The Genetic Origin of the Indo-Europeans

Iosif Lazaridis[1,2*#], Nick Patterson[1,3*#], David Anthony[1,4*#], Leonid Vyazov[5#], Romain Fournier[6], Harald Ringbauer[1,7], Iñigo Olalde[1,8,9], Alexander A. Khokhlov[10], Egor P. Kitov[11], Natalia I. Shishlina[12], Sorin C. Ailincăi[13], Danila S. Agapov[14], Sergey A. Agapov[14], Elena Batieva[15], Baitanayev Bauyrzhan[16], Zsolt Bereczki[17], Alexandra Buzhilova[18], Piya Changmai[5], Andrey A. Chizhevsky[19], Ion Ciobanu[20], Mihai Constantinescu[21], Marietta Csányi[22], János Dani[23,24], Peter K. Dashkovskiy[25], Sándor Évinger[26], Anatoly Faifert[27], Pavel N. Flegontov[1,5,28], Alin Frînculeasa[29], Mădălina N. Frînculeasa[30], Tamás Hajdu[31], Tom Higham[32,33], Paweł Jarosz[34], Pavol Jelínek[35], Valeri I. Khartanovich[36], Eduard N. Kirginekov[37], Viktória Kiss[38], Alexandera Kitova[39], Alexeiy V. Kiyashko[40], Jovan Koledin[41], Arkady Korolev[10], Pavel Kosintsev[42,43], Gabriella Kulcsár[38], Pavel Kuznetsov[10], Rabadan Magomedov[44], Mamedov Aslan Malikovich[16], Eszter Melis[38], Vyacheslav Moiseyev[36], Erika Molnár[17], Janet Monge[45], Octav Negrea[29], Nadezhda A. Nikolaeva[46], Mario Novak[47,48], Maria Ochir-Goryaeva[49], György Pálfi[17], Sergiu Popovici[50], Marina P. Rykun[51], Tatyana M. Savenkova[52], Vladimir P. Semibratov[53], Nikolai N. Seregin[54], Alena Šefčáková[55], Mussayeva Raikhan Serikovna[16], Irina Shingiray[56], Vladimir N. Shirokov[57], Angela Simalcsik[20,58], Kendra Sirak[1,2], Konstantin N. Solodovnikov[60], Judit Tárnoki[22], Alexey A. Tishkin[53], Viktor Trifonov[61], Sergey Vasilyev[62], Ali Akbari[1,2], Esther S. Brielle[1], Kim Callan[2,63], Francesca Candilio[64], Olivia Cheronet[32,33], Elizabeth Curtis[2,63], Olga Flegontova[5,28], Lora Iliev[2,63], Aisling Kearns[2], Denise Keating[65], Ann Marie Lawson[2,63], Matthew Mah[2,3,63], Adam Micco[2], Megan Michel[1,2,7,63], Jonas Oppenheimer[2,63], Lijun Qiu[2,63], J. Noah Workman[2,63], Fatma Zalzala[2,63], Anna Szécsényi-Nagy[66], Pier Francesco Palamara[6,67], Swapan Mallick[2,3,63], Nadin Rohland[2], Ron Pinhasi[32,33#], David Reich[1,2,3,63#]

* Contributed equally
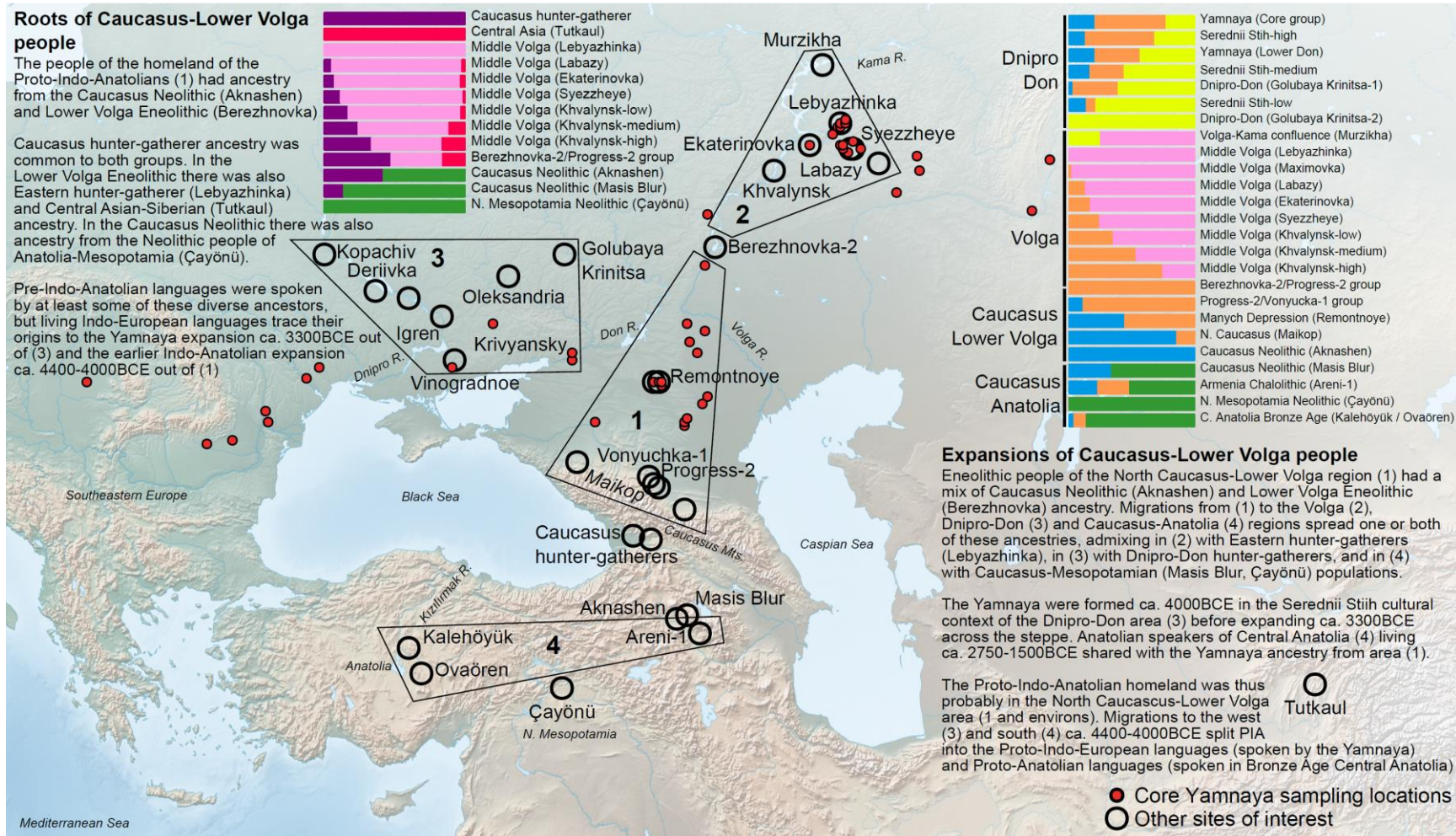
[#] To whom correspondence should be addressed:
lazaridis@fas.harvard.edu, nickp@broadinstitute.org, AnthonyD@hartwick.edu,
l.a.vyazov@gmail.com, ron.pinhasi@univie.ac.at, reich@genetics.med.harvard.edu

1   Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA
2   Department of Genetics, Harvard Medical School, Boston, MA, USA
3   Broad Institute of Harvard and MIT, Cambridge, MA, USA
4   Hartwick College, Dept. of Anthropology, USA
5   Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava, Czechia
6   Department of Statistics, University of Oxford, Oxford, UK
7   Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
8   BIOMICs Research Group, Department of Zoology and Animal Cell Biology, University of the Basque Country UPV/EHU,Vitoria-Gasteiz, Spain
9   Ikerbasque-Basque Foundation of Science, Bilbao, Spain
10   Samara State University of Social Sciences and Education, Samara, Russia

11   Center of Human Ecology, Institute of Ethnology and Anthropology, Russian Academy of Science, Moscow, Russia

12   State History Museum, Department of Archaeology, Moscow, Russia

13   "Gavrilă Simion" Eco-Museum Research Institute, Tulcea, Romania

14   Samara Regional Public Organization "Historical, ecological and cultural Association "Povolzje"

15   Azov History, Archaeology and Palaeontology Museum-Reserve, Azov, Russia

16   Institute of Archaeology named after A.Kh Margulan, Almaty, Kazakhstan

17   Department of Biological Anthropology, Institute of Biology, University of Szeged, Szeged, Hungary

18   Research Institute and Museum of Anthropology, Moscow, Russia

19   Institute of Archeology named after A. Kh. Khalikov Tatarstan Academy of Sciences, Kazan, Russia

20   Orheiul Vechi Cultural-Natural Reserve, Institute of Bioarchaeological and Ethnocultural Research, Chișinău, Republic of Moldova

21   Fr. I Rainer Institute of Anthropology, University of Bucharest, Bucharest, Romania

22   Damjanich János Museum, Szolnok, Hungary

23   Department of Archaeology, University of Szeged, Szeged, Hungary

24   Déri Museum, 4026 Debrecen, Hungary

25   Department of Regional Studies of Russia, National and State-Confessional Relations, Altai State University, Barnaul, Russia

26   Hungarian Natural History Museum, Department of Anthropology, Budapest, Hungary

27   Research Institute GAUK RO "Don Heritage", Rostov-on-Don, Russia

28   Institute of Parasitology, Biology Centre of the Czech Academy of Sciences, České Budějovice, Czechia

29   Prahova County Museum of History and Archaeology, Ploiești, Romania

30   Department of Geography, Faculty of Humanities, University Valahia of Târgoviște, Târgoviște, Romania

31   Eötvös Loránd University (Department of Biological Anthropology, Institute of Biology, Eötvös Loránd University, Budapest, Hungary

32   Department of Evolutionary Anthropology, University of Vienna, Vienna, Austria

33   Human Evolution and Archaeological Sciences, University of Vienna, Vienna, Austria

34   Department of Mountain and Highland Archaeology, Institute Archaeology and Ethnology Polish Academy of Science, Kraków, Poland

35   Slovak National Museum-Archaeological Museum, Bratislava, Slovak Republic

36   Peter the Great Museum of Anthropology and Ethnography, Department of Physical Anthropology, St. Petersburg, Russia

37   State Autonomous Cultural Institution of the Republic of Khakassia "Khakassian National Museum of Local Lore named after L.R. Kyzlasova", Republic of Khakassia, Abakan, Russia

38   Institute of Archaeology, HUN-REN Research Centre for the Humanities, Budapest, Hungary

39   Centre for Egyptological Studies of the Russian Academy of Sciences, Russian Academy of Sciences, Moscow, Russia

40 Department of Archaeology and History of the Ancient World of the Southern Federal University, Rostov-on-Don, Russia

41 Museum of Vojvodina, Novi Sad, Serbia

42 Department of History of the Institute of Humanities, Ural Federal University, Ekaterinburg, Russia

43 Institute of Plant and Animal Ecology, Urals Branch of the Russian Academy of Sciences, Yekaterinburg, Russia

44 Institute of History, Archaeology and Ethnography, Dagestan branch of the Russian Academy of Science, Makhachkala. Dagestan, Russia

45 Independent Researcher, 106 Federal Street, Philadelphia PA, USA

46 Department of General History, Historical and Literary Institute of the State University of Education, Ministry of Education Moscow, Moscow, Russia

47 Centre for Applied Bioanthropology, Institute for Anthropological Research, Zagreb, Croatia

48 Department of Archaeology and Heritage, Faculty of Humanities, University of Primorska, Koper, Slovenia

49 Kalmyk Scientific Centre of the Russian Academy of Sciences, Elista, Republic of Kalmykia, Russia

50 National Agency for Archaeology, Chişinău , Republic of Moldova

51 National Research Tomsk State University, Tomsk, Russia

52 V.F. Voino-Yasenetsky Krasnoyarsk State Medical University, Krasnoyarsk, Russia

53 Department of Archaeology, Ethnography and Museology, Altai State University, Barnaul, Russia

54 Laboratory of Ancient and Medieval Archaeology of Eurasia, Altai State University, Barnaul, Russia

55 Slovak National Museum-Natural History Museum, Bratislava, Slovak Republic

56 University of Oxford, Faculty of History, Oxford, United Kingdom

57 Center for Stone Age Archeology, Institute of History and Archaeology, Ural Branch of the Russian Academy of Sciences, Ekaterinburg, Russia

58 Olga Necrasov Centre for Anthropological Research, Romanian Academy, Iași Branch, Iași, Romania

59 Institute of Bioarchaeological and Ethnocultural Research, Chişinău, Republic of Moldova

60 Tyumen Scientific Center of the Siberian Branch of Russian Academy of Sciences, Institute of Problems of Northern Development, Tyumen, Russia

61 Institute for the History of Material Culture, Russian Academy of Sciences, St Petersburg, Russia

62 Russian Academy of Sciences, Institute of Ethnology and Anthropology, Moscow, Russia

63 Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA

64 Museo delle Civiltà, Italian Ministry of Culture, Rome, Italy

65 School of Archaeology, University College Dublin, Ireland

66 Institute of Archaeogenomics, HUN-REN Research Centre for the Humanities, Budapest, Hungary

67 Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK

The Yamnaya archaeological complex appeared around 3300BCE across the steppes north of the Black and Caspian Seas, and by 3000BCE reached its maximal extent from Hungary in the west to Kazakhstan in the east. To localize the ancestral and geographical origins of the Yamnaya among the diverse Eneolithic people that preceded them, we studied ancient DNA data from 428 individuals of which 299 are reported for the first time, demonstrating three previously unknown Eneolithic genetic clines. First, a "Caucasus-Lower Volga" (CLV) Cline suffused with Caucasus hunter-gatherer (CHG) ancestry extended between a Caucasus Neolithic southern end in Neolithic Armenia, and a steppe northern end in Berezhnovka in the Lower Volga. Bidirectional gene flow across the CLV cline created admixed intermediate populations in both the north Caucasus, such as the Maikop people, and on the steppe, such as those at the site of Remontnoye north of the Manych depression. CLV people also helped form two major riverine clines by admixing with distinct groups of European hunter-gatherers. A "Volga Cline" was formed as Lower Volga people mixed with upriver populations that had more Eastern hunter-gatherer (EHG) ancestry, creating genetically hyper-variable populations as at Khvalynsk in the Middle Volga. A "Dnipro Cline" was formed as CLV people bearing both Caucasus Neolithic and Lower Volga ancestry moved west and acquired Ukraine Neolithic hunter-gatherer (UNHG) ancestry to establish the population of the Serednii Stih culture from which the direct ancestors of the Yamnaya themselves were formed around 4000BCE. This population grew rapidly after 3750-3350BCE, precipitating the expansion of people of the Yamnaya culture who totally displaced previous groups on the Volga and further east, while admixing with more sedentary groups in the west. CLV cline people with Lower Volga ancestry contributed four fifths of the ancestry of the Yamnaya, but also, entering Anatolia from the east, contributed at least a tenth of the ancestry of Bronze Age Central Anatolians, where the Hittite language, related to the Indo-European languages spread by the Yamnaya, was spoken. We thus propose that the final unity of the speakers of the "Proto-Indo-Anatolian" ancestral language of both Anatolian and Indo-European languages can be traced to CLV cline people sometime between 4400-4000 BCE.

Roots of Caucasus-Lower Volga people

The people of the homeland of the Proto-Indo-Anatolians (1) had ancestry from the Caucasus Neolithic (Aknashen) and Lower Volga Eneolithic (Berezhnovka)

Caucasus hunter-gatherer ancestry was common to both groups. In the Lower Volga Eneolithic there was also Eastern hunter-gatherer (Lebyazhinka) and Central Asian-Siberian (Tutkaul) ancestry. In the Caucasus Neolithic there was also ancestry from the Neolithic people of Anatolia-Mesopotamia (Çayönü).

Pre-Indo-Anatolian languages were spoken by at least some of these diverse ancestors, but living Indo-European languages trace their origins to the Yamnaya expansion ca. 3300BCE out of (3) and the earlier Indo-Anatolian expansion ca. 4400-4000BCE out of (1)

Left barplot legend:
Caucasus hunter-gatherer
Central Asia (Tutkaul)
Middle Volga (Lebyazhinka)
Middle Volga (Labazy)
Middle Volga (Ekaterinovka)
Middle Volga (Syezzheye)
Middle Volga (Khvalynsk-low)
Middle Volga (Khvalynsk-medium)
Middle Volga (Khvalynsk-high)
Berezhnovka-2/Progress-2 group
Caucasus Neolithic (Aknashen)
Caucasus Neolithic (Masis Blur)
N. Mesopotamia Neolithic (Çayönü)

Map labels:
Murzikha, Kama R., Lebyazhinka, Ekaterinovka, Syezzheye, Labazy, Khvalynsk, Berezhnovka-2, Kopachiv, Deriivka, Golubaya Krinitsa, Oleksandria, Igren, Krivyansky, Don R., Volga R., Vinogradnoe, Remontnoye, Vonyuchka-1, Progress-2, Maikop, Caucasus hunter-gatherers, Caucasus Mts, Aknashen, Masis Blur, Kalehöyük, Areni-1, Ovaören, Anatolia, Çayönü, N. Mesopotamia, Southeastern Europe, Black Sea, Caspian Sea, Mediterranean Sea, Dnipro R., Kızılırmak R., Tutkaul

Right barplot legend:
Yamnaya (Core group)
Serednii Stih-high
Yamnaya (Lower Don)
Serednii Stih-medium
Dnipro-Don (Golubaya Krinitsa-1)
Serednii Stih-low
Dnipro-Don (Golubaya Krinitsa-2)
Volga-Kama confluence (Murzikha)
Middle Volga (Lebyazhinka)
Middle Volga (Maximovka)
Middle Volga (Labazy)
Middle Volga (Ekaterinovka)
Middle Volga (Syezzheye)
Middle Volga (Khvalynsk-low)
Middle Volga (Khvalynsk-medium)
Middle Volga (Khvalynsk-high)
Berezhnovka-2/Progress-2 group
Progress-2/Vonyucka-1 group
Manych Depression (Remontnoye)
N. Caucasus (Maikop)
Caucasus Neolithic (Aknashen)
Caucasus Neolithic (Masis Blur)
Armenia Chalcolithic (Areni-1)
N. Mesopotamia Neolithic (Çayönü)
C. Anatolia Bronze Age (Kalehöyük / Ovaören)

Right barplot group labels: Dnipro Don, Volga, Caucasus Lower Volga, Caucasus Anatolia

Expansions of Caucasus-Lower Volga people

Eneolithic people of the North Caucasus-Lower Volga region (1) had a mix of Caucasus Neolithic (Aknashen) and Lower Volga Eneolithic (Berezhnovka) ancestry. Migrations from (1) to the Volga (2), Dnipro-Don (3) and Caucasus-Anatolia (4) regions spread one or both of these ancestries, admixing in (2) with Eastern hunter-gatherers (Lebyazhinka), in (3) with Dnipro-Don hunter-gatherers, and in (4) with Caucasus-Mesopotamian (Masis Blur, Çayönü) populations.

The Yamnaya were formed ca. 4000BCE in the Serednii Stiih cultural context of the Dnipro-Don area (3) before expanding ca. 3300BCE across the steppe. Anatolian speakers of Central Anatolia (4) living ca. 2750-1500BCE shared with the Yamnaya ancestry from area (1).

The Proto-Indo-Anatolian homeland was thus probably in the North Caucascus-Lower Volga area (1 and environs). Migrations to the west (3) and south (4) ca. 4400-4000BCE split PIA into the Proto-Indo-European languages (spoken by the Yamnaya) and Proto-Anatolian languages (spoken in Bronze Age Central Anatolia)

● Core Yamnaya sampling locations
○ Other sites of interest

Summary Figure: The origin of Indo-Anatolian and Indo-European languages. Genetic reconstruction of the ancestry of Pontic-Caspian steppe and West Asian populations points to the North Caucasus-Lower Volga area as the homeland of Indo-Anatolian languages and to the Serednii Stih archaeological culture of the Dnipro-Don area as the homeland of Indo-European languages. The Caucasus-Lower Volga people had diverse distal roots, estimated using the *qpAdm* software on the left barplot, as Caucasus hunter-gatherer (purple), Central Asian (red), Eastern hunter-gatherer (pink), and West Asian Neolithic (green). Caucasus-Lower Volga expansions, estimated using *qpAdm* on the right barplot as disseminated Caucasus Neolithic (blue)-Lower Volga Eneolithic (orange) proximal ancestries, mixing with the inhabitants of the North Pontic region (yellow), Volga region (yellow), and West Asia (green).

Introduction

Between 3300-1500 BCE, people of the Yamnaya archaeological complex and their descendants, in subsequent waves of migration, spread over large parts of Eurasia, contributing to the ancestry of people of Europe, Central and South Asia, Siberia, and the Caucasus. The spread of Indo-European language and culture[1-7] transformed all these regions. Despite the centrality of the Yamnaya expansion to the human story of Bronze Age Eurasia, their ancestral origins are poorly understood. A first challenge has been the sparse sampling of the Yamnaya themselves across their enormous geographic distribution. The remarkable long-range mobility of the Yamnaya, quickly spreading over a vast region, adds further difficulties to tracing, from radiocarbon dating, the origins of their material culture and associated genetic profile. Nor can these origins be traced to the numerous earlier Eneolithic cultures that preceded the Yamnaya, and among whom their ancestors must be sought, as these have been sampled even more poorly and unsystematically.

The first formal study of the origins of the Yamnaya identified two disparate sources of ancestry: a northern, "Eastern Hunter-Gatherer" (EHG) source from far eastern Europe, and a southern, West Asian source related to present-day Armenians.[2] The latter source was revealed, by ancient DNA, to be related to some of the region's earliest inhabitants: Paleolithic-Mesolithic "Caucasus Hunter-Gatherers" (CHG) of Georgia,[8] and Neolithic people of the Zagros[9] and South Caucasus.[6,10,11] Additional discoveries further complicated the stories of both the northern and southern ancestors of the Yamnaya. First, it was noted that both CHG and EHG were part of an interaction sphere across the boundary between West Asia and eastern Europe,[9] suggesting the existence of intermediate populations and raising the question of when and where these came together to form the Eneolithic antecessors of the Yamnaya. Second, it was recognized that the steppe itself was an admixture zone of EHG with "Western Hunter-Gatherers" (WHG[12]). Mesolithic hunter-gatherers from Ukraine were succeeded by more WHG-admixed Neolithic hunter-gatherers in the Dnipro valley,[13] representing a local reshuffling within the European portion of a ~7,000km-long trans-Eurasian cline of boreal hunter-gatherers.[14] What was the relative contribution of the EHG (who were present in the Volga River at Lebyazhinka[2] ca. 5660-5535 BCE) and these more western Ukraine Neolithic hunter-gatherers (UNHG) of the Dnipro to later populations? Third, it was discovered that the Yamnaya had not only CHG-related, but also Anatolian Neolithic ancestry, absent in the early known steppe inhabitants, and derived from European farmer neighbors west of the steppe[5]. This ancestry was later shown to be of rather Anatolian-Levantine-Mesopotamian origin, and to be mediated not from Europe but from the Caucasus neighbors south of the steppe.[6] Such ancestry must have been added following the expansion of Neolithic farmers into the Caucasus, introduced thence into the steppe as a later exogenous element, distinct from the earlier CHG-related one. Finally, it was recognized that European steppe populations were formed not only by northern-southern admixture, but included, in at least some Eneolithic and Bronze Age people of the North Caucasus, contributions related to Siberians from further east.[5] What was the extent of the spread of this eastern ancestry and did the Yamnaya themselves possess it?

