Supplementary Materials

The Formation of Human Populations in South and Central Asia

Table of Contents

S1 Description of Four Online Tables and the Online Data Visualizer 7

S2 Archeological descriptions of 523 newly reported ancient individuals 12

S2.1 Overview 12

S2.2 Individuals from the Forest Zone / Steppe..13

S2.2.1 Individuals from the Forest Zone in the Neolithic period 13

S2.2.1.1 West Siberian Hunter-Gatherers 13

S2.2.1.1.1 Sosnoviy-Ostrov, Tyumen Oblast, western Siberia, Russia (n=1)............................... 13

S2.2.1.1.2 Mergen Settlements, Tyumen Oblast, western Siberia, Russia (n=2)........................... 14

S2.2.2 Sites from the Steppe Zone in Early to Middle Bronze Age Cemeteries 16

S2.2.2.1 Kumsay (Kyryk Oba), Kazakhstan (n=4) 16

S2.2.2.2 Mereke, Kazakhstan (n=3)... 17

S2.2.2.3 Dali, Bayan-Zherek Valley, Kazakhstan (n=1)... 18

S2.2.2.4 Yamnaya Pastoralists from Samara, Russia (n=1).. 19

S2.2.2.5 Afanasievo pastoralists of the Altai mountains and Yenisei river basin (n=26).................. 19

S2.2.2.5.1 Saldyar-1 cemetery, Altai Mountains, Russia (n=1) 20

S2.2.2.5.2 Ust'-Kuyum cemetery, Altai Mountains, Russia (n=1)................................ 21

S2.2.2.5.3 Kaminnaya Cave, Altai Mountains, Russia (n=1).. 21

S2.2.2.5.4 Karasuk III, Russia, cluster B, kurgans 1 and 2 (n=8) 21

S2.2.2.5.5 Afanasieva Gora, Middle Yenisei region, Russia (n=5)............................... 24

S2.2.2.5.6 Elo 1, Russia (n=3)... 26

S2.2.2.5.7 Elo Bashi, Russia (n=2).. 26

S2.2.2.5.8 Lower Tyumechin 1, western Siberia, Russia (n=3) 27

S2.2.2.5.9 Krasnoyarsk Krai, Minusinsk, Podsukhanikha II (n=1)................................ 28

S2.2.2.5.10 Inskoy Dol, Russia (n=1)... 29

S2.2.2.6 Poltavka: Pastoralists from the Poltavka culture (n=4) 31

S2.2.2.6.1 Grachevka cluster, Sok River, Samara Oblast, Russia (n=1)........................... 31

S2.2.2.6.2 Samara River valley and steppes to the south (n=1) 32

S2.2.2.6.3 Zhuravlikha I, Bolshoi Irgiz River, Samara oblast (n=1).............................. 32

S2.2.2.6.4 Nikolaevka III, Volzhskiy region, Samara Oblast (n=1)................................ 32

S2.2.2.7 Potapovka culture pastoralists from the Grachevka kurgan cemeteries (n=2) 33

S2.2.3 Sites from the Steppe Zone in the MLBA 34

S2.2.3.1 Corded Ware Complex from the Western Steppe (n=7)...................................... 34

S2.2.3.1.1 Radovesice, Czech Republic (n=3) 34

S2.2.3.1.2 Bílina, Czech Republic (n=3)... 35

S2.2.3.1.3 Velké Žernoseky, Czech Republic (n=1) 37

S2.2.3.2 Northeastern Kazakhstan, Kanai Cemetery (n=1) 37

S2.2.3.3 Petrovka culture, Stepnoe VII cemetery, Russia (n=3)... 37

S2.2.3.4 Sintashta culture from Kamennyi Ambar 5 cemetery, Russia (n=50) ... 39
S2.2.3.5 Satan, Central Kazakhstan (n=1) ... 49
S2.2.3.6 Maitan, Central Kazakhstan (n=9) ... 50
S2.2.3.7 Karagash 2, Central Kazakhstan (n=3) ... 51
S2.2.3.8 Ak-Moustafa, Central Kazakhstan (n=1) ... 52
S2.2.3.9 Dali, Bayan-Zherek Valley, Kazakhstan (n=4) ... 53
S2.2.3.9.1 Middle to Late Bronze Age Dali 1 cemetery, burial 1, Kazakhstan (n=3) ... 53
S2.2.3.9.2 Early Bronze Age Dali settlement, Kazakhstan (n=1) ... 55
S2.2.3.10 Lisakovskiy, Northern Kazakhstan (n=2) ... 56
S2.2.3.11 Kairan, Kazakhstan (n=10) ... 56
S2.2.3.12 Kyzyl Bulak, Kazakhstan (n=2) ... 60
S2.2.3.13 Multiple sites near Krasnoyarsk, Western Siberia (n=18) ... 61
S2.2.3.13.1 Potroshilovo II cemetery, Russia (n=3) ... 62
S2.2.3.13.2 Ust-Bir IV cemetery, Russia (n=3) ... 62
S2.2.3.13.3 Chumysh-Perekat-1 cemetery (n=1) ... 63
S2.2.3.13.4 Orak burial place, Russia (n=11) ... 63
S2.2.3.14 Priobrazhenka 3, Russia (n=1) ... 65
S2.2.3.15 Srubnaya culture, Samara Steppes, Samara oblast, Russia (n=1) ... 66
S2.2.3.16 Solyanka Vodokhranilische, western Kazakhstan (n=1) ... 67
S2.2.3.17 Aktogai, Kazakhstan (n=5) ... 67
S2.2.3.18 Kazakh Mys, Kazakhstan (n=4) ... 69
S2.2.3.19 Dashti-Kozy, Tajikistan (n=3) ... 71
S2.2.3.20 Kokcha 3, Kokcha, Uzbekistan (n=3) ... 72
S2.2.3.21 Oy-Dzhaylau III, Kazakhstan, (n=7) ... 73
S2.2.3.22 Taldysay, Central Kazakhstan (n=2) ... 75
S2.2.3.23 Talapty II, South Kazakhstan (n=1) ... 76
S2.2.3.24 Kashkarchi Burial Mound, Ferghana Province, Uzbekistan (n=2) ... 77
S2.2.3.25 Bylkyldak, Central Kazakhstan (n=1) ... 79
S2.2.3.26 Shoendykol, Baianaul District, Pavlodar Region, Eastern Kazakhstan (n=3) ... 79
S2.2.4 Sites from the Steppe Zone in the LBA ... 81
S2.2.4.1 Chanchar, Kazakhstan (n=2) ... 81
S2.2.4.2 Georgievsky Bugor, Kazakhstan (n=2) ... 82
S2.2.4.3 Guruldek, Kazakhstan (n=1) ... 82
S2.2.4.4 Priobrazhenka 3, Russia (n=1) ... 83
S2.2.4.5 Zevakinskiy stone fence, Kazakhstan (n=8) ... 83
S2.2.4.6 Tasbas, Byan Zherek (n=3) ... 87
S2.2.4.7 Malaya Krasnoyarka, Ul’ken Narym, Katon Karagai, East Kazakhstan (n=1) ... 88
S2.2.4.8 Myrzhyk, Kazakhstan (n=1) ... 88

S2.3 Sites from Iran/Turan ... 88

S2.3.1 Sites from Mesolithic Iran/Turan ... 88

S2.3.1.1 Belt Cave in the Alborz Mountains, Iran (n=1) ... 88
S2.3.2 Sites from Neolithic Iran/Turan ... 89
S2.3.2.1 Ganj Dareh, Zagros Mountains, Iran (n=9) ... 89
S2.3.3 Sites from Copper Age/Eneolithic Iran/Turan ... 91
S2.3.3.1 Hajji Firuz, Zagros Mountains, Iran (n=7) ... 91
S2.3.3.2 Aigyrzhal, Kyrgyzstan (n=2) ... 93
S2.3.3.3 Sarazm, Tajikistan (n=2) ... 94
S2.3.3.4 Geoksyur, Turkmenistan (n=25) ... 94
S2.3.3.5 Tepe Anau, Turkmenistan (n=3) ... 99
S2.3.3.6 Parkhai I and II, Sumbar Valley, Turkmenistan (n=9) ... 99
S2.3.3.7 Tepe Hissar, Damghan Plain, Iran (n=13) ... 103
S2.3.4 Sites from Turan in the BA ... 106
S2.3.4.1 Darra-i-Kur Cave, Afghanistan (n=1) ... 106
S2.3.4.2 Farkhor, Tajikistan (n=2) ... 107
S2.3.4.3 Zaman Baba, Uzbekistan (n=1) ... 107
S2.3.4.4 Shahr-i-Sokhta, Seistan, Iran (n=22) ... 108
S2.3.4.5 Gonur Tepe, Bactria Margiana Archaeological Complex (BMAC), Turkmenistan (n=47) 115
S2.3.4.6 Jarkutan, post-BMAC, Uzbekistan (n=10) ... 126
S2.3.4.7 Sappali-Tepe, post-BMAC, Uzbekistan (n=14) ... 128
S2.3.4.8 Bustan, mostly post-BMAC, Uzbekistan (n=14) ... 131
S2.3.4.9 Sumbar II, Sumbar Valley, Southern Turkmenistan (n=1) ... 134
S2.3.5 Sites from historical period Turan ... 135
S2.3.5.1 Kushan empire context individuals from Ksirov, Tajikistan (n=5) ... 135

S2.4 Sites from South Asia ... 136

S2.4.1 Late Bronze-Iron Age Swat Protohistoric Graves (Gandhara Grave Culture), Swat Valley, Pakistan (n=99) ... 136
S2.4.1.1 Udegram, Swat Protohistoric Graves, Pakistan (n=16) ... 137
S2.4.1.2 Gogdara, Swat Protohistoric Graves (n=2) ... 143
S2.4.1.3 Loebanr, Swat Protohistoric Graves, Pakistan (n=38) ... 144
S2.4.1.4 Katelai, Swat Protohistoric Graves, Pakistan (n=33) ... 153
S2.4.1.5 Arkotkila (Arkot-kala), Swat Protohistoric Graves, Pakistan (n=1) ... 159
S2.4.1.6 Barikot, Swat Protohistoric Graves, Pakistan (n=3) ... 160
S2.4.1.7 Aligrama, Swat Protohistoric Graves, Pakistan (n=3) ... 161
S2.4.1.8 Butkara, Swat Protohistoric Graves, Pakistan (n=5) ... 162
S2.4.2 Individuals from South Asia in the early historic and historic periods ... 163
S2.4.2.1 Aligrama, Early Historic Period, Pakistan (n=3) ... 163
S2.4.2.2 Saidu Sharif Buddhist monastery, early historic graves in the Swat Valley, Pakistan (n=12) 165
S2.4.2.3 Butkara, Early Historic Period, Swat Valley, Pakistan (n=5) ... 168
S2.4.2.4 Barikot: Historic Period Graves in the Swat Valley, Pakistan (n=4) ... 169
S2.4.2.5 Medieval graves in Udegram, Swat Valley, Pakistan (n=3) ... 171

S2.4.2.6 Parwak, Chitral District, Khyber Pakhtunkhwa Province, Pakistan (n=3) 172
S2.4.2.7 Shah Mirandeh Graves, Singoor, Chitral District, Khyber Pakhtunkwha Province, Pakistan (n=1) 174

S3 Sample classification and nomenclature 176

S3.1 Classification of ancient samples from South and Central Asia 176

S3.2 Classification of 1,789 individuals from 246 groups in South Asia 177

S3.2.1 Austroasiatic-Related Cluster (n=41) 179
S3.2.2 East Asian-Related Cluster (n=10) 180
S3.2.3 Other groups not included in any analysis cluster (n=21) 180

S4 Population Genetic Characterization of Ancient Individuals 182

S4.1 Terminology and acronyms 182

S4.1.1 Time periods 182
S4.1.2 Population labels 182

S4.2 Population modeling strategy 182

S4.2.1 First line of analysis – PCA and ADMIXTURE 183
S4.2.2 Second line of analysis: f_3- and f_4-statistics 183
S4.2.3 Third line of analysis - admixture modeling with qpAdm 183
S4.2.4 “Distal” modeling of each group from the Copper Age period and onward 184

S4.3 “Proximal” modeling of each group from the Copper Age and onward 185

S4.4 Population modeling 188

S4.4.1 Iran and Turan 188
S4.4.1.1 Admixture and movement in the Neolithic and Copper Age Periods 189
S4.4.1.2 The Bronze Age period of Turan and the formation of the BMAC 200
S4.4.2 The Forest Zone and Steppe 224
S4.4.2.1 The Neolithic and Eneolithic periods in the Steppe and Russian Forest Zone 226
S4.4.2.2 The Early Bronze Age: Formation and decline of Steppe EMBA groups 233
S4.4.2.3 The Middle and Late Bronze Age of the Western and Central Steppe 243
S4.4.2.4 The Late Bronze Age and Early Iron Age: Large scale admixture of East and West 255
S4.4.3 The historical period in Turan and the Steppe 260
S4.4.4 South Asia 262
S4.4.4.1 The Late Bronze and Iron Age and Historical Period in South Asia 262

S5 The Genomic formation of South Asia in light of ancient DNA 271

S5.1 The clinal nature of ancient and present-day individuals from South Asia and the rationale for our modeling framework 271

S5.2 Hierarchical modeling of ancient and modern clines 273

S5.2.1 Developing a hierarchical model for estimating ancestry along admixture clines 273
S5.2.2 The Indus Periphery Cline: An Early to Middle Bronze Age ancestry cline 276
S5.2.3 The Steppe Cline: A South Asian Ancestry Gradient formed by a Mixture of a point on the Indus Periphery Cline and MLBA Steppe Pastoralists 280
S5.2.4 The Modern Indian Cline: A Mixture Between a Point on the Indus Periphery Cline (ASI) and a point on the Steppe Cline (ANI) 284

S5.3 The Modern Indian cline is not a perfect mixture of two ancestral populations 291

S5.4 Evidence for excess Steppe pastoralist-related ancestry in groups of traditionally priestly status 293

S5.5 Additional groups with poor fits to the 2-way admixture model 294

S5.6 A single joint model for the 3 clines determines the ASI and the ANI 295

S5.7 Simple statistics confirm Iranian farmer-related ancestry in the ASI 296

S5.8 No significant evidence for AHG-related gene flow into Iranian farmers 300

S5.9 Admixture graph relating South Asians to non-South Asians 303

S5.10 High proportions of AASI ancestry in present-day Austroasiatic speakers 304

S5.11 Sex bias in the spread of Steppe ancestry into South Asia 305

S6 Dating of population mixture events in present-day and ancient individuals 307

S6.1 Admixture timing estimates on the Modern Indian Cline..307

S6.2 Ancient Individuals 309

S6.2.1 DATES: Theory.. 309
S6.2.2 DATES: Software / Web Resources.. 310
S6.2.3 Testing the performance of DATES 310
S6.2.4 Application of DATES 312
S6.2.4.1 Indus_Periphery_Pool... 312
S6.2.4.2 South Asia: Early Late Bronze-Iron Age Swat Valley individuals 313

S7 Uniparental marker analysis 315

S7.1 Y-chromosome haplogroup determination 315

S7.2 Interpreting the Y-chromosome data in light of our autosomal admixture models ..317

S7.3 Interpreting mitochondrial DNA (mtDNA) assignments in light of our autosomal admixture models 320

S1 Description of Four Online Tables and the Online Data Visualizer

We begin by listing four online tables that are too large to include within this supplement.

Table S 1 Summary information on the archaeological context and genetic results for all individuals analyzed in this study

Table S 2 Laboratory methods and summary of genetic results for each library newly generated for this study

Table S 3 Information about genetic kinship between pairs of individuals detected up to the 3rd degree

Table S 4 Details on 269 newly reported radiocarbon dates and associated stable isotope data used for quality control

Table S 5 246 present-day South Asian groups whose data we analyzed and analyses for a subset of 140 groups on the Modern Indian Cline

We have also made freely available available an Online Data Visualizer based on the Tableau software. The Online Data Visualizer allow users to explore the genetic, temporal, and geographic relationships among the ancient and modern individuals analyzed in this study. A link to the visualizer can be found here:

https://public.tableau.com/profile/vagheesh#!/vizhome/TheGenomicFormationofSouthandCentralAsia/Fig_1

It has four tabs.

The first tab visualizes all qualitative statistics described in the paper, including the PCA and ADMIXTURE plots.

The second visualizes information about uniparental markers.

The third and fourth tabs visualize f_3- and f_4- statistics respectively.

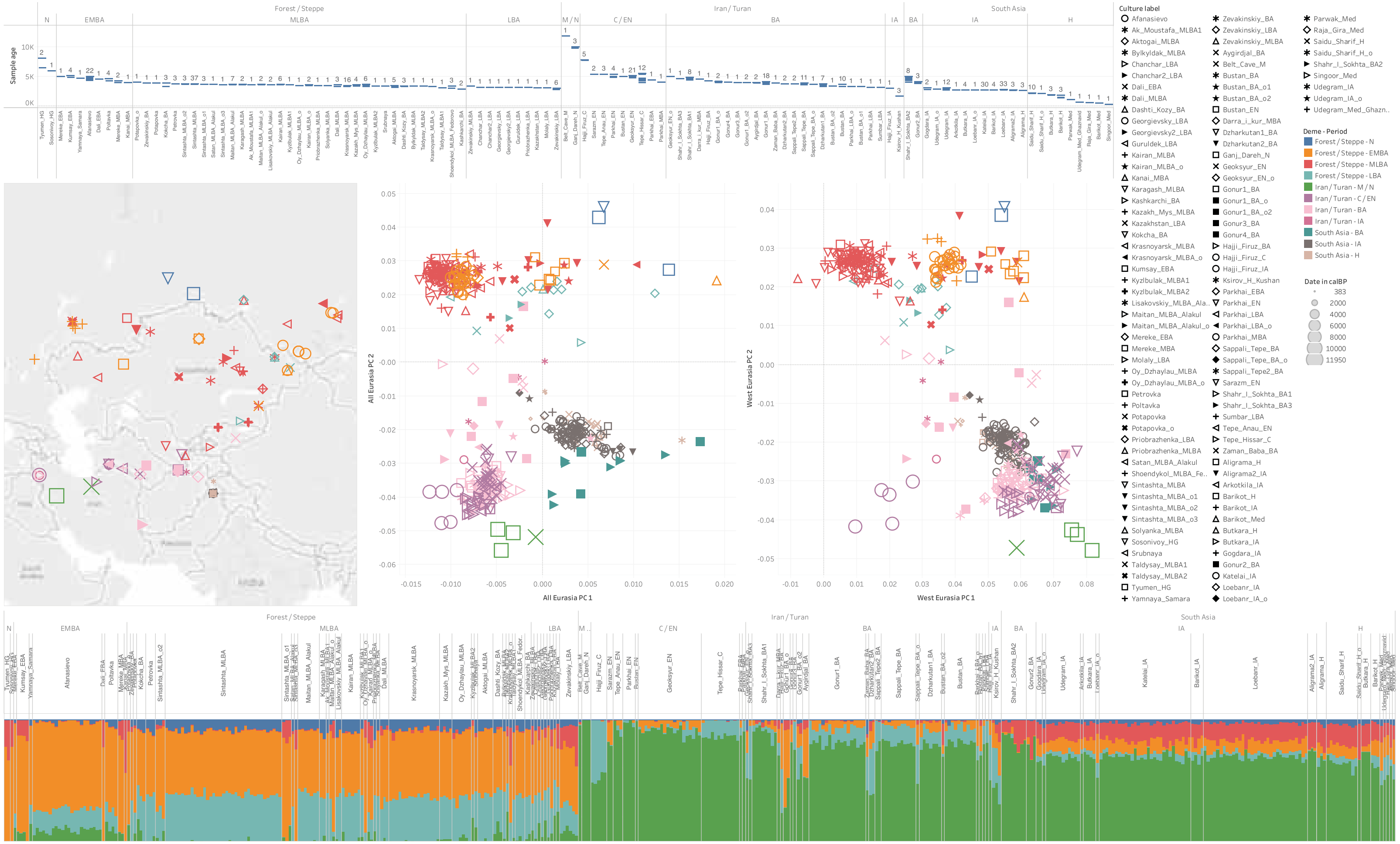

Fig S 1 Online Data Visualizer Tab 1 - Qualitative genetic results

Top: The number and age distribution of individuals. Radiocarbon dates or archeological context dates are shown.

Middle Left: Geographic location color0coded by broad chronological / archaeological period and area. Marker size is proportional to age.

Middle Center: First two principal components of all individuals projected on a basis of all samples from Eurasia.

Middle Right: First two principal components of all individuals projected on a basis of samples from West Eurasia.

Y chromosome haplogroups

	Forest / Steppe			Iran / Turan				South Asia		
	EMBA	MLBA	LBA	M/N	C/EN	BA	IA	BA	IA	H
E1a									0.021	
E1b						0.061			0.149	
G						0.030				
G2a						0.030			0.021	
H1a								0.333	0.085	
I2a					0.050				0.043	
J					0.250	0.091		0.167	0.021	0.083
J1					0.150	0.030				0.250
J2									0.021	
J2a		0.014			0.100	0.333		0.333	0.064	0.083
J2b					0.100				0.021	
L1									0.021	
L1a						0.091			0.234	0.083
L2					0.050					
N1a	0.111									
P						0.030				
Q					0.050					
Q1a	0.111	0.072	0.100			0.030			0.021	
Q1b						0.030			0.043	0.083
R		0.043			0.050	0.030		0.167	0.021	
R1		0.072								
R1a		0.725	0.800				0.333		0.043	0.333
R1b	0.778	0.072	0.100			0.061	0.333			
R2							0.333		0.021	
R2a				1.000	0.100	0.121			0.149	0.083
T					0.050	0.030				
T1a					0.050					

mtDNA haplogroups

	Forest / Steppe				Iran / Turan				South Asia		
	N	EMBA	MLBA	LBA	M/N	C/EN	BA	IA	BA	IA	H
HV1						0.116	0.068			0.069	
M65										0.069	
M30									0.111	0.057	0.143
W3a				0.067		0.023	0.017		0.222	0.057	
M5a										0.046	0.048
U2b										0.046	0.048
U2c									0.111	0.046	
H						0.023				0.034	
R30							0.017			0.034	0.048
T1a		0.028	0.136	0.067			0.017			0.034	
U7a					0.250	0.023	0.068			0.034	
I1c						0.047	0.017			0.023	
M										0.023	
N1a			0.016	0.067						0.023	
T2g						0.023				0.023	
U1a						0.047	0.068		0.111	0.023	
U2e	0.333		0.120				0.017	0.500		0.023	
U7b										0.023	
U8b										0.023	
A+1										0.011	
D4j										0.011	
H13			0.008			0.023				0.011	0.095
H14							0.017			0.011	
H2a		0.028	0.024				0.051			0.011	
H3a										0.011	
HV2						0.023	0.068			0.011	
HV6			0.008							0.011	
J1b			0.024			0.023	0.068			0.011	0.048
J1d						0.093				0.011	
K1a			0.040			0.047	0.034	0.250		0.011	
M35										0.011	

Allele frequencies of phenotypic markers

	Forest / Steppe			Iran / Turan				South Asia		
	EMBA	MLBA	LBA	M/N	C/EN	BA	IA	BA	IA	H
ASIP/PIGU-rs2378249	0.741	0.865	0.889	1.000	0.875	0.931		0.000	0.893	0.733
EDAR-rs3827760	0.115	0.056	0.167	0.000	0.000	0.000	0.000	0.000	0.024	0.048
EXOC2-rs4959270	0.448	0.405	0.154	0.000	0.192	0.250	0.000	0.000	0.263	0.350
IRF4-rs12203592	0.111	0.038	0.100	0.000	0.043	0.000	0.000	0.000	0.000	0.000
KITLG-rs12821256	0.000	0.037	0.000	0.000	0.000	0.042		0.000	0.059	0.083
LCT-rs4988235	0.000	0.054	0.083	0.000	0.000	0.000	0.000	0.000	0.025	0.050
MC1R-rs885479	0.034	0.047	0.000	0.000	0.074	0.054		0.000	0.105	0.048
MC1R-rs1110400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MC1R-rs1805005	0.000	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.050
MC1R-rs1805006	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MC1R-rs1805008	0.033	0.022	0.083	0.000	0.000	0.026	0.000	0.000	0.023	0.048
MC1R-rs1805009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MC1R-rs2228479	0.091	0.108	0.077	0.000	0.043	0.083	0.000	0.000	0.091	0.111
MC1R-rs11547464	0.000	0.020	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
OCA2-HERC2-rs12913832	0.167	0.590	0.083	0.000	0.000	0.186	1.000	0.333	0.150	0.150
OCA2-rs1800407	0.074	0.118	0.000	0.000	0.083	0.000	0.000	0.000	0.056	0.045
SLC24A4-rs2402130	0.917	0.753	0.750	1.000	0.760	0.771		1.000	0.902	0.944
SLC24A4-rs12896399	0.250	0.294	0.667	1.000	0.004	0.100			0.238	0.273
SLC45A2-rs28777	0.222	0.755	1.000	0.000	0.214	0.231		0.500	0.242	0.158
SLC45A2-rs16891982	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TYR-rs1042602	0.000	0.253	0.357	0.000	0.040	0.020	1.000	0.000	0.073	0.100
TYR-rs1393350	0.036	0.124	0.091	0.000	0.083	0.111	0.000	0.000	0.026	0.050
TYRP1-rs683		0.333	0.000		1.000	0.500			0.000	

Fig S 2 Online Data Visualizer Tab 2 - Frequency distribution of uniparental markers and selected phenotypically relevant markers

Left Panel: Y chromosome haplogroup frequencies by broad chronological / archaeological period and geographic area.
Middle Panel: mitochondrial DNA (mtDA) haplogroup frequencies by broad chronological / archaeological period and geographic area.
Right Panel: Allele frequencies of selected phenotypically relevant markers by broad chronological / archaeological period and geographic area.

