figure 2. Median network of U4 haplotypes. Colors of nodes represent region of Europe (A) and archaeological culture (B) where a particular haplotype was found. Superscript numbers in haplotype names indicate different individuals that are described in detail in supplementary electronic material, table S7. Population abbreviations: NEUk - Neolithic Ukraine; PWC - Pitted Ware culture; NEBL - Neolithic Baltic; UNC - Únětice culture; EBACz - Early Bronze Age Czech; CABA - Chalcolitic Balkans; BABI - Bronze Age Siberia; NESC – Neolithic Scandinavia; CWC - Corded Ware culture; HGE - East European hunter-gatherers; TCC - Trzciniec cultural circle; YAM - Yamnaya culture; LONG - Longobards; BBC - Bell Beaker culture; HGS - South European hunter-gatherers; BARu - Bronze Age Russia; EBAGe - Early Bronze Age Germany; HGN - North European hunter-gatherers.

Figure 3. PCA based on mtDNA haplogroup frequencies with k-means clustering. The two principal components explained 50.38% of the total variance. Loading vectors, representing mt haplogroup contributions, are highlighted as grey arrows. Populations are grouped into nine clusters according to k-means. Population abbreviations are as follows: AND - Andronovo culture; BABA - Bronze Age Balkans; BAC - Baalberge culture; BANE - Bronze Age Near East; BASI - Bronze Age Siberia; BEC
- Bernburg culture; CAT - Catacomb culture; CABA - Chalcolithic Balkans; CHAHu - Chalcolithic Hungary; CWCCe - Corded Ware culture, EBACz - Early Bronze Age Czech; EBAG - Early Bronze Age Germany; FBCce - Central East Funnel Beaker culture; FBCnw – North West Funnel Beaker culture; GAC - Globular Amphora culture; HGC - Central European hunter-gatherers; HGE - East European hunter-gatherers; HGS - South European hunter-gatherers; IAG - Iron Age Germany; IAK – Iron Age Kazakhstan; IASI - Iron Age Syberia - Aldy Bel culture; LBK - Linear Pottery culture; LDN - Late Danubian Neolithic; MCC- Mierzanowice culture; MIC - Minoans; MNG - Middle Neolithic Germany; NEBA - Neolithic Balkans; NEUk - Neolithic Ukraine; PPNE - Pre-Pottery Near East; PWC - Pitted Ware culture; SCA - Scytho-Siberian; SCG - Schöningen group; SCR - Rostov-Scythians, Samara; SCU - Scythians from Moldova and Ukraine; SMC - Salzmünde culture; SRU - Srubnaya culture; STC - Strzyżów culture; TAG - Tagar culture; TCC - Trzciniec Cultural circle; UNC - Únětice culture; YAM - Yamnaya culture; ZLO - Złota group. Detailed descriptions and references of comparative populations are provided in supplementary electronic material Table S2.

**Figure 4.** t-SNE results colored according the $k$-means clustering with $k=6$. Population abbreviations are as in Figure 3.

**Figure 5.** MDS plot based on SHD pairwise distance estimated from mt haplogroups. Populations are colored according to the $k$-means clustering with $k=6$. Population abbreviations are as in Figure 3.