Here we present a unified population genetic analysis of 372 newly reported individuals dating from 6400-2000 BCE, as well as increased quality data for 61 individuals. The present study serves as the formal technical report for 299 of the newly reported individuals and 55 of the individuals with increased quality data; more than 80% of the individuals are from Russia, but the dataset is also significant in including dozens of individuals from westward expansion of

47   Steppe cultures along the Danube (Supplementary Information, section 1, Online Table 1).
48   Technical details of the 803 ancient DNA libraries that are the basis for the newly reported data
49   (and an additional 195 libraries that failed our screening) are presented in Online Table 2, while
50   details of 198 newly generated radiocarbon dates on these individuals are presented in Online
51   Table 3. A parallel study[15] presents a combined archaeological and genetic analysis of population
52   transformations in the North Pontic Region (Ukraine and Moldova) and serves as the formal
53   report for the data from the other 73 of the newly analyzed individuals and the other 5
54   individuals with increased quality data, with both studies co-analyzing the full dataset. We
55   grouped individuals into analysis labels based on geographical and temporal information,
56   archaeological context, and genetic clustering (Online Table 4 lists all individuals used for
57   analyses, with their labels). The potential of the combined dataset for shedding light on this
58   period can be appreciated from the fact that it adds 79 analyzed Eneolithic people from the
59   steppe and its environs (from Russia or Ukraine, west of 60E longitude and south of 60N
60   latitude, between 5000-3500BCE) to 82 published[5,7,13,15-20] and a total of 286
61   Yamnaya/Afanasievo individuals compared to 75 in the literature.[2,4-6,13,21-29]
62
63   **Discovery of three pre-Bronze Age genetic clines that collapsed after Yamnaya expansion**
64
65   Principal Component Analysis (PCA) of ancient individuals from the Pontic-Caspian steppe and
66   adjacent areas of Southeastern Europe, the Caucasus and West Asia reveals that most of the
67   Eneolithic people of the steppe as well as the later Bronze Age Yamnaya fall on non-overlapping
68   genetic gradients (Figure 1). Visual impressions from a two-dimensional PCA must be evaluated
69   cautiously, as populations at intermediate PCA positions, may not, in fact, be mixtures of more
70   extreme ones, and the plot may suggest alternative ways of modeling each population of interest.
71   For example, PC1 correlates (from left to right) to the differentiation between inland West
72   Asians (Caucasus and Iran) to East Mediterranean (Anatolian-European) populations[10], but also
73   to the differentiation between Siberians and European hunter-gatherers[14]. On the other hand, PC2
74   differentiates between Neolithic and earlier populations from northern Eurasia (top; including
75   Europe and Siberia) and West Asia (bottom: Anatolia-Mesopotamia-Caucasus-Iran). The
76   Eneolithic and Bronze Age populations occupy the middle of the PCA: how did the earlier
77   groups surrounding them across these orthogonal directions combine to form them?
78
79   To answer these questions in a statistically rigorous way, we implemented a new model
80   competition tournament framework around qpWave/qpAdm methods[2,30] to fit and distinguish
81   among alternative models (Methods; Supplementary Information, section 2). Briefly
82   summarized, the idea of this methodology is that an admixture model X that includes a set of
83   sources describes the admixture history of a target population T well if it: (i) reconstructs the
84   shared genetic drift of T with both distant outgroup populations and the sources of alternative
85   competing models, but also (ii) renders these competing models infeasible by showing that they
86   cannot model this shared drift with the sources of X. In our framework, models are evaluated
87   against a conservative set of distant outgroups as an initial filtering step; if they fit poorly, they
88   are rejected; if not, they are further evaluated by comparing them against each other in
89   symmetrical fashion (all-against-all) to identify a smaller set of promising models.
90
91   With this note of caution, we observe that in the PCA in the Eneolithic-Bronze Age steppe there
92   are three clines (geographically denoted as "Volga", "Dnipro", and "Caucasus-Lower Volga"),

93   which diverge, in PCA space, from an area that includes populations enclosed by the Lower Don
94   (at the site of Krivyansky), Lower Volga (at Berezhnovka-2), and north Caucasus mountains (at
95   Progress-2, Vonyuchka-1, and Sharakhalsun[5]). From these similar beginnings the three clines
96   extend outward into distinct directions corresponding to their geographical neighbors: both
97   towards the EHG and UNHG representing the pre-Eneolithic people that lived in the Volga-Don-
98   Dnipro area of eastern Europe, and towards the CHG and Caucasus Neolithic representing the
99   pre-Eneolithic people that lived in the Caucasus and West Asia. In what follows, we introduce
100   the key populations of each of the three clines and show how these can be modeled in terms of
101   proximate sources. We also infer the ancestry origins of the people of the three clines to discover
102   what is shared among all of them and unique to each of them.

104   *Volga Cline:* The "Volga Cline" consists of sites on waterways that drain into the Caspian Sea and
105   is suggestive of a zone of ongoing human contact within its region. The Eneolithic individuals fall
106   at positions that correlate extraordinarily well to their position on the Volga River as one moves
107   downstream: the Volosovo-attributed Sakhtysh (in the Upper Volga) and Murzikha (near the
108   Kama-Volga confluence)[14] constitute the upriver portion of the cline, situated in PCA space
109   between EHG and UNHG. The Volga Cline then distinctly "bends" in PCA space and the knee of
110   the cline is occupied by EHG groups, including those sampled in the northwest of Russia in
111   Karelia[2,19] and those of the Middle Volga, suggesting that this widely dispersed set of hunter-
112   gatherers, which has also been called the Sidelkino Cluster based on its oldest representative[19,22]
113   were the major population of much of eastern Europe. Past the knee, in the downriver portion of
114   the cline, the hunter-gatherer affinity decreases starting at the Middle Volga: Labazy, Lebyazhinka,
115   Ekaterinovka, Syezzheye, then Khvalynsk (4500-4350 BCE) and Khlopkov Bugor, finally
116   reaching the Lower Volga at Berezhnovka (4450-3960 BCE) (Fig. 1a). The decrease of hunter-
117   gatherer affinity is counterbalanced by increased affinity towards populations of the Caucasus,
118   suggesting that it is generated by an unsampled CHG-related source—that existed somewhere
119   between Georgia (where the known CHG individuals were sampled[8]) and the Lower Volga—
120   interacting with the northern EHG natives. Archaeological correlates for such south-north
121   interactions do exist, and begin with the expansion of the Seroglazovo forager culture around the
122   Lower Volga estuary ~6200 BCE, with some ceramic and lithic typological parallels with
123   Caucasus cultures, and continue to the unsampled North Caucasus Neolithic cemetery dated ~4800
124   BCE near Nalchik.[31,32]

126   At the end of the cline, the four individuals from the newly reported Lower Volga site of
127   Berezhnovka-2 can be grouped with the PG2004 individual of the Progress-2[5] site in the north
128   Caucasus into a "Berezhnovka-2-Progress-2 cluster" (abbreviated to "BPgroup"). This proves
129   that the CHG-related ancestry found at Progress-2 extended well into the steppe in the Lower
130   Volga. The second individual from Progress-2 (PG2001) is grouped with another north Caucasus
131   individual from Vonyuchka-1[5] into a related "Progress-2-Vonyuchka-1 cluster" (abbreviated as
132   "PVgroup"). PVgroup and BPgroup are distinct (p=0.0006), but their genetic differentiation was
133   small in magnitude (F_{ST}=-0.002±0.002; Extended Data Table 1) suggesting movement between
134   the north Caucasus piedmont and Lower Volga sites. The two locations also shared a distinctive
135   burial pose on the back with raised knees, later typical of Yamnaya and currently dated earliest
136   in the Samara region at Lebyazhinka-5 and in a few graves at Ekaterinovka dated before 4500
137   BCE. It is clear from the PCA (Fig. 1b) that BPgroup differs from PVgroup in that the former is
138   shifted towards the Afontova Gora-3 Upper Paleolithic individual from Siberia,[33] West Siberian

139 hunter-gatherers,[4] and Central Asians such as a 7,500-year old Neolithic individual from Tutkaul
140 (TTK) in Tajikistan.[19] We will see below that Siberian/Central Asian ancestry was one of the
141 constitutive elements of the Lower Volga-North Caucasus Eneolithic population represented by
142 the two groups.
143
144 A natural interpretation of the Volga cline is that upriver EHG-related ancestors and downriver
145 Berezhnovka-related ones came together to form communities along the length of the river,
146 resulting in a highly variable set of sampled individuals along the genetic gradient. While the
147 origin of the upriver EHG ancestry is clear, as it has antecedents in eastern Europe for thousands
148 of years,[19] that of the downriver Berezhnovka group is less so, as (i) no earlier individuals from
149 the Lower Volga have been sequenced, (ii) the genetic position of the Berezhnovka people is
150 distinct from that of all preceding groups, and (iii) the BPgroup cannot be modeled as a clade
151 with any contemporary or earlier groups (p<0.001). Whatever the origin of BPgroup, a point to
152 which we will return below, we can use it as a proximate source and test Volga cline populations
153 and individuals for consistency with a history of mixture of people related to the BPgroup and
154 EHG (using Karelia[2,19] as an EHG source well outside the Volga area and unlikely to be part of
155 the riverine mating network), as suggested by the PCA. Seven Volga cline populations fit this
156 model (p-values of 0.04 to 0.72) with the only consistently poor fits for Upper Volga, Murzikha,
157 Maximovka, and "Klo" (the Khvalynsk individuals with low Berezhnovka relatedness) (p-values
158 of 1e-66 to 0.006). Three of these populations (other than Klo which we discuss below) are
159 arrayed in the upriver portion of the Volga cline, before its PCA "bend" (from EHG towards the
160 UNHG). Individuals along the downriver portion of the cline can be well-modeled with only the
161 two sources (BPgroup and EHG) (Fig. 1c).
162
163 People on the Volga Cline buried at the Ekaterinovka cemetery likely died between 5050-4450
164 BCE (based on radiocarbon dates on three herbivore bones including a domesticated sheep in the
165 graves of individuals we analyzed that are not expected to be affected by marine reservoir
166 effects; Online Table 1). The Ekaterinovka people were already in the process of mixing with
167 BPgroup-related people from the Lower Volga (24.3±1.3% on average). This contrasts to the
168 earlier hunter-gatherer from Lebyazhinka, who had the lowest estimate of Berezhnovka ancestry
169 on the Volga Cline of only 7.9±3.6%, providing a baseline of this component prior to the
170 Eneolithic and which can also be modeled with only EHG-related ancestry (p=0.21) while
171 Ekaterinovka cannot (p=2e-4). Mixing intensified over time so that 100-200 years later at the site
172 of Khvalynsk[34] which is ~120km from Ekaterinovka (date range of 4500-4350 BCE based on
173 two herbivore bones in the graves of individuals we analyzed), we observe a continuous gradient
174 of admixture which we divide for convenience into three groups: "Khavlynsk high (Khi)"
175 (76.8±1.9% BPgroup), "Khvalynsk medium (Kmed)" (57.3±1.7% BPgroup), and "Khalynsk low
176 (Klo)" (41.2±1.6% BPgroup). Individuals on the downriver portion of the Volga cline exhibited
177 a range of Berezhnovka ancestry from ~14-89% (Fig. 1c) and thus were not clearly dominated by
178 either the old EHG ancestors of the region or the Lower Volga newcomers. Genetic
179 differentiation between Lower Volga (BPgroup) and Ekaterinovka was strong (F_{ST}=0.030±0.001;
180 Extended Data Table 1) and quite probably reflected at least two different linguistic-cultural
181 communities interacting with each other.
182
183 A genetically Volga Cline individual not from the Volga Basin is from Csongrád-Kettőshalom in
184 Hungary, whose direct date is 4331-4073 cal BCE. This individual is estimated to have

9

185 87.9±3.5% of its ancestry from the BPgroup (Fig. 1c) comparable to the most extreme
186 "Khvalynsk high" individuals. The Csongrád individual is one among a group of steppe-like
187 graves that appeared in Southeastern Europe in the late 5th millennium BCE including a cemetery
188 at Giurgiuleşti,[35] Moldova, from which one individual (I20072; 4330-4058BCE) is consistent
189 with being a clade (p=0.90) with BPgroup, and another cemetery at Mayaky, Ukraine.[36]
190 Archaeological analysis has documented long-distance movement of Balkan copper to the
191 Volga-Cline site of Khvalynsk,[34] and the Csongrád and Mayaky individuals were plausibly part
192 of the cultural exchange that mediated this process—a process our results show has no evidence
193 of being contributed to genetically by people with ancestry typical of the Dnipro and Don basins.
194 As we will now see, migrants with ancestry from the Lower Volga Eneolithic populations at the
195 southern extreme of the Volga Cline did settle in the Dnipro area and generated the second major
196 cline of the steppe.
197
198 *Figure 1: Three Eneolithic clines and their neighbors in space and time. (a) Map with*
199 *analyzed sites. (b) PCA analysis using axes formed by a set of ancient West European hunter-*
200 *gatherer (WHG), Siberian, West Asian, and European farmer populations. Selected individuals*
201 *relevant to this study are projected[37] (Methods). (c) qpAdm models fitted on individuals of the*
202 *populations of the three clines. The Volga Cline is generated by admixture between Lower Volga*
203 *(BPgroup) people with upriver Eastern hunter-gatherers (EHG). Populations of the Dnipro*
204 *Cline have UNHG or UNHG+EHG admixture relative to the Core Yamnaya (the hunter-*
205 *gatherer source along this cline is significantly variable). The Caucasus-Lower Volga Cline is*
206 *generated by admixture between lower Volga people with those from the Neolithic Caucasus*
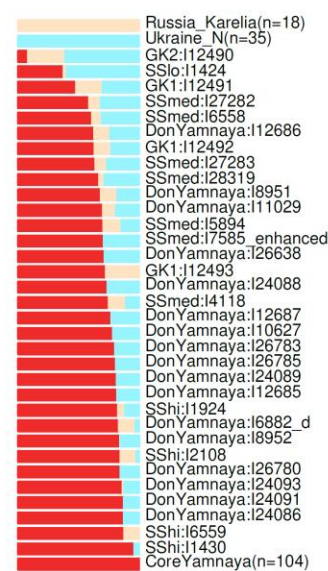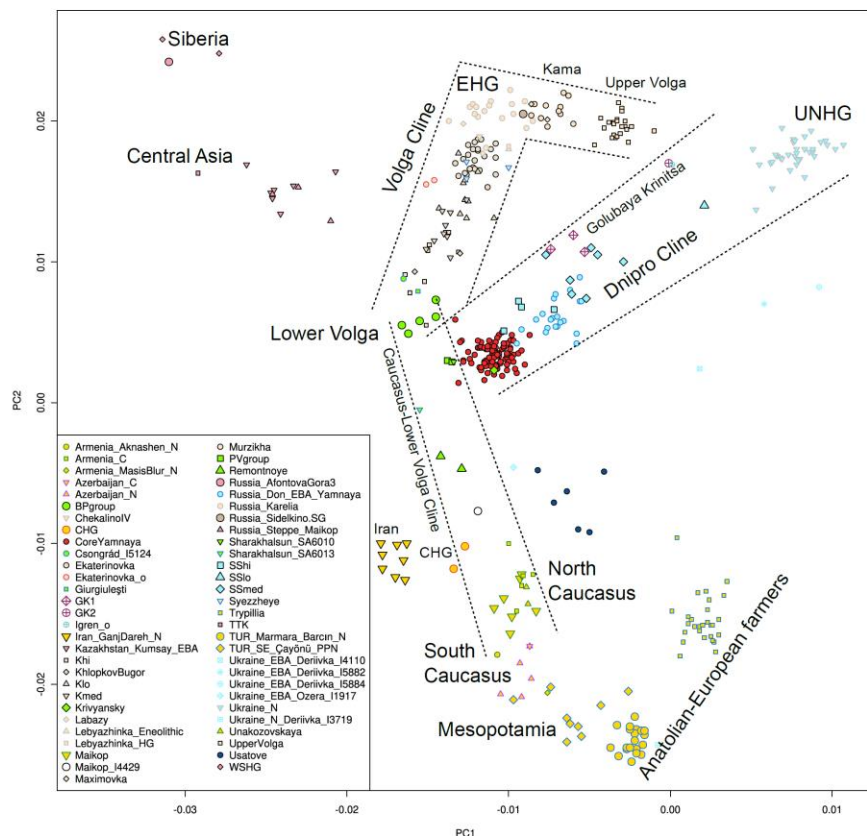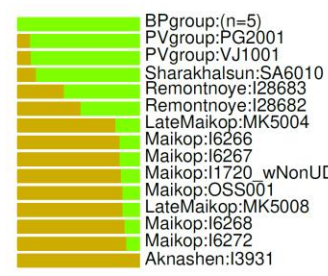207 *(Aknashen-related).*
208
209
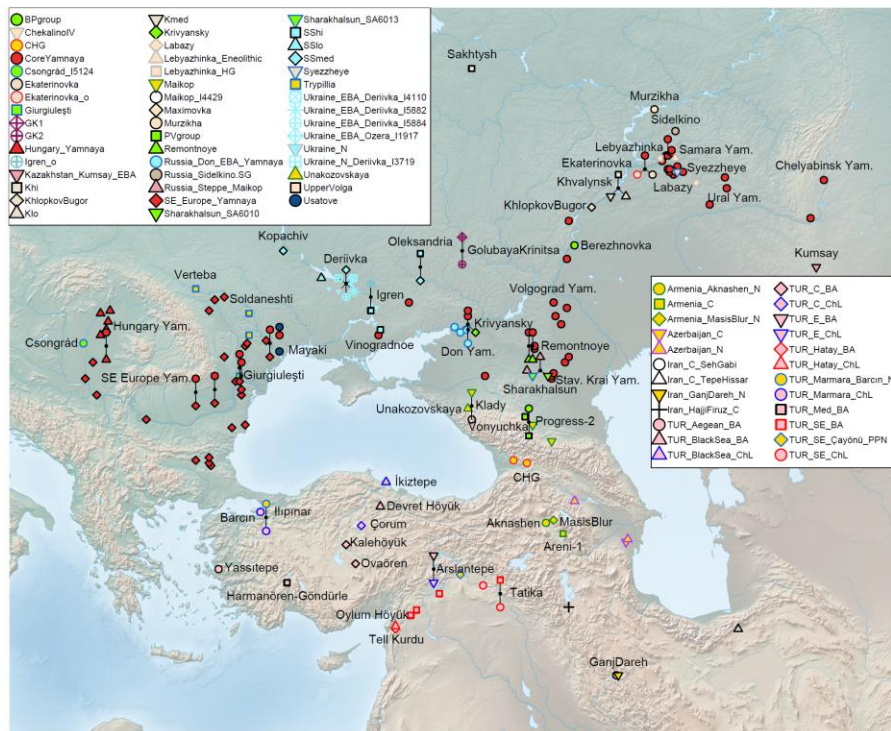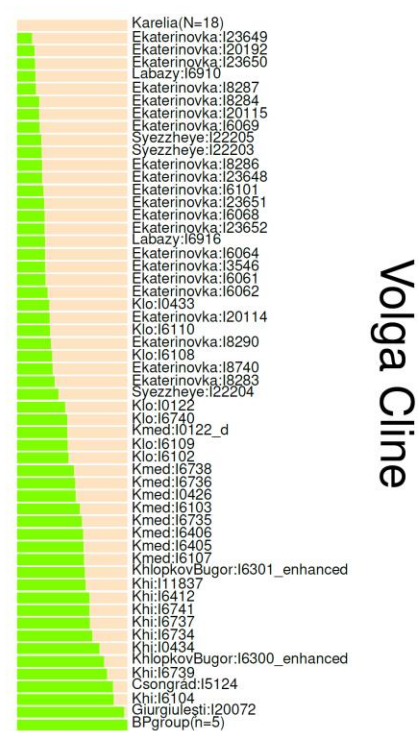
211   *(2) Dnipro Cline:* The Dnipro Cline is formed at one end by Neolithic individuals living along
212   the Dnipro River rapids whose union of calibrated radiocarbon dates is 6242-4542 BCE
213   (UNHG), and at the other end by the Serednii Stih population represented by 13 individuals with
214   good quality data whose union of radiocarbon date ranges uncorrected for freshwater reservoir
215   effects are 4996-3372 BCE. The Dnipro Cline also includes the great majority of later Yamnaya
216   individuals who expanded widely, most of whom are from a genetically homogeneous subset,
217   and we used a large group of these individuals that have high quality data (n=104) to represent
218   "Core Yamnaya" (Supplementary Information, section 2). Close to the Core Yamnaya in PCA
219   are two Eneolithic groups: the Serednii Stih individual from Krivyansky in the Lower Don
220   (4359-4251 BCE), and the PVgroup from the north Caucasus we discussed above as related to
221   the Berezhnovka Lower Volga population. Nonetheless, the Core Yamnaya cannot be modeled
222   as derived from either of these two earlier sources or indeed any other single source (p<1e-4).
223   Their ancestry must have involved some admixture as their position along the highly variable
224   Dnipro/Serednii Stih-associated cline also suggests. People from the Dnipro Cline as a whole are
225   also fully distinct from those of the Volga Cline in PCA, and no pair of populations from the
226   Volga and Dnipro clines form a genetic clade (p<1e-7). This distinctiveness spans a period of
227   three millennia, beginning with earlier groups from Ukraine (UNHG), continuing with those of
228   the Eneolithic Serednii Stih culture, and ending with the Yamnaya at the beginning of the Bronze
229   Age, documenting the distinctiveness of the communities of these two great eastern European
230   rivers and the relative lack of migration between them. A more geographically localized
231   Yamnaya population of the Lower Don (n=23), many (n=17) of which are from the site of
232   Krivyansky, bear no affinity to the Eneolithic individual from the area (Fig. 1). The Yamnaya
233   can thus be traced neither to the north Caucasus (PVgroup), nor to the Lower Don (Krivyansky),
234   nor to the Volga (BPgroup and the rest of the Volga cline). Yet, their position on the Dnipro
235   cline, generated by populations of UNHG ancestry suggests that they emerged there, as a
236   descendant community of people of the Serednii Stih culture.

237

238   The genetic heterogeneity of the Serednii Stih contrasts with the homogeneity of the Core
239   Yamnaya (Fig. 1) which occupies one end of the Dnipro cline. The Core Yamnaya homogeneity
240   is remarkable given that this cluster includes individuals sampled across 5,000 km from Hungary
241   to southern Siberia, a vast slice of Eurasia across which the Yamnaya expanded but, for whatever
242   reason, hardly admixed, at least initially, and at least for the elite subset of people afforded burial
243   in kurgans, with any of the people that previously occupied it. Individuals of the Serednii Stih
244   culture are arrayed along the Dnipro Cline with individuals of high or low Yamnaya affinity
245   found at different sites. Closest to the Core Yamnaya genetically is a Serednii Stih individual
246   from Vinogradnoe from the coast of the Azov Sea which we group with two other individuals
247   from Oleksandria and one from Igren into an "SShi" cluster of greatest Yamnaya affinity. The
248   sampled SShi group does not form a clade with the Core Yamnaya (p=2×10^{-7}). A female from
249   Kopachiv (I7585)[38], represented by a long bone found loose in a Trypillia phase BI-II settlement,
250   is part of a second "SSmed" cluster that is further along the Dnipro Cline; this group also
251   includes three individuals from Oleksandria and three from Deriivka. The SShi and SSmed
252   subsets are largely contiguous with each other, but individual I1424 from Moliukhiv Bugor
253   ("SSlo") is much further apart and close to the UNHG. The true variation within the Serednii
254   Stih plausibly included individuals that fill gaps along the cline, e.g., between SSlo and SSmed,
255   and even extended beyond the sampled variation, occupying the position of the Core Yamnaya
256   itself. The Don Yamnaya largely overlap the Serednii Stih individuals, and the Don Yamnaya are

257 discontinuous with the earlier Eneolithic individual from that location (p=7e-15). An interesting
258 material correlate is seen in settlement continuity at stratified sites of the Konstantinovka culture
259 on the Lower Don where the Don Yamnaya continued to settle in the same place as the earlier
260 Serednii Stih, a continuity not seen in the Volga-Ural steppes, where most Eneolithic settlement
261 sites exhibited no re-use by the Yamnaya.
262
263 *qpAdm* analysis reveals that all groups visually on the Dnipro Cline in the PCA can be well
264 modeled with either UNHG or GK2 (individual I12490 from Golubaya Krinitsa in the Middle
265 Don dated 5610-5390 BCE) at one extreme, and Core Yamnaya on the other (p-values between
266 0.07 and 0.85). Some populations of the cline (SSmed) can be modeled as Core Yamnaya and
267 either GK2 (p=0.43) or UNHG (p=0.27); others, like the Don Yamnaya, can be modeled only as
268 Core Yamnaya and UNHG (p=0.08) but not GK2 (p=0.0001); and others, like SShi, as Core
269 Yamnaya and GK2 (p=0.08) but not UNHG (p=0.003). Thus, the hunter-gatherer end of the
270 Dnipro Cline is not clearly UNHG or GK2. We therefore model individuals of the Cline with
271 ancestry from any population from the UNHG-EHG cline (Fig. 1c), observing that individuals
272 can be modeled as a mix in which UNHG ancestry predominates but EHG ancestry is also
273 present in individuals (similar to GK2). This reflects the admixture of Caucasus-Lower Volga
274 ancestry with hunter-gatherers of the Dnipro-Don (or UNHG-GK2) area, rather than other areas
275 of eastern Europe (such as the Volga area) in which the hunter-gatherer population was EHG.
276 Using Core Yamnaya as a source for the Serednii Stih is, of course, ahistorical, as they postdate
277 the Serednii Stih, and so the model of Core Yamnaya + UNHG/GK2 admixture must be
278 interpreted as admixture between local Neolithic residents of the Dnipro-Don area with a second,
279 unsampled, Eneolithic source, which together account for the ancestry of the Core Yamnaya
280 and—with even more UNHG/GK2 ancestry—of the Dnipro cline as a whole.
281
282 The Don, situated geographically between the Dnipro and Volga, is represented in our data by
283 individuals from Golubaya Krinitsa (in the Middle Don) and Krivyansky (in the Lower Don).
284 Golubaya Krinitsa contained two archaeologically contrasting styles of graves, one compared to
285 Dnipro Neolithic graves and the other like Serednii Stih.[39] The GK2 individual can be modeled
286 as 66.6±4.7% UNHG and 33.4±4.7% EHG (p=0.39), suggesting that intermediate populations
287 between the Dnipro hunter-gatherers (represented by UNHG) and the EHG existed not only in
288 the Upper Volga (the upriver portion of the Volga cline), but also in the Middle Don. When we
289 examine populations using the most ancient sources (Karelia, UNHG, and CHG) of the steppe
290 and Caucasus, we see that the Eneolithic population of the Lower Don at Krivyansky and
291 Neolithic individuals from Golubaya Krinitsa can all be well modeled with variable proportions
292 of CHG-related ancestry (Fig. 2a). The most CHG-related ancestry is seen at Krivyansky
293 (58.9±2.4%); there is less (25.3±2.1%) in three individuals which (Fig. 1) we group as GK1; and
294 individual GK2 is consistent with having none or very little (4.0±2.2%), fitting the simpler
295 EHG+UNHG model mentioned above. Thus, the Neolithic and Eneolithic individuals of the Don
296 were a mixture of European hunter-gatherer ancestries (intermediate between the Dnipro-
297 sampled UNHG and the Volga-sampled EHG, paralleling the intermediate geographic position
298 of the Don) and southern CHG-related ancestry (Fig. 2a). When did the CHG-related ancestry
299 reach the Don area? Its presence in a ^{14}C-dated individual of the GK1 group (I12491/5557-5381
300 BCE) and others from the region[7] suggest it was present there as early as the Neolithic. However,
301 its absence from GK2 of similar ^{14}C age proves that it was not a general feature of the Neolithic
302 population. Both GK1/GK2 dates may be too early given that archaeologists of Golubaya

303 Krinitsa interpreted people of the site as in contact with people of the much later Eneolithic
304 Serednii Stih Culture.[40] Moreover, an outlier Serednii Stih individual from Igren (I27930; 4337-
305 4063 cal BCE) is consistent with all its ancestry coming from GK2; this could be an example of
306 long-distance migration from the Don to the Dnipro, but also casts some doubt on the much older
307 date of the GK2 individual, as genetic identity across more than a millennium in two different
308 locations seems implausible given the diverse admixtures taking place throughout the steppe
309 during the Eneolithic. The interpretation of the Golubaya Krinitsa population is further
310 complicated by uncertainties as to their date due to freshwater reservoir effects in individuals
311 who have a diet heavily reliant on freshwater fish. This can make nominal dates up to a
312 millennium too old in this region.[41] Further sampling along the Don would shed light on the
313 distinctive processes and temporality of the ancestry change along this major river and place both
314 the Golubaya Krinitsa individuals and those of Krivyansky on the Don mouth in their proper
315 context.