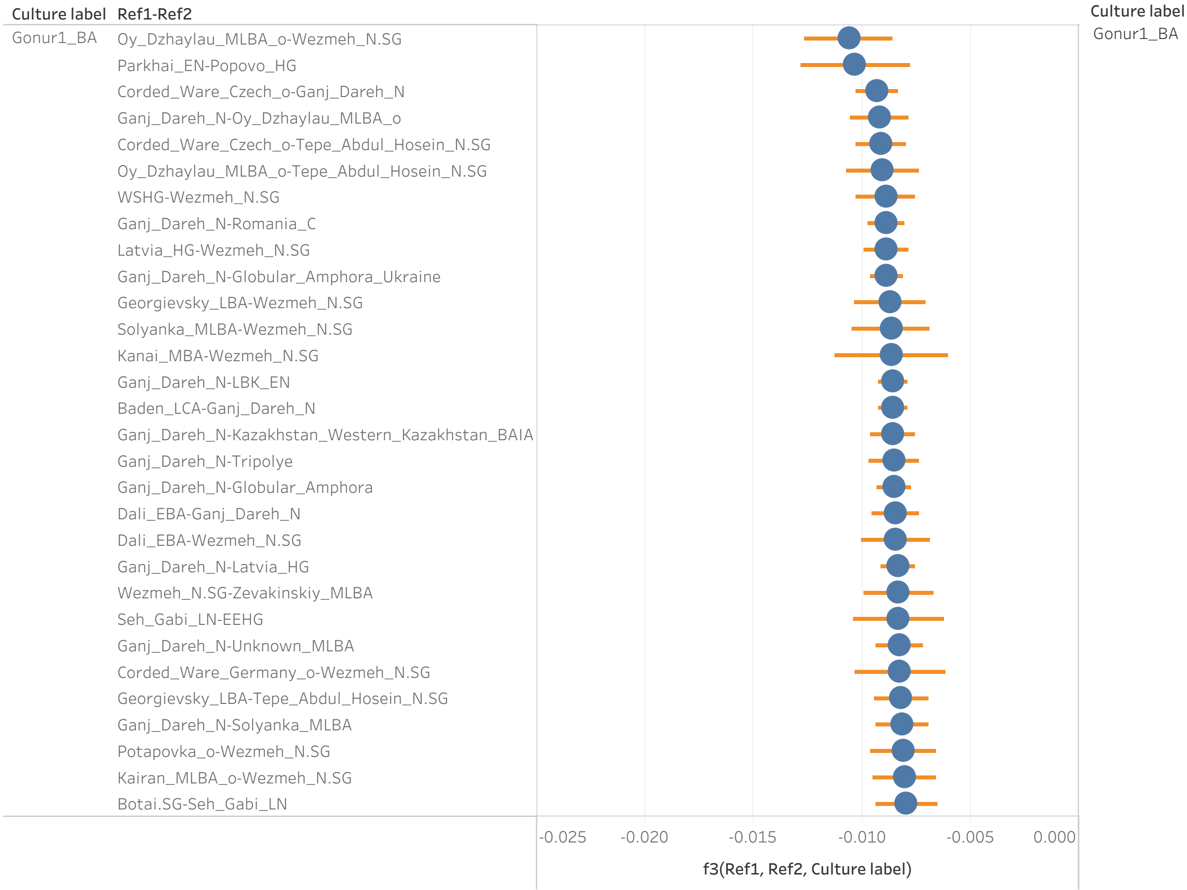

Fig S 3 Online Data Visualizer Tab 3 - Top admixture-f_3 statistics for all populations

Notes: For each population we show the mean (blue circles) for the most highly significant admixture-f_3 statistics for a particular population compared to all other pairs of populations in our dataset. Orange bars show one standard error. This visualization is restricted to significant (Z>2) results, and groups with at least 2 individuals with at least 100,000 SNPs.

Interactivity: Mousing over the blue circles reveals the number of SNPs used for the statistic. The drop down menu on the right accesses other statistics involving that population.

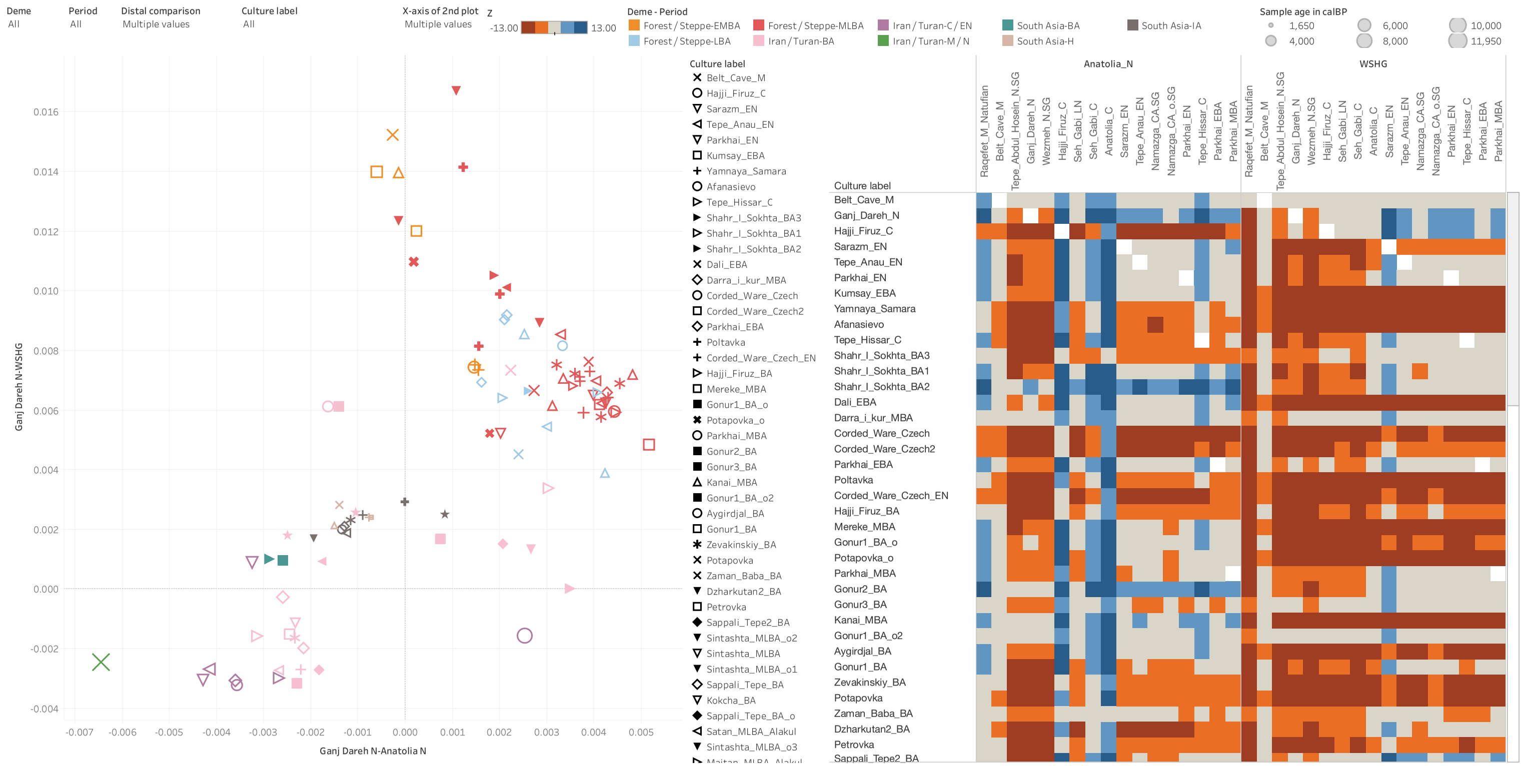

Fig S 2 Online Data Visualizer Tab 4 - All possible computations of f_4-statistics

Left panel: "Pre-Copper Age Affinity" f_4-statistics plotting *f_4(Ethiopia_4500BP.SG, Test; PreCopperAge1, PreCopperAge2)* vs. *f_4(Ethiopia_4500BP.SG, Test; PreCopperAge3, PreCopperAge4)*. Sizes of points are proportional to the sample age, and colors represent chronological / archaeological period and geographic area.

Right panel: "Two Population Comparison" statistics of the form *f_4(Ethiopia_4500BP.SG, PreCopperAge; Test1, Test2)* showing differences in ancestry between pairs of *Test* populations in comparison to a reference population. Orange squares indicate |Z|>3 for sharing more alleles with the *Test* population compared to a population from the columns, with respect to an outgroup. Blue squares indicate significantly less sharing at the same threshold. Darker orange and blue squares show sharing at |Z|>6.

Interactivity: Mouse over reveals the number of SNPs involved in the computation of each statistic as well as metadata associated with each population from Error! Reference source not found.. The drop down menus on the top left allow the selection of the Pre-Copper Age populations as well as an "easy viewing" option that allows selection of all individuals by broad chronological / archaeological period and geographic area.

S1 Archeological descriptions of 523 newly reported ancient individuals

S1.1 Overview

We present genome-wide data from 523 ancient individuals for which such data have never previously been reported, and also add data for 19 previously reported individuals. The total analysis dataset of ancient individuals consists of the 916 listed in Error! Reference source not found. after merging with data from previously reported individuals. The ancient individuals are also summarized in **Fig. 1** and **Supplementary Materials** Error! Reference source not found..

We also report 269 direct AMS radiocarbon dates. For individuals with direct dates, we use the following format: 2452-2140 calBCE (3815±35 BP, Poz-83491), giving the laboratory code, conventional uncalibrated radiocarbon date, and 95.4% calibrated confidence interval obtained using OxCal4.2 (*92*) and the IntCal13 calibration curve (*90*). For some individuals we also add additional notes in cases there is a correction for marine reservoir effect or if there is another issue that is important to take into account when considering the calibrated date.

For individuals without a direct date, we use a range from archaeological context in the format “2500-1700 BCE” (the lack of “cal” indicates a date from archaeological context).

When a single date is used to represent an individual, we use the average of the high and low extremes of the ranges (acknowledging that this is an imperfect summary of the true probability distribution which may be multi-modal). When a single date is used to represent a group, we average the point estimates for all individuals in our analysis dataset.

The individuals for which we have obtained data come from a mix of freshly excavated individuals as well as those that are from museum collections. In some cases, information about the specific graves associated with an individual from a site is either not indicated or is no longer readable on the original descriptions written from archeological excavations from the previous century. In such settings, we list the descriptions associated with the burials but a [?] to indicate that we no longer have the specific grave number within the site.

In what follows, we provide additional archaeological details for the 523 individuals for whom we generated genome-wide ancient DNA data. This list is organized into 3 main sections, one for each geographical area: Forest Zone / Steppe, Iran / Turan and South Asia. These are then organized into the various time periods for which we have data in each of these geographical areas and grouped by their genetic clusters within them. For each reported individual, we present a bullet point summary (also given in tabular form in **Table S 1**). In the event that a single site contains individuals through a range of time periods, we provide the site descriptions in order of time, with the oldest individuals described first.

Beyond the 523 reported individuals that have not previously been reported in the literature, there are 19 individuals for whom data have previously been reported and for whom we add additional data in the present study:

- 13 of these individuals are from sites where all the individuals for which we are reporting genome-wide data have other previously published data (*19*) and hence we do not present archaeological summaries for these sites and associated bullet points listing the individuals. These sites are all in the Samara region of Russia and include Barinovka I (n=1), Spiridonovka II (n=3), Spiridonovka IV (n=5), Utyevka VI (n=2), and Uvarovka I (n=1). We refer readers to (*19*) and **Table S 1** for more information on these sites and individuals.

S2.2 Individuals from the Forest Zone / Steppe

S2.2.1 Individuals from the Forest Zone in the Neolithic period

S2.2.1.1 <u>West Siberian Hunter-Gatherers</u>

These individuals fall in the *WSHG* analysis label.

S2.2.1.1.1 <u>Sosnoviy-Ostrov, Tyumen Oblast, western Siberia, Russia (n=1)</u>
Description by Sergey M. Slepchenko and Dmitry Enshin

The main site of Sosnovyi Ostrov occupies the northwestern part of Sosnovyi Island, which is a sand dune towering above the level of the marshy part of Lake Sredniy Tarman (3 km south from the village of Yamino, Nizhnetavdinsky district, Tyumen Oblast) in the forest zone of the Middle Trans-Urals. Research at the site was carried out in 1966 under the guidance of V.D. Viktorova, and in 2004 by a team of researchers of the Institute of the Problems of Northern Development, Siberian Branch of the Russian Academy of Sciences and Ural State University. During the work, outlines of several constructions were deduced, and three cultural and chronological groups were distinguished: the Middle Neolithic (Kozlov Mys culture, 6th millennium BCE), the Late Neolithic (Sosnovyi Ostrov culture, 5th millennium BCE), and the Eneolithic (Bairyk culture, 4rd millennium BCE). The findings include fragments of ceramic vessels and stone tools. In addition, excavations revealed three human burials. Based on the analysis of objects from them (ceramic vessels) and on stratigraphic data, it was provisionally determined that the burials correlate with the period of the existence of the Sosnovyi Ostrov culture. These individuals were part of the anthropological collection of Tomsk State University and the inventory numbers from the collection are provided along with that of the burial.

• Tomsk10, inventory number: 3079, burial 1 (I5766): Date of 4230-3983 calBCE (5261±33 BP, OxA-33486, d15N=+12.8 permil possible marine influence). Genetically female.

S2.2.1.1.2 Mergen Settlements, Tyumen Oblast, western Siberia, Russia (n=2)

Description by Sergey M. Slepchenko and Dmitry Enshin

The settlement of Mergen 6 is located 6 km southwest of Ishim city, on the northeastern shore of Lake Mergen in the forest-steppe zone of Western Siberia (Ishim district, Tyumen Oblast). The site is located on a peninsula between the lake and the Mergenka channel. The research was carried out during seven seasons from 1990 to 2011 (the research supervisors were V.A. Zakh, S.N. Skochina, D.N. Enshin). As a result of the work, it was determined that this part of the lake coast had been inhabited for a long time (Neolithic, Eneolithic, and the transitional time from the Bronze Age to the Early Iron Age). However, the main complex of materials at the site correlates with the period of the Early Neolithic (late 7th millennium BCE). At that time, bearers of the Boborykino and Koshkino cultural traditions built a settlement that was planned as a rough circle (13 buildings). The dwellings were from 35 to 100 m^2 with a post-and-frame structure. Ceramic vessels, and numerous stones, bone and horn artifacts (weapons, tools, etc.) are among the findings. In addition, a significant paleozoological collection correlates with that time, including bones of elk, deer, bear, wolf, fox, beaver, goose, duck, and Cyprinidae fish. On the whole, the findings allow us to characterize economic activities of the villagers as complex; they were hunters and fishermen. The ritual complex at the settlement, represented by burials of animal and fish parts (skulls, jaws, etc.), by tools for the main craft activities coated with ocher and white clay, and by human burials in dwellings are a distinctive feature of the burials from this time (*93–95*).

The settlement of Mergen 7 is located 7 km southwest of Ishim city, on the second lake terrace of the northeastern shore of Lake Mergen, in the forest-steppe zone (Ishim district, Tyumen region). There are 11 housing pits within the territory of the site. Excavations were carried out in 2011 under the guidance of D.N. Enshin. The two-chamber dwelling 1 (total area of about 120 m^2) was completely explored during the excavations. Two more constructions were partially examined. As a result of the research, it was determined that the site is multilayered (within the Neolithic - Eneolithic). Small groups of ceramics similar to the Boborykino, Kokuy, Ekaterininskiy and Andreevskiy types were distinguished. The complex associated with dwelling 1 is the most substantial, with direct analogies in the Kozlov culture of the Trans-Urals. The findings are represented by ceramic vessels and numerous stone artifacts. The lifetime of dwelling 1 was determined by a series of radiocarbon dates within the mid-second half of the 5th millennium BCE (Neolithic). Additionally, burial ground 1 was discovered near the housing complex and the remains of two women were found at the bottom. Chaotic arrangement of the bones suggests a secondary burial. One woman was 30 to 40 years old, the other one was 17 to 21 years old. During site clearance, fragments of two skulls, two jaws, ribs, an ulna, phalanges, pelvic bones, a femur and a tibia were found. There was an accompanying inventory in the grave, placed near the waists of the deceased. Two plain knives, an article made of an animal rib and a fragment of a tool were found near skeleton 1; a serrated arrowhead with a groove and a fragment of a chisel were found near skeleton 2. Four tools were deliberately broken in antiquity. The interpretation is that this inventory reflects tools related to leather production, wood processing and hunting. Preliminary research suggests that this burial dates to the Neolithic period (*96*, *97*), an interpretation that is confirmed by two direct radiocarbon dates. These samples were all part of the anthropological collection of the Tyumen Scientific center.

• Tyumen1, Kurgan 1, Mergen 6 (Building No. 21 - burial 1) (I1958): Date of 4723-4558 calBCE (5805±25 BP, PSUAMS-2359). Genetically female. This was a child burial.

• Tyumen50, Kurgan 6, Mergen 6 (Building No 15) (I1960): Two dates whose calibrated span is 6361-6071 calBCE [6335-6071 calBCE (7330±40 BP, Poz-82198), 6361-6086 calBCE (7355±40 BP, OxA-33489, d15N=+15.3 permil possible marine influence)]. Genetically female. This burial was a human skull in the center of the dwelling near the hearth.

S2.2.2 Sites from the Steppe Zone in Early to Middle Bronze Age Cemeteries

S2.2.2.1 Kumsay (Kyryk Oba), Kazakhstan (n=4)
Description by Arman A. Bissembaev, Aslan M. Mamedov, M.N. Duisengali, Talgat B. Mamirov, E. P. Kitov and Michael Frachetti

The cemetery of Kumsay (or Kyryk Oba) is located 11 km south-southeast of the town of Uil, Kazakhstan, in the Aktobe province. The burials are located on ancient terraces of the Uil River. The burial ground consists of over 168 earthen mounds distributed over a flat area spanning roughly 400 × 500 m. It has been hypothesized to be a southern outlier site of the Yamnaya culture.

These individuals fall within the *Central_Steppe_EMBA* analysis label and the *Kumsay_EBA* split label.

• Kurgan 6, Burial 10 (I11501): Date of 3341-3097 calBCE (4495±25 BP, PSUAMS-4604). Genetically female. Kurgan 6 is located in the eastern portion of the cemetery. The earthern mound had a diameter of roughly 26m and a height of 1.75m. Fifteen burials were recovered from kurgan 6, reflecting a range of grave orientations and positioning of the skeletons, including flexed and extended interments. Burial number 10 was found in the central part of the mound and contained an adult female lying on the right side in a flexed position. A fragmented child’s skull was also recovered at the foot of the adult skeleton. The second infant skeleton was not sampled for ancient DNA. This individual is genetically detected to be a 2nd to 3rd degree relative of all other individuals from Kumsay for which we have obtained data.

• Kurgan 4, Burial 1, skeleton 1 (I11531): Date of 3087-2919 calBCE (4385±20 BP, PSUAMS-4617). Genetically female. Kurgan 4 is located in the eastern part of the cemetery, was found in a disturbed condition. The mound was roughly 1.5 m high with a diameter of 26m. Six discrete burials were recovered under the mound. Burial 1 contained two individuals, an adult and an infant. Individual 1 was an adult placed on its back in a compressed flexed position within a small shallow pit, head to the southwest. All 4 individuals from Kumsay from which we have obtained data are genetically detected to be a 2nd to 3rd degree relative of each other.

• Kurgan 4, Burial 1, skeleton 2 (I11734): Date of 3022-2911 calBCE (4360±20 BP, PSUAMS-4619). Genetically male. This individual was an infant 0-6 months old, whose skeletal remains included a highly fragmented skull and a few teeth (all poorly preserved) and was buried in the same Kurgan as I11531. All 4 individuals from Kumsay from which we have obtained data are genetically detected to be a 2nd to 3rd degree relative of each other.

• Kurgan 1, Burial 2 (I11732): Date of 3078-2912 calBCE (4365±20 BP, PSUAMS-4751). Genetically female. Excavated in 2010, Kurgan 1 is located in the eastern part of the cemetery. The kurgan was constructed as an earthen mound, round in shape with a diameter of 33m and a height of 1.7m. Burial 2 was located on the southeast portion of the mound, a mere 52 centimeters below the general land surface. The skeleton was lying on its stomach (anterior side down), with the head facing southwest. The skeleton was apparently covered in a dark reddish substance, possibly ocher. No artifacts were recovered with the skeleton. Osteological analysis suggests that the interred individual was a woman, roughly 35-45 years old. All 4 individuals from Kumsay from which we have obtained data are genetically detected to be a 2nd to 3rd degree relative of each other.

Relatedness summary:

• 4-person family: All four individuals from Kumsay (I11501, I11734, I11732, and I11531) are genetically detected to be 2nd to 3rd degree relatives of each other.

S2.2.2.2 <u>Mereke, Kazakhstan (n=3)</u>

Description by A.P. Krygin, U.A. Utepbaev and M. Frachetti

Mereke is a mound burial located in the Taskala district (Zapadno-Kazahstanskaya Oblast, Kazakshtan). Within Kurgan 1 there are 17 burials in all, most dating to the Bronze Age. The mound itself measures 1.75m high with a diameter of 34m. More intact burials in Kurgan 1 contained skeletons positioned on their left side in a flexed position, head to the north and facing eastward. A notable feature of the kurgan was a ritually interred horse head placed at the southeastern sector of the mound. On the basis of the material inventory from burials 14 and 17, and details of the cemetery as a whole, the authors attribute Mereke 1 to the "Poltavka culture". Given the practice of large mound burials, sometimes as large as 300 cubic meters and the material assessment of the traits and artifacts recovered, Mereke 1 was dated stylistically to the second half of the 3rd millennium BCE.

These individuals fall within the *Central_Steppe_EMBA* analysis label and the *Mereke_MBA* split label.

- Kurgan 1, burial 2 (I11735): Date of 2462-2299 calBCE (3885±20 BP, PSUAMS-4901). Genetically male.

- Kurgan 1, burial 10 (I11737): Date of 2195-2032 calBCE (3710±20 BP, PSUAMS-4902). Genetically male.

This individual is an outlier from the main cluster of individuals and has an earlier radiocarbon date, so we assign it an analysis label of *Mereke_EBA_Yamnaya*.

- Kurgan 1, Burial 4 (I11736): Date of 3307-2928 calBCE (4425±20 BP, PSUAMS-4944) BCE. Genetically female. Burial 4 was located in the northwest sector of the kurgan. Oriented north-to-south, the burial was roughly 70 cm deep and measured 58×37 cm. The rounded grave pit contained a poorly preserved skull, placed into the cist with the crown pointing upward and the face oriented to the north. This individual is from an earlier period than most of the other individuals from the same site based on its radiocarbon date, and is genetically similar to samples from the Yamnaya and Afansievo cultures.

S2.2.2.3 Dali, Bayan-Zherek Valley, Kazakhstan (n=1)

See the section “Steppe Zone MLBA” for a description of this site and individual (we have data for a total of three individuals from Dali, 3 from the MLBA and 1 from the EBA). As this individual’s ancestry is distinctive we use the specific analysis label *Dali_EBA.*

S2.2.2.4 Yamnaya Pastoralists from Samara, Russia (n=1)

Some of the oldest radiocarbon dates from Yamnaya culture graves have been obtained from Yamnaya graves in the Samara oblast. The sample published here is not among the oldest dates but adds to the body of Yamnaya individuals from this region. Based on its genetic affinities, we designate this individual as having the *Western_Steppe_EMBA* analysis label and *Yamnaya_Samara* split label.

- 58, Utyevka V, 1/1 (I7489): Context date of 3300-2500 BCE. Genetically female. The area around Utyevka has more Bronze Age kurgan cemeteries than any other part of the Samara River valley. The cemetery designated Utyevka V, northeast of the modern village, was excavated by P. Kuznetsov in 1995. By grave ritual and form, kurgan 1, grave 1 appears to be a Yamnaya culture grave. In the grave was a female 30 to 35 years old and a child with red ochre. The sampled individual was the adult female.

S2.2.2.5 Afanasievo pastoralists of the Altai mountains and Yenisei river basin (n=26)
Overview description by James Mallory and Svetlana Svyatko

We present genome-wide data from 26 individuals from the Afanasievo culture (approximately 3100-2500 BCE), adding to the 6 individuals previously reported from this culture (*8*). In addition, we add data from one of the individuals from this culture that was previously reported (independent ~1.2 million SNP enrichment analysis for an individual for which ancient DNA data was previously generated by shotgun sequencing). All these individuals are labeled in the *Afanasievo* genetic cluster for analysis (which we distinguish from the genetically extremely similar *Western_Steppe_EMBA* analysis label in which the Yamnaya culture individuals largely fall).

The Afanasievo culture is associated with the eastern spread of domestic livestock (sheep, cattle, and possibly horses), copper metallurgy, and tumulus burials from the European steppe to southern Siberia. The sites are clustered in two main regions: the Altai Mountains where there are about 77 known cemeteries, 40 settlements and several ritual and mining sites, and the Middle Yenisei / Minusa Basin region where there are approximately 35 known cemeteries and 12 settlements. On the evidence of material culture, ritual behavior, physical anthropology and ancient DNA, the ancestry of the people of the Afanasievo culture seems to have been spread by long-distance migration from the European steppe (Yamnaya and related cultures) to the east.

S2.2.2.5.1 Saldyar-1 cemetery, Altai Mountains, Russia (n=1)

Description by James Mallory, Svetlana Svyatko, and Nadezhda Stepanova

One individual came from the Saldyar-1 burial ground in the Altai Mountains, located approximately 6 km from the Malyi Yaloman village on the right bank of the Katun River, in a valley bounded from the north, east, and south by ridges of the Saldzhar (Saldyar) Ridge, and from the west by the Katun River. Excavations of the site were undertaken by O.V. Larin in 1988-1991 (*98, 99*). The Saldyar-1 site differs from other Afanasievo cemeteries in the sense of having a relatively small number of grave goods. Within the cemetery, most structures have a ring made of vertically positioned slabs of large stones and blocks, or of fragmentary stones. The burials at the site were in ground pits, about 25% of them overlaid by massive slabs. The dead were on their backs with knees bent, and sometimes on their right sides with their heads mainly oriented towards the west. Goods were found only in half of burials (pointed, round and flat clay vessels, pestles, arrowhead, beads, buttons, etc.).

- SSSal1/36 (I1829): Calibrated span of dates of 3316-2915 calBCE [Date of 2915-2878 calBCE (4270±25 BP, PSUAMS-2109) and previously reported date of 3316-2915 calBCE (4409±34 BP, UBA-29307)]. Genetically female. Kurgan 36 is an enclosure made of upright slabs. The grave pit, overlaid with massive plates, was located in the center of the enclosure. At the bottom (at a depth of 1.3 m compared to the natural ground level), three skeletons were found: a 14 to 17-year-old teenager (SSSal1/36, the one from whom ancient DNA was successfully obtained) and two children of approximately 4 and 1.5-2 years old. The individuals were on their backs with heads turned towards the west, knees bent, legs collapsed to the right, and hands slightly bent at the elbows. Jewelry made of animal teeth was found at their chests, and a small accumulation of coal was found at the legs of an adult.

