



OPEN

# Corded Ware cultural complexity uncovered using genomic and isotopic analysis from south-eastern Poland

Anna Linderholm<sup>1✉</sup>, Gülşah Merve Kılınc<sup>2,3</sup>, Anita Szczepanek<sup>4,5</sup>, Piotr Włodarczak<sup>4</sup>, Paweł Jarosz<sup>4</sup>, Zdzisław Belka<sup>6</sup>, Jolanta Dopieralska<sup>7</sup>, Karolina Werens<sup>8</sup>, Jacek Górski<sup>9</sup>, Mirosław Mazurek<sup>10</sup>, Monika Hozer<sup>11</sup>, Małgorzata Rybicka<sup>12</sup>, Mikołaj Ostrowski<sup>13</sup>, Jolanta Bagińska<sup>14</sup>, Wiesław Koman<sup>15</sup>, Ricardo Rodríguez-Varela<sup>2</sup>, Jan Stora<sup>16</sup>, Anders Götherström<sup>2</sup> & Maja Krzewińska<sup>2✉</sup>

During the Final Eneolithic the Corded Ware Complex (CWC) emerges, chiefly identified by its specific burial rites. This complex spanned most of central Europe and exhibits demographic and cultural associations to the Yamnaya culture. To study the genetic structure and kin relations in CWC communities, we sequenced the genomes of 19 individuals located in the heartland of the CWC complex region, south-eastern Poland. Whole genome sequence and strontium isotope data allowed us to investigate genetic ancestry, admixture, kinship and mobility. The analysis showed a unique pattern, not detected in other parts of Poland; maternally the individuals are linked to earlier Neolithic lineages, whereas on the paternal side a Steppe ancestry is clearly visible. We identified three cases of kinship. Of these two were between individuals buried in double graves. Interestingly, we identified kinship between a local and a non-local individual thus discovering a novel, previously unknown burial custom.

The Neolithic Stone Age of continental Europe saw important demographic changes and population events which in recent years have been demonstrated by numerous archaeogenomic studies<sup>1–6</sup>. The genetic ancestry, affinity and admixture processes between human groups have been traced by analyses of individuals from different time periods and geographical contexts which contextually are associated to cultural complexes with differing lifeways, burial customs and material culture expressions. The identification of the demographic event and successive development associated with the Neolithization of Europe demonstrated the importance of mobility and migration in the process of population turnover and transition towards changed lifeways (e.g.<sup>7–9</sup>). The process took different paths in different areas of Europe which is evident in the admixture patterns between the hunter-gatherer (HG) and the farming groups but also between farming and pastoralist groups, depending on the time period. The significant transitional process around 3000 cal BCE and the appearance of the (Pontic Steppe) Yamnaya cultural complex in eastern Europe is associated with a wave of migration from East that had marked impact on the demographic (a Steppe ancestry component), cultural and social as well as linguistic development in the third Millennium BCE<sup>1,10</sup>.

<sup>1</sup>The BiG lab (Bioarchaeology and Genomics Laboratory), Department of Anthropology, Texas A&M University, College Station, USA. <sup>2</sup>Centre for Palaeogenetics, 10691, Stockholm, Sweden. <sup>3</sup>Department of Bioinformatics, Graduate School of Health Sciences, Hacettepe University, 06100, Ankara, Turkey. <sup>4</sup>Institute of Archaeology and Ethnology, Polish Academy of Sciences, Cracow, Poland. <sup>5</sup>Department of Anatomy, Jagiellonian University, Medical College, Cracow, Poland. <sup>6</sup>Isotope Laboratory, Adam Mickiewicz University, Poznań, Poland. <sup>7</sup>Poznań Science and Technology Park, Poznań, Poland. <sup>8</sup>School of Archaeology 34-36 Beaumont Street, Oxford, OX1 2PG, United Kingdom. <sup>9</sup>Department of History and Cultural Heritage, University of Pope Jan Paweł II, Cracow, Poland. <sup>10</sup>Archaeological Studio Mirosław Mazurek, Rzeszów, Poland. <sup>11</sup>Łańcut Castle Museum, Łańcut, Poland. <sup>12</sup>Institute of Archaeology, University of Rzeszów, Rzeszów, Poland. <sup>13</sup>Archaeological Society "ARGO", Cracow, Poland. <sup>14</sup>Regional Museum in Tomaszów Lubelski, Tomaszów Lubelski, Poland. <sup>15</sup>Provincial Office for the Protection of Cultural Heritage, Zamość, Poland. <sup>16</sup>Osteoarchaeological Research Laboratory, Department of Archaeology and Classical Studies, Stockholm University, Stockholm, Sweden. ✉email: [linderholm@tamu.edu](mailto:linderholm@tamu.edu); [maja.krzewinska@arklab.su.se](mailto:maja.krzewinska@arklab.su.se)

The appearance of the Yamnaya (around 3100-3000 BCE) complex in southern Europe roughly coincides with the appearance of the Corded Ware complex (CWC) (around 2900-2800 BCE) further to the North. Earlier archaeogenomic analyses have shown that the CWC individuals, exhibit Steppe ancestry<sup>1,10-12</sup>. Contemporaneous individuals analysed from this region (i.e. central Europe) have however shown a varying degree of Steppe ancestry<sup>1,10</sup>. This component has not been demonstrated among individuals associated with the Globular Amphora culture<sup>11,13,14</sup>. Although our knowledge about the demographic prehistory in continental Europe has increased immensely over the last five years<sup>1-6</sup> the regional resolution of the events is insufficient and, thus, our understanding of the social processes is still lacking. The population dynamics and interpretations of the archaeogenomic data of the Yamnaya and the Corded Ware complexes and their relationships have been intensely debated and reviewed (see e.g.<sup>15-18</sup>).

Around 5400 BCE farming was introduced in central Poland which coincided with the appearance of the Linear Pottery culture (LBK) complex marking the transition to the Early Neolithic period. The HG population living in a neighbourhood of the LBK area upheld their lifeways and for almost a millennium the dispersal of the Neolithic lifeways halted. However, there were evidently contacts between the farming and HG groups seen in admixture patterns<sup>11</sup>. These relations are clearly visible in the Brześć Kujawski group (of the Lengyel complex) found in the region of Kuyavia which for a long time was the northern limit of the (post LBK) farming communities<sup>11,19</sup>.

During the next millennium, from around 4000 BCE, i.e. the Middle Eneolithic, the Funnel Beaker culture (TRB) emerges and the Eneolithic lifeways dispersed to the Northern parts of central Europe, including parts of Scandinavia, and now to most of the Polish area. Individuals from TRB contexts in Poland exhibit admixture with the HG-populations which were mostly linked to the west European HG group (WHG) (e.g.<sup>11</sup>). Around 3100 BCE the TRB manifestations in Poland decline with the appearance of the Baden culture and the Globular Amphora culture (GAC, c.3400/3100-c.2800 BCE) which replaced the earlier TRB complex in many areas of Poland. The GAC was a rather short-lived phenomenon and after 2800 BCE the Corded Ware Complex (CWC) manifestations became dominant, and continued for another 500 years before that complex disappears<sup>11,20</sup>.

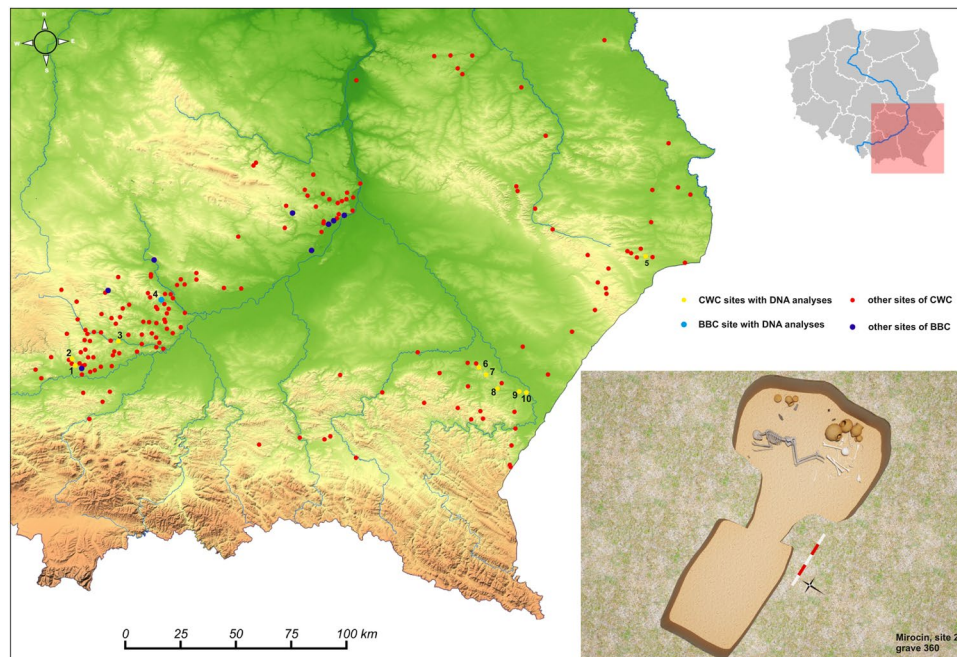
Individuals of the GAC and CWC have very different ancestry. Individuals from CWC contexts in Poland exhibit a marked input from the eastern Yamnaya Steppe pastoralist whereas the individuals of the GAC complex hardly exhibit any Steppe ancestry<sup>11,13,19</sup>. Around 2400 BCE the Bell Beaker culture (BBC) complex enters this region, from the southwest, and yet another major genetic component is added to the local population<sup>5,21</sup>. In south-eastern Poland, the BBC disappears around 2200 BCE<sup>22,23</sup>. The development in Poland clearly shows the extent of admixture that has taken place in continental Europe and the importance of mobility in the demographic and social developments<sup>10,11</sup>. (Supplementary Information - The Corded Ware and the Bell Beaker societies in the Małopolska uplands).