316
317 It has been suggested[7] that the Yamnaya were formed by a substantial contribution of ~65%
318 Golubaya Krinitsa people from the Middle Don, that already had ~20-30% CHG-related
319 ancestry, with an additional ~35% CHG-related ancestry. This scenario implies that they were
320 formed in the Don area as the result of the CHG-related admixture observed there. Our results
321 contradict this as the Core Yamnaya do not fit models with CHG-related and either GK1/GK2
322 sources (p<1e-6), suggesting that they have ancestry not accounted for by the model of ref.[7] To
323 understand the source of this ancestry, we fit the model of Fig. 2a (with the most ancient sources:
324 Karelia, UNHG, and CHG) and observed that its failure ($p=2x10^{-20}$) is explained by the fact that
325 it severely underestimates their shared genetic drift with both Afontova Gora-3 from Upper
326 Paleolithic Siberia (Z=-5.2) and Anatolian Neolithic (Z=-6.8).[6] Thus, the Yamnaya must have
327 Siberian- and Anatolian-related ancestry and cannot be a simple mixture of Caucasus- and
328 Middle Don hunter-gatherers. A Volga source of the Siberian-related ancestry is strongly
329 suggested by the fact that the Volga cline is shifted away from the Dnipro cline precisely in the
330 direction of Siberian populations (Fig. 1b). That the Volga cline populations had such ancestry is
331 proven by the fact that the model of Fig. 2a fails them precisely for the same reason as it does the
332 Core Yamnaya as it also underestimates shared drift with Afontova Gora-3, e.g., for BPgroup
333 ($p=1x10^{-8}$ and Z=-4.5). This extra ancestry in BPgroup is also affirmed positively by the fact that
334 it can be modeled as a mixture of Krivyansky and ~24% Central Asian (Siberian-related)
335 Tutkaul[19] ancestry (p=0.13). When we fit both Krivyansky and the BPgroup with the same
336 model that includes all relevant ancestries (Fig. 2b)—CHG, GK2, and Tutkaul—we see that
337 indeed Krivyansky has little to no Central Asian ancestry (5.1±3.6%) but it can be fitted as
338 56.7±2.6% CHG-related and 43.3±2.6% GK2 alone (p=0.37), while BPgroup does have
339 29.3±2.2% Tutkaul ancestry. The model of Fig. 2b corrects for the missing Siberian-related
340 ancestry in the Yamnaya, predicting shared genetic drift with Afontova Gora-3 reasonably
341 accurately (Z=-1.7), but still fails (p=1e-9) as it does not predict shared drift with Anatolian
342 Neolithic (Z=-6.1). Thus, while ancestry from the Volga can explain the Siberian relatedness of
343 the Core Yamnaya it cannot explain the Anatolian Neolithic relatedness as this was not a
344 component of Volga cline populations.

345
346 Our new data resolve the extent of the spread of eastern "Central Asian" or "Siberian" ancestry
347 into the Pontic-Caspian steppe. It was present, during the Eneolithic, on the Volga and in the
348 North Caucasus Steppe, but further west on the Don there still existed populations without much

349   or any of it like those at Krivyansky and Golubaya Krinitsa. When we repeat our modeling of the
350   Volga Cline as a mixture of BPgroup and EHG sources but add either a western (UNHG) or
351   eastern (Tutkaul) source (Fig. 2c,d) we see that individuals on the cline remain largely well-
352   modeled as linear combinations of the two groups: Fig. 2c shows the characteristic "bend" of the
353   Volga Cline with a portion showing variable Berezhnovka ancestry and the other (including
354   many individuals from the Upper Volga and Murzikha) showing variable UNHG ancestry which
355   increases further still in the GK2 individual from the Don. Fig. 2d shows that individuals of the
356   Volga Cline have more Tutkaul ancestry than is explained by the simpler Berezhnovka-Karelia
357   model; however, the deviations are small (4.4±2.6% Tutkaul ancestry for "Khi"). The Eneolithic
358   Volga was an admixture zone between downriver BPgroup people with upriver EHG ones that
359   included Central Asian ancestry mainly via BPgroup. Crucially, the Core Yamnaya fail all
360   models of Fig. 2a-d (p<1e-8), and thus its origins must include a different blend of ancestry than
361   the CHG-EHG-UNHG-Tutkaul ancestries involved in these models. As we will now see, this
362   ancestry came from a third cline formed between the Caucasus Neolithic populations and those
363   of the Lower Volga.
364
365   ***Figure 2. The three Eneolithic clines in the context of Eneolithic and Bronze Age admixture.***
366   *Six 3-source models elucidate a complex history of admixture. Individuals plotted at the triangle*
367   *edge fit (p>0.05); the simpler 2-source model is plotted for individuals with a negative*
368   *coefficient from one of the three sources. The corners of each triangle represent the sources.*
369   *Unplotted individual all give fits at P<0.05 and so should be viewed as poorly described by the*
370   *model. (a) Caucasus and European hunter-gatherer admixtures in the "Old Steppe": Krivyansky*
371   *on the Lower Don received much more CHG-related admixture than upriver people of the*
372   *Middle Don at Golubaya Krinitsa. In the Middle and Upper Volga and the Kama River,*
373   *populations belonged to the old EHG cline with negligible CHG-related influence. (b) The*
374   *"Don-Volga" difference. On the Lower Volga and North Caucasus piedmont, the BPgroup did*
375   *receive CHG-related ancestry like its western Lower Don counterpart at Krivyansky; but, unlike*
376   *Krivyansky, it also received ancestry from Central Asia; this eastern influence was higher still in*
377   *the Bronze Age Steppe Maikop. (c) The "Volga Cline" vis-à-vis the Don: populations at*
378   *Khvalynsk, Klopkov Bugor, and Ekaterinovka are clinal between the Berezhnvoka cluster on the*
379   *Lower Volga and the upriver EHG-like populations of the Middle Volga (Labazy and*
380   *Lebyazhinka). (d) the "Volga Cline" vis-à-vis Central Asia: a slight excess of Central Asian*
381   *ancestry in the Khi subset of Khvalynsk. (e) the "Dnipro" cline: the Core Yamnaya are on one*
382   *end of a cline that also includes the Don Yamnaya and Serednii Stih populations. The cline is*
383   *formed by admixture from the "Caucasus-Lower Volga" (CLV) cline that is formed by*
384   *differential admixture of Neolithic Caucasus and BPgroup people. The CLV Cline includes*
385   *diverse people buried in kurgans at Berezhnovka, Progress-2, Remontnoye, and Maikop sites*
386   *Klady and Dlinnaya-Polyana ~5000-3000 BCE. (f) "West Asian": CLV ancestry first appears in*
387   *the Chalcolithic population at Areni-1 in Armenia and is also present in the Bronze Age at*
388   *Maikop. The majority of the ancestry in both populations is from West Asian sources from the*
389   *Mesopotamia-Caucasus (or Çayönü-Masis Blur-Aknashen) cline. Chalcolithic and Bronze Age*
390   *Anatolians lack CLV ancestry but traces of it can be found in Bronze Age Central Anatolians.*

15
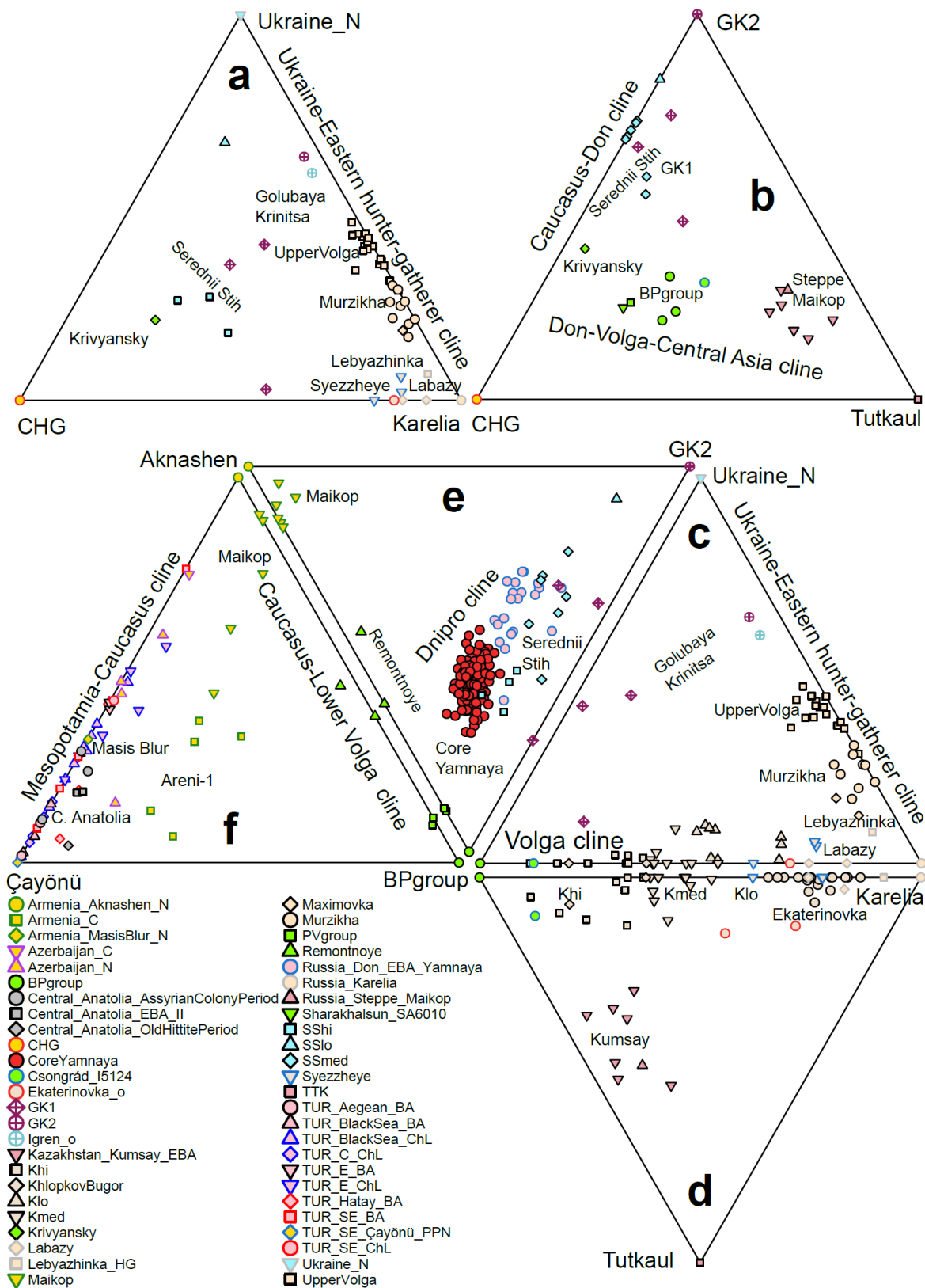
391
392

393   *(3) Caucasus-Lower Volga Cline (CLV):* The Yamnaya are on the edge of the Dnipro cline,
394   having less UNHG/GK2-related ancestry than other cline populations; thus, they cannot be
395   modeled in terms of them alone (Fig. 1), but must have possessed more of a second source of
396   ancestry. We found that the only consistently fitting (p=0.67) two-way model for the Core
397   Yamnaya involved 73.7±3.4% of the SShi subset of the Serednii Stih population and 26.3±3.4%
398   from a population represented by a sample of two individuals from Eneolithic burial sites at
399   Sukhaya Termista I (I28682) and Ulan IV (I28683), dated 4152-3637 BCE near the village of
400   Remontnoye, north of the Manych Depression on the watershed between the Lower Don and
401   Caspian. The Remontnoye population is on neither the Volga nor Dnipro clines and is neither
402   genetically close (Fig. 1) nor forms a clade (p<1e-10) to any other single sampled population.
403   We determined that it had at least two sources: a southern one from the Caucasus—either
404   descendants of the Aknashen Neolithic in Armenia[6], or ancestors of people of the Bronze Age
405   Maikop[5] culture—and a northern one from a population from the low-EHG end of the Volga
406   Cline such as the BPgroup. The Caucasus component is about half when using either Aknashen
407   (44.6±2.7%; p=0.66) or Maikop (48.1±2.9%; p=0.44) as the proxy for the southern source. We
408   also observed that the main cluster of Maikop individuals, including those buried in kurgans in
409   Klady and Dlinnaya-Polyana, can be modeled as having 86.2±2.9% (p=0.50) Aknashen ancestry.
410   Thus, there exists a Caucasus-Lower Volga (CLV) cline: Aknashen-Maikop-Remontnoye-
411   Berezhnovka. These four populations are arrayed in order of decreasing Caucasus Neolithic
412   component, concordant with their south-to-north geographical location. However, there were
413   also populations of the CLV cline that bucked this latitudinal trend, such as the people of the
414   North Caucasus at Progress-2 and Vonyuchka-1 that, unlike their Maikop neighbors, had little
415   Caucasus Neolithic ancestry and were most like the people of Berezhnovka-1 in the Lower
416   Volga. These violations of the genetic-geographic pattern prove long-range connectivity across
417   the CLV area; they also caution us not to easily interpret genetic position along the CLV cline as
418   predictive of position within the CLV geography.
419
420   What was the proximal source for the southern ancestry of the intermediate populations of the
421   CLV cline? Aknashen makes a poor choice, as it is both geographically remote from the steppe
422   and earlier by two millennia (5985-5836 BCE) than Remontnoye. Neither is Maikop a good
423   proximal source; it is geographically closer, but postdates (3932-2934 BCE) Remontnoye.
424   Settlements at Meshoko and Svobodnoe, dated 4466-3810 BCE,[42] provide a temporally,
425   geographically, and archaeologically plausible source, as they exhibit exchanges of exotic stone,
426   copper, and stone mace heads with Volga Cline sites, setting the context for the expansion of
427   Aknashen-like ancestry northward and Berezhnovka-like ancestry southward. These settlements
428   are temporally earlier than Maikop and later than two individuals from Eneolithic
429   Unakozovskaya (ref.[5] 4607-4450 BCE, and this study) in the North Caucasus; however, unlike
430   Aknashen and Maikop, the Unakozovskaya population is not a good genetic source for
431   Remontnoye, as the model BPgroup+Unakozovskaya fails (p<0.001) by overestimating (Z=3.8)
432   shared genetic drift with the CHG. The Unakozovskaya was not exactly the same genetically as
433   the Maikop who succeeded them (p=2e-11) but were genetically similar (Fig. 1) and can be
434   modeled as 95.3±6.3% Maikop and 4.7±6.3% CHG (p=0.46). Thus, there were three elements of
435   ancestry in the North Caucasus in the Eneolithic: (i) Aknashen-related ancestry was dominant,
436   representing the spread of the Neolithic from the south across the Caucasus mountains; (ii) there
437   was some variation in CHG-related ancestry as suggested by the Maikop-Unakozovskaya
438   contrast; and (iii) there was also a small component of northern Lower Volga ancestry of about

17

439 one seventh in the Maikop on average. Thus, in the north Caucasus there lived, side by side, both
440 "high steppe" ancestry people genetically close to the Lower Volga Berezhnovka population
441 (individuals at Progress-2 and Vonyuchka-1), as well as "low steppe" ancestry people in which
442 the Lower Volga ancestry had been diluted by the greater contribution of the (Aknashen-related)
443 Caucasus Neolithic.
444
445 The Remontnoye and Berezhnovka people, like the Maikop people, were buried in kurgans.
446 Thus, the kurgan burial rite was widespread 5000-3000 BCE among people of diverse ancestry
447 from both the edges and middle of the CLV Cline, suggesting that—regardless of its ultimate
448 origin and whether it was culturally adopted or spread by migration—it was common among the
449 people of the CLV region.[22] In contrast, a distinctive position of the body on the back with knees
450 raised and the floor of the burial pit covered with red ochre was shared by all the steppe groups
451 including Serednii Stih, groups on the Volga Cline, and Remontnoye, while the Maikop burial
452 position was contracted on one side. Thus, some funeral customs united Maikop with the steppes
453 and others separated them.
454
455 The discovery of the CLV Cline suggests a solution to the question of the origin of the Dnipro
456 Cline and thus the genetic origins of the Yamnaya. Most of their ancestors were people of the
457 CLV Cline, similar to the sampled Remontnoye individuals. These CLV ancestors were drawn
458 into the Dnipro-Don region and mixed with local groups to form Serednii Stih people and
459 eventually the Yamnaya. It must be emphasized that the CLV and Dnipro-Don sources need not
460 have been identical to the sampled Remontnoye and SShi populations or have lived close to the
461 sampling locations of these two populations. The Dnipro Cline can be fit (Fig. 2e) by a 3-way
462 model in which the GK2 admixed with groups of mixed Aknashen and Berezhnovka ancestry.
463 We note the aforementioned caveat that either of GK2 or UNHG could be contributing to the
464 Dnipro Cline, but chose GK2 in Fig. 2e as this model has a higher p-value (p=0.93) for the Core
465 Yamnaya than the alternative with UNHG as the source (p=0.04); however, we do not take this
466 as evidence that the GK2 population was a better source than the UNHG as we have far better
467 data for UNHG (n=35 individuals) than GK2 (n=1), which provides more power to detect slight
468 but qualitatively unimportant oversimplifications in models. Note also, that GK2 is itself ~2/3
469 UNHG in ancestry, and that the proportion of either GK2 (22.5±1.8%) or UNHG (17.7±1.3%) is
470 similar, and about one fifth. A full exploration of 3-way models (Supplementary Information
471 section 2) reveals that the Yamnaya could have been formed from diverse (but similar) distal
472 sources which include populations of (i) Neolithic or Chalcolithic age from Armenia[6,9] and
473 Azerbaijan[43,44] representing the "Caucasus Neolithic", (ii) GK2, UNHG, or Serednii Stih
474 representing the Dnipro-Don area, and (iii) BPgroup or PVgroup representing the Lower Volga-
475 north Caucasus Eneolithic. What is invariant among the class of 2- and 3-way models for the
476 Core Yamnaya is that they posit their descent from people of the CLV Cline (the remaining four
477 fifths of their ancestry) who admixed with Dnipro-Don people of substantial UNHG ancestry.
478
479 Our results show that movement of people and culture we document as having occurred along the
480 CLV Cline was the vector by which Caucasus-derived ancestry like that present in the Aknashen
481 Neolithic population flowed into the steppe and into the ancestors of the Yamnaya[45]. Crucially,
482 the successful Remontnoye+SShi model predicts shared genetic drift with the Anatolian Neolithic
483 outgroup well (Z=-0.8). CLV cline populations can account for both Siberian-related (via the
484 Lower Volga component) and Anatolian Neolithic-related (via the Caucasus Neolithic component)

485   affinities of the Yamnaya. Archaeological evidence shows that Balkan copper was traded during
486   the late 5th millennium BCE across the steppes to North Caucasus farmer sites (Svobodnoe) and
487   to the Volga (Khvalynsk), while Neolithic pots like those from Svobodnoe appeared in Dnipro-
488   Don steppe sites connected with the Seredni Stih culture (Novodanilovka), documenting an active
489   period of cultural exchange that was the context for the movement of groups of mixed
490   BPgroup/Aknashen-related ancestry into the Dnipro-Don steppes.
491
492   *CLV impact in the Caucasus and Anatolia*: CLV Cline people also had an impact further south,
493   in Armenia and Anatolia (Fig. 2f). The earliest evidence of steppe ancestry south of the Caucasus
494   is at Areni-1 in Chalcolithic Armenia around 4000 BCE[9], documenting its southward penetration
495   which parallels the incursion of Caucasus ancestry generating the Volga/Dnipro clines on the
496   steppe. Our analysis (Supplementary Information section 2) clarifies that in Areni-1 the Lower
497   Volga ancestry (26.9±2.3% BPgroup) admixed with a local "Masis Blur"-related Neolithic
498   substratum, in contrast to the North Caucasus (at Maikop) where it combined with an
499   "Aknashen"-related Neolithic substratum. The Aknashen/Masis Blur distinction of the Neolithic
500   population of Armenia reflected the dilution of the native CHG ancestry that was higher in
501   Aknashen than in Masis Blur.[6] We can model Masis Blur as 33.9±8.6% Aknashen and
502   66.1±8.6% Çayönü ancestry (p=0.47) associated with the Pre-Pottery Neolithic of the Tigris
503   Basin of Mesopotamia[46], thus documenting the spread of early Neolithic ancestry into the
504   Caucasus that formed a cline of diminishing Mesopotamian-related and increasing CHG-related
505   ancestry: Çayönü-Masis Blur-Aknashen. Using CHG as the source, we see that the two
506   populations from Armenina differed indeed in their retention of CHG ancestry, with more
507   (42.0±3.8%) in Aknashen than in Masis Blur (13.7±4.0%). Some Anatolian Chalcolithic and
508   Bronze Age groups can be derived entirely from this north-south Caucasus-Mesopotamian cline
509   (Fig. 2f), while others also have ancestry from the east-west Mesopotamian-Anatolian cline,
510   lacking any steppe ancestry.[22,43,45,47,48]
511
512   The discovery of the Mesopotamian-Caucasus cline allows us to study the ancestry of the
513   population of Bronze Age Central Anatolia[22] from the Early Bronze Age (2750-2500 BCE),
514   Assyrian Colony (2000-1750 BCE), and Old Hittite (1750-1500 BCE) periods. We cannot be
515   certain of what languages were spoken by these individuals in what may well have been
516   multilingual societies, but we document for the first time that they had a small amount of CLV
517   cline ancestry combined with Mesopotamian (Çayönü) ancestry (Supplementary Information,
518   section 2; Fig. 2f; Extended Data Fig. 1). The inferred amount of ancestry from the CLV or
519   CLV-influenced source depends on the amount of "dilution" of this ancestry in the source: more
520   such ancestry is required from populations of higher dilution. For example, it is estimated as
521   10.8±1.7% ancestry (p=0.14) from the BPgroup, or about double 19.0±2.4% from Remontnoye
522   (p=0.19)—whose own ancestry is about half from the BPgroup—or 33.5±4.8% of Armenia_C
523   ancestry (p=0.10)—where the BPgroup ancestry is lower.
524
525   The exact source of the steppe ancestry in Anatolia cannot be precisely determined, but it is
526   noted that all fitting models involve some of it (Extended Data Fig. 1a). Some of the steppe-
527   related sources can be rejected on chronological grounds; for example, the Core Yamnaya itself
528   (12.2±2.0%; p=0.10) as well as western Yamnaya-derived populations from Southeastern Europe
529   such as from Boyanovo or Mayaky Early Bronze Age[36] (Extended Data Fig. 1b). Moreover,
530   when we consider pairs of steppe sources (and can thus place the steppe ancestry at varying

19

531 points along the Volga, Dnipro, and CLV clines), we observe a negative hunter-gatherer
532 contribution (-3.4±2.6% EHG) on the Volga cline, and also on the Dnipro cline (-2.3±2.7%
533 UNHG or -3.9±3.5% GK2); thus, there is no evidence that the admixing population had more
534 EHG/UNHG/GK2 ancestry than the BPgroup/Core Yamnaya endpoints of these two clines
535 (Supplementary Information section 2). The admixing population in this analysis contributed a
536 significant amount of BPgroup ancestry (8.8±2.7%) from the CLV cline and was consistent with
537 being on that cline (p=0.129). Thus, a model in which the steppe ancestry is derived from the
538 Caucasus-Lower Volga Eneolithic is not only geographically and chronologically plausible but
539 also genetically so. The steppe+Mesopotamian class of models fit the Central Anatolian Bronze
540 Age but do not fit any of the Chalcolithic/Bronze Age Anatol0ian regional subsets (p<0.001; the
541 BPgroup+Çayönü model is shown in Extended Data Fig. 1c), indicating that their success is not
542 due to their general applicability. Moreover, the steppe ancestry in the Central Anatolian Bronze
543 Age is observed in all individuals of the three periods (Extended Data Fig. 2d) and is thus not
544 driven by any outlier individuals within the population. Its presence in both Early Bronze Age
545 individuals from Ovaören south of the Kızılırmak river and in Middle Late Bronze Age
546 individuals from Kalehöyük just within the bend of the river is consistent with the idea that the
547 Kızılırmak formed an Anatolian-Hattic linguistic boundary that was crossed some time before
548 the ca. 1730 BCE conquest of Hattusa by the Hittites.[49] Regardless of the linguistic identity of
549 the sampled individuals, the truly unique blend of CLV and Mesopotamian ancestries found in
550 the Central Anatolia Bronze Age calls for an explanation.
551
552 How and when did this blend reach Central Anatolia? We note that populations along the path
553 from the steppe to Central Anatolia can all be modeled with BPgroup ancestry and distinctive
554 substratum ancestries along the north-south / Caucasus-Mesopotamia cline: Aknashen-related in
555 the North Caucasus Maikop; Masis Blur-related in the South Caucasus Chalcolithic population
556 of Armenia at Areni-1; and Mesopotamian Neolithic for the Central Anatolian Bronze Age
557 (Extended Data Fig. 1e, f). This series of admixtures had certainly begun by ca. 4300-4000BCE
558 (the date range of the Armenia_C population[9]) and can be dated using DATES to 4382±63BCE
559 (Extended Data Fig. 2f). The Pre-Pottery Neolithic population of Çayönü was itself genetically
560 halfway between that of Mardin[10], 200km to the east, and the Central Anatolian pottery Neolithic
561 at Çatalhöyük along the east-west / Mesopotamian-Anatolian cline. Chalcolithic/Bronze Age
562 people from Southeastern and Central Anatolia all had ancestry from the same Çatalhöyük-
563 Mardin continuum and such populations may have been proximal sources for the Çayönü-related
564 ancestry of the Central Anatolian Bronze Age population (Supplementary Information section 2).
565 If the Proto-Anatolian population was formed in this region by the admixture of CLV cline
566 people with Mesopotamian ones then their descendants may have been present there at the
567 unknown site of Armi whose Anatolian personal names are recorded by their neighbors in the
568 kingdom of Ebla in Syria.[50] We thus propose the following hypothesis: that CLV cline people
569 migrated southwards ca. 4400BCE, or about a millennium before the appearance of the
570 Yamnaya, (admixing with different substratum populations along the way) and then westwards
571 before finally reaching Central Anatolia.
572
573 We in fact find Y-chromosome evidence that is consistent with the autosomal evidence. Sporadic
574 instances of the steppe-associated Y-chromosome haplogroup R-V1636 in West Asia occurred at
575 Arslantepe[43] in Eastern Anatolia and Kalavan[9] in Armenia in the Early Bronze Age (~3300-2500
576 BCE) among individuals without detectible steppe ancestry[45] and these could be remnants of the

577   dilution process. This haplogroup was found in the male individual from Remontnoye, both
578   individuals from Progress-2[5] and two of three males from Berezhnovka, in addition to its
579   occurrence in eleven individuals of the Volga Cline and thus was a prominent lineage of the pre-
580   Yamnaya steppe. Isolated instances have also been found beyond the steppe in Corded Ware
581   individuals from Esperstedt in Germany[17] and Gjerrild in Denmark.[51] The expansive distribution
582   of R-V1636 on the steppe and beyond contrasts with its disappearance on the steppe after the
583   Yamnaya arrived on the scene: a single individual (SA6010; 2886-2671 BCE) from
584   Sharakhalsun[5] has it, with a genetic profile consistent with CLV ancestry (Fig. 2), the last
585   detected holdout of this once pervasive population (Fig. 3).

