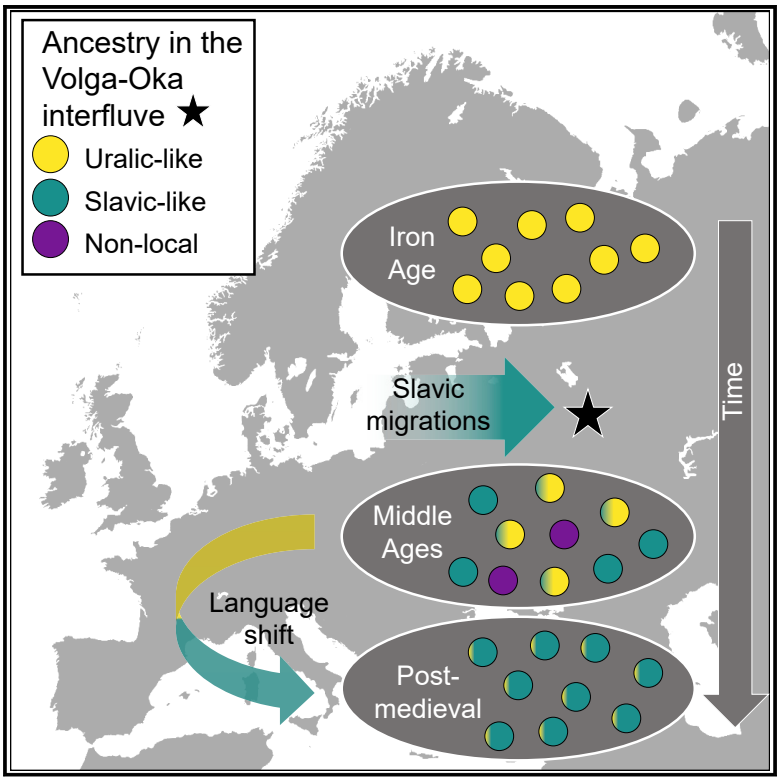
Report

Current Biology

Genetic admixture and language shift in the medieval Volga-Oka interfluve

Graphical abstract



Highlights

- Iron Age inhabitants of Suzdal were genetically unique but close to Uralic speakers
- A shift in the local gene pool coincided with Slavic migrations and a language shift
- Genetic changes mirror the insights from historical linguistics and written records
- Outliers suggest far-reaching contacts during the medieval times

Authors

Sanni Peltola, Kerttu Majander, Nikolaj Makarov, Maria Dobrovolskaya, Kerkko Nordqvist, Elina Salmela, Päivi Onkamo

Correspondence

sanni.peltola@helsinki.fi (S.P.), paivi.onkamo@utu.fi (P.O.)

In brief

Peltola et al. present ancient DNA data of 31 individuals dating from 200 to 1,800 CE from the Suzdal region in the Volga-Oka interfluve. They describe a previously unsampled Iron Age population and show that a genetic shift in the medieval times coincided with the historically recorded language shift from Uralic to Slavic.

Peltola et al., 2023, Current Biology *33*, 174–182
January 9, 2023
https://doi.org/10.1016/j.cub.2022.11.036

CellPress

CellPress
OPEN ACCESS

Current Biology

Report

Genetic admixture and language shift in the medieval Volga-Oka interfluve

Sanni Peltola,[1,2,7,*] Kerttu Majander,[3] Nikolaj Makarov,[4] Maria Dobrovolskaya,[4] Kerkko Nordqvist,[5] Elina Salmela,[1,2,6] and Päivi Onkamo[6,*]

[1]Faculty of Biological and Environmental Sciences, University of Helsinki, 00014 Helsinki, Finland
[2]Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany
[3]Department of Evolutionary Anthropology, University of Vienna, 1030 Vienna, Austria
[4]Institute of Archaeology, Russian Academy of Sciences, 117292 Moscow, Russia
[5]Department of Cultures, Archaeology, University of Helsinki, 00014 Helsinki, Finland
[6]Department of Biology, University of Turku, 20014 Turku, Finland
[7]Lead contact
*Correspondence: sanni.peltola@helsinki.fi (S. P.), paivi.onkamo@utu.fi (P. O.)
https://doi.org/10.1016/j.cub.2022.11.036

SUMMARY

The Volga-Oka interfluve in northwestern Russia has an intriguing history of population influx and language shift during the Common Era. Today, most inhabitants of the region speak Russian, but until medieval times, northwestern Russia was inhabited by Uralic-speaking peoples.[1–3] A gradual shift to Slavic languages started in the second half of the first millennium with the expansion of Slavic tribes, which led to the foundation of the Kievan Rus' state in the late 9th century CE. The medieval Rus' was multicultural and multilingual—historical records suggest that its northern regions comprised Slavic and Uralic peoples ruled by Scandinavian settlers.[4–6] In the 10th–11th centuries, the introduction of Christianity and Cyrillic literature raised the prestige status of Slavic, driving a language shift from Uralic to Slavic.[3] This eventually led to the disappearance of the Uralic languages from northwestern Russia. Here, we study a 1,500-year time transect of 30 ancient genomes and stable isotope values from the Suzdal region in the Volga-Oka interfluve. We describe a previously unsampled local Iron Age population and a gradual genetic turnover in the following centuries. Our time transect captures the population shift associated with the spread of Slavic languages and illustrates the ethnically mixed state of medieval Suzdal principality, eventually leading to the formation of the admixed but fully Slavic-speaking population that inhabits the area today. We also observe genetic outliers that highlight the importance of the Suzdal region in medieval times as a hub of long-reaching contacts via trade and warfare.

RESULTS AND DISCUSSION

Genetic admixture between populations leaves traces in the genomes of subsequent generations, whereas linguistic contacts appear as alterations in languages.[7] In human history, the spread of languages has often coincided with the movement of people, but each can occur without the other.[8] Thus, inferring past population contacts based on either the current distributions of languages or genetic ancestry components alone can be misleading. Ancient DNA studies allow us to directly observe changes in a population's gene pool through time and to resolve the correlations between genetic, archaeological, and historical linguistic data.

Studies on modern genomes have shown that most present-day Uralic speakers, ranging from the Baltic Sea to western Siberia, share a modern-Siberian-like ancestry component.[9,10] One notable feature in the Uralic languages' current spatial distribution is the gap in the northwestern and central European Russia (Figure 1A). Evidence from historical linguistics suggests that Uralic speakers inhabited this area too, before the spread of Slavic, and historical sources name many of these groups.[1–3] Despite the language extinction, present-day northwestern Russians show pronounced affinity to their Uralic-speaking neighbors, suggesting a genetic contribution from the preceding Uralic-speaking population.[9,11]

We studied 32 ancient individuals from six archaeological sites located in the Suzdal region in the Volga-Oka interfluve, using DNA sequencing and radiocarbon dating (Figures 1A, 1B, and S1A–S1C; Data S1A; STAR Methods): Bolshoye Davydovskoye 2 (BOL), representing an Iron Age culture (3rd–4th cc.; n = 9)[14]; Shekshovo 9 (SHE), a burial site of a large medieval settlement (10th–12th cc.; n = 9)[15]; Shekshovo 2 (SHK), a later burial ground of the same settlement (late 12th–13th cc.; n = 2); and post-medieval burials from Kibol 3 (KBL) (18th c.; n = 3), Kidekscha (KED) (15th–18th cc.; n = 4), and Krasnoe 3 (KRS) (14th c.; n = 1). Additionally, we included one kurgan burial (GOR) (12th c.) and three medieval flat burials (GOS) (12th–13th cc.) from the town of Gorokhovets in the eastern part of the Vladimir region. We also measured stable isotope ratios of carbon and nitrogen from Bolshoye Davydovskoye 2 and Shekshovo 2 and 9 to reconstruct changes in diet and lifestyle (Figure S1D; STAR Methods).

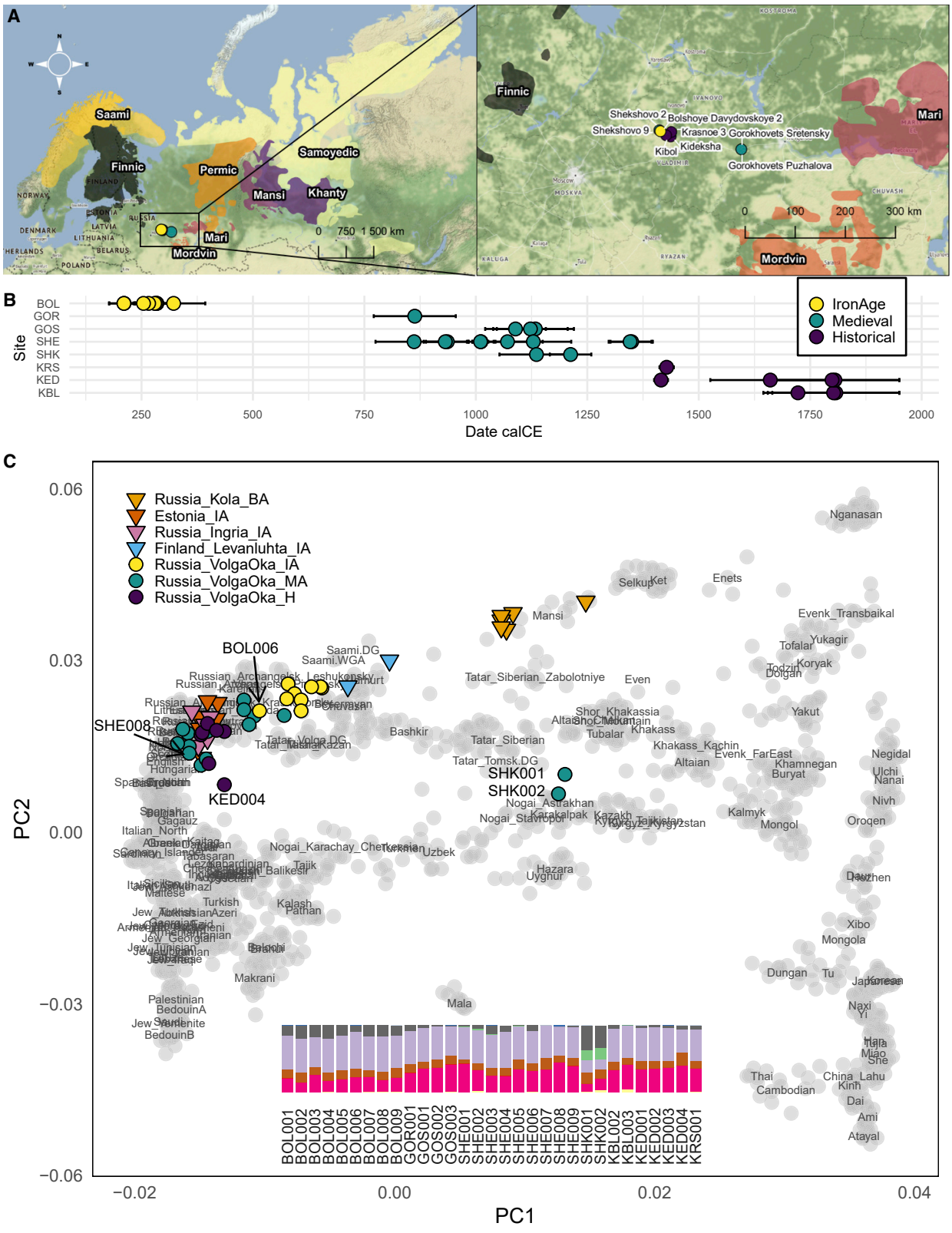


Figure 1. Geographic locations, radiocarbon dates, and PCA and ADMIXTURE results of the 32 samples from the Volga-Oka interfluve

(A) Locations of the archaeological sites included in this study and the speaker areas of extant Uralic branches.[12,13] The Hungarian speaker area resides south of the map.

(legend continued on next page)

Genetic shift in the Volga-Oka region in the Iron Age-medieval transition

We used an in-solution capture of 1.2 million genome-wide markers and obtained 0.06–4× on-target coverage from 31 samples that had sufficient DNA preservation (STAR Methods). To view our samples in the context of present-day genetic variation, we performed principal-component analysis (PCA). In a Eurasian-wide PCA, PC1 separates West Eurasia from East Asia, which are connected by three genetic clines, separated by PC2 (Figure 1C). These clines roughly correspond to ecoregions of Central Eurasia; the uppermost cline follows to the forest-tundra biome, and most Uralic-speakers fall on that cline.[10] Consistent with their geography, the Iron Age individuals from Bolshoye Davydovskoye fall on this cline, between the Russians from the coast of the White Sea (Arkhangelsk region) and the present-day Volga populations. The results of ADMIXTURE[16] also showed that Bolshoye Davydovskoye individuals carried Siberian ancestry, as do most present-day Uralic-speakers (Figures 1C and S2). This ancestry component is maximized in present-day Nganasans from the Taymyr Peninsula, but it may have been more widespread in the past.

