The transition from Stone to Bronze Age in Central and Western Europe was a period of major population movements originating from the Ponto-Caspian Steppe. Here, we report new genome-wide sequence data from 30 individuals north of this area, from the understudied western part of present-day Russia, including 3 Stone Age hunter-gatherers (10,800 to 4250 cal BCE) and 26 Bronze Age farmers from the Corded Ware complex Fatyanovo Culture (2900 to 2050 cal BCE). We show that Eastern hunter-gatherer ancestry was present in northwestern Russia already from around 10,000 BCE. Furthermore, we see a change in ancestry with the arrival of farming—Fatyanovo Culture individuals were genetically similar to other Corded Ware cultures, carrying a mixture of Steppe and European early farmer ancestry. Thus, they likely originate from a fast migration toward the northeast from somewhere near modern-day Ukraine—the closest area where these ancestries coexisted from around 3000 BCE.

INTRODUCTION

The western part of the present territory of Russia has been a focal point of several prehistoric processes yet remains heavily underrepresented in ancient DNA (aDNA) studies. Some of the oldest genetically studied individuals from Europe come from this region (1–3), but overall, ancient genetic information is sparse.

The colonization of the eastern and northern European forest belt took place in two extensive waves during the end of the Paleolithic and the beginning of the Mesolithic period 13th to 9th millennium cal BCE (bordering ca. 9700 cal BCE) (4, 5). In both cases, groups of peoples with similar material cultures to those spread in wide areas of Europe took part in the colonization process. In regard to the Mesolithic settlements in the area, a number of distinct archaeological cultures (Butovo, Kunda, Veretey, Suomusjärvi, etc.) have been identified (6–8). In older stages of habitation, the material culture is so similar that it has also been handled as a single cultural area. However, from the middle of the 9th millennium cal BCE, local population groups with clearly distinguished cultural differences already existed in the area (4). Despite a series of small changes occurring during the Mesolithic period (partly the Early Neolithic period according to the Russian periodization based on pottery production not agriculture), the cultural continuities as a general trend of those groups are observable across time until the beginning of the 5th millennium, in some areas up to the beginning of the 4th millennium cal BCE, when the so-called Pit-Comb Ware and Comb Ware cultures formed in wide areas of Europe (9). In the territory of the Volga-Oka interfluvial area in Russia, Lyalovo Culture with pit-comb pottery and its local variants were described (10). It is likely that the people from this cultural realm gave the starting boost to a series of developments in archaeological cultures specific to the 4th to 3rd millennium cal BCE, in particular to the hunting-gathering Volosovo Culture, distinguishable in large areas of Russia (11).

European Mesolithic hunter-gatherers (HG) can be divided into groups based on their genetic ancestry. The so-called Western group (WHG) was spread from Iberia to the Balkans and reached as far as the Late Mesolithic Eastern Baltic (12–14). The Eastern group (EHG) had genetic influences from further east (a genetic connection to modern Siberians) and so far includes six individuals from western Russia (14–17). The genomes of four of these individuals have been previously studied from Karelia in the northwest from 7500 to 5000 BCE (14–16) and two from the Samara region in the eastern part of European Russia from 9400 to 5500 BCE (15, 17).

Genetic studies have shown that the people associated with the Yamnaya cultural complex spread out of the Steppe region of the East European Plain and contributed substantially to the ancestry of the European populations (15, 18) that started to produce Corded Ware around 2900 to 2800 cal BCE (19). For brevity, we will use culture names when talking about individuals whose archaeological context—funerary practices and/or time period—has been associated with a certain culture. It is important to stress that, in reality, the link between culture and genetic ancestry is not to be assumed. The migration of the Yamnaya population has been estimated to have been two times faster than the Anatolian early farmer (EF) migration into Europe a few thousand years earlier and coincided with a decrease in broad-leaf forests and increase in grasslands/pastures in Western Europe (20). The Corded Ware cultural complex (CWC) was spread on a wide area, reaching Tatarstan in the East; the southern parts of Finland, Sweden, and Norway in the North; Belgium and The Netherlands in the West; and Switzerland and Ukraine in...
the South (21, 22). Its easternmost extension—Fatyanovo Culture—is a prominent Eastern European CWC branch that was spread over a large area in European Russia and introduced animal husbandry and probably crop cultivation into the forest belt (23, 24). So far, only 14 radiocarbon dates have been published for Fatyanovo Culture, placing it to 2750 to 2500 (2300) cal BCE (24). The burial customs characteristic of the culture included the placement of the dead in flat earth graves (less often in barrows), mostly flexed and on their side—men mostly on the right and women on the left side—with shaft-hole stone axes, flint tools, ceramic vessels, etc., as grave goods (23, 25).

The Yamnaya complex pastoralists shared ancestry with EHG and Caucasian HGs (CHG) (26). Genetic studies have revealed that CWC individuals, with predominantly Yamnaya ancestry, showed some admixture with European EF of Anatolian ancestry and were most similar to modern populations from Eastern and Northern Europe (14–16, 18, 27). Lactase persistence, frequent in contemporary Central and Northern Europe, was still at low frequency in the CWC individuals (14, 16, 18, 28) but underwent a fast frequency increase soon after (14, 28, 29). It has also been shown that the Yamnaya expansion was male-biased (30), while the Anatolian EF ancestry in the CWC individuals was acquired more through the female lineage (27) [see discussion in (31, 32)].

In this study, we aim to shed light on the demographic processes accompanying the change from fishing-hunting-gathering to productive livelihoods in the forest belt of northeast Europe and to look into the genetic changes involved in the transition from the Stone to the Bronze Age in the western part of present-day Russia. We add 28 new radiocarbon dates from western Russia and characterize the genetic affinities of the HG and Fatyanovo Culture farmers. As part of the study, we set out to examine whether and how the major population movements seen in other parts of Europe during the Holocene have affected this area—what was the ancestry of the settlers of northwestern Russia and was the Fatyanovo Culture population the result of a direct migration from the East European Steppe or is European EF ancestry involved similarly to more western CWC groups. In addition, our aim is to shed light on local processes like the potential admixture between Volosovo and Fatyanovo Culture people suggested by archaeological evidence (21).

RESULTS

Samples and archaeological background

In this study, we have extracted DNA from the apical tooth roots of 48 individuals from 18 archaeological sites in modern-day western Russia and Estonia (Fig. 1, data S1, table S1, and text S1). The 30 individuals with the highest preservation yielded 10 to 78% endogenous DNA and <3% contamination (table S1). We have shotgun-sequenced these individuals to an average genomic coverage of >0.01× (n = 2), >0.1× (n = 18), >1× (n = 9), and 5× (PES001) (Table 1 and table S1). The presented genome-wide data are derived from 3 Stone Age HGs (WeRuHG; 10,800 to 4250 cal BCE) and 26 Bronze Age Fatyanovo Culture farmers from western Russia (Fatyanovo; 2900 to 2050 cal BCE) and 1 Corded Ware culture individual from Estonia (ESTCWC; 2850 to 2500 cal BCE) (Fig. 1, data S1, table S1, and text S1). We analyzed the data in the context of published ancient and modern populations (tables S3 and S4).

In the case of radiocarbon dating, it is possible that fish from rivers and lakes consumed by Stone Age fisher-hunter-gatherers may cause a notable reservoir effect. This means that the radiocarbon dates obtained from the human bones and teeth can be hundreds but not thousands of years older than the actual time these people lived (33). Unfortunately, we do not yet have data to estimate the size of the reservoir effect for each specific case. However, this does not change the overall picture regarding the Stone Age individuals of this study.

Affinities of Western Russian HGs

First, we assessed the mitochondrial DNA (mtDNA) and Y chromosome (chrY) variation of the three Stone Age HGs from western Russia. The oldest individual PES001 belongs to mtDNA haplogroup (hg) U4 (Table 1 and table S1), which has been found before in EHG and Scandinavian HG individuals (13–15, 27, 34). The other two represent mtDNA hgs T2 and K1 (Table 1, fig. S1, and table S1), which is noteworthy because hg U was, by far, the most frequent in European HG before the spread of farming. However, hgs H11 and T2 have also been found previously in HG individuals (13, 14), and the estimated ages of the lineages determined for the Western Russian individuals (T2a1b1 and K1) are ~11,000 ± 2800 and ~22,000 ± 3300 years, respectively (35), likely predating the ages of the individuals (~8500 to 8300 and ~6500 to 6300 years). The chrY lineages carried by PES001 and BER001 are R1a5–YP1272 and Q1–L54, respectively (Table 1 and tables S1 and S2); both hgs have also been found previously in EHG individuals (13–16, 27).