586
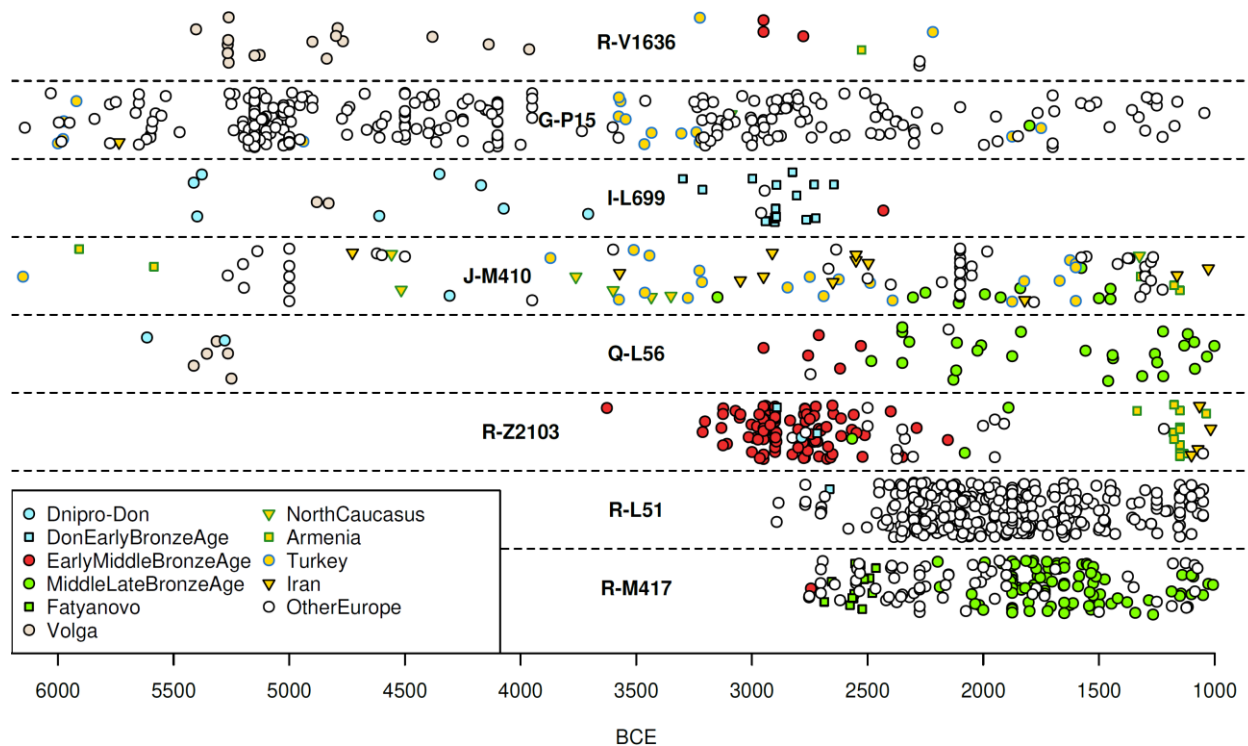
587
588   *Figure 3. Patrilineal succession. Temporal distribution of key Y-chromosome haplogroups from*
589   *Kazakhstan, Kyrgyzstan, Mongolia, Russia, Turkmenistan, Ukraine, Uzbekistan, and*
590   *comparative regions of Europe and West Asia 6000-1000 BCE. The Early and Middle Bronze*
591   *Age group includes the Yamnaya, Afanasievo, Poltavka, Catacomb, Chemurchek, and North*
592   *Caucasus cultures; the Middle and Late Bronze Age group individuals of diverse cultures down*
593   *to 1000 BCE including those of the Sintashta, Andronovo, Potapovka, and Srubnaya cultures.*

594

595   **The Yamnaya expansion broke correlations between geography and genetics**

596

597   We have traced the origins of the Yamnaya to the Dnipro Cline and the populations of the
598   Serednii Stih culture: the Yamnaya were formed as people of the CLV cline admixed with
599   people of the Dnipro-Don area having UNHG ancestry. Deeper in time, the CLV cline was
600   formed by the admixture of Aknashen-related and BPgroup-related people who, in turn, were
601   formed by earlier mixtures still: the Caucasus Neolithic represented at Aknashen by the
602   admixture of CHG people with Neolithic farmers of the Fertile Crescent[6,10] and the lower Volga
603   Eneolithic people represented by BPgroup had ancestries that were related to CHG, EHG, and

21

604 people from Siberia or Central Asia. Dating this complex sequence of admixtures could be done
605 by generating time transects of fine resolution in all relevant areas from which the ancestors of
606 the Yamnaya were drawn across the millennia until they finally combined to form the Yamnaya
607 genetic profile somewhere in the territory of the Serednii Stih culture: seeing the admixture "as it
608 happened" through the lens of ancient DNA. Our study has revealed the outlines of this
609 millennia-long process and future studies may fill in the details.
610
611 A different way is to date the admixture itself in the genomes of the Yamnaya using methods like
612 DATES[52] to measure the average sizes of stretches of ancestry related to UNHG/EHG hunter-
613 gatherer populations on the one hand, and West Asian/Caucasus-related populations on the other,
614 as this reflects the number of generations elapsed since mixture began and stretches of ancestry
615 broke down. This population contrast aligns to the differentiation along PC2 (Fig. 1). We would
616 also like to model the Core Yamnaya in terms of ancestry along the Dnipro cline itself (their last
617 and most proximal admixture event), but unfortunately this is challenging given that the
618 Yamnaya themselves are the end of the Dnipro cline (Fig. 1). The inferred date of 4038±48 BCE
619 (Extended Data Fig. 2a) should thus be viewed with caution given the complex history of the
620 ancestors of the Yamnaya, and admixture may have taken place both before and after this date.
621 Nonetheless, an Eneolithic time frame (with a small standard error of <2 generations) proves that
622 the admixture derived using qpAdm and observed visually in PCA did not occur in the remote
623 past, but corresponds, at least in part, to the efflorescence of the Serednii Stih culture that our
624 reconstruction points to as ancestral to the Yamnaya.
625
626 Uncertainty about where, exactly, within the territory of the Serednii Stih culture the ancestors of
627 the Core Yamnaya lived contrasts with their expansive distribution after the formation of the
628 Yamnaya archaeological horizon: individuals we identified as "Core Yamnaya" (Extended Data
629 Table 2) cluster in a small portion of the PCA (Fig. 1) and are from several countries: China,
630 Hungary, Kazakhstan, Moldova, Romania, Ukraine (Extended Data Table 2), and 15 different
631 locations in Russia (Fig. 4a). The homogeneity is also evident in a mean F_{ST} of 0.005,
632 comparable to that between modern northern Europeans (Extended Data Table 3). This
633 remarkable homogeneity across vast geographical distances of the "eastern" expansion of the
634 Yamnaya shows that many of them mixed very little if at all with any of the people that inhabited
635 the Eurasian steppe before them. The Don Yamnaya (Fig. 4a) are distinctive and can be modeled
636 with 79.4±1.1% Core Yamnaya and 20.6±1.1% UNHG ancestry; the actual proportion of Core
637 Yamnaya ancestry may be lower if, as is plausible, the Core Yamnaya admixed with a Serednii
638 Stih population of partial UNHG ancestry (e.g., 40.0±4.7% with SSmed as the Serednii Stih
639 source). The Don Yamnaya were formed in the late 4th millennium BCE (Extended Data Fig.
640 2b), a time during which unmixed UNHG, after a millennium or more of the Serednii Stih
641 culture, would be rare if they existed at all.
642
643 The western expansion of the Core Yamnaya also brought them into southeastern Europe;
644 Yamnaya there or other individuals of "high steppe ancestry" can be found as far west and south
645 as Albania and Bulgaria.[6] Many western Yamnaya cluster with the Core Yamnaya, but many
646 also deviate in the direction of Neolithic and Chalcolithic populations of southeastern and central
647 Europe (Fig. 4b) and can be modeled with admixture from such populations (Extended Data
648 Table 4). This admixture also took place in the late 4th millennium BCE (Extended Data Fig. 2c),
649 after the sporadic early Chalcolithic migrations into southeastern Europe from the steppe.[36] It is

650 interesting that after the Don Yamnaya formed they participated little or not at all in the Core
651 Yamnaya expansion to either the Altai or SE Europe, and thus the Lower Don represented a cul-
652 de-sac for the Yamnaya expansion.
653
654 The late 4th millennium BCE admixtures with European farmers and UNHG-admixed
655 populations frame the Dnipro-Don region from west and east, providing another line of evidence
656 for the formation of the Yamnaya within this region. Y chromosome haplogroup sharing—which
657 traces the entirely male line and is of particular interest in societies that have patrilineal
658 traditions—(Fig. 3) is less informative for tracing the origins of the Core Yamnaya, but proves
659 continuity of the Don Yamnaya with their Serednii Stih ancestors. Haplogroup I-L699 was an
660 important lineage in the Dnipro area since the Neolithic hunter-gatherer period, continued to be
661 prevalent among the Serdenii Stih, and in the Don Yamnaya was dominant (17/20 instances).
662 The Core Yamnaya belonged primarily to haplogroup R-M269 (49/51 instances) most of which
663 could be determined as belonging to the Z2103 sub-lineage (41/51). This lineage is
664 unprecedented in our sampling of the steppe before the Yamnaya period; its closest relative is the
665 L51 lineage which dominated the Beaker group[3] and mainland Europe outside the steppe (Fig.
666 3), with a slightly more distant relative in the R-PF7563 lineage found in Pylos in Mycenaean
667 Greece.[45] With an estimated time of formation of ~4450 BCE (https://www.yfull.com/tree/R-
668 L23/; v11.04.00), the R-L23 lineage unifies Beaker, Yamnaya, and Mycenaean Y-chromosomes
669 within an Eneolithic timeframe, which is consistent with the ancestors of these three groups
670 being part of a single population in the Yamnaya period itself since population divergences are
671 always lower than the genetic divergences of specific haplotypes. It is a challenge for future
672 ancient DNA studies to find the population in which the Eneolithic R-L23 founder lived and to
673 trace his R-Z2103 descendants. Their absence from the Eneolithic record, together with the
674 evidence (discussed below) for isolation in the formative period of the Yamnaya suggest that he
675 might have been part of a small group not yet sampled.
676
677 That the Core Yamnaya are part of the Dnipro cline may suggest an origin in the Dnipro basin
678 itself, but (a) the Dnipro cline is generated by admixture with Dnipro-Don people (UNHG/GK2-
679 related), and (b) the Yamnaya on the Don are also part of this cline, so an alternative origin in the
680 Don area cannot be excluded. An origin of the Core Yamnaya further east, in the Caucasus-
681 Volga region is unlikely given that they are not part of the Volga or CLV Clines. Conversely,
682 placing Yamnaya origins west of the Dnipro is implausible as the Core Yamnaya are the
683 population of the Dnipro Cline that is maximally derived from the eastern CLV Cline and they
684 also do not have the European farmer-derived ancestry of western populations such as the
685 Usatove (Fig. 1b).[15] The Core Yamnaya share ancestry with people of the whole Dnipro-Don-
686 Volga-Caucasus region, but their ancestral mix includes all components also found in the
687 Serednii Stih, while these are lacking elsewhere (Extended Data Fig. 3). A more western origin
688 of the Core Yamnaya would also bring their latest ancestors in proximity to the place of origin of
689 the Corded Ware complex whose origin is itself in question but must have certainly been in the
690 area of central-eastern Europe occupied by the Globular Amphora culture west of the Core
691 Yamnaya. The Corded Ware population, which could trace a large part of its ancestry to the
692 Yamnaya,[2] was formed by admixture concurrent with the Yamnaya expansion[52] (Extended Data
693 Fig. 2d), shared segments of IBD proving connections within a shallow genealogical timeframe,
694 and had a balance of ancestral components from the Caucasus and eastern Europe
695 indistinguishable from the Yamnaya.[6] In combination, these lines of evidence suggests that it

696 was formed indeed by early 3rd millennium BCE admixture with Yamnaya, or, at the very least,
697 genetically Yamnaya ancestors that need not have been Yamnaya in the archaeological sense.
698 The geographical homelands of the Corded Ware and Yamnaya would then conceivably be in
699 geographical proximity to allow for their synchronous emergence and shared ancestry. The
700 Dnipro-Don area of the Serednii Stih culture fits the genetic data, as it explains the ancestry of
701 the nascent Core Yamnaya and places them in precisely the area from which both Corded Ware,
702 and Southeastern European Yamnaya (in the west) and the Don Yamnaya (in the east) could
703 have emerged by admixture of the Core Yamnaya with European farmers and UNHG
704 respectively.

705

706 *Figure 4: Population structure in people with a Yamnaya cultural affiliation. Individuals are*
707 *projected in the same space as in Fig. 1. (a) showing that the Core Yamnaya cluster (black*
708 *symbols) from diverse sites is differentiated from the Don Yamnaya (blue) who tend towards the*
709 *UNHG. (b) Yamnaya individuals in the West (Ukraine, Hungary, Slovakia, and Southeastern*
710 *Europe) include a tight cluster of individuals as well as others that tend towards the direction of*
711 *European Neolithic and Chalcolithic groups from Romania and Hungary. Individuals from*
712 *Russia are shown in grey circles in panel (b).*
713

714
715
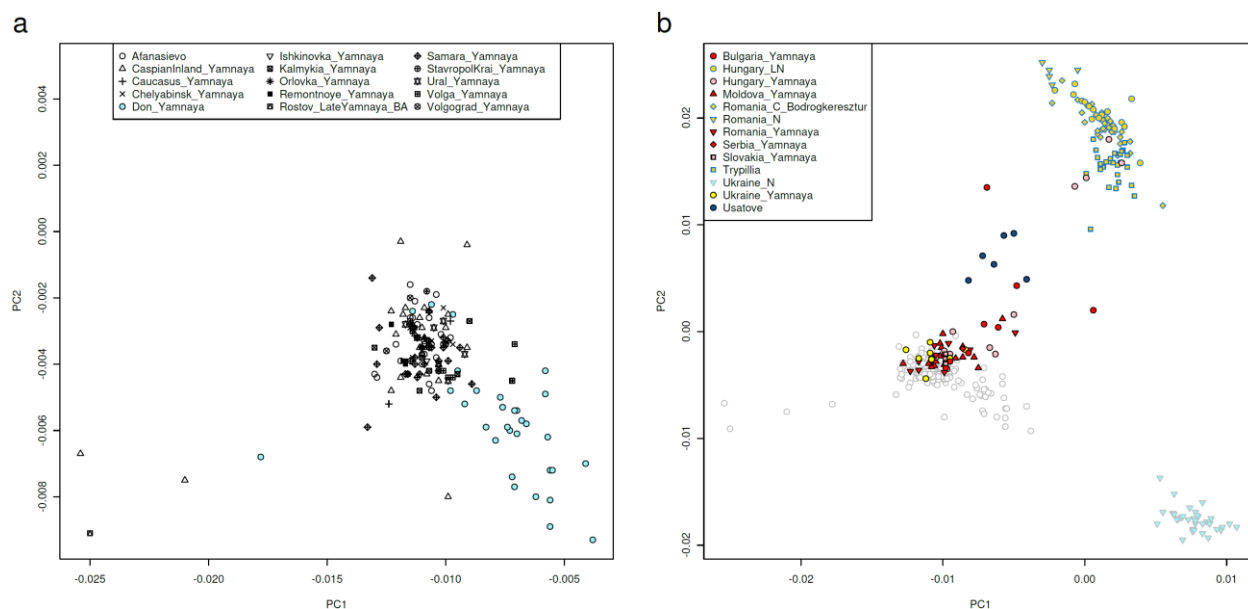
From Serednii Stih to Yamnaya: the 4th millennium BCE
717 We estimated the population growth trajectory of Core Yamnaya using HapNe-LD, a
718 methodology that can infer effective population size fluctuations in low-coverage ancient DNA
719 data.[53] Figure 5 shows the results separately analyzed for Core Yamnaya dating to the first three
720 hundred years of our sampling (n=25) who produce a 95% confidence interval of 3829-3374
721 BCE for the time before growth, and 3642-3145 BCE for Core Yamnaya groups from the later
722 three hundred years (n=26). In both cases, these correspond to growth from an effective number
723 of reproducing individuals of a few thousand people. These intervals overlap at 3642-3374 BCE,
724 corresponding to the late Serednii Stih period. Taken together with the admixture dating, these
725 findings point to a scenario where the Serednii Stih were largely formed by admixture before
726 4000 BCE likely somewhere within the geographic span of the Dnipro-Don Cline. Half a
727 millennium later, a subgroup of them developed cultural innovations that allowed them to

24

728 expand dramatically, manifesting in a way that can be detected in the archaeological record
729 around 3300 BCE in both the Pontic and Caspian Steppes.
730
731 ***Figure 5: Trajectory of the Yamnaya expansion.*** *We use HapNe-LD to estimate the changes in*
732 *effective population size over time of Yamnaya ancestors, performing the computation separately*
733 *for the individuals from the earlier three hundred years (a) of our sampling, and the later three*
734 *hundred years (b); shading shows uncertainty intervals. We infer an extraordinary population*
735 *expansion (c) after 3642-3374 BCE (intersection of 95% confidence intervals for the two*
736 *analyses for the minimum), from a time when the effective size is a few thousand to an order of*
737 *magnitude larger.*
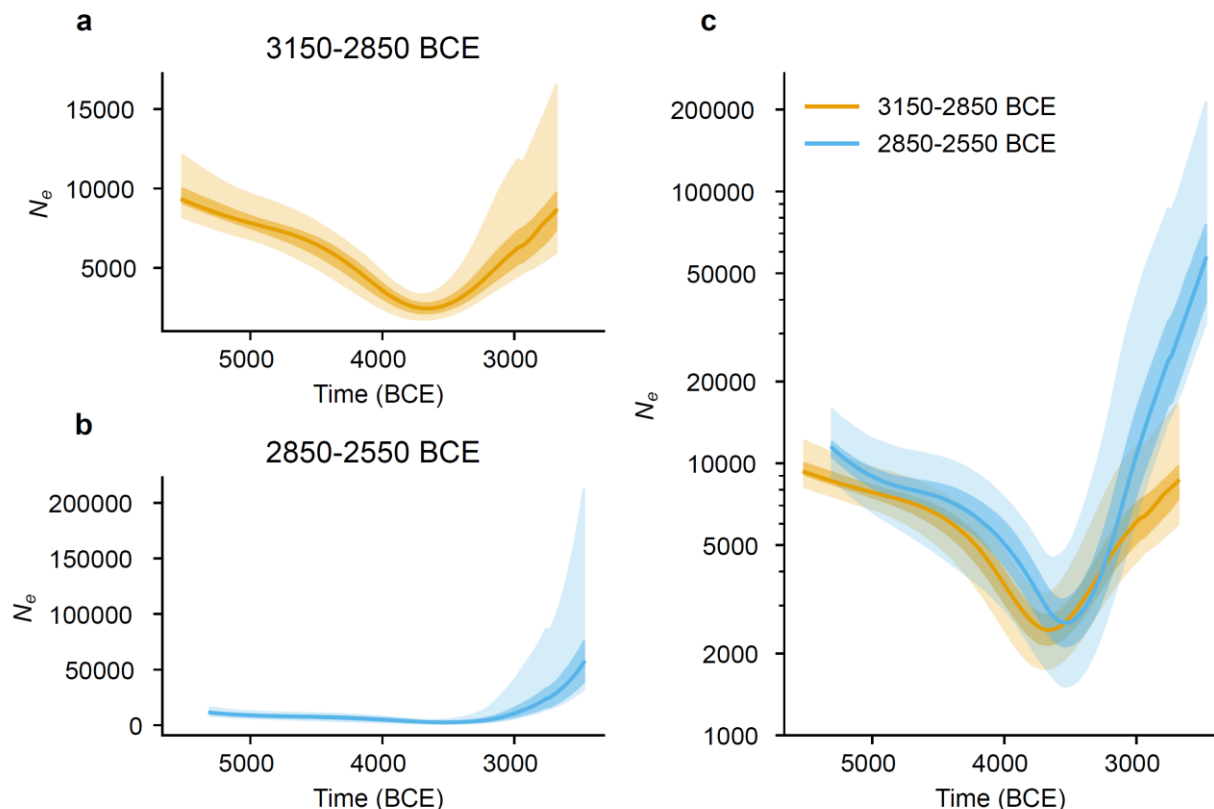
738
739
740
741 We tested for segments of the genome Identical-By-Descent (IBD) between pairs of
742 individuals[54], and found that the Yamnaya expansion transformed the interconnectedness of
743 steppe populations. Before the Yamnaya, IBD links of ≥20cM did exist between regional
744 populations (Fig. 6a), but this network of connections expanded dramatically in the Yamnaya
745 period (Fig. 6b). Prior to the Yamnaya period, the rate of IBD links for individuals separated by
746 more than 500km was vanishingly low (Fig. 6c), but in Yamnaya times, it was measurably non-
747 zero (at a few percent) for distance separations between 500-5000km (Fig. 6d). We also studied
748 close genetic relatives, defined as sharing at least three ≥20cM segments or a total sum of IBD
749 ≥100cM. Both before and during the Yamnaya period, close relatives are only detected living
750 within 500km, with a greatly elevated rate in the same cemetery (Fig. 6e, f). We examined
751 Yamnaya-Afanasievo individuals in kurgans or kurgan cemeteries represented by at least two
752 individuals (Fig. 6g), and found that around 14.4% of individual pairs were close relatives within

25

753   kurgans and 7.4% of individual pairs were close relatives across kurgans of the same cemetery.
754   These patterns are general across Yamnaya kurgan cemeteries (they are not dominated by one or
755   a few sites with large numbers of samples). The observed rate of close relatives is much less than
756   the 29.0% rate among pairs of individuals in Hazleton North chambered tomb in Neolithic
757   Britain ~3700BCE[55] (p=0.00075; Fisher's exact test), where 27 of 35 sequenced individuals were
758   all found to be part of the same genetically tightly connected pedigree. These findings disprove
759   theories that kurgans were "family tombs"[56] of biological relatives. Instead, kurgan cemeteries
760   largely included individuals that were biological kin only in the sense of sharing common
761   descent for a population that lived many centuries in the past; if there were kinship links within
762   the same kurgan, they were largely non-biological ones.