The post-Iron Age samples from Shekshovo 9, Gorokhovets, Kibol, Kideksha, and Krasnoe fall closer to the West European cluster on the PCA than the Bolshoye Davydovskoye group, indicating a genetic shift after the Iron Age (Figures 1C and S3). However, approximately half of the medieval individuals still fall close to Bolshoye Davydovskoye individuals, whereas the other half cluster with present-day East Slavs. This pattern suggests an ongoing admixture or presence of two distinct populations at the time. Post-medieval individuals fall largely among modern south-central European Russians and did not show similar variation in ancestry as their predecessors.

The results from the dietary isotope analysis also reflect a shift between the Iron Age and medieval times (Figure S1D). The Bolshoye Davidovskoye individuals showed a diet with a high protein consumption and C4 plant use (likely millet), whereas the Shekshovo 9 individuals had a C3-plant-based diet. Despite genetic scattering of the Shekshovo 9 samples, isotopic values indicate a similar diet within this group.

For downstream analyses, we assigned our samples to four analysis clusters based on archaeological context, radiocarbon date, and the PCA and ADMIXTURE results (Figure S3; Data S1A): Iron Age (VolgaOka_IA, n = 7), medieval Iron Age-like (VolgaOka_MA1, n = 4), medieval East European-like (VolgaOka_MA2, n = 6), and post-medieval individuals from Kideksha and Kibol (VolgaOka_H, n = 4) (Figure S3). Two samples were excluded from the analysis clusters due to their close genetic relatedness to other individuals—one due to contamination and seven because they were genetic or chronological outliers.

Iron Age Volga population shows genetic affinities to Siberia

To measure allele sharing between our analysis clusters and present-day populations, we calculated outgroup f_3 statistics. The results indicated high levels of allele sharing with Lithuanians for all analysis clusters (Data S2A). For each analysis cluster, we tested the symmetry of relatedness (cladality) with relevant present-day populations ("Target") by calculating f_4 statistics "f_4(Mbuti, Modern_group; Analysis_cluster, Target)" (Figure 2A; Data S2B; STAR Methods). We found that VolgaOka_MA2 was cladal with non-Russian East Slavs (Belarusians, Ukrainians, and North Ukrainians) and present-day Russians from the Ryazan region, whereas VolgaOka_H was cladal only with the Ryazan Russians. VolgaOka_IA and VolgaOka_MA1 were not cladal with any of the tested Targets; however, they had the smallest number of significantly non-zero estimates with present-day Russians from Archangelsk region, suggesting that northern European Russians are their closest contemporary relatives.

The cladality tests revealed an allele sharing boundary between the East Eurasians and VolgaOka_IA and VolgaOka_MA1 (Figure 2A). The boundary approximately corresponds to the Ural Mountains: VolgaOka_IA and VolgaOka_MA1 shared significantly more alleles with the populations east of the Urals than present-day Lithuanians. Nganasans gave the lowest f_4 estimate for both VolgaOka_IA and VolgaOka_MA1, suggesting that the observed eastern affinity stems from the same Siberian ancestry we saw in the ADMIXTURE analysis (Figure S2; Data S2B). A similar but more subtle pattern in allele sharing was also observed for VolgaOka_H, but not for VolgaOka_MA2, indicating that VolgaOka_MA2 does not carry additional Siberian ancestry.

To study the relative affinity to Siberian ancestry (using a Bronze Age southern Siberian genome[17] as a proxy) in our study populations, we calculated "f_4(Mbuti, Krasnoyarsk_Krai_BA; Test, Lithuanian)", where "Test" was substituted with selected modern and ancient populations, including our analysis clusters. The results show that in terms of allele sharing with Krasnoyarsk_Krai_BA, VolgaOka_IA and VolgaOka_MA1 fall in the same range with North Russians from Archangelsk and Vologda, Veps, Karelians, Finns, and Mordovians (Figure 2B; Data S2C).

The Volga-Oka region is close to the area that was inhabited by East European hunter-gatherers (EEHGs) during the Mesolithic. This group contributed ancestry to many later populations. To examine the relative affinity to EEHG, we calculated "f_4(Mbuti, EEHG; Test, Lithuanian)". The results indicated excess EEHG ancestry in VolgaOka_IA, Estonia_BA, and Russia_Kola_BA, although the sharing was statistically significant only for Russia_Kola_BA (Figure 2B; Data S2C).

Central Russian population formed as a product of local admixture

To model the ancestry composition of our analysis clusters and other relevant populations, we used qpAdm. First, we constructed a consistent distal model with five sources (Krasnoyarsk_Krai_BA, West European hunter-gatherers [WEHGs], EEHG, Yamnaya_Samara, and European Neolithic

(B) Calibrated radiocarbon dates of the samples (see also Figures S1A–S1C and Data S1A). Bars indicate mean calibrated radiocarbon date ± 2σ.
(C) PCA of 164 present-day Eurasian groups (gray). Labels indicate median coordinates of the groups. Ancient populations (colored symbols) are projected on the present-day variation. Circles indicate 31 new samples from this study; upward triangles indicate previously published samples (see also Figure S3 and Data S1B). Vertical bars show the ADMIXTURE result at K = 9 for the 31 samples (full results in Figure S2). BA, Bronze Age; IA, Iron Age; MAs, Middle Ages; H, historical.

CellPress
OPEN ACCESS

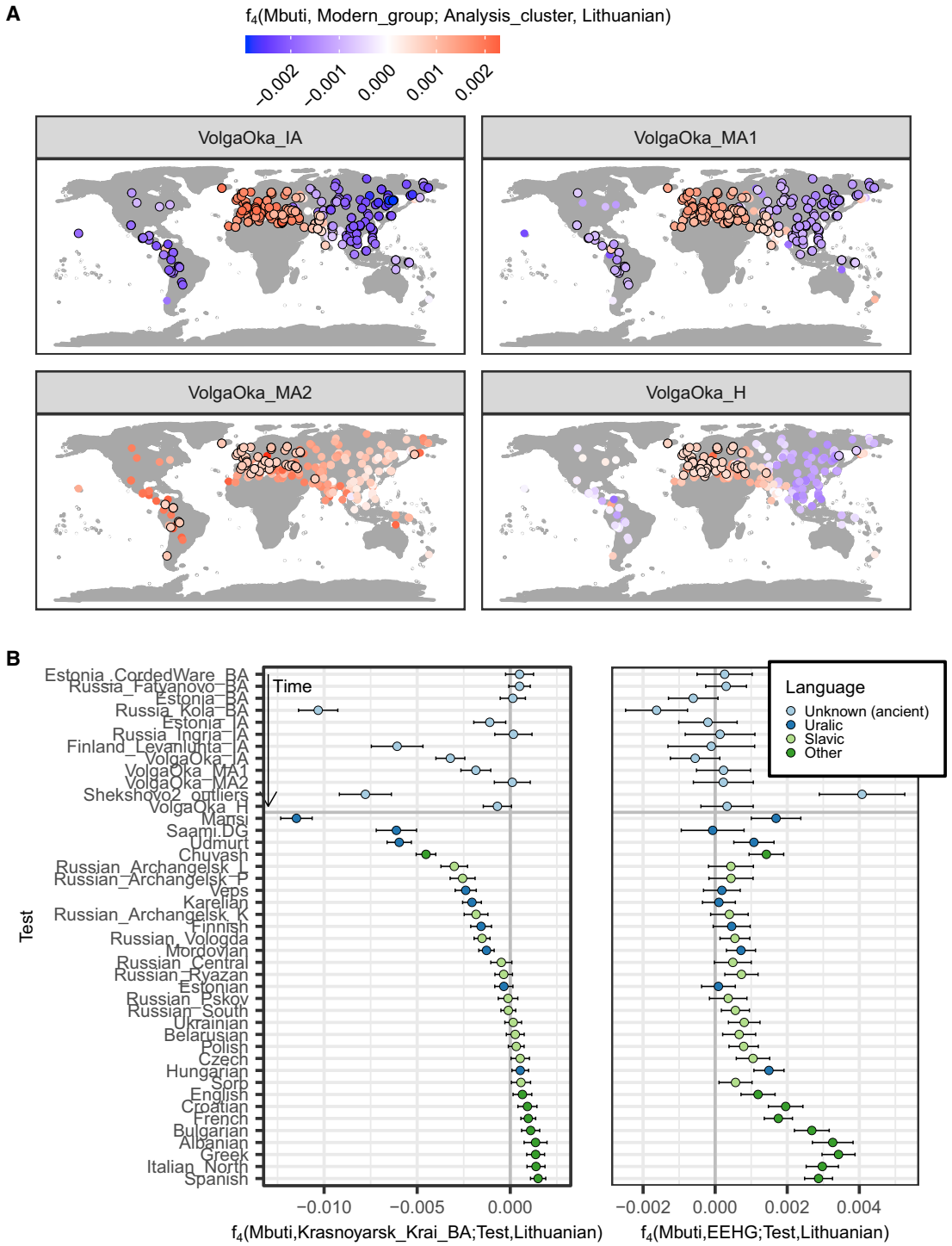


Figure 2. Results from f_4 statistics

(A) Tests for cladality with Lithuanians compared with 400 present-day populations. Outlined circles indicate statistically significant ($|Z| \geq 3$) estimates of f_4 (see also Data S2B).

Blue hues indicate that the analysis cluster shares more alleles with the tested modern group than present-date Lithuanians, and red hues indicate vice versa. (B) Relative affinity of ancient and present-day groups to Siberian ancestry (Krasnoyarsk_Krai_BA; left panel) and EEHG ancestry (right panel) in ancient and modern groups (see also Data S2C). Present-day populations are colored by their language group. Ancient groups are ordered by time and present-day groups by their Siberian affinity. Error bars indicate three standard errors to show statistically significant deviations from 0.

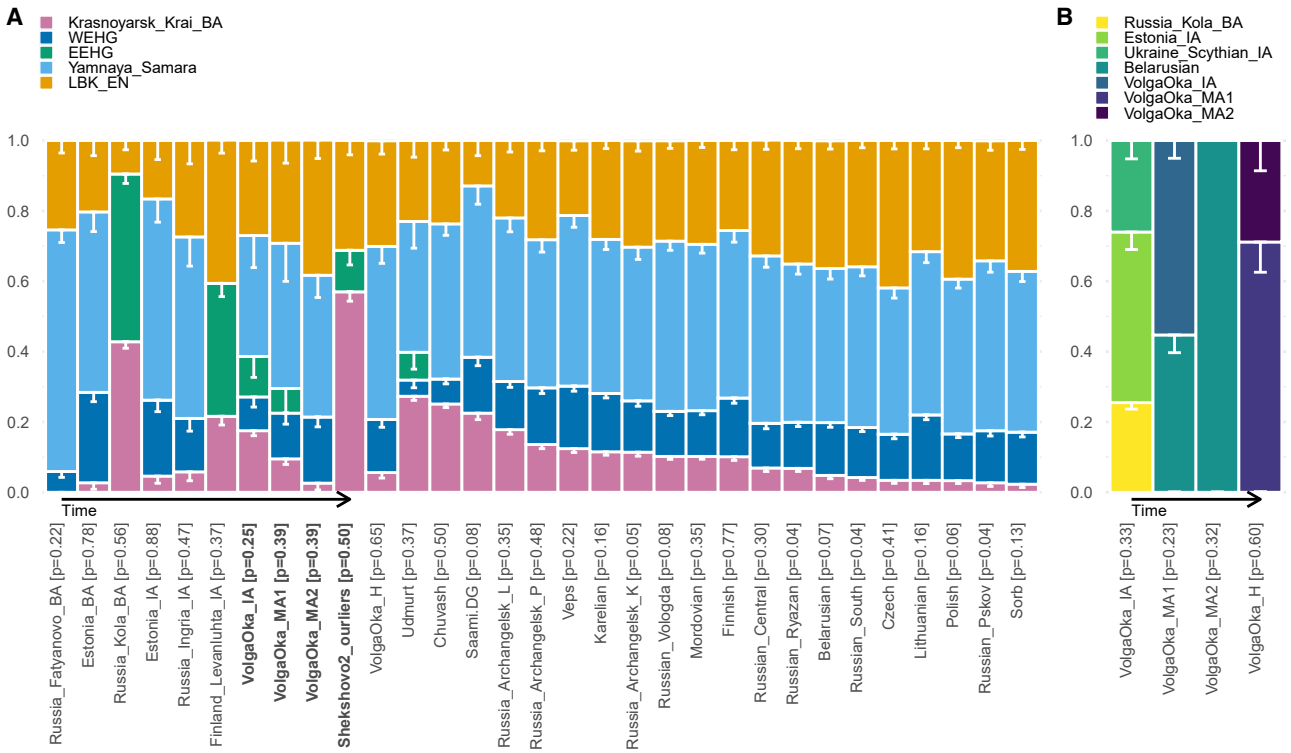


Figure 3. Results from qpAdm analyses

(A) Distal model fitted on selected ancient and present-day groups. Ancient populations are in the order from the oldest to youngest as indicated by an arrow. Present-day groups are in the descending order by Siberian ancestry. Error bars indicate one standard error. p values from chi-square test for each model are shown inside the square brackets. We show models that have a p value $\geq$ 0.01. Right populations used: Ethiopia_4500BP.SG, CHG, Raqefet_M_Natufian, Onge, Villabruna, ANE, Mixe, and SHG (see also Data S3A).