S2.2.2.5.2 Ust'-Kuyum cemetery, Altai Mountains, Russia (n=1)

Description by Anatoly Derevianko and Tatiana Chikisheva

One individual came from the Ust'-Kuyum cemetery, located on the left side of the Kuyum River close to its confluence with the Middle Katun at the Iolgo Ridge, 25 km from the beginning of the Chemal highway. The cemetery represents separate tombs marked by stone pavements. Tomb 10 was excavated by Elisaveta Bers in 1965. On the basis of the burial features, the tomb was attributed to the Afanasievo culture.

- 230/6, tomb 10 (I2071): Calibrated span of dates of 3331-2704 calBCE [date of 2926-2704 calBCE (4260±35 BP, Poz-83510) and previously reported dates of 3322-2923 calBCE (4423±29 BP, OxA-31219) and 3331-2935 calBCE (4442±29 BP, OxA-31220)]. Genetically male and skeletally male. The body was placed in a stone cist built with vertical slabs supported by a stone pavement, 0.4 m deep from the ancient surface. The cist was 1.3 – 1.5 m long, 0.87 m wide and 0.6 m high and covered with capstones. The man was placed in a supine position with flexed legs, head turned towards the southwest. Traces of ocher were in the soil filling the cist.

S2.2.2.5.3 Kaminnaya Cave, Altai Mountains, Russia (n=1)

Description by Anatoly Derevianko and Tatiana Chikisheva

One individual came from Kaminnaya Cave, from a grave excavated by Yuri Grichan in 1983. Kaminnaya Cave is in the Ust'-Kanski District in the Altai Autonomous Republic, located at 1100 m above sea level in the Karakol River valley in the northwestern part of the Altai Mountains.

- 230/4 (I2069): Date of 3331-2922 calBCE (4430±40 BP, Poz-83425). Genetically and skeletally female. The woman was placed in an earthen grave in a flexed position on the left side with head turned north-northwest. The bones were pigmented with ocher, and the tomb was attributed to the Afanasievo culture based on features of its burial rites.

S2.2.2.5.4 Karasuk III, Russia, cluster B, kurgans 1 and 2 (n=8)

Description by Vyacheslav Moiseyev

A total of seven individuals came from the Karasuk III site, a group of 11 kurgans excavated by M.P. Gryaznov in 1961-1963 and extending 350 meters along the right bank of the Karasuk tributary of the Yenisey River. The kurgans form three clusters, which may be independent burial places (*99*). All individuals with DNA are from cluster B, which includes two kurgans.

Kurgan 1 (Graves 1, 2, and 3): The structure of Kurgan 1 is similar to that of other Afanasievo burial places. At the base of the kurgan there is a stone-circled enclosure with several grave pits inside. The enclosure of Kurgan 1 is made of four to five rows of stone plates with a diameter of about 10 m and walls of 0.6 m width and 0.25 m high. In the center of the enclosure there are two grave pits. One is about 7 m^2 (2.8 x 2.6 m) and another about 3.5 m^2 (1.8 x 1.8 m). The depths of the graves are about 1 m from the enclosure floor. The third shallower grave is joined to the wall of the enclosure. Graves 1 and 2 are archaeologically defined as summer graves and grave 3 is a winter one (*100*).

Grave 1 contained the remains of three individuals, and we obtained DNA from all 3 who turned out to be a father and his two sons. Pieces of copper, polypod vessels (kuritel'nitsy) and rare ornaments on egg-shape pottery vessels may point to a high rank of individuals buried in the grave:

- StPet48, collection 6612, individual 2 (I3950): Date of 2878-2636 calBCE (4160±25 BP, PSUAMS-1955). Genetically male. This individual is skeletally a male of about 35 to 40 years old. This individual is father of two sons buried in the same grave (I3949 who was anthropologically determined to be 6 to 8 years old, and I6714 who was anthropologically determined to be 16 to 18 years old). He is the brother of I3951.

- StPet47, collection 6612, individual 1 (I3949): Date 2837-2498 calBCE (4075±20 BP, PSUAMS-2292). Genetically male. Child of I3388 (his mother) and I3950 (his father), brother of I6714, and nephew of I3951. This individual, 6 to 8 years old, was buried in the same grave as his genetic father and brother.

- StPet49, collection 6612, individual 3 (I6714): Date of 2617-2472 calBCE (4020±25 BP, PSUAMS-3909) [son of I3950 at 2878-2636 calBCE (4160±25 BP, PSUAMS-1955)]. Genetically male. Son of I3388 and I3950, brother of I3949, and nephew of I3951. This individual, 16 to 18 years old, was buried in the same grave as his genetic father and brother.

• StPet50, collection 6612, individual 4 (I3951): Date of 2879-2639 calBCE (4165±25 BP, PSUAMS-1956). Genetically female. This individual was buried in Grave 2 of Kurgan 1, the central grave of the kurgan, and is skeletally determined to be a female about 50 years old. This individual is sister of I3950, and aunt of I3949 (son) and I6714 (son). Skeletally a 50-year-old female.

We obtained DNA from one individual from Grave 2, the central grave of Kurgan 1:

• StPet51, collection 6612, individual 5 (I3952): Date of 2866-2579 calBCE (4120±30 BP, PSUAMS-1957). Genetically male. This individual was skeletally determined to be a male about 50 years old.

We obtained DNA from one individual from Grave 3, located at the periphery of Kurgan 1.

• StPet53, collection 6612, kurgan 1, individual 6 (I3388): Context date of 2700-2600 BCE [mate of I3950 at 2878-2636 calBCE (4160±25 BP, PSUAMS-1955), mother of I3951 at 2879-2639 calBCE (4165±25 BP, PSUAMS-1956), and mother of I6714 at 2617-2472 calBCE (4020±25 BP, PSUAMS-3909)]. Genetically female. Skeletally a 30 to 40 years old female. Two vessels, one ceramic and one wooden, were on the legs of the skeleton. Gryaznov suggested that the woman was killed by a bronze knife, which was found in her left scapula (*100*). Our visual studies of the scapula confirmed the presence of bronze oxide on its outer surface, but we found no evidence of damage to the bone structures. Thus, if the individual was truly killed, it was likely a ritual killing. Genetically this individual was the mother of I3949 and I6714 in Grave 1.

We also obtained DNA from two individuals from Kurgan 2 who are detected as 2nd or 3rd degree relatives of each other.

• StPet52, collection 6612, individual 8 (I3954): Date of 2872-2625 calBCE (4140±25 BP, PSUAMS-2293). Genetically female and a 2nd or 3rd degree relative of I6715. Skeletally this is a female of at least 50 years of age.

• StPet54, collection 6612, individual 7 (I6715): Context date of 3300-2500 BCE. Genetically female and a 2nd or 3rd degree relative of I3954.

Relatedness summary:

• 5-person family: Our analysis revealed that five of the individuals are part of a closely related family, consisting of I3950 (father) and I3388 (mother) who are the parents of I3949 and I6714. In addition, I3951 is the sister of I3950.

• 2-person family: Our analysis revealed that I3954-I6715 are 2nd to 3rd degree relatives.

S2.2.2.5.5 Afanasieva Gora, Middle Yenisei region, Russia (n=5)

Description by Vyacheslav Moiseyev

Genome-wide ancient DNA data from a total of 9 individuals from this site has now been reported: 3 individuals from (*8*) and data from 5 previously unreported individuals in this study (data from one individual: a female RISE509 / I10565, was generated independently in the two studies, and is not listed). We did not include data from RISE510 in the analysis in this study.

The Afanasieva Gora Eneolithic burial ground is located on the left bank of the Yenisey River near the abandoned village of Bateni in the Bograd district, Karasnoyarsk region, Russia. The site was excavated by S. A.Teploukhov in the early 1920s. The unique archaeological characteristics of the burials of Afanasieva Gora which belong to the earliest Siberian pastoralists made it a reference site for the Afanasievo culture after its discovery (*101*). Regrettably no photos or regular drawings were made during excavation. However, a short description of the graves is given in 124-125 and 301 of ref. (*99*) based on the field diary of Teploukhov. Some of the graves were also excavated in the 1960s by M. P. Gryaznov and these are well documented (*99*, *100*). The burial ground consists of 40 kurgans, 35 of which are attributed to the Afanasievo culture. Like most Afanasievo cemeteries, Afanasieva Gora consists of single burials as well as small collective burials with the deceased usually flexed on their backs in a pit. The burial pits were made in enclosures that were usually rectangular although sometimes circular. It has been argued that the burials represent family plots with four or five enclosures constituting the local social group which is supported by results of our study (see the description of the Karasuk III site). The large size

(3750 square meters) occupied by the Afanasieva Gora site and the existence of separate groups of burial features may reflect chronological differences in use of the burial ground so that the site can potentially be viewed as a series of separate Afanasievo cemeteries. See page 121 of ref. (*99*).

- StPet42, collection 6136, individual 4 (I6711): Context date of 2950-2650 BCE [based on being genetically detected as the son of previously reported RISE511 who has a date of 2909-2679 calBCE (4224±36 BP, OxA-31568)]. Genetically male.

- StPet45, collection 6136, individual 8 (I6712): Date of 2858-2505 calBCE (4095±25 BP, PSUAMS-3883). Genetically male. This individual is the son of I15065 (who is the same individual as the previously reported RISE509. This individual is also a 2nd to 3rd degree relative of I3387.

- StPet46, collection 6136, individual 10 (I6713): Context date of 3000-2000 BCE. Genetically female.

- StPet41, collection 6136, individual 3 (I3387): Context date of 2950-2600 BCE [based on being genetic detected as the sibling of the previously reported RISE511 who is directly dated]. Genetically male. In addition to being genetically detected as the sibling of the previously reported RISE511, this individual is genetically detected as a 2nd to 3rd degree relative of I6712, and as the uncle of I6711 (RISE511's son).

- StPet40, collection 6136, individual 2 (I10564): Context date of 3300-2500 BCE. Genetically female.

Relatedness summary:

- 5-person family: We detect an extended family that consists of two related nuclear families. In the first nuclear family, the previously reported RISE511 is the mother, I6711 is her son, and I3387 is her brother. In the second nuclear family, RISE509/I10565 and I6712 have a mother-daughter relationship. The two nuclear families are themselves related to each other through I3387 and I6712 who are detected as having a 2nd to 3rd degree relationship.

S2.2.2.5.6 Elo 1, Russia (n=3)

Description by Sergey M. Slepchenko and Dmitry Enshin

The burial ground Elo-1 is situated in the Altai Mountains, 1 to 1.5 km from the village of Elo, on the right bank of the Elo River, opposite the mouth of the Tobotoi River. The exact number of burial constructions is not determined. Two groups of structures are distinguished and located 0.5 km from each other. Five fences were excavated from 1976 to 1977 under the guidance of V.A. Posrednikov (three fences in the western group and two fences in the eastern group). Burial places within two fences showed evidence of robbery. The objects of the western group are distinguished by construction of sepulchral structures, position of the buried, and inventory. They include constructions made of vertically placed slabs with a diameter of 5.6 to 6.2 m. The deceased were laid on the right side with their knees bent, with their heads oriented in a southwestern or west southwestern direction and sprinkled with ocher. Each grave had one ceramic vessel. The eastern group of sepulchral structures consists of fences made of rubble laid horizontally with the deceased inside lying on their backs with their legs bent, oriented towards northeast and west southwest, and painted with ocher. Ceramic vessels, adornments of red deer teeth, and copper adornments were found in the burials. The materials of the burial ground are attributed to the Afanasievo culture of the Early Metal Age of the Altai Mountains, 3rd millennium BCE (*102, 103*). These individuals were all part of the anthropological collection of Tomsk State University and the inventory numbers from the collection are provided along with that of the burial.

- Tomsk_1950, inventory number 1950, burial 3 (I5269): Date of 3016-2899 calBCE (4335±25 BP, PSUAMS-2350). Genetically male.

- Tomsk_1951, inventory number 1951, burial No 4. (I5270): Date of 3322-2939 calBCE (4435±20 BP, PSUAMS-2405). Genetically male.

- Tomsk_1952, inventory number 1952, burial 7 (I5271): Date of 3013-2901 calBCE (4335±20 BP, PSUAMS-2406). Genetically female.

S2.2.2.5.7 Elo Bashi, Russia (n=2)

Description by Sergey M. Slepchenko and Dmitry Enshin

The burial ground of Elo-Bashi is situated in the Altai Mountains, 3.5 km from the village of Elo, on the left bank of the Elo River. The researchers of the site have identified about 30 objects, including 15 fences made of vertically placed slabs. The site is divided into several territorial groups without clear boundaries between them. Five Afanasievo fences made of vertically placed slabs and one fence of rubble stone were excavated from 1977-1978 and in 1980 by archaeological expeditions of Altai State University under the guidance of V.A. Posrednikov. All of the graves of the eastern group are covered with slabs and the remains of a preserved wooden floor was found in fences 4 and 5. The deceased were laid on their backs with their knees bent and their head oriented towards west and southwest; their bodies were painted with ocher. Ceramic vessels, metal and stone artifacts were found in the graves. The materials of the burial ground are attributed to the Afanasievo culture of the Early Metal Age of the Altai Mountains, 3rd millennium BCE (*104*). These individuals were all part of the anthropological collection of Tomsk State University and the inventory numbers from the collection are provided along with those of the burials.

• Tomsk_1955, inventory number 1955, burial 1, skeleton 2 (I5272): Date of 3003-2887 calBCE (4305±20 BP, PSUAMS-2351). Genetically female.

• Tomsk_2099, inventory number 2099, Kurgan 5 (I5277): Calibrated span of 3264-2929 calBCE [date of 3083-2916 calBCE (4375±20 BP, PSUAMS-2353) and date of 3264-2929 calBCE (4420±20 BP, PSUAMS-2368)]. Genetically male.

S2.2.2.5.8 Lower Tyumechin 1, western Siberia, Russia (n=3)

Description by Sergey M. Slepchenko and Dmitry Enshin

The burial ground of Nizhniy Tyumechin-1 is situated in the Altai Mountains, 0.5 to 0.6 km to the southeast of the village of Elo, on the right bank of the Kaerlyk River, near the place where it merges with the Elo River, forming the Ursul River. The funerary complexes are located in clusters forming a chain oriented in the northwest-southeast direction along the foot of the mountain. Fifteen fences of the Afanasievo culture were excavated during three field seasons (1977-1979) by expeditions of Altai State University under the guidance of V.A. Posrednikov. Construction of all graves is characterized by a ring of vertically placed slabs. Most of the bone chambers are covered with large slabs; remains of a wooden construction are found in two of them. As a rule, the deceased were put on their backs with their knees bent, more rarely they were placed on their right side, with their heads oriented to southwest, west southwest, and sprinkled with ocher. The findings in the burials are represented by ceramic vessels, and less often by stone, bone and metal artifacts. The site is distinguished by the following features: almost all of the graves are covered with slabs, the buried are oriented in a consistent direction, and the majority of them were buried with an inventory. Products made of animal horns are more often found within the fences or over the graves than at other burial grounds. The materials of the burial ground are attributed to the Afanasievo culture of the Early Metal Age of the Altai Mountain, 3rd millennium BCE (*104*). These individuals are all part of the anthropological collection of Tomsk State University and the inventory numbers from the collection are provided along with those of the burials.

- Tomsk_1959, inventory number 1959, burial 1, Lower Tyumechin 1 burial ground (I5273): Date of 3011-2887 calBCE (4310±25 BP, PSUAMS-2352). Genetically male.

- Tomsk_2101, inventory number 2101, burial 8, Lower Tyumechin 1 burial ground (I5278): Date of 3084-2911 calBCE (4370±25 BP, PSUAMS-2354). Genetically male. Genetically detected to be a 2^{nd} to 3^{rd} degree relative of I5279.

- Tomsk_2102, inventory number 2102, burial 9, Lower Tyumechin 1 burial ground (I5279): Date of 3012-2897 calBCE (4330±20 BP, PSUAMS-2355). Genetically male. Genetically detected to be a 2^{nd} to 3^{rd} degree relative of I52778.

Relatedness summary:
- 2-person family: I5278-I5279 are genetically detected to be 2^{nd} to 3^{rd} degree relatives.

S2.2.2.5.9 Krasnoyarsk Krai, Minusinsk, Podsukhanikha II (n=1)

Description by Tatyana M. Savenkova & Valery S. Zubkov

The Podsukhanikha site is located on an 8-10-meter floodplain terrace on the right bank of the Yenisei River, 15 km southwest of the village of Nikolo-Petrovka in the Minusinsk District of Krasnoyarsk Region (Russia). The burial ground was discovered by A. A. Gavrilova in 1955, at which time one enclosure was excavated in order to check its cultural attribution. In 1960, A. A. Gavrilova, as part of the Krasnoyarsk archaeological expedition, excavated six more enclosures, and a plan of the burial ground was composed. After the inundation of the Yenisei floodplain by the Krasnoyarsk Reservoir, the Afanasievo kurgans began to disintegrate during the maximum rise of water levels. In 1987, N. V. Leont'ev excavated a destroyed burial, and also, in 1995-1996, together with the expedition of the German Archaeological Institute, excavated 7 more graves in the destroyed and half-ruined enclosures. In 2000, V. S. Zubkov and K. G. Kotozhekov as part of the same expedition investigated the largest kurgan of the Afanasievo burial ground (No. 19A) (*105*). With an open-area excavation of 800 m^2, the remains of a ring-shaped dry-stone structure of flat-laid sandstone slabs was exposed, the width of which had expanded because of landslides. The outer diameter of the ring was about 17.5 m, the width of the ring was 1.3 m, and its height was 0.4 m.

In the center of the enclosure, a large earthen grave had been constructed. The dimensions of the grave were 3.5 x 2.4 m, with a depth from the level of the ancient surface of 1.4 m. Three burials were found in the grave pit: burial 1 was the primary burial and the earliest and is associated with the Afanasievo Culture, while burials 2 and 3 were more recent inhumations inserted later. Burial 1 had been looted in ancient times. Originally, it had been covered with wooden boards or logs, and on top of them sandstone slabs. In the infill of the grave and at its bottom were found scattered skeletal remains belonging to a man of mature age (determination by T. M. Reis-Savenkova). The accompanying inventory consisted of several fragments of large earthenware vessels (korchagi), 10 bone studs, a spindle-shaped bone object, and a drilled astragalus of a sheep. At the bottom of the pit lay an ovoid vessel and a stone pestle.

On the outer side of the enclosure, two rounded pieces of hearth foundation stonework were identified, also fashioned from sandstone slabs.

- Podsukhanikha II, kurgan 19A, burial 1 (I11112): Context date of 3000-2700 BCE [2926-2779 calBCE (4271±30 BP, Bln-5280), based on a sample of wood from this grave). Genetically male, and anthropologically a mature male.

S2.2.2.5.10 Inskoy Dol, Russia (n=1)

Description by Petr K. Dashkovskiy and Alexey A. Tishkin

The kurgan burial ground, designated as Inskoy Dol, was discovered by Petr K. Dashkovskiy, of Altai State University (Barnaul, Russia) in 2010, in the process of studying various archaeological sites within the Chinetinsky archaeological district located in the Krasnoshchekovsky district of the Altai Krai, Russia. The archaeological site of Inskoy Dol is located on the second terrace above the floodplain on the left bank of the Yin River (in the southern part of its valley), 2 km east-southeast of the village of Chineta. The river Hankara flows 0.8 km to the north of Inskoy Dol. Approximately 0.7 km to the southeast of the site are the remains of the foundation of the pioneer camp building located at the edge of the river bank.

The mounds of the Eneolithic period are found in the northern and southern parts of the Inskaya Dol burial site. At the level of the modern surface, no signs of the presence of archaeological objects were initially visible due to the high degree of sodding. In this case, the search for mounds was carried out using a metal probe. The use of this method made it possible to fix the presence of barrows No. 6–8 on a small elevation in the northern part of the site. The second group of burial mounds (No. 4–5, 9) of the same period was revealed in the southern part of the necropolis 200 m south of the previous one.

Mounds number 6-8 were investigated by a single excavation. After stripping the stone mounds of the kurgans, it was discovered that the southern edge of mound 6 and the northern edge of mound 7 overlapped each other, practically forming a single structure. The diameter of mound 6, which was predominantly composed of one layer of stones, reached 9 m. The height was 0.35 m. The external portion of the mound was composed of small and medium stones in the shape of a ring, 1.5–2 m wide. In the central part of the mound, stones were virtually absent. While excavating the stone ring in the northwestern section, an unornamented fragment of ceramic was found. After complete removal of the stone structure, a grave pit was revealed, which was oriented by a long axis with slight deviations along the southwest-northeast line. The grave pit had the following dimensions at the level of the ancient horizon: 2 x 1.2 x 0.94 m. A human skeleton was found in the grave, which was laid in an extended position on its back with its legs bent toward the left side and oriented with its head to the east. The skeleton was heavily sprinkled with red ocher. No items of accompanying inventory were found.

The features noted above are characteristic of the so-called Kurotin type of burials (highlighted by NF Stepanova). In particular, the material culture objects of the group of mounds No. 6–8 belong to the Saldyar culture of the Altai Afanasievo cultural and historical community. A radiocarbon date from a person buried in barrow No. 9 dates the objects to 2922-2704 calBCE (4255±35 BP, UBA-26406). This date fits the chronological span of the Afanasievo culture which includes the above interval.

A woman was buried in mound number 6. Anthropological definitions were performed by Ph.D. K.N. Solodovnikov (Tyumen).

- BARN-039, Inskoy Dol, kurgan 6 (I11752): Context date of 2900-2500 BCE. Genetically female.

S2.2.2.6 *Poltavka:* Pastoralists from the Poltavka culture (n=4)
Descriptions by David Anthony and Alexander Khokhlov

These individuals are all assigned to the *Western_Steppe_EMBA* analysis cluster.

S2.2.2.6.1 Grachevka cluster, Sok River, Samara Oblast, Russia (n=1)

Grachevka is one of six kurgan cemeteries dated to the Early and Middle Bronze Age (Yamnaya and Poltavka cultures) recorded within a 20-km section of the Sok River, north of Samara, between Grachevka and Orlovka villages. This unusually dense cluster of E/MBA cemeteries represents the northwestern border of their distribution in the Volga-Ural steppes, and thus marked a cultural and economic border. This cluster has previously reported some of the oldest radiocarbon dates for the Yamnaya culture generally (ca. 3300-3100 BCE) at Nizhnjaya Orlyanka I, k. 4/gr.2 and Lopatino I, k. 31/gr.1.

- 41, Grachevka II 2/1 (I6294): Date of 2876-2666 calBCE (4160±20 BP, PSUAMS-2956). Genetically male. Kurgan 2 at Grachevka II was excavated by P.F. Kuznetsov, A.A. Khokhlov, and O.G. Mochalov from Samara in 1998. Grave 1 contained a typical Poltavka ceramic vessel, a tubular bone handle or flute, and the bones of a male 50-55 years old.

S2.2.2.6.2 Samara River valley and steppes to the south (n=1)

The riverine marshes of the lower Samara River are 6-8 km wide in the Bobrovka district, providing valuable fodder for cattle and sheep through the cold winter. On the southern side of the Samara across from Bobrovka was a long-used series of kurgan cemeteries at Spiridonovka that contained unusually well-equipped graves and could have been the burying place of local chiefs. Bobrovka is a Middle and Late Bronze Age kurgan cemetery in the Kinel' region of Samara oblast excavated by V.A. Skarbovenko in 2006.

- Bobrovka kurgan 1/grave 1 (I7671): Date of 2885-2680 calBCE (4185±20 BP, PSUAMS-4275). Genetically male. Kurgan 1/grave 1 contained the skeleton of a male 50-60 years old. The skull is dolichocranial, of a southern "Europoid" type, according to A. Khokhlov. Genetically he is also male. The grave exhibits a Poltavka culture (MBA) funeral ritual and grave form, in agreement with the radiocarbon date.

S2.2.2.6.3 Zhuravlikha I, Bolshoi Irgiz River, Samara oblast (n=1)

Zhuravlivkha I is a kurgan cemetery in the arid steppes south of the Samara River in Samara oblast, about 100 km from the border with Kazakhstan. Kurgan 1 was excavated in 1998 by P. Kuznetsov.

- SVP HB 55, K. 1, gr. 10 (I0442): Context date of 2800-2000 BCE. Genetically female. Kurgan 1 contained a founding EBA Yamnaya grave (gr. 18) and a secondary MBA Poltavka grave (gr. 10). Grave 10 contained the skeleton of a woman aged 18-20 years.

S2.2.2.6.4 Nikolaevka III, Volzhskiy region, Samara Oblast (n=1)

Nikolaevka III is a kurgan cemetery located on the north side of the Volga River near its junction with the Sok River, a tributary of the Volga. It is not far from Grachevka II described below. The cemetery contained six Bronze Age kurgans and was excavated in 1995 by V.A. Skarbovenko.

• Nikolaevka III K.1 B.3 (I8745): Context date of 2500-2200 BCE. Genetically male. Kurgan 1, grave 3 contained a male skeleton aged 40-55. The grave is characterized by a Poltavka culture (MBA) funeral rite and grave form.

S2.2.2.7 Potapovka culture pastoralists from the Grachevka kurgan cemeteries (n=2)
Description by David Anthony and Alexander Khokhlov

As described above, Grachevka I and Grachevka II are kurgan cemeteries located south of the Sok River in the Samara oblast, 55 km northeast of Samara, Russia, described above. Some kurgans were constructed during the Early and Middle Bronze Age (EMBA) by the Yamnaya and Poltavka cultures, and others were added during the late MBA (MLBA) by the Poptapovka culture, closely related to the Sintashta culture.

A first individual is from Grachevka I, a kurgan cemetery located near Grachevka II on the Sok River, was excavated in 1998 by P.F. Kuznetsov, A.A. Khokhlov, and O.G. Mochalov from Samara. It is assigned on genetic grounds to the *Western_Steppe_MLBA* analysis label, consistent with other previously reported samples from the Potapovka culture and many other Steppe Middle to Late Bronze Age individuals. We assign this individual to the *Potapovka* split label.

• Grachevka I, kurgan 1, grave 1 (I7670): Date of 2121-1942 calBCE (3640±20 BP, PSUAMS-4274). Genetically male. Kurgan 1, grave 1 contained a male with grave type and artifacts of the Potapovka culture. Two culturally similar groups located southeast (Sintashta) and southwest (Potapovka) of the Urals had similar grave rituals, ceramics, projectile point types, and metal weapon types, and our genetic analysis shows that they were similar genetically as well.

A second individual is assigned to the *Steppe_MLBA_oWSHG* genetic analysis cluster, meaning that our genetic analysis reveals it as having ancestry similar to other western Steppe Middle to Late Bronze Age groups, with additional *WSHG*-related admixture.