Next, we compared the WeRuHG individuals to a set of available ancient and modern populations (tables S3 and S4) using autosomal data. We performed principal components analysis (PCA), by projecting ancient individuals onto components calculated on Western Eurasian individuals from the Human Origins (HO) dataset (https://reich.hms.harvard.edu/downloadable-genotypes-present-day-and-ancient-dna-data-compiled-published-papers) (tables S3 and S4). The PCA revealed that all three WeRuHG individuals cluster together with individuals positioned at the EHG end of the European HGcline (Fig. 2A). We then projected ancient individuals onto a worldwide modern sample set from the HO dataset (tables S3 and S4) using ADMIXTURE analysis. We ran the calculations on K = 3 to K = 18 (fig. S2, C and D) but discuss K = 9 (Fig. 2B and fig. S2, A and B). This K level had the largest number of inferred genetic clusters for which >10% of the runs that reached the highest log likelihood values yielded very similar results. The analysis again shows that WeRuHG individuals are most similar to EHG, being made up of mostly the component maximized in WHG (blue) and considerable proportions of the components most frequent in modern Russian Far East and ancient Caucasus/Iran (orange and olive, respectively) (Fig. 2B and fig. S2, A and B).

Next, we used FST, outgroup β, and D statistics to compare the genetic affinity of WeRuHG to those of other relevant populations (tables S3 and S4). We found that WeRuHG and EHG are similar in their genetic affinities both to other ancient and to modern populations (Fig. 3A, figs. S3 and S4A, and tables S5 to S8). On the other hand, when comparing WeRuHG to the later Fatyanovo, we found that WeRuHG shares relatively more genetic drift with EHG-like populations, West Siberian HG, ancient Iranians, and modern populations from Siberia, while Fatyanovo shares more with most of ancient European and Steppe populations and modern populations from the Near East, the Caucasus, and Europe (Fig. 3B, fig. S4B, and tables S5 to S8).

We studied the genetic affinities of the higher-coverage (5×) HG PES001 further using outgroup β statistics by comparing the affinities of three higher-coverage Russian Mesolithic HGs—PES001.
Fig. 1. Map of the geographical locations of the individuals of this study and timeline showing the dates of individuals and cultures. Numbers in parentheses beside site names indicate the number of individuals included from this site (if more than one). Shaded areas mark the spread of each cultural group indicated on the map. Arrow indicates the proposed direction of migration of the predecessors of the Fatyanovo Culture people. On the timeline, each individual of this study is marked by a dot using the midpoint of the 95% calibrated date estimate, and the periods of the associated cultures are shown in shaded bars.

(Peschanitsa; 10,785 to 10,626 cal BCE), I0061 (Yuzhnyy Oleni Ostrov; 6773 to 5886 BCE), and Sidelkino (Sidelkino; 9386 to 9231 cal BCE)—to Mesolithic and Paleolithic HGs from different areas of Europe and Siberia (Fig. 2C and table S9). All three are most similar to individuals from >30,000 cal BCE— to Mesolithic and Paleolithic HGs from different areas of Europe and Siberia (Fig. 2C and table S9). All three are most similar to individuals from the European part of Russia or from Siberia who lived within 10,000 years of each other—to each other, the West Siberia Neolithic population and the Afontova Gora 3 individual. These are followed by individuals from Central Europe from the same time window. Geographically close Paleolithic Sunghir and Kostenki individuals from >30,000 cal BCE share less than temporally close HGs from Central Europe with the Russian Mesolithic individuals. We also tried modeling PES001 as a mixture of WHG and either CHG, Mal’ta, or Afontova Gora 3 using qpAdm, but all three models were rejected (P value of 1.96 × 10^{−85} to 7.93 × 10^{−13}) (table S10).

**EF ancestry in Fatyanovo Culture individuals**

Then, we turned to the Bronze Age Fatyanovo Culture individuals and determined that they carry maternal (subclades of mtDNA hg U5, U4, U2e, H, T, W, J, K, I, and N1a) and paternal (chrY hg R1a1-M417) lineages (Table 1, fig. S1, and tables S2 to S4) that have also been found in CWC individuals elsewhere in Europe (14–16, 18, 27). In all individuals for which the chrY hg could be determined with sufficient depth (n = 6), it is R1a2-Z93 (Table 1 and tables S1 and S2), a lineage now spread in Central and South Asia, rather than the R1a1-Z283 lineage that is common in Europe (36). R1a2-Z93 is also not rejected for the individuals that were determined with less depth due to missing data on more apical markers (table S2).

On the PCA, the Fatyanovo individuals (and the Estonian CWC individual) group together with many European Late Neolithic/ Bronze Age (LNBA) and Steppe Middle/Late Bronze Age individuals on top of modern Northern and Eastern Europeans (Fig. 2A). This ancient cluster is shifted toward Anatolian and European EF compared to Steppe Early/Middle Bronze Age populations, including the Yamnaya. The same could be seen in ADMIXTURE analysis where the Fatyanovo individuals are most similar to LNBA Steppe
ancestry populations from Central Europe, Scandinavia, and the Eastern Baltic (Fig. 2B and fig. S2). These populations are composed of the "WHG" (blue) and "ancient Caucasus/Iran" (olive) component and small amounts of the "Russian Far East" (orange) component, similarly to Yamnaya populations. However, the European LNBA populations (including Fatyanovo) also display a component most frequent in Anatolian and European EF populations (light green), which is not present in the Yamnaya from Russia.

We studied the affinities of the Fatyanovo individuals by comparing $F_{ST}$, outgroup $f_3$, and $D$ statistics’ results of different populations and found that Fatyanovo shares more with European EF populations and modern Near Easterners than Yamnaya Samara does (Fig. 3C; figs. S3, S4C, and S5C; and tables S5 to S8). This signal can also be seen when using either autosomal or X chromosome (chrX) positions from the 1240K dataset (https://reich.hms.harvard.edu/downloadable-genotypes-present-day-and-ancient-dna-data-compiled-published-papers) instead of the autosomal positions of the HO dataset with less single-nucleotide polymorphisms (SNPs) (fig. S5, A and B, and tables S11 and S12). We also compared the affinities of Yamnaya Samara and Fatyanovo directly with $D$ statistics and saw that Fatyanovo is significantly more similar ($Z > 3$) to most EF populations than to Yamnaya Samara, and the latter, in turn, is significantly more similar to most Steppe populations than to Fatyanovo (table S13). We studied the apparent EF input in Fatyanovo further using admixture $f_3$ statistics and got significant results ($Z < −3$) for admixture between different Yamnaya groups and a wide variety of EF populations (table S14). Furthermore, when comparing outgroup $f_3$ and $D$ statistics’ results for Fatyanovo

### Table 1. Archaeological information, genetic sex, mtDNA and Y chromosome haplogroups, and average genomic coverage of the individuals of this study.