(B) Sequential proximal models for the Volga-Oka time transect groups. Right populations: Ethiopia_4500BP.SG, Krasnoyarsk_Krai_BA, WEHG, EEHG, Yamnaya_Samara, and LBK_EN.

LBK_EN, European Neolithic farmers; EEHGs, East European hunter-gatherers; WEHGs, West European hunter-gatherers (see also Data S3B and S3C).

farmers [LBK_EN]) across all groups, followed by proximal models customized for each analysis cluster (STAR Methods).

In the distal model, only VolgaOka_IA, VolgaOka_MA1, and present-day Udmurt had the best fit with all five sources (Figure 3A; Data S3A). For most populations, a four-way model without EEHG was a better fit. Finland_Levanluhta_IA and Russia_Kola_BA also harbored EEHG ancestry, but unlike VolgaOka_IA, VolgaOka_MA1, and Udmurt, they were best modeled without WEHG and Steppe ancestry. The EEHG component is also present in Shekshovo 2 outliers, but we caution that our model is targeted to populations of north-western Eurasia and may not provide meaningful results for these outliers.

To pinpoint temporally or geographically proximate populations that may have contributed to VolgaOka_IA, we tested several potential models. We used proxy sources due to sparse sampling of the area. We used Russia_Kola_BA as the source of Siberian ancestry because it harbored both Siberian and EEHG components, which we detected in VolgaOka_IA. As Western sources, we tested ancient Baltic populations and Fatyanovo, and for Steppe ancestry, we used Iron Age and Middle-Late Bronze Age Steppe groups from surrounding regions. The best-fitting models indicated that VolgaOka_IA shared approximately half of its ancestry with a population related to Baltic Iron Age individuals (800 BCE–50 CE), and 25% of its ancestry related to Russia_Kola_BA and 25% to Iron Age Steppe (Figure 3B; Data S3B). Thus, the Volga-Oka interfluve appears to have been at the crossroads of gene flow from several directions, although we do note that the proximal origin of these components in the Suzdal Iron Age gene pool may lie elsewhere. Interestingly, Fatyanovo did not provide a feasible ancestry source for VolgaOka_IA in any of the tested models, indicating a lack of genetic contribution from the Fatyanovo people who inhabited the Volga-Oka region in the Bronze Age.

We used VolgaOka_IA in turn as a source to model the medieval admixture between early Slavs and Uralic-speaking groups. Lacking ancient DNA data from early Slavs, we tested modern populations to approximate Slavic ancestry: Belarusian, Ukrainian, and North Ukrainian (East Slavs) and Polish and Sorb (West Slavs). East Slavs generally provided a better source of Slavic ancestry than West Slavs. We found that VolgaOka_MA2 could be sufficiently modeled with Slavs as the sole source, supporting the idea that VolgaOka_MA2 represents an unadmixed early Slavic population (Figure 3B; Data S3B). On the contrary, VolgaOka_MA1 required a substantial contribution from VolgaOka_IA. Post-medieval VolgaOka_H also required VolgaOka_IA-related ancestry, but in a smaller proportion. Alternatively, VolgaOka_H could be modeled as a mixture of

VolgaOka_MA1 and VolgaOka_MA2 clusters, consistent with local admixture. According to this model, the Slavic-like VolgaOka_MA2 contributed approximately 70% of ancestry to the post-medieval population.

We extended the above mixture models to individual level to assess variation in the admixture proportions in post-Iron Age individuals. Consistent with the cluster-level models, most individuals assigned to the VolgaOka_MA2 cluster could be modeled without any VolgaOka_IA ancestry (Data S3D). However, individuals assigned to the VolgaOka_MA1 cluster all carried varying proportions (35%–75%) of VolgaOka_IA ancestry.

Finally, we dated the arrival of the Siberian ancestry into our analysis clusters using DATES.[18] We used Nganasans and Lithuanians as proxy sources for Siberian and European ancestries. VolgaOka_IA and VolgaOka_MA1 had admixture times that corresponded to Bronze Age and Iron Age, respectively (Data S3E). Meanwhile, VolgaOka_MA2 and VolgaOka_H had more recent admixture times: the mean admixture dates (assuming a 29-year generation time) fell on the 8th and 9th centuries, respectively, corresponding well with the assumed beginning of the Slavic settlement in the Volga-Oka region.

Outliers highlight the connectedness of the medieval Suzdal

We detected several genetic outliers in our dataset. On the West Eurasian PCA, one individual from Bolshoye Davydovskoye 2 (BOL006) falls closer to their medieval successors than the main VolgaOka_IA group (Figure S3). Similarly, one individual from Shekshovo 9 (SHE008) falls closer to Central and West Europeans than the rest of the Shekshovo 9 individuals. Most strikingly, both individuals from Shekshovo 2 (SHK001 and SHK002) fell far from other individuals in our dataset, closer to East Asia and the "forest-steppe cline" in the Eurasian-wide PCA (Figure 1C). Their closest PCA neighbors were Kazakhs, Karakalpaks, Siberian Tatars, and other Turkic-speaking groups from Central Asia and Siberia. Both individuals also carried mitochondrial haplogroups that are more common in Asia than in Europe.[19,20] In cladality f_4 tests, we found the smallest number of significantly non-zero estimates with Karakalpaks, suggesting genetic similarity (Data S2B). In a previous study, their strontium values indicated a non-local origin,[21] further supporting the interpretation that these individuals had moved to the Volga area in their adulthood. These two men, who died at a young age, may represent Turkic-speaking groups whose members were in the military service in Kievan Rus', mostly guarding its southern borders.[22,23]

Lastly, the earliest individual from Kideksha (KED004) carried an intriguing mixture of ancestries. On the West Eurasian PCA, this individual falls within the space between European and Iranian clines (Figure S3). This individual also carried a geographically unusual mitochondrial haplogroup F2e, mostly found in present-day East Asians and mainland Southeast Asians.[24] ADMIXTURE suggested that this individual has genetic affinity to ancient and present-day Iranians, and thus we fitted qpAdm models with temporally proximate Iranian-related sources (Figure S2; Data S3D). The best model included medieval Alans from northern Caucasus in addition to VolgaOka_IA and Slavic-like ancestry. These findings highlight the connectedness and importance of the Volga-Oka interfluve during the medieval era.

Parallel development of genes and languages in the Volga-Oka interfluve

The past 1,500 years of history of the Volga-Oka interfluve are characterized by a gradual language shift from Uralic to Slavic.[3] A corresponding pattern emerged from our genetic data. The Iron Age inhabitants of the Volga-Oka interfluve carried Siberian ancestry, which places them on the same genetic continuum with most present-day Uralic-speaking populations. In our modeling framework, the local Iron Age group provided a fitting source of Siberian ancestry for most of our medieval individuals. However, archaeological evidence suggests that the Bolshoye Davydovskoye people represented a unique culture that disappeared already by the 7th century,[11] making them an unlikely candidate for the direct ancestors of medieval groups. Thus, although the Bolshoye Davydovskoye group may have contributed to the later population, the Uralic-speaking people who lived in the Volga-Oka interfluve at the time of the Slavic migration likely represented a closely related but separate group. One such group could be Meryans, a now-extinct Uralic-speaking group mentioned in the Chronicles, whose reconstructed speaker area covered the Suzdal region.[25,26]

Slavic migrations in the latter half of the first millennium shaped the linguistic landscape of northwestern Russia.[1,5,27] In the 10th–12th centuries, Slavic and Uralic-speaking groups often formed multilingual communities in the northeastern regions of Kievan Rus', where Suzdal lies. Concordantly, our dataset captures the arrival of the Slavic ancestry component and the medieval coexistence of Slavic-like and Uralic-like groups. In Shekshovo 9, we detected approximately equal numbers of individuals from both genetic groups, and their burial placement showed no apparent distinction between them. Moreover, some individuals with Uralic-like ancestry were buried with "Slavic" grave goods or a mixture of Slavic and "Uralic" items, indicating cultural integration of the groups. However, our model suggests that the Slavic-like group contributed a major proportion (70%) of ancestry to the later population. Obviously, our medieval sample may be too small to be fully representative, but the difference could also suggest additional contribution from the surrounding Slavic population in the Late Middle Ages.

Whereas historical sources indicate strong Scandinavian influence in early Rus', we did not detect Scandinavian ancestry in our medieval individuals, which may suggest that the majority of the population in medieval Suzdal comprised of Uralic and Slavic peoples. Alternatively, the individuals with Scandinavian ancestry may have been less frequently buried in the cemeteries we sampled.

Conclusions

Our unique time transect of ancient DNA data shows that the Volga area has been at the crossroads of population interaction for the last two millennia. The local Iron Age gene pool carried three main ancestry components resembling populations from Iron Age eastern Baltics, Iron Age Steppe, and Bronze Age Kola Peninsula. Intriguingly, these sources also connected the Iron Age individuals to local Mesolithic hunter-gatherers. Meanwhile, their gene pool seemed largely discontinuous from the nearby Bronze Age population of Fatyanovo.[28] The early Middle Ages in turn saw a shift in diet and the arrival of a Slavic-like genetic component, which tightly mirrors insights from historical

linguistics and written records. The medieval genetic diversity was further bolstered by long-distance migrants with genetic affinity to Central Asia and Iran, underlining the region's long-distance connections. Admittedly, the dynamics we have discovered may be very local: the genetic structure in present-day Russians suggests that the details of the Slavic admixture process may have varied by area.[9,11] Nevertheless, our results indicate that ancient DNA may also provide indirect evidence of language history when linguistic data are sparse.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Bolshoye Davydovskoye 2
  - Shekshovo 9
  - Shekshovo 2
  - Gorokhovets, Puzhalova gora
  - Gorokhovets, Sretensky monastery
  - Krasnoe 3
  - Kibol 3
  - Kideksha
- METHOD DETAILS
  - Radiocarbon dating and calibration
  - Stable isotope analysis
  - Ancient DNA sample processing and quality control
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data processing and quality control
  - Ancient DNA authentication
  - Biological sex determination and uniparental markers
  - Runs of homozygosity and biological relatedness
  - Genotyping and merging with reference data
  - Principal component analysis
  - Admixture analysis
  - Analysis cluster assignment
  - F_3 and f_4 statistics
  - qpAdm modelling
  - DATES analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2022.11.036.

ACKNOWLEDGMENTS

We thank Nelli-Johanna Saari for carrying out laboratory work, the sequencing team at the Max Planck Institute for the Science of Human History, and the lab technicians involved in this project. We thank Prof. Dr. Johannes Krause for resources and comments on the manuscript and Dr. Outi Vesakoski for helpful discussion and comments on an earlier version of the manuscript. The Finnish Center of Scientific Computing (CSC) and the Max Planck Institute provided computational resources. This project was funded by the Emil Aaltonen Foundation (S.P. and K.M.), Ella and Georg Ehrnrooth Foundation (S.P. and E.S.), Jenny and Antti Wihuri Foundation (E.S.), Jane and Aatos Erkko Foundation, the Kone Foundation and the Max Planck Society. This work is part of the SUGRIGE project (https://sites.utu.fi/paleogenetics/).