By identifying mtDNA, Y-chromosomal DNA and nuclear DNA; ancestry, origin and admixture can be ascertained<sup>24-27</sup>. It has been shown that Yamnaya pastoralists contributed Y chromosome R1a and R1b haplogroups to continental Europe almost entirely replacing the previously wide-spread G2a haplogroup<sup>1,2</sup>. The mtDNA contribution of Yamnaya to CWC individuals is associated with the appearance of U2 and W haplogroups. This large-scale contribution, close to complete replacement is very evident among CWC individuals that have retained very little genetic ancestry from the Mesolithic HG and Early Neolithic farmer groups. The study of individuals associated with the CWC complex detected ancestry and admixture patterns with the HG groups, individuals of the Steppe culture groups such as Afanasievo and Yamnaya complexes. Moving forward in time evidence of admixture with individuals of the succeeding BBC complex appears, suggesting that this region of prehistoric Europe was an important social arena and melting pot of human genetics from both the east and the west<sup>5</sup>. Archaeogenomic analyses have the best possibilities to decipher demographic processes and the individuals associated to the CWC complex provide important clues to the development in the 3<sup>rd</sup> Millennium BCE.

One of the most characteristic features of the CWC complex are the funeral rituals that quickly spread over a large part of the Central Europe<sup>28</sup>. Attempts have been made to identify and understand the migration patterns within this vast cultural complex, and whether there were regional subgroups or other kinds of subdivisions<sup>11,12,20,29</sup>. To identify and to understand how different subgroups of the Corded Ware complex interacted with each other and the surrounding populations, an in-depth analysis of individuals representing this cultural complex during a highly important period in prehistory (2800-2300 BCE) is of great importance<sup>29</sup>. The present study goes deeper by also examining kinship of the individuals, which will aid the understanding of the intricate networks and social structures of the CWC subgroups<sup>28</sup>. By further investigating the genomic signatures among the regional CWC population in south-eastern Poland, a more complete yet complex image will emerge with several admixture events from different cultural groups.

## Materials

Samples of the petrous portion of the temporal bone were taken from 50 individuals representing the Corded Ware Complex (CWC) and 3 individuals representing the Bell Beaker culture (BBC) buried in graves in south-eastern Poland. We have obtained genetic data from 19 individuals (16 of CWC and 3 of BBC). All examined individuals come from three geographical regions (Fig. 1; Table 1; Table S1): the Rzeszów Foothills (part of the Subcarpathian Region; sites of Szczytna, Chłopice, Mirocin and Święte), the Małopolska Upland (Mistrzejowice, Proszowice, Bosutów, Pełczyska) and the Sokal Ridge (the western part of Volhynian Upland – site of Łubcze) (Supplementary Information – Materials; Figs. S1–S13). All burials are of similar type exhibiting the same funeral rite with some differences concerning grave goods and their radiocarbon dates coincide (Table 1; Supplementary Information – 14 C Dating; Table S2, Fig. S15). The Strontium isotope (Sr) analyses included human enamel samples obtained from 16 individuals who were selected for sampling based on preserved dental enamel and availability of contextual information. Among them were females and males, predominantly of adult and mature individuals, and also children. Enamel was taken from first molars (M1) whenever

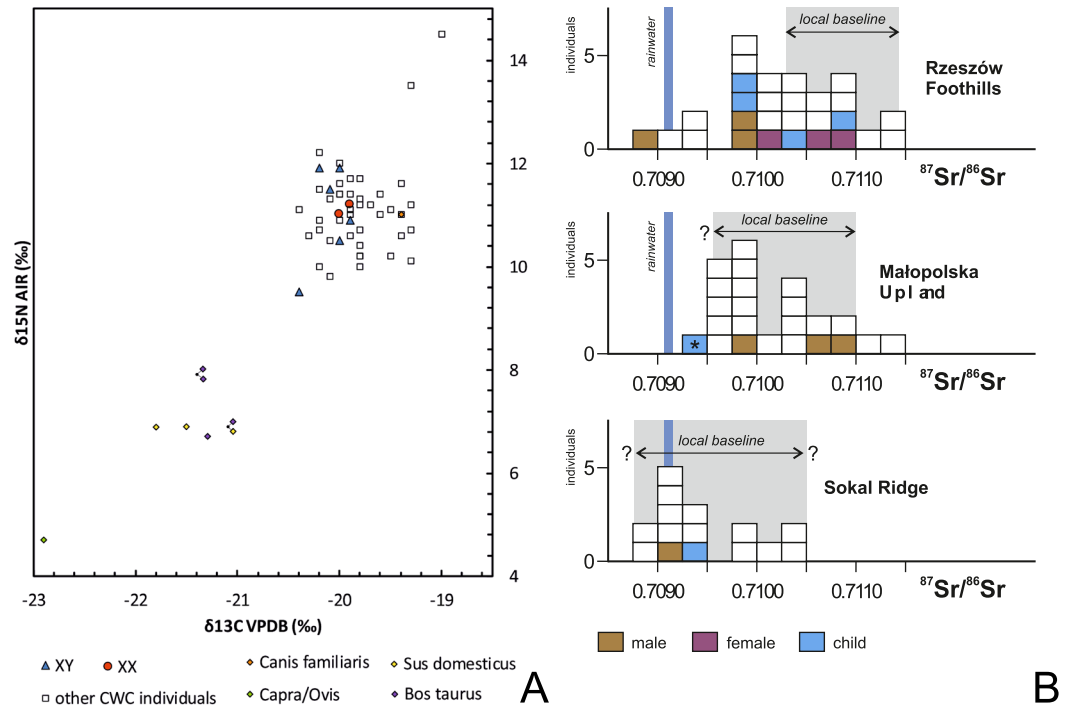


**Figure 1.** The relief map of south-eastern Poland with marked location of archaeological sites of the CWC and BBC cultures. 1 – Kraków-Mistrzejowice, 2 – Bosutów, 3 – Proszowice, 4 – Pełczyńska, 5 – Łubcze, 6 – Mirocin, 7 – Szczytna, 8 – Chłopice, 9 – Skołoszów, 10 – Święte. Reconstruction of niche grave 360 from Mirocin, site 27 (drawn by K. Rosińska-Balik).

Site/feature	Atlas ID	Site grouping	Sr Grouping	Genome coverage	Mol. Sex	MtDNA Hg	Y chromosome Hg	Lab no.	14 C (BP)
Szczytna 6/84	pcw110	Rzeszów Foothills (CWC) RFCWC <b>Group I</b>	Rzeszów Foothills (CWC) <b>Group Ia</b>	0.04	XY	T1	R1b1a1a2		
Chłopice 26/11	pcw212			0.12	XX	H2a2b	—	Poz-90881	3985 ± 35
	pcw211			0.96	XX	H2a2b	—		
Święte 15/408b	pcw062			0.02	XX	HV2a	—		
Mirocin 27/360	pcw160		0.04	XY	I2	NA	Poz-54043	3870 ± 35	
Święte20/40 A	pcw040		2.67	XY	U5b2b1a1	R1b1a1a2	Poz-90777	3950 ± 35	
Święte 20/43/I	pcw041		0.31	XY	J1c2c	R1b1a1a2	Poz-90778	3950 ± 35	
Święte 15/408a	pcw061		0.77	XX	H7a	—	Poz-90780	3890 ± 35	
Święte 11/876	pcw070		0.79	XY	I2	R1b1a1a2	Poz-90875	3890 ± 35	
Skołoszów 7/256	pcw191		0.09	XX	T2b11	—	Poz-55335	3830 ± 35	
Łubcze 2/2	pcw350	Sokal Ridge (CWC) SRCWC <b>Group IV</b>	Sokal Ridge & nonlocals from the Rzeszów Foothills (CWC) <b>Group IVa</b>	0.23	XY	T1a1	R1b1a1a2	Poz-90898	3865 ± 35*
Łubcze 25/3	pcw361			3.28	XY	K1a4b	R1b1a1a2a1a		
	pcw362			5.47	XY	U5a1a2a	R1b1a1a2a1a	Poz-90899	3875 ± 35*
Mistrzejowice 85/1311	pcw250			Malopolska Upland (CWC) MUCWC <b>Group II</b>		0.08	XY	T1a1	NA
Proszowice 1/2	pcw420	0.03	XY			L3c'd	R		
Bosutów/1	pcw430	0.13	XY			U4b1b2	R		
Pełczyńska 6/12	pcw260	Malopolska Upland (BBC) MUBBC <b>Group III</b>		1.59	XX	T2b	—	Poz-34734	3830 ± 35
Pełczyńska 6/13	pcw270			0.08	XX	H	—		
Pełczyńska 6/25	pcw280			0.41	XX	U5a1	—		