<table>
<thead>
<tr>
<th>Individual</th>
<th>Site Location</th>
<th>Group Date (cal BCE)</th>
<th>Morph. sex</th>
<th>Gen. sex</th>
<th>MT hg</th>
<th>Y hg</th>
<th>Av. cov.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BER001</td>
<td>Berendeyevo</td>
<td>Yaroslavl, RUS</td>
<td>WeRuHG 4447–4259</td>
<td>Male</td>
<td>XY</td>
<td>K1</td>
<td>0.13</td>
</tr>
<tr>
<td>KAR001</td>
<td>Karavaikha 1</td>
<td>Vologda, RUS</td>
<td>WeRuHG 6457–6258</td>
<td>Female?</td>
<td>XX</td>
<td>T2a1</td>
<td>0.19</td>
</tr>
<tr>
<td>PES001</td>
<td>Peschantitsa</td>
<td>Arkhangelsk, RUS</td>
<td>WeRuHG 10785–10626</td>
<td>Male</td>
<td>XY</td>
<td>U4a1</td>
<td>5.03</td>
</tr>
<tr>
<td>BOL001</td>
<td>Bolshevo 3</td>
<td>Tver, RUS</td>
<td>Fatyanovo 2829–2460</td>
<td>Male</td>
<td>XY</td>
<td>H1b</td>
<td>0.11</td>
</tr>
<tr>
<td>BOL002</td>
<td>Bolshevo 3</td>
<td>Tver, RUS</td>
<td>Fatyanovo 2468–2211</td>
<td>Female</td>
<td>XX</td>
<td>J1c1</td>
<td>0.12</td>
</tr>
<tr>
<td>BOL003</td>
<td>Bolshevo 3</td>
<td>Tver, RUS</td>
<td>Fatyanovo 2571–2345</td>
<td>Male</td>
<td>XY</td>
<td>H41a</td>
<td>1.13</td>
</tr>
<tr>
<td>GOL001</td>
<td>Goluzinovo</td>
<td>Yaroslavl, RUS</td>
<td>Fatyanovo 2575–2349</td>
<td>Male</td>
<td>XY</td>
<td>T2b</td>
<td>0.12</td>
</tr>
<tr>
<td>HAL001</td>
<td>Khaldeevoy</td>
<td>Yaroslavl, RUS</td>
<td>Fatyanovo 2832–2473</td>
<td>Male</td>
<td>XY</td>
<td>N1a1</td>
<td>1.35</td>
</tr>
<tr>
<td>HAN002</td>
<td>Khanevo</td>
<td>Moscow, RUS</td>
<td>Fatyanovo 2859–2495</td>
<td>Male</td>
<td>XY</td>
<td>U5a1</td>
<td>1.28</td>
</tr>
<tr>
<td>HAN004</td>
<td>Khanevo</td>
<td>Moscow, RUS</td>
<td>Fatyanovo 2835–2471</td>
<td>Male</td>
<td>XY</td>
<td>H6a1</td>
<td>1.25</td>
</tr>
<tr>
<td>IVA001</td>
<td>Ivanovogorsky</td>
<td>Moscow, RUS</td>
<td>Fatyanovo 2864–2496</td>
<td>Female</td>
<td>XX</td>
<td>U4a1</td>
<td>1.29</td>
</tr>
<tr>
<td>MIL001</td>
<td>Miloslava</td>
<td>Ivanovo, RUS</td>
<td>Fatyanovo 2624–2474</td>
<td>Female</td>
<td>XX</td>
<td>U5b2</td>
<td>0.17</td>
</tr>
<tr>
<td>MIL002</td>
<td>Miloslava</td>
<td>Ivanovo, RUS</td>
<td>Fatyanovo 2287–2047</td>
<td>Female</td>
<td>XX</td>
<td>H</td>
<td>0.11</td>
</tr>
<tr>
<td>NAU001</td>
<td>Naumovskoye</td>
<td>Yaroslavl, RUS</td>
<td>Fatyanovo 2836–2573</td>
<td>Male</td>
<td>XY</td>
<td>T2a1</td>
<td>1.58</td>
</tr>
<tr>
<td>NAU002</td>
<td>Naumovskoye</td>
<td>Yaroslavl, RUS</td>
<td>Fatyanovo 2836–2469</td>
<td>Male</td>
<td>XY</td>
<td>U5b2</td>
<td>1.14</td>
</tr>
<tr>
<td>NIK002</td>
<td>Nikultsino</td>
<td>Yaroslavl, RUS</td>
<td>Fatyanovo 2865–2500</td>
<td>Male</td>
<td>XY</td>
<td>U5a1</td>
<td>0.12</td>
</tr>
<tr>
<td>NIK003</td>
<td>Nikultsino</td>
<td>Yaroslavl, RUS</td>
<td>Fatyanovo 2522–2298</td>
<td>Male</td>
<td>XY</td>
<td>H41a</td>
<td>0.12</td>
</tr>
<tr>
<td>NIK004</td>
<td>Nikultsino</td>
<td>Yaroslavl, RUS</td>
<td>Fatyanovo 2876–2620</td>
<td>Male</td>
<td>XX</td>
<td>T2a1</td>
<td>0.18</td>
</tr>
<tr>
<td>NIK005</td>
<td>Nikultsino</td>
<td>Yaroslavl, RUS</td>
<td>Fatyanovo 2881–2581</td>
<td>Female</td>
<td>XX</td>
<td>J1c1</td>
<td>0.12</td>
</tr>
<tr>
<td>NIK007</td>
<td>Nikultsino</td>
<td>Yaroslavl, RUS</td>
<td>Fatyanovo 2862–2466</td>
<td>Female</td>
<td>XX</td>
<td>U5a1</td>
<td>0.12</td>
</tr>
<tr>
<td>NIK008AB</td>
<td>Nikultsino</td>
<td>Yaroslavl, RUS</td>
<td>Fatyanovo 2834–2472</td>
<td>Male</td>
<td>XY</td>
<td>H5b</td>
<td>1.34</td>
</tr>
<tr>
<td>RDT002</td>
<td>Nikolo-Perevoz</td>
<td>Moscow, RUS</td>
<td>Fatyanovo 2562–2299</td>
<td>NA</td>
<td>XX</td>
<td>T2b</td>
<td>0.02</td>
</tr>
<tr>
<td>TIM001</td>
<td>Timofeyevka</td>
<td>Ivanovo, RUS</td>
<td>Fatyanovo NA</td>
<td>Male</td>
<td>XY</td>
<td>K1b1</td>
<td>0.02</td>
</tr>
<tr>
<td>TIM006</td>
<td>Timofeyevka</td>
<td>Ivanovo, RUS</td>
<td>Fatyanovo 2833–2470</td>
<td>Female</td>
<td>XX</td>
<td>W6</td>
<td>0.12</td>
</tr>
<tr>
<td>TIM008</td>
<td>Timofeyevka</td>
<td>Ivanovo, RUS</td>
<td>Fatyanovo 2832–2473</td>
<td>Male</td>
<td>XY</td>
<td>K1c1</td>
<td>0.12</td>
</tr>
<tr>
<td>VOD001</td>
<td>Volosovo-</td>
<td>Yaroslavl, RUS</td>
<td>Fatyanovo 2570–2299</td>
<td>Female</td>
<td>XX</td>
<td>J1c1</td>
<td>0.24</td>
</tr>
<tr>
<td>SOP002</td>
<td>Sope</td>
<td>Ida-Viru, EST</td>
<td>EstCWC 2864–2495</td>
<td>Female</td>
<td>XX</td>
<td>R1b1</td>
<td>0.12</td>
</tr>
</tbody>
</table>

modern populations and the mixing proportions (Fig. 4 and tables of the Fatyanovo Culture individuals was a result of admixture being compared Russian Mesolithic HGs PES001, Sidelkino, and I0061 to European and Siberian Paleolithic and }

Fig. 2. Principal component, ADMIXTURE analyses’, and HG outgroup f3 statistics’ results. (A) PCA results of modern West Eurasians with ancient individuals projected onto the first two components (PC1 and PC2). (B) ADMIXTURE analysis results for a selection of ancient population averages at K9 with ancient individuals projected onto the modern genetic structure. (C) Outgroup f3 results comparing Russian Mesolithic HGs PES001, Sidelkino, and I0061 to European and Siberian Paleolithic and Mesolithic HGs (listed from youngest to oldest on the plot). LNBAIA, Late Neolithic/Bronze Age/Iron Age/Middle Ages; ML, Mesolithic; LBK, Linearbandkeramik, Linear Pottery Culture; TRB, Trichter(-rand)-becherkultur, Funnel(-neck-)beaker Culture.

and Central CWC, there are no clear differences in their affinities to different ancient or modern population groups (Fig. 3D, fig. S3D, and tables S5 to S8).