AUTHOR CONTRIBUTIONS

P.O. and N.M. conceived the study. P.O. acquired funding. P.O. and E.S. supervised the work. P.O., N.M., and M.D. assembled the collection of archaeological specimens, and N.M., M.D., and K.N. provided archaeological expertise. M.D. analyzed the stable isotope data. K.M. participated in the laboratory work. S.P. and K.M. performed bioinformatic processing. S.P. analyzed the genetic data and wrote the original draft of the manuscript. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 1, 2022
Revised: September 23, 2022
Accepted: November 17, 2022
Published: December 12, 2022

REFERENCES

1. Leontyev, A.E. (1996). Arkheologiya meri. K predystorii Severo-Vostochnoy Rusi (Archaeology of the Merya People. Towards the Pre-History of Northeastern Rus') (Geoeko).
2. Vasmer, M. (1935). Beiträge zur historischen Völkerkunde Osteuropas. III: Merja und Tscheremissen (Verlag d. Akad. d. Wiss.).
3. Saarikivi, J. (2006). Substrata Uralica - Studies on Finno-Ugrian Substrate. PhD thesis (University of Helsinki).
4. Lind, J.H. (2006). Problems of ethnicity in the interpretation of written sources on early Rus'. Slavicization of the Russian North: Mechanisms and Chronology (University of Helsinki), pp. 246–258.
5. Saarikivi, J. (2009). Itämerensuomalais-slaavilaisten kontaktien tutkimuksen nykytilasta. In The Quasquicentennial of the Finno-Ugrian Society (Suomalais-Ugrilaisen Seuran Toimituksia = Mémoires de la Société Finno-Ougrienne 258), J. Ylikoski, ed. (Société Finno-Ougrienne), pp. 109–160.
6. Makarov, N.A. (2017). Historiche Zeugnisse und archäologische Realien: Auf der Suche nach Übereinstimmungen. In Die Rus' im 9-10 Jahrhundert: ein archäologisches Panorama (Wachholtz), pp. 456–469.
7. Thomason, S.G., and Kaufman, T. (1992). Language Contact, Creolization, and Genetic Linguistics (University of California Press).
8. Hunley, K. (2015). Reassessment of global gene-language coevolution. Proc. Natl. Acad. Sci. USA *112*, 1919–1920.
9. Tambets, K., Yunusbayev, B., Hudjashov, G., Ilumäe, A.M., Rootsi, S., Honkola, T., Vesakoski, O., Atkinson, Q., Skoglund, P., Kushniarevich, A., et al. (2018). Genes reveal traces of common recent demographic history for most of the Uralic-speaking populations. Genome Biol. *19*, 139.
10. Jeong, C., Balanovsky, O., Lukianova, E., Kahbatkyzy, N., Flegontov, P., Zaporozhchenko, V., Immel, A., Wang, C.C., Ixan, O., Khussainova, E., et al. (2019). The genetic history of admixture across inner Eurasia. Nat. Ecol. Evol. *3*, 966–976.
11. Kushniarevich, A., Utevska, O., Chuhryaeva, M., Agdzhoyan, A., Dibirova, K., Uktveryte, I., Möls, M., Mulahasanovic, L., Pshenichnov, A., Frolova, S., et al. (2015). Genetic heritage of the Balto-Slavic speaking populations: a synthesis of autosomal, mitochondrial and Y-chromosomal data. PLoS One *10*, e0135820.
12. Rantanen, T., Vesakoski, O., Ylikoski, J., and Tolvanen, H. (2021). Geographical database of the Uralic languages. (v1.0) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.4784188.

13. Rantanen, T., Tolvanen, H., Roose, M., Ylikoski, J., and Vesakoski, O. (2022). Best practices for spatial language data harmonization, sharing and map creation – a case study of the Uralic languages. PLoS One *17*, e0269648.

14. Makarov, N., Krasnikova, A., and Zaytseva, I. (2010). Mogilnik Bolshoye Davydovskoye 2 – pogrebalniy pamyatnik pervoy poloviny I tys. n.e. v Suzdalskom Opolye (The Cemetery of Bolshoye Davydovskoye 2 – A Burial Site from the First Half of the First Millennium CE). Rossiyskaya Arkheologiya *1*, 41–52.

15. Makarov, N., Krasnikova, A., Zaitseva, I., and Dobrovolskaya, M. (2021). New evidence on the late Viking Age burial rituals in the Volga-Oka region: excavations at Shekshovo. Archaol. Korrespondenzblatt *51*, 287–308.

16. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.

17. Kılınç, G.M., Kashuba, N., Koptekin, D., Bergfeldt, N., Dönertaş, H.M., Rodríguez-Varela, R., Shergin, D., Ivanov, G., Kichigin, D., Pestereva, K., et al. (2021). Human population dynamics and *Yersinia pestis* in ancient northeast Asia. Sci. Adv. *7*, 1–14.

18. Narasimhan, V.M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., Mallick, S., Lazaridis, I., Nakatsuka, N., Olalde, I., Lipson, M., et al. (2019). The formation of human populations in South and Central Asia. Science *365*, eaat7487.

19. Peng, M.S., Palanichamy, M.G., Yao, Y.G., Mitra, B., Cheng, Y.T., Zhao, M., Liu, J., Wang, H.W., Pan, H., Wang, W.Z., et al. (2011). Inland post-glacial dispersal in East Asia revealed by mitochondrial haplogroup M9a'b. BMC Biol. *9*, 2.

20. Fedorova, S.A., Reidla, M., Metspalu, E., Metspalu, M., Rootsi, S., Tambets, K., Trofimova, N., Zhadanov, S.I., Hooshiar Kashani, B.H., Olivieri, A., et al. (2013). Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. BMC Evol. Biol. *13*, 127.

21. Dobrovolskaya, M.V., Makarov, N.A., and Samorodova, M.A. (2020). Mobility of the Suzdal Opolye settlers in 900–1150 AD. Arheol. ètnogr. antropol. Evrazii. *48*, 106–115.

22. Pletneva, S.A. (1982). Kochevniki srednevekovya. Poisk istoricheskikh zakonomernostey (The Nomads of the Middle Ages. The Search for Historical Patterns) (Nauka), p. 192.

23. Pletneva, S.A. (1990). Polovtsy (Cuman People) (Nauka).

24. Derenko, M., Malyarchuk, B., Denisova, G., Perkova, M., Rogalla, U., Grzybowski, T., Khusnutdinova, E., Dambueva, I., and Zakharov, I. (2012). Complete mitochondrial DNA analysis of eastern Eurasian haplogroups rarely found in populations of northern Asia and Eastern Europe. PLoS One *7*, e32179.

25. Frog, M., and Saarikivi, J. (2015). De situ linguarum fennicarum aetatis ferreae. RMN Newsletter *9*, 64–115.

26. Rahkonen, P. (2013). South-eastern contact area of Finnic languages. PhD thesis (University of Helsinki).

27. Makarov, N.A. (2006). Cultural identity of the Russian North Settlers in the 10th–13th centuries: archaeological evidence and written sources. The Slavicization of the Russian North. Mechanisms and Chronology (Helsinki University Press), pp. 259–281.

28. Saag, L., Vasilyev, S.V., Varul, L., Kosorukova, N.V., Gerasimov, D.V., Oshibkina, S.V., Griffith, S.J., Solnik, A., Saag, L., D'Atanasio, E., et al. (2021). Genetic ancestry changes in Stone to Bronze Age transition in the east European plain. Sci. Adv. *7*, eabd6535.

29. Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes *9*, 88.

30. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. BMC Bioinformatics *15*, 356.

31. Jun, G., Wing, M.K., Abecasis, G.R., and Kang, H.M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. Genome Res. *25*, 918–925.

32. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

33. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

34. Neukamm, J., Peltzer, A., and Nieselt, K. (2021). DamageProfiler: fast damage pattern calculation for ancient DNA. Bioinformatics *37*, 3652–3653.

35. Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047–3048.

36. Fellows Yates, J.A., Lamnidis, T.C., Borry, M., Andrades Valtueña, A.A., Fagernäs, Z., Clayton, S., Garcia, M.U., Neukamm, J., and Peltzer, A. (2021). Reproducible, portable, and efficient ancient genome reconstruction with nfcore/eager. PeerJ *9*, e10947.

37. Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Pääbo, S., Krause, J., and Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proc. Natl. Acad. Sci. USA *111*, 2229–2234.

38. Daley, T., and Smith, A.D. (2013). Predicting the molecular complexity of sequencing libraries. Nat. Methods *10*, 325–327.

39. Okonechnikov, K., Conesa, A., and García-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics *32*, 292–294.

40. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

41. Lamnidis, T.C., Majander, K., Jeong, C., Salmela, E., Wessman, A., Moiseyev, V., Khartanovich, V., Balanovsky, O., Ongyerth, M., Weihmann, A., et al. (2018). Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. Nat. Commun. *9*, 5018.

42. Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biol. *16*, 224.

43. Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A., and Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. *44*, W58–W63.

44. Poznik, G.D. (2016). Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. Preprint at bioRxiv. https://doi.org/10.1101/088716.

45. Nakatsuka, N., Harney, É., Mallick, S., Mah, M., Patterson, N., and Reich, D. (2020). ContamLD: estimation of ancient nuclear DNA contamination using breakdown of linkage disequilibrium. Genome Biol. *21*, 199.

46. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

47. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190.

48. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093.

49. Ringbauer, H., Novembre, J., and Steinrücken, M. (2021). Parental relatedness through time revealed by runs of homozygosity in ancient DNA. Nat. Commun. *12*, 5425.

50. Monroy Kuhn, J.M., Jakobsson, M., and Günther, T. (2018). Estimating genetic kin relationships in prehistoric populations. PLoS One *13*, e0195491.

51. Frei, K.M., Makarov, N., Nosch, M.-L., Skals, I., Berghe, I. Vanden, and Zaytseva, I. (2016). An 11th century 2/2 Twill from a Burial in Shekshovo in Russia. Archaeol. Text. Rev. *58*, 34–42.

52. Zaytseva, I.E., and Stolyarova, E.K. (2018). Ob odnom pogrebenii v mogilnike Shekshovo v Suzdalskom Opolye [On one of the burials at the Shekshovo burial site in the Suzdal Opolye]. Zemli Rod. Minuvshaya Sudba, 73–91.

53. Reimer, P.J., Austin, W.E.N., Bard, E., Bayliss, A., Blackwell, P.G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R.L., Friedrich, M., et al. (2020). The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP). Radiocarbon *62*, 725–757.

54. Bronk Ramsey, C. (2009). Bayesian analysis of radiocarbon dates. Radiocarbon *51*, 337–360.

55. DeNiro, M.J. (1985). Postmortem preservation and alteration of in vivo bone collagen isotope ratios in relation to palaeodietary reconstruction. Nature *317*, 806–809.

56. Ambrose, S.H. (1993). Isotopic analysis of paleodiets: methodological and interpretive considerations. In Investigations of Ancient Human Tissue. Chemical Analysis in Anthropology, M.K. Stanford, ed. (Breach Science Publishers), pp. 59–130.

57. van Klinken, G.J. (1999). Bone collagen quality indicators for palaeodietary and radiocarbon measurements. J. Archaeol. Sci. *26*, 687–695.

58. Neumann, G., Valtuena, A.A., Yates, J.F., Stahl, R., and Brandt, G. (2020). Tooth sampling from the inner pulp chamber for ancient DNA extraction. protocols.io. https://doi.org/10.17504/protocols.io.bqebmtan.

59. Orfanou, E., Himmel, M., Aron, F., and Haak, W. (2020). Minimally-invasive sampling of pars petrosa (os temporale) for ancient DNA extraction. protocols.io. https://doi.org/10.17504/protocols.io.bqd8ms9w.

60. Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.L., and Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl. Acad. Sci. USA *110*, 15758–15763.

61. Velsko, I., Skourtanioti, E., and Brandt, G. (2020). Ancient DNA extraction from skeletal material. protocols.io. https://doi.org/10.17504/protocols.io.baksicwe.

62. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*. pdb.prot5448.

63. Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. *40*, e3.

64. Aron, F., Neumann, G., and Brandt, G. (2020). Half-UDG treated double-stranded ancient DNA library preparation for illumina sequencing. protocols.io. https://doi.org/10.17504/protocols.io.bmh6k39e.

65. Stahl, R., Warinner, C., Velsko, I., Orfanou, E., Aron, F., and Brandt, G. (2021). Illumina double-stranded DNA dual indexing for ancient DNA. protocols.io. https://doi.org/10.17504/protocols.io.bvt8n6rw.

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

67. Gansauge, M.T., Gerber, T., Glocke, I., Korlević, P., Lippik, L., Nagel, S., Riehl, L.M., Schmidt, A., and Meyer, M. (2017). Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. Nucleic Acids Res. *45*, e79.

68. Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature *528*, 499–503.

69. Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H.A., Kelso, J., and Pääbo, S. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. Proc. Natl. Acad. Sci. USA *110*, 2223–2227.

70. Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K.E., Rasmussen, S., Albrechtsen, A., Skotte, L., Lindgreen, S., Metspalu, M., Jombart, T., et al. (2011). An aboriginal Australian genome reveals separate human dispersals into Asia. Science *334*, 94–98.

71. Furtwängler, A., Reiter, E., Neumann, G.U., Siebke, I., Steuri, N., Hafner, A., Lösch, S., Anthes, N., Schuenemann, V.J., and Krause, J. (2018). Ratio of mitochondrial to nuclear DNA affects contamination estimates in ancient DNA analysis. Sci. Rep. *8*, 14075.

72. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74.

73. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature *513*, 409–413.

74. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient near East. Nature *536*, 419–424.

75. Saag, L., Laneman, M., Varul, L., Malve, M., Valk, H., Razzak, M.A., Shirobokov, I.G., Khartanovich, V.I., Mikhaylova, E.R., Kushniarevich, A., et al. (2019). The arrival of Siberian ancestry connecting the eastern Baltic to Uralic speakers further East. Curr. Biol. *29*, 1701–1711.e16.