**Table 1.** Basic sample info, stats, division, dating, isotopes. Cells highlighted in grey represent non-local individuals from Rzeszów Foothills which constitute Group V. \*unpublished earlier.

possible and occasionally from first premolars (P1) (Supplementary Information – Strontium Isotopes; Table S4, Fig. 2). Moreover carbon ( $\delta^{13}\text{C}_{\text{coll}}$ ) and nitrogen ( $\delta^{15}\text{N}_{\text{coll}}$ ) stable isotope analyses were performed for 8 individuals (Supplementary Information – Stable isotopes – diet; Table S5, Fig. 2A). We investigated patterns in the genetic variation in relation to geographical subgroupings based on archaeological information but also the genetic data in relation to the other results obtained through the stable isotope (Sr) analyses and radiocarbon dating. The groups are listed in Table 1. In short, we first divided individuals based on geographical and archaeological context into four groups (Groups I-IV), and in addition introduce a division based on results from the strontium



**Figure 2.** (A) Variation in carbon and nitrogen isotope compositions of the bone collagen of humans and animals of the Corded Ware Complex from south-eastern Poland. Individuals with DNA profiles (XX, XY) are indicated. (B) Strontium isotope composition ( $^{87}\text{Sr}/^{86}\text{Sr}$ ) of human enamel of the CWC populations from the Rzeszów Foothills, Małopolska Upland and the Sokal Ridge. A child from the BBC grave at Pełczyńska is asterisked. Individuals investigated during the present study are coloured. Other isotopic data and local baselines (in grey) are from Belka *et al.*<sup>41</sup>, Szczepanek *et al.*<sup>29</sup> and results of unpublished investigations of the authors. The Sr isotope composition of rainwater is indicated.

analyses (Groups Ia, IVa and V): For more details regarding tested groupings and obtained results, see section *Grouping individuals according to their provenance (Sr isotopes)*.

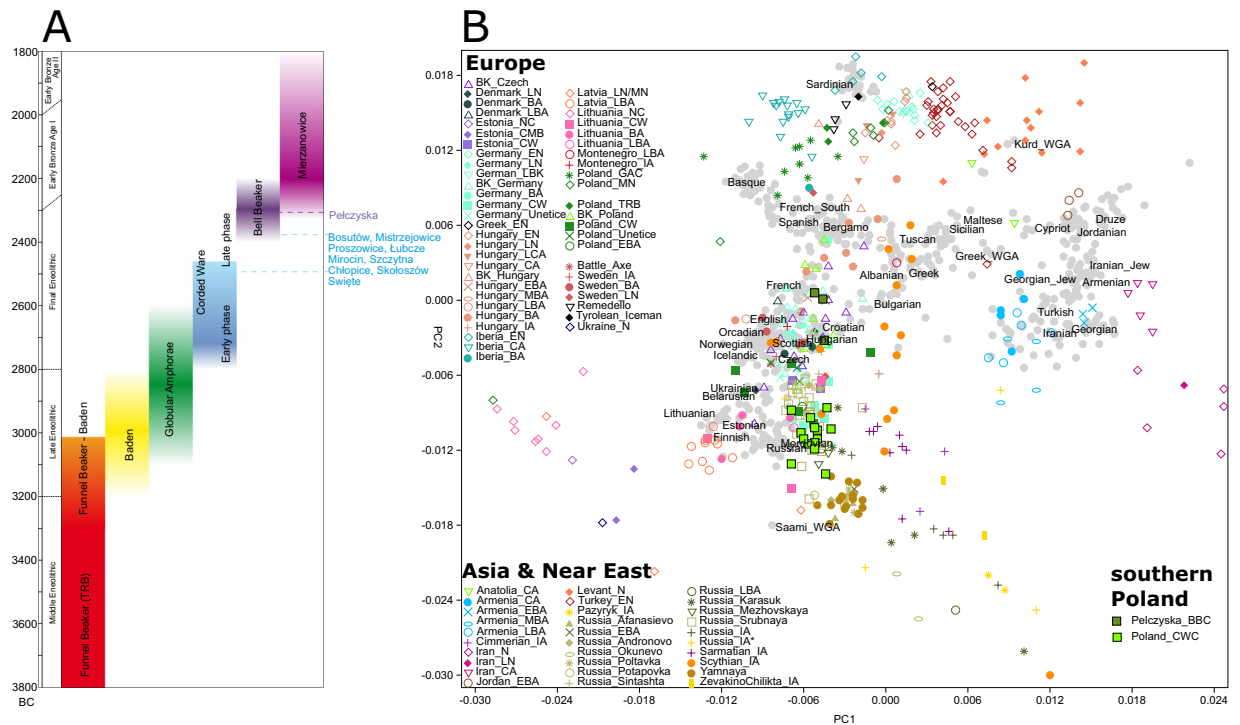
## Results

We collected bone samples from 53 individuals and produced low to medium coverage whole genome sequence data with coverages ranging between 0.02x to 5.4x for 19 Final Eneolithic/Bronze Age individuals from the ten sites in southern Poland. Ten of 19 sequenced individuals were radiocarbon dated from  $3985 \pm 35$  to  $3830 \pm 35$  BP (Fig. S14, Tables S2, S7). Obtained genetic sequences displayed cytosine deamination patterns characteristic of ancient DNA (Fig. S17)<sup>8,30–32</sup>. Mitochondrial (mtDNA) DNA-based contamination levels were calculated based on private polymorphism characterization (13 individuals with sufficient mtDNA coverage)<sup>33</sup>, and sequence mapping likelihood estimation<sup>34</sup>. The obtained contamination estimates varied between 0–3.6% (95% CI of 0–7.3%) and all individuals carried sequences with >90% (>97% in 14 individuals) probability of being authentic (Table S10). Thus, the obtained genomic data were deemed authentic and all sequenced individuals were included in population analyses.

All Eneolithic individuals from Poland carried mtDNA lineages of European or West Eurasian origin<sup>35</sup> including H (including H2 and H7), HV, I2, J1, K1, T1, T2, U4, and U5 (Table S8). Individuals from Pełczyńska exhibited the same mtDNA haplogroups identified at other CWC sites. In contrast to individuals of the Yamnaya complex the CWC and BBC individuals from southern Poland carried I2 and J1 lineages, but lacked the mtDNA haplogroups W and U2, often found in Yamnaya individuals<sup>1,10,36</sup>.

Molecular sex was assigned in all 19 individuals<sup>37</sup> of which ten were sub-adults and therefore lacking prior osteological sex assessments. Eight individuals were female (XX) and 11 individuals were male (XY). The Y chromosome haplogroup was assigned in nine males of which all belonged to macrohaplogroup ‘R’ (Table S9). In seven individuals the Y chromosome haplogroup was further narrowed down to lineage R1b-M269 or R-L11 characteristic of Yamnaya and Bell Beaker individuals<sup>5,10</sup> and particularly widespread throughout Eurasia since the Bronze Age<sup>38</sup>.

In order to investigate mutual relations between individuals we employed conditional nucleotide diversity estimates which is calculated between all pairs of individuals investigated in this study<sup>7</sup> (Table S11). Here an average number of mismatches between pairs of individuals was estimated based on sites in Human Origins dataset and in Yoruban individuals from 1000 Genomes Project (Table S11)<sup>39</sup>. The results did not reveal strong structuring between sites but highlighted closer relationships between a number of individuals (pcw040-pcw041, pcw061-pcw062 and pcw211-pcw212). Additionally, reduced conditional diversity was observed between an individual from Proszowice (pcw420) and individuals from Święte (pcw062, pcw110) and Skołoszów (pcw191).