• SVP62, Grachevka II, kurgan 5, grave 3 (I0244): Union of dates of 2465-1981 calBCE (previously published date of 2341-1981 calBCE (3752±52 BP, AA-53806) and previously published date of 2465-2054 calBCE (3815±60 BP, Le-6545)). Genetically female. Kurgan

5 was excavated by P. F. Kuznetsov, A. A. Khokhlov and O. G. Mochalov from Samara in 2002. Grave 3 is a female aged 35 to 45 years, oriented to the south, with grave type and artifacts of the Potapovka culture (MLBA). The radiocarbon date of this individual is among the earlier dates for the Potapovka culture, and genetically she is an outlier with respect to other individuals from the Potapovka culture in that she appears to have admixture related to that of the *Central_Steppe_EMBA* individuals where were located further East. In her grave were sacrificed skulls of a cow and a goat ram. Her craniofacial type was unusual and exhibited eastern features associated with “Uralic” or “Mongoloid” populations, according to analysis by A.A. Khokhlov.

S2.2.3 Sites from the Steppe Zone in the MLBA

S2.2.3.1 Corded Ware Complex from the Western Steppe (n=7)

S2.2.3.1.1 Radovesice, Czech Republic (n=3)

Description by Miroslav Dobeš and Petr Velemínský

Four isolated Corded Ware Complex-associated graves were discovered in the 70s and 80s of the 20th century, amongst many others that were destroyed by the activity of a brown coal mine. These graves were spaced about 200m to 1000m apart, and three were included in the study and described in detail below.

- RDVS_04/81, Radovesice XVIII, "Na vyhlídce", Grave 4/81, National Museum No. P7A 17 333 (I7207): Date of 2559-2340 calBCE (3935±25 BP, PSUAMS-3887). Genetically male. Grave 4/81 was excavated by P. Budinský in 1981. In the burial pit, the skeleton lay in a right-sided crouched position, head toward the west. The grave furnishings consisted of a beaker, amphora, jug, another vessel, a stone battle-axe and a flat axe. The skeleton including grave equipment was heavily damaged by bulldozer pressure. According to the archaeological indicators (position of the body and grave goods) this is the burial of a male dated to the younger stage of the Corded Ware Complex. According to the skeletal morphology, these are the remains of a male who died between 30-50 years of age (*106*, *107*).

• RDVS_40/78, Radovesice III, "Za kostelem”, Grave 40/78, National Museum No. P7A 9319 , Radovesice (I7208): Date of 2464-2295 calBCE (3885±25 BP, PSUAMS-3888). Genetically male. The grave 40/78 was excavated by J. Muška in 1978. In the burial pit, the skeleton lay in a right-sided crouched position, head toward the west. The grave equipment consisted of a beaker, amphora, stone flat axe and blade. According to the archaeological indicators, this is the burial of a male dated to the younger stage of the Corded Ware Complex. Similarly, according to the skeletal morphology, this is also probably the remains of a male who died between the ages of 18-20 (*106*, *108*, *109*).

• RDVS_5/79, Radovesice X, "U bílinské silnice", Grave 5/79, National Museum No. P7A 9322 (I7209): Date of 2458-2207 calBCE (3850±25 BP, PSUAMS-4026). Genetically male. The grave 5/79 was excavated by J. Muška in 1979. In the burial pit, the skeleton lay in the right-sided crouched position, head toward the west. The grave equipment consisted of a beaker, amphora, jug, stone club and flat axe. According to the archaeological indicators, this is the burial of a male dated to the younger stage of the Corded Ware Complex. Similarly, according to the skeletal morphology, this is a male who died between the ages of 40-50 (*106*, *108*, *109*).

S2.2.3.1.2 Bílina, Czech Republic (n=3)

Description by Miroslav Dobeš and Petr Velemínský

These individuals were found in two sandpits (Tizler’s sandpit and Köhler’s sandpit) in the territory of the town of Bílina in northern Bohemia (district of Teplice).

Tizler’s sandpit was part of a cemetery disturbed by the long-term mining of sand. A number of graves were destroyed in the process, but a large number of graves dating from the Eneolithic to the Early Middle Ages have been partially rescued. These graves were mainly uncovered at the turn of the 19th to the 20th century, when the anthropological material was not carefully treated, and therefore came into the collection of museums only occasionally. Despite this, the size of the Corded Ware burial ground can be estimated from several dozen graves.

Köhler’s sandpit is more than 1 km from the Titzler`s sandpit, and as with the previous site, a large number of graves were destroyed and the bones that were taken were not stored in a professional manner. This was a Bronze Age settlement and a few Corded Ware graves were randomly evaluated during the progressive mining process (*106*).

Genetically all three individuals appeared to be distinct, and we clustered them accordingly.

- BILI_139, National Museum No. P7A 7564 (I6695): Context date of 2900-2350 BCE. Genetically female. This skeleton was delivered to the Teplice Museum on February 19th. 1903, together with three vessels belonging to the middle-to-late stage of the Corded Ware Complex. Further archaeological circumstances are not known, and dating of skeletal material is therefore not entirely reliable. Only the skull has preserved from the skeleton (*106*, *110*). This individual was anthropologically determined to be an adult aged over 30 years (30-40 years?). Based on the genetic ancestry of this individual we assigned it the *Corded_Ware_Czech_EN* split label, and the *Western_Steppe_MLBA* analysis label.

- BILI_4, National Museum No. P7A 7558 (I6677): Context date of 3800-3400 BCE. Genetically male. This grave without any further description was discovered on April 19th. 1910 by the amateur archeologist Fassl. According to the finder's documentation, there were two skeletons in a crouched (?) position - a man and woman - lying probably in a burial pit next to each other, in an east-west orientation. Only the male is included in our study (I6677; P7A 7558), who died as an adult (aged 30-50 years). A Baalberg Amphora was found next to the skull of the male skeleton (Baalberg stage of the Funnel Beaker Culture, ~ 3500 BCE). The anthropological evaluation of these remains revealed that besides the skull of a man and a woman, there is also a child's skull, which the finder does not mention (*109–111*). Based on the archaeological context as well as the genetic clustering we assign this individual to the *Bilina_N_Baalberge* split label and analysis label.

- BILI_7958, National Museum No. P7A 7958 (I7949): Context date of 2200-800 BCE. Genetically male. The skull comes from an archeologically undated burial pit (grave ?, No 15), which was excavated in 1941. According to the dating of other archeological features on the locality, the skeleton probably belongs to the Bronze Age (about 2200-800 BCE), and this is supported by the genetic clustering patterns as well so we assigned it a *Bilina_BA* analysis label and split label. Only the skull has been preserved (*110*). This individual was anthropologically determined to be an adult (40-50 years).

S2.2.3.1.3 Velké Žernoseky, Czech Republic (n=1)

Description by Miroslav Dobeš and Petr Velemínský

This site was located on the right-bank terrace of the Elbe river and was explored at the turn of the 19th and 20th centuries. The burial grounds were damaged by the mining of the porphyry. In total, several dozen Funnel Beaker Culture graves (ca. 3700-3500 BCE) and Únětice culture (2200-1700 BC) were uncovered. This sample was assigned the *Corded_Ware_Czech_EN* split label and the *Western_Steppe_MLBA* analysis label.

- VEZE_27-III, Nicolaus Parthe's quarry, Grave 27, National Museum No. P7A 6589 (I6696): Context date of 2900-2350 BCE. Genetically male. The solitary grave 27 was excavated by R. R. Weinzierl in the March in 1892. The skeleton lay in a crouched position in the southeast-northwest orientation. In the grave, a bone pin and two antler belt clasps were found. According to the grave furnishings, this burial is dated to the older stage of the Corded Ware Complex. Anthropologically, this individual was determined to be a male who died over the age of 50. Only the skull has preserved from the skeleton (*109*, *110*, *112*).

S2.2.3.2 Northeastern Kazakhstan, Kanai Cemetery (n=1)

The individual is assigned to its own *Kanai_MBA* genetic analysis label.

- StPet37, collection 6095, individual 9 (I6708): Date of 2193-2031 calBCE (3705±20 BP, PSUAMS-2916). Genetically male. The genetic affinity of this individual is similar to the *Okunevo.SG* individuals we analyze in this study, which are also geographically close to this site.

S2.2.3.3 Petrovka culture, Stepnoe VII cemetery, Russia (n=3)

Description by Elena Kupriyanova and Bryan Hanks

These individuals are assigned to the *Western_Steppe_MLBA* genetic analysis label.

The Stepnoye (Stepnoe) VII cemetery is situated in the Chelyabinsk region of Russia on the border of the steppe and forest-steppe zones of the Southern Trans-Urals. The site is located along the Ui River and 180 km southwest of the city of Chelyabinsk. Archaeological research was carried out at this site from 2000-2005 by D. Zdanovich and E. Kupriyanova and in 2016 by E. Kupriyanova. Eight burial complexes were excavated including more than 100 burials and sacrificial pits. All complexes have circular planning with one to three central burials in the styles of the Petrovka and Alakul cultures. Most adult burials were disturbed in antiquity; however, five burials were untouched and contained a diverse array of grave goods including ornaments, weapons, tools, and horse harness components.

In total, human skeletal remains recovered from the cemetery include 63 individuals. Remains of 34 individuals are from the Alakul' phase, 28 from the Petrovka phase, and one is indeterminate. In the Alakul' burials, most skeletons represent neonates and children before the age of one year. However, in Petrovka burials, groups of children and sub-adults have statistically similar percentages (38.8% and 32.3%). Adults and children from 1 to 10 years of age represent a small percentage of the total number of burials excavated (**Table S6**).

	Age					
	0-1	**1-4**	**4-10**	**10-17**	**17-25**	**25+**
Petrovka culture	11 38.8%	0 0%	4 14.3%	9 3.2%	3 10.7%	1 3.6%
Alakul' culture	30 88%	0 0%	1 3%	1 3%	0 0%	2 6%

Table S6 Ages of individuals associated with the Alakul' and Petrovka phases at Stepnoye VII

Adjacent archaeological sites that are related to the Stepnoye VII cemetery include the fortified settlement of Stepnoye and the Stepnoye-1 cemetery (Sintashta, Petrovka and Alakul' cultures), which are situated 1.5 km to the northwest. These sites were situated at the border zone between the forest-steppe ecological zone and the steppe and thus represent an important contact zone between different cultural and economic systems. Populations that inhabited the Stepnoye settlement, and utilized the Stepnoye-1 and Stepnoye VII cemeteries, may have mediated metalwork technology and metal resources (particularly tin) between the northern and central areas of the Urals and central steppes region.

• StepVIIS-1, burial 17, skeleton 1 (I0944): Date of 2023-1704 calBCE (3540±52 BP, AA-90949). Genetically female. This individual is skeletally 15 to 17 years old and female.

• StepVIIS-2, burial 17, skeleton 2 (I0945): Context date of 2200-1700 BCE. Genetically female. This individual is skeletally 15 to 17 years old and female.

• StepVIIS-3, burial 17, skeleton 3 (I0946): Date of 2130-1765 calBCE (3584±55 BP, AA-90948). Genetically female. This individual is skeletally 45 to 55 years old and female.

S2.2.3.4 <u>Sintashta culture from Kamennyi Ambar 5 cemetery, Russia (n=50)</u>
Description by Bryan Hanks, Andrey Epimakhov, Alexander Khokhlov and David Anthony

We present genome-wide ancient DNA data from 50 individuals from the Kamennyi Ambar 5 cemetery in western Siberia, which we co-analyzed for genetic purposes with an individual from another site (Grachevka in Samara) and that clustered with them genetically (this individual was associated to the Potapovka culture that was culturally similar to and contemporary to the Sintashta culture albeit further to the west).

The Kamennyi Ambar 5 cemetery is located 260 km south of the city of Chelyabinsk, Russia, and is situated within the steppe zone of the Trans-Urals region. V.P. Kostyukov and A.V. Epimakhov conducted archaeological research at this site over nine seasons between 1990 and 2014 (*113*). The cemetery consists of four barrows dating to the Bronze Age (Sintashta culture), which are related to the population that inhabited the Kamennyi Ambar fortified settlement on the opposite bank of the Karagaily-Ayat River (Fig. S1.1) (*114*). Despite the barrows being of relatively small size (30 m in diameter and height of <1 m), some of the burial mounds contained up to 17 grave pits with one or more inhumations. Direct dates of skeletons from the site fall in the range of 2030-1660 BCE. Excavation indicated that the barrow complexes were formed gradually, and all large grave pit features had individual surface structures (**Fig S 5**). Stratigraphically later burials have features of the Early Srubnaya culture type. A total of 23 of 38 pits contained collective burials that included up to eight individuals buried at the same time. The main ritual treatment of the deceased was inhumation. More than half of the graves had been disturbed by subsequent human activities and burrowing animals and therefore many individuals were recovered as incomplete skeletons. The remains of about 130 individuals (predominantly subadults) were discovered. According to the authors of the excavations, the individuals interred within the burial mounds represent only a fraction of the population that would have inhabited the settlement. Grave goods included numerous metals, stone and bone items, as well as pottery. Among the finds are weapons, tools, horse harness components, stone mace heads, jewelry, and utilitarian items. Nearly all burials were accompanied by animal sacrifices that included domestic animals such as sheep, goats, cattle, horses and dogs.

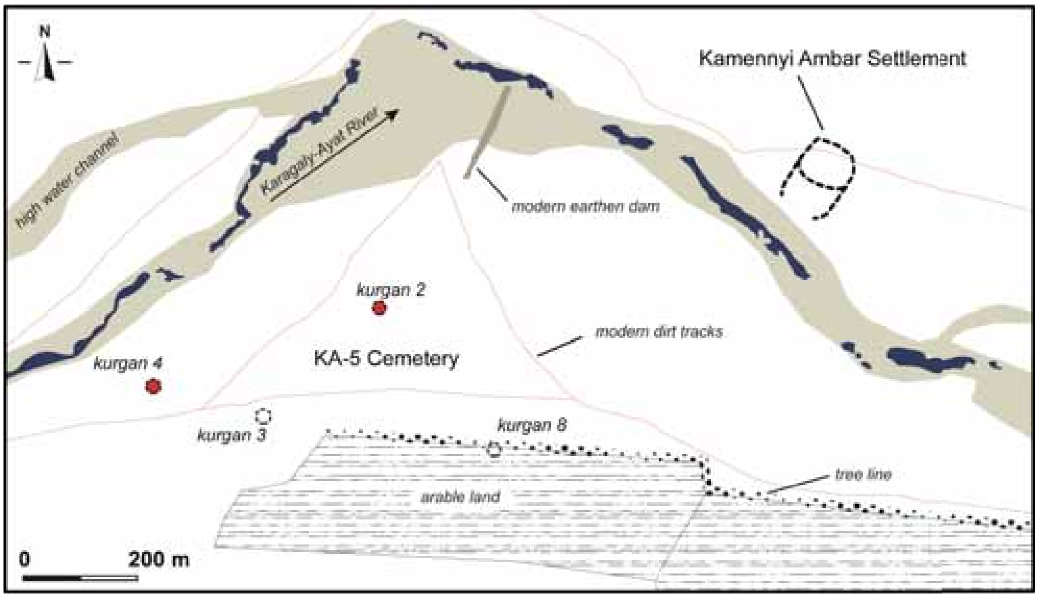

Fig S 5 Geographic layout of the Kamennyi Ambar 5 cemetery

We combined the data we generated from Kamennyi Ambar 5 with previously reported data from 5 Sintashta individuals (*8*). We observed a main cluster of 41 *Sintashta* individuals that was genetically similar to *Srubnaya*, *Potapovka*, and *Andronovo* in being well modeled as a mixture of *Yamnaya*-related and *Anatolia_N* (European farmer-related) ancestry (**Table S 60**) (*9*), and that we assigned to the *Western_Steppe_MLBA* cluster for genetic analysis.

The Kamennyi Ambar 5 individuals also included 10 outliers with distinctive ancestry of different types. The first set consists of 4 individuals, *Sintashta_o1* is genetically similar to LBA individuals with *WSHG*-related admixture, an ancestry profile that is also seen in outlier individuals from other sites we analyzed found further to the east. The next set of 4 outliers, *Sintashta_o2*, consists of individuals with admixture from *Western_Steppe_EMBA*-like groups (**Fig S3.21**). The third set, includes 2 individuals that we denote *Sintastha_o3*, who have much less Anatolian farmer- and Iranian farmer-related ancestry and are genetically similar to Eneolithic populations from Khvalynsk. All these ancestry outliers were not obviously correlated to any archaeological features of the cemetery and the direct dates that we obtained were in the range of other individuals from the cemetery. The fact that these genetic outliers were interred simultaneously in the same grave pits with individuals from the main cluster of Sintashta individuals highlights the genetic heterogeneity of Sintashta communities that were nevertheless organized as single social groups. For analysis, we removed 6 relatives and 9 outliers (2 individuals were both relatives and outliers), producing 37 individuals to represent the main cluster. The locations of the analyzed individuals within the analyzed kurgans (2 and 4) are shown in **Fig S 6**.

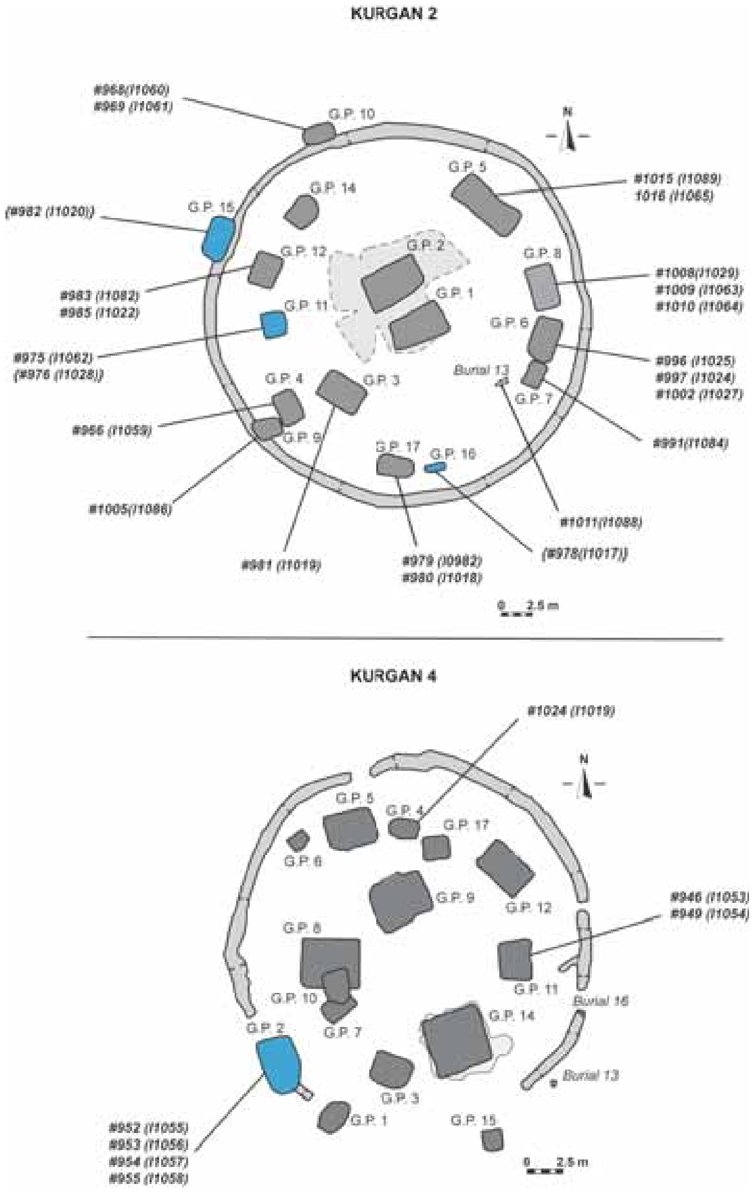

Fig S 6 Plan of the two excavated Sintashta Kurgans at Kamennyi Ambar 5. The analyzed individuals are labeled, and blue coloring indicates grave containing genetic outliers.

We first list the 41 *Sintashta* individuals in the main genetic cluster (including four relatives of other high coverage individuals in the dataset that we excluded from our main analysis). We assign these individuals the *Sintashta_MLBA* split label.

Kamennyi Ambar 5, Russia (n=41; main genetic cluster)

• 979, kurgan 2, burial 17, skeleton 1 (I0982): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically female. This individual is skeletally an adult female.

• 980, kurgan 2, burial 17, skeleton 2 (I1018): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically male. This individual is skeletally 6 to 10 years old and of indeterminate sex.

• 981, kurgan 2, burial 3 (I1019): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically male. This individual is skeletally a 45 to 59-year-old male.

• 985, kurgan 2, burial 12, skeleton 3 (I1022): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically female. This individual is skeletally a 38 to 49-year-old female.

• 993, kurgan 2, burial 7, skeleton 6 (I1024): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically female. This individual is skeletally 6 to 10 years old and of indeterminate sex.

• 1000, kurgan 2, burial 6, skeleton 6 / 1002, kurgan 2, burial 6, skeleton 8 / 1006, kurgan 2, burial 9, skeleton 2 (I1027): Date of 1962-1775 calBCE (3550±25 BP, PSUAMS-1954). Genetically male. This individual is skeletally 3 to 9 months old and of indeterminate sex. The genetic data represent a merge of ancient DNA from three different individuals that we genetically determined to be from the same individual.

- 1008, kurgan 2, burial 8, skeleton 2 (I1029): Previously reported date of 1973-1772 calBCE (3549±29 BP, OxA-12531). Genetically female. This individual is skeletally 7 to 11 years old and of indeterminate sex.

- 946, kurgan 4, burial 11, skeleton 2 (I1053): Date of 1922-1763 calBCE (3520±25 BP, PSUAMS-2064). Genetically male. This individual is genetically a relative (brother) of 949 (I1054), and they were both buried in the same grave (G.P. 11 in Kurgan 4). This individual is skeletally <1-year-old and of indeterminate sex.

- 946, kurgan 4, burial 11, skeleton 2 (I1054): Date of 1891-1746 calBCE (3495±25 BP, PSUAMS-1952). Genetically male. This individual is excluded from analysis as a first degree relative (brother) of 946 (I1053) buried in the same grave (G.P. 11 in Kurgan 4). This individual is skeletally 2 to 4 years old and of indeterminate sex.

- 952, kurgan 4, burial 2, skeleton 3 (I1055): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically male. This individual is skeletally 3 to 5 years old and of indeterminate sex.

- 968, kurgan 2, burial 10, skeleton 1 (I1060): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically female. This individual is skeletally 3 to 5 years old and of indeterminate sex.

- 969, kurgan 2, burial 10, skeleton 2 (I1061): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically female. This individual is skeletally 2 years old (±8 months) and of indeterminate sex.

- 975, kurgan 2, burial 11, skeleton 1 (I1062): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically female. This individual is skeletally 6 to 10 years old and of indeterminate sex.

- 1009, kurgan 2, burial 8, skeleton 3 (I1063): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically female. This individual is skeletally 2 to 4 years old and of indeterminate sex.

- 1010, kurgan 2, burial 8, skeleton 4 (I1064): Context date of 1879-1694 calBCE (3460±20 BP, PSUAMS-2102). Genetically male. This individual is skeletally a neonate of indeterminate sex.

- 1016, kurgan 2, burial 5, skeleton 4 (I1065): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically male. This individual is skeletally a 10 to 12-year-old of indeterminate sex.

- 983, kurgan 2, burial 12, skeleton 1 (I1082): Context of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically male. This individual is skeletally a 5 to 9-year-old and of indeterminate sex.

- 991, kurgan 2, burial 7, skeleton 3 (I1084): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically male. This individual is skeletally 4.5 to 8.5 years old and of indeterminate sex. Genetic analysis indicates that I0985 and I1084 are first-degree relatives and indeed both were buried in the same grave (G.P. 7 of Kurgan 2).

- 1005, kurgan 2, burial 9, skeleton 1 (I1086): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically male. This individual is skeletally 10 years old (±8 months) and of indeterminate sex. He is genetically determined to be a first-degree relative of I1010; the two must be brothers as neither is of reproductive age. The two were buried in immediately adjacent graves in Kurgan 2 (G.P. 4 for I1010, and G.P. 9 for I1086).

- 1011, kurgan 2, burial 13 (I1088): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically female. This individual is skeletally an adult, likely a female.

- 1015, kurgan 2, burial 5, skeleton 3a (I1089): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically female. This individual is skeletally a <1-year-old of indeterminate sex. Genetic analysis indicates that I0987 and I1089 are second-to-third-degree relatives and indeed both were buried in the same grave (G.P. 5 of Kurgan 2).

- 1024, kurgan 4, burial 4 (I1090): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically female. This individual is skeletally 4 to 8 years old and of indeterminate sex.

- 1012, kurgan 2, burial 5, skeleton 1 (I0942): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically male. This individual is skeletally 15 to 18 years old and of indeterminate sex.

- 1017, kurgan 2, burial 5, skeleton 1 (I0943): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically female. This individual is skeletally 10 to 12 years old and of indeterminate sex.

- 11018, kurgan 2, burial 5, skeleton 5a (I0989): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically male. This individual is skeletally 14 to 18 years old and possibly male.

- 937, kurgan 4, burial 3 (I0937): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically female. This individual is skeletally 22 to 25 years old and female.

- 939, kurgan 4, burial 1 (I0938): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically female. This individual is skeletally 12 to 18 years old and female.

- 945, kurgan 4, burial 9, skeleton 1 (I1003): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically male. This individual is skeletally 9.5 to 14.5 years old and of indeterminate sex.

- 948, kurgan 4, burial 9, skeleton 2 (I0939): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically female. This individual is skeletally 12 to 18 years old and of indeterminate sex.

- 950, kurgan 4, burial 2 (I0940): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically female. This individual is skeletally 7.5 to 12.5 years old and of indeterminate sex.

- 956, kurgan 4, burial 5, skeleton 1 (I1006): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically male. This individual is skeletally an adult and male.

- 960, kurgan 4, burial 5, skeleton 5 (I1008): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically male. This individual is skeletally 5 to 6 years old and of indeterminate sex.

- 964, kurgan 2, burial 4, skeleton 3 (I1010): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically male. This individual is skeletally 5 to 7 years old and of indeterminate sex. He is genetically determined to be a first-degree relative of I1086; the two must be brothers as neither is of reproductive age. The two were buried in immediately adjacent graves in Kurgan 2 (G.P. 4 for I1010, and G.P. 9 for I1086).

- 965, kurgan 2, burial 4, skeleton 4 / 966, kurgan 2, burial 4, skeleton 5 (I1011): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically male. This individual is skeletally 6 to 10 years old and of indeterminate sex. Specific context of the two individuals is indeterminate archaeologically; however, both bone samples represent a single individual genetically that was recovered from a single kurgan complex. This individual is skeletally 1 to 2.5 years old and of indeterminate sex.