76. Gallego Llorente, M., Jones, E.R., Eriksson, A., Siska, V., Arthur, K.W., Arthur, J.W., Curtis, M.C., Stock, J.T., Coltorti, M., Pieruccini, P., et al. (2015). Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. Science *350*, 820–822.

77. Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkhoshbacht, N., Candilio, F., Cheronet, O., et al. (2018). The genomic history of southeastern Europe. Nature *555*, 197–203.

78. Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., et al. (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. Nature *551*, 368–372.

79. Jones, E.R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R.L., Gallego Llorente, M., Cassidy, L.M., Gamba, C., et al. (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nat. Commun. *6*, 8912.

80. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. Nature *534*, 200–205.

81. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Jr., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature *505*, 87–91.

82. Olalde, I., Brace, S., Allentoft, M.E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., et al. (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. Nature *555*, 190–196.

83. Järve, M., Saag, L., Scheib, C.L., Pathak, A.K., Montinaro, F., Pagani, L., Flores, R., Guellil, M., Saag, L., Tambets, K., et al. (2019). Shifts in the genetic landscape of the western Eurasian steppe associated with the beginning and end of the Scythian dominance. Curr. Biol. *29*, 2430–2441.e10.

84. Unterländer, M., Palstra, F., Lazaridis, I., Pilipenko, A., Hofmanová, Z., Groß, M., Sell, C., Blöcher, J., Kirsanow, K., Rohland, N., et al. (2017). Ancestry and demography and descendants of Iron Age nomads of the Eurasian Steppe. Nat. Commun. *8*, 14615.

85. Damgaard, P.B., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliussen, T., Moreno-Mayar, J.V., Pedersen, M.W., Goldberg, A., Usmanova, E., et al. (2018). 137 ancient human genomes from across the Eurasian steppes. Nature *557*, 369–374.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Archaeological human remains	Russian Academy of Sciences	N/A
Critical commercial assays		
High Pure Viral Nucleic Acid Large Volume Kit	Roche	Cat# 5114403001
MinElute PCR Purification Kit	Qiagen	Cat# 28006
HiSeq 4000 SBS Kit (50/75 cycles)	Illumina	Cat# FC-410-1001/2
Deposited data		
FASTQ and BAM	This study	ENA: PRJEB57974
Software and algorithms		
AdapterRemoval v2.3.1	Schubert et al.[29]	https://github.com/MikkelSchubert/adapterremoval
ANGSD v0.933	Korneliussen et al.[30]	https://github.com/ANGSD/angsd
bamUtil v1.0.14	Jun et al.[31]	https://genome.sph.umich.edu/wiki/BamUtil
bedtools v2.29.2	Quinlan and Hall[32]	https://github.com/arq5x/bedtools2/releases
BWA v0.7.17-r1188	Li and Durbin[33]	http://bio-bwa.sourceforge.net/bwa.shtml
DamageProfiler v0.4.9	Neukamm et al.[34]	https://github.com/Integrative-Transcriptomics/DamageProfiler
eigenstrat_snp_coverage v1.0.2	https://github.com/TCLamnidis/EigenStratDatabaseTools	https://github.com/TCLamnidis/EigenStratDatabaseTools
endorS.py v0.4	https://github.com/aidaanva/endorS.py	https://github.com/aidaanva/endorS.py
FastQC v0.11.9	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
MTNucRatioCalculator v0.7	https://github.com/apeltzer/MTNucRatioCalculator	https://github.com/apeltzer/MTNucRatioCalculator
MultiQC v1.10.1	Ewels et al.[35]	https://multiqc.info/
nf-core/eager v2.3.3	Fellows Yates et al.[36]	https://nf-co.re/eager/2.3.3
Picard MarkDuplicates v2.22.9	N/A	http://broadinstitute.github.io/picard/
pmdtools v0.50	Skoglund et al.[37]	https://github.com/pontussk/PMDtools
Preseq v2.0.3	Daley and Smith[38]	http://smithlabresearch.org/software/preseq/
Qualimap v2.2.2-dev	Okonechnikov et al.[39]	http://qualimap.bioinfo.cipf.es/
Samtools v1.9	Li et al.[40]	http://www.htslib.org/
sequenceTools v1.4.0.5	https://github.com/stschiff/sequenceTools	https://github.com/stschiff/sequenceTools
sexdeterrmine v1.1.2	Lamnidis et al.[41]	https://github.com/TCLamnidis/Sex.DetERRmine.git
Schmutzi	Renaud et al.[42]	https://github.com/grenaud/schmutzi
Haplogrep-2	Weissensteiner et al.[43]	https://haplogrep.i-med.ac.at/
Yhaplo	Poznik[44]	https://github.com/23andMe/yhaplo
ContamLD	Nakatsuka et al.[45]	https://github.com/nathan-nakatsuka/ContamLD
Plink v1.9	Purcell et al.[46]	https://www.cog-genomics.org/plink/
smartpca v13050	Patterson et al.[47]	https://github.com/DReichLab/EIG
ADMIXTOOLS v6.0	Patterson et al.[48]	https://github.com/DReichLab/AdmixTools
Admixtools 2 v.2.0.0	Patterson et al.[48]	https://github.com/uqrmaie1/admixtools
ADMIXTURE v1.3.0	Alexander et al.[16]	https://dalexander.github.io/admixture/
hapROH v0.4a1	Ringbauer et al.[49]	https://github.com/hringbauer/hapROH
READ	Monroy Kuhn et al.[50]	https://bitbucket.org/tguenther/read/src/master/
DATES	Narasimhan et al.[18]	https://github.com/priyamoorjani/DATES

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Sanni Peltola (sanni.peltola@helsinki.fi).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The newly reported sequences reported in this study are available in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB57974.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

In this paper, we aim to associate linguistic shifts to genetic turnover. However, we want to emphasise that a spoken language or ethnic identity cannot be inferred from genetic data. Similarly, the link between material culture and ethnicity is a very complex one and we cannot assume that ethnicity or language would automatically override all other social or local group affiliations expressed in the material culture. For the purpose of this study, we have used simplified terminology: we use terms Slavic-like and Uralic-like to describe the genetic nature of the individuals, i.e. their similarity to present-day Slavic and Uralic-speaking population, and terms Slavic-speaking and Uralic-speaking to refer to linguistic groups.

Suzdal region is the historical core of northwestern Russia and it played an important role in political history of the Volga-Oka interfluve in the 11th–13th centuries. It is characterized by a continuous existence of rural settlements that emerged in the 10th–12th centuries, often in territories peopled already in the 1st millennium CE. The region is considered part of the area inhabited by Uralic-speaking groups until the 9th–10th centuries, while the Slavic colonization began in the 10th century.[25,27] The Russian Primary Chronicles associate Meryans, an extinct Uralic-speaking group, as the first inhabitants of the region.

Medieval cities (Suzdal, Vladimir, Yuryev) and rural settlements in Suzdal region have been thoroughly studied by archaeologists. The material culture of these is found to represent classical Ancient Rus' tradition of the 11th–13th cc The preceding period is characterized by a complex and multi-component material culture, combining elements of both Slavic and Volga Finnic groups. Thus, the ethnocultural history of the Suzdal land in the Middle Ages can be viewed as one of the formation and consolidation of a regional group of medieval Rus' (referred to in the chronicles as "Suzdal People") on the territory previously occupied by the Volga Finns. Nonetheless, the details of the interaction and merging of Slavic and Finnic populations are largely unknown. The necropolises of the 10th–12th centuries in the Suzdal region are mainly known from the excavations conducted in the mid-19th century, and their documentation does not give a clear idea of their burial rite and dating. The settlements of the second half of the first millennium, i.e. the period preceding the Slavic habitation, have hardly been studied and no burial grounds are known from this time either.

For the archaeogenetic study, we selected bone materials from six burial sites excavated in the Suzdal region in the 2000s and 2010s and from two burial sites in Gorokhovets, located on the eastern outskirts of the Suzdal lands (about 140 km from the Suzdal region). These burial sites belong to different periods, spanning from the Iron Age (3rd c. CE) to the historical era (19th c.). All the investigated Suzdal burial grounds are located within a small area: the greatest distance between two burial sites is no more than 20 km. Thus, we have a unique opportunity to characterize genetic changes in different periods in the Volga-Oka region.

Bolshoye Davydovskoye 2

The samples from Bolshoye Davydovskoye 2 burial ground represent the Iron Age inhabitants of the Suzdal region. It is the only currently known necropolis of the first half of the 1st millennium CE in the Suzdal region. The site is located 20 km northwest of Suzdal and was fully excavated in 2007–2010. The burial ground consists of flat inhumation graves, covered by the occupational deposits of a later settlement (10th–12th cc.). Excavations unearthed 18 burial pits, arranged in two rows and containing the remains of 22 individuals, most of whom were richly furnished. Metal clothing ornaments date the cemetery to the second half of the 3rd and 4th centuries.[14] The character of the burial ritual and the female costume connect the burial site with the Ryazan-Oka culture, one of the Finnic Iron Age cultures. Its core area was in the Middle Oka region and it dates from the 2nd to early 7th centuries CE – Bolshoye Davydovskoye 2 can be regarded the northern outpost of this culture.

Grave 2 (BOL001)

Skeletal remains from rectangular burial pit, oriented east-west, partly destroyed by looters. Shallow (up to 15 cm deep) pit contained disturbed remains of two individuals: a young male and a female of 20–29 aged (BOL001). Fragmented remains of the skull of the latter individual were found in the eastern part of the grave. Grave goods: fragments of temporal rings (near the skull and in mixed soil) and details of headband (in mixed soil).

Grave 3-1 (BOL002)

A wooden structure made of linden in a large (300 ч 120 cm) rectangular east–west oriented burial pit contained skeletal remains of five individuals in poor state of preservation. BOL002 was discovered in the northeastern part of the pit, lying in supine position with its head to the east and limbs slightly flexed. Skeletal remains belong to a female individual in the age of 25–34 years. Grave goods: glass

pendant, temporal rings (near the skull), neck ring, two spindle whorls, two buckles, lash handle and remains of a lash, enameled horseshoe-shaped brooch (on the chest), arm ring, finger ring, dress plaques.

Grave 3-5 (BOL003)

Grave details as for BOL002. Skeletal remains of a child aged 7–8 years in poor state of preservation were discovered in the north-western part of the pit, lying supine with its head to the west, near burial 3-3 and partly covering it. Grave goods: neck ring (on the neck), glass beads, metal plaques, brooch, two arm rings, finger ring, three jingling pendants, duck pendant, iron knife.

Grave 6 (BOL004)

Burial in an oval pit up to 29 cm deep with remains of constructions made of linden. Skeletal remains of an 8–10-year-old child in poor state of preservation lay supine with its head to the east. Grave goods: two neck rings, pendants, round brooch (on the chest), plaques.

Grave 9 (BOL005)

Burial in a rectangular pit up to 22 cm deep, skeletal remains in poor state of preservation. An adult individual lying supine with its head to the east. Grave goods: temporal ring (near the skull), brooch (on the chest), iron artefact (near the right foot).

Grave 15 (BOL006)

Burial in a rectangular pit up to 27 cm deep contained poorly preserved human remains (a skull) and remains of a timber construction. A female of the age of 30–39 years was lying in supine position with her head to the east. Grave goods: two temporal rings (near the skull), finger ring, clips and plaques, glass beads, round brooch, iron knife, spindle whorl, fragmented clay vessel.

Grave 17 (BOL007)

Burial in a rectangular pit 25 cm deep contained poorly preserved remains of two individuals: a male aged 20–29 years and a teenager of 12–14 years. Skeletal remains were strongly disturbed. BOL007 is a male individual lying supine with his head to the east. Grave goods: plaques, iron brooch, iron knife.

Grave 18 (BOL008)

Rectangular burial pit in southeast–northwest orientation up to 49 cm deep with remains of a timber construction. The burial contained poorly preserved skeletal remains of a female individual of the age of 30–39 years lying supine with her head to the east. Grave goods: two temporal rings (on both sides of the skull), metal plaques, remains of a headband, two neck rings, arm ring, pendants, clay spindle whorl, iron awl, two round brooches, glass beads, finger ring.

Grave 20 (BOL009)

Rectangular burial pit in southeast–northwest orientation up to 47 cm deep with remains of a timber construction. Skeletal remains in the burial were heavily disturbed and disarticulated. The burial contained remains of a female individual, aged 30–39, lying supine oriented to the east with her right arm on the chest. Grave goods: two temporal rings (near the skull and on the opposite side of the pit), neck ring, glass beads and metal plaques, round brooch, pendants, finger ring, iron awl, iron knife, iron needle, iron buckle, clay spindle whorl.