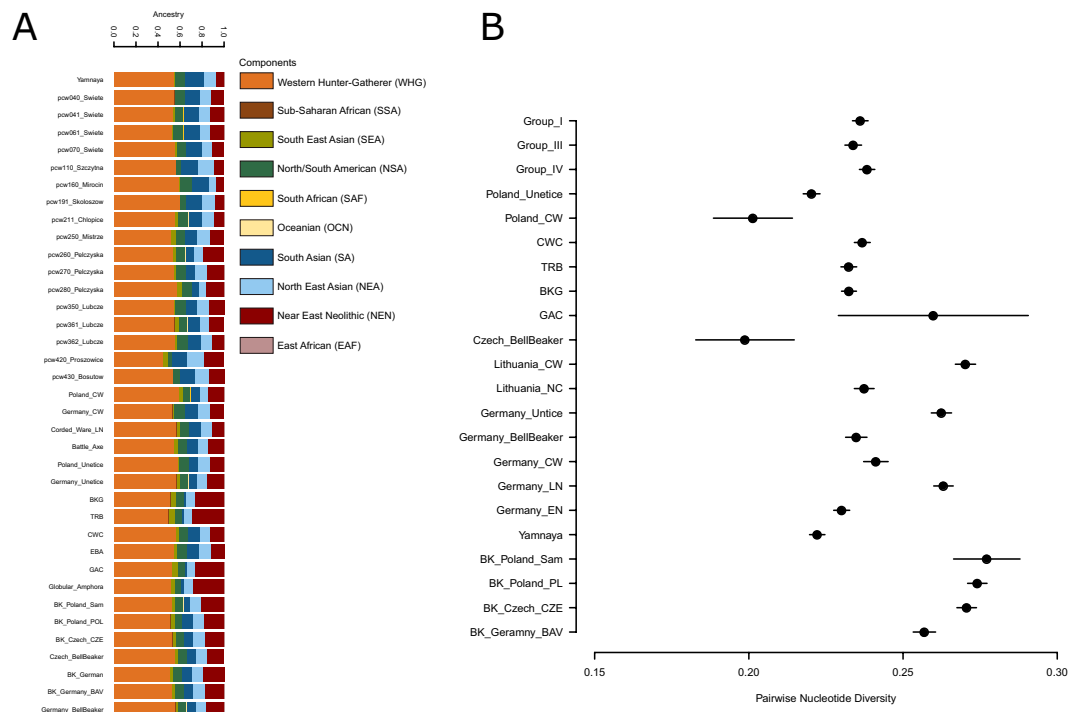


**Figure 3.** Cultural timeline approximately visualising occurrence of different cultural complexes and their dates with the southern Poland samples marked on the left (A); PCA visualising genetic variation of modern European populations marked as grey dots while ancient samples (spanning Neolithic to Iron Age) are following key presented in the legends (B).

**Principal Component Analysis (PCA).** In order to investigate and visualize genetic relations between our Polish Corded Ware individuals and both present-day and ancient populations we performed PCA on the autosomal genomic data (Figs. 3B, S21, S22). The PCA suggests that (a) despite geographical proximity there is a distinct genetic separation between CWC and BBC individuals from Southern Poland. (b) the genetic variation of CWC individuals from southern Poland overlaps with the majority of the published CWC individuals from Germany while the eight published CWC individuals from Poland<sup>10,11</sup> show a closer similarity to BBC representatives (Fig. S20) (c) the genetic variation of BBC individuals from southern Poland overlaps with the broad variation of BBC individuals from Central Europe (Czech Republic, Germany, Poland and Hungary) (Fig. S21). Based on the archaeological, geographical and genetic results we further ran tests according to cultural division (two groups: CWC and BBC) and geographical distribution of sites (four groups: Group I-IV) (Table 1).

**Admixture.** In order to trace ancestral whgs of the samples we performed ADMIXTURE analyses between 2K and 10K using 10 replicates per run with the total of 214 world-wide modern populations and 485 ancient individuals grouped into 123 populations/individual samples. As expected the Admixture analyses showed that Polish Eneolithic individuals were carriers of three major components: West and North European HG (WHG; orange), Near East Neolithic (NEN; red), blue and green components of Asian origins (SA – South Asian - navy blue; NEA - North East Asian - light blue; SEA – South East Asian - light green and NSA –North/South American – dark green). In that respect, they were similar to earlier published Admixture analyses<sup>11</sup> and revealed that individuals from Święta, Szczytna and Miocin were carriers of larger NEN and SEA components than the representatives of Yamnaya. Polish CWC individuals also had traces of yet another component which is most prominent in Sub-Saharan and other modern African populations. According to earlier research the component was found to be more evident in Neolithic European populations than those with Steppe ancestry. In terms of differences between groups of individuals it seems that members of Group I carried larger proportion of SA and less of SEA component, while individuals in Groups III and IV had larger SEA and NEN components. This variation is mirrored by variation and component distribution patterns observed in previously published individuals from central European Neolithic (Figs. 4A, S23). Our findings point to variable local admixture patterns between earlier Neolithic populations from southern Poland and incoming Steppe nomads as well as structuring between CWC groups from different parts of present-day Poland.

**$f_2$ ,  $f_3$ - and  $f_4$ -statistics.** We tested with outgroup  $f_3$ -statistics, which ancient population/individuals shared more genetic drift with the ancient samples generated here individually. The different population size between the reference ancient population/individuals and the lack of significant results between closely related individuals and populations make it difficult to extract conclusions from these analyses.



**Figure 4.** A pruned visualisation of the admixture run of the whole dataset at  $K = 10$  (Fig. S23). The bar plot shows newly published data from Poland as well as previously published individuals from Poland, Yamnaya, Battle Axe cultures and Corded Ware, Bell Beaker individuals from Germany and Czech Republic (A). Conditional nucleotide diversity in Groups I, II and IV compared to diversity estimates from other closely related population groups (B).

Therefore, we grouped our ancient individuals into four groups (groups I-IV) based on the aforementioned geographical distribution of the sites and compared these groups between them and also with the ancient reference individuals and populations. Ancient reference individuals were merged into groups following the same criteria as in the ADMIXTURE analyses. Overall, although non-significant the results suggested a trend where the four groups share more genetic drift with Russia\_Afanasievo than with Yamnaya and Groups I, II and III share more genetic drift with Poland\_CW than with Russian\_Afanasievo (Table S14 & Fig. S20). This pattern was also mirrored by the  $f_2$ -statistics.

To test for structuring in our sample set we performed multiple  $f_4$ -statistics of a form  $f_4(O, X; A, B)$  where O is the outgroup (Yoruba), X represent our test Eneolithic individuals divided into groups I-IV, and A & B represent populations tested for admixture. First, we tested for mutual relations between Polish Eneolithic groups from this study. While most of the results were statistically insignificant, we observed that group II was closer to group I & IV to the exclusion of group III ( $Z > |2|$ ) confirming that Pelczyska would be a distinct population group. We observed that the significant  $f_4$  statistics were largely consistent with the grouping pattern on PCA (Table S16). When tested with other published European Neolithic populations we observed similar pattern in Late Eneolithic and Final Eneolithic populations, but not earlier ones, e.g. Globular Amphora. Groups I-IV are genetically closer to Russia\_Afanasievo than to Yamnaya individuals ( $Z > |2.9|$ ) and are also closer to Russia\_Afanasievo than to each other ( $Z > |2.9|$ ). Even when we compare two individuals from the same group, same chronology and same site with Russian Afanasievo with  $f_4$ -statistics they share significant more genetic drift with Russian Afanasievo than with each other. These results indicate that there might be a bias in  $f_3$ - and  $f_4$ -statistics towards Russian Afanasievo. We do not observe this bias in the  $f_2$ -statistics where individuals within groups are closer to each other than to Russian Afanasievo as expected. Due to low sample size of Afanasievo group ( $n = 5$ ), which was further reduced by removing RISE507, likely the same as RISE508<sup>11</sup> the observed affinity to the Russian Afanasievo should be interpreted with caution. To further investigate our observations and trace any bias that might result from the way reference data was generated, we perform the same  $f_4$ -statistics calculation with division of Yamnaya dataset into whole-genome sequenced dataset (WGS) and enriched dataset (CP), we only trust the results of  $F_4$  stats using WGS Yamnaya data. We find that we still observe similar results to those obtained while using mixed reference dataset albeit with more similarities within WGS generated dataset (Table S16). Selected  $f_4$ -statistics results are listed in Table S16.

**Kinship analyses (READ).** In order to check for kinship relations between individuals, we used Relationship Estimation from Ancient DNA (READ) between all pairs of individuals tested<sup>40</sup> (Table S12). We identified two pairs of first-degree relatives in Święte (between the male pcw040 and boy pcw041 as well as the female pcw061 and the female pcw062), and a pair of second-degree relatives in Chlopice (the children in pcw211 and pcw212)

(Tables S6). In further population based statistical analyses we exclude the individuals with lower coverage genomes from each related pair (i.e. pcw041, pcw062 and pcw212).

**Pairwise Nucleotide Diversity.** We used pairwise nucleotide diversity to test for differentiation between pairs of individuals, groupings (pcw040/pcw070, pcw260/pcw280 and pcw361/pcw362 representative of Groups I, III and IV respectively) and compared the values with diversity estimated in between pairs of individuals representing various Neolithic cultural groups (Figs. 4, S2). On individual level, we found that pairwise nucleotide diversity is in line with results from READ. In other words, we pickup closer relations between three pairs of individuals which were also found to be related according to READ analyses (pcw04/pcw041, pcw061/pcw062 and pce211/pcw212). On group level, we discovered that nucleotide diversity is similar between the groups and is on par with diversity found in other Eneolithic cultures (Fig. 4B). We do not include Group II as it consists of individual from different sites and could result in artificially inflated diversity estimates.