- 967, kurgan 2, burial 4, skeleton 6 / 971, kurgan 2, burial 8, skeleton (I1012): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically male. This individual is skeletally adult and male. The specific context of the two samples is indeterminate archaeologically, but the genetic data indicate that the bone samples derive from a single individual and this is unsurprising given that this is consistent with the bones being recovered from a single kurgan complex.

- 970, kurgan 2, burial 10, skeleton 3 (I1013): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically female. This individual is skeletally 5 to 9 years old and of indeterminate sex.

- 992, kurgan 2, burial 7, skeleton 4 / 996, kurgan 2, burial 6, skeleton 2 (I0984): Date of 2023-1782 calBCE (3572±29 BP, OxA-12530). Genetically male. Specific context of these individuals is indeterminate archaeologically, however, both bone samples represent a single individual genetically that was recovered from a single kurgan complex.

- 994, kurgan 2, burial 7 (I0985): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically female. Genetic analysis indicates that I0985 and I1084 are first-degree relatives and indeed both were buried in the same grave (G.P. 7 of Kurgan 2).

- 995, kurgan 2, burial 6, skeleton 1 (I0986): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically male. This individual is skeletally adult and possibly male.

- 39, kurgan 4, grave 8, individual 1 (I7480): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically female. Grave 8 of Kurgan 4 contained partial bones of four individuals, buried with seven ceramic vessels, two pairs of antler cheekpieces, two projectile points, and the skulls of three sacrificed horses, two cattle, and one sheep-goat. Skeleton 1 was a woman of mature age. Like Grachevka II k5/3 described earlier (I0244), her craniofacial type was unusual and exhibited eastern features associated with Uralic or Mongoloid populations, according to analysis by A.A. Khokhlov.

- 1013, kurgan 2, burial 5, skeleton 2 (I0987): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically male. This individual is skeletally 9.5-12 years old and of indeterminate sex. Genetic analysis indicates that I0987 and I1089 are second-to-third-degree relatives and indeed both are buried in the same grave (G.P. 5 of Kurgan 2).

Kamennyi Ambar 5, Russia (n=9; genetic outliers)

In what follows, we label the 10 outlier individuals from Kamennyi Ambar by their split analysis labels.

Sintashta_MLBA_oWSHG: These outliers have *WSHG*-related affinity and in this way resemble Steppe populations in the EMBA (n=8).

- 972, kurgan 2, trench 5 (I0980): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically male. This individual is skeletally adult and possibly male.

- 957, kurgan 4, burial 5, skeleton 2 (I1007): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically male. This individual is skeletally 12 to 18 years old and possibly female.

- 978, kurgan 2, burial 16 (I1017): Date of 1929-1753 calBCE (3520±30 BP, Beta-436294). Genetically male. This individual is skeletally 12 to 18 years old of indeterminate sex.

- 986, kurgan 2, burial 12, skeleton 4 (I0983): Context date of 2050-1650 BCE (range of directly dated individuals from the same site). Genetically female. This individual is skeletally 20 to 24 years old and female.

- 982, kurgan 2, burial 15 (I1020): Previously reported date of 2012-1774 calBCE (3555±31 BP, OxA-12533). Genetically male. This individual is skeletally a 16 to 18-year-old male.

- 953, kurgan 4, burial 2, skeleton 4 (I1056): Date of 1886-1695 calBCE (3475±30 BP, PSUAMS-2101). Genetically male. This individual is one of a trio of brothers of similar outlier ancestry—953 (I1056), 954 (I1057), and 955 (I1058) —all buried in the same grave (G.P. 2 of Kurgan 4). This individual is skeletally 3 to 5 years old and of indeterminate sex.

- 954, kurgan 4, burial 2, skeleton 5 (I1057): Date of 1949-1772 calBCE (3540±25 BP, PSUAMS-1953). Genetically male. This individual is one of a trio of brothers of similar outlier ancestry—953 (I1056), 954 (I1057), and 955 (I1058) —all buried in the same grave (G.P. 2 of Kurgan 4). This individual is skeletally a 3-9-month-old and of indeterminate sex.

• 955, kurgan 4, burial 2, skeleton 6 (I1058): Date of 1906-1743 calBCE (3500±30 BP, Beta-436363). Genetically male. This individual is one of a trio of brothers of similar outlier ancestry—953 (I1056), 954 (I1057), and 955 (I1058)—all buried in the same grave (G.P. 2 of Kurgan 4). This individual is skeletally 3 to 5 years old and of indeterminate sex.

Sintashta_MLBA_o3: Outlier genetically resembling *Khvalynsk_EN* (n=1)

• 976, kurgan 2, burial 11, skeleton 2 / 1003, kurgan 2, burial 1 (I1028): Date of 1878-1664 calBCE (3440±30 BP, Beta-436293) obtained for sample 976, which was skeletally determined to be a 12 to 18-year-old male genetically identical to the data from 1003.

Relatedness summary:
• 3-person family: I1056-I1057-I1058 are a trio of brothers who are also ancestry outliers.
• 2-person family: I1053-I1054 are brothers.
• 2-person family: I1010-I1086 are brothers (they are genetically first-degree relatives and neither is of reproductive age).
• 2-person family: I0985-I1084 are first-degree relatives.
• 2-person family: I0940-I1055 are first-degree relatives.
• 2-person family: I0987-I1089 are a pair of 2^{nd} to 3^{rd} degree relatives.

S2.2.3.5 Satan, Central Kazakhstan (n=1)
Description by Sergey M. Slepchenko and Dmitry Enshin

This individual is assigned on genetic grounds to the *Satan_MLBA_Alakul* analysis label.

The burial ground of Satan is situated on a slope of a bald mountain on the left bank of the Talda River in the steppe zone of Central Kazakhstan (near the village of Frunze, Taldinsky district, Karaganda oblast). The research was carried out in 1980 under the guidance of V.V. Evdokimov. As a result of the work, six burial constructions were studied, all of which consisted of mounds and circular fences without mounds. Large ground pits, sometimes with log constructions, were in the center of the plots under the mounds. Graves were either stone boxes or ground pits located inside fences made of vertically-oriented granite slabs placed into dug grooves. The burials are represented by individual, double and group graves. The dead were buried in a crouched position with their hands at their faces. Ceramic vessels, parts of a horse harness, bronze and gold adornments, and an arrowhead were found. In addition, remains of a wooden chariot were found in one of the mounds. Researchers attribute the population that built the Satan burial ground to the Nurtai culture of Central Kazakhstan (Andronovo-associated) of the Bronze Age in the first half of the 2nd millennium BCE. This was a society in complex cattle and sheep husbandry (*115*). This individual we analyzed was part of the anthropological collection of Tomsk State University and the inventory numbers from the collection are provided along with that of the burial.

- Tomsk_4371, inventory number 4371, Kurgan 4, burial 1, Satan burial ground (I6799): Date of 1876-1688 calBCE (3445±20 BP, PSUAMS-2981). Genetically male.

S2.2.3.6 Maitan, Central Kazakhstan (n=9)
Description by Sergey M. Slepchenko and Dmitry Enshin

These individuals are assigned on genetic grounds to the *Western_Steppe_MLBA* analysis label.

The burial ground of Maitan is situated in the steppe zone on the Ashi Su River, the right tributary of the Nura River (Central Kazakhstan, Karaganda oblast, Ulyanovsk district). Research at this site was conducted from 1984 to 1991 under the guidance of A.A. Tkachev. The site was completely excavated and consisted of 113 ring-shaped fences (and annexes) constructed by placing stone slabs on edge into dug grooves, inside which graves were organized. The deceased were placed in a stone box, mainly on their left or right side in a contracted position with the hands in front of the face. Based on analysis of ceramic vessels and bronze artifacts, the burial ground was attributed to the Atasu culture (Andronovo-associated) by the researchers. Its bearers were shepherds and farmers of the Bronze Age between the 15th and 13th centuries BCE (*116*, *117*). These individuals were all part of the anthropological collection of Tomsk State University and the inventory numbers from the collection are provided along with that of the burial.

- Tomsk_4335, inventory number 4335, Kurgan 6, burial 1 (I6789): Date of 1877-1693 calBCE (3455±20 BP, PSUAMS-2922). Genetically male.

- Tomsk_4337, inventory number 4337, Kurgan 8, burial 1 (I6790): Date of 1872-1684 calBCE (3435±20 BP, PSUAMS-2980). Genetically female.

- Tomsk_4341, inventory number 4341, Kurgan 17, burial 1 (I6791): Date of 1862-1664 calBCE (3425±20 BP, PSUAMS-2923). Genetically female.

- Tomsk_4346, inventory number 4346, Kurgan 27a, burial 1 (I6793): Date of 1745-1636 calBCE (3395±20 BP, PSUAMS-2925). Genetically male.

- Tomsk_4347, inventory number 4347, Kurgan 29b, burial 1 (I6794): Date of 1876-1688 calBCE (3445±20 BP, PSUAMS-2926). Genetically male.

- Tomsk_4351, inventory number 4351, Kurgan 40, burial 2 (I6796): Date of 1882-1748 calBCE (3485±20 BP, PSUAMS-2928). Genetically female.

- Tomsk_4352, inventory number 4352, Kurgan 40, burial 6 (I6797): Date of 1872-1684 calBCE (3435±20 BP, PSUAMS-2929). Genetically female.

The following two individuals are outliers from the main cluster and resemble individuals from the LBA, with significant amounts of *WSHG* ancestry. We assign them the analysis label *Steppe_MLBA_oWSHG*.

- Tomsk_4344, inventory number 4344, Kurgan 23d, burial 1 (I6792): Date of 1876-1691 calBCE (3450±20 BP, PSUAMS-2924). Genetically female.

- Tomsk_4350, inventory number 4350, Kurgan 36a, burial 1 (I6795): Date of 1749-1642 calBCE (3405±20 BP, PSUAMS-2927). Genetically female.

S2.2.3.7 Karagash 2, Central Kazakhstan (n=3)

These individuals are assigned on genetic grounds to the *Western_Steppe_MLBA* analysis label.

• KZ-KAR-002, Karagash 1994, KB. 4Г. Grave 3 (I4262): Date of 1881-1695 calBCE (3465±25 BP, PSUAMS-2122). Genetically male. Individual I4262 and I4778 have a father-son relationship (it is not clear from the genetic data which one is the father).

• KZ-KAR-003, KV. 1V, Grave 2, 1994 (I4778): Date of 1728-1546 calBCE (3345±20 BP, PSUAMS-2611). Individual I4262 and I4778 have a father-son relationship (it is not clear from the genetic data which one is the father).

• KZ-KAR-004, Karagash 1994, KB. 1A. Grave 1 (I4263): Date of 1861-1639 calBCE (3415±25 BP, PSUAMS-2123). Genetically male.

Relatedness summary:

• 2-person family: I4262-I4778 have a father-son relationship (it is not clear from the genetic data which one is the father).

S2.2.3.8 Ak-Moustafa, Central Kazakhstan (n=1)

Description by Gian Luca Bonora, Stefania Sarno and Davide Pettener

This individual is assigned on genetic grounds to the *Western_Steppe_MLBA* analysis label.

The Ak-Mustafa settlement and burial ground are located in Central Kazakhstan, on the southern bank of the Atasu river, about 240 km south-southwest of Karaganda and 245 km northwest of Balqash. The Ak-Mustafa settlement was composed of 49 Bronze Age pit-houses with evidence of copper smelting and metal production. The burial ground, located opposite the settlement, was excavated between 1975 and 1979, and identified 13 burials dating to the Late Bronze Age. The research continued in 1987 when numerous funerary structures, belonging to different periods, were excavated. By the time the research had been completed, 89 funerary fences and 27 burial constructions of the Bronze Age, as well as some barrows of the Early Iron Age and later periods, had been discovered. The Bronze Age funerary structures are represented by both round and quadrangular fences, composed of vertical slab stones and stone cists without fences. An extremely large barrow surrounded by a stone fence (number 95) was identified on the southeastern edge of the graveyard. Ak-Mustafa is a typical Bronze Age complex of central Kazakhstan, wherein varying combinations of Fedorovo and Alakul’ pottery traditions have been identified, testifying that active interaction between different groups of populations took place for an extended period of time. These interactions may have reflected the result of the rich copper resources of the region, which could have attracted the economic interest of many communities from the ancient farming villages of southern Central Asia.

- Al-1, kurgan 2 (I3767): Date of 1869-1665 calBCE (3430±20 BP, PSUAMS-2496). Genetically and morphologically male, age 30 to 35 years old.

S2.2.3.9 Dali, Bayan-Zherek Valley, Kazakhstan (n=4)
Descriptions by Michael Frachetti, Egor Kitov, and Gaziz Akhatov

We present genome-wide ancient DNA from four individuals from Dali. The site complex of "Dali" is situated across a sloping ravine at roughly 1500 m elevation on the eastern spur of the Bayan-Zherek range. The site was documented in 2010 under the aegis of the collaborative Dzhungar Mountains Archaeology Project (DMAP, PI: M. Frachetti and A. Mar'yashev), roughly 900 m northwest from the prehistoric settlement of Tasbas (*118*, *119*). The site complex includes numerous multi-phase habitations with the earliest Bronze Age occupation strata dating to the first quarter of the third millennium BCE and the later occupation strata to the first half of the second millennium BCE. In addition to settlements, Dali contains two Bronze Age cemeteries distributed in discrete clusters, Dali 1 and Dali 2. Burial excavations were carried out in 2012 and 2014.

Human remains were recovered from formal burials excavated from the Dali 1 and Dali 2 cemeteries. Dali 1, the older of the two cemeteries, is located at the base of the site area, where human remains date from roughly 1865-1640 calBCE. The upper burial area, Dali 2, is later chronologically, dating between 1690-1505 calBCE. In addition, displaced human remains were recovered from within early Bronze Age (3000-2500 BCE) layers of the Dali settlement, directly dated to the first quarter of the third millennium BCE.

S2.2.3.9.1 Middle to Late Bronze Age Dali 1 cemetery, burial 1, Kazakhstan (n=3)

Dali 1, burial 1 was excavated in the lower cemetery cluster at Dali in 2012 by Dr. Mark Meyer (bioarcheologist) and Dr. Michael Frachetti (Project PI). The grave was constructed of large granite flagstones, arranged in a rectilinear enclosure that defined the outer boundary of the grave structure. Within this outer stone "fence" was a layer of stones, also arranged in an irregular and inconsistent oval ring. The grave was clearly disturbed and robbed in antiquity, as illustrated by displaced stones from both the outer and inner part of the burial; only a fraction of the skeletons and a few broken ceramic fragments were recovered within the excavated strata. The burial cist was located ~1.90 m below the surface at the center of the outer ring and consisted of thin granite stones aligned to form a 1.6 x 0.9 m chamber. The stone edges of the cist were badly weathered and crumbled upon excavation. Skeletal remains of a minimum of two individuals were recovered within the grave, indicated by divergent estimates of size and stature for each individual. Bone samples in the form of disarticulated metatarsals were taken from each and analyzed for ancient DNA. We use a single date from the co-mingled bones to represent both individuals.

Both individuals are assigned on genetic grounds to the *Steppe_MLBA_oWSHG* analysis label, reflecting ancestry that is genetically intermediate between group like *Western_Steppe_MLBA* or *Central_Steppe_MLBA*, and *WSHG.*

- KzHS7, individual 1, Dali 1, burial 1 (I0507): Date of 1871-1636 calBCE (3420±30 BP, Beta-391199). Genetically male. On the basis of the relative size and morphology of the bones recovered, the majority of skeletal elements in Dali 1, burial 1 belonged to a middle-aged male of relatively large stature. The small number of recovered bones precludes a more extensive physical characterization of the individual or its burial placement, though given the limited size of the grave cist, an adult-sized human would presumably have been positioned in a flexed position, as was typical of burials in this region during the early second millennium BCE.

- DL-BR1-B, CTX 2-B3, #21, individual 2, Dali 1, burial 1 (I3448): Context date of 1900-1600 BCE, based on the direct date for another individual in the same burial (I0507, direct date of 1871-1636 calBCE (3420±30 BP, Beta-391199), see previous entry). Genetically female. This bone sample was recovered in the upper levels of the burial stratigraphy and was labeled at the time of excavation as "individual 2" on account of its dimorphic size in comparison with the other skeletal elements found within the grave. The occurrence of this second individual within burial 1 either represents a co-interment (join burial), in which case the bone was displaced during the ancient grave robbery, or it represents a displaced bone fragment that washed into the upper fill levels of burial 1 from a nearby burial, also likely to have been disturbed in ancient times.

• DL-Br-2, sample KZBR-2-012, individual 1 (I1931): Date of 1495-1300 calBCE (3130±30 BP, Beta-391198). Genetically female. Dali burial 2 is located in the burial cluster on the northwestern extent (upslope) of the Dali site complex. Oval stone arrangements are visible on the surface and appear as ring enclosures with diameters of 2 to 3 m. Excavations were initiated on the most prominent stone ring in Dali 2, designated DL-BR2. Burial 2 was excavated in 2014 by Taylor Hermes, Michael Frachetti, and Alexei Mar'yashev. The burial was constructed with an outer stone oval, within which a stone-lined slab cist was built at a depth of roughly 1.8 m. As with the burials in the downslope cluster, Burial 2 was also looted and disturbed in antiquity, leaving the human remains and material grave goods fragmentary and ex-situ. Based on the size of the stone cist (0.8 x 1.6 m) it is logical that the body would have originally been interred in a flexed position, though the orientation cannot be reconstructed. The remaining disarticulated human remains were strewn throughout the upper levels of the grave's stratigraphy, though the ancient grave robbers were meticulous to reseal their trench with a dense layer of sun-dried mudbrick. In total 93 bone elements were recovered within the grave context. All were consistent with a middle-aged female individual based on mandible and dentition. In addition to the skeletal remains, bronze beads were recovered within the burial cist along with a broken bone knife handle and a small handmade ceramic pot with etched geometric design, all typical of eastern steppe materials in the middle/late Bronze Age. An inverted sheep or goat skull was ritually deposited in the burial, consistent with a desecration ritual performed at the time of the burial robbery.

S2.2.3.9.2 Early Bronze Age Dali settlement, Kazakhstan (n=1)

This individual is genetically assigned to its own *Dali_EBA* analysis label.

• DL-OP2-B, #41 (I3447): Date of 2850-2495 calBCE (4075±25 BP, PSUAMS-2071). Genetically female. This specimen reflects the earliest human remains recovered at Dali, consisting of a fragmented parietal bone recovered within the lower strata of a third millennium BCE pit-house in 2012. The skeletal fragment was not associated with any burial construction and was disarticulated from the rest of the body. Thus, we interpret this find as an ex-situ secondary deposit. Radiocarbon dates from stratigraphic layers above and below the deposit correlate with the direct date of the bone itself, indicating it was deposited in the settlement context in the first half of the third millennium BCE.

S2.2.3.10 Lisakovskiy, Northern Kazakhstan (n=2)

Description by Sergey M. Slepchenko and David Anthony

These individuals are assigned genetically to the *Western_Steppe_MLBA* analysis label.

The burial ground of Lisakovskiy I is situated on the left bank of the Tobol River near the town of Lisakovsk (Kustanai oblast, Republic of Kazakhstan). A total of 233 burials were completely excavated from 1986 to 1996 under the guidance of E.R. Usmanova. The burials were located in funerary constructions (fenced burial mounds). The placement of the funeral complex on the banks of the Tobol is associated with migration processes and increased mobility associated with the development of nomadic cattle breeding in the Bronze Age. Based on the analysis of funeral rituals, ceramic vessels and bronze articles, the materials of the burial ground were attributed to the Alakul and Fedorovo archaeological cultures (Andronovo antiquities). These features date activity at Lisakovskiy I from 1780 to 1660 BCE (*120–122*). These individuals were all part of the anthropological collection of Tomsk State University and the inventory numbers from the collection are provided along with that of the burial.

- Tomsk_4114, inventory number 4114, Kurgan 3, burial3 (I6788): Date of 1862-1664 calBCE (3425±20 BP, PSUAMS-2921). Genetically female.

- Tomsk_4387, inventory number 4387, Kurgan 1, burial1 (I6800): Date of 1767-1658 calBCE (3415±20 BP, PSUAMS-2942). Genetically female.

S2.2.3.11 Kairan, Kazakhstan (n=10)

Descriptions by Michael Frachetti, Egor Kitov, and Gaziz Akhatov

We present genome-wide ancient DNA data from 10 individuals discovered at Kairan 1, a burial ground located in the vicinity of the Ak-Koitas and Kara-Koitas mountains in the Karaganda region of Kazakhstan. Kairan 1 is distributed across two low hills and the lowland located between them. Stone fences of round and rectangular shape stretch along a line northwest to southeast. These stone enclosure burials, well known in the steppe during the early and middle 2nd millennium BCE, range in type and variety. In total, 12 compact groups of individual burial grounds have been identified in the Ak-Koitas and Kara-Koitas mountains. Samples of Bronze Age human remains recovered from cist burials (identified at the time of excavation) were taken for ancient DNA analysis.

The materials obtained from the Kairan 1 necropolis have their strongest cultural affinity to the burial grounds of central Kazakhstan, Ayapbergen and Nurtay. A diversity of decorative elements in materials recovered at Kairan 1 may reflect a complex process of interactions between regional Bronze Age communities. On the basis of the material remains and cultural typological relationships, Kairan 1 has been dated to the 14th to 12th centuries BCE. However, AMS radiocarbon dating carried out on the specimens reported here produces dates from the start of the 19th to the 16th centuries BCE. The chronological and cultural affinities of this site should be reconsidered in light of this chronological shift of roughly 400 years.

The main group of seven individuals falls in the *Western_Steppe_MLBA* analysis label.

- KZ-KAN-006, Kairan I, Enclosure 1, Grave 1 (I4567): Date of 1743-1631 calBCE (3390±20 BP, PSUAMS-2545). Genetically male. Stone enclosure 1 is a round-shaped arrangement of flagstones with a diameter of 3.5 m. Within the enclosure two grave cists were found, and a limited number of fragments of handmade decorated pottery were recovered in the fill areas around the graves. Grave 1 (the central grave) is represented by a rectangular stone cist oriented along the E-W axis, built of four stone slabs erected vertically on their lateral edges. Within the cist were the fragmentary remains of an adult individual; fragments of a ceramic vessel; remains of a necklace strung with canines, molars and the heel bones of a fox; bone beads of cylindrical shape ornamented with a spiral ornament in the form of carved lines; and other ceramic beads. Animal bones such as sheep astragali, horse bones, and fragments of shells were also recovered. Although the grave was apparently robbed in antiquity, a small assortment of bronze fragments including beads and pendants were recovered. Judging by the location and position of the few in situ bones, the body was interred in a flexed position on the left side, with the head oriented to the northwest.

- KZ-KAN-002 + KZ-KAY-002, Kairan I, Enclosure 7b (I4318): Date of 1767-1658 calBCE (3415±20 BP, PSUAMS-2961). Genetically female. Burial enclosure 7b is rectangular in shape and oriented in the W-E direction. It shares its northern wall with burial enclosure 7a. Within the enclosure is a central grave cist, consisting of four stone slabs erected vertically on their lateral edges. Within the cist, fragmentary remains of a human skeleton and one fragment of a ceramic vessel were found. Other materials include a bronze bead of cylindrical shape. Judging by the few in situ bones, the interred individual was placed in a flexed position lying on her left side.

- KZ-KAR-012, Kairan I, Enclosure 11, Grave 1 (I5761): Date of 1931-1772 calBCE (3530±20 BP, PSUAMS-3099). Genetically female. Burial enclosure 11, grave cist 1 is rectangular in shape and oriented along the E-W axis. The grave cist consists of four stone slabs erected vertically on their lateral edges. Within the cist, disarticulated remains of a human skeleton and ceramic fragments were recovered. Other materials include a bronze bead of cylindrical shape. Judging by the few in situ bones, the interred individual was placed in a flexed position lying on her left side, with the head facing to the west.

- KZ-KAN005, Kairan I, Enclosure 5, Grave 2 (I4776): Date of 1767-1658 calBCE (3415±20 BP, PSUAMS-2912). Genetically female. Burial enclosure 5 is a rectangular stone arrangement with two stone-lined cist graves within, oriented in a N-S alignment. Grave 2 (to the south) is represented by a rectangular stone cist oriented along the E-W axis, built of four stone slabs erected vertically on their lateral edges. Fragmentary remains of an adult individual were recovered within the cist, as well as round and cylindrical shaped beads, and rounded bronze beads with knobs at the edges.

- KZ-KAR-006, Kairan I, Enclosure 11, Grave 3 (I4568): Date of 1745-1636 calBCE (3395±20 BP, PSUAMS-2546). Genetically male. Burial enclosure 11, grave 3 (southern most cist) is rectangular in shape and oriented along an E-W axis. The grave cist is built using four stone slabs erected vertically on their lateral edges. In the eastern part of the burial, two round white beads and one ring-shaped bronze artifact were recovered. In the western part of the burial, fragments of a pottery vessel were found with geometric ornamentation. Within the fill of the grave cist, disarticulated human bones of (at least) two humans were found, I4568 and I4779 (described below), along with ceramic fragments and materials including a bronze bead of cylindrical shape. Due to the lack of in situ bones, it is impossible to judge the posture and orientation of the skeleton.

• KZ-KAR005, Kairan I, Enclosure 11, Grave 3 (I4779): Date of 1745-1636 calBCE (3395±20 BP, PSUAMS-2939). Genetically female. The disarticulated remains of this individual were found in Burial enclosure 11, grave 3 along with the remains of individual I4568 (described above). Due to the lack of in situ bones, it is impossible to judge the posture and orientation of the skeleton.

• KZ-KAR-014, Kairan I, Enclosure 6, Grave 1 (I5762): Context date of 1600-1500 BCE. Genetically male.

These 3 individuals deviate from the main genetic grouping of Kairan individuals in harboring significant proportions of ancestry related to *WSHG*, and so we use a split label of *Kairan_MLBA_o* to designate them here. These were then merged with other MLBA individuals from the steppe and assigned the *Steppe_MLBA_oWSHG* analysis label.

• KZ-KAN-004, Kairan I, Enclosure 7a, Grave 2 (I4566): Date of 1729-1563 calBCE (3350±20 BP, PSUAMS-2991). Genetically female. Grave 2 was located on the northern portion of the rectangular burial enclosure 7a. The rectangular-shaped cist was oriented along an E-W axis and consists of four vertically placed plates standing on their lateral edges. Within the fill of the grave cist, several vertebrae and ribs of a human skeleton were recovered. At the bottom of the grave bones and objects were found disturbed. As a result, the position of the buried cannot be determined.