Because the previous analyses suggested that the genetic makeup of the Fatyanovo Culture individuals was a result of admixture between migrating Yamnaya individuals and contemporary European populations, we used two complementary methods (qpAdm and ChromoPainter/NLNS) to determine suitable proxies for the admixing populations and the mixing proportions (Fig. 4 and tables S15 to S17). We tested qpAdm models including Yamnaya from Samara or Kalmykia and a variety of EF populations one at a time and found that the two EF populations, with the highest P values with both Yamnaya populations, are Globular Amphora and Trypillia (P values of 0.02/0.16 and 0.06/0.26, respectively) (table S15). The admixture proportions are 65.5%/66.9% Yamnaya Samara/Kalmykia + 34.5%/33.1% Globular Amphora and 65.5%/69.6% Yamnaya Samara/Kalmykia + 34.5%/30.4% Trypillia, respectively (Fig. 4 and table S15). The proportions are similar (69 to 75% Yamnaya + 25 to 31% EF) for Central and Baltic CWC (Fig. 4 and table S15). We also tested these four models with the preceding Volosovo Culture HG
BER001 added to “right” populations to see whether there is shared drift between Volosovo and Fatyanovo that would cause the models with admixing populations without this drift to be rejected. All four models are still not rejected ($P$ values of 0.97/0.57 and 0.98/0.59, respectively) (table S15), suggesting that no Volosovo contribution is needed to model Fatyanovo. Ancestry proportions were also calculated using the ChromoPainter/NNLS pipeline with the results 37%/38% Yamnaya Samara + 63%/62% Globular Amphora/Trypillia for Fatyanovo and 51 to 60% Yamnaya + 40 to 49% EF for Central/Baltic CWC (table S16). Although the estimated proportion of EF ancestry is higher for Fatyanovo compared to the other groups in both cases, the difference is significant ($P$ value of 0.005 to 0.03) only with Trypillia in the model. Note that qpAdm calculates admixture between populations, while ChromoPainter/NNLS uses single individuals as sources, which might influence the results. Although two-way admixture between Yamnaya and an EF population is enough to explain the genetic variation in Fatyanovo, qpAdm models with an HG population added are also not rejected with EHG, WeRuHG, and Volosovo ($P$ values of 0.14 to 0.98) (table S17). Fatyanovo can be modeled as 60 to 63% Yamnaya Samara + 33 to 34% Globular Amphora + 3 to 6% HG. The results are similar for Central and Baltic CWC with 2 to 11% of HG ancestry (except with Volosovo as a source for Central CWC, which gets a negative mixture coefficient).

We estimated the time of admixture for Yamnaya and EF populations to form the Fatyanovo Culture population using DATES (37) as 13 ± 2 generations for Yamnaya Samara + Globular Amphora and 19 ± 5 generations for Yamnaya Samara + Trypillia. If a generation time of 25 years and the average calibrated date of the Fatyanovo individuals (~2600 cal BCE) are used, this equates to the admixture happening ~3100 to 2900 BCE.

Next, we investigated the possible difference in affinities between Fatyanovo and other CWC groups (Central and Baltic CWC)

Fig. 3. Outgroup $f_3$ statistics’ results of comparisons with ancient populations. Outgroup $f_3$ statistics’ values of form $f_3$(Yorubas; study population, ancient) plotted against each other for two study populations (blue and red axes): (A) Western Russian HGs (WeRuHG) and Eastern HGs (EHG), (B) WeRuHG and Fatyanovo, (C) Yamnaya Samara (YamSam) and Fatyanovo, and (D) Central Corded Ware Culture (CeCWC) and Fatyanovo. The trend line for all datapoints is shown in black. EMBA, Early/Middle Bronze Age; MLBA, Middle/Late Bronze Age; IMA, Iron/Middle Ages; IA, Iron Age; MA, Middle Ages.
and found that the populations are similar to each other because a one-way qpAdm model cannot be rejected for Fatyanovo vs Central CWC (P value of 0.26) or Central CWC vs Baltic CWC (P value of 0.48) (table S14). On the other hand, there is considerable variation in admixture proportions within populations, visible on PCA (Fig. 2) and ADMIXTURE (fig. S2A) and confirmed by per-individual qpAdm models showing 4 to 47% Globular Amphora ancestry in Fatyanovo (Fig. 4B and table S15) and 7 to 55% in the other two groups (table S15). We tested whether the variation in ancestry correlates with time using the second PC component (PC2) or the qpAdm ancestry proportions and the calibrated radiocarbon dates of the individuals. We found that there is no correlation between time and ancestry proportions in Fatyanovo (P value of 0.92/0.23, respectively), but there is a significant change toward more EF ancestry in Baltic CWC using PC2 (P value of 0.0003) and in both Central and Baltic CWC using qpAdm proportions (P value of 0.02 for both).

Furthermore, we confirmed the presence of sex-biased admixture previously seen in CWC individuals from Estonia, Poland, and Germany (27, 38, 39) in the Fatyanovo. To that end, we first compared autosomal and chrX outgroup β3 results (fig. S5D and tables S11 and S12). Two-sample two-tailed t tests assuming unequal variances showed that the mean β3 value for EF populations is significantly lower than for HG/Steppe ancestry populations based on autosomal positions (P value of 0.000001) but not based on chrX positions (P value of 0.14). Next, we calculated admixture proportions on chrX data using qpAdm and the same models as with autosomal data (table S15). Only two of the four models presented for autosomal data (Yamnaya Samara + Globular Amphora/Trypillia) yielded a significant P value (0.13/0.36, respectively) due to the low number of chrX SNPs available. The confidence intervals (CIs) were extremely wide with Trypillia, but the chrX data showed 40 to 53% Globular Amphora ancestry in Fatyanovo, in contrast with the 32 to 36% estimated using autosomal data. The sex-biased admixture is also supported by the presence of mtDNA hg N1a in two Fatyanovo individuals—an hg frequent in Linear Pottery Culture (LBK) EFs but not found in Yamnaya individuals so far (16, 18, 40)—and by all males carrying chrY hg R1a-M417, which appeared in Europe after the Steppe migration (15, 18).

DISCUSSION

After the last glacial period, at the end of the 10th and the beginning of the 9th millennium BC, vast areas in the Eastern Baltic, Finland,
and northern parts of Russia were populated relatively quickly by HG groups (42-44). Flint originating from several places in the Eastern Baltic and the European part of Russia and the similarities in lithic and bone technologies and artifacts suggest the existence of extensive social networks in the forest belt of Eastern and Northern Europe after it was populated (43, 45, 46). This has led to the hypothesis that the descendants of Final Paleolithic HG from eastern Poland to the central areas of European Russia took part in the process (8, 47) and remained connected to their origin, creating a somewhat stretched social network (4). However, these connections ceased after a few centuries, as can be witnessed from the production of stone tools from mostly local materials, and new geographically smaller social units appeared in the middle of the 9th millennium BCE (4). aDNA studies of this area have included Mesolithic individuals that are from the 8th millennium BC or younger and reveal two genetic groups: WHG in the Eastern Baltic and EHG in northwestern Russia (17,14). Flint originating from several places in the Eastern Baltic and the European part of Russia and the similarities in lithic and bone technologies and artifacts suggest the existence of extensive social networks in the forest belt of Eastern and Northern Europe after it was populated (43, 45, 46). This has led to the hypothesis that the descendants of Final Paleolithic HG from eastern Poland to the central areas of European Russia took part in the process (8, 47) and remained connected to their origin, creating a somewhat stretched social network (4). However, these connections ceased after a few centuries, as can be witnessed from the production of stone tools from mostly local materials, and new geographically smaller social units appeared in the middle of the 9th millennium BCE (4). aDNA studies of this area have included Mesolithic individuals that are from the 8th millennium BC or younger and reveal two genetic groups: WHG in the Eastern Baltic and EHG in northwestern Russia (17,14). Flint originating from several places in the Eastern Baltic and the European part of Russia and the similarities in lithic and bone technologies and artifacts suggest the existence of extensive social networks in the forest belt of Eastern and Northern Europe after it was populated (43, 45, 46). This has led to the hypothesis that the descendants of Final Paleolithic HG from eastern Poland to the central areas of European Russia took part in the process (8, 47) and remained connected to their origin, creating a somewhat stretched social network (4). However, these connections ceased after a few centuries, as can be witnessed from the production of stone tools from mostly local materials, and new geographically smaller social units appeared in the middle of the 9th millennium BCE (4). aDNA studies of this area have included Mesolithic individuals that are from the 8th millennium BC or younger and reveal two genetic groups: WHG in the Eastern Baltic and EHG in northwestern Russia (17,14). Flint originating from several places in the Eastern Baltic and the European part of Russia and the similarities in lithic and bone technologies and artifacts suggest the existence of extensive social networks in the forest belt of Eastern and Northern Europe after it was populated (43, 45, 46). This has led to the hypothesis that the descendants of Final Paleolithic HG from eastern Poland to the central areas of European Russia took part in the process (8, 47) and remained connected to their origin, creating a somewhat stretched social network (4). However, these connections ceased after a few centuries, as can be witnessed from the production of stone tools from mostly local materials, and new geographically smaller social units appeared in the middle of the 9th millennium BCE (4). aDNA studies of this area have included Mesolithic individuals that are from the 8th millennium BC or younger and reveal two genetic groups: WHG in the Eastern Baltic and EHG in northwestern Russia (17,14). Flint originating from several places in the Eastern Baltic and the European part of Russia and the similarities in lithic and bone technologies and artifacts suggest the existence of extensive social networks in the forest belt of Eastern and Northern Europe after it was populated (43, 45, 46). This has led to the hypothesis that the descendants of Final Paleolithic HG from eastern Poland to the central areas of European Russia took part in the process (8, 47) and remained connected to their origin, creating a somewhat stretched social network (4). However, these connections ceased after a few centuries, as can be witnessed from the production of stone tools from mostly local materials, and new geographically smaller social units appeared in the middle of the 9th millennium BCE (4). aDNA studies of this area have included Mesolithic individuals that are from the 8th millennium BC or younger and reveal two genetic groups: WHG in the Eastern Baltic and EHG in northwestern Russia (17,14). Flint originating from several places in the Eastern Baltic and the European part of Russia and the similarities in lithic and bone technologies and artifacts suggest the existence of extensive social networks in the forest belt of Eastern and Northern Europe after it was populated (43, 45, 46).