**Functional SNPs.** Finally, we selected the seven individuals with best genome coverage (pcw040, pcw061, pcw070, pcw211, pcw361 and pcw362), to test a number of functional SNPs, especially those associated with lactase persistence, eye and hair colour and a number of traits associated with Asian ancestry (Table S17). None of the individuals was a carrier of Asian ancestry alleles nor was any of the individuals lactose tolerant (Table S17). For those individuals who had enough coverage at pigmentation associated SNPs allowing for HirisPlex prediction, they were predicated to have been brown-eyed and dark (brown or black) haired (Table S18).

**Grouping individuals according to their provenance (Sr isotopes).** *Local and nonlocal individuals.* According to  $^{87}\text{Sr}/^{86}\text{Sr}$  values some of individuals buried at sites in the Rzeszów Foothills are non-locals with signatures below 0.7103<sup>41</sup> (Fig. 2B). While the dietary patterns of those individuals are consistent with CWC background, the archaeological and isotopic data indicate that their place of provenience (e.g. where they spent their childhood) could have been partly the Sokal Ridge or Ukraine. Using these data, and to test possible relations between the groupings of individuals we performed  $F_{ST}$  analyses using popstats<sup>42</sup> between: A) the four groups as used for analyses in this paper (Groups I-IV); B) new grouping of four in which individuals from the Rzeszów Foothills were divided into local and non-local individuals, and where the latter were combined with individuals buried at sites in the Sokal Ridge (Group Ia, Group II-III and Group IVa); C) finally, we repeated the analyses by dividing the whole dataset into five groups (Group Ia, Group II-V), where group V consisted only of non-local individuals from the Rzeszów Foothills. We tested all possible groupings using SNP panels merged both with Human Origins and 1000 Genomes Project (Fig. S19). We find that inclusion of non-local individuals from Rzeszów Foothills in Group I makes the distance between that and the other groups lesser (Fig. S19A) than when the individuals are included in the alternative Group IVa (Fig. S19B). The general pattern of genetic distance remains the same and the results are reproducible between the two reference datasets used (Human Origins and 1000 Genome Project). When the dataset was divided into five groups (with individuals from Rzeszów Foothills divided into local and non-local group) we indeed observed that non-local individuals from Rzeszów Foothills are closest to individuals from the Sokal Ridge and both present very similar patterns in relation to other groups (Fig. S19C). However, since we detect no evidence of Group V constituting a uniform well-defined subgroup based on nucleotide diversity tests (Fig. S18) we find that grouping A, based on geography and archaeological context is most relevant for downstream genetic analyses and is thus used across different analyses.

## Discussion

All individuals studied here are associated to burials of the CWC and BBC complexes. Based on the geographic location and the strontium isotope analysis we identified four different territorial sub groups, and based on the genetics we identified at least two different groups within our sample set. The mtDNA and the Y-chromosome data provide a slightly different picture of the genetic variation in the region contrasting earlier studies of individuals from corresponding archaeological contexts from other regions of Central Europe. In contrast to observations by Juras *et al.*<sup>12</sup> we did not find mitochondrial lineages specifically linked to Yamnaya pastoralists, instead most of mtDNA lineages found in our sample may be associated with European Neolithic farming groups as is the case for the Western Corded Ware sample in the earlier study<sup>12</sup>. Our results would indicate a stronger continuity with earlier Neolithic populations than previously observed. In other words, our study detected traces of an evident “incorporation” of local individuals into the migrating groups. However, the funerary rituals seem to have been affected in limited extent as the burials exhibit the typical CWC pattern in all cases examined. The Y chromosome haplogroup lineage R1b-M269 or R-L11 are characteristic of Yamnaya and Bell Beaker individuals<sup>5,10</sup> and they were particularly widespread throughout Eurasia in the Bronze Age and thereafter<sup>38</sup>. Curiously, the haplogroup is uncommon among other published Corded Ware Complex individuals from Europe (Germany, Poland, Bohemia, Estonia, Lithuania)<sup>6</sup> and is associated with the later Bell Beaker communities<sup>5</sup>. We see the inclusion of the Yamnaya genetic signals but again in a different manner than what has been shown in adjacent regions. These results indicate a higher level of CWC continuity with earlier Neolithic individuals than those previously studied. The result also shows that the CWC groups exhibit an influence of the Steppe world, i.e. in the individuals with specific Y chromosome. Later the influence of the BBC communities was stronger.

The PCA revealed that despite geographical proximity there is a distinct genetic separation between CWC and BBC individuals from southern Poland. The genetic variation of CWC individuals from southern Poland overlaps with the majority of previously published CWC individuals from Germany while the eight published CWC individuals from the Polish lowland<sup>10,11</sup> more closely resemble BBC individuals (Fig. S21). This fact is not unexpected if we consider the CWC communities in Polish lowlands as representatives of north-western parts of the CWC world called as the Single-Grave culture (see supplementary information). The genetic variation of BBC individuals from south-eastern Poland overlaps with the broad variation of BBC individuals from Central

Europe (Bohemia, Moravia, Germany, south-western Poland and Hungary) (Fig. S22) which corresponds well with archaeological data. The results are in line with Admixture analyses. To determine whether the structuring can be detected on even more regional level we divided our sample into regional subgroups to test their relations.

According to  $f_4$ -statistics on individual level CWC individuals from south-eastern Poland are equally different to published CWC and Yamnaya individuals, while the three Pełczyska individuals tend to select German, Estonian, Lithuanian and Polish western CWC to the exclusion of Yamnaya (Table S16). This however is not the case for CWC individuals from the Kuyavia region published by Fernandes<sup>11</sup>. CWC (but not BBC) from south-eastern Poland tends to pick Yamnaya over CWC from Kuyavia region. Our results emphasize the different impacts the Yamnaya migration event had on different populations across Europe, i.e. the genetic legacy that the Yamnaya process left varies greatly between regions and cultures.

We obtained statistically significant values for group II being closer to group I & IV to the exclusion of group III confirming that BBC Pełczyska are a distinct population group. This suggests structuring not only within the Eneolithic in southern Poland but also between groups representing the same (Archaeological) cultural horizon. Compared to the Early and Middle Neolithic samples it seems that the CWC groups I, II and IV are equally distant from the Yamnaya pastoralists and from most of the earlier published Corded Ware groups from Estonia, Germany, Lithuania and central Poland (Table S16). The south-eastern Polish CWC individuals are significantly more closely related to Yamnaya than to CWC individuals from the Polish lowland supporting the differentiation between various CWC groups from Central Europe. This is in coincidence with archaeological finds that show differences between lowlands and uplands materials of CWC<sup>23</sup>. Bell Beaker individuals from Pełczyska mostly favour German, Polish (lowland) and Estonian CWC as well as German and Czech Bell Beaker populations over Steppe ancestors (Table S16). Interestingly, in contrast to CWC individuals from south-eastern Poland (Group I, II and IV), they share significantly closer affinity to Neolithic Iberian, Italian, Hungarian, Swedish, Polish TRB and Brześć Kujawski group populations (and nonsignificant but positive affinity to Polish Globular Amphora) than Yamnaya, pointing to possible continuity between this group and earlier populations. The genetic specificity of the population associated with this process shows similarity to the features of the BBC complex in Central Europe dated ca. 150–200 years later<sup>5</sup>.

Building on the idea that the CWC complex identity is founded on the type of burial rituals performed, the study of graves and the double burials in particular can be illuminating for interpretations of the local social structures of the CWC groups regarding possible kinship structures. Positive results were obtained from 3 double graves, all containing individuals of the same sex. Kinship was observed in the graves from Chłopice grave 11 (pcw211 and pcw212) and from Święte, grave 408 (pcw061 and pcw062). The former burial represented a second-degree kinship while the latter was a first degree one. The two young boys buried in the double grave from Łubcze (pcw361 and pcw362) were not closely related according our READ analysis. However, their similar Y chromosome haplogroup (Haplogroup R1b1a1a2a1a) leaves room for a possible shared ancestry on the male lineage. In Chłopice both young females died and were buried at the same time but the cause of their death could not be established. Skeletons from Łubcze were too badly preserved to provide information on time or cause of death. Interestingly in Święte (grave 408), the younger female (pcw061) probably had died earlier than the older one (pcw062) as her remains probably were exhumed from another place and added to the grave with the older female. The niche construction made “revisits” and secondary depositions possible, however this practice was part of an uncommon funeral rite in which the kinship between the two females might have had relevance. Interestingly, these closely related females spent their childhood in different places as according to strontium isotopes analysis pcw061 was a non-local and pcw062 – a local individual in the area. The observations are particularly surprising and significant considering the CWC burial customs and also social organisation. They show that in some cases close relatives were buried together in spite of different time of death. Closely related individuals were also found in graves 40 A (pcw040) and 43 (pcw041) in Święte, although not placed in a double grave they had been placed in burials in close proximity of each other. Both individuals were non-locals. The first-degree kinship identified here most likely is the father and his son, as they both have the same radiocarbon date. The close proximity of the graves indicate that kinship was an important part of the society and may have been manifested even in the funeral customs of the Corded Ware Complex communities.