• KZ-KAR009, Kairan I, Enclosure 7a, Grave 1 (I4780): Date of 1754-1642 calBCE (3410±20 BP, PSUAMS-2913). Genetically female. Burial Enclosure 7a is a construction consisting of vertically placed slabs arranged in a rectangle. Two grave cists were documented within the enclosure built from large stone slabs erected vertically on their lateral edges. The interred body of a male was placed in a flexed position, lying on his left side and his head oriented to the west. Between the front of the skull and the northern longitudinal wall of the box was an open form ceramic vessel with geometric ornamentation. Additional grave goods include two fragmented bronze bracelets of convex-concave shape and flattened spiral arms, and ceramic beads.

• KZ-KAR-010, Kairan I, Enclosure 10 (I4319): Date of 1746-1630 calBCE (3395±25 BP, PSUAMS-2543). Genetically male. Burial enclosure 10 is a round-shaped arrangement of stones with an internal diameter of 2.85 m. A rectangular grave cist was uncovered in the center of the enclosure, oriented along a southwest-northeast axis. The grave cist was built using four stone slabs erected vertically on their lateral edges. At the bottom of the grave in the eastern part, the bones of the lower half of the human skeleton were found in situ. Judging from these remains, the buried individual was interred in a flexed position on his left side, his head oriented in a western direction. In addition to human remains, fragments of a ceramic vessel were found.

S2.2.3.12 Kyzyl Bulak, Kazakhstan (n=2)

Description by Michael Frachetti

We present genome-wide ancient DNA from two individuals from Kyzyl Bulak, a multifeatured complex located approximately 50 km east of Almaty in the Zailiskii Alatau range. Kyzyl bulak-1 has a stone layout with four areas containing six wooden frame burials. Wooden frame burials were found in nearly every room, and contained ceramics, beads, and earrings. Seven burials were interpreted as cremations. Only four bodies, including one of an infant, were found in these wooden boxes.

Kyzlbulak_MLBA1 – This individual is part of the main Steppe Middle to Late Bronze Age genetic cluster, so we assign them to *Central_Steppe_MLBA* for population genetic analysis.

• KZ-KUZ-002, Kyzylbulak 1, burial 45 (I4323): Date of 1741-1627 calBCE (3385±20 BP, PSUAMS-2963). Genetically female.

Kyzlbulak_MLBA2 – This individual had evidence of admixture with groups related to those of BA Turan, and so we do not pool them with other individuals and instead assign the split label *Kyzlbulak_MLBA2.* For analysis we combine this individual with other individuals that have admixture with farmers from Turan and assign it the analysis label *Steppe_MLBA_oBMAC*.

• KZ-KUZ-001, Grave 46 (I4784): Date of 1618-1513 calBCE (3290±20 BP, PSUAMS-2613). Genetically male.

S2.2.3.13 Multiple sites near Krasnoyarsk, Western Siberia (n=18)

Description by James Mallory, Svetlana Svyatko, Alisa Zubova, Vyacheslav Moiseyev

We present genome-wide data from 18 individuals assigned to the northwestern parts of the range of territory occupied by the Andronovo archaeological complex, which flourished from 2000-900 BCE. Our data add to the genome-wide ancient DNA from four individuals reported previously from this region and cultural complex (*8*).

Andronovo is a complex material culture phenomenon, encompassing a large number of archaeological populations from the vast steppe zone between the Ural Mountains in the West and Minusinsk Basin in the East. The most important local Andronovo cultures are often considered as falling into Fedorovo, which occupied territory from Central Kazakhstan to the Yenisei River in present-day Russia, and Alakul', which occupied territory in the Southern Urals and Central Kazakhstan (*123*). According to archaeological data, both Fedorovo and Alakul' sites in Siberia have radiocarbon dates mostly between 1900 to 1400 BCE (*124*). Morphological data has been interpreted as suggesting that both Fedorovka and Alakul' skeletons are similar to Sintashta groups, which in turn may reflect admixture of Neolithic forest HGs and steppe pastoralists, descendants of the Catacomb and Poltavka cultures (*125–127*).

Genetic analysis indicates that the individuals in our study classified as falling within the Andronovo complex are for the most part genetically similar to the main clusters of *Potapovka*, *Sintashta*, and *Srubnaya* in being well modeled as a mixture of Yamnaya/*Western_Steppe_EMBA*-related and early European farmer- or Anatolian farmer-related ancestry (**Supplementary Materials S4**).

We assign the main cluster of 17 individuals the *Krasnoyask_MLBA* split label and the *Central_Steppe_MLBA* analysis label as they are genetically homogenous to other sites across the Central Steppe.

S2.2.3.13.1 Potroshilovo II cemetery, Russia (n=3)

Three individuals from which we successfully obtained ancient DNA were from the Potroshilovo II cemetery, excavated in 1989 by N. V. Leontyev. The cemetery is located at the mouth of the Tuba River, 3 to 3.5 km southeast from the former settlement of Potroshilovo. The cemetery consists of 11 enclosures (round 2.1-9.0 m in diameter, and rectangular 2.5-9.4 × 2.5-4.4 m), and 23 graves (log-house tombs 160-185 × 75-135 × 40-195 cm, cists 57-226 × 35-180 × 30-160). Most graves contained a single skeleton. Individuals were laid on the left or right side with heads turned to the southwest, and some were cremated. Grave goods consisted of pottery, awls, bone comb, bronze beads, rings, remains of leather, wood (possibly a vessel lid) and textile, cattle, sheep and roe-deer bones (*128*).

- SS8993-8, Potroshilovo II, enclosure 5, grave 1 (I1821): Date of 1612-1506 calBCE (3275±20 BP, PSUAMS-2435). Genetically male. The individual is skeletally determined to have been male (disagreeing with the genetic sex determination as female) and 25 to 35 years old at the time of death.

- SS8993-10, Potroshilovo II, enclosure 7 (I1853): Date of 1611-1503 calBCE (3270±20, PSUAMS-2437). Genetically male. The individual is skeletally determined to have been male and at least 40 years old at the time of death.

- SS8993-9, Potroshilovo II, enclosure 5, grave 3 (I1856): Date of 1685-1531 calBCE (3330±25 BP, PSUAMS-2067). Genetically male. The individual is skeletally determined to have been male and at least 40 years old at the time of death.

S2.2.3.13.2 Ust-Bir IV cemetery, Russia (n=3)

Three individuals from which we successfully obtained ancient DNA came from the Ust-Bir IV, cemetery, excavated between 1982-1983 by N. V. Leontyev. The cemetery is located on the right shore of the Birya River mouth. It consists of 28 graves (cists and log house tombs). One grave is an oval pit faced with vertical logs, and one grave is covered by Okunevo stele. Remains of cremations have also been found. For the most part there is a single individual in each grave, but in some cases, there are two individuals in a grave. Grave goods include pottery, bronze earrings and beads.

• SS7800-23, Ust-Bir IV, grave 26 (I1828): Date of 1611-1503 calBCE (3270±20 BP, PSUAMS-2960). Genetically female. The individual is skeletally determined to have been female and 20 to 25 years old at the time of death.

• SS7800-24, Ust-Bir IV, grave 28 (I1851): Date of 1611-1459 calBCE (3255±20 BP, PSUAMS-2436). Genetically male. The individual is skeletally determined to have been male and elderly at the time of death.

• SS7800-9, Ust-Bir IV, grave 10 (I1852): Date of 1623-15s and 8 calBCE (3295±20BP, PSUAMS-2865). Genetically female. The individual is skeletally determined to have been a probable male (disagreeing with the genetic sex determination) and 40 to 50 years old at the time of death.

S2.2.3.13.3 Chumysh-Perekat-1 cemetery (n=1)

Description by Alexey V. Fribus, Sergey P. Grushin and Alexey A. Tishkin

The cemetery of “Chumysh-Perekat” is located in the Zalesovski region (Altaiisky Kray, Russia), roughly 10.3 km southwest of the town Muravei. The burials are situated in the foothills of the Salair ridge on a high flat terrace along the right bank of the Chumysh River. Archaeological excavations were carried out under the auspices of a joint expedition of the Altaic and Kemerovo Universities, from 2014-2018. Excavations revealed a number of different archaeological complexes dating to diverse periods of prehistory (*129*).

• 138, BARN-014 (I11214): Context date of 2000-900 BCE. Genetically female . Anthropologically this appears to be a child burial.

S2.2.3.13.4 Orak burial place, Russia (n=11)

Eleven individuals for which we successfully obtained DNA were taken from skeletons excavated in 1926 and 1928 by G.P. Sosnovski at the Orak burial place. The site is located near Orak Ulus village in the Krasnoyarsk region of Russia. Most burials of the site are attributed to the Fedorovo type of the Andronovo complex while some are attributed to the Karasuk culture. The cemetery consisted of several oval kurgans. Square stone enclosures were built on the mounds. The enclosures touching each other formed a complex network (*130*).

The archaeological context of the skeletons excavated in 1926 is not well recorded, as the numbers of the burials in the publications do not match the documents in the museum catalog of the Peter the Great Museum of Anthropology and Ethnography where the collection is housed.

- StPet56, collection 3390, individual 4 (I3389): Context date of 1900-1400 BCE. Genetically male. The individual is skeletally determined to be male and 16 to 20 years old.

- StPet57, collection 3390, individual 2 (I3390): Context date of 1900-1400 BCE. Genetically female. The individual is skeletally determined to be female and 35 to 40 years old.

- StPet59, collection 3390, individual 6 (I3391): Context date of 1900-1400 BCE. Genetically male. The individual is skeletally determined to be 30 to 35 years old.

- StPet60, collection 3390, individual 7 (I3392): Context date of 1900-1400 BCE. Genetically male. The individual is skeletally determined to be male and 35 to 40 years old.

- StPet64, collection 3864, individual 4 (I6718): Date of 1616-1501 calBCE (3275±25 BP, PSUAMS-3911). Genetically male. This individual is genetically determined to be the father or son of I6718.

- StPet55, collection 6141, individual 12 (I6716): Calibrated date span of 1740-1534 calBCE [date of 1740-1614 calBCE (3370±25 BP, PSUAMS-3910) and date of 1686-1534 calBCE (3335±20 BP, PSUAMS-4254)]. Genetically female.

An additional three skeletons for which we obtained ancient DNA were excavated in kurgan 2 near Orak Ulus in 1928. The kurgan contained eight Andronovo and three Karasuk burials (assignment based on the archaeological context). The Andronovo-associated skeletons were found in crouched positions on their left side with their heads to the southwest and were placed in stone boxes, cists or pit-graves covered by logs. The artifacts found in the burials include ornamented ceramic vessels, and scattered copper beads in the female burials. There were also fragments of woolen cloth in two burials (N32 and 35) (*130*). The individuals from kurgan 2 used for genetic analysis are:

• StPet61, collection 3864, individual 1 (I3393): Context date of 1900-1400 BCE. Genetically male. This individual is a second-to-third-degree relative of I3394. This individual was found with a knitted woolen cap on his skull and is skeletally a male of approximately 50 years in age. This individual is father or son of I6718 and a second-to-third degree relative of I3394.

• StPet62, collection 3864, individual 2 (I3394): Context date of 1900-1400 BCE. Genetically male. This individual is a second-to-third-degree relative of I3393. The individual was genetically determined to be a male and is skeletally 27 to 30 years in age.

• StPet63, collection 3864, individual 3 (I3395): Context date of 1900-1400 BCE. Genetically female. This individual is skeletally determined to be a female of 25 to 27 years in age.

A final individual was found in a different Orak kurgan labeled as "kurgan on the swamp."

• StPet65, collection 3864, individual 5 (I3396): Context date of 1900-1400 BCE. Genetically male. Skeletally the individual is a female of 15 to 20 years in age.

Finally, we identified one genetic outlier, who we designate by the *Krasnoyarsk_MLBA_o* split label and the *Steppe_MLBA_oESHG* analysis label because of evidence for East Asian-related admixture.

• StPet58, collection 3390, individual 5 (I6717): Date of 1497-1406 calBCE (3160±20 BP, PSUAMS-4447). Genetically male. This individual is a genetic outlier in PCA or ADMIXTURE from the main *Krasnoyarsk_MLBA* group and hence we remove it from our primary analyses.

Relatedness summary:

- 3-person family: I6718 and I3393 are related as a father and son (order undetermined), while I3394 is a second-to-third degree relative of I3393.

S2.2.3.14 Priobrazhenka 3, Russia (n=1)

Description by Sergey M. Slepchenko and Dmitry Enshin

We assign this individual genetically to the *Western_Steppe_MLBA* analysis label.

The archaeological complex of Priobrazhenka 3 is situated in the forest-steppe zone of Baraba (3 km to the west from the village of Staraya Priobrazhenka, Chanovsky district of Novosibirsk region, on the right bank of the Om River) and consists of a settlement and a burial ground. The burial ground was studied in 1968 under the guidance of V.I. Molodina. There are 81 burials of the Andronovo culture and 113 individual and collective burials of the Irmen culture. The settlement complex is represented by Krotovo materials. Ceramic vessels, bronze and bone adornments, bronze tools (knives, arrowheads, etc.) are among the findings. The complex existed in the 2nd millennium BCE. Krotovo materials are the earliest, Andronovo antiquities chronologically come after them, and materials of the Irmen culture correlate with the late stage of the Bronze Age (*131*). This individual was part of the anthropological collection of Tomsk State University and the inventory numbers from the collection are provided along with that of the burial.

- Tomsk_2125, inventory number 2125, Kurgan 6 (I6048): Date of 1729-1563 calBCE (3350±20 BP, PSUAMS-2918). Genetically female.

S2.2.3.15 <u>Srubnaya culture, Samara Steppes, Samara oblast, Russia (n=1)</u>
Description by David Anthony and Alexander Khokhlov

We assign this individual genetically to the *Western_Steppe_MLBA* analysis label.

Novoselki is a kurgan cemetery of the Srubnaya or Timber-Grave culture dated to the Late Bronze Age (LBA), ca. 1800-1200 BCE. It is located 50 km north of Ulyanovsk, Russia, in a naturally forested environment rather than in the steppes 150 km to the south or in the transitional forest-steppe, where the majority of Srubnaya settlements and cemeteries are distributed. Novoselki is interesting because it is among the most northern of the Srubnaya sites, in a region where substantial interaction with Uralic populations might be expected. Here we co-analyze the data from an individual from this site (I0233) for which data are reported in this study with that of another individual from the same site for whom data was previously reported (I0232) and for whom we add additional data in the present study.

• SVP HB 13, Novoselki, Kurgan 2, Grave 4 (I0233): Context date of 1850-1200 BCE. Genetically male.

S2.2.3.16 Solyanka Vodokhranilische, western Kazakhstan (n=1)

Description by Gian Luca Bonora, Stefania Sarno and Davide Pettener

We assign this individual genetically to the *Western_Steppe_MLBA* analysis label.

The burial ground of Solyanka Vodokhranilische is located in western Kazakhstan, along the lower course of the Solyanka river, left tributary of the Ural river in its middle course, about 106 km south of Ural'sk and 15 km east of Chapaevo (*132*). A few funerary barrows were identified here and all were poorly preserved. Only barrow number 4 was excavated in 1983.

• DI No 59, Kurgan 4, Grave 3 or 4 (I3864): Date of 1728-1546 calBCE (3345±20 BP, PSUAMS-3183). Genetically female. Morphologically male, 25 to 30 years old. Barrow number 4 had a diameter of 32 m, while its height was 0.32 m. The excavation of the fill of the barrow brought to light some animal bones and numerous fragments of burnt wood. Six burial pits were identified, all of which had been plundered in antiquity. During the excavation led by G.A. Kushaev, archaeologists determined that only burials number 3 and 4 contained human osteological remains. All other graves were void of both of human osteological remains and funerary inventory.

S2.2.3.17 Aktogai, Kazakhstan (n=5)

Descriptions by Michael Frachetti, Egor Kitov, and Gaziz Akhatov

We assign these individuals genetically to the *Western_Steppe_MLBA* analysis label.

We present genome-wide ancient DNA data for 5 individuals from the burial ground of Aktogai (area-7), located on a small natural elevation in Kazakhstan. Aktogai consists of four stone burial enclosures: 2, 3, and 4 are aligned in the north-northeast to south-southwest direction, and enclosure 1 is located to the east of the other three. The funerary sites at Aktogai reflect constructions typical of the Late Bronze Age: round and circular stone burial enclosures containing internal grave cists lined with vertically placed stone slabs. The number of grave cists within each burial enclosure ranges from one to four. The graves resemble both stone boxes and stack-cists (boxes constructed of stone tiles of small size stacked one on top of the other) and capped with two to four layers of overlapping stone slabs.

The interred individuals are typically buried in a flexed position on their left side with their head oriented to the west. Individuals buried at Aktogai include both adults and children. The funerary inventory is represented by open form handmade ceramic vessels, often with comb-impressed geometric designs. Grave goods also included bronze objects such as needles and twisted bronze/gold foil hair pendants that are typical in burials of the late Bronze Age in the region. Part of Aktogai was robbed in antiquity. On the basis of radiocarbon dating of the skeletal remains, the cemetery likely dates from the early 17th to late 16th centuries BCE.

- KZ-AKT-001, object 7, Burial enclosure 2, Grave 1 (I4264): Date of 1691-1528 calBCE (3335±30 BP, PSUAMS-2124). Genetically female. Burial enclosure 2 consists of vertically placed stone slabs arranged in a rectangular shape, roughly 10 m from north to south, 7 m from west to east, and 0.2 m in height. Three stone grave cists, each containing a single individual, were recovered within the burial enclosure and aligned in a north-south orientation. Grave 1 in this burial enclosure is defined by a stone-lined cist of rectangular shape, oriented with longitudinal edges along the east-west axis. The funerary cist was constructed by composite techniques of layering and aligning flag-stone slabs. The longitudinal walls consist of solid rectangular granite slabs, while the upper portions of the end-walls were formed by stacking rectangular shaped stones on top of each other in three to four rows. Beneath these stacks of stones, the lower portion of the cist consisted of vertically placed plates. Individual I4264 was found at the bottom of the cist, lying in a flexed position on her left side, with head oriented to the northwest. Here hands were positioned in front of her face and her legs were strongly bent and pressed to her chest. A handmade ceramic pot with incised tree-styled ornamentation was recovered from behind the head of the interred individual, near the northern wall. In addition, two fragments of bronze wrapped with gold foil, a bell-shaped bronze earring, and bronze beads were recovered.

- KZ-AKT-002, Object 7, Burial enclosure 2, Grave 2 (I4773): Date of 1618-1513 calBCE (3290±20 BP, PSUAMS-2607). Genetically male. Burial enclosure 2, grave 2, is located in the northern part of the burial ground, oriented with longitudinal edges along the E-W axis.

The grave 2 cist was built using granite slabs, erected vertically on their lateral edges. Within the cist the skeleton was positioned in an elongated position along the southern wall, the head oriented to the east. The arms are stretched along the body, with the left arm bent at the elbow and placed upon the pelvic bone. The head is slightly raised due to a stone placed under it. The bones of the vertebrae are slightly curved to tilt towards the north.

- KZ-AKT-003, Object 7, Burial enclosure 2, Grave 3 (I4265): Date of 1640-1527 calBCE (3310±20 BP, PSUAMS-2511). Genetically male. Burial enclosure 2, grave 3, is located in the southern part of the burial ground. Grave 3 is a rectangular stone cist built from four solid rectangular granite slabs placed upright along their lateral edges. At the bottom of the grave, poorly preserved and disarticulated remains of an adolescent were recovered along with a poorly preserved ceramic. Judging by the position of the bones, the individual was positioned on his left side, head oriented to the west.

- KZ-AKT-006, Object 7, kurgan 3, burial 1 (south) (I10140): Context date of 1400-1200 BCE. Genetically male.

- KZ-AKT-008, Object 7, Kurgan 4 (I4774): Date of 1615-1509 calBCE (3285±20 BP, PSUAMS-2608). Genetically male. Burial enclosure 4 is built with an ovular outer ring of stones and contains a central grave cist composed of stacked stone walls. Within the grave cist, the body was placed in a flexed position on its left side, head oriented to the west. A ceramic vessel was placed near the head and other grave goods included a bronze needle.

S2.2.3.18 Kazakh Mys, Kazakhstan (n=4)

Descriptions by Michael Frachetti, Gaziz Akhatov, Erlan Kazizov, and Egor Kitov

We assign these individuals genetically to the *Central_Steppe_MLBA* analysis label.

The site of Kazakh Mys is located roughly 22 km east of the modern city of Aktogai, Kazakhstan (in the vicinity of the Bronze Age Aktogai cemetery described above). The burial ground is defined by a group of four funerary constructions of the Late Bronze Age and 11 “kurgans” of the Early Iron Age. Human bones were sampled from collected remains of the Bronze Age graves, defined at the time of excavation based on the presence of rectangular and round stone enclosures. In each burial enclosure there were from one to four grave cists.

Graves were built either as slab cists with large stones stood on their lateral sides, or as stacked cists constructed using stacked flagstones, typically with three to four layers. Human remains were usually placed in a flexed position on their left side with their head oriented toward the west.

The inventory of grave goods included open form handmade ceramic vessels, often with comb-impressed geometric designs, bronze objects such as needles, as well as twisted bronze/gold foil hair pendants; all typical in burials of the late Bronze Age in the region. Part of the burial site was robbed in antiquity. On the basis of radiocarbon dating of the skeletal remains at Kazakh Mys, the Bronze Age portion of the cemetery most likely dates from the late 17th to early 15th centuries BCE, roughly contemporary to Bronze Age Aktogai.

- KZ-KAZ-005, Kazakh Mys 2015, Area 1, Grave 1 (I4322): Date of 1611-1503 calBCE (3270±20 BP, PSUAMS-2544). Genetically male. Area 1, grave 1 is a burial cist made of stacked flagstones. The skeletal remains were disarticulated and poorly preserved, lying in a flexed position on the left side, head oriented to the west. The few skeletal elements that remained *in situ* included the skull, pelvis, femora and tibia. Near the skull a ceramic vessel was recovered, decorated with a comb impressed stamping of oblique lines, horizontal lines, and herringbone lines in distinct sections across the vessel body.

- KZ-KAZ-006, Kazakh Mys 2015, Area 3, Grave 1 (I4783): Date of 1610-1454 calBCE (3250±20 BP, PSUAMS-2612). Genetically male. Area 3, grave 1 is a stacked stone cist, built with roughly three to six layers of flagstones. The human remains were poorly preserved, lying in a flexed position on their left side, head oriented to the northwest. Near the skull a ceramic vessel was recovered, decorated with comb impressed stamps of fluted lines around the neck and herringbone lines in distinct sections across the vessel body.

- KZ-KAZ-003, Kazakh Mys 2015, Area 6, Grave 1 (I4321): Date of 1640-1527 calBCE (3310±20 BP, PSUAMS-2962). Genetically female. Area 6 is defined by a rectangular stone burial enclosure with central burial cists aligned in a row in N-S orientation. Area 6, grave 1 is the most northern stone cist in the row. The rectangular cist is constructed with stacked flag stones, built in four to six rows, as well as larger slab stones on the eastern and western ends of the grave. At the bottom of the funerary construction, the disorderly and poorly preserved bones of a human skeleton are spread throughout its area. Funeral inventory was absent. Because of a robbery and the complete absence of *in situ* bones, it is impossible to determine the pose and orientation of the buried individual.

- KZ-KAZ-004, Kazakh Mys 2015, Area 6, Grave 1 (I4782): Date of 1736-1621 calBCE (3370±20 BP, PSUAMS-2915). Genetically female. Area 7, grave 1 was constructed as a stone slab cist, with four large stones erected vertically along its lateral edge in an E-W orientation. The skeletal remains were apparently placed in a flexed position with the head facing west. Grave goods included an open form, handmade ceramic vessel with no decoration.

S2.2.3.19 Dashti-Kozy, Tajikistan (n=3)

Description by Michael Frachetti

We assign these individuals genetically to the *Steppe_MLBA_oBMAC* analysis label to reflect the fact that they have ancestry that is intermediate between the *Western_Steppe_MLBA/Central_Steppe_MLBA* cluster and the *BMAC* cluster.

We present genome-wide ancient DNA data from three individuals from the Bronze Age cemetery of Dashti-Kozy, located on the northern bank of the Zeravshan River, roughly 50 km to the east of present-day Pendzhikent in the Republic of Tajikistan. The burials were first supervised by A. I. Isakov (1983-1984) and primarily studied and later published by T. Potyomkina (1985-1986) (*133*). The cemetery spans an area roughly 400 m^2, and contains 27 graves (21 single graves, 4 two person graves, 1 three person grave, and one collective grave), with a total of 39 buried individuals (*134*). Dashti-Kozy has been interpreted as one of the most southern examples of "steppe" burials, on account of the mixed grave goods that have material culture analogues amongst steppe communities in central Kazakhstan and along the Inner Asian Mountain Corridor (*50*, *134*). It is also important to note that within the material assemblage recovered from the graves are ceramic forms and material types commonly recovered among farming villages of southern Central Asia, indicating that the population buried at Dashti-Kozy was interacting with a diversity of neighboring groups, at least in terms of their material belongings. The genetic ancestry of these individuals resonates with the archeological evidence in showing admixture between ancestry typical of the *Central_Steppe_MLBA* genetic cluster and the main *BMAC* cluster.

- UZ-DK-001, Dashti-Kozy, 44-33, burial 15 (I4258): Date of 1620-1506 calBCE (3285±25 BP, PSUAMS-2121). Genetically female. Burial 15 is located in the southwestern quarter of the cemetery (Grid G/9) and consisted of an oval burial pit roughly 1.6 m across. The remains of a single inhumation were recovered within the burial pit: a mature female lying *in situ* on her left side in a flexed position facing north. The body was adorned in jewelry including bronze bracelets, earrings, and beads typical of steppe designs seen throughout central and southeastern Kazakhstan from the same time period (*51*). The grave also contained a small hand-thrown ceramic pot with incised geometric designs, typical of several steppe cultural types and time periods.

- UZ-DK-003, Dashti-Kozy 1985, Burial 12, 44-34, individual 1 (I4257): Date of 1518-1434 calBCE (3215±20 BP, PSUAMS-2510). Genetically female. Burial 12 at Dashti-Kozy is located in the southwestern quarter of the cemetery (Grid G/8). The remains of three individuals were recovered within grave 12. The body positions of individual 1 (I4257) and individual 2 were disturbed (likely in antiquity), though the burial position of individual 3 (I4160) was more intact (described below). Although the original burial context was apparently disturbed before excavation, the archaeologist recovered 76 small beads as well as an incised ceramic vessel, analogous to findings from a number of steppe material typologies and time periods.