763

764   *Figure 6:* **IBD analysis of the Yamnaya and their predecessors.** Pairs of individuals linked by
765   at least one IBD segment ≥20cM in length reveal a sparse and highly connected network in the
766   Pre-Yamnaya (a) and Yamnaya (b) groups. No detectable IBD is found in the Pre-Yamnaya
767   period beyond the scale of 1000km (c); Yamnaya share more IBD with each other at short
768   distance scales but IBD sharing extends all the way to the ~6000km scale of their geographical
769   distribution. However, closely related individuals only occur at short distance scales in both Pre-
770   Yamnaya (e) and Yamnaya (f) groups, indicating that the IBD sharing in the Yamnaya was a
771   legacy of their common origin. (g) In a set of 9 Yamnaya cemeteries, and a total of 25 kurgans
772   closely or distantly related individuals are virtually absent in inter-cemetery comparisons, more
773   are found in inter-kurgan/within-cemetery comparisons, and more still in intra-kurgan
774   comparisons; nonetheless, most Yamnaya individuals in all comparisons were unrelated. Kurgan
775   burial of close kin was less common than in the case of a local patrilineal dynasty as at a
776   Neolithic long cairn at Neolithic Hazleton North,[55] but more common than in Neolithic
777   monuments of Neolithic Ireland.[57]

778

779
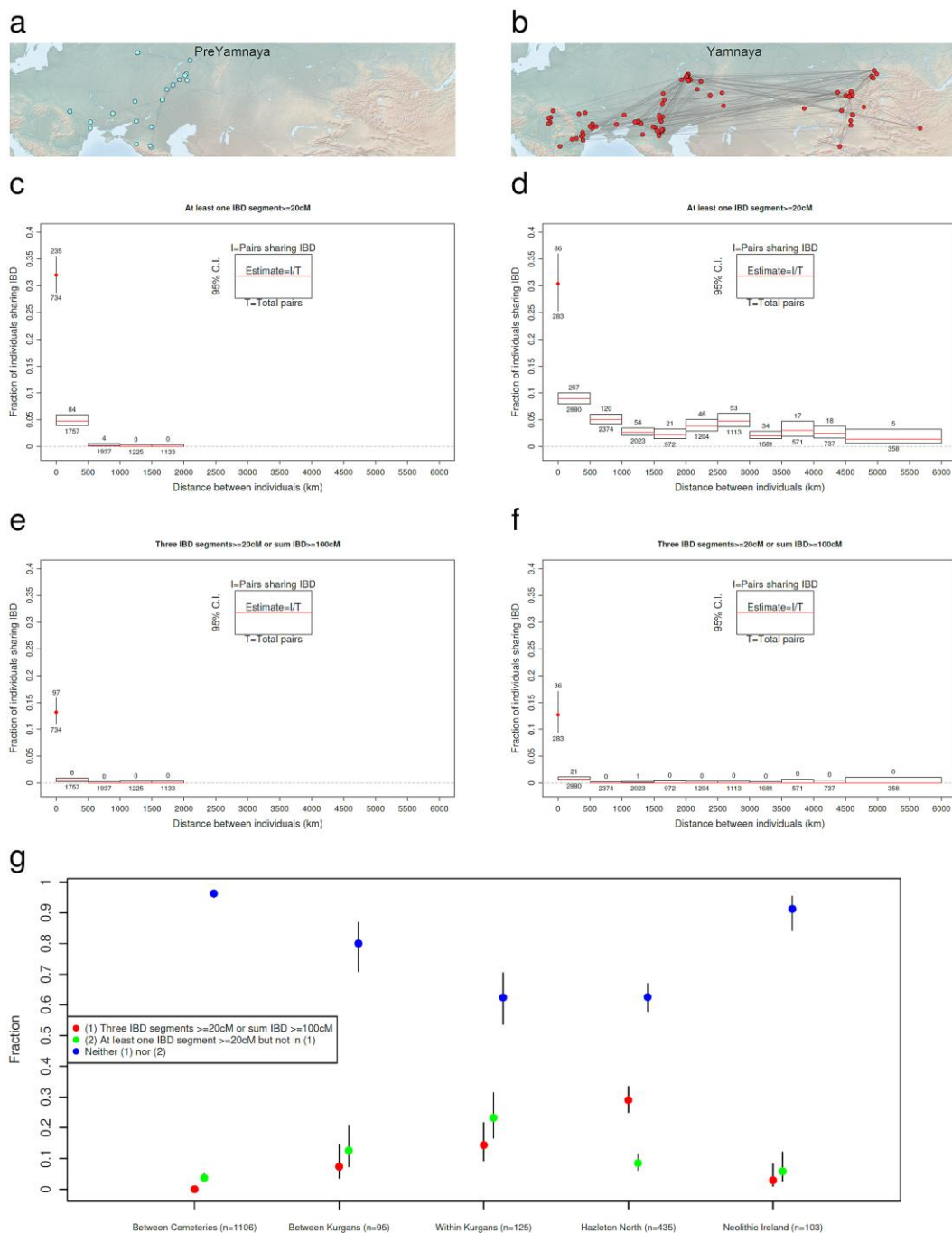
The origin and spread of the first speakers of Indo-Anatolian languages

Different terminologies exist to designate the linguistic relationship of Anatolian and Indo-European languages. The traditional view includes both within an "Indo-European" (IE) group in which Anatolian languages usually represent the first split[58,59]. An alternative terminology, which we use here, names the entire linguistic group "Indo-Anatolian" (IA) and uses IE to refer to the set of related non-Anatolian languages such as Tocharian, Greek, Celtic, and Sanskrit.[6,49] Dates between 4300-3500 BCE have been proposed for the time of IA split[49,59-61] predating both the first attestation of the Hittite language in Central Anatolia (post-2000 BCE[49]) and the expansion of the Yamnaya archaeological culture (post-3300 BCE). We identify the Yamnaya

791 population as Proto-IE for several reasons. First, the Yamnaya were formed by admixture ~4000
792 BCE and began their expansion during the middle of the 4th millennium BCE, corresponding to
793 this linguistic split date between IE and Anatolian. Second, the Yamnaya were the source of the
794 Afanasievo migration to the east[62] a leading candidate for the split of the ancestral form of
795 Tocharian, widely recognized as the second split after that of Anatolian.[63] Third, the Yamnaya
796 can be linked to the languages of Armenia[45] via both autosomal and Y-chromosome ancestry
797 after ~2500 BCE, and to the languages of the Balkans[13] such as Greek.[45,47] Fourth, the Yamnaya
798 can be linked indirectly to other IE speakers via the demographically and culturally
799 transformative Corded Ware and Beaker archaeological cultures of the 3rd millennium BCE that
800 postdate it by centuries. Most people of the Corded Ware culture of central-northern Europe had
801 about three quarters of Yamnaya ancestry,[2] a close connection within a few generations that can
802 be traced to the late 4th millennium BCE. The Beaker archaeological culture of central-western
803 Europe also shared a substantial amount of autosomal ancestry with the Yamnaya and were also
804 linked to them by their possession of R-M269 Y-chromosomes.[3] The impact of these derivative
805 cultures in Europe leaves no doubt that they were linguistically Indo-European as most later
806 Europeans were; the Corded Ware culture itself can also be tentatively linked via both autosomal
807 ancestry and R-M417 Y-chromosomes with Indo-Iranian speakers via a long migratory route that
808 included Fatyanovo[20] and Sintashta[4,22] intermediaries. A recent study proposed a much deeper
809 origin of IA/IE languages[64] to ~6000 BCE or about two millennia older than our reconstruction
810 and the consensus of other linguistic studies. The technical reasons for these older dates will
811 doubtlessly be debated by linguists. From the point of view of archaeogenetics, we point out that
812 the post-3000 BCE genetic transformation of Europe by Corded Ware and Beaker cultures on the
813 heels of the Yamnaya expansion is hard to reconcile with linguistic split times of European
814 languages consistently >4000 BCE as no major pan-European archaeological or migratory
815 phenomena that are tied to the postulated South Caucasus IA homeland ~6000 BCE can be
816 discerned.
817
818 The Yamnaya culture stands as the unifying factor of all attested Indo-European languages. Yet,
819 the homogeneity of the Yamnaya patrilineal community was formed out of the admixture of
820 diverse ancestors, via proximal ancestors from the Dnipro and CLV clines (Fig. 2e). Yamnaya
821 and Anatolians share ancestry from the CLV Cline (Fig. 2e,f), and thus, if the earliest IA
822 language speakers shared any genetic ancestry at all—the possibility of an early transfer of
823 language without admixture must not be discounted—then the CLV Cline is where this ancestry
824 must have come from. On the Anatolian side, we see that ancestry from the southern Caucasus
825 Neolithic end of the CLV Cline was impactful during the Chalcolithic and Bronze Ages[45] and
826 Bronze Age Central Anatolians over the time span of Hittite presence there also had traces of
827 Lower Volga-related ancestry which implies an origin north of the Caucasus (Fig. 2f; Extended
828 Data Fig. 1). On the steppe side, we see that mixed Lower Volga/Caucasus Neolithic ancestry
829 was present in the Dnipro Cline and maximized in the Yamnaya population along that cline (Fig.
830 2e). IBD analysis identifies long (≥30cM) segments shared by Eneolithic individuals from
831 Berezhnovka-2 in the Lower Volga with Khvalynsk, Igren-8 Serednii Stih, and Areni-1
832 Armenian Chalcolithic populations, providing strong direct evidence for the impact of Lower
833 Volga ancestry on the Middle Volga, Dnipro, and South Caucasus regions, and active gene flow
834 among these regions around the time the sampled individuals lived (Extended Data Table 5). The
835 individual from Vonyucka-1 in the North Caucasus, in fact, has an IBD link (15.2cM) with an
836 early Bronze Age Anatolian from Ovaören. Indo-Anatolian languages must have been spread

837   widely by people carrying CLV cline ancestry (Fig. 2) >4000BCE. However, only two
838   descendant groups transmitted their languages to later groups: the Yamnaya in the Dnipro-Don
839   area, aided by the mobility of their horse-wagon technology, and the Proto-Anatolians in the
840   south, surviving in the diverse linguistic landscape of ancient Western Asia long enough for their
841   languages to be recorded in writing after 2000BCE. Whatever their deeper origins in time out of
842   the diverse constituents of CLV cline populations, the Indo-Anatolians must have been part of
843   that cline. Genetics has little to say whether within this cline the IA languages were first spoken
844   in the Caucasus end of the cline and spread into the steppe along with the spread of Caucasus
845   ancestry, or vice versa, or even if a linguistic unity uncoupled with ancestry existed within the
846   CLV continuum. DNA has traced back the ancestors of both Anatolian and IE speakers to the
847   part of the CLV Cline that was north of the Caucasus mountains, bringing them into proximity
848   with each other and uncovering their common CLV ancestry. However, it cannot adjudicate, on
849   its own, who among the proximate and diverse distal ancestors of the CLV people were Pre-IA
850   speaking. Future studies of the dynamics and temporality of intra-CLV contacts (to which
851   genetics may add its information) and of the cultures of CLV people (as reconstructed by
852   archaeology and linguistics) may decide who among them were most likely to have been the
853   "original" Indo-Anatolians.
854
855   Linguistic evidence has been advanced in favor of different solutions of the Proto-IE origins
856   problem for more than two centuries and we review some recent proposals relevant to our
857   reconstruction of early IA/IE history.
858
859   First, the presence of some cereal terminology in IA languages and even more in IE was
860   suggested to reflect a subsistence strategy that relied in part on agriculture; this was interpreted
861   as providing evidence against a geographic origin of the populations that spread Indo-European
862   languages east of the Dnipro valley, the easternmost point in which agriculture was used (along
863   with foraging and herding) during the Eneolithic.[65] Our genetic findings are consistent with this
864   constraint. If a Caucasus Neolithic population like that at Aknashen spread IA languages to the
865   north (via the CLV cline to the Dnipro-Don area) it would almost certainly have had a cereal
866   vocabulary, and then this vocabulary would have been retained during the Serednii Stih culture
867   of the Eneolithic down to the time of the Yamnaya as agriculture continued to be used there.[65]
868
869   Second, the fact that Anatolian languages are attested largely in western Anatolia has been
870   interpreted as evidence for entry into Anatolia from the west (via the Balkans),[49] and thus we
871   need compelling genetic evidence to provide a strong synthetic case for an eastern route. In fact,
872   however, our genetic data does provide such a strong case, greatly increasing the plausibility of
873   scenarios of an eastern entry of Proto-Anatolian speaking ancestors into Anatolia.[66] This is
874   because we find that Central Anatolian Early Bronze Age people who were plausibly speakers of
875   Anatolian languages based on their archaeological contexts, were striking genetic outliers from
876   their neighbors due to having a minority component of their ancestry from the CLV (plausibly
877   from the people who brought the ancestral form of Anatolian languages to Anatolia), the
878   majority of their ancestry from Mesopotamian Neolithic farmers, and little or no ancestry from
879   the Neolithic and Chalcolithic Anatolians who were overwhelming the source populations of
880   other Early Bronze Age Anatolians. Mesopotamian Neolithic ancestry almost certainly had an
881   eastern geographic distribution, while the Central Anatolian Bronze Age people had no evidence
882   of the European farmer or European hunter-gatherer ancestry that CLV have encountered if they

29

883 had migrated to Anatolia from the west, so the genetic data favor an eastern route. How then
884 could it be that there is no linguistic evidence of Anatolian speakers in eastern Anatolia? We
885 propose that the archaeologically momentous expansion of the Kura-Araxes archaeological
886 culture in the Caucasus and eastern Anatolia after around 3000BCE may have driven a wedge
887 between steppe and West Asian speakers of IA languages, isolating them from each other and
888 perhaps explaining their survival in western Anatolia into recorded history. That the expansion of
889 the Kura-Araxes archaeological culture could have had a profound enough demographic impact
890 to have pushed out Anatolian-speakers, is attested by genetic evidence showing that in Armenia,
891 the spread of the Kura-Araxes culture was accompanied by the complete disappearance of CLV
892 ancestry that had appeared there in the Chalcolithic (Fig. 2f).[9,45,67]
893
894 The Kura-Araxes culture may not be the only reason for the IA split. The ancestors of the
895 Yamnaya did not only become separated from their Anatolian linguistic relatives but from other
896 steppe populations as well. The homogenization of the Yamnaya ancestral population during the
897 4th millennium BCE, both in terms of its autosomal ancestry, and in terms of its Y-chromosome
898 lineage, attest to a period of relative isolation and the cessation of admixture. Such isolation
899 would foster linguistic divergence of the languages spoken in the pre-Yamnaya community with
900 those of their linguistic relatives on the steppe. This isolation must have persisted even after the
901 sudden appearance of the Yamnaya archaeological horizon. Mobility and geographical dispersal
902 provided ample opportunities for the resumption of admixture, yet the genetic homogeneity of
903 the "Core Yamnaya" across much of the steppe leaves little room for the absorption of any pre-
904 existing steppe communities: they all seem to disappear in the face of the Yamnaya juggernaut.
905 Did mixing occur between the segment of the Yamnaya population not buried in kurgans and
906 locals they encountered while the kurgan-buried elite largely avoided it with some exceptions?[15]
907 The rise of the Yamnaya in the Steppe at the expense of their predecessors was followed by their
908 demise after a thousand years (Fig. 3), displaced by descendants of people of the Corded Ware
909 culture. Was this the demise of the kurgan elites of the Yamnaya or of the population as a whole?
910 The steppe was dominated by many and diverse groups later still, such as the Scythians and
911 Sarmatian nomads of the Iron Age. These groups are certainly very diverse genetically, but their
912 kurgans scattered across the steppe attest to the persistence of at least some elements of culture
913 that began in the Caucasus-Volga area seven thousand years ago before blooming, in the Dnipro-
914 Don area, into the Yamnaya culture that first united the steppe and impacted most of Eurasia. To
915 what symbolic purpose did the Yamnaya and their precursors erect these mounds we may not
916 ever fully know. If they aimed to preserve the memory of those buried under them, they did
917 achieve their goal, as the kurgans, dotting the landscape of the Eurasian steppe, drew generations
918 of archaeologists and anthropologists to their study, and enabled the genetic reconstruction of
919 their makers' origins presented here.

Methods

Terminology for archaeological cultures and geographic locations: For archaeological cultures and geographic locations that span more than one modern country, we used the prevalent term in the archaeological and genetic literature, for example using "Yamnaya" which is the common term used in Russia and most of Eastern Europe instead of the Ukrainian "Yamna". For archaeological cultures and locations that are confined to a single country, we generally use the local terminology, for example we refer to the archaeological cultures of "Usatove" and "Trypillia" and "Serednii Stih" and the river "Dnipro" with the Ukrainian terms rather than the corresponding Russian terms "Usatovo", "Tripolye," "Sredni Stog" and "Dniepr".

Sampling ancient individuals: The skeletal remains analyzed here were almost all sampled in ancient DNA clean rooms either at Harvard Medical School, the University of Vienna or the Institute for Archaeogenomics in Budapest. If available and accessible, we prioritized sampling petrous bones, taking bone powder from the cochlea by sandblasting and milling[68], or directly drilling into the cochlea after physical surface cleaning, or drilling through the cranial base to minimize damage to intact skulls[69]. If we could not sample from the cochlea, we sought to sample a tooth, prioritizing the cementum layer after physical surface cleaning[70]. If neither a cochlea nor a tooth was available, we sought to sample a dense cortical bone, which we analyzed by drilling and collecting powder after physical surface cleaning. For some samples that could not leave the museum, we sampled on site, either drilling directly into the cochlea, the tooth root, or bone after physical surface removal. We sometimes dislodged auditory ossicles during sandblasting or drilling into the cochlea. When this happened during the cleaning procedure, we generally stopped the destructive sampling and collected the ossicle(s)[71]. As suggested in the study that recognized the high preservation of DNA in ossicles, we cleaned the ossicle with 10% Bleach and radiated it ultraviolet light for 10 minutes before submerging it in extraction buffer without attempting to produce powder.

Ancient DNA data generation: The samples we studied were processed in our laboratories between 2013 and 2023 and therefore were analyzed with changing protocols. Details and protocols used for each library can be found in Online Table 2. At Harvard Medical School, where the majority of wet laboratory work was done, we initially carried out all DNA extractions and Illumina library preparations manually, using small batches of samples and silica columns for DNA cleanup[72-74]. Since 2018, we used automated liquid handlers (Agilent Bravo Workstations) for both DNA extraction[75] and library preparation with magnetic beads (see supplementary material in [76] for automated double-stranded library preparation, and ref. [77] for automated single-stranded library preparation). We treated DNA extracts with USER (NEB) during library preparation to cut DNA at uracils; this treatment is inefficient at terminal uracils and leaves a damage pattern expected for ancient DNA at the terminal bases that can be filtered out for downstream analysis while allowing a library to be authenticated as old. All libraries were either dual barcoded through double-stranded ligation or dual indexed through indexing PCR at the end of single-stranded library preparation to allow pooling before sequencing.

Before 2015, we screened libraries for mitochondrial DNA before attempting to capture nuclear loci[78]. In the next couple of years, we added an increasing number of nuclear SNPs (between 10 and 4000) as targets into the screening capture since mitochondrial DNA quality does not always

31

966 correlate well with nuclear DNA quality and quantity. We later increased the number of targeted
967 SNPs in our nuclear capture from about 390,000 (390k) [2,79] to about 1.24 million (1240k)[80] for
968 libraries passing the mitochondrial capture with nuclear spike-in. Later, we dropped the
969 screening capture altogether and added the mitochondrial probes to the 1240k probes (1240k+).
970 In 2022, we switched from the 1240k homebrew capture to a kitted capture product available
971 from Twist Biosciences[81].

973 For ancient DNA data generated in the Budapest at the Institute of Archaeogenomics, HUN-REN
974 Research Centre for the Humanities, we followed the protocol described in [82].

976 **Bioinformatic processing:** All ancient DNA libraries were sequenced with paired-end reads.
977 We then performed the following steps: preprocessing, alignment and post-alignment filtering
978 for variant calling. The goal of preprocessing is to take raw sequenced products and create
979 merged sequences for alignment. We demultiplex reads, binning these to whichever library each
980 read belongs to using the identifying barcodes and indices, trim these identifying markers as well
981 as any residual adapter sequences, and merge each paired-end read into a single molecule using
982 the overlap of the paired-end reads as a guide, employing a modified version of *SeqPrep*
983 (https://github.com/jstjohn/SeqPrep). The resulting single-ended reads are aligned to both the
984 *hg19* human genome reference (https://www.internationalgenome.org/category/grch37/) and the
985 inferred ancestral Reconstructed Sapiens Reference Sequence (RSRS) mitochondrial sequence[83]
986 using the *samse* aligner of *bwa*[84]. Duplicate molecules are marked by barcode bin, based on the
987 same start/stop positions and orientation. The computational pipelines with specific parameters
988 used are publicly available on GitHub at https://github.com/dReichLab/ADNA-Tools and
989 https://github.com/dReichLab/adna-workflow.

991 We used a 'pseudohaploid genotyping' approach to determine a randomly selected allele at SNP
992 sets of interest. To represent the allele at each SNP, we randomly selected sequences from a pool
993 of all sequences covering that position with a minimum data quality; our criteria were a
994 minimum mapping quality of at least 10, and a base quality of at least 20, after trimming
995 sequences by 2 base pairs at both 5' and 3' ends to remove damage artifacts. We assessed
996 ancient DNA authenticity by using *contamMix-1.0.1051*[85] to search for heterogeneity in
997 mitochondrial DNA sequences which are expected to be non-variable in uncontaminated
998 individuals, and also ANGSD to teset for heterogeneity in X chromosome sequences which are
999 expected to be homozygous in male individuals.[86] We also evaluated authenticity of the ancient
1000 samples by using *pmdtools*[87] to measure the rate of cytosine-to-thymine mutations in the first and
1001 last nucleotides (in untrimmed sequences) which is expected for genuine ancient DNA[73], and by
1002 computing the ratio of Y chromosome to the sum of X and Y chromosome sequences which is
1003 expected to be very low for females and to have a very much higher value for males. We
1004 determined a consensus for mitochondrial DNA using *bcftools*
1005 (https://github.com/samtools/bcftools) and *SAMTools*[88] requiring a minimum of 2-fold coverage
1006 to call the nucleotide and a majority rule to determine its value. We used *HaploGrep2* to
1007 determine mitochondrial haplogroups based on the phylotree database (mtDNA tree build
1008 17).[89,90]

1010 **Principal Components Analysis:** Individuals in Fig. 1b are projected analysis in *smartpca*[37] using
1011 parameters newshrink: YES and lsqporject: YES:

1012 on a PCA space whose axes are formed by the following set of populations: OberkasselCluster
1013 (set of trans-Alpine WHG individuals identified in[19]), Russia_Firsovo_N, Iran_HajjiFiruz_C[4],
1014 Iran_C_SehGabi[9], Iran_C_TepeHissar[91], Israel_C[92], Germany_EN_LBK[2,17,82,93]
1015
1016 **F_{ST} estimation:** F_{ST} was computed in *smartpca*[37] with parameters inbreed: YES and fstonly:
1017 YES.[94]
1018
1019 **Visualizing the three Eneolithic Clines:** Three models are fitted for Eneolithic cline
1020 populations using qpAdm[2] and with OldAfrica, Russia_AfontovaGora3, CHG,
1021 Iran_GanjDareh_N, Italy_Villabruna, Russia_Sidelkino.SG, Turkey_N set of Right populations
1022 (Fig. 1c).
1023
1024 **Model competition with qpAdm/qpWave:** We use qpWave/qpAdm methods[2,30] on diverse
1025 target and source populations from the steppe and adjacent areas (Supplementary Information
1026 section 2). We use OldAfrica, Russia_AfontovaGora3, CHG, Iran_GanjDareh_N,
1027 Italy_Villabruna, Russia_Sidelkino.SG, Turkey_N as the set of Right populations for most
1028 analyses. For the analysis of Anatolian populations, we expanded this set to OldAfrica, CHG,
1029 Iran_GanjDareh_N, Italy_Villabruna, Russia_AfontovaGora3, Russia_Sidelkino.SG,
1030 TUR_Marmara_Barcın_N, TUR_C_Boncuklu_PPN, TUR_C_Çatalhöyük_N, Natufian to gain
1031 leverage for differentiating between different West Asian sources. For faster computation, we ran
1032 qpWave/qpAdm on precomputed output from qpfstats runs
1033 (https://github.com/DReichLab/AdmixTools/blob/master/qpfs.pdf) with poplistname that
1034 includes Han.DG, and all target, source, and Right populations, and parameters allsnps: YES,
1035 inbreed: NO. Separate qpWave/qpAdm runs directly on genotype files were performed as needed
1036 when the target or source populations were not present in the qpfstats output with parameter
1037 basepop: Han.DG. Feasible models are identified as having p>0.05, all standard errors ≤0.1, and
1038 admixture proportions within ≤2 standard errors from 0 and 1. Target or source populations are
1039 removed from the Right set. Competition of models A and B involves two qpWave/qpAdm runs
1040 in which all sources of A \ B and B \ A (\ denotes set difference) are placed on the Right set.
1041 Details of all analyses can be found in Supplementary Information section 2.
1042
1043 **Y-chromosome haplogroup inference:** We used the methodology described in ref. [6] which used
1044 the YFull YTree v. 8.09 phylogeny
1045 (https://github.com/YFullTeam/YTree/blob/master/ytree/tree_8.09.0.json) to denote Y-
1046 chromosome haplogroups in terminal notation.[95]
1047
1048 **Estimates of dates of admixture:** We used DATES[4,52] to estimate a date of admixture for the
1049 Core Yamnaya, Don Yamnaya, Eastern European Yamnaya, Corded Ware, and Caucasus-
1050 Anatolian populations (Extended Data Fig. 2). For the Core Yamnaya and Caucasus-Anatolian
1051 populations, we used sets of diverse West Asian and European hunter-gatherer populations as the
1052 two sources. For the Don Yamnaya we used the Core Yamnaya and UNHG as the two sources.
1053 For the Eastern European Yamnaya we used the Core Yamnaya and a diverse set of
1054 Neolithic/Chalcolithic "European farmers" from Fig. 4b. For the Corded Ware we used the Core
1055 Yamnaya and Globular Amphora as the two sources. It is more important to use many source
1056 samples even if they are not identical to the true ones; picking the wrong sources does not bias
1057 the date estimate[52].

1058

1059 **Identity-by-Descent (IBD) segment detection:** We used ancIBD[54] to detect IBD segments of
1060 length ≥8cM.

1061

1062 **Geographical distance estimation:** To study the decay of IBD with geographical distance, we
1063 estimate distance between sites based on their latitude and longitude (Online Table 2) using the
1064 Haversine distance as implemented in distHaversine[96] of the package *geosphere* in R.