## Conclusion

Social processes in prehistory are hard to identify let alone interpret. Using ancient DNA and genomic analysis we can detect several levels of structuring and population dynamics within the CWC complex as well as between the CWC and BBC cultures present in Eneolithic south-eastern Poland. By evaluating the admixture between groups, the impact of earlier known demographic events like the Steppe expansions can be detected in our CWC individuals. Not surprisingly we also discover strong connections to the population in Central Germany and the CWC subpopulation living on the German lowlands. These ties are not new but they might indicate an interesting southern affiliation that has not been as clear in other previously studied samples from Poland. Furthermore, we detect less association between CWC groups from the Kuyavia region and southern Poland revealing fine scale routes of the spread of CWC traditions. The most unusual signal identified is the one between the CWC and the Afanasievo complex. This genetic incorporation from a Steppe population further east than the Yamnaya culture, is novel for these parts and suggests a CWC population structure and history more complex than previously thought. Although, our results should be treated with caution due to not just low number of samples, but the appearance of the same signal in individuals that predate steppe expansions, and are geographically more widespread. Our findings are in alignment with recent archaeological reviews suggesting lesser impact of Yamnaya event than estimated in earlier genomic studies (e.g.<sup>16,17</sup>). The CWC ancestry exhibits links both to common mtDNA lineages of the initial Neolithic but also to those assimilating and replaced by the Yamnaya pastoralists. Moving further forward in time we detect genetic reminiscence of south-eastern CWC in BBC gene pool. This region was an important social area in the 3<sup>rd</sup> Millennium BCE, a true prehistoric melting pot of human



groups with different origin, which may have witnessed emergence of typical BBC genomics almost 200 years earlier than in other parts of Europe<sup>5</sup>.

By using strontium isotopes, we can detect locals and non-locals in our material, which can shed light on how different subgroups within the CWC interacted. Using these characterisations as a foundation allows for deeper look into kinship amongst the buried individuals. The identification of three different types of kinship amongst the individuals studied and especially in the context of an unusual burial customs expands our understanding of rituals and traditions underlying the social structures of the CWC complex.

## Methods

**Isotopic analyses.** All carbon ( $\delta^{13}\text{C}_{\text{coll}}$ ) and nitrogen ( $\delta^{15}\text{N}_{\text{coll}}$ ) stable isotope analyses were performed at the Isotope Dating and Environment Research Laboratory at the Institute of Geological Sciences of the Polish Academy of Sciences in Warsaw<sup>43</sup>. Collagen was earlier extracted for AMS  $^{14}\text{C}$  dating in the Poznań Radiocarbon Laboratory. Stable isotope composition was determined using a Thermo Flash EA 1112HT elemental analyser connected to a Thermo Delta V Advantage isotope ratio mass spectrometer in a Continuous Flow system. The results of the carbon and nitrogen isotope analyses are shown in Fig. 2A and are listed in Table S5 (Supplementary Information). The stable isotope variation of the CWC individuals was compared to earlier published data on other CWC individuals from south-eastern Poland<sup>29</sup>. The Sr isotope analyses for 16 individuals were carried out in the Isotope Laboratory of the Adam Mickiewicz University at Poznań, Poland. The procedure included chemical separation of Sr and measurements of Sr isotope ratios. Strontium was analysed in dynamic collection mode on a Finnigan MAT 261 mass spectrometer. During this study, the NIST SRM 987 Sr standard yielded  $^{87}\text{Sr}/^{86}\text{Sr} = 0.710236 \pm 12$  ( $2\sigma$  mean on 10 analyses). Total procedure blanks were less than 80 pg. The  $^{87}\text{Sr}/^{86}\text{Sr}$  values were corrected to  $^{86}\text{Sr}/^{88}\text{Sr} = 0.1194$ . The Sr results for samples were normalized to certified value of NIST SRM 987 = 0.710240 (Fig. 2B; Supplementary Information – Strontium Isotopes; Table S4).

**Ancient DNA extraction.** 53 archaeological bones were decontaminated prior to analysis by wiping them with a 1% sodium hypochlorite solution and DNA-free water. Furthermore, all surfaces were UV irradiated ( $6\text{ J}/\text{cm}^2$  at 254 nm). After removing 1 mm of the surface, a piece of the petrous bone was cut (50–150 mg) after which the bone was placed in an Eppendorf tube and left to dissolve in 1 ml extraction buffer (0.5 M EDTA and 10 mg/ml Proteinase K), at 55 °C in hybridization oven. The buffer was changed twice and all the aliquots were collected and combined. After one-week DNA was extracted following silica-based methods as in<sup>44</sup> with modifications as in<sup>45,46</sup> and eluted in  $2 \times 55 \mu\text{l}$  of EB buffer (Qiagen). One extraction was made from each sample and one extraction blank with water instead of bone powder was included per six to 10 extracts. Blanks were carried along the whole process until quantitative PCR (qPCR) and/or PCR and subsequent quantification.

**Library preparation and sequencing.** DNA libraries were prepared using 20  $\mu\text{l}$  of extract, with blunt-end ligation coupled with P5 and P7 adapters and indexes as described in<sup>46,47</sup>. From each extract one double-stranded library was built. Since aDNA is already fragmented, the shearing step was omitted from the protocol. Library blank controls (extraction blanks), were carried along during every step of library preparation. In order to determine the optimal number of PCR cycles for library amplification, qPCR was performed. Each reaction was prepared in a total volume of 25  $\mu\text{l}$ , containing 1  $\mu\text{l}$  of DNA library, 1X MaximaSYBRGreen mastermix, and 200 nM each of IS7 and IS8<sup>47</sup> reactions were set up in duplicates. The amplification reactions had a total volume of 25  $\mu\text{l}$ , with 3  $\mu\text{l}$  DNA library, and the following in final concentrations: 1 X AmpliTaq Gold Buffer, 2.5 mM  $\text{MgCl}_2$ , 250  $\mu\text{M}$  of each dNTP, 2.5 U AmpliTaq Gold (Thermo Fisher Scientific, Waltham, MA), and 200 nM each of the IS4 primer and index primer<sup>47</sup>. PCR was done with the following conditions: an activation step at 94 °C for 10 min followed by 8–20 cycles of 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 45 s, and a final elongation step of 72 °C for 10 min. For each library, six amplifications with the same indexing primer were pooled and purified with AMPure XP beads (Agencourt; Beckman Coulter, Brea, CA). The quality and quantity of libraries was checked using BioAnalyzer using the High Sensitivity Kit (Agilent Technologies, Cary, NC). None of the blanks showed any presence of DNA comparable to that of a sample and were therefore not further analysed. For initial screening, 10–20 libraries were pooled at equimolar concentrations for sequencing on an Illumina HiSeq X ( $2 \times 150\text{ bp}$ , PE) at NGI Stockholm. After evaluation of factors such as clonality, proportion of human DNA, and genomic coverage samples were selected for resequencing, aiming to yield as high coverage as possible for each library. In this study we used single indexed libraries. While we acknowledge the issues that might be caused by index-hopping on Illumina HiSeq X flowcell, we strive to monitor for any issues by screening for presence of contamination

**Sequencing data processing.** The sequencing reads were demultiplexed according to the sample associated index sequence. Paired-end reads were merged requiring an overlap of minimum 11 bp and minimum mapping quality of 30 and thereafter trimmed using MergeReadsFastq\_cc.py<sup>48</sup>. Merged reads were mapped to the human reference genome (hs37d5) using BWA (v 0.7.8) with non-standard parameters (-l 16500 -n 0.01 -o 2)<sup>49</sup>. Files representing different sequencing runs were merged using samtools (v 1.8)<sup>50</sup>. Sequences with more than 10% mismatches, shorter than 35 bp and representing PCR duplicates were removed using FilterUniqSAMCons\_cc.py<sup>48</sup>. Using R<sub>s</sub> method which is based on the proportion of reads mapping to sex chromosomes genetic sex of the individuals were identified<sup>37</sup>. We re-mapped the published whole genome sequenced and SNP-captured ancient genome sequence data using the same parameters.