The formation of Fatyanovo Culture is one of the main factors that affected the population, culture, and lifestyle of the previously hunter/fisher-gatherer culture of the Eastern European forest belt. The Fatyanovo Culture people were the first farmers in the area, and the arrival of the culture has been associated with migration (21, 24). This is supported by our results as the Stone Age HG and the Bronze Age Fatyanovo individuals are genetically clearly distinguishable. The sample size of our HG is low, but the three individuals form a genetically homogenous group with previously reported EHG individuals (14, 16), and the newly reported PES001 is the highest-coverage whole-genome sequenced EHG individual so far, providing a valuable resource for future studies. What is more, the Fatyanovo Culture individuals (similarly to other CWC people) have not only mostly Steppe ancestry but also some EF ancestry that was not present in the area before and thus excludes the northward migration of Yamnaya Culture people with only Steppe ancestry as the source of Fatyanovo Culture population. The strongest connections for Fatyanovo Culture in archaeological material can be seen with the Middle Dnieper Culture (23, 48) spread in present-day Belarus and Ukraine (49, 50). The territory of what is now Ukraine is where the most eastern individuals with European EF ancestry and the most western Yamnaya Culture individuals are from based on published genomic data (13, 51) (Fig. 1 and data S1). Furthermore, archaeological finds show that LBK reached western Ukraine around 5300 BCE (52), and the Yamnaya complex (burial mounds) arrived in southeastern Europe around 3000 BCE and spread further as far as Romania, Bulgaria, Serbia, and Hungary (53). This is in accordance with our genetic results as the two populations that proved to be plausible mixture sources for Fatyanovo, with the

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Western Russian HGs</th>
<th>Latvian HGs</th>
<th>Estonian and Latvian Corded Ware complex farmers</th>
<th>Fatyanovo</th>
<th>Estonian Late Bronze Age</th>
<th>Estonian/Ingrian Iron Age</th>
<th>Estonian Middle Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue eyes</td>
<td>0.00</td>
<td>0.60</td>
<td>0.29</td>
<td>0.21</td>
<td>0.70</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Brown eyes</td>
<td>1.00</td>
<td>0.40</td>
<td>0.71</td>
<td>0.79</td>
<td>0.30</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Blond + dark blond hair</td>
<td>0.00</td>
<td>0.20</td>
<td>0.00</td>
<td>0.04</td>
<td>0.10</td>
<td>0.78</td>
<td>0.75</td>
</tr>
<tr>
<td>Red hair</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.25</td>
<td>0.00</td>
</tr>
<tr>
<td>Brown/dark brown hair</td>
<td>0.00</td>
<td>0.00</td>
<td>0.14</td>
<td>0.21</td>
<td>0.30</td>
<td>0.11</td>
<td>0.00</td>
</tr>
<tr>
<td>Dark brown + black hair</td>
<td>1.00</td>
<td>0.80</td>
<td>0.86</td>
<td>0.75</td>
<td>0.60</td>
<td>0.11</td>
<td>0.00</td>
</tr>
<tr>
<td>Very pale + pale skin</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.10</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Intermediate skin</td>
<td>0.33</td>
<td>0.60</td>
<td>0.57</td>
<td>0.58</td>
<td>0.80</td>
<td>0.67</td>
<td>0.50</td>
</tr>
<tr>
<td>Mixed/unpredictable very pale-pale</td>
<td>0.67</td>
<td>0.40</td>
<td>0.43</td>
<td>0.42</td>
<td>0.00</td>
<td>0.11</td>
<td>0.00</td>
</tr>
<tr>
<td>Intermediate-dark + dark + black skin</td>
<td>0.00</td>
<td>0.00</td>
<td>0.07</td>
<td>0.17</td>
<td>0.40</td>
<td>0.50</td>
<td>0.75</td>
</tr>
<tr>
<td>Lactase persistence (rs4988235)</td>
<td>1.00</td>
<td>0.90</td>
<td>0.57</td>
<td>0.83</td>
<td>0.45</td>
<td>0.56</td>
<td>0.13</td>
</tr>
</tbody>
</table>
other source being either of the two Russian Yamnaya groups (Kalmykia or Samara), were Globular Amphora that includes individuals from Ukraine and Poland, and Trypillia that is composed of individuals from Ukraine. These findings suggest present-day Ukraine as the possible origin of the migration leading to the formation of the Fatyanovo Culture and of the Corded Ware cultures in general.

The exact timing of and processes involved in the emergence of the Fatyanovo Culture in European Russia and the local processes following it have also remained unclear. Until recently, the Fatyanovo Culture was thought to have developed later than other CWC groups and over a longer period of time (21, 23). However, radiocarbon dates published last year (24), in combination with the 25 new dates presented here and the estimate for Yamnaya and EF admixture ~300 to 500 years before the Fatyanovo individuals of this study lived, point toward a fast process, similar in time to CWC people reaching the Eastern Baltic and southern Fennoscandia (54–56). The archaeological cultures are clearly differentiated between the areas. What is more, it has been suggested that the Fatyanovo Culture people admixed with the local Volosovo Culture HG after their arrival in European Russia (21, 57, 58). Our results do not support this as they do not reveal more HG ancestry in the Fatyanovo people compared to two other CWC groups; the three groups are shown to be similar by nonrejected one-way qpAdm models, and correlating radiocarbon dates with PC values or qpAdm ancestry proportions reveals no change in ancestry proportions of the Fatyanovo people during the period covered by our samples (2900 to 2050 BCE).

Last, allele frequency changes in western Russia and the Eastern Baltic revealed similar patterns in both areas: The frequencies of alleles in the MCM6 and FADS1-2 genes, which have been hypothesized to have changed due to dietary shifts from the Neolithic onward (16, 29, 59), change significantly during the Bronze Age, although the first signals of change can be seen already from the Neolithic. In accordance with a recent publication on lactase persistence (60), we find a low frequency of the rs4988235A allele in the initial Steppe ancestry samples [90% CI, 0 to 2.7% in (60), 0 to 33.8% in this study]. This suggests that the factors affecting these allele frequency shifts over time were complex and may have involved several environmental factors and genetic forces, as has already been suggested previously (14, 16, 18, 28, 29, 60, 61).

### MATERIALS AND METHODS

#### Experimental design

The teeth used for DNA extraction were obtained with relevant institutional permissions from the Institute of Ethnology and Anthropology of Russian Academy of Sciences (Russia), Cherepovets Museum Association (Russia), and Archaeological Research Collection of Tallinn University (Estonia). DNA was extracted from the teeth of 48 individuals: 3 from Stone Age HGs from western Russia (WeRuHG; 10,800 to 4250 cal BCE), 44 from Bronze Age Fatyanovo Culture individuals from western Russia (Fatyanovo; 2900 to 2050 cal BCE), and 1 from a Corded Ware Culture individual from Estonia (EstCWC; 2850 to 2500 cal BCE) (Fig. 1, data S1, table S1, and text S1). Petrous bones of 13 of the Fatyanovo Culture individuals have been sampled for another project. More detailed information about the archaeological periods and the specific sites and burials of this study is given below.