1065

1066 **Estimates of effective population sizes:** We ran HapNe-LD (version 1.20230726 [18]) using
1067 default parameters and providing pseudo-haploid genotypes as input. Briefly, HapNe-LD uses a
1068 summary statistic measuring long-range correlations between markers to infer fluctuations in the
1069 effective population size (defined as the inverse of the coalescence rate) through time. We
1070 studied two distinct sets of unrelated individuals all of which had a coverage of at least 0.7x on
1071 the target autosomal SNPs and with a standard deviation on their estimated date smaller than 180
1072 years (~6 generations). The first group consists of 25 Core Yamnaya individuals with estimated
1073 dates ranging between 4500 and 4800 BP. The second group contains 26 Core Yamnaya
1074 individuals ranging from 4800 to 5100 BP.

1075

1076 If no evidence of effective population size fluctuations can be detected in the data, HapNe-LD
1077 produces a flat line. An output containing fluctuations should thus be interpreted as the detection
1078 of changes in historical effective population sizes. Recent admixture between highly
1079 differentiated populations ($F_{st} > 0.1$) might lead to biases in LD-based analyses that induce
1080 fluctuations similar to a population bottleneck. However, HapNe implements a test to flag the
1081 presence of recent structure in the data, which was not detected in both sample sets (approximate
1082 $p \geq 0.1$), suggesting that the observed signal instead reflects variation in the effective population
1083 size of these groups.

1084

1085 In our analyses, the effective population size is defined as the inverse of the instantaneous
1086 coalescence rate. This quantity corresponds to twice the number of breeding individuals in an
1087 idealized population. We note that, in addition to changes in the number of individuals in the
1088 population (census size), several factors, such as changes in population structure, selection, and
1089 cultural practices,[97] can have an influence on the effective population size. These additional
1090 factors may in part be responsible for the effective size fluctuations observed in the Core
1091 Yamnaya.

1092

1093 Approximate confidence intervals were obtained using bootstrap with different chromosome
1094 arms as resampling units. The beginning of the expansion was determined by using the location
1095 of the minimum of each bootstrapped trajectory. We converted the results into years by
1096 assuming 28.6 years per generation for the median minimum location and 25.6 and 31.5 years
1097 per generation for the lower and upper bounds, respectively.[98] We used these values,
1098 corresponding to the estimated number of years per generation for males (31.5) and females
1099 (25.6) to account for uncertainty in the conversion factor.

1100

1101 **Data Access**
1102 Genotype data for individuals included in this study can be obtained from the Harvard Dataverse
1103 repository through the following link (XXX).The DNA sequences reported in this paper are

1104 deposited in the European Nucleotide Archive under the accession number XXX. Other newly
1105 reported data such as radiocarbon dates and archaeological context information are included in
1106 the manuscript and supplementary files.
1107

Acknowledgments

1137

Author Contributions

1139 I.L., N.P., D.A., L.V., and D.R. wrote the manuscript and supplementary materials with input
1140 from all co-authors. A.S.-N., P.F.P., S.M., N.R., R.P. and D.R. supervised different aspects of the
1141 study. I.L. and N.P. carried out the main genetic analyses. I.L., R.F., H.R., I.O., P.F.P. analyzed
1142 genetic data. D.A. and L.V. edited archaeological information. D.A., A.Kh., E.K., N.Sh., S.C.A.,
1143 S.V.A., E.Ba., Z.B., A.B., P.C., A.C., I.C., M.Co., M.Cs, J.D., N.E., S. É, A.Fa., P.F., A.Fr.,
1144 M.F., A.G., T.Ha., T.Hi., P.Je., V.Kh., V.Ki., A.Kit., A.Ko., G.K., N.N., B.B., P.D., P.Ja., E.Ki.,
1145 A.Kiy., J.K., P.K., R.M., M.M., E.Me., V.M., E.Mo., J.M., N.M., M.N., O.N., M.R., M.O-G.,
1146 G.P., S.P., T.S., N.Se., A. Š, M.R.S., I.S., V.Sh., A.Si., K.S., K.N.S., J.T., A.A.T., V.T., S.V.,
1147 O.F., A.Sz-N., and R.P. sampled anthropological remains and/or contributed to the creation of
1148 the archaeological supplement. A.A., E.S.B., M.Ma., A.Mi., and S.M. carried out bioinformatic

1149 data processing. K.C., F.C., O.Ch., E.C., L.I., A.Ke., D.K., A.M.L., M.Mi., J.O., L.Q., J.N.W.,
1150 F.Z., N.R., and carried out wet laboratory work.
1151
1152 **Conflict of Interest Statement**
1153 The authors declare no competing interests.
1154
1155 **Ethics Statement**
1156 The individuals studied here were all analyzed with the goal of minimizing damage to their
1157 skeletal remains, with permission from local authorities in each location from which they came.
1158 Every sample is represented by stewards such as archaeologists or museum curators, who are
1159 either authors or thanked in the Acknowledgments. Open science principles require making all
1160 data used to support the conclusions of a study maximally available, and we support these
1161 principles here by making fully publicly available not only the digital copies of molecules (the
1162 uploaded sequences) but also the molecular copies (the ancient DNA libraries themselves, which
1163 constitute molecular data storage). Those researchers who wish to carry out deeper sequencing of
1164 libraries published in this study should make a request to corresponding author D.R. We commit
1165 to granting reasonable requests as long as the libraries remain preserved in our laboratories, with
1166 no requirement that we be included as collaborators or co-authors on any resulting publications.

Extended Data Table 1: F_{ST} values among select populations of the Dnipro, Don, Volga, and Caucasus areas. F_{ST} values are shown below the diagonal and their standard errors above it.

	BPgroup	CoreYamnaya	Ekaterinovka	GK1	Khi	KhlopkovBugor	Klo	Kmed	Labazy	Maikop	Maximovka	Murzikha	PVgroup	Remontnoye	Russia_Caucasus_LateMaikop	Russia_Don_EBA_Yamnaya	SShi	SSmed	Syezzheye	Ukraine_N	Unakozovskaya	UpperVolga
BPgroup		0.001	0.001	0.003	0.001	0.002	0.001	0.001	0.002	0.001	0.002	0.001	0.002	0.002	0.005	0.001	0.001	0.001	0.003	0.001	0.003	0.001
CoreYamnaya	0.011		0.000	0.003	0.000	0.002	0.000	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.004	0.000	0.001	0.001	0.002	0.001	0.002	0.000
Ekaterinovka	0.030	0.032		0.003	0.001	0.002	0.000	0.001	0.002	0.001	0.002	0.001	0.002	0.002	0.004	0.000	0.001	0.001	0.003	0.001	0.003	0.000
GK1	0.042	0.041	0.045		0.003	0.007	0.003	0.003	0.005	0.004	0.006	0.003	0.005	0.005	0.018	0.003	0.004	0.005	0.009	0.003	0.006	0.003
Khi	0.007	0.014	0.019	0.039		0.002	0.001	0.001	0.002	0.001	0.002	0.001	0.002	0.002	0.004	0.001	0.001	0.001	0.003	0.001	0.003	0.001
KhlopkovBugor	0.010	0.017	0.022	0.037	0.008		0.002	0.002	0.003	0.003	0.003	0.002	0.003	0.003	0.009	0.002	0.003	0.003	0.005	0.002	0.004	0.002
Klo	0.018	0.022	0.008	0.041	0.009	0.013		0.001	0.002	0.001	0.002	0.001	0.002	0.002	0.004	0.001	0.001	0.001	0.003	0.001	0.003	0.001
Kmed	0.014	0.018	0.015	0.042	0.006	-0.002	0.002		0.002	0.001	0.002	0.001	0.002	0.002	0.005	0.001	0.001	0.001	0.003	0.001	0.003	0.001
Labazy	0.032	0.034	0.009	0.048	0.021	0.027	0.010	0.016		0.002	0.003	0.002	0.003	0.003	0.007	0.002	0.002	0.003	0.004	0.002	0.004	0.002
Maikop	0.031	0.025	0.064	0.064	0.037	0.043	0.052	0.045	0.067		0.003	0.001	0.002	0.002	0.006	0.001	0.002	0.002	0.003	0.001	0.003	0.001
Maximovka	0.044	0.041	0.021	0.048	0.033	0.033	0.021	0.028	0.021	0.076		0.002	0.003	0.003	0.007	0.002	0.003	0.003	0.004	0.002	0.003	0.002
Murzikha	0.056	0.053	0.034	0.065	0.044	0.047	0.034	0.039	0.034	0.088	0.018		0.002	0.002	0.004	0.001	0.001	0.001	0.003	0.001	0.003	0.001
PVgroup	-0.002	0.012	0.035	0.046	0.010	0.012	0.024	0.018	0.038	0.025	0.048	0.061		0.003	0.006	0.002	0.002	0.003	0.004	0.002	0.003	0.002
Remontnoye	0.012	0.011	0.040	0.041	0.015	0.020	0.028	0.024	0.046	0.012	0.052	0.065	0.011		0.006	0.002	0.002	0.002	0.004	0.002	0.003	0.002
Russia_Caucasus_LateMaikop	0.025	0.020	0.058	0.065	0.033	0.037	0.048	0.041	0.059	-0.001	0.063	0.081	0.026	0.002		0.004	0.006	0.007	0.011	0.004	0.007	0.004
Russia_Don_EBA_Yamnaya	0.014	0.005	0.029	0.040	0.014	0.019	0.019	0.018	0.030	0.030	0.037	0.048	0.016	0.015	0.025		0.001	0.001	0.003	0.001	0.003	0.001
SShi	0.009	0.011	0.027	0.034	0.013	0.014	0.017	0.017	0.029	0.030	0.036	0.048	0.010	0.016	0.034	0.011		0.002	0.004	0.001	0.003	0.001
SSmed	0.011	0.010	0.021	0.034	0.011	0.012	0.015	0.015	0.021	0.030	0.030	0.041	0.013	0.014	0.019	0.008	0.004		0.004	0.001	0.003	0.001
Syezzheye	0.045	0.047	0.022	0.059	0.034	0.035	0.026	0.033	0.029	0.082	0.043	0.050	0.049	0.056	0.077	0.042	0.040	0.034		0.003	0.004	0.003
Ukraine_N	0.046	0.039	0.036	0.047	0.040	0.042	0.032	0.037	0.036	0.063	0.038	0.048	0.049	0.055	0.029	0.031	0.017	0.055			0.003	0.001
Unakozovskaya	0.059	0.057	0.094	0.090	0.068	0.069	0.083	0.076	0.096	0.034	0.107	0.117	0.058	0.039	0.030	0.060	0.062	0.061	0.107	0.092		0.003
UpperVolga	0.044	0.040	0.021	0.048	0.033	0.035	0.019	0.028	0.019	0.073	0.015	0.027	0.049	0.051	0.067	0.033	0.035	0.026	0.038	0.029	0.103	

Extended Data Table 2: Extraordinary Genetic Homogeneity in the Core Yamnaya. We tested all populations and individuals for cladality with Samara Yamnaya. We list populations for which this is not rejected (p>0.05) and populations that include individuals that fit Core Yamnaya selection criteria (p>0.2, at least 300k SNPs, and Yamnaya or Afanasievo culture).

Population	P-value	Included in Core Yamnaya	Total individuals
Populations that are a clade with Samara Yamnaya			
China_Xinjiang_G218_BA_Afanasievo_oWestEurasia	9.7E-01	1	1
Russia_Chelyabinsk_EBA_Yamnaya	9.5E-01	5	5
Russia_Volgograd_EBA_Yamnaya	9.0E-01	3	5
Russia_Ural_EBA_Yamnaya_contam	8.2E-01	0	1
Usatove_EBA	7.9E-01	0	1
Russia_Ural_EBA_Yamnaya	7.3E-01	5	7
Russia_Afanasievo_Yenisei	6.7E-01	1	1
Russia_MBA_Poltavka	6.5E-01	0	6
Romania_EBA_Catacomb	6.3E-01	0	2
Russia_Orlovka_EBA_Yamnaya	5.5E-01	1	1
Ukraine_MBA	5.1E-01	0	1
Russia_Samara_EBA_Yamnaya_possible	5.0E-01	0	1
Kazakhstan_EBA_Yamnaya.SG	4.8E-01	1	1
Ukraine_EBA_Yamnaya_contam	4.4E-01	0	1
Russia_LowerVolga_EBA_Yamnaya	3.9E-01	0	1
Moldova_Crasnoe_Eneolithic	3.9E-01	0	1
Russia_EBA_o1.SG	3.6E-01	0	1
Ukraine_EBA_Catacomb	3.5E-01	0	2
Ukraine_MBA_Catacomb_o1	3.4E-01	0	1
Moldova_Eneolithic	3.3E-01	0	1
Russia_BA_WestManych_Catacomb	3.2E-01	0	1
Mongolia_Chalcolithic_Afanasievo_1	2.9E-01	0	1
Russia_Kalmykia_EBA	2.8E-01	0	1
Russia_Afanasievo.SG	2.8E-01	0	2
Russia_UpperOb_Eneolithic_Afanasievo	2.8E-01	6	6
Russia_Volgograd_EBA_Yamnaya_o	2.8E-01	0	1
Russia_Ishkinovka_EBA_Yamnaya	2.5E-01	1	1
Usatove_Yamnaya	2.1E-01	0	1
Latvia_LN_CordedWare	1.7E-01	0	1
Hungary_EBA_Yamnaya_1drel.I3510_contam	1.7E-01	0	1
Brillenhohle.pmd	1.6E-01	0	1
Russia_Steppe_Catacomb	1.6E-01	0	4
Russia_Volga_EBA_Yamnaya	1.3E-01	4	5
Russia_Kalmykia_EasternManych_EMBA	1.3E-01	0	2
Russia_N_BA_possible	1.2E-01	0	1
Russia_Afanasievo	1.2E-01	18	29
Moldova_Eneolithic_Suvorove	1.1E-01	0	1
Russia_Afanasievo_Khakassia_possible	9.0E-02	0	1
BOY_EBA	8.8E-02	0	5
Russia_Rostov_Steppe_NorthCaucasus_BA	7.4E-02	0	1
Russia_LowerDon_EBA_Yamnaya	6.6E-02	0	1
Moldova_EBA_Yamnaya	6.5E-02	4	16
Ukraine_EBA_Catacomb.SG	6.2E-02	0	1
Russia_Afanasievo_contam	5.8E-02	0	2
Romania_Brailita_EBA_Yamnaya	5.8E-02	0	1
Slovakia_EBA_Yamnaya	5.1E-02	0	2
Ukraine_EBA_Yamnaya	5.1E-02	4	9
Populations that are not a clade with Samara Yamnaya but include at least one individual that is			
Romania_EBA_Yamnaya	3.9E-02	2	8
Russia_Remontnoye_EBA_Yamnaya	3.5E-02	5	6
Russia_Kalmykia_EBA_Yamnaya.SG	1.8E-02	2	6
Russia_Caucasus_EBA_Yamnaya	1.6E-02	1	3
Hungary_EBA_Yamnaya	1.5E-04	1	5
Russia_CaspianInland_EBA_Yamnaya	1.2E-04	12	26
Russia_UpperYenisey_Eneolithic_Afanasievo	2.3E-05	1	4
Russia_Don_EBA_Yamnaya	2.8E-50	2	23

Extended Data Table 3: F_{ST} values among populations that include Core Yamnaya individuals. F_{ST} values are shown below the diagonal and their standard errors above it.

	Hungary_EBA_Yamnaya	Moldova_EBA_Yamnaya	Romania_EBA_Yamnaya	Russia_Afanasievo	Russia_CaspianInland_EBA_Yamnaya	Russia_Caucasus_EBA_Yamnaya	Russia_Chelyabinsk_EBA_Yamnaya	Russia_Don_EBA_Yamnaya	Russia_Kalmykia_EBA_Yamnaya.SG	Russia_Remontnoye_EBA_Yamnaya	Russia_Samara_EBA_Yamnaya	Russia_UpperOb_Eneolithic_Afanasievo	Russia_UpperYenisey_Eneolithic_Afanasievo	Russia_Ural_EBA_Yamnaya	Russia_Volga_EBA_Yamnaya	Russia_Volgograd_EBA_Yamnaya	Ukraine_EBA_Yamnaya
Hungary_EBA_Yamnaya		0.001	0.001	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Moldova_EBA_Yamnaya	0.001		0.001	0.000	0.000	0.001	0.001	0.000	0.001	0.001	0.000	0.001	0.001	0.001	0.001	0.001	0.001
Romania_EBA_Yamnaya	0.001	0.001		0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Russia_Afanasievo	0.006	0.004	0.005		0.000	0.001	0.001	0.000	0.001	0.001	0.000	0.001	0.001	0.001	0.001	0.001	0.001
Russia_CaspianInland_EBA_Yamnaya	0.004	0.003	0.002	0.006		0.001	0.001	0.000	0.001	0.001	0.000	0.001	0.001	0.001	0.001	0.001	0.001
Russia_Caucasus_EBA_Yamnaya	0.001	0.002	0.001	0.003	0.003		0.002	0.001	0.002	0.002	0.001	0.002	0.002	0.002	0.002	0.002	0.002
Russia_Chelyabinsk_EBA_Yamnaya	0.008	0.009	0.009	0.010	0.009	0.009		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Russia_Don_EBA_Yamnaya	0.006	0.005	0.006	0.008	0.006	0.005	0.012		0.001	0.001	0.000	0.001	0.001	0.001	0.001	0.001	0.001
Russia_Kalmykia_EBA_Yamnaya.SG	0.007	0.005	0.004	0.005	0.001	0.004	0.011	0.007		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Russia_Remontnoye_EBA_Yamnaya	0.004	0.004	0.003	0.004	0.000	0.003	0.010	0.006	-0.049		0.001	0.001	0.001	0.001	0.001	0.001	0.001
Russia_Samara_EBA_Yamnaya	0.003	0.002	0.003	0.005	0.003	0.003	0.008	0.005	0.005	0.004		0.001	0.001	0.001	0.001	0.001	0.001
Russia_UpperOb_Eneolithic_Afanasievo	0.006	0.005	0.004	0.002	0.003	0.006	0.010	0.008	0.001	0.003	0.004		0.001	0.001	0.001	0.001	0.001
Russia_UpperYenisey_Eneolithic_Afanasievo	0.011	0.010	0.008	0.009	0.009	0.009	0.015	0.012	0.009	0.007	0.010	0.006		0.001	0.001	0.001	0.001
Russia_Ural_EBA_Yamnaya	0.002	0.002	0.001	0.004	0.003	0.003	0.006	0.005	0.004	0.003	0.001	0.003	0.008		0.001	0.001	0.001
Russia_Volga_EBA_Yamnaya	0.004	0.004	0.003	0.005	0.005	0.005	0.007	0.007	0.008	0.007	0.003	0.007	0.011	0.003		0.001	0.001
Russia_Volgograd_EBA_Yamnaya	0.005	0.003	0.004	0.007	0.005	0.004	0.009	0.007	0.007	0.005	0.004	0.007	0.009	0.003	0.006		0.001
Ukraine_EBA_Yamnaya	0.003	0.001	0.001	0.004	0.002	0.002	0.008	0.005	0.003	0.003	0.002	0.004	0.009	0.001	0.004	0.004	

Extended Data Table 4: *qpAdm* **models that fit non-Core Yamnaya.** We use the following sources to model Yamnaya-related populations other than the Core and Don Yamnaya: CoreYamnaya, Romania_C_Bodrogkeresztur, Romania_N, Serbia_IronGates_Mesolithic, Trypillia, Ukraine_N, Usatove. The Baden individuals from Hungary represent a reburial into a kurgan[82] and are predominantly of European farmer, not Yamnaya, ancestry. The Riltsi individual is shown with Usatove ancestry here and can also be modeled with about half Remontnoye ancestry, as the Usatove have ancestry from the CLV cline.[15]

Modeled group	A	B	P-value	A	B	S.E.
Bulgaria_C_ProtoYamnaya	Serbia_IronGates_Mesolithic	Usatove	0.062	17.5%	82.5%	2.4%
Bulgaria_EBA_Yamnaya	CoreYamnaya	Romania_C_Bodrogkeresztur	0.883	85.4%	14.6%	1.4%
Bulgaria_EBA_Yamnaya	CoreYamnaya	Romania_N	0.682	86.7%	13.3%	1.3%
Bulgaria_EBA_Yamnaya	CoreYamnaya	Trypillia	0.719	82.9%	17.1%	1.7%
Bulgaria_EBA_Yamnaya	CoreYamnaya	Usatove	0.396	66.7%	33.3%	3.3%
Bulgaria_Riltsi_EBA_Yamnaya	Romania_N	Usatove	0.079	25.3%	74.7%	5.8%
Hungary_LateC_EBA_Baden_Yamnaya	CoreYamnaya	Romania_C_Bodrogkeresztur	0.113	3.4%	96.6%	1.5%
Hungary_LateC_EBA_Baden_Yamnaya	Romania_C_Bodrogkeresztur	Serbia_IronGates_Mesolithic	0.309	95.9%	4.1%	1.4%
Hungary_LateC_EBA_Baden_Yamnaya	Romania_C_Bodrogkeresztur	Trypillia	0.114	77.6%	22.4%	9.7%
Hungary_LateC_EBA_Baden_Yamnaya	Romania_C_Bodrogkeresztur	Ukraine_N	0.197	96.2%	3.8%	1.5%
Hungary_LateC_EBA_Baden_Yamnaya	Romania_C_Bodrogkeresztur	Usatove	0.099	94.2%	5.8%	2.5%
Hungary_LateC_EBA_Baden_Yamnaya	Romania_N	Serbia_IronGates_Mesolithic	0.323	87.0%	13.0%	1.7%
Hungary_LateC_EBA_Baden_Yamnaya	Romania_N	Ukraine_N	0.094	86.9%	13.1%	1.7%
Moldova_EBA_Yamnaya	CoreYamnaya	Romania_C_Bodrogkeresztur	0.724	93.8%	6.2%	0.9%
Moldova_EBA_Yamnaya	CoreYamnaya	Romania_N	0.571	94.3%	5.7%	0.8%
Moldova_EBA_Yamnaya	CoreYamnaya	Trypillia	0.675	92.7%	7.3%	1.0%
Moldova_EBA_Yamnaya	CoreYamnaya	Usatove	0.367	86.0%	14.0%	2.2%
Moldova_GlobularAmphora_Yamnaya	CoreYamnaya	Trypillia	0.053	88.7%	11.3%	2.6%
Romania_Brailita_EBA_Yamnaya	CoreYamnaya	Romania_C_Bodrogkeresztur	0.561	91.6%	8.4%	2.4%
Romania_Brailita_EBA_Yamnaya	CoreYamnaya	Romania_N	0.524	92.4%	7.6%	2.1%
Romania_Brailita_EBA_Yamnaya	CoreYamnaya	Trypillia	0.601	90.1%	9.9%	2.8%
Romania_Brailita_EBA_Yamnaya	CoreYamnaya	Usatove	0.509	81.0%	19.0%	5.6%
Romania_EBA_Yamnaya	CoreYamnaya	Romania_N	0.096	95.8%	4.2%	1.0%
Romania_EBA_Yamnaya	CoreYamnaya	Usatove	0.143	89.1%	10.9%	2.5%
Serbia_EBA_Yamnaya	CoreYamnaya	Romania_C_Bodrogkeresztur	0.196	87.3%	12.7%	2.0%
Serbia_EBA_Yamnaya	CoreYamnaya	Romania_N	0.097	88.6%	11.4%	1.9%
Serbia_EBA_Yamnaya	CoreYamnaya	Trypillia	0.200	85.1%	14.9%	2.3%
Ukraine_EBA_Yamnaya	CoreYamnaya	Romania_C_Bodrogkeresztur	0.561	93.3%	6.7%	1.3%
Ukraine_EBA_Yamnaya	CoreYamnaya	Romania_N	0.481	94.0%	6.0%	1.2%
Ukraine_EBA_Yamnaya	CoreYamnaya	Trypillia	0.756	92.0%	8.0%	1.5%
Ukraine_EBA_Yamnaya	CoreYamnaya	Usatove	0.534	84.5%	15.5%	3.1%

Extended Data Table 5: Cross-regional shared Identity-by-Descent (IBD) segments. We list all segments≥12cM shared between individuals from two different regions defined as follows: "Dnipro cline": CoreYamnaya, GK1, GK2, Russia_Don_EBA_Yamnaya, SShi, SSlo, SSmed, Ukraine_N; "Volga cline": Ekaterinovka, Khi, KhlopkovBugor, Klo, Kmed, Labazy, Lebyazhinka_HG, Maximovka, Murzikha, Syezzheye, UpperVolga; "Caucasus-Lower Volga Eneolithic": BPgroup, PVgroup; "CLV-South": Remontnoye, Maikop, Unakozovskaya, Armenia_C, TUR_C_Kalehöyük_MLBA, TUR_C_Ovaören_EBA