- UZ-DK-004, Dashti-Kozy 1985, Burial 12, 44-35, individual 3 (I4160): Context date of 1650-1400 BCE. Genetically female. Individual 3 in Burial 12 (described above) was intact enough to determine she was interred in a flexed position, lying on her left side. Indiv-3 was not directly dated.

S2.2.3.20 Kokcha 3, Kokcha, Uzbekistan (n=3)

We assign all three individuals to the *Steppe_MLBA_oBMAC* analysis label to reflect the fact that they have ancestry that is intermediate between the *Western_Steppe_MLBA/Central_Steppe_MLBA* cluster and the *BMAC* cluster.

- MOS309, 1955, Kokcha 3 1955, Grave N77, IE-60-I9 (I8506): Context date of 2500-1500 BCE. Genetically male.

• MOS310, Kokcha 3 1954, Grave N5, Skull N1, IE-60-20 (I8507): Context date of 2500-1500 BCE. Genetically male.

• MOS308, Kokcha 3 1951, Grave N90, IE-60-8, 659 (I12499): Context date of 1600-1300 BCE. Genetically male.

S2.2.3.21 Oy-Dzhaylau III, Kazakhstan, (n=7)

Descriptions by Gian Luca Bonora, Stefania Sarno, and Davide Pettener

We present genome-wide ancient DNA from eight individuals from the site of Oy-Dzhaylau (sometimes transliterated as Oi-Dzhailau, Oi-Zhailau) burial ground, located in southern Kazakhstan, 40 km west of the Otrar railway station, in the Kurday district of the Zhambyl region. The funerary structures are spread on a plateau, which sits at about 1200 m above sea level. The plateau extends approximately 6 to 7 km from east to west and 4 to 5 km from north to south and different clusters of graves have been identified.

The burial ground of Oy-Dzhaylau III, located in the western part of the plateau, covers an area of about 1500 sq. m., 860 of which were excavated in the late 80s (*135*). A total of 51 Bronze Age burials, represented by quadrangular stone fences containing a cist burial with the inhumation of an adult individual, were identified. There were also a few exceptions that contained infant burials, namely pit graves 25, 34 and 35 (without a stone fence).

The main group of 6 Oy-Dhzaylau individuals is assigned to the *Central_Steppe_MLBA* genetic analysis label.

• KZ-DJA-005, Oy-Dzhaylau 7, Grave 4 (I7060): Date of 1608-1437 calBCE (3235±25 BP, PSUAMS-3728). Genetically female.

• KZ-DJA-001 + KZ-DJA-002, Oy-Dzhaylau 7, Construction 3, Grave 1, bones without skull (I4789): Date of 1617-1498 calBCE (3270±25 BP, PSUAMS-2964). Genetically female.

• KZ-DJA-004, Oy-Dzhaylau 7a, Grave 7 (Construction 3) (I4790): Date of 1527-1439 calBCE (3220±20 BP, PSUAMS-2547). Genetically male. This individual is genetically detected as a 2nd to 3rd degree relative of I4791.

• KZ-DJA-006, Oy-Dzhaylau 7, Grave 5 (Construction 3) (I4791): Date of 1609-1450 calBCE (3245±20 BP, PSUAMS-2548). Genetically male. This individual is genetically detected as a 2nd to 3rd degree relative of I4790.

• BI No 22, Grave 4 (I3861): Date of 1872-1684 calBCE (3435±20 BP, PSUAMS-3115). Genetically female. Morphologically male, 35 to 40 years old. Grave 4 was located in the northern section (square V2) of the graveyard ((*136*): 12, Fig 5). The burial is represented by a rectangular stone fence (1.5 x 1.1 m), whose main axis is positioned east to west (*137*). The burial (1.05 x 0.65 m), which faces the same direction as the fence, was found at a depth of 0.1 m from the modern surface. The pit was 0.55 m deep and the skeleton of a sub-adult individual was buried therein on its left side in a flexed position, with the skull oriented towards the west and the head facing north. The arms were bent at the elbow with the right arm placed very close to the ribs and the left moved slightly away from the chest. The hands were 10 to 15 cm from the frontal bone of the skull. The legs were closely bent at the knees and the feet were near the pelvic bone. A hand-made vessel in dark greyish ware found in the northwest corner of the burial was the only artifact of the funerary inventory. The body of the vessel was decorated with two carved parallel lines enclosing a series of vertical signs in the shape of wolf teeth.

• BI No 21, Burial 34 (I3788): Date of 1683-1532 calBCE (3330±20 BP, PSUAMS-2499). Genetically male. Burial 34 was in section E-3 of the burial ground ((*136*): 12, fig. 5). It was represented by a stone cist (2 x 1.7 m) which was built adjacent to stone cist 35. In fact, burials 34 and 35 shared one wall, represented by a stone slab set vertically in the ground (*137*). Burial 34 was positioned in a northeast to southwest direction and contained two skeletons: a 17 to 18-year-old sub-adult male and an 18 to 20-year-old young female. The male skeleton, well preserved, was located near the southern wall of the cist lying on its left side in a flexed position, with the skull oriented toward west. No grave goods were found in association with the male skeleton. The skull of the female individual was not found and her lower post-cranial bones were extremely poorly preserved. This second corpse was located along the northern wall of the stone cist, presumably placed on her left side, in a flexed position, with the skull facing west. Two hand-made fragmentary vessels were found near the now absent skull. One of them bore no decoration while the second was adorned with two large grooves below the rim. The wrists of the female skeleton were bedecked with metal bracelets with spirals at their ends; two bracelets were around her left wrist and one around the right. Bronze and paste beads adorned the hands and the ankles of the young woman.

One individual has genetic evidence of admixture between ancestry typical of the *Central_Steppe_MLBA* analysis label and the main *BMAC* cluster, and so for analysis we assign an analysis label of *Steppe_MLBA_oWSHG*.

- BI No 20 Grave 24, Section D-4 (I3860): Date of 1734-1617 calBCE (3365±20 BP, PSUAMS-2492). Genetically and morphologically male, 20 to 25 years old. Burial 24 was located in the northeastern part of section D-4. It was poorly preserved and archaeologists were unable to identify traces of the burial pit (*138*). On account of this, it has been hypothesized that the burial was plundered and partially destroyed in the past and the skeleton was extracted from the pit. The osteological remains of the arms, legs, ribs, pelvis and vertebrae were gathered in a heap at a depth of 0.5 m from the modern surface, while the skull was located southwest of the pile, slightly moved away. No funerary inventory was found in association with this skeleton.

Relatedness summary:
- 2-person family: I4790-I4791 are genetically detected as 2nd to 3rd degree relatives.

S2.2.3.22 Taldysay, Central Kazakhstan (n=2)

Descriptions by Michael Frachetti, Egor Kitov, Gaziz Akhatov, and Antonina Yermolayeva

We present genome-wide ancient DNA data from two individuals found at the ancient settlement of Taldysay. First studied by A.S. Yermolayeva, Taldysay is located in the Ulytau administrative district of the Karaganda region, 77 km from the city of Zhezkazgan in the direction of the Ulytau Mountains, in the homonymous tract at the confluence of the Bala Zhezdy and Ulken Zhezdy rivers. Beginning in 1994 and up to the present time, Taldysay has been investigated by the Central Kazakhstan archaeological expedition.

Taldysay_MLBA1: This individual falls in the *Central_Steppe_MLBA* analysis label.

• KZ-UKZ-002, Grave (?) 1, KV E7 (I4787): Date of 1379-1196 calBCE (3015±20 BP, PSUAMS-2614). Genetically male. In 2010, excavation site 2 uncovered a child's burial in square G/11, located on the western outskirts of the settlement. This part of the settlement is constantly flooded by the rivers in spring, destroying surface layers and exposing a gray colored metallurgical layer. The burial was located in the metallurgical layer along with cultural remains. The skeleton of the baby was lying on the left side with the head positioned to the north, arms bent at the elbows. The skull was crushed, both hands were not preserved, and the spine was disturbed. Behind the skull on the northern side stood a pottery-can jar without ornamentation; inside the jar was a fragment of a vessel with ornamentation typical of early Alakul dishes (dating to the beginning of the 2nd millennium BCE). Here the AMS chronology illustrates a discordance between the formal characteristics of the ceramics and their typological attribution. The AMS date obtained directly from the human bone fixes the period of death somewhere between 1379-1196 calBCE, roughly 500 years later than the purported chronology of Alakul cultural materials.

Taldysay_MLBA2: Genetically, this individual has admixture between ancestry typical of the *Central_Steppe_MLBA* analysis label and the main *BMAC* analysis label, and so we designate it for analysis as *Steppe_MLBA_oBMAC*.

• KZ-TAL-001, Taldysay 1, Grave N5, Central Kazakhstan Archaeology (project) 7.02 (?) (I4794): Context date of 1600-1400 BCE. Genetically male. This individual is genetically systematically different from I4787. The grave is associated with a metallurgical pit and chimney. Two skeletons of cattle were laid on the slabs of the chimney block. On both sides of the chimney lay separate parts of the skeletons of animals. On the northern side of the chimney there were two human skeletons, both of which lay in twisted positions. The southern skeleton lay twisted on the right side, the skull to the south with a slight deviation to the west. The northern skeleton lay at the feet of the southern one also on the right side facing down, with a slight deviation to the east. The humans and animal skeletons have been interpreted as sacrifices related to metallurgical production.

S2.2.3.23 Talapty II, South Kazakhstan (n=1)

Description by Gian Luca Bonora, Stefania Sarno and Davide Pettener

We assign this individual on genetic grounds to the *Steppe_LBA* analysis label (characterized by substantial East Asian-related admixture).

The burial ground of Talapty II is located in the Koksu river valley within the Eshkiol’mes Mountains, along the western edge of the Dzhungar Alatau of Southern Kazakhstan. Talapty II is located 400 m west of the Talapty I burial ground, which lies on the upper fluvial terrace of the Koksu. Thirteen graves were excavated here by A.N. Mar'yashev and A.A. Goryachev (*135*). They are represented by simple pits (in three cases), a ground pit with a rectangular structure at its bottom (one case) and stone cists (nine cases). The funerary inventories found in the graves of this burial ground included a variety of pottery shapes and decorations that have been attributed to the Fedorovo, Alakul' and Sargary material cultural traditions. The first of these traditions, in its late variant, is predominant as evidenced by the co-occurrence of two diverse funerary rituals: cremation and inhumation. These observations suggest that the dating to the second half of the 2nd millennium BCE is too conservative. It is reasonable to suggest a chronological attribution to the Late Bronze Age between 1650/1600 and 1400/1350 BCE.

- BI-25 (I3769): Context date of 1400-1000 BCE. Genetically female. Stone fence number 2 was square in shape (each side was 3.8 m long) with the corners positioned facing the four cardinal points (*139*). At the center of the fence, at a depth of 0.7 m, a simple rectangular pit was found (2 x 0.95 m). It was positioned in an east-west direction and contained a partially preserved skeleton. According to the few bones found *in situ*, the corpse was buried on its left side, in a flexed position, with the skull facing west. The arms were bent at the elbow and the hands were in front of the face. The funerary inventory consisted of two bronze earrings and a hand-made vessel in rough ware tempered with sand and quartz. It was set between the skull and the northwestern corner of the pit.

S2.2.3.24 Kashkarchi Burial Mound, Ferghana Province, Uzbekistan (n=2)
Description by Michael Frachetti

Genetically, these two individuals fall within the *Central_Steppe_MLBA* analysis label.

The Kashkarchi Burial Mound is located in the Kashkarchi Area of the Ferghana Province, 6 km to the north of the town of Ferghana (near the road from Ferghana to Andijan). The site was visited by B.Z. Gamburg and N.G. Gorbunova in 1954, but G.P. Ivanov (1988) excavated and published the site, which consists of six burials (generally aligned in a row from north to south) (*140*). On the surface the burials are identified as rock piles covering an area of 50 x 25 m, each with a height of 2 m. Oval burial pits (2.80 x 1.50 m) were dug in the layer of gravel conglomerate aligned (individually) from northwest to southeast, with a general depth of 1.40 m, located very close to each other, so that sometimes the shaft edges overlap, while others have a distance of 40 to 60 cm.

On the northern side of the bottom of each burial shaft there are round holes with a diameter of 50 cm, filled with rocks and vertically leading down to the burial chambers located to the north of the shafts. The chambers had an oval shape but were merged due to being very close to each other, thus becoming a unified underground "corridor" 13.5 m long, 2 m wide and with a vault (arch) height of 160 cm. All the burials were robbed in ancient times. At a depth of 370 cm, the skeletons of ten adults and eight children were found interred on their right side in a flexed position, heads turned to northeast and facing southward towards the exit from the chamber. Some bones are colored with red ochre. Grave goods include ceramic vessels, golden, bronze and silver spiraled and ring-shaped earrings, a round bronze button, bronze awls, knives, beads, bracelets, a scraper, and a cosmetic stick made of stone. The ceramic vessels are handmade of gray clay with inclusion of limestone. The vessels are mainly flat-bottomed pots with rounded shoulders and narrow necks, with their top edges bent outward and scratched geometric ornamentation on the outside surface.

According to the author, similar burial constructions are known for Ferghana and other parts of Central Asia, and based on this the author interpreted it as a cemetery of an extended family/clan of mobile pastoralists. The pottery is synchronized with a late stage of the Kayrakkum Culture (the Ferghana variant of Andronovo Culture). The type of bronze knives found here were dated by Kuzmina E.E. to the late 2nd to early 1st millennium BCE; similar earrings were dated by Avanessova N.A. to the same period (13th to 12th centuries BCE).

- UZ-KKC-001, Kashkarchi, burial N. 1, 14-46 (I4153): Context date of 1200-1000 BCE. Genetically male. This individual is genetically detected as being a 2nd or 3rd degree relative of I4255.

• UZ-KKC-003, Kashkarchi, N. 3, 14-47 (I4255): Context date of 1200-1000 BCE. Genetically male. This individual is genetically detected as being a 2^{nd} or 3^{rd} degree relative of I4153.

Relatedness summary:

• 2-person family: I4153-I4255 are genetically detected as 2^{nd} to 3^{rd} degree relatives.

S2.2.3.25 Bylkyldak, Central Kazakhstan (n=1)

We used the analysis and split label *Bylkyldak_MLBA*.

The Andronovo cultural burial ground of Bylkyldak 1 is located 35 km to the northeast of Aksu-Ayuly village, in the Karagana region on the left bank of Bylkyldak River. The site was excavated in 1951 by A. H. Margulan. The burial ground consists of 70 stone enclosures. Only 15 of them were excavated, of which 14 were attributed to Atasu period of Andronovo culture and one to the Iron Age.

• StPet36, collection 6037, individual 3, engraving 6, skeleton 9 (I6707): Date of 1596-1439 calBCE (3230±20 BP, PSUAMS-2978). Genetically female. Skeletally a 17-20-year-old female.

S2.2.3.26 Shoendykol, Baianaul District, Pavlodar Region, Eastern Kazakhstan (n=3)

Description by Sergey Slepchenko and Dmitry Enshin

We assign these three individuals genetically to the *Central_Steppe_MLBA* analysis label and the *Shoendykol_MLBA_Fedorovo* split label.

Shoendykol burial ground is situated on a flat, turfed site 200 meters north of the shore of Lake Dzhasybai, Baianaul District, in the Pavlodar Region of the Republic of Kazakhstan. Five fences made of stone slabs, rising 0.1-0.15 m above the contemporary surface, were found. The cemetery was studied within two continuous excavations (*141*). Three stone fences were opened during the first excavation, and one stone fence and 15 Andronovo archaeological culture burials were opened in subsurface pits during the second excavation.

The archaeological analysis revealed that in ancient times, fence 1 was a rectangular structure made of granite slabs which were dug into the ground to a depth of about 8 cm. Three granite 1.6 m long beams were found at the corners of the structure, and it is possible that they were placed vertically at the corners of the structure in ancient times. One burial in the form of a stone box was discovered inside the fence. The box was made of untreated slabs up to 6 cm thick and dug 55 cm below the surface. The burial dimensions are 1.15 x 0.65 m at the surface, and 0.8 x 0.5 m at the bottom of the burial. A vessel of the Alekseevka-Sargary culture was found in the burial. No human bones were found there.

Fence 2 was made of stone slabs with the dimensions of 3.4 x 3 m. The slabs formed an oval structure of 2.6 x 2.1 m in ancient times. A stone box was found inside the fence close to its northern wall. Its dimensions at the surface were 1.7 x 1.0 m, decreasing towards the bottom where its dimensions were 1.5 x 0.7 m. A skeleton of a 40 to 50 year old woman, buried in a crouched position on the right side, was found at the bottom of the grave. An unornamented pot-shaped vessel was found in the corner of the box, closer to the head.

Similar to fence 2, fence 3 was oval-shaped. It had dimensions 3.4 x 2.1 m. The construction method was similar to fence 2, but the original dimensions could not be established. The burial box was made of granite slabs and had the dimensions of 1.68 x 1.1 m at the surface level and internal dimensions of 1.67 x 0.85 m. The box was dug 0.75 m deep. A skeleton of a man 50 to 60 years old was found at the bottom of the box in a crouched position. Ornamented pot-shaped vessels were found on both sides of the head.

According to the ornamented ceramics found in the burials inside the fences of Shoendykol burial ground, they belonged to representatives of the Alekseevka-Sargary culture. The subsurface burials belonged to the Andronovo archaeological culture. These individuals were all part of the anthropological collection of Tomsk State University and the inventory numbers from the collection are provided along with those of the burials.

- Tomsk_4376, inventory number 4376, Kurgan 2, burial 2 (I10110): Context date of 1400-900 BCE. Genetically female.

• Tomsk_4377, inventory number 4377, Kurgan 3, burial 3 (I10111): Context date of 1400-900 BCE. Genetically male.

• Tomsk_4379, inventory number 4379, burial 7 (I10112): Context date of 1600-1400 BCE. Genetically male.

S2.2.4 Sites from the Steppe Zone in the LBA

S2.2.4.1 Chanchar, Kazakhstan (n=2)
Description by Vitaliy V. Tkachev and Michael Frachetti

The site of Chanchar (or Shanshar) is located in western Kazakhstan in Aktyubinsk province. The site is situated on the interfluvial terrace between the Zhaksy-Kargala River and its tributary, the Shanshar River. The cemetery consists of 38 stone and earth kurgans, with stone constructions in both round and rectilinear forms. The chronology of the kurgans at Chanchar spans from the Middle Bronze Age to the Early Iron Age. On the basis of their form, the construction of the stone burial enclosures, and in some cases the ceramics, the majority of the burials are dated to the Bronze Age and associated with the so-called "Kozhumberdyn cultural group," which is a sub-category of Late Bronze Age "Alakul" culture characterized by the presence of integrated "Fedorovo" stylistic components (*142*).

• Kurgan 2 (I11537): Context date of 1500-1100 BCE. Genetically male. This burial is located in the northeastern burial group. Beneath the kurgan mound which was 15m in diameter and 1m in height, excavators recovered a large stone circle roughly 11m in diameter and a central burial cist built of massive stone boulders dug into the soil. This construction appears to be associated with burial 4, which was nearly completely destroyed during subsequent burials of the early Iron Age. Within the central burial cist, human skeletal remains along with undecorated pottery fragments were recovered. We assign this individual genetically to the *Steppe_LBA* analysis label.

• Kurgan 2, Burial 1 (I11538): Context date of 1500-1100 BCE. Genetically female. This burial was located to the south of the concentration of stones, in a discrete pit. The authors dated Burial 1 to the Early Iron Age, although diagnostic archaeological materials were not reported. We assign this individual genetically to the *Chanchar_LBA* analysis label (because it is an outlier from the *Steppe_LBA* analysis label).

S2.2.4.2 Georgievsky Bugor, Kazakhstan (n=2)

• Burial 3 (I11540): Context date of 1500-1100 BCE. Genetically male. We assign this individual genetically to the *Steppe_LBA* analysis label.

• Burial 8 (I11541): Context date of 1500-1100 BCE. Genetically male. We assign this individual genetically to the *Central_Steppe_MLBA* analysis label.

S2.2.4.3 Guruldek, Kazakhstan (n=1)
Description by Arman A. Bissembaev and Vitaliy V. Tkachev

The Late Bronze Age cemetery of Guruldek (or Guryuldek) is located in the north of the Aktyubinsk region (Kazakhstan). The site consists of 14 kurgan burials each with a single burial chamber, four of which were investigated. The burial chambers within the mounds were built mostly of stone and timber, and grave goods included pottery and personal ornaments. The artifacts recovered at Guryuldek share stylistic qualities with Alakul-type sites, which establish the chronology of the cemetery within the developed stages of the Late Bronze Age. The site is attributed to the Kozhumberdy cultural group, which is represented by a combination of the Alakul and Fedorovka characteristics .

• Kurgan 1, Burial 2, right skeleton (I11542): Context date of 1500-1100 BCE. Genetically female. Kurgan 1 is the largest of the mound burials at Guryuldek, with a diameter of 16m and a height of 1.37m. Within the mound, the grave was a stone enclosure made from cobblestones, and was 1m in length and roughly 60cm in width. The skeletal remains of an adult couple were recovered at a depth of 65cm. The skeletons were aligned facing one another: the “left skeleton” lay on its left side in a flexed position, with arms and legs bent upward toward the torso. The second (or “right”) individual was an adult anthropologically determined to be male (despite the genetic determination as female) placed on the right side, with the skull displaced from the post-crania presumably due to taphonomic disturbances (i.e. this was probably not the original burial position). Within the grave there were three hand made ceramic vessels, all with flat bottoms. One vessel was decorated with incised geometric patterns typical of the Alakul material assemblage. While no direct dating was performed on these skeletons, the diagnostic nature of the ceramics serves to date this site roughly to the middle centuries of the 2nd millennium BCE in accordance with other "Alakul" type sites of the region. We assign this individual genetically to the *Western_Steppe_MLBA* analysis label.

S2.2.4.4 Priobrazhenka 3, Russia (n=1)

The archaeological description is given earlier (associated with individual I6048 from the same site who is radiocarbon dated to about 350 years earlier in the MLBA).

- Tomsk_2124, inventory number 2124, Kurgan 19, burial 5 (I6047): Date of 1384-1213 calBCE (3025±20 BP, PSUAMS-2917). Genetically male. We assign this individual genetically to the *Steppe_LBA* analysis label.

S2.2.4.5 Zevakinskiy stone fence, Kazakhstan (n=8)

Description by Michael Frachetti, Egor Kitov, and Gaziz Akhatov, Davide Pettener, Gian Luca Bonora, and Stefania Sarno

We present genome-wide data from six individuals originating from a burial ground 5 km northwest of Zevakino in eastern Kazakhstan, also known as Zevakinskiy. The site is located on the right bank of the Irtysh river, near the confluence with the Marinka stream and at the foot of the Ubinskiy Mountain range in Eastern Kazakhstan. The cemetery was first investigated by F.H. Arslanova, who documented more than 500 burials ranging from the Bronze Age to the Middle Ages (*143*). The individuals analyzed in the present study belong to Bronze Age, and individuals were excavated from surface-visible stone enclosures associated with this period. Within the stone enclosures are stone cist graves containing individuals lying typically on their right side (less often on the left) in a flexed position, with their head oriented to the west. Occasional orientations to the north or south are also observed.

Ceramic vessels found in the cist graves are divided into two types: jars with a smooth profile line and straight or bent rim, and vessels with a convex body, no definable neck, and a straight or concave rim. Pottery decoration was typically in the form of comb-impressed designs, sometimes covering the majority of the vessel surface. Primary decorative elements include zigzags, horizontal ribs, grooves, and geometric patterns of triangles, characteristic of dishes made during the Fedorovo stage of East Kazakhstan and the Yenisei Valley. We note that the radiocarbon dates obtained from 6 out of the 8 human remains fall in the 12th to 10th centuries BCE, whereas Federovo cultural assemblages are now well dated to the 18th to 17th centuries BCE. Thus, the cultural typological attribution of the grave goods published at the time of the excavation may need revision.

Zevakinsikiy_BA (n=1):

This individual is assigned genetically to the *Central_Steppe_MLBA* analysis label.

- CII-52, stone fencing 90 (I3770): Date of 2132-1940 calBCE (3645±25 BP, PSUAMS-2079). Genetically and morphologically male, 25 to 30 years old. Stone fence number 90 had a diameter of 9 m and was located 20 m from stone fence number 91 excavated in 1971 (*137*). The excavation brought to light, at a depth of 0.4 m, a rectangular pit (2.45 x 1.45 m; depth: 1.75 m) with rounded corners, oriented in a west-to-east direction. A stone cist composed of four stone wall slabs and a fifth cover slab were found at a depth of 1.05 m. The cist measured 1.7 x 0.7 m, and was 0.7 m high. Human bones were found at the bottom of the stone cist including a skull, femurs, ribs, pelvis and a mandible. A hand-made vessel base, manufactured out of blackish-ware tempered with quartz, was found near the northern wall, 30 cm from the eastern wall. This artifact has been attributed to the Early Andronovo Culture.

Zevakinsikiy_MLBA (n=1):

This individual is assigned genetically to the *Steppe_LBA* analysis label.

- CII-47, stone fencing 35, grave 1 (I3763): Date of 1609-1443 calBCE (3240±25 BP, PSUAMS-2502). Genetically and morphologically male, 40 to 45 years old. Stone fence 35 was oval-shaped (1.5 x 1.3 m; depth: 1 m), its main axis oriented north-to-south. The burial contained three skeletons: all were on their right sides in crouched position, with the skull oriented towards the south. The first skeleton (grave 1), located near the western wall, was extremely poorly preserved. Sheep ribs, as well as a bronze knife with a handle decorated with a D-shaped end, were recovered near its skull (Table I, 1 of ref. (*144*)). The bones of the second and third skeletons (graves 2 and 3) were in disarray. The two skulls, cervical vertebrae, and the bones of the right leg which was located near the eastern wall, were all that was found in situ. A second bronze knife characterized by a ring-shaped end, was found near the skull of the third skeleton (Table I, 5 of ref. (*144*)). From a chronological point of view, bronze knives with D- and ring-shaped ends are characteristic of the second half of the 2nd millennium BCE, and specifically to the later centuries of this millennium.

Zevakinsikiy_LBA (n=6)

These six individuals are all assigned genetically to the *Steppe_LBA* analysis label.