All of the laboratory work was performed in dedicated aDNA laboratories of the Institute of Genomics, University of Tartu. The library quantification and sequencing were performed at the Institute of Genomics Core Facility, University of Tartu. The main steps of the laboratory work are detailed below.

#### Laboratory methods

**DNA extraction**

The teeth of 48 individuals were used to extract DNA. One individual was sampled twice from different teeth. Apical tooth roots were cut off with a drill and used for extraction because root cementum has been shown to contain more endogenous DNA than crown dentine (62). The root pieces were used whole to avoid heat damage during powdering with a drill and to reduce the risk of cross-contamination between samples. Contaminants were removed from the surface of tooth roots by soaking in 6% bleach for 5 min, then rinsing three times with Milli-Q water (Millipore), and finally soaking in 70% ethanol for 2 min, shaking the tubes during each round to dislodge particles. Last, the samples were left to dry under an ultraviolet light for 2 hours.

Next, the samples were weighed, [20 * sample mass (mg)] μl of EDTA and [sample mass (mg) / 2] μl of proteinase K were added, and the samples were left to digest for 72 hours on a rotating mixer at 20°C to compensate for the smaller surface area of the whole root compared to powder. Undigested material was stored for a second DNA extraction if need be.

The DNA solution was concentrated to 250 μl (Vivaspin Turbo 15, 30,000 MWCO PES, Sartorius) and purified in large-volume columns (High Pure Viral Nucleic Acid Large Volume Kit, Roche) using 2.5 ml of PB buffer, 1 ml of PE buffer, and 100 μl of EB buffer (MinElute PCR Purification Kit, Qiagen). The adaptor ligation module was implemented using 10 μl of buffer, 5 μl of T4 ligase, and 5 μl of adaptor mix (63), incubating at 20°C for 15 min. The samples were purified as in the previous step and eluted in 30 μl of EB buffer (MinElute PCR Purification Kit, Qiagen). The adaptor fill-in module was implemented using 13 μl of water, 5 μl of buffer, and 2 μl of Bst DNA polymerase, incubating at 37°C for 30 min and at 80°C for 20 min. The libraries were amplified, and both the indexed and universal primers (NEBNext Multiplex Oligos for Illumina, New England Biolabs) were added by polymerase chain reaction (PCR) using HGS Taq DNA polymerase (Eurogentec). The samples were purified and eluted in 35 μl of EB buffer (MinElute PCR Purification Kit, Qiagen). Three verification steps were implemented to make sure library preparation was successful and to measure the concentration of double-stranded DNA/sequencing libraries—fluorometric quantitation (Qubit, Thermo Fisher Scientific), parallel capillary electrophoresis (Fragment Analyzer, Agilent Technologies), and quantitative PCR. One sample (TIM004) had a DNA concentration lower than our threshold for sequencing and was...
hence excluded, leaving 48 samples from 47 individuals to be sequenced.

**DNA sequencing**

DNA was sequenced using the Illumina NextSeq 500 platform with the 75–base pair (bp) single-end method. First, 15 samples were sequenced together on one flow cell. Later, additional data were generated for some samples to increase coverage.

**Statistical analysis**

**Mapping**

Before mapping, the sequences of adaptors and indexes and poly-G tails occurring due to the specifics of the NextSeq 500 technology were cut from the ends of DNA sequences using cutadapt 1.11 (66). Sequences shorter than 30 bp were also removed with the same program to avoid random mapping of sequences from other species. The sequences were mapped to reference sequence GRCh37 (hs37d5) using Burrows-Wheeler Aligner (BWA 0.7.12) (67) and command mem with reseeding disabled.

After mapping, the sequences were converted to BAM format, and only sequences that mapped to the human genome were kept with samtools 1.3 (68). Next, data from different flow cell lanes were merged and duplicates were removed with picard 2.12 (http://broadinstitute.github.io/picard/index.html). Indels were realigned with GATK 3.5 (69), and lastly, reads with mapping quality under 10 were filtered out with samtools 1.3 (68).

The average endogenous DNA content (proportion of reads mapping to the human genome) for the 48 samples is 29% (Table S1). The endogenous DNA content is variable as is common in aDNA studies, ranging from under 1 to around 78% (Table S1).

**aDNA authentication**

As a result of degrading over time, aDNA can be distinguished from modern DNA by certain characteristics: short fragments and a high frequency of C→T substitutions at the 5′ ends of sequences due to cytosine deamination. The program mapDamage2.0 (70) was used to estimate the frequency of 5′ C→T transitions.

mtDNA contamination was estimated using the method from (71). This included calling an mtDNA consensus sequence based on reads with mapping quality of at least 30 and positions with at least 5× coverage, aligning the consensus with 311 other human mtDNA sequences from (71), mapping the original mtDNA reads to the consensus sequence, and running contamMix 1.0-10 with the reads mapping to the consensus and the 312 aligned mtDNA sequences while trimming seven bases from the ends of reads with the option trimBases. For the male individuals, contamination was also estimated on the basis of chrX using the two contamination estimation methods first described in (72) and incorporated in the ANGSD software (73) in the script contamination.R.

The samples show 10% C→T substitutions at the 5′ ends on average, ranging from 6 to 17% (Table S1). The mtDNA contamination point estimate for samples with >5× mtDNA coverage ranges from 0.03 to 2.02% with an average of 0.4% (Table S1). The average of the two chrX contamination methods of male individuals with average chrX coverage of >0.1× is between 0.4 and 0.87% with an average of 0.7% (Table S1).

**Kinship analysis**

A total of 4,375,438 biallelic single-nucleotide variant sites, with minor allele frequency (MAF) > 0.1 in a set of more than 2000 high-coverage genomes of Estonian Genome Center (EGC) (74), were identified and called with ANGSD (73) command --doHaploCall from the 25 BAM files of 24 Fatyanovo individuals with coverage of >0.03×. The ANGSD output files were converted to .tped format as an input for the analyses with READ script to infer pairs with first- and second-degree relatedness (41).

The results are reported for the 100 most similar pairs of individuals of the 300 tested, and the analysis confirmed that the two samples from one individual (NIK008A and NIK008B) were indeed genetically identical (fig. S6). The data from the two samples from one individual were merged (NIK008AB) with samtools 1.3 option merge (68).

**Calculating general statistics and determining genetic sex**

Samtools 1.3 (68) option stats was used to determine the number of final reads, average read length, average coverage, etc. Genetic sex was calculated using the script sexing.py from (75), estimating the fraction of reads mapping to chrY out of all reads mapping to either X or Y chromosome.

The average coverage of the whole genome for the samples is between 0.00004× and 5.03× (Table S1). Of these, 2 samples have an average coverage of >0.01×, 18 samples have >0.1×, 9 samples have >1×, 1 sample have around 5×, and the rest are lower than 0.01× (Table S1). Genetic sexing confirms morphological sex estimates or provides additional information about the sex of the individuals involved in the study. Genetic sex was estimated for samples with an average genomic coverage of >0.005×. The study involves 16 females and 20 males (Table 1 and table S1).

**Determining mtDNA hgs**

The program bcftools (76) was used to produce VCF files for mitochondrial positions; genotype likelihoods were calculated using the option mpileup, and genotype calls were made using the option call. mtDNA hgs were determined by submitting the mtDNA VCF files to HaploGrep2 (77, 78). Subsequently, the results were checked by looking at all the identified polymorphisms and confirming the hg assignments in PhyloTree (78). Hgs for 41 of the 47 individuals were successfully determined (Table 1, fig. S1, and table S1).

No female samples have reads on the chrY consistent with a hg, indicating that levels of male contamination are negligible. Hgs for 17 (with coverage of >0.005×) of the 20 males were successfully determined (Table 1 and tables S1 and S2).

**chrY variant calling and hg determination**

In total, 113,217 haplogroup informative chrY variants from regions that uniquely map to chrY (36, 79–82) were called as haploid from the BAM files of the samples using the --doHaploCall function in ANGSD (73). Derived and ancestral allele and hg annotations for each of the called variants were added using BEDTools 2.19.0 intersect option (83). Hg assignments of each individual sample were made manually by determining the hg with the highest proportion of informative positions called in the derived state in the given sample. chrY haplogrouping was blindly performed on all samples regardless of their sex assignment.

**Genome-wide variant calling**

Genome-wide variants were called with the ANGSD software (73) command --doHaploCall, sampling a random base for the positions that are present in the 1240K dataset (https://reich.hms.harvard.edu/downloadable-genotypes-present-day-and-ancient-dna-data-compiled-published-papers).