Individual 1	Individual 2	Group 1	Group 2	Segment length (cM)
I22201	I1924	BPgroup	SShi	35.8
I22202	I6734	BPgroup	Khi	32.1
I1634	I22199	Armenia_C	BPgroup	31.4
I6300_enhanced	I22202	KhlopkovBugor	BPgroup	22.0
I6406	I22200	Kmed	BPgroup	20.1
PG2004	I11837	BPgroup	Khi	18.4
I6301_enhanced	I22199	KhlopkovBugor	BPgroup	18.2
I6301_enhanced	PG2001	KhlopkovBugor	PVgroup	17.6
I28683	PG2004	Remontnoye	BPgroup	16.6
I10567	I28682	Russia_CaspianInland_EBA_Yamnaya	Remontnoye	16.2
PG2001	I3950	PVgroup	Russia_Afanasievo	15.9
PG2001	I6062	PVgroup	Ekaterinovka	15.9
I22199	I8282	BPgroup	Ekaterinovka	15.8
I22201	I10208	BPgroup	Moldova_EBA_Yamnaya	15.5
I1924	I20188	SShi	Klo	15.4
I32501	I8448	Russia_UpperYenisey_Eneolithic_Afanasievo	Murzikha	15.4
I12637	I8457	Moldova_EBA_Yamnaya	Murzikha	15.4
I32821	I8449	Russia_UpperOb_Eneolithic_Afanasievo	Murzikha	15.4
MA2213_wNonUDG.SG	VJ1001	TUR_C_Ovaören_EBA	PVgroup	15.2
I32501	I8455	Russia_UpperYenisey_Eneolithic_Afanasievo	Murzikha	15.2
I6301_enhanced	I22199	KhlopkovBugor	BPgroup	14.9
I8411_enhanced	I26785	UpperVolga	Russia_Don_EBA_Yamnaya	14.9
I22201	I1924	BPgroup	SShi	14.8
I22199	I28682	BPgroup	Remontnoye	14.8
I0122	I22202	Klo	BPgroup	14.6
I32501	I8454	Russia_UpperYenisey_Eneolithic_Afanasievo	Murzikha	14.5
I22199	I6734	BPgroup	Khi	14.5
I22201	I11752	BPgroup	Russia_Afanasievo	14.3
I6064	I22199	Ekaterinovka	BPgroup	14.2
I0122	I22199	Klo	BPgroup	14.2
I1634	I1924	Armenia_C	SShi	13.9
I6301_enhanced	I22201	KhlopkovBugor	BPgroup	13.9
I6918	I8446	Russia_Volgograd_EBA_Yamnaya	Maximovka	13.9
I22202	I6739	BPgroup	Khi	13.9
PG2004	I23651	BPgroup	Ekaterinovka	13.7
I0357	I11842	Russia_Samara_EBA_Yamnaya	Murzikha	13.7
I22202	I3952	BPgroup	Russia_Afanasievo	13.7
I0122	I20190	Klo	Russia_Samara_EBA_Yamnaya	13.6
I8951	I11842	Russia_Don_EBA_Yamnaya	Murzikha	13.5
PG2004	I8290	BPgroup	Ekaterinovka	13.4
I0231	I8456	Russia_Samara_EBA_Yamnaya	Murzikha	13.4
I25159	I22199	Russia_Afanasievo	BPgroup	13.3
I4111	I6109	Ukraine_N	Klo	13.3
I22199	I26787	BPgroup	Russia_Don_EBA_Yamnaya	13.3
I6301_enhanced	PG2004	KhlopkovBugor	BPgroup	12.9
I8449	I2105	Murzikha	Ukraine_EBA_Yamnaya	12.9
I20189	I22200	Ekaterinovka	BPgroup	12.8
I6297	I22201	Russia_Orlovka_EBA_Yamnaya	BPgroup	12.8
I6705	I28682	Russia_Samara_EBA_Yamnaya	Remontnoye	12.8
I32821	I22200	Russia_UpperOb_Eneolithic_Afanasievo	BPgroup	12.7
I32501	I8449	Russia_UpperYenisey_Eneolithic_Afanasievo	Murzikha	12.6
I22201	I6739	BPgroup	Khi	12.4
I0231	I28682	Russia_Samara_EBA_Yamnaya	Remontnoye	12.3
PG2004	I6739	BPgroup	Khi	12.3
I6918	I22200	Russia_Volgograd_EBA_Yamnaya	BPgroup	12.3
I22201	I3952	BPgroup	Russia_Afanasievo	12.2
I6406	I1450	Kmed	Russia_Samara_EBA_Yamnaya	12.2
I22199	I5273	BPgroup	Russia_Afanasievo	12.1
I4114	I12964	Ukraine_N	UpperVolga	12.1
I11838	I23651	Russia_Volga_EBA_Yamnaya	Ekaterinovka	12.0
I6907	I11841	Russia_Samara_EBA_Yamnaya	Murzikha	12.0
I22201	I1924	BPgroup	SShi	12.0

Extended Data Figure 1: The origin of Central Anatolian Bronze Age. (a) Fitting models include Mesopotamian (Çayönü) and steppe ancestry. (b) Models with western sources from Southeastern Europe fail except those with Mayaky or Boyanovo EBA sources both of which are Yamnaya-derived. (c) The steppe (BPgroup)+Çayönü model fails all Chalcolithic/Bronze Anatolians except Central Anatolian Bronze Age. (d) Steppe (BPgroup) ancestry observed in all individuals of the Central Anatolian Bronze Age (±3s.e. shown). (e) BPgroup-related ancestry admixed with different substrata: Aknashen-related in the North Caucasus Maikop, Masis Blur-related in Chalcolithic Armenia, and Mesopotamian-related (Çayönü) in the ancestors of the Central Asian Bronze Age, following the route (f) from the North Caucasus to Anatolia.
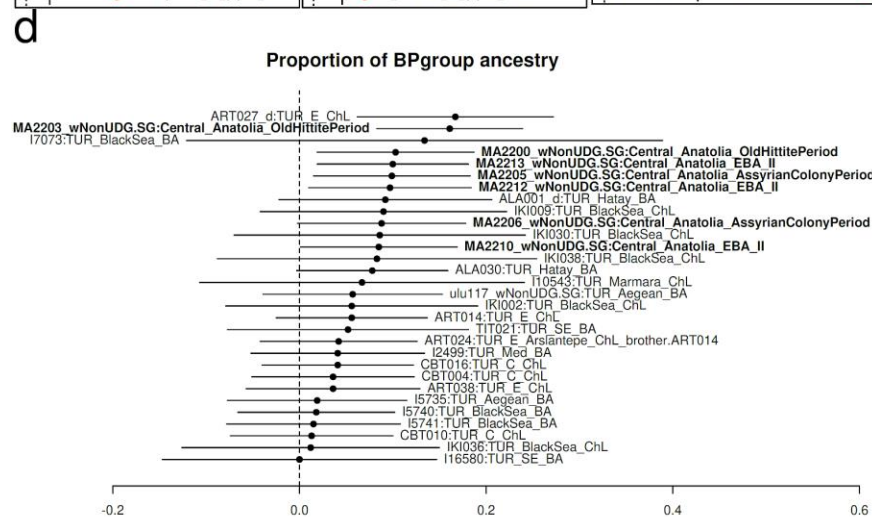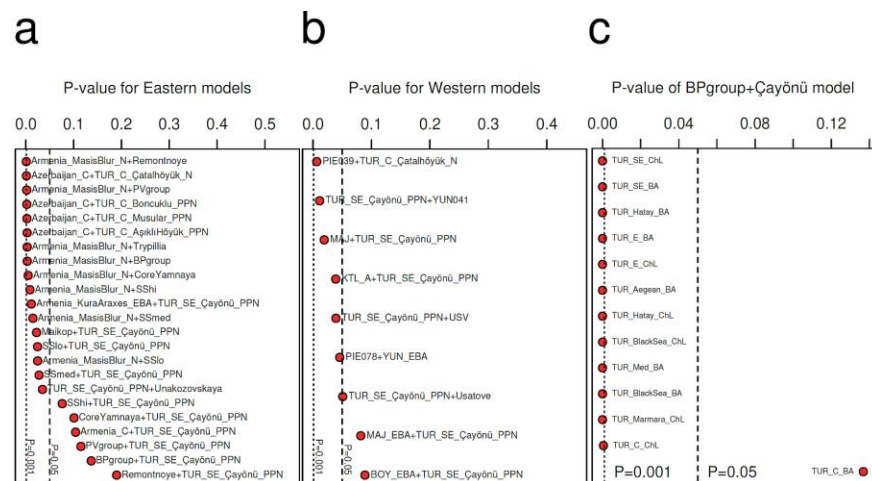
Test	Substratum	P-value	BPgroup Proportion	Std. error
Maikop	TUR_SE_Çayönü_PPN	1.07E-36	34.4%	1.8%
Maikop	Armenia_MasisBlur_N	1.02E-08	26.9%	2.3%
Maikop	**Armenia_Aknashen_N**	**1.70E-01**	**14.5%**	**2.9%**
Armenia_C	TUR_SE_Çayönü_PPN	6.32E-04	31.1%	1.8%
Armenia_C	**Armenia_MasisBlur_N**	**7.82E-01**	**27.2%**	**2.2%**
Armenia_C	Armenia_Aknashen_N	9.06E-10	12.0%	3.2%
TUR_C_BA	**TUR_SE_Çayönü_PPN**	**1.37E-01**	**10.8%**	**1.7%**
TUR_C_BA	Armenia_MasisBlur_N	2.92E-03	5.9%	2.4%
TUR_C_BA	Armenia_Aknashen_N	2.10E-17	-15.8%	3.8%
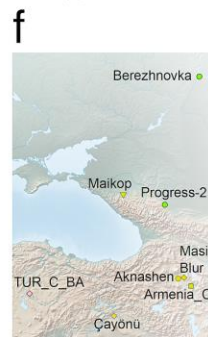
Extended Data Figure 2: Admixture time estimates. We estimate admixture times for the Core Yamnaya as a mixture of European hunter-gatherer and West Asian populations (a), for the Don Yamnaya as a mixture of Core Yamnaya and UNHG (b), for the Bulgaria-Moldova-Romania-Serbia (BMRS) Yamnaya as a mixture of Core Yamnaya and European Neolithic/Chalcolithic farmers (c), for the Corded Ware as a mixture of Core Yamnaya and Globula Amphora (d), and for Caucasus-Anatolia populations (Maikop-Armenia_C-TUR_C_BA) as a mixture of European hunter-gatherer and West Asian populations which occurred ca. 4400BCE (e). The Core Yamnaya were formed ca. 4000BCE, followed by admixture ca. 3350 BCE with UNHG and European farmers in the east and west of the Dnipro-Don region and <3000BCE in central-eastern Europe.
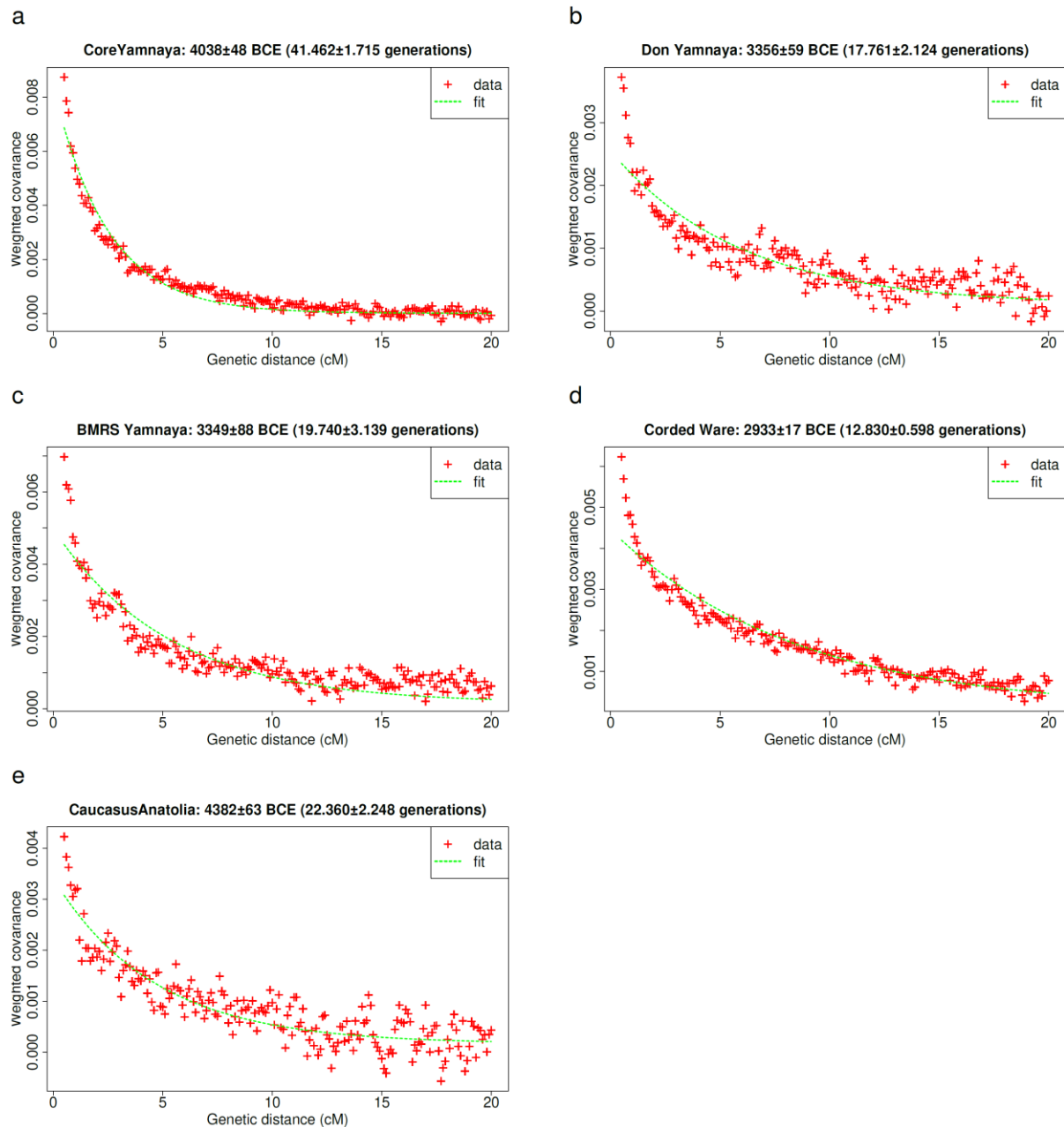
Extended Data Figure 3: A 4-way model for the entire Dnipro-Don-Volga-Caucasus region. Error bars show ±1 standard error.
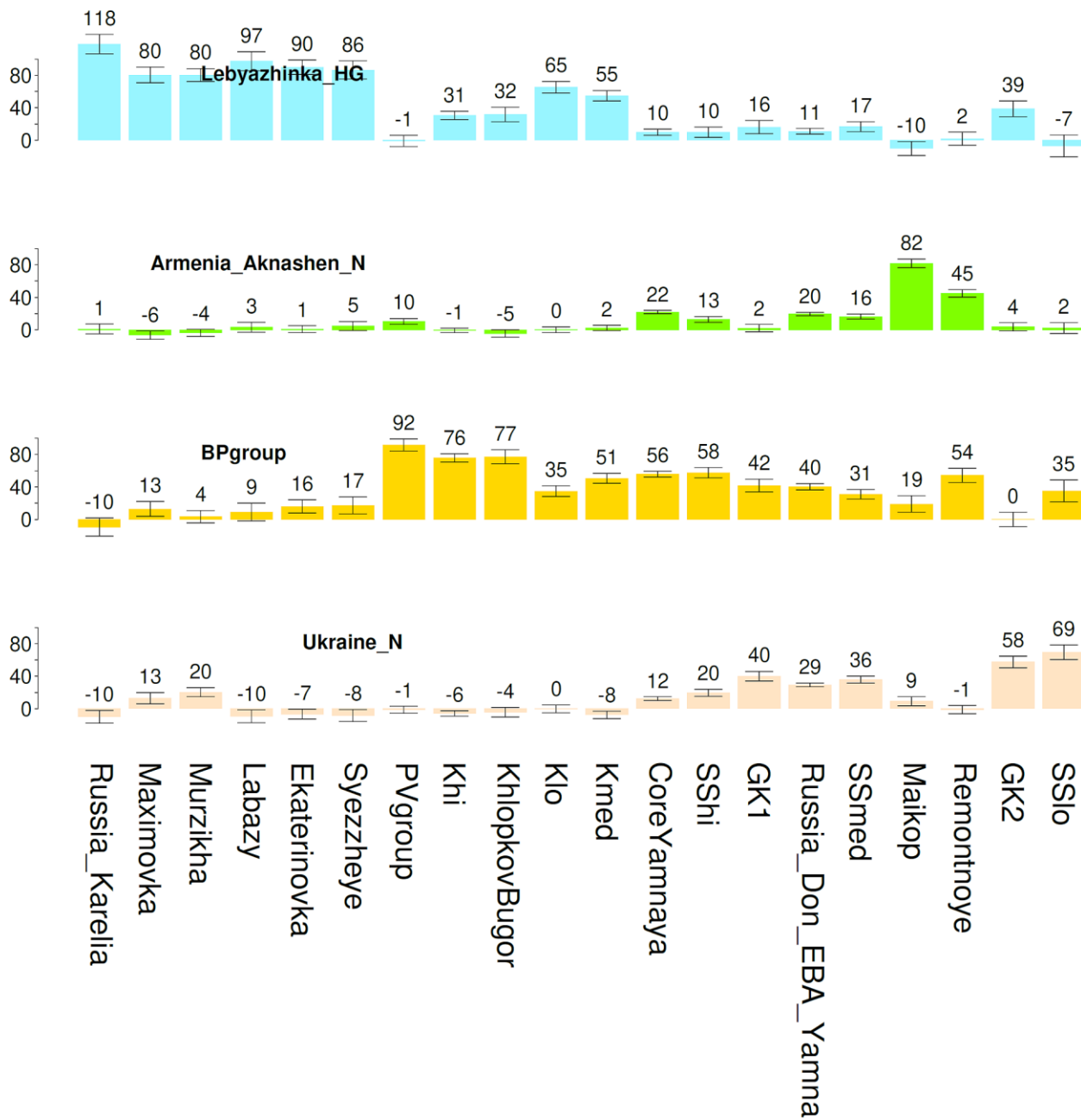
Online Tables

Online Table 1: Ancient individuals with newly reported genome-wide data.

Online Table 2: Technical details of newly reported ancient DNA libraries.

Online Table 3: Newly reported direct radiocarbon dates.

Online Table 4: All ancient individuals including in genome-wide analysis.

References

1167   1   Anthony, D. W. *The horse, the wheel, and language : how bronze-age riders from the*
1168   *Eurasian steppes shaped the modern world.* (Princeton University Press, 2007).

1169   2   Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European
1170   languages in Europe. *Nature* **522**, 207-211, doi:10.1038/nature14317 (2015).

1171   3   Olalde, I. *et al.* The Beaker phenomenon and the genomic transformation of northwest
1172   Europe. *Nature* **555**, 190-196, doi:10.1038/nature25738 (2018).

1173   4   Narasimhan Vagheesh, M. *et al.* The formation of human populations in South and
1174   Central Asia. *Science* **365**, eaat7487, doi:10.1126/science.aat7487 (2019).

1175   5   Wang, C.-C. *et al.* Ancient human genome-wide data from a 3000-year interval in the
1176   Caucasus corresponds with eco-geographic regions. *Nature Communications* **10**, 590,
1177   doi:10.1038/s41467-018-08220-8 (2019).

1178   6   Lazaridis, I. *et al.* The genetic history of the Southern Arc: A bridge between West Asia
1179   and Europe. *Science* **377**, eabm4247, doi:10.1126/science.abm4247 (2022).

1180   7   Allentoft, M. E. *et al.* Population genomics of post-glacial western Eurasia. *Nature* **625**,
1181   301-311, doi:10.1038/s41586-023-06865-0 (2024).

1182   8   Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians.
1183   *Nat Commun* **6**, 8912, doi:10.1038/ncomms9912 (2015).

1184   9   Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near East.
1185   *Nature* **536**, 419-424, doi:10.1038/nature19310 (2016).

1186   10   Lazaridis, I. *et al.* Ancient DNA from Mesopotamia suggests distinct Pre-Pottery and
1187   Pottery Neolithic migrations into Anatolia. *Science* **377**, 982-987,
1188   doi:10.1126/science.abq0762 (2022).

1189   11   Skourtanioti, E. *et al.* Genomic History of Neolithic to Bronze Age Anatolia, Northern
1190   Levant, and Southern Caucasus. *Cell* **181**, 1158-1175 e1128,
1191   doi:10.1016/j.cell.2020.04.044 (2020).

1192   12   Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for
1193   present-day Europeans. *Nature* **513**, 409-413, doi:10.1038/nature13673 (2014).

1194   13   Mathieson, I. *et al.* The genomic history of southeastern Europe. *Nature* **555**, 197-203,
1195   doi:10.1038/nature25778 (2018).

1196   14   Tian Chen, Z. *et al.* Postglacial genomes from foragers across Northern Eurasia reveal
1197   prehistoric mobility associated with the spread of the Uralic and Yeniseian languages.
1198   *bioRxiv*, 2023.2010.2001.560332, doi:10.1101/2023.10.01.560332 (2023).

1199   15   Nikitin, A. G., Lazaridis, I. & others. A genomic history of the North Pontic Region from
1200   the Neolithic to the Bronze Age. *In submission* (2024).

1201   16   Gelabert, P. *et al.* Genomes from Verteba cave suggest diversity within the Trypillians in
1202   Ukraine. *Scientific Reports* **12**, 7242, doi:10.1038/s41598-022-11117-8 (2022).

1203   17   Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*
1204   **528**, 499-503, doi:10.1038/nature16152 (2015).

1205   18   Mattila, T. M. *et al.* Genetic continuity, isolation, and gene flow in Stone Age Central
1206   and Eastern Europe. *Communications Biology* **6**, 793, doi:10.1038/s42003-023-05131-3
1207   (2023).

1208   19   Posth, C. *et al.* Palaeogenomics of Upper Palaeolithic to Neolithic European hunter-
1209   gatherers. *Nature* **615**, 117-126, doi:10.1038/s41586-023-05726-0 (2023).

1210   20   Saag, L. *et al.* Genetic ancestry changes in Stone to Bronze Age transition in the East
1211   European plain. *Science Advances* **7**, eabd6535, doi:10.1126/sciadv.abd6535 (2021).

21   Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167-172, doi:10.1038/nature14507 (2015).

22   de Barros Damgaard, P. *et al.* The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* **360**, doi:10.1126/science.aar7711 (2018).

23   Järve, M. *et al.* Shifts in the Genetic Landscape of the Western Eurasian Steppe Associated with the Beginning and End of the Scythian Dominance. *Current Biology* **29**, 2430-2441.e2410, doi:https://doi.org/10.1016/j.cub.2019.06.019 (2019).

24   Jeong, C. *et al.* A Dynamic 6,000-Year Genetic History of Eurasia's Eastern Steppe. *Cell* **183**, 890-904.e829, doi:10.1016/j.cell.2020.10.015 (2020).

25   Kumar, V. *et al.* Bronze and Iron Age population movements underlie Xinjiang population history. *Science* **376**, 62-69, doi:10.1126/science.abk1534 (2022).

26   Patterson, N. *et al.* Large-scale migration into Britain during the Middle to Late Bronze Age. *Nature* **601**, 588-594, doi:10.1038/s41586-021-04287-4 (2022).

27   Olalde, I. *et al.* The genomic history of the Iberian Peninsula over the past 8000 years. *Science* **363**, 1230, doi:10.1126/science.aav4040 (2019).

28   Zhang, F. *et al.* The genomic origins of the Bronze Age Tarim Basin mummies. *Nature* **599**, 256-261, doi:10.1038/s41586-021-04052-7 (2021).

29   Wang, C.-C. *et al.* Genomic insights into the formation of human populations in East Asia. *Nature* **591**, 413-419, doi:10.1038/s41586-021-03336-2 (2021).

30   Reich, D. *et al.* Reconstructing Native American population history. *Nature* **488**, 370-374, doi:10.1038/nature11258 (2012).

31   Vybornov, A. *et al.* Diet and Chronology of Neolithic-Eneolithic Cultures (from 6500 to 4700 cal BC) in the Lower Volga Basin. *Radiocarbon* **60**, 1597-1610, doi:10.1017/RDC.2018.95 (2018).

32   Gimbutas, M. *The prehistory of eastern Europe*. (Peabody Museum, 1956).

33   Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200-205, doi:10.1038/nature17993 (2016).

34   Anthony, D. W. *et al.* The Eneolithic cemetery at Khvalynsk on the Volga River. *Praehistorische Zeitschrift* **97**, 22-67, doi:doi:10.1515/pz-2022-2034 (2022).

35   Govedarica, B. & Manzura, I. The Giurgiulesti cemetery in chronological and cultural context of Southeastern and Eastern Europe. *Eurasia Antiqua* **22**, 1-39 (2016).

36   Penske, S. *et al.* Early contact between late farming and pastoralist societies in southeastern Europe. *Nature* **620**, 358-365, doi:10.1038/s41586-023-06334-8 (2023).

37   Patterson, N., Price, A. L. & Reich, D. Population Structure and Eigenanalysis. *PLOS Genetics* **2**, e190, doi:10.1371/journal.pgen.0020190 (2006).

38   Nikitin, A. G. I., S.; Culleton, B.J.; Potekhina, I.; Reich, D. New radiocarbon and stable isotope data from the Usatove culture site of Mayaky in Ukraine. *SSRN Electronic Journal*, doi:doi:10.2139/ssrn.4236123 (2023).

39   Skorobogatov, A. M. Pamyatniki Neolita I Eneolita v Ust'e Chernoi Kalitvy. *Trudy Voronezhskogo Oblastnogo Kraevedcheskogo Muzeiya Vyp.* **3**, 47-53 (2019).

40   Skorobogatov, A. M. & Smol'janinov, R. V. Srednestogovskie materialy v bassejne Verhnego i Srednego Dona. *Rossiyskaya arkheologiya* **2013**, 126-136 (2013).

41   Shishlina, N. I. *et al.* Paleoecology, Subsistence, and 14C Chronology of the Eurasian Caspian Steppe Bronze Age. *Radiocarbon* **51**, 481-499, doi:10.1017/S0033822200055879 (2009).

42   Korenevskii, S. *Rozhdenie Kurgana [Origins of Kurgans]* (2012).

43 Skourtanioti, E. *et al.* Genomic History of Neolithic to Bronze Age Anatolia, Northern Levant, and Southern Caucasus. *Cell* **181**, 1158-1175.e1128, doi:10.1016/j.cell.2020.04.044 (2020).

44 Guarino-Vignon, P. *et al.* Genome-wide analysis of a collective grave from Mentesh Tepe provides insight into the population structure of early neolithic population in the South Caucasus. *Communications Biology* **6**, 319, doi:10.1038/s42003-023-04681-w (2023).

45 Lazaridis, I. *et al.* The genetic history of the Southern Arc: A bridge between West Asia and Europe. *Science* **377**, eabm4247, doi:10.1126/science.abm4247 (2022).