Shekshovo 9

Shekshovo 9 is one of the largest medieval burial sites in Suzdal region, located 21 km northwest of Suzdal and 1.5 km from Bolshoye Davydovskoye. It is the necropolis of a large unfortified settlement, one of the local centers of the 10th to early 12th centuries. The burial site was first discovered in 1852. A.S. Uvarov excavated 244 barrows with cremations and inhumations, many richly furnished. Excavations conducted in 2011–2017 revealed an area of 2,500 square meters with 26 inhumation graves, remains of at least 20 cremations, 14 barrows and over 2,500 medieval artefacts. The Shekshovo 9 burial site displays a wide variety of burial rituals. Cremations constitute the earliest part of the cemetery, and the transition to inhumations occurred during the first half of the 11th century. Thus, Shekshovo 9 represents the transition from the multi-component culture of the 10th–early 11th century to the Rus' culture of the 11th–early 12th century. Elements associated with the Volga Finnic people formed an important component of 10th-century culture at Shekshovo 9, while the burials of the 11th century follow common traditions of Rus' funeral rite. Sets of grave goods, including Oriental, Byzantine and Western European coins, and fragments of prestigious metal items testify to the wealth and high social standing of many Shekshovo settlers. The last burials at Shekshovo 9 date to the second half of the 12th century.[15,51,52]

Grave 2 (SHE001)

Discovered in 2013. A flat burial in a large (300ч130 cm) rectangular pit contained skeletal remains of a 40–49 years old male lying supine on his back, arms along the body, his head oriented west. Grave goods: finger ring (on a finger of the left hand), iron buckle (on the waist), two arrowheads, iron knife in scabbard, remains of a leather purse (?), iron key, firesteel, flint, silver ring (near the left hand), handmade clay vessel (near the left foot). Dating: the 11th century.

Grave 5 (SHE002)

Discovered in 2013. A flat burial in a rectangular, 27 cm deep pit with rounded corners contained skeletal remains of a female individual lying supine, right arm straight, left arm on the waist. Grave goods: 14 temporal rings, 55 glass beads (near the neck), finger-ring, two iron knives. The set of dress ornaments is typical for the Rus' ("Slavic") costume. Dating: the second half of the 11th century.

Grave 7 (SHE003)

Discovered in 2014. A flat burial in a large (300 ч 100 cm), 15–20 cm deep oval pit. The burial pit cut through an older cremation: the soil in the pit contained cremated bones, fragments of ceramics and melted fragments of metal ornaments. SHE003 was a 40-49 year

old male with the preserved skeletal remains lying supine with his head to the west and arms along the body. Grave goods: bronze belt ring (on the waist), iron firesteel, three flints, iron knife, iron arrowhead (near the right arm and waist), two handmade clay vessels (near the right foot). Dating: the 11th century.

Grave 12 (SHE004)

Discovered in 2015. A flat burial in a large (340 x 110 cm), 30 cm deep rectangular pit contained poorly preserved skeletal remains of a 30–39 years old female. The deceased lay in supine position with the head to the west and arms along the body. Grave goods: four temporal rings (near the skull), 67 glass beads (on the neck), coin pendant (England, 991–997), finger ring, iron knife (near the right hand), two handmade clay vessels (in the eastern part of the pit). The set of dress ornaments is typical for the Rus' ("Slavic") costume. Dating: the first half of the 11th century.

Grave 5 (SHE005)

Discovered in 2017. A flat burial in a rectangular, 20 cm deep pit contained poorly preserved remains of a 20–25 year old male lying supine with his head to the west and arms along the body. Grave goods: half of a silver Cufic coin (994/995) (near the waist), iron knife (near the right hand), handmade clay vessel (near the right foot). Dating: the first half of the 11th century.

Grave 6 (SHE006)

Discovered in 2017. A flat burial in a large (260 x 120 cm) rectangular pit 20 cm deep contained remains of an approximately 20-year-old female lying supine with her head to the west, arms beside the body, and with an exceptionally richly furnished grave. Grave goods: eight temporal rings of wire (near the skull and on the chest), iron neck ring, 102 glass and stone beads, two Cufic coin pendants (Bulgarian imitations of the second half of the 10th century), Byzantine silver coin pendant (945–959) (on the neck and chest), triangle-shaped pendant with jingling pendants (to the right of the skull), four bronze bells (to the right of the skull), three horseshoe-shaped brooches (on the chest), two arm rings on both hands, five finger rings and one glass finger ring (on the fingers of both hands, near the right arm, near the waist and near the feet), two pendants with jingling pendants (near both feet, probably shoe decoration), iron knife (near the right arm), handmade clay vessel, iron awl (near the feet). The dress assemblage includes ornaments of both "Finnic" and "Slavic" style, including jingling pendants and metal shoe ornaments typical of the culture of the Volga Finns. Dating: the set of artefacts dates to the beginning of the 11th century.

Grave 8/1 (SHE007)

Discovered in 2017. A flat inhumation grave in a rectangular pit 30 cm deep, oriented south–north, contained skeletal remains of a 35-45 year old male lying in supine position with his head to the south and his hands on the stomach. Two iron nails in the opposite corners of the pit probably belonged to a coffin. Grave foods: firesteel, flint, iron ring, awl, two iron needles (all to the left of the skull). Dating: the 12th century.

Grave 8/2 (SHE008)

Discovered in 2017. Poorly preserved separate human skull in the northwestern corner of the burial pit of inhumation grave 8/1. The skull of a 20–35 year old adult male(?) individual originates from a disturbed burial. Dating: the 11th–12th century.

Grave 10 (SHE009)

Discovered in 2018 in an excavation trench in the northern periphery of the burial ground. A flat inhumation grave in a rectangular burial pit, 30 cm deep, contained remains of a wooden coffin and skeletal remains of a 40–49 year old male. The deceased was lying supine with his head to the west (with a deviation to the south) and arms bent, right hand on the chest near shoulder, left hand on the chest.

Shekshovo 2

Shekshovo 2 is a large unfortified settlement where the cultural layer of the 10th–13th centuries covers an area of 30 ha. The burials at Shekshovo 2 are located 150 meters from the Shekshovo 9 burial site. Flat inhumation burials were discovered in 2007 and 2011 in the northern outskirts of the settlement site. The burial pits cut cultural deposits from the late 10thto the first half of the 12th century. It is unclear whether these burials are part of a larger cemetery or a smaller cluster of burials. The burial ritual of the graves follows the common Rus' tradition of the 12th–13th centuries, including the adoption of Christian practices.

Grave 1 (SHK001)

Discovered in 2007. An inhumation in a large (280ч130 cm), 30 cm deep oval pit. Skeletal remains of a 25–35-year-old male were lying in a supine position with his head to the southwest and arms flexed, left hand on the stomach, right hand near the waist. Grave goods: two bronze buttons near the collarbone.

Grave 2 (SHK002)

Discovered in 2011. An inhumation in a large (270ч110 cm), 30 cm deep rectangular pit. Skeletal remains of a 25–35 year old male were lying in a supine position with his head to the west (with a deviation to the south) and arms bent, left hand on the stomach, right hand near the waist.

Gorokhovets, Puzhalova gora

Puzhalova gora is a burial ground with 105 barrows on the southern outskirts of the Gorokhovets town on a high terrace of the Klyazma River. Excavations in 1956, 1976, 2015 and 2020 revealed 30 burials under the barrows and flat inhumations not marked with the burial mounds. The burial pits contained skeletal remains lying supine and their heads to the west, most of them unfurnished. Grave goods discovered in several graves in 2020 included items used in the second half of the 12th to the first half of the 13th century. The burial ritual represents the common Rus' tradition of the 12th century. The barrow cemetery at Puzhalova gora is known as one of

the easternmost sites with burial mounds and marks the eastern edge of the territory where the barrow ritual of medieval Rus' spread in the 11th–12th centuries.

Grave 1 (GOR001)

Discovered in 2015 under a mound (no. 80, 50 cm high, 5 m in diameter). A rectangular burial pit 50 cm deep contained skeletal remains of an approximately 50-year-old male individual. The deceased was lying supine with his head to the west and arms bent, hands on the chest. The burial ritual is similar to other graves of Puzhalova gora, including those furnished with ornaments from the second half of the 12th– the first half of the 13th centuries.

Gorokhovets, Sretensky monastery

Cemetery of a medieval town in Gorokhovets, located on the territory of the Sretensky monastery, under the building of cells constructed in the end of the 17th century. The cemetery represents an ordinary necropolis of a small medieval town and had no connection with the monastery founded in 1658. Inhumation burials were discovered in the rescue excavations in 2018. The burial pits cut settlement deposits of the second half of the 12th– the early 13th centuries and were covered by cultural layers of the 16th–17th centuries. The burial ritual with west-oriented, unfurnished inhumations in pits follows the common medieval tradition of the Rus'. The dating of the graves to the 13th–15th centuries is defined by their stratigraphic position.

Grave 1 (GOS001)

A rectangular burial pit contained well-preserved skeletal remains of a male(?) individual of about 50 years old, lying supine, the head to the southwest, arms bent, left hand on the stomach, right hand on the waist.

Grave 2 (GOS002)

A rectangular burial pit (230x70 cm) contained remains of a coffin and skeletal remains of a male(?) individual over 40 years of age lying supine, head to the southwest, arms slightly bent, hands near the waist.

Grave 4 (GOS003)

A rectangular burial pit contained poorly preserved skeletal remains of a male individual of about 30 years old lying supine, head to the southwest, right arm along the body.

Krasnoe 3

Krasnoe 3 is a medieval settlement and early modern period burial site on the right bank of the Nerl River, 4 km east of Suzdal, in the Krasnoe village. The burial ground was discovered in 2006. The excavation trench uncovered an area of 7.5 square meters with eight inhumation graves in rectangular pits at the depth of 50–70 cm from the modern surface. The inhumations had their heads oriented to the southeast, were lying in supine position and had their arms bend, with hands on the chests and stomachs. Mixed soil from the pits contained iron nails from coffins and a fragment of a bronze cross pendant common in the 16th–17th centuries. The burial ritual with narrow grave pits, a dense arrangement of the graves and the position of the deceased with bent arms indicate an early modern period dating.

Grave 3 (KRS001)

A rectangular burial pit (180x70 cm) 30 cm deep contained well-preserved skeletal remains of a 25–45-year-old female individual with thoracic and spinal pathologies. The deceased lay supine with her head to the southwest, her arms bent and hands on the chest.

Kibol 3

A medieval dwelling site and an early modern period cemetery in the Kibol village, 3 km northwest of Suzdal. Kibol 3 is a settlement site with cultural layers of the 10th–15th centuries covering an area of 11 ha and with continuous settlement until the present day. The medieval settlement was discovered in 2002 and excavations conducted in 2005–2017 uncovered an area of about 1,300 square meters with medieval deposits and structures. The early modern period cemetery was discovered in 2013 during rescue excavations in the northeastern part of the settlement site. The excavations revealed 32 inhumations with the heads oriented to the west. Several graves contained metal crosses and remains of leather shoes, dating the burials to the 18th–19th century.

Grave 1 (KBL001)

A rectangular burial pit (220x90 cm), 90 cm deep, oriented east–west, contained skeletal remains of a 20–25-year-old female in a wooden coffin fastened with an iron brace(?). The deceased was lying in supine position with the head to the west, arms bent and hands on the chest and on the stomach.

Grave 4-1 (KBL002)

A rectangular burial pit, 110 cm deep, cut a larger burial pit of Grave 4-2. It contained skeletal remains of a 1.5–2-year-old child, lying supine, head to the west (with a deviation to the north), arms bent and hands on the chest.

Grave 4-2 (KBL003)

A rectangular burial pit 120 cm deep with rounded corners (200x90 cm) contained skeletal remains of a 20–29-year-old male. The deceased lay supine with his head to the west (with a deviation to the north), limbs flexed, hands on the chest and stomach and right leg slightly bent.

Kideksha

The medieval settlement site Kideksha is located on the territory of the village of the same name on the right bank of the Nerl River, 4 km east of Suzdal. Kideksha is known as a large settlement with medieval occupational layers covering an area of 16 ha. It is famous

for its limestone church dedicated to Boris and Gleb, one of the earliest stone buildings in northeastern Rus', constructed in 1152. The first excavations in Kideksha were conducted in 1851 on the territory of the churchyard. Subsequent excavations and surveys revealed a settlement since the early 11th century, with expansion in the 12th–13th centuries and continuing up to modern times. Excavations conducted in 2011–2012 revealed 49 burials from the 18th – the first half of the 19th century, cutting through the settlement deposits of the 16th–17th centuries in the churchyard. Samples for genomic analysis were selected from a trench that contained disarticulated remains, including well-preserved skulls marked as Grave A (KED001), Grave B (KED002), Grave C (KED003) and Grave D (KED004). The general archaeological context indicates that the remains date to the early modern period.