**Preparing the datasets for autosomal analyses**

The HO array dataset (https://reich.hms.harvard.edu/downloadable-genotypes-present-day-and-ancient-dna-data-compiled-published-papers) was used as the modern DNA background. Individuals
from the 1240K dataset (https://reich.hms.harvard.edu/downloadable-genotypes-present-day-and-ancient-dna-data-compiled-published-papers) were used as the aDNA background. The data of the comparison datasets and of the individuals of this study were converted to BED format using PLINK 1.90 (http://pngu.mgh.harvard.edu/purcell/plink/) (84), and the datasets were merged. Two datasets were prepared for analyses: one with HO and 1240K individuals and the individuals of this study, where 584,901 autosomal SNPs of the HO dataset were kept; the other with 1240K individuals and the individuals of this study, where 1,136,395 autosomal and 48,284 chrX SNPs of the 1240K dataset were kept.

Individuals with <10,000 SNPs overlapping with the HO autosomal dataset were removed from further autosomal analyses, leaving 30 individuals of this study to be used in autosomal analyses. These included 3 from WeRuHG, 26 from Fatyanovo, and 1 from EstCWC (table S1).

Principal components analysis

To prepare for PCA, a reduced comparison sample set composed of 813 modern individuals from 53 populations of Europe, Caucasus, and Near East and 737 ancient individuals from 107 populations was assembled (tables S3 and S4). The data were converted to EIGENSTRAT format using the program convertf from the EIGENSOFT 5.2.0 package (85). PCA was performed with the program smart pca from the same package, projecting ancient individuals onto the components constructed based on the modern genotypes using the option lsqproject and trying to account for the shrinkage problem introduced by projecting by using the option autoshrink.

Admixture analysis

For Admixture analysis (86), the same ancient sample set was used as for PCA, and the modern sample set was increased to 1861 individuals from 144 populations from all over the world (tables S3 and S4). The analysis was carried out using ADMIXTURE 1.3 (86) with the P option, projecting ancient individuals into the genetic structure calculated on the modern dataset due to missing data in the ancient samples. The HO dataset of modern individuals was pruned to decrease linkage disequilibrium using the option indep-pairwise with parameters 1000 250 0.4 in PLINK 1.90 (http://pngu.mgh.harvard.edu/purcell/plink/) (84). This resulted in a set of 269,966 SNPs. Admixture was run on this set using \(K = 3\) to \(K = 18\) in 100 replicates. This enabled us to assess convergence of the different models. \(K = 10\) and \(K = 9\) were the models with the largest number of inferred genetic clusters for which >10% of the runs that reached the highest log likelihood values yielded very similar results. This was used as a proxy to assume that the global likelihood maximum for this particular model was indeed reached. Then, the inferred genetic cluster proportions and allele frequencies of the best run at \(K = 9\) were used to run Admixture to project the aDNA individuals, for which the intersection with the LD pruned modern dataset yielded data for more than 10,000 SNPs, on the inferred clusters. The same projecting approach was taken for all models for which there is good indication that the global likelihood maximum was reached (K3 to 18). We present all ancient individuals in fig. S2 but only population averages in Fig. 2B. The resulting membership proportions to \(K\) genetic clusters are sometimes called “ancestry components,” which can lead to overinterpretation of the results. The clustering itself is, however, an objective description of genetic structure and hence a valuable tool in population comparisons.

Outgroup \(f_3\) statistics

For calculating autosomal outgroup \(f_3\) statistics, the same ancient sample set as for previous analyses was used, and the modern sample set included 1177 individuals from 80 populations from Europe, Caucasus, Near East, Siberia and Central Asia, and Yoruba as outgroup (tables S3 and S4). The data were converted to EIGENSTRAT format using the program convertf from the EIGENSOFT 5.0.2 package (85). Outgroup \(f_3\) statistics of the form \(f_3(\text{Yoruba}; \text{West Siberia}_N/\text{EHG}/\text{CentralRussiaHG}/\text{Fatyanovo}/\text{Yamnaya_Samara}/\text{Poland_CWC}/\text{Baltic_CWC}/\text{Central_CWC}, \text{modern}/\text{ancient})\) were computed using the ADMIXTOOLS 6.0 program qp3Pop (87).

To allow chrX versus autosomal comparison for ancient populations, outgroup \(f_3\) statistics using chrX SNPs were computed. To allow the use of the bigger number of positions in the 1240K over the HO dataset, Mbuti from the Simons Genome Diversity Project (88) was used as the outgroup. The outgroup \(f_3\) analyses of the form \(f_3(\text{Mbuti}; \text{West Siberia}_N/\text{EHG}/\text{CentralRussiaHG}/\text{Fatyanovo}/\text{Yamnaya_Samara}/\text{Poland_CWC}/\text{Baltic_CWC}/\text{Central_CWC}, \text{ancient})\) were run both using not only 1,136,395 autosomal SNPs but also 48,284 chrX positions available in the 1240K dataset. Because all children inherit half of their autosomal material from their father but only female children inherit their chrX from their father, then in this comparison chrX data give more information about the female and autosomal data about the male ancestors of a population.

The autosomal outgroup \(f_3\) results of the two different SNP sets were compared to each other and to the results based on the chrX positions of the 1240K dataset to see whether the SNPs used affect the trends seen. Outgroup \(f_3\) analyses were also run with the form \(f_3(\text{Mbuti}; \text{PE}5001/\text{I0061}/\text{Sidelkino}, \text{Paleolithic/Mesolithic HG})\) and admixture \(f_3\) analyses with the form \(f_3(\text{Fatyanovo}; \text{Yamnaya}, \text{EF})\) using the autosomal positions of the 1240K dataset.

D statistics

\(D\) statistics of the form \(D(\text{Yoruba}, \text{West Siberia}_N/\text{EHG}/\text{CentralRussiaHG}/\text{Fatyanovo}/\text{Yamnaya_Samara}/\text{Poland_CWC}/\text{Baltic_CWC}/\text{Central_CWC}, \text{Russian, modern/ancient})\) were calculated on the same dataset as outgroup \(f_3\) statistics (tables S3 and S4) using the autosomal positions of the HO dataset. The ADMIXTOOLS 6.0 package program qpDstat was used (87).

In addition, \(D\) statistics of the form \(D(\text{Mbuti}, \text{ancient}; \text{Yamnaya_Samara}, \text{Fatyanovo}/\text{Baltic_CWC}/\text{Central_CWC})\) and \(D(\text{Mbuti}, \text{ancient}; \text{Poland_CWC}/\text{Baltic_CWC}/\text{Central_CWC}, \text{Fatyanovo})\) were calculated using the autosomal positions of the 1240K dataset. However, comparing very similar populations directly using \(D\) statistics seems to be affected by batch biases—Central_CWC comes out as significantly closer to almost all populations than Fatyanovo, while this is not the case when comparing less similar Fatyanovo and Yamnaya_Samara. Because of this, the results of \(D(\text{Mbuti}, \text{ancient}; \text{Poland_CWC}/\text{Baltic_CWC}/\text{Central_CWC}, \text{Fatyanovo})\) are not discussed in the main text, but the data are included in table S19.

\(F_{ST}\)

Weir and Cockerham pairwise average \(F_{ST}\) (89) was calculated for the dataset used for outgroup \(f_3\) and \(D\) statistics using the autosomal positions of the HO dataset using a custom script.

\(qPAdm\)

The ADMIXTOOLS 6.0 (87) package programs qpWave and qpAdm were used to estimate which populations and in which proportions are suitable proxies of admixture to form the populations or individuals of this study. The autosomal positions of the 1240K dataset were used. Only samples with more than 100,000 SNPs were used in the analyses. Mota, Ust-Ishim, Kostenki14, GoyetQ116, Vestonice16, MA1, AfontovaGora3, ELMiron, Villabruna, WHG,
Ehgh, CHG, Iran_N, Natufian, Levant_N, and Anatolia_N (and Volosovo in some cases indicated in table S15) were used as right populations. Yamnaya_Samara or Yamnaya_Kalmykia was used as the left population representing Steppe ancestry. Levant_N, Anatolia_N, LBK_EN, Central_MN, Globular_Amphora, Trypillia, Ukraine_Eneolithic, or Ukraine_Neolithic was used as the left population representing EF ancestry. In some cases, WHG, EH, WesternRussiaHG, or Volosovo was used as the left population representing HG ancestry. Alternatively, one-way models between Fatyanovo, Baltic_CWC, and Central_CWC were tested. Also, PES001 was modeled as a mixture of WHG and AfontovaGora3, MA1, or CHG.