**Damage & Contamination.** PMDtools<sup>8</sup> was used in order to verify presence and quantify the proportion of 5' cytosine deamination patterns which are characteristic of ancient DNA<sup>30,32,51,52</sup>. All individuals displayed damage patterns consistent with presence of ancient DNA (Fig. S17). To further evaluate data integrity, we estimated contamination levels by the analysing polymorphic site distribution in mitochondrial sequences and by investigating the likelihood of mitochondrial sequences mapping to more than one reference<sup>33</sup> (Table S10). In addition

to this, we estimated the degree of contamination in our samples using a likelihood-based method implemented in *contamMix* software<sup>34</sup> that checks the matches of reads with 311 reference mitochondrial genome sequences from different worldwide populations.

**Mitochondrial DNA analysis.** Sequencing reads with base quality of minimum 30 and aligning to mitochondrial DNA were filtered out using *Samtools* (v. 1.8)<sup>50</sup>. Consensus mitochondrial sequences were called using *mpileup* and *vcfutils.pl* (*vcf2fq*). Mitochondrial haplogroups were identified using *HAPLOFIND*<sup>53</sup> and *mthap* v. 0.19b (<https://dna.jameslick.com/mthap/>) based on SNPs reported in rSRS and rCRS reference sequences respectively. All mutations and resulting haplogroup calls are listed in the supplement (Table S8).

**Y chromosome analysis.** Sequencing reads with base quality of minimum 30 and aligning to Y chromosome were filtered out using *Samtools* (v. 1.8)<sup>50</sup>. Y chromosome sequences were called using *mpileup*. The resulting pileup file was then merged with haplogroup definitions from the *PyloTree* Y<sup>54</sup>. Y chromosome haplotype calls were reported using both *PhyloTree* Y nomenclature and the nomenclature of the International Society of Genetic Genealogy ISOGG database (<https://www.isogg.org>) v. 11.349 (Table S9). In order to avoid misidentification resulting from DNA damage whenever possible we base our haplogroup calls only on derived transversions.

**Reference datasets.** We prepared 2 different datasets: Pseudo-haploid genotype calls of ancient individuals prepared by randomly choosing one read per position were merged with present day individuals from Human Origins data set<sup>55,56</sup> and with 108 Yoruban individuals of the 1000 Genomes Project phase 3 data<sup>39</sup>. Human Origins dataset was used in PCA and ADMIXTURE analyses while 1000 Genomes Dataset was used in *F*-statistics. To prepare Human Origins Dataset, we discovered the genotypes of ancient individuals for a total of 594,924 autosomal SNPs genotyped for 2,730 modern-day individuals from 203 different populations and merged these ancient individuals as well as 485 ancient individuals collected from published data and processed with same parameters applied to our newly generated data (Table S13) with the present-day individuals. To prepare 1000 Genomes dataset, we discovered the pseudohaploid genotypes of ancient individuals for 1,938,919 transversion type SNPs with a minor allele frequency of 10% in Yoruba individuals in 1000 Genomes Project. To prepare this dataset, in brief, we first downloaded the VCF and BAM files of the African Yoruba individuals sequenced in phase 3 of the 1000 genomes project (<ftp.1000genomes.ebi.ac.uk>). Then we used *vcftools* and extracted all SNPs with a minor allele frequency of at least 10% in the Yoruba population. After which we excluded transition type SNPs to prevent potential bias due to postmortem damage in the analysis and lead us to retain a total of 1,938,919 SNPs<sup>8,46</sup>. For both datasets, the analyses were restricted to nucleotide positions with minimum mapping and base quality of 30. Indels and transition were excluded to avoid bias due to post-mortem damage.

**READ analysis.** In order to identify possible kinship relations between individuals and to be able to exclude related individuals from downstream population analyses we performed READ (Relationship Estimation for Ancient DNA) analysis<sup>40</sup>. In order to guarantee the highest possible number of overlapping SNPs between individuals to ensure highest possible resolution of the analyses we merged our ancient individuals with a sub-sampled set of ten fully sequenced individuals from Estonian Genome Diversity Project (EGDP)<sup>57</sup>. The analyses were restricted to transversions only. After identification of three kin pairs (Table S12) from each pair we removed the individual with lower coverage to avoid bias in population analyses.

**Principal Component Analysis.** The principal component analysis was performed using *smartpca* module in EIGENSOFT (v.6.0.1)<sup>58</sup> with a selection of populations from the Human Origins dataset merged with ancient reference material<sup>55,56</sup>. We used *smartpca* with *lsqproj*:YES and *shrinkmode*:YES options and calculated the eigenvectors using only present-day populations. The resulting *vec* files were visualised using *Past* 3.22<sup>59</sup>.

**ADMIXTURE analysis.** In order to identify ancestral clusters, we performed unsupervised clustering using ADMIXTURE<sup>60</sup>. For this analysis, we used our data merged with the published ancient individuals as well as modern individuals from Human Origins dataset. Prior to analysis, we filtered the Human Origins dataset for linkage disequilibrium using *PLINK*<sup>61</sup> with parameters “indep-pairwise 200 25 0.4”. To prevent bias due to difference in number of overlapping SNPs between ancient individuals we carried out ADMIXTURE analysis as described in<sup>62,63</sup>. First, we calculated ancestry components for present-day populations. Using the resulting ancestral allele frequencies, we estimated the cluster membership of each ancient individual. We run ADMIXTURE for  $K = 2$  to  $K = 10$  in 10 independent runs with different random seeds for modern individuals and used each of the resulting allele frequency file (P files) to determine clusters of the ancient individuals. We merged the resulting Q files for modern and ancient individuals and used *Pong*<sup>64</sup> in greedy mode to align clusters between different runs and to identify common modes between various runs for each K run. The results were visualised using R (Figs. S22 and S23).

**$f_2$ ,  $f_3$ - and  $f_4$ -statistics.** In order to formally test for admixture, we calculated  $f_3$ -statistics and  $f_4$ -statistics using *qp3Pop* and *qpDstat* options of *AdmixTools* (v.5.0-20171024). The outgroup  $f_3$ -statistics of the form (O; A, B) were used to check for shared genetic affinity between test populations, A and B<sup>55</sup>. We compared our ancient samples with 165 ancient populations/individuals (including the ones generated here) using Yoruban population from the 1000 Genomes data set as outgroup (O). The analyses were performed with 1938916 SNPs and 559 jackknife blocks (Table S14). The  $f_4$ -statistics of the form  $f_4(O, X; A, B)$  were used to check for tracing admixture between tested populations (Table S16). We computed  $F_4$  statistics setting the  $|Z| > = 2$  as threshold. Since we compared population groups which are genetically very close to each other, we set the threshold as  $|Z| > = 2$  and evaluated the genetic affinities between groups. Both  $f_3$ -statistics and  $f_4$ -statistics were calculated separately for all

newly published individuals and for the same individuals grouped according to geographical clustering presented in Table 1. In all analyses, we exclude transition sites.

Finally, we used popstats (<https://github.com/pontussk/popstats>) to calculate nucleotide diversity between pairs of our individuals, FST between different groupings of our individuals (Fig. S18),  $f_2$ -statistics between individuals, their groupings and different populations groups (Table S15) as well as conditional nucleotide diversity between pairs of CWC individuals representing three (out of four) different groups compared to a pairs of individuals representing selected contemporary population samples (Fig. 4B, Table S11)<sup>42</sup>. Standard errors were calculated using block jackknife over blocks of 500 SNPs (Tables S5, Figs. S17, S19, 4B).

## Data Availability

The sequence data was deposited in European Nucleotide Archive (ENA) under the following study accession number: PRJEB34091 (ERP116945).

Received: 6 August 2019; Accepted: 24 March 2020;