METHOD DETAILS

Radiocarbon dating and calibration

The samples were radiocarbon dated by the Klaus-Tschira Laboratory for Physical Age Determination in Mannheim, Germany (MAMS in Data S1A) and the Laboratory of Chronology, Finnish Museum of Natural History LUOMUS in Helsinki, Finland (Hela in Data S1A). We calibrated the results according to the IntCal20 atmospheric curve using Oxcal 4.4.4.[53,54] In addition, we modelled start and end boundaries for Bolshoye Davydovskoye 2 and Shekshovo 9.

Stable isotope analysis

Stable isotope analyses were carried out in the Center for Collective Use at the Institute of Ecology and Evolution of the Russian Academy of Sciences. Approximately 500 mg of compact bone fragments were used to extract collagen. The samples were first washed in bidistilled water and brought to constant weight during 48 hours. Once dry, the fragments were demineralized in 1M hydrochloric acid in a ratio of at least 1 g per 50 ml at +3 °C. Then, the samples were repeatedly washed with bidistilled water until they reached pH = 7. The samples were placed in 0.1M alkaline solution for 24 hours, after which they were washed again till pH = 7. Then, the samples were placed in a solution of hydrochloric acid with pH = 2.5 (10 ml) and transferred to a thermostat where collagen dissolves at a temperature of +65 °C. The liquid collagen was purified via centrifugation. The solution was evaporated in a thermostat at +40 °C. The resulting sample weight for the analysis was approximately 400 μg. Collagen preservation was assessed based on the C/N ratio (2.9–3.6)[55] and the percent of carbon (from 30 to 47%) and nitrogen (from 11.0 to 17.3%) in the samples.[56,57] Finally, the isotopic composition of the samples was measured using a mass spectrometer (Thermo Finnigan Delta V Plus).

Ancient DNA sample processing and quality control

Sampling and DNA extraction

Specimens were sampled in a dedicated ancient DNA laboratory in the Max Planck Institute for the Science of Human history, Jena, Germany. All samples were photographed and documented before sampling. We used minimally invasive sampling methods for petrous bone and teeth.[58,59] Approximately 50 μg of bone powder was drilled from the dental pulp chamber inside the teeth, or the inner ear canal part of the petrous bone. DNA was extracted from the bone powder with extraction buffer containing 0.45 M EDTA and 0.25 mg/ml Proteinase K using modified protocol from Dabney et al.[60,61]

Library preparation, target enrichment and sequencing

Extracted DNA from samples from Bolshoye Davydovskoye 2, Shekshovo 2, Krasnoe 3, Kideksha and Kibol 3, and samples SHE001-004 from Shekshovo 9 were converted into double-stranded DNA libraries barcoded with unique, standard indexes for Illumina platforms, attached to each side of DNA fragments. The libraries went through a partial USER enzyme treatment, to remove most of the DNA damage from the sequenced reads,[62–66] but conserve enough of it to authenticate the data as ancient. Samples from Gorokhovets and samples SHE005-009 from Shekshovo 9 were converted into single-stranded libraries without the USER-treatment.[67] We sequenced all libraries to 5 million reads with Illumina Hiseq 4000 platform to assess ancient DNA preservation. Out of 32 samples, 31 had sufficient endogenous DNA preservation (>0.01%) and were processed further with whole-genome capture enrichment.

Selected libraries were enriched for 1,237,207 ancestry-informative single-nucleotide polymorphisms using an in-solution capture protocol.[68,69] Captured libraries were sequenced to 20 million reads with Illumina HiSeq 4000 using 50 paired-end cycles or 75 single-end cycles.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data processing and quality control

Sequenced reads from enriched libraries were processed with nf-core/eager pipeline v. 2.3.2.[36] Adapters were clipped using AdapterRemoval v2.3.1,[29] with option –preserve5p to keep 5-prime ends intact, and without length filtering. For paired-end sequences, reads were merged by the sequence overlap and unmerged reads were discarded. Reads were mapped to the human reference genome hs37d5 using bwa aln v0.7.17-r1188[33] with seeding disabled (-l 16500) and –n 0.01. Duplicates were removed using Picard MarkDuplicates v2.22.9 (http://broadinstitute.github.io/picard/). We used DamageProfiler v0.4.9[34] to calculate the cytosine deamination frequencies on the terminal positions of the fragments. To avoid errors arising from spurious mapping, mapping reads shorter than 30 bp were excluded from downstream processing. Aligned sequences from UDG-half treated libraries were trimmed from terminal positions by masking two terminal bases from both ends of each sequence using bamUtil v1.0.14.[31]

Ancient DNA authentication

To estimate ancient DNA authenticity, we first investigated the deamination patterns in the terminal positions of the fragments. Most samples showed a distinct pattern of high C-to-T transition in the terminal positions. The signal was faint in some UDG-half libraries, but because they came from recent historical time periods and thus were expected to have accumulated only relatively low amount of DNA damage, we did not exclude them from further analyses.

Next, we used Schmutzi[42] to obtain estimates of mitochondrial contamination and ANGSD method to estimate autosomal contamination in male samples.[30,70] With these two analyses, all samples had point estimates of contamination below 4%. We caution that mitochondrial estimates are only reliable for samples which have mt/nuc ratio $\leq$ 200.[71] Because most of the sampled individuals were females, we also used the program ContamLD[45] to obtain additional autosomal contamination estimates. ContamLD utilizes the fact that contamination breaks down the expected patterns of linkage disequilibrium in the endogenous genome. First, we used outgroup f_3 statistics to determine which present-day population in the 1000 Genomes panel[72] was closest to each ancient individual.[72] We then used PMDtools[37] to filter out reads that did not show evidence of *post mortem* damage (using PMD score threshold 0). To run ContamLD, we piled up allele count files for 1240K SNPs from both PMD-filtered and non-filtered BAMs for each individual using a custom Python script. ContamLD flagged one sample, KBL003, as contaminated and thus we decided to not include it in our analysis clusters.

Biological sex determination and uniparental markers

We used SexDetERRmine[41] and 1240K panel positions to infer genetic sex of the individuals. SexDetERRmine software measures the coverage of both sex chromosomes relative to autosomes. All individuals were assigned as either XX or XY (Data S1A). We note that all individuals from Bolshoye Davydovskoye 2 were classified as XX, contradicting some of the previous archaeological sex assignment.

Mitochondrial haplotypes were called from consensus sequences produced with the Schmutzi pipeline.[42] The Schmutzi log file of each sample was converted to a fasta file using base call quality threshold 50. Haplotypes were then called with HaploGrep2.[43] Y-chromosomal haplogroups were called using Yhaplo[44] (https://github.com/23andMe/yhaplo) and lineage-informative SNPs from the International Society of Genetic Genealogy 2016 tree (https://isogg.org/tree/2016/index16.html).

Runs of homozygosity and biological relatedness

We used hapROH[49] to detect signals of recent inbreeding or background relatedness (indicative of small effective population size). The program was run with default parameters for all individuals. We did not detect any long ROHs, which would have implied inbreeding between close relatives.

To estimate biological relatedness, we used READ.[50] First, we ran the analysis across all individuals and sites to detect potential cross-site relatives (and to spot potential sample mix-ups). Next, we repeated the analysis for subsets of individuals, including only those from the same site or time interval. Monomorphic sites were removed before running the analysis for any group. We detected two pairs of related individuals in Bolshoye Davydovskoye 2: BOL002 and BOL004 (first degree) and BOL003 and BOL004 (second degree). BOL002 and BOL004 also shared a maternal haplogroup. The third kin pair was detected in Shekshovo 9, between SHE002 and SHE006 (second degree). Unfortunately, the low coverage of SHE002 did not allow us to infer the mitochondrial haplogroup for this individual and thus we were unable to conclude whether these two individuals shared maternal lineage or not.

Genotyping and merging with reference data

Pseudohaploid genotypes were called using samtools[40] mpileup and pileupCaller (https://github.com/stschiff/sequenceTools), applying mapping-quality and base-quality thresholds of 30. Trimmed bams were used as a genotyping source for double-stranded UDG-half libraries. For the single-stranded non-UDG libraries, pileupcaller was run in the SingleStrandMode, which ignores C/T polymorphisms in the reads aligned to forward strand.

Genotypes were merged with published modern and ancient individuals using Poseidon package tool trident (https://github.com/poseidon-framework). We retrieved all individuals from public Poseidon repository on August 6, 2021, and merged them with access-restricted populations from Human Origins panel[73,74] and our newly genotyped individuals (see Data S1B). We generated two datasets: one that contained all SNPs available for each individual ("1240K") and one that contained only the ~600,000 SNPs covered in Human Origins panel ("HO").

We coanalysed our data with relevant previously published ancient individuals. Most notable samples came from Bronze Age Bolshoy Oleniy Ostrov in Kola Peninsula, Russia (Russia_Kola_BA)[41]; these individuals are the oldest known carriers of Siberian ancestry west of the Urals. Other points of comparisons for our data were the Bronze Age Fatyanovo individuals found from the same geographic region as our samples (Russia_Fatyanovo_BA).[28] Temporally proximate individuals – genetically close to present-day Saami – were available from the Levänluhta water burial in Isokyrö, Finland (Finland_Levanluhta_IA).[41] We also used Bronze Age and Iron Age samples from Estonia (Estonia_BA and Estonia_IA) and Ingria, Russia (Russia_Ingria_IA)[75]; Estonian and Ingrian Iron Age individuals carry moderate amounts of Siberian ancestry and are linked to *tarand* burial phenomenon.

To achieve higher geographical resolution in our analyses, we also split present-day Russians into eight groups based on their geographical origin: Russian_Archangelsk_Krasnoborsky, Russian_Archangelsk_Leshukonsky, Russian_Archangelsk_Pinezhsky, Russian_Vologda, Russian_Central (Tver and Yaroslavl regions), Russian_Pskov, Russian_Ryazan and Russian_South (Orel, Smolensk, Kaluga, Belgorod and Kursk regions).

Principal component analysis

We used smartpca from the Eigensoft package[47] to calculate principal components of genetic variation for a selected set of (a) Eurasian (Figure 1C) and (b) West Eurasian (Figure S1) populations (https://github.com/DReichLab/EIG). We used the HO dataset and option "lsqproject: YES" to project ancient individuals on the PCs.

Populations used for spanning the PCA in (a): Abazin, Abkhasian, Adygei, Albanian, Altaian, Altaian_Chelkan, Ami, Armenian, Armenian_Hemsheni, Atayal, Avar, Azeri, Balkar, Balochi, Bashkir, Basque, BedouinA, BedouinB, Belarusian, Brahui, Besermyan, Bulgarian, Buryat, Cambodian, Canary_Islander, Chechen, China_Lahu, Chuvash, Circassian, Croatian, Cypriot, Czech, Dai, Darginian, Daur, Dolgan, Druze, Dungan, Enets, English, Estonian, Even, Evenk_FarEast, Evenk_Transbaikal, Ezid, Finnish, French, Gagauz, Georgian, Greek, Han, Hazara, Hezhen, Hungarian, Icelandic, Ingushian, Iranian, Italian_North, Italian_South, Itelmen, Japanese, Jew_Ashkenazi, Jew_Georgian, Jew_Iranian, Jew_Iraqi, Jew_Libyan, Jew_Moroccan, Jew_Tunisian, Jew_Turkish, Jew_Yemenite, Kabardinian, Kaitag, Kalash, Kalmyk, Karachai, Karakalpak, Karelian, Kazakh, Ket, Khakass, Khakass_Kachin, Khamnegan, Kinh, Korean, Koryak, Kumyk, Kurd, Kyrgyz_Kyrgyzstan, Kyrgyz_Tajikistan, Lak, Lebanese, Lezgin, Lithuanian, Makrani, Mala, Maltese, Mansi, Miao, Mongol, Mongola, Mordovian, Nanai, Naxi, Negidal, Nganasan, Nivh, Nogai_Astrakhan, Nogai_Karachay_Cherkessia, Nogai_Stavropol, Norwegian, Orcadian, Oroqen, Ossetian, Palestinian, Pathan, Polish, Russian_Archangelsk_Krasnoborsky, Russian_Archangelsk_Leshukonsky, Russian_Archangelsk_Pinezhsky, Russian_Central, Russian_Pskov, Russian_Ryazan, Russian_South, Russian_Vologda, Saami.DG, Saami.WGA, Sardinian, Saudi, Scottish, Selkup, She, Shor_Khakassia, Shor_Mountain, Sicilian, Spanish, Spanish_North, Surui, Tabasaran, Tajik, Tatar_Kazan, Tatar_Mishar, Tatar_Siberian, Tatar_Siberian_Zabolotniye, Tatar_Tomsk.DG, Tatar_Volga.DG, Thai, Todzin, Tofalar, Tu, Tubalar, Tujia, Turkish, Turkish_Balikesir, Turkmen, Udmurt, Ukrainian, Ulchi, Uyghur, Uzbek, Veps, Xibo, Yakut, Yi, Yukagir.