To look at sex bias, four models that were not rejected using autosomal data were also tested using the 48,284 chrX positions of the 1240K dataset. The same samples were used as in the autosomal modeling.

ChromoPainter/NNLS
To infer the admixture proportions of ancient individuals, the ChromoPainter/NNLS pipeline was applied (28). Because of the low coverage of the ancient data, it is not possible to infer haplotypes, and the analysis was performed in unlinked mode (option -u). The autosomal positions of the HO dataset were used. Only samples with more than 20,000 SNPs were used in the analyses. Because ChromoPainter (90) does not tolerate missing data, every ancient target individual was iteratively painted together with one representative individual from potential source populations as recipients. All the remaining modern individuals from the sample set used for Admixture analysis were used as donors (tables S3 and S4). Subsequently, we reconstructed the profile of each target individual as a combination of two or more ancient individuals, using the non-negative least square approach. Let $Xg$ and $Yp$ be vectors summarizing the proportion of DNA that source and target individuals copy from each of the modern donor groups as inferred by ChromoPainter. $Yp = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_g X_g$. $Xg$ was reconstructed using a slight modification of the nnls function in R (91) and implemented in GlobeTrotter (92) under the conditions $\beta_g \geq 0$ and $\sum \beta = 1$. To evaluate the fitness of the NNLS estimation, we inferred the sum of the squared residual for every tested model (93). Models identified as plausible with qPAdm with Yamnaya_Samara and Globular_Amphora/Trypillia as sources were used. The resulting painting profiles, which summarize the fraction of the individual’s DNA inherited by each donor individual, were summed over individuals from the same population.

DATES
The time of admixture between Yamnaya and EF populations forming the Fatyanovo Culture population was estimated using the program DATES (37). The autosomal positions of the 1240K dataset were used.

Phenotyping
To predict eye, hair, and skin color in the ancient individuals (tables S20 to S22), the HirisPlex-S variants from 19 genes in nine autosomes were selected (94–96), and the region to be analyzed was selected adding 2 Mb around each SNP, collapsing in the same region the variants separated by less than 5 Mb. A total of 10 regions (2 for chromosome 15 and 1 for each of the remaining autosomes) were obtained, ranging from about 6 to about 1.5 Mb. Similarly, to analyze the other phenotype-informative markers (diet, immunity, and diseases), 2 Mb around each variant was selected, and the overlapping regions were merged, for a total of 47 regions (45 regions in 17 autosomes and 2 regions on chrX). For the local imputation, we used a two-step pipeline (97) as follows: (i) variant calling, (ii) first imputation step using a reference panel as much similar as possible to the target samples, (iii) variant filtering, (iv) second imputation step using a larger worldwide reference panel, and (v) final variant filtering. This pipeline has been validated by randomly downsampling a high-coverage Neolithic sample (NE1) (98) to 0.05x and comparing the imputed variants in the low-coverage version with the called variants from the original genome. For a local imputation approach on 2 Mb, we obtained a concordance rate higher than 90% for all the variants, a figure that increased to 99% for frequent variants (MAF ≥ 0.3). The variants were called using ATLAS v0.9.0 (99) (task = call and method = MLE commands) (step 1) at biallelic SNPs with a MAF ≥ 0.1% in a reference panel composed of more than 2000 high-coverage Estonian genomes (EGC) (74). The variants were called separately for each sample and merged in one VCF file per chromosomal region. The merged VCFs were used as input for the first step of our two-step imputation pipeline (genotype likelihood update; -gl command on Beagle 4.1 (100)), using the EGC panel as reference (step 2). Then, the variants with a genotype probability (GP) less than 0.99 were discarded (step 3), and the missing genotype was imputed with the -gt command of Beagle 5.0 (101) using the large HRC as reference panel (102), with the exception of variants rs333 and rs2430561 [not present in the HRC (Haploptype Reference Consortium)], imputed using the 1000 Genomes as reference panel (step 4) (103). Last, a second GP filter was applied to keep variants with GP ≥ 0.85 (step 5). Then, the 113 phenotype-informative SNPs were extracted, recoded, and organized in tables, using VCFTools (104), PLINK 1.9 (http://pngu.mgh.harvard.edu/purcell/plink/) (84), and R (91) (tables S21 and S22). The HirisPlex-S variants were uploaded on the HirisPlex webtool (https://hirisplex.erasmusmc.nl/) to perform the pigmentation prediction, after tabulating them according to the manual of the tool. Out of 41 variants of the HirisPlex-s system, two markers were not analyzed, namely, the rs312262906 indel and the rare (MAF = 0) in the HRC rs201326893 SNP, because of the difficulties in the imputation of such variants.

The 28 samples analyzed here were compared with 34 ancient samples from surrounding geographical regions from literature, gathering them in seven groups according to their region and/or culture: (i) 3 Western Russian Stone Age HGs (present study); (ii) 5 Latvian Mesolithic HGs (34); (iii) 7 Estonian and Latvian Corded Ware Culture farmers [present study and (27, 34)]; (iv) 24 Fatyanovo Culture individuals [present study]; (v) 10 Estonian Bronze Age individuals (28); (vi) 9 Estonian and Ingrain Iron Age individuals (28); (vii) 4 Estonian Middle Age individuals (28). For each variant, an analysis of variance (ANOVA) test was performed between the seven groups, applying Bonferroni’s correction by the number of tested variants to set the significance threshold (table S20). For the significant variant, a Tukey test was performed to identify the significant pairs of groups.

SUPPLEMENTARY MATERIALS
Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/4/eabd6535/DC1
View/request a protocol for this paper from Bio-protocol.

REFERENCES AND NOTES
110. 7, 43950 (2017).
111. 8, 141 (2016).
112. 98, 116–126 (2016).
113. 16883 (2019).
123. N. A. Кренке, Московские памятники фатьяновской культуры, в Российская Археология (2014), pp. 5–18.
125. E. V. Volkova, Новые материалы по Фатьяновской культуре (Институт археологии Российской академии наук, 2010).
126. O. N. Bader, К истории бассейна Москвы, в Советская Эпоха, 1 (1950), pp. 64–85.
133. 0. S. Gadzhiyatskaia, Милославский фатьяновский могильник, в Памятники Каменного и бронзового веков Евразии, O. N. Bader, Ed. (Hayka, 1964).
134. O. N. Bader, Минусинский фатьяновский могильник, в Краткие сообщения о докладах и полевых исследованиях Института истории материальной культуры, 75 (Издательство Академии наук СССР, 1959), pp. 143–153.


Acknowledgments: We thank M. Keller for help with coordinates. Analyses were carried out using the facilities of the High Performance Computing Center of the University of Tartu.

Funding: This work was funded by research projects of the Estonian Research Council IUT20-7 (to A.K.) and PRG243 (to M.M., Lauri Saag, C.L.S., Lehti Saag, and A.S.); the EU European Regional Development Fund 2014-2020.01.16-0030, 2014-2020.01.16-0125, and 2014-2020.01.15-0012; and Arheograator Ltd. (to L.V. and A.K.). We would like to thank the University of Tartu Development Fund for support to the Collegium for Transdisciplinary Studies in Archaeology, Genetics and Linguistics. Author contributions: Lehti Saag, A.K., and M.M. conceived the study. S.V.V., L.V., N.V.K., D.V.G., S.V.O., and A.K. assembled skeletal samples and performed osteological analyses. Lehti Saag, S.I.G., K.T., C.L.S., and A.S. performed aDNA extraction and sequencing. Lehti Saag, Lauri Saag, E.D., E.M., M.R., S.R., and T.K. analyzed data. Lehti Saag, L.V., N.V.K., D.V.G., T.K., K.T., A.K., and M.M. wrote the manuscript with input from the remaining authors. Competing interests: The authors declare that they have no competing interests. Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The DNA sequences are available through the data depository of the EBC (http://evolbio.ut.ee) and through the European Nucleotide Archive (accession code PRJEB40698). Additional data related to this paper may be requested from the authors.

Submitted 3 July 2020
Accepted 1 December 2020
Published 20 January 2021
10.1126/sciadv.abd6535