Published online: 14 April 2020

## References

1. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207 (2015).
2. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499 (2015).
3. de Barros Damgaard, P. *et al.* 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, 369 (2018).
4. Lipson, M. *et al.* Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* **551**, 368 (2017).
5. Olalde, I. *et al.* The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**, 190 (2018).
6. Wang, C.-C. *et al.* Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. *Nature communications* **10**, 590 (2019).
7. Skoglund, P. *et al.* Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336**, 466–469 (2012).
8. Skoglund, P. *et al.* Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proceedings of the National Academy of Sciences* **111**, 2229–2234 (2014).
9. Loog, L. *et al.* Estimating mobility using sparse data: Application to human genetic variation. *Proceedings of the National Academy of Sciences* **114**, 12213–12218 (2017).
10. Allentoft, M. E. *et al.* Population genomics of bronze age Eurasia. *Nature* **522**, 167 (2015).
11. Fernandes, D. *et al.* A genomic Neolithic time transect of hunter-farmer admixture in central Poland. *Scientific reports* **8**, 14879 (2018).
12. Juras, A. *et al.* Mitochondrial genomes reveal an east to west cline of steppe ancestry in Corded Ware populations. *Scientific reports* **8**, 11603 (2018).
13. Tassi, F. *et al.* Genome diversity in the Neolithic Globular Amphorae culture and the spread of Indo-European languages. *Proceedings of the Royal Society B: Biological Sciences* **284**, 20171540 (2017).
14. Schroeder, H. *et al.* Unraveling ancestry, kinship, and violence in a Late Neolithic mass grave. *Proceedings of the National Academy of Sciences* **116**, 10705–10710 (2019).
15. Klejn, L. S. *et al.* Discussion: Are the Origins of Indo-European Languages Explained by the Migration of the Yamnaya Culture to the West? *European Journal of Archaeology* **21**, 3–17 (2018).
16. Furholt, M. Massive Migrations? The impact of recent aDNA studies on our view of third millennium Europe. *European Journal of Archaeology* **21**, 159–191 (2018).
17. Heyd, V. Kossinna's smile. *Antiquity* **91**, 348–359 (2017).
18. Kristiansen, K. *et al.* Re-theorising mobility and the formation of culture and language among the Corded Ware Culture in Europe. *Antiquity* **91**, 334–347 (2017).
19. Lorkiewicz, W. *et al.* Between the Baltic and Danubian worlds: the genetic affinities of a middle neolithic population from central Poland. *PLoS one* **10**, e0118316 (2015).
20. Sjögren, K.-G., Price, T. D. & Kristiansen, K. Diet and mobility in the corded ware of central Europe. *PLoS one* **11**, e0155083 (2016).
21. de Barros Damgaard, P. *et al.* The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* **360**, eaar7711 (2018).
22. Włodarczak, P. Chronologia absolutna cmentarzysk późno-i schyłkoweolitycznych na Wyżynie Lubelskiej. *Schyłek neolitu na Wyżynie Lubelskiej, Kraków*, 537–548 (2016).
23. Włodarczak, P. Battle-axes and beakers: The Final Eneolithic societies. *The Past Societies. Polish Lands from the First Evidence of Human Presence to the Early Middle Ages* **2**, 5500–2000 (2017).
24. Haber, M., Mezzavilla, M., Xue, Y. & Tyler-Smith, C. Ancient DNA and the rewriting of human history: be sparing with Occam's razor. *Genome Biology* **17**, 1 (2016).
25. Harris, E. E. Demic and cultural diffusion in prehistoric Europe in the age of ancient genomes. *Evolutionary Anthropology: Issues, News, and Reviews* **26**, 228–241 (2017).
26. Kivisild, T. The study of human Y chromosome variation through ancient DNA. *Human Genetics* **136**, 529–546 (2017).
27. Zeng, T. C., Aw, A. J. & Feldman, M. W. Cultural hitchhiking and competition between patrilineal kin groups explain the post-Neolithic Y-chromosome bottleneck. *Nature communications* **9**, 2077 (2018).
28. Bourgeois, Q. & Kroon, E. The impact of male burials on the construction of Corded Ware identity: Reconstructing networks of information in the 3rd millennium BC. *PLoS one* **12**, e0185971 (2017).
29. Szczepanek, A. *et al.* Understanding Final Neolithic communities in south-eastern Poland: New insights on diet and mobility from isotopic data. *PLoS one* **13**, e0207748 (2018).
30. Hansen, A. J. *et al.* Statistical evidence for miscoding lesions in ancient DNA templates. *Molecular Biology and Evolution* **18**, 262–265 (2001).
31. Gilbert, M. T. P. *et al.* Distribution patterns of postmortem damage in human mitochondrial DNA. *American Journal of Human Genetics* **72**, 32–47 (2003).
32. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences* **104**, 14616–14621 (2007).
33. Green, R. E. *et al.* A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416–426 (2008).
34. Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Current Biology* **23**, 553–559 (2013).
35. Behar, D. M. *et al.* A “Copernican” reassessment of the human mitochondrial DNA tree from its root. *The American Journal of Human Genetics* **90**, 675–684 (2012).
36. Wilde, S. *et al.* Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proceedings of the National Academy of Sciences* **111**, 4832–4837 (2014).

37. Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *Journal of Archaeological Science* **40**, 4477–4482 (2013).
38. Balaesque, P. *et al.* Y-chromosome descent clusters and male differential reproductive success: young lineage expansions dominate Asian pastoral nomadic populations. *European Journal of Human Genetics* **23**, 1413 (2015).
39. Genomes Project, C. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
40. Kuhn, J. M. M., Jakobsson, M. & Günther, T. Estimating genetic kin relationships in prehistoric populations. *PLoS one* **13**, e0195491 (2018).
41. Belka, Z., Dopieralska, J., Szczepanek, A. & Jarosz, P. Human mobility in the Final Eneolithic population of Święte, south-eastern Poland: Evidence from strontium isotope data. *Baltic-Pontic Studies* **23** (2018).
42. Skoglund, P. *et al.* Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104 (2015).
43. Goslar, T. *et al.* Builders and Users of Ritual Centres, Yampil Barrow Complex: Studies of Diet Based on Stable Carbon and Nitrogen Isotope Composition. *Baltic-Pontic Studies* **22**, 91–125 (2018).
44. Yang, D. Y., Eng, B., Wayne, J. S., Dудар, J. C. & Saunders, S. R. Improved DNA extraction from ancient bones using silica-based spin columns. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists* **105**, 539–543 (1998).
45. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences* **110**, 15758–15763 (2013).
46. Günther, T. *et al.* Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proceedings of the National Academy of Sciences* **112**, 11917–11922 (2015).
47. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* **2010**, pdb.prot5448 (2010).
48. Kircher, M. In *Ancient DNA* 197–228 (Springer, 2012).
49. Li, H. & Durbin, R. J. B. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
50. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
51. Hofreiter, M., Jaenicke, V., Serre, D., Haeseler, A. V. & Pääbo, S. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic acids research* **29**, 4793–4799 (2001).
52. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Pääbo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS one* **7**, e34131 (2012).
53. Vianello, D. *et al.* HAPLOFIND: A New Method for High-Throughput mt DNA Haplogroup Assignment. *Human mutation* **34**, 1189–1194 (2013).
54. Van Oven, M., Van Geystelen, A., Kayser, M., Decorte, R. & Larmuseau, M. H. *Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome*. **35**, 187–191 (2014).
55. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
56. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409 (2014).
57. Pagani, L. *et al.* Genomic analyses inform on migration events during the peopling of Eurasia. *Nature* **538**, 238 (2016).
58. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS genetics* **2**, e190 (2006).
59. Hammer, Ø., Harper, D. A. T. & Ryan, P. D. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* **4**(1), 9 (2001).
60. Alexander, D. H., Novembre, J. & Lange, K. J. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**, 1655–1664 (2009).
61. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics* **81**, 559–575 (2007).
62. Sikora, M. *et al.* Population genomic analysis of ancient and modern genomes yields new insights into the genetic ancestry of the Tyrolean Iceman and the genetic structure of Europe. *PLoS genetics* **10**, e1004353 (2014).
63. Kılınc, G. M. *et al.* The demographic development of the first farmers in Anatolia. *Current Biology* **26**, 2659–2666 (2016).
64. Behr, A. A., Liu, K. Z., Liu-Fang, G., Nakka, P. & Ramachandran, S. pong: fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics* **32**, 2817–2823 (2016).

## Acknowledgements

The authors wish to thank Vendela Kempe Lagerholm from Archaeological Research Laboratory, Stockholm University for valuable laboratory assistance, Torsten Günther for his help with data analysis and Arielle R. Munters at Department of Organismal Biology, Uppsala University, for data management and processing of raw sequence data. We would like also to thank the Foundation of the Rzeszów Archaeological Centre, Regional Museum in Rzeszów, Institute of Archaeology and Ethnology, Polish Academy of Sciences, Regional Museum in Tomaszów Lubelski and Archaeological Museum in Cracow for providing the research materials. The project was supported by the Swedish Research Council (2013-4959), Knut and Alice Wallenberg Foundation (1000 Ancient Genome Project), Riksbankens Jubileumsfond (M13\_0904:1 and P16-0553:1) and the National Science Centre, Poland, Grants No. NCN 2012/05/B/HS3/04138, 2015/19/B/HS3/02149 and 2016/20/S/HS3/00307. Sequencing was conducted at the Uppsala University SNP&SEQ Technology Platform. Computations were performed at UPPMAX resources (Uppsala Multidisciplinary Centre for Advanced Computational Science) under the projects slstore2017020, uppstore2017094 and snic2017-7-271. Open access funding provided by Stockholm University.

## Author contributions

The study was conceived and designed by A.S., P.W., P.J., M.K., A.G., J.S., A.L.; P.W., P.J., J.G., M.M., M.H., M.R., M.O., M.R., J.B., W.K. provided archaeological background information and bone material for analysis. Skeletal material was sampled by A.S., M.K. A.L., M.K. performed laboratory analyses. J.D. and Z.B. provided and interpreted Sr isotope data. Bioinformatic and population genetic analyses were performed by G.M.K., M.K. and R.R.-V. The manuscript was written by A.L. with input from all the authors and M.K., G.M.K., R.R.-V., Z.B., J.D., A.S., P.J., P.W., K.W. prepared the figures. A.G., M.K., J.S., A.S. coordinated the study.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-63138-w>.

**Correspondence** and requests for materials should be addressed to A.L. or M.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020