Populations used for spanning the PCA in (b): Abkhasian, Adygei, Albanian, Armenian, Balkar, Basque, BedouinA, BedouinB, Belarusian, Bulgarian, Canary_Islander, Chechen, Chuvash, Croatian, Cypriot, Czech, Druze, English, Estonian, Finnish, French, Georgian, Greek, Hungarian, Icelandic, Iranian, Italian_North, Italian_South, Jew_Ashkenazi, Jew_Georgian, Jew_Iranian, Jew_Iraqi, Jew_Libyan, Jew_Moroccan, Jew_Tunisian, Jew_Turkish, Jew_Yemenite, Jordanian, Kumyk, Lebanese, Lezgin, Lithuanian, Maltese, Mordovian, Norwegian, Orcadian, Palestinian, Polish, Russia_NorthOssetian.DG, Russian_South, Russian_Vologda, Sardinian, Saudi, Scottish, Sicilian, Spanish, Spanish_North, Syrian, Turkish, Ukrainian.

Admixture analysis

We ran unsupervised ADMIXTURE v.1.3.0[16] on 62 modern and 24 ancient populations and our newly sequenced individuals using the merged HO dataset. We used PLINK 1.90[46] to prune the data: we excluded variants with minor allele frequency below 0.01 and performed LD pruning using window size of 200, step size 5 and R^2 threshold of 0.5, leaving us with 234,558 variants for the analysis. We ran five replicates of each K value ranging from 2 to 16. The K value with the lowest CV error across replicates was 9.

Populations and individuals used for ADMIXTURE analysis were: Ami, Ami.DG, Armenian, Atayal, Atayal.DG, Balochi, Basque, BedouinB, Belarusian, Brahmin_Tiwari, Brahui, Chuvash, Croatian, Cypriot, Czech, English, Estonian, Even, Finnish, Finnish.DG, French, Greek, GujaratiB, Hadza, Han, Hungarian, Icelandic, Kalash, Karelian, Karitiana, Lithuanian, Makrani, Mala, Mansi, Mansi.DG, Mari.SG, Mbuti, Mbuti.DG, Mixe, Mordovian, Nganasan, Norwegian, Onge, Orcadian, Papuan, Pima, Polish, Russian_Archangelsk_Krasnoborsky, Russian_Archangelsk_Leshukonsky, Russian_Archangelsk_Pinezhsky, Russian_Central, Russian_Pskov, Russian_Ryazan, Russian_South, Russian_Vologda, Saami.DG, Sardinian, Scottish, Selkup, Sorb, Spanish, Udmurt, Ukrainian, Ukrainian_North, Ulchi, Veps, Yoruba, Ethiopia_4500BP.SG, Villabruna, Raqefet_M_Natufian, CHG, ANE, DevilsCave_N.SG, Ganj_Dareh_N, Anatolia_N, Krasnoyarsk_Krai_BA, Yamnaya_Samara, WEHG, LBK_EN, EEHG, WSHG, ESHG, Russia_Fatyanovo_BA, Estonia_BA, Estonia_IA, Russia_Ingria_IA, Russia_Kola_BA, Finland_Levanluhta_IA, Russia_EasternScythian_SouthernUrals_IA, Ukraine_Scythian_IA, Russia_EarlySarmatian_IA, Ganj_Dareh_H, Russia_Alan.SG, BOL001, BOL002, BOL003, BOL004, BOL005, BOL006, BOL007, BOL008, BOL009, GOR001, GOS001, GOS002, GOS003, SHE001, SHE002, SHE003, SHE004, SHE005, SHE006, SHE007, SHE008, SHE009, SHK001, SHK002, KBL002, KBL003, KED001, KED002, KED003, KED004, KRS001.

Analysis cluster assignment

We assigned our samples into analysis clusters for downstream allele frequency-based analyses (Data S1A). Clustering was primarily based on radiocarbon dating and archaeological context, and PCA and ADMIXTURE results were used to refine the groupings (Figures 1B, 1C, and S1–S3) We excluded individuals that were radiocarbon (SHE008, SHE009, KED004 and KRS001) and/or genetic (BOL006, SHE008, KED004, SHK001 and SHK002) outliers relative to their respective groups. In addition, we divided the medieval group into two clusters based on PCA observations. We note that the first of these two medieval groups, VolgaOka_MA1, contains an individual SHE003 that could be considered as an outlier based on PCA; however, we chose not to exclude the sample due to small number of samples in the cluster – it is possible that our sampling just covers the extremes of a genetic group. We further pruned our analysis clusters based on our kinship results. We removed BOL004 and SHE002 from our analysis clusters to obtain analysis clusters without closely related individuals. Finally, we excluded KBL003 due to possible contamination flagged by ContamLD.

The final analysis clusters included 21 individuals: Iron Age (VolgaOka_IA, n=7), medieval Iron Age-like (VolgaOka_MA1, n=4), medieval East European-like (VolgaOka_MA2, n=6), and post-medieval individuals from Kideksha and Kibol (VolgaOka_H, n=4).

F_3 and f_4 statistics

We calculated outgroup f_3 statistics using ADMIXTOOLS v.6.0[48] and the merged 1240K dataset. The level of allele sharing with 528 present-day groups was measured by calculating a set of outgroup f_3 statistics in the form of *f3(Modern_group,Analysis_cluster;Mbuti)*. Higher estimates of f_3 indicate higher level of allele sharing. We selected Lithuanians as a point of comparison for further f_4 tests, because all analysis clusters consistently had Lithuanians among their three highest f_3 estimates (Data S2A). Outgroup f_3 statistics were also calculated for the Shekshovo 2 outliers and the results indicated that these two individuals shared most alleles with Nganasans.

We calculated cladality tests in the form of *f4(Mbuti,Modern_group;Analysis_cluster,Lithuanian)* for all analysis clusters to measure their allele sharing with 400 present-day non-African groups relative to Lithuanians. For Sheksohovo 2 outliers, we used Nganasans instead of Lithuanians. If the analysis cluster was cladal with their f_3-based comparison population, we would expect to see no significantly non-zero estimates of f_4 (Z score $\geq$ |3|) when we compare them to a set of present-day populations. However, no analysis cluster was cladal with their comparison population (Data S2B). We extended the cladality tests by manually selecting populations that showed genetic similarity with our study populations on PCA and ADMIXTURE analyses. For VolgaOka_IA and VolgaOka_MA1, who fell close to the present-day Uralic-speakers on PCA, we selected Saami.DG, Karelian, Veps, Mordovian, Chuvash, Mari.SG, Russian_Archangelsk_Krasnoborsky, Russian_Archangelsk_Leshukonsky and Russian_Archangelsk_Pinezhsky. For VolgaOka_MA2 and VolgaOka_H, who fell close to present-day Slavs, we used Czech, Sorb, Polish, Belarusian, Ukrainian_North, Ukrainian, Russian_South, Russian_Pskov and Russian_Ryazan, and ran f_4 cladality tests for them as described above.

We measured the relative affinity of our analysis clusters to Siberian ancestry and East European hunter-gatherer (EEHG) ancestry by computing a set of f_4 statistics in the form of *f4(Mbuti,Krasnoyarsk_Krai_BA/EEHG;Test,Lithuanian)*. Although previous studies have shown that Nganasans are the best proxy for the Siberian ancestry in Uralic-speaking populations,[10,41] we preferred to use an ancient proxy in our analyses. We found that out of all published ancient populations, Nganasans shared most alleles with a Bronze Age individual from Kraskoyarsk Krai, Southern Siberia, in outgroup f_3 tests (data not shown). They also have identical ancestry profiles in our ADMIXTURE analysis (Figure S2). We also ran f_4 statistics on using either Nganasan or Krasnoyarsk_Krai_BA and found that they produced very similar results (data not shown).

All f_4 statistics were calculated from genotype data using Admixtools 2.0.0 package (https://github.com/uqrmaie1/admixtools/) in R 4.0.4. Mbuti was used as outgroup in all calculations.

qpAdm modelling

All qpAdm models were computed using the merged 1240K dataset and Admixtools 2.0.0 package with functions *qpadm_rotate()* and *qpadm_multi()*. Group-level models were calculated directly from genotype files using default parameters; for individual-level models, we used the option allsnps=TRUE, which uses all available SNPs to calculate each individual f_4 statistic. This option allowed us to use more data, but may lead to biased results because each f_4 statistic is calculated on a different set of SNPs.

We adopted a two-level approach for the qpAdm modelling: First, we constructed distal models for the populations of interest using five rotating sources and seven fixed right populations. Second, we designed sequential proximal models, which used individuals from a preceding time point as sources to model the target individuals in the next time point. For proximal models, we used the five sources of the distal models as right populations, adding a true outgroup (6,500 years old individual from Ethiopia[76]). In addition to our study populations, we included as targets four previously published ancient groups: Russia_Fatyanovo_BA,[28] Russia_Kola_BA,[41] Russia_Ingria_IA[75] and Finland_Levanluhta_IA.[41] Modern target populations included Lithuanians, Slavic-speaking Belarusians, Ukrainians, North Ukrainians, Polish, Sorbs, and Czech, Uralic-speaking Saami, Udmurt, Mordovian, Veps, Karelian, and Finnish, and the geographically close Chuvash, and Russians. As sources for distal models, we used Krasnoyarsk_Krai_BA[17] as a proxy for Siberian ancestry, Serbian and Romanian Iron Gates hunter-gatherers as WEHG,[18,77] hunter-gatherers from Karelia and Samara as EEHG,[68] Yamnaya from Samara as Early-Middle Bronze Age Steppe[68] and Early Neolithic LBK as early European farmer ancestry.[68,77,78] Right populations for distal models were: Ethiopia_4500BP.SG, CHG,[79] Raqefet_M_Natufian,[74] Onge, Villabruna,[80] ANE[80,81], Mixe and Sweden_HG_Motala.[68] We implemented a rotating scheme to compare alternative models: when a model was run without a given source, we moved the unused source into the right populations. For each target group, we chose the model that had the highest p-value among all feasible models. Additionally, we required all component weights to be fully within an interval [0, 1] and have z scores above one. Finally, we calculated p-values, weights and standard errors for the selected models using consistent right populations. We used p-value $\geq$0.01 as an indication for a fitting model. The results are shown in Figure 3A and Data S3.

For proximal models, we selected geographically and temporally appropriate ancient and modern populations for each of the test populations as potential proximal sources and applied the same model selection criteria as above with the exception of using a p-value $\geq$0.05 as a cut-off criterion for a fitting model. We modelled VolgaOka_IA as a mixture of up to three sources of ancestry: a Siberian-related source (Russia_Kola_BA,[41] Finland_Levanluhta_IA,[41] Mansi or Nganasan); western/local source (Russia_Fatyanovo_BA,[28] Estonia_BA,[75] or Estonia_IA[75]); and a Steppe source (Central_Steppe_MLBA,[18] Western_Steppe_MLBA,[18,68,82] Ukraine_Scythian_IA,[83] or Russia_EarlySarmatian_IA[84]). We used qpWave to ensure that our right populations had enough power to distinguish between potential sources in each source group (data not shown). All working models required contribution from all three source groups. To narrow down the model space further, we performed model competition experiment as described by Narasimhan et al.[18] In brief, we compared model pairs with one different source by moving the alternative source to the right populations. However, with this approach we could exclude only one more model (Data S3C). In Figure 3B, we show the model that had the highest p-value in the original comparison.

Next, VolgaOka_MA2, VolgaOka_MA1 and VolgaOka_H were modelled as a two-way mixture of VolgaOka_IA and Belarusian, Polish, Sorb, Ukrainian or Ukrainian_North. Finally, VolgaOka_H was also modelled as two-way mixtures of VolgaOka_MA2 and VolgaOka_MA1. We also extended the latter two models to individual level: All sequenced post-Iron Age individuals, excluding the two Shekshovo 2 outliers, regardless of their cluster membership, were modelled individually as a mixture of VolgaOka_IA and Slavic ancestry, and post-mediaval individual were additionally modelled as a mixture of VolgaOka_MA2 and VolgaOka_MA1. We found one or more working two-way model for all but two individuals – KED006, who required a third, Iranian-related source of ancestry (Russia_Alan.SG[85] or Ganj_Dareh_H[74]) and SHE006 (Data S3D).

DATES analysis

We used DATES[18] to infer admixture times from our data. To obtain more reliable results, we used present-day populations (Lithuanian and Nganasan) as proxy sources for North European and Siberian ancestry. Admixture time was inferred for each analysis cluster using these to populations as sources and by running the software on recommended parameters (see https://github.com/priyamoorjani/DATES). The targets included our four analysis clusters as well as present-day Russians. We used the results of the software to calculate admixture times in years by using a mean radiocarbon date for each cluster and assuming a generation time of 29 years (Data S3E).