46 Altınışık, N. E. *et al.* A genomic snapshot of demographic and cultural dynamism in Upper Mesopotamia during the Neolithic Transition. *Science Advances* **8**, eabo3609, doi:10.1126/sciadv.abo3609.

47 Lazaridis, I. *et al.* Genetic origins of the Minoans and Mycenaeans. *Nature* **548**, 214-218, doi:10.1038/nature23310 (2017).

48 Lazaridis, I. The evolutionary history of human populations in Europe. *Current Opinion in Genetics & Development* **53**, 21-27, doi:https://doi.org/10.1016/j.gde.2018.06.007 (2018).

49 Kloekhorst, A. in *The Indo-European Puzzle Revisited: Integrating Archaeology, Genetics, and Linguistics* (eds Eske Willerslev, Guus Kroonen, & Kristian Kristiansen) 42-60 (Cambridge University Press, 2023).

50 Kroonen, G., Barjamovic, G. & Peyrot, M. Linguistic supplement to Damgaard et al. 2018 : Early Indo-European languages, Anatolian, Tocharian and Indo-Iranian. (2018). <https://zenodo.org/doi/10.5281/zenodo.1240523>.

51 Egfjord, A. F.-H. *et al.* Genomic Steppe ancestry in skeletons from the Neolithic Single Grave Culture in Denmark. *PLOS ONE* **16**, e0244872, doi:10.1371/journal.pone.0244872 (2021).

52 Chintalapati, M., Patterson, N. & Moorjani, P. The spatiotemporal patterns of major human admixture events during the European Holocene. *eLife* **11**, e77625, doi:10.7554/eLife.77625 (2022).

53 Fournier, R., Tsangalidou, Z., Reich, D. & Palamara, P. F. Haplotype-based inference of recent effective population size in modern and ancient DNA samples. *Nature Communications* **14**, 7945, doi:10.1038/s41467-023-43522-6 (2023).

54 Ringbauer, H. *et al.* Accurate detection of identity-by-descent segments in human ancient DNA. *Nature Genetics* **56**, 143-151, doi:10.1038/s41588-023-01582-w (2024).

55 Fowler, C. *et al.* A high-resolution picture of kinship practices in an Early Neolithic tomb. *Nature* **601**, 584-587, doi:10.1038/s41586-021-04241-4 (2022).

56 Shishlina, N. *Reconstruction of the Bronze Age of the Caspian steppes: Life styles and life ways of pastoral nomads*. Vol. 1876 (Archaeopress, 2008).

57 Cassidy, L. M. *et al.* A dynastic elite in monumental Neolithic society. *Nature* **582**, 384-388, doi:10.1038/s41586-020-2378-6 (2020).

58 Olander, T. Indo-European cladistic nomenclature. **124**, 231-244, doi:doi:10.1515/if-2019-0008 (2019).

59 Anthony, D. W. & Ringe, D. The Indo-European Homeland from Linguistic and Archaeological Perspectives. *Annual Review of Linguistics* **1**, 199-219, doi:10.1146/annurev-linguist-030514-124812 (2015).

1303   60   Pronk, T. C. Indo-European secondary products terminology and the dating of Proto-
1304   Indo-Anatolian. *Journal of Indo-European Studies* **49**, 141-170 (2022).
1305   61   Kassian, A. S. *et al.* Rapid radiation of the inner Indo-European languages: an advanced
1306   approach to Indo-European lexicostatistics. *Linguistics* **59**, 949-979,
1307   doi:doi:10.1515/ling-2020-0060 (2021).
1308   62   Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167-172,
1309   doi:10.1038/nature14507 (2015).
1310   63   Ringe, D., Warnow, T. & Taylor, A. Indo-European and Computational Cladistics.
1311   *Transactions of the Philological Society* **100**, 59-129, doi:https://doi.org/10.1111/1467-
1312   968X.00091 (2002).
1313   64   Heggarty, P. *et al.* Language trees with sampled ancestors support a hybrid model for the
1314   origin of Indo-European languages. *Science* **381**, eabg0818,
1315   doi:10.1126/science.abg0818.
1316   65   Kroonen, G., Jakob, A., Palmér, A. I., van Sluis, P. & Wigman, A. Indo-European cereal
1317   terminology suggests a Northwest Pontic homeland for the core Indo-European
1318   languages. *PLOS ONE* **17**, e0275744, doi:10.1371/journal.pone.0275744 (2022).
1319   66   Kristiansen, K. in *Dispersals and Diversification: Linguistic and Archaeological*
1320   *Perspectives on the Early Stages of Indo-European* Vol. 19 *Brill's Studies in Indo-*
1321   *European Languages & Linguistics* (eds M. Serangeli & Thomas Olander) 157-165
1322   (Brill, 2019).
1323   67   Fernandes, D. M. *et al.* The spread of steppe and Iranian-related ancestry in the islands of
1324   the western Mediterranean. *Nature Ecology & Evolution* **4**, 334-345,
1325   doi:10.1038/s41559-020-1102-0 (2020).
1326   68   Pinhasi, R., Fernandes, D. M., Sirak, K. & Cheronet, O. Isolating the human cochlea to
1327   generate bone powder for ancient DNA analysis. *Nat Protoc* **14**, 1194-1205,
1328   doi:10.1038/s41596-019-0137-7 (2019).
1329   69   Sirak, K. A. *et al.* A minimally-invasive method for sampling human petrous bones from
1330   the cranial base for ancient DNA analysis. *Biotechniques* **62**, 283-289,
1331   doi:10.2144/000114558 (2017).
1332   70   Damgaard, P. B. *et al.* Improving access to endogenous DNA in ancient bones and teeth.
1333   *Sci Rep* **5**, 11184, doi:10.1038/srep11184 (2015).
1334   71   Sirak, K. *et al.* Human auditory ossicles as an alternative optimal source of ancient DNA.
1335   *Genome research* **30**, 427-436, doi:10.1101/gr.260141.119 (2020).
1336   72   Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave
1337   bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy*
1338   *of Sciences of the United States of America* **110**, 15758-15763,
1339   doi:10.1073/pnas.1314445110 (2013).
1340   73   Korlević, P. *et al.* Reducing microbial and human contamination in DNA extractions
1341   from ancient bones and teeth. *Biotechniques* **59**, 87-93, doi:10.2144/000114320 (2015).
1342   74   Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-
1343   glycosylase treatment for screening of ancient DNA. *Philos Trans R Soc Lond B Biol Sci*
1344   **370**, 20130624, doi:10.1098/rstb.2013.0624 (2015).
1345   75   Rohland, N., Glocke, I., Aximu-Petri, A. & Meyer, M. Extraction of highly degraded
1346   DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nat*
1347   *Protoc* **13**, 2447-2461, doi:10.1038/s41596-018-0050-5 (2018).

76   Prendergast, M. E. *et al.* Ancient DNA reveals a multistep spread of the first herders into sub-Saharan Africa. *Science* **365**, doi:10.1126/science.aaw6275 (2019).

77   Gansauge, M. T., Aximu-Petri, A., Nagel, S. & Meyer, M. Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. *Nat Protoc* **15**, 2279-2300, doi:10.1038/s41596-020-0338-0 (2020).

78   Maricic, T., Whitten, M. & Paabo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PloS one* **5**, e14004, doi:10.1371/journal.pone.0014004 (2010).

79   Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 2223-2227, doi:10.1073/pnas.1221359110 (2013).

80   Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216-219, doi:10.1038/nature14558 (2015).

81   Rohland, N. *et al.* Three assays for in-solution enrichment of ancient human DNA at more than a million SNPs. *Genome research* **32**, 2068-2078, doi:10.1101/gr.276728.122 (2022).

82   Lipson, M. *et al.* Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* **551**, 368-372, doi:10.1038/nature24476 (2017).

83   Behar, D. M. *et al.* A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am J Hum Genet* **90**, 675-684, doi:10.1016/j.ajhg.2012.03.002 (2012).

84   Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

85   Fu, Q. *et al.* A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes. *Current Biology* **23**, 553-559, doi:https://doi.org/10.1016/j.cub.2013.02.044 (2013).

86   Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356, doi:10.1186/s12859-014-0356-4 (2014).

87   Skoglund, P. *et al.* Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proceedings of the National Academy of Sciences* **111**, 2229, doi:10.1073/pnas.1318934111 (2014).

88   Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

89   Weissensteiner, H. *et al.* HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res* **44**, W58-63, doi:10.1093/nar/gkw233 (2016).

90   van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation* **30**, E386-E394, doi:10.1002/humu.20921 (2009).

91   Shinde, V. *et al.* An Ancient Harappan Genome Lacks Ancestry from Steppe Pastoralists or Iranian Farmers. *Cell* **179**, doi:10.1016/j.cell.2019.08.048 (2019).

92   Harney, É. *et al.* Ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation. *Nature Communications* **9**, 3336, doi:10.1038/s41467-018-05649-9 (2018).

93   Rivollat, M. *et al.* Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers. *Science Advances* **6**, eaaz5344, doi:10.1126/sciadv.aaz5344 (2020).

94   Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461**, 489-494, doi:10.1038/nature08365 (2009).

95   Adamov, D., Gurianov, V. M., Karzhavin, S., Tagankin, V. & Urasin, V. Defining a New Rate Constant for Y-Chromosome SNPs based on Full Sequencing Data. *Russian Journal of Genetic Genealogy* **7**, 1920-2997 (2015).

96   Sinnott, R. W. Virtues of the Haversine. *Sky and telescope* **68**, 158 (1984).

97   Charlesworth, B. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* **10**, 195-205, doi:10.1038/nrg2526 (2009).

98   Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am J Phys Anthropol* **128**, 415-423, doi:10.1002/ajpa.20188 (2005).

Roots of Caucasus-Lower Volga people

The people of the homeland of the Proto-Indo-Anatolians (1) had ancestry from the Caucasus Neolithic (Aknashen) and Lower Volga Eneolithic (Berezhnovka).

Caucasus hunter-gatherer ancestry was common to both groups. In the Lower Volga Eneolithic there was also Eastern hunter-gatherer (Lebyazhinka) and Central Asian-Siberian (Tutkaul) ancestry. In the Caucasus Neolithic there was also ancestry from the Neolithic people of Anatolia-Mesopotamia (Çayönü).

Pre-Indo-European languages were spoken by at least some of these diverse ancestors, but living Indo-European languages trace their origins to the Yamnaya expansion ca. 3300BCE out of (3) and the earlier Indo-Anatolian expansion ca. 4400-4000BCE out of (1).

Caucasus hunter-gatherer
Central Asia (Tutkaul)
Middle Volga (Lebyazhinka)
Middle Volga (Labazy)
Middle Volga (Ekaterinovka)
Middle Volga (Syezzheye)
Middle Volga (Khvalynsk-low)
Middle Volga (Khvalynsk-medium)
Middle Volga (Khvalynsk-high)
Berezhnovka-2/Progress-2 group
Caucasus Neolithic (Aknashen)
Caucasus Neolithic (Masis Blur)
N. Mesopotamia Neolithic (Çayönü)

Murzikha
Kama R.
Lebyazhinka
Ekaterinovka
Syezzheye
Labazy
Khvalynsk
Berezhnovka-2
2

Kopachiv
Deriivka
Golubaya Krinitsa
Oleksandria
Igren
Krivyansky
Vinogradnoe
3

Don R.
Volga R.
Remontnoye
1

Vonyuchka-1
Progress-2
Maikop

Caucasus hunter-gatherers

Dnipro R.

Southeastern Europe
Black Sea
Caspian Sea

Aknashen
Masis Blur
Kalehöyük
Areni-1
Ovaören
Çayönü
4

Anatolia
Kızılırmak R.
Caucasus Mts.
N. Mesopotamia
Mediterranean Sea

Yamnaya (Core group)
Serednii Stih-high
Yamnaya (Lower Don)
Serednii Stih-medium
Dnipro-Don (Golubaya Krinitsa-1)
Serednii Stih-low
Dnipro-Don (Golubaya Krinitsa-2)
Volga-Kama confluence (Murzikha)
Middle Volga (Lebyazhinka)
Middle Volga (Maximovka)
Middle Volga (Labazy)
Middle Volga (Ekaterinovka)
Middle Volga (Syezzheye)
Middle Volga (Khvalynsk-low)
Middle Volga (Khvalynsk-medium)
Middle Volga (Khvalynsk-high)
Berezhnovka-2/Progress-2 group
Progress-2/Vonyucka-1 group
Manych Depression (Remontnoye)
N. Caucasus (Maikop)
Caucasus Neolithic (Aknashen)
Caucasus Neolithic (Masis Blur)
Armenia Chalcolithic (Areni-1)
N. Mesopotamia Neolithic (Çayönü)
C. Anatolia Bronze Age (Kalehöyük / Ovaören)

Dnipro Don
Volga
Caucasus Lower Volga
Caucasus Anatolia

Expansions of Caucasus-Lower Volga people

Eneolithic people of the North Caucasus-Lower Volga region (1) had a mix of Caucasus Neolithic (Aknashen) and Lower Volga Eneolithic (Berezhnovka) ancestry. Migrations from (1) to the Volga (2), Dnipro-Don (3) and Caucasus-Anatolia (4) regions spread one or both of these ancestries, admixing in (2) with Eastern hunter-gatherers (Lebyazhinka), in (3) with Dnipro-Don hunter-gatherers, and in (4) with Caucasus-Mesopotamian (Masis Blur, Çayönü) populations.

The Yamnaya were formed ca. 4000BCE in the Serednii Stiih cultural context of the Dnipro-Don area (3) before expanding ca. 3300BCE across the steppe. Anatolian speakers of Central Anatolia (4) living ca. 2750-1500BCE shared with the Yamnaya ancestry from area (1).

The Proto-Indo-Anatolian homeland was thus probably in the North Caucasus-Lower Volga area (1 and environs). Migrations to the west (3) and south (4) ca. 4400-4000BCE split PIA into the Proto-Indo-European languages (spoken by the Yamnaya) and Proto-Anatolian languages (spoken in Bronze Age Central Anatolia).

Tutkaul

● Core Yamnaya sampling locations
○ Other sites of interest

a

Ukraine_N

Ukraine-Eastern hunter-gatherer cline

Golubaya
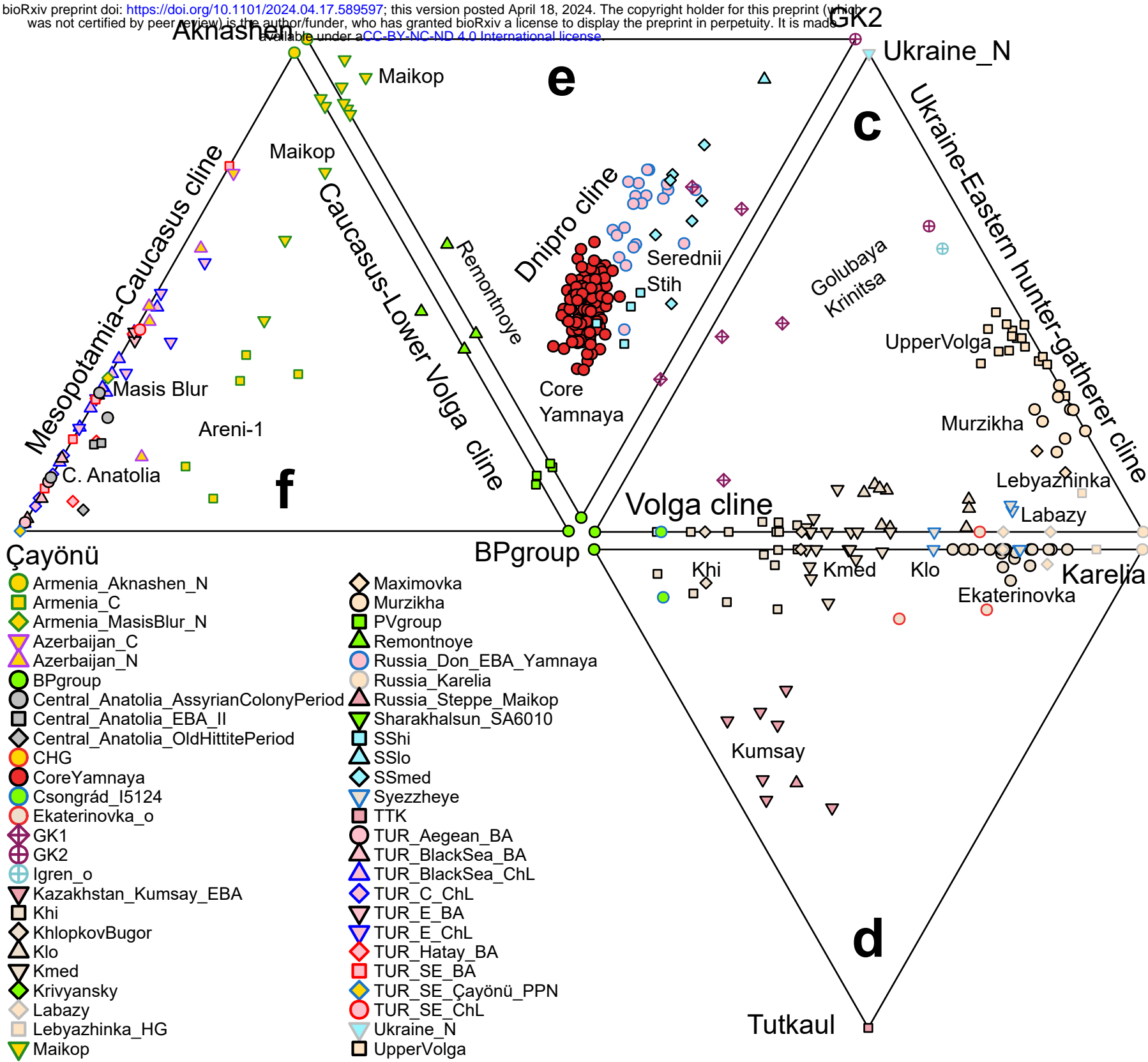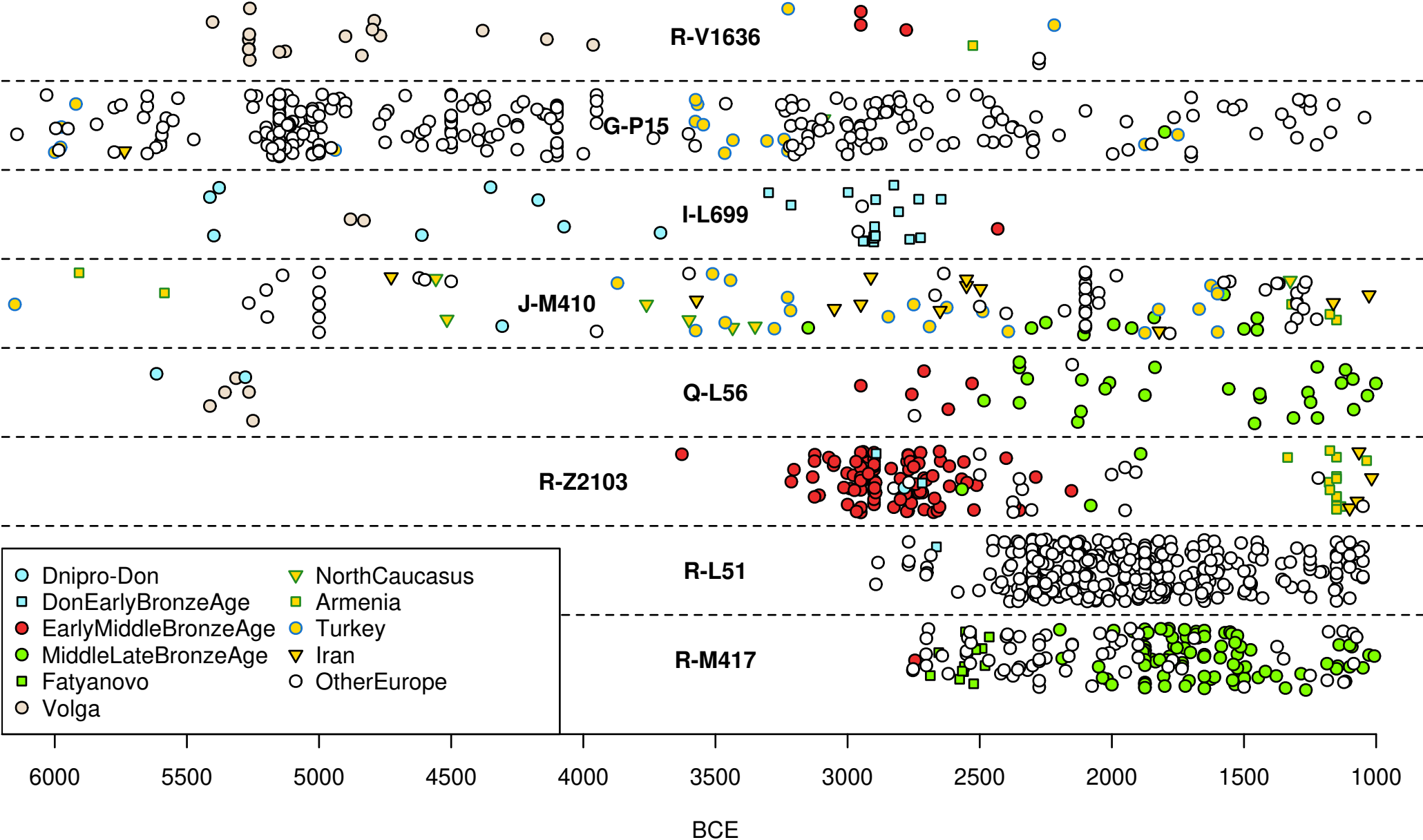Krinitsa

UpperVolga

Serednii Stih

Murzikha

Krivyansky

Lebyazhinka

Syezzheye   Labazy

CHG   Karelia

b

GK2

Caucasus-Don cline

Serednii Stih   GK1

Krivyansky

BPgroup

Steppe
Maikop

Don-Volga-Central Asia cline

CHG

Tutkaul

e

Aknashen   GK2

Maikop   Ukraine_N

Maikop

Mesopotamia-Caucasus cline

Caucasus-Lower Volga cline

Remontnoye

Dnipro cline

Serednii
Stih

Ukraine-Eastern hunter-gatherer cline

Golubaya
Krinitsa

UpperVolga

Core
Yamnaya

Masis Blur

Areni-1

Murzikha

C. Anatolia

Lebyazhinka

Labazy

f

Çayönü

Volga cline

Khi   Kmed   Klo   Karelia

BPgroup

Ekaterinovka

c

Kumsay

d

Tutkaul

Legend:

- Armenia_Aknashen_N
- Armenia_C
- Armenia_MasisBlur_N
- Azerbaijan_C
- Azerbaijan_N
- BPgroup
- Central_Anatolia_AssyrianColonyPeriod
- Central_Anatolia_EBA_II
- Central_Anatolia_OldHittitePeriod
- CHG
- CoreYamnaya
- Csongrád_I5124
- Ekaterinovka_o
- GK1
- GK2
- Igren_o
- Kazakhstan_Kumsay_EBA
- Khi
- KhlopkovBugor
- Klo
- Kmed
- Krivyansky
- Labazy
- Lebyazhinka_HG
- Maikop
- Maximovka
- Murzikha
- PVgroup
- Remontnoye
- Russia_Don_EBA_Yamnaya
- Russia_Karelia
- Russia_Steppe_Maikop
- Sharakhalsun_SA6010
- SShi
- SSlo
- SSmed
- Syezzheye
- TTK
- TUR_Aegean_BA
- TUR_BlackSea_BA
- TUR_BlackSea_ChL
- TUR_C_ChL
- TUR_E_BA
- TUR_E_ChL
- TUR_Hatay_BA
- TUR_SE_BA
- TUR_SE_Çayönü_PPN
- TUR_SE_ChL
- Ukraine_N
- UpperVolga

a

Symbol	Group
○	Afanasievo
△	CaspianInland_Yamnaya
+	Caucasus_Yamnaya
×	Chelyabinsk_Yamnaya
⊙	Don_Yamnaya
▽	Ishkinovka_Yamnaya
⊠	Kalmykia_Yamnaya
✳	Orlovka_Yamnaya
■	Remontnoye_Yamnaya
⊞	Rostov_LateYamnaya_BA
◆	Samara_Yamnaya
⊕	Stavropolkrai_Yamnaya
▽	Ural_Yamnaya
⊡	Volga_Yamnaya
⊛	Volgograd_Yamnaya

b

Symbol	Group
●	Bulgaria_Yamnaya
○	Hungary_LN
⬡	Hungary_Yamnaya
▲	Moldova_Yamnaya
◆	Romania_C_Bodrogkeresztur
▽	Romania_N
▼	Romania_Yamnaya
◆	Serbia_Yamnaya
⊡	Slovakia_Yamnaya
○	Trypillia
▽	Ukraine_N
●	Ukraine_Yamnaya
●	Usatove

a 3150-2850 BCE

b 2850-2550 BCE

c
- 3150-2850 BCE
- 2850-2550 BCE

a

PreYamnaya

b

Yamnaya

c

At least one IBD segment≥≥20cM

235
734

84
1757

4
1937

0
1225

0
1133

d

At least one IBD segment≥≥20cM

86
283

257
2880

120
2374

54
2023

21
972

46
1204

53
1113

34
1681

17
571

18
737

5
358

e

Three IBD segments≥20cM or sum IBD≥100cM

97
734

8
1757

0
1937

0
1225

0
1133

f

Three IBD segments≥20cM or sum IBD≥100cM

36
283

21
2880

0
2374

1
2023

0
972

0
1204

0
1113

0
1681

0
571

0
737

0
358

g

(1) Three IBD segments >=20cM or sum IBD >=100cM
(2) At least one IBD segment >=20cM but not in (1)
Neither (1) nor (2)

Between Cemeteries (n=1106)   Between Kurgans (n=95)   Within Kurgans (n=125)   Hazleton North (n=435)   Neolithic Ireland (n=103)