• CII No 51, stone fencing 43, skull 1 (I3753): Date of 1111-941 calBCE (2860±20 BP, PSUAMS-2540). Genetically and morphologically male, 40 to 45 years old. Stone fence 43 (*145*) was rectangular in shape (3.5 x 2.5 m) and contained an oval-shaped burial (2.1 x 1.3 m; depth: 0.8 m) which was found at a depth of 0.4 m from the contemporary surface. The burial was sealed by a layer of stones 10 cm thick. The removal of the stones brought to light two skeletons lying on their right side, both in crouched position. The skull of the northern skeleton (number 1, I3753) was oriented towards the south; its arms were bent at the elbow with the hands in front of the face. The legs were bent at the knees at 90° and the femurs were perpendicular to the spinal column. A bone artifact was found near the right hand. The skull of the southern skeleton (number 2) faced southwest and most of its bones were below the northern skeleton.

• CII No 43, stone fencing 1, extension 1 (I3976): Date of 1191-1010 calBCE (2900±20 BP, PSUAMS-2506). Genetically and morphologically male, 50 to 55 years old. Stone fence number 1 (*137*) contained the main burial in the center of the funerary structure and a second burial (I3976) in the southern part of the stone fence, named extension 1 by archaeologist F. H. Arslanova. The stone fence was round in shape with a diameter of 6 m. The main burial (1.9 x 1.5 m; depth: 1.75 m) was found at a depth of 0.6 m and was positioned in a west-to-east direction. Different layers of the fill contained fragments of hand-made blackish pottery, decorated with comb-stamped patterns and horizontal zigzag lines. The main burial contained a skeleton buried on its left side in a flexed position, with the skull oriented toward the east. Round bronze earrings covered with gold foil (diameter: 2.7 and 2.9 cm), as well as two rhomboidal and two oval-shaped pendants were found near the temporal bones of the skull. These artifacts can be attributed to the transition period between the Final Bronze Age and the Early Iron Age. The second burial (1.1 x 0.85; depth: 1.05 m) which was the one from which we obtained ancient DNA (I3976) was entirely covered by stone slabs and was positioned from southwest to northeast. The skeleton was inhumated in a flexed position with the skull oriented toward southwest. No funerary inventory was found in association with this skeleton.

• CII No 44, stone fencing 27 (I3977): Date of 1126-1000 calBCE (2885±20 BP, PSUAMS-2507). Genetically and morphologically male, 40 to 45 years old. Stone fence 27 was oval in shape (4.8 x 2.8 m), its main axis positioned from northeast to southwest (*145*). The burial was also oval in shape (1.25 x 0.7 m; depth: 0.7 m) and contained a skeleton inhumated lying on its right side in contracted position, with the skull oriented towards the south with a 15°-degree inclination towards the west. The face was oriented east. Both arms were bent at the elbow with the hands in front of the face, while the legs were bent at a right angle and the femur bones were perpendicular to the spinal column. The annular finger (digit IV) of the right hand was adorned with a ring made of a bronze wire with a quadrangular section ((*144*): table. III, 11). This artifact is attributed to the Dongal culture of the Final Bronze Age.

• CII-56, stone fencing 34, Skeleton 2 (I3772): Date of 1025-901 calBCE (2810±25 BP, PSUAMS-2080). Genetically and morphologically female, 20 to 25 years old. Stone fence 34 was oval-shaped (2.15 x 1.85 m; depth: 1 m), its main axis positioned southwest-to-northeast. Four skeletons have been found in the burial located at the center of the fence. All were lying on their right side in a flexed position. The skulls of three skeletons were oriented towards the south and only one skull, the fourth (ordered from east-to-west), toward the west. All the skeletons had their hands in front of the face, while the legs were sharply bent at the knees (*145*). Several sheep's ribs were found near the skull of the third skeleton, and some vessel sherds were discovered near the elbow of the fourth skeleton ((*144*): table. III, 8). This ceramic vessel can be assigned to the Dongal culture of the Final Bronze Age.

• KZ-KP-001, CMK, KP27390 (Цмк, кп27390), 9a (I4295): Date of 1211-1056 calBCE (2935±20 BP, PSUAMS-2515). Genetically male.

• KZ-KP-004, CMK, KP27390 (Цмк, кп27390), 6a (I4267): Date of 1193-1013 calBCE (2905±20 BP, PSUAMS-2512). Genetically male.

S2.2.4.6 Tasbas, Byan Zherek (n=3)
Description by Michael Frachetti

The prehistoric settlement site of Tasbas is located in the Byan-Zherek Valley, Semirech'ye (45.13427 N, 79.36794 E, elevation 1492m). The site was first excavated in 2001 by Alexei Mar'yashev (*119*). The site was restudied in 2011 under the directorship of Michael Frachetti and Alexei Mar'yashev, with excavations performed by Paula Doumani Dupuy (*118*). Tasbas is a small pit-house settlement with its first foundation in the Early Bronze Age, dating to roughly 2800-2600 BCE (*118*). A significant finding within the settlement context was a cist burial containing cremated human remains as well as the earliest documented archaeobotanical evidence for domesticated wheat in the Eurasian steppe region. Human remains from the cremation were in poor condition and thus were not sampled for DNA. However, human skeletal elements were recovered in later (upper) stratigraphic layers – unassociated with specific burials. The three individuals analyzed from Tasbas are associated with stratigraphic levels containing late Bronze Age ceramic forms, dated by AMS/14C to the final centuries of the 2nd millennium BCE or the beginning of the 1st millennium BCE which is the beginning of the Iron Age (*146*). At this time, the inhabitants of Tasbas were engaged in farming of wheat, barley, and legumes as well as animal pastoralism (*147*).

We designate two individuals by the *Tasbas_LBA* analysis and split label.
• 14d, ctx 20, level 2, ZA-0238 (I3453): Previously reported date of 1264-1095 BCE (2940±30 BP, OS-97173, layer date). Genetically male. This individual is assigned genetically to the *Tasbas_LBA* analysis label.

• TB ctx 129. Bag ZA-0255 (I3456): Date of 1117-976 calBCE (2870±20 BP, PSUAMS-2538). Genetically male. This individual is assigned genetically to the *Tasbas_LBA* analysis label.

We designate one individual whose radiocarbon date is slightly later to the *Steppe_LBA* analysis label and *Tasbas_IA* split label.

• TB-13/14E22 (below/in)-SF (I3451): Date of 929-807 calBCE (2720±35 BP, OS-97172). Genetically female. This individual is assigned genetically to the *Steppe_LBA* analysis label.

S2.2.4.7 Malaya Krasnoyarka, Ul'ken Narym, Katon Karagai, East Kazakhstan (n=1)

• StPet38, collection 6095, burial place 1, grave 5g, individual 13 (I6709) Date of 1391-1196 calBCE (3025±25 BP, PSUAMS-2940). Genetically female. Skeletally a child, 5 to 6 years old. This individual is genetically to the *Steppe_LBA* analysis label.

S2.2.4.8 Myrzhyk, Kazakhstan (n=1)

• KZ-MRK-001, Myrzhik 6, kurgan 3 (I4785) Context date of 1600-500 BCE. Genetically male. This individual is to the *Myrzhyk_MLBA_LowCov* analysis label (because it is a single representative of a site and has low coveage).

S2.3 Sites from Iran/Turan

S2.3.1 Sites from Mesolithic Iran/Turan

S2.3.1.1 Belt Cave in the Alborz Mountains, Iran (n=1)
Description by Janet Monge

We give this individual in its own *Belt_Cave_M* analysis label. The individual has genetic affinity to early Iranian farmers from the Zagros Mountains as well as to Caucasus HGs (**Supplementary Materials S4)**.

We present genome-wide ancient DNA data from an individual from Belt Cave from the Alborz Mountains in northeastern Iran. Belt Cave was excavated by Carleton Coon in 1949 and 1951 with a large groups of Iranian colleagues in the span of nine months (*148*, *149*). The nearby and larger Hotu Cave was excavated in the 1951 season after limestone quarry activity exposed ancient cave sediments. Both limestone rock shelters/caverns are located side-by-side along the southern rim of the Caspian Sea. The caves were exposed during Pleistocene climatic fluctuations and were alternatively occupied and abandoned, or inundated by rising sea levels, over the next several thousand years. Coon was interested specifically in the documentation of the transition to agriculture and village life in western Asia, that is, the Mesolithic-to-Neolithic transition.

Examination of publications, archival field and laboratory notes at both the University of Pennsylvania Museum, Harvard University and Smithsonian Institution archives by Michael Gregg (*150*) integrates the profiles of the 1949 and 1951 excavations at Belt Cave and the position of the skeletal materials in each trench (A, B and C). Radiocarbon dates were obtain by Libby (*151*) and Ralph (*152*) on a sequence of individuals from the cave.

The minimum number of individuals recorded from the skeletal materials at the University of Pennsylvania Museum from the site is estimated to be at least 7 from the combined Mesolithic and Neolithic levels. New work detailing this site suggest that there are discrepancies between the Coon description of the deposition of the skeletons and the recent reconstructions of their stratigraphic position. Belt 4 is referred to as Skeleton 4 in the Trench C profile.

- Belt4S (I2312_all_d): Context date of 12000-8000 BCE and genetically too low in coverage to determine sex. The individual is skeletally 6 to 8 years old.

S2.3.2 Sites from Neolithic Iran/Turan

S2.3.2.1 Ganj Dareh, Zagros Mountains, Iran (n=9)

Description by Deborah Merrett and Christopher Meiklejohn

We present data from 5 individuals for which genome-wide ancient DNA data has not previously been reported from Ganj Dareh an early Neolithic site in the western Zagros Mountains of present-day Iran. Ancient DNA from 6 other individuals from this site was previously reported in the ancient DNA literature (*9*), and we supplement the data from 4 of these with additional libraries to produce higher quality data (these individuals are not listed below). We give all these samples the *Ganj_Dareh_N* analysis label.

The small mound of Ganj Dareh, at an elevation of 1400 m above sea level, was excavated between 1967 and 1974 by Phillip E. L. Smith. Dates from goat skeletal remains indicate the site was occupied over a period as short as 100 to 200 years (*153*). Radiocarbon dates from human remains of three of the individuals included here (GD22, GD40, GD41) and three others place the occupation of Levels C and D (3rd and 4th from the surface, respectively) at 8150 to 7850 calBCE, confirming the direct archaeological association between the humans and goats (*154*). All individuals for the present study are from Level D. GD20, GD22, GD25, and GD40 were within the same structure; all but GD40 were below-house-floor burials. GD41, also a below-house-floor burial, was recovered from a burnt alcove structure several rooms to the east of the others (*155*).

- GD20 (I1946): Context date of 8250-7850 BCE. Genetically male. This individual from level D is largely complete but fragmentary and exhibits intentional cranial deformation. The individual is skeletally determined to be an adult male (*156*), and is the father of both GD22/I1947 and GD40/I1952.

- GD22 (I1947): Previously reported date of 8210-7845 calBCE (8860±30 BP, Beta-432800) (*154*). Genetically male. This individual from level D is represented by a mostly complete but fragmentary infracranial skeleton and incomplete cranium, morphologically consistent with an older adult male. Antemortem tooth loss and well developed entheses are both consistent with the older adult age estimation. He is son of GD20/I1946 and brother of GD40/I1952.

- GD40 (I1952): Previously reported date of 8219-7761 calBCE (8850±50 BP, Poz-81114). Genetically male. This individual from level D, represented by a fragmentary skull with almost complete facial skeleton, and largely complete infracranial skeleton is morphologically consistent with a young male. He is the son of GD20/I1946 and brother of GD40/I1947.

• GD25 (I7527): Context date of 8200-7700 BCE. Genetically female. An infact, ca. one year in age based on tooth formation, a largely complete skull and fragmentary infracranial skeleton. Ectocranial areas of porosity adjacent to the lambdoidal suture are present.

• GD41 (I1954): Previously reported date of 8294-7992 calBCE (9000±40 BP, Beta-436170) (*154*). Genetically male. This 16 to 19-year-old individual from level D represented by a reasonably complete but fragmentary skull and very incomplete infracranial skeleton, morphologically consistent with being an adolescent male. Remains are partly calcined and were recovered from an alcove adjacent to room with sheep skulls embedded into an alcove (*157*).

Relatedness summary:

- 3-person family: At Ganj Dareh, genetic analysis detects three males from a nuclear family: I1946 (the father) and his two sons (I1947 and I1952), genetically confirmed to be full siblings of each other.

S2.3.3 Sites from Copper Age/Eneolithic Iran/Turan

S2.3.3.1 <u>Hajji Firuz, Zagros Mountains, Iran (n=7)</u>
Description by Janet Monge

We present genome-wide ancient DNA data from seven individuals from the site of Hajji Firuz, from which ancient DNA has not been previously reported. These individuals all have strong genetic affinity to early Iranian farmers from the Zagros Mountains (**Supplementary Materials S4**).

Hajji Firuz Tepe is located in the Lake Urmia Basin in the eastern Solduz Valley in the western Azerbaijan province of northwestern Iran. After an early surface site survey performed in 1936, test excavations were conducted in 1958 and 1961, which led to more extensive exploration of the mound in 1968. The primary purpose of the 1968 excavation was to enrich archaeologically understanding of the origins of agriculture in the Neolithic within this region and to determine the relationship of this area to the sites and peoples of eastern Anatolia, northern Mesopotamia and the central Zagros region. Hajji Firuz was chosen for closer investigation because it was confirmed to contain well stratified sediments that dated to the last part of the sixth millennium BCE (*158*). The excavation levels are placed in time-ordered Phases from “A” to “L” (latest to earliest phases of site occupation). The Hajji Firuz Tepe skeletal materials used in the analysis were excavated in 1968 by Mary Voigt and Robert Dyson (University of Pennsylvania Museum). Burials at the site are all associated with architectural features and each burial unit contains either multiple individuals in a pit, bin or jar ossuary, or single floor burials. Individuals within the ossuary context are commingled but, in several cases, the skeletons were in partial positional articulation. Bones from Units K10 (I2327), F10 (I2323, I2328, I4241) and F11 (I4243, I14349, I4351) were sampled. Preliminary study of the skeletal material indicates that approximately 28 to 40 individuals in total were excavated during the 1968 field season. The six individuals in the main genetic cluster are as follows:

Hajji_Firuz_C (n=5):

- FH5, F10 B1 prenatal (I2323): Date of 6060-5851 calBCE (7090±50 BP, Poz-81115). Genetically female. Second-to-third-degree relative of I4249 and I4351. This is a bin ossuary burial in Phase A3, which is skeletally prenatal or neonatal.

- FH9, F10 B1 skull 7 (I2328): Date of 6013-5898 calBCE (7075±30 BP, PSUAMS-2345). Genetically male. This is a bin ossuary burial in Phase A3, which is skeletally an adult male.

- F10, B1 S3 (I4241): Date of 6016-5899 calBCE (7080±30 BP, PSUAMS-2163). Genetically male. This is a bin ossuary burial in Phase A3, which is skeletally a young adult male.

- F11, 4 + F11, B3 3 (I4349): Date of 5887-5724 calBCE (6915±40 BP, PSUAMS-2126). Genetically male. Second-to-third-degree relative of I2323 and I4351.

- F11, B3 1 (I4351): Date of 6056-5894 calBCE (7100±45 BP, PSUAMS-2151). Genetically female. Second-to-third-degree relative of I2323 and I4249. This is a bin ossuary burial in Phase A3, which was skeletally determined to be a 35 to 45-year-old male (the genetic determination that this individual is a female is not surprising given that older females often are skeletally assigned as males).

Relatedness summary:

- 3-person family: I2323, I4249, and I4351 are all genetically detected to be second-to-third degree relatives of each other.

We also identified two individuals that we assigned different analysis labels because radiocarbon dating placed them in later periods.

Hajji_Firuz_BA (n=1):

- F11, 3 (I4243): Date of 2465-2286 calBCE (3875±25 BP, PSUAMS-2113). Genetically female. This individual is a genetic outlier and is also an intrusive burial from the Bronze Age based on its radiocarbon dating. This individual has additional Steppe pastoralist related admixture when compared with the other individuals from the Copper Age.

Hajji_Firuz_IA (n=1):

- FH8, K10 B1 (I2327): Date of 1193-1019 calBCE (2910±20 BP, PSUAMS-4413). Genetically male. The sample is taken from a bin ossuary burial in Phase F-G, which is skeletally a 25 to 35-year-old male. However, their radiocarbon date places them in a much later time period so we do not co-analyze this individual with the others and give them a distinctive analysis label. This individual has additional Steppe pastoralist related admixture when compared with the other individuals from the Copper Age.

Relatedness summary:

- 3-person family: At Hajji Firuz we have genetic data from three individuals who we detect as related within the 2^{nd} to 3^{rd} degree in all combinations: I2323, I4349 and I4351.

S2.3.3.2 Aigyrzhal, Kyrgyzstan (n=2)

- Aigyrzhal K.67 (I11526): Date of 2203-2041 calBCE (3735±20 BP, PSUAMS-4607). Genetically male.

• Aigyrzhal K.105 (I11527): Date of 2114-1928 calBCE (3630±20 BP, PSUAMS-4750). Genetically male.

S2.3.3.3 Sarazm, Tajikistan (n=2)
Description by Michael Frachetti

Sarazm is an agricultural village site located along the lower Zerafshan River valley in Tajikistan and is the northernmost site in central Asia with direct evidence for both farming and herding in the 4th millennium BCE. The site's chronology spans from 3500-1800 BCE. The early phases of the site's habitation, around 3500 BCE, correlate with an expansion to the north of "Geoksyur type" material forms and can be broadly related to the Copper Age expansion of Kopet Dag and Geoksyur farming villages roughly dated to the late Namazga II period (Yalangach Phase in Geoksyur oasis). This is attested in a diversity of material forms as well as directly derived chronologies from *in situ* botanical remains (*159*).

A few burials were excavated at Sarazm. Third-millennium-BCE pastoralism in the region can be inferred only from a few "steppe-type" burials, one recovered directly at Sarazm (*160*) and another recently documented at the site of Jukov (*161*). Material from both sites reflects atypical ceramics categorized by the analysts as broadly "Afanasievo type" with local admixtures and dating to the mid-fourth millennium BCE.

• UZ-SZ-001, Sarazm 85 (1), burial 44-38 (I4290): Context date of 3700-3300 BCE. Genetically female.

• UZ-SZ-002 Sarazm 85 (2), burial 44-37 (I4910): Date of 3636-3521 calBCE (4765±20 BP, PSUAMS-2624). Genetically female.

S2.3.3.4 Geoksyur, Turkmenistan (n=25)
Description by Sergey Vasilyev

Geoksyur is the largest agricultural settlement of the Geoksyur Oasis located in southeastern Turkmenia (Turkmenistan) in the delta of the river Tedzhen in the Eneolithic period (Geoksyur I, second half of the 4th millennium BCE). The settlement represents the culture of the East Anau group of tribes in theLate Eneolithic, which is the final documented stage of the Anau culture. The area is approximately 12 hectares and the population is estimated have been 2000 people. Geoksyur was likely to have been the political center of the oasis. Geoksyur was found by A.A. Marushchenko in 1939, and was excavated by V.I. Sarianidi in 1955-57, 1960, and 1962-65. In 1962 V.M. Masson distinguished the stages of the development of the early agricultural culture of southeastern Turkmenia (Turkmenistan) based on the stratigraphy of Geoksyur and the other settlements of the oasis.

Geoksyur is constructed of mudbrick architecture arranged in housing blocks, often with courtyards formed by adjoining two-room rectangular houses made of mud brick, complete with ritual spaces and hearth altars. There is a burial ground with domed tombs – burial vaults for the larger families – on the settlement edge. Findings within the settlement include Geoksyur-type ceramics decorated with two-color paintings (red paint inside, black and red paint outside), crosses, zigzags, rhombs, and geometrized animal figures. Female terracotta figurines were also found. Among the numerous female statuettes there were a few male figurines wearing battle helmets. It has been hypothesized that the female statuettes were the images of the mother-goddess while the masculine ones represented chieftains or leaders.

Geoksyur was surrounded by a network of canals (length 3 km, width 2.5-5 m), which irrigated 50 to 70 hectares near the settlement. The inhabitants of Geoksyur began to abandon the settlement at the beginning of the 3rd millennium BCE probably for environmental reasons. The ethno-cultural ties of the Geoksyur tribes spread to parts of present-day Tajikistan and Afghanistan. It is hypothesized that the ancestors of the Geoksyur population arrived there due to dispersal from southeastern Iran. The settlement of Sarazm in Tajikistan, also represented by individuals for whom data is generated in the study, was culturally similar and documents the widespread dispersal of the Geoksyur-type tribes.

Geoksyur_EN (n=24).

The following individuals are assigned a *Geoksyur_EN* analysis label:

- MOS324, Geoksyur 1, 1964, 172, вяпе, IE-10-23 (I8527): Context date of 3500-2800 BCE. Genetically male.

- MOS323, Geoksyur 1, 1964, 188, Tolos "Щ", IE-10-15 (I8526): Context date of 3500-2800 BCE. Genetically male.

- MOS326, Geoksyur 1, 1965, N206, IE-10-8 (I8529): Context date of 3500-2800 BCE. Genetically male.

- MOS330, Geoksyur 1, 1963, Skull N107, Tolos "Ч", IE-10-4 (I8532): Context date of 3500-2800 BCE. Genetically female. This individual is genetically determined to be part of a 7-person family, and is directly detected as being a 2nd or 3rd degree relative of I12487.

- MOS317, Geoksyur 1, 1964, 200 бяия, IE-10-20 (I8510): Date of 3341-3094 calBCE (4490±25 BP, PSUAMS-3987). Genetically female.

- MOS295, Geoksyur 1, 1963, N 139, Tolos "Ц", IE-10-35 (I8504): Context date of 3500-2800 BCE. Genetically male.

- MOS319, Geoksyur 1, 1963, Skull N124, IE-10-14 (I8524): Context date of 3500-2800 BCE. Genetically male. This individual is genetically determined to be part of a 7-person family, and is directly detected as being a 1st or 2nd to 3rd degree relative of I8502.

- MOS291, Geoksyur 1, 1963, Tolos "Ц", N144, IE-10-27 (I8502): Date of 3309-2923 calBCE (4420±25 BP, PSUAMS-4038). Genetically female. This individual is genetically determined to be part of a 7-person family, and is directly detected as being a first-degree relative of I8505, a 1st or 2nd degree relative of I8524, and a 2nd or 3rd degree relative of I12483.

- MOS325, Geoksyur 1, 1963, N43, Tolos "К", IE-10-18 (I8528): Context date of 3500-2800 BCE. Genetically female.

- MOS298, Geoksyur 1, 1963, N129, Tolos "Ф", IE-10-46 (I8505): Context date of 3500-2800 BCE. Genetically undetermined sex. This individual is genetically determined to be part of a 7-person family, and is directly detected as being a first-degree relative of I8502.

• MOS327, Geoksyur, 1963, grave 92 ч, Tolos "P", GS-1, вяие, IE-10-1 (I8530): Context date of 3500-2800 BCE. Genetically female.

• MOS332, Geoksyur 1, 1963, 64, Tolos "O", IE-10-5 (I8534): Date of 3010-2881 calBCE (4300±25 BP, PSUAMS-3988). Genetically female. This individual is genetically determined to be the mother or daughter of I12486.

• MOS293, Geoksyur 1, 163, IE-10-29 (I8503): Context date of 3500-2800 BCE. Genetically female. This individual is genetically determined to be part of a 4-person family, and is directly detected as being a first-degree relative of I8531, and a 2nd to 3rd degree relative of I12481.

• MOS328, Geoksyur 1, 1964, 164, IE-10-2 (I8531): Context date of 3500-2800 BCE. Genetically female. This individual is genetically determined to be a first-degree relative of I8503. This individual is genetically determined to be part of a 4-person family, and is directly detected as being a first-degree relative of I8503.

• MOS321, IE-10-22 (I12478): Context date of 3500-2800 BCE. Genetically female.

• MOS322, IE-10-16 (I12479): Context date of 3500-2800 BCE. Genetically female. This individual is genetically determined to be part of a 7-person family, and is directly detected as being a 2nd or 3rd degree relative of I12483 and I12487.

• MOS318, IE-10-13 (I12480): Context date of 3500-2800 BCE. Genetically female.

• MOS312, IE-10-52 (I12481): Context date of 3500-2800 BCE. Genetically male. This individual is genetically determined to be part of a 4-person family, and is directly detected as being a 2nd to 3rd degree relative of both I8503 and I12481.

• MOS297, IE-10-48 (I12482): Context date of 3500-2800 BCE. Genetically male.

• MOS313, IE-10-51 (I12483): Context date of 3500-2800 BCE. Genetically female. This individual is genetically determined to be part of a 7-person family, and is directly detected as being a 2nd or 3rd degree relative of I8502, I12479, and I12487.

• MOS296, IE-10-47 (I12484): Context date of 3500-2800 BCE. Genetically female.

• MOS333, IE-10-6 (I12485): Context date of 3500-2800 BCE. Genetically female.

• MOS290, IE-10-33 (I12486): Context date of 3500-2800 BCE. Genetically female. This individual is genetically determined to be the mother or daughter of I8534.

• MOS329, IE-10-9 (I12487): Context date of 3500-2800 BCE. Genetically male. This individual is genetically determined to be part of a 7-person family, and is directly detected as being a 2nd or 3rd degree relative of I12483, I12479, and I8532.

Geoksyur_EN_o (n=1)

The following individual is assigned to the *Geoksyur_EN_o* analysis label and *BMAC_o2* split label. On the basis of a large number of individuals that we have from Copper Age and Bronze Age Turan, we reassign this individual to the Bronze Age.

• MOS320 MOS320, Geoksyur 1, 1963, N 130, Tolos "S", IE-10-21 (I8525): Context date of 3500-2800 BCE. Genetically undetermined sex. This is an outlier from the main *Geoksyur_EN* genetic cluster and hence is analyzed separately.

Relatedness summary:

• 2-person family: I8534-I12486 have a mother-daughter relationship (the order is unknown)
• 4-person family: I8503-I8351 are a pair of first-degree relatives. In addition, I8503 is a 2nd to 3rd degree relative of I12481, who in turn is a 2nd to 3rd degree relative of I8510.
• 7-person family: I8502 is directly detected as having a 1st degree relationship to I8505, a 1st or 2nd degree relative to I8524, and a 2nd to 3rd degree relative to I12483. The latter individual I12483 is part of a trio of individuals who are directly detected as having 2nd to 3rd degree relatives to each other (I12483-I12479-I12487). Finally, the last individual in this trio I12487 has a 2nd to 3rd degree relationship to a seventh individual I8532.

S2.3.3.5 Tepe Anau, Turkmenistan (n=3)

These 3 individuals are assigned to the *Parkhai_Anau_EN* analysis label, a term that combines the genetic data from Eneolithic individuals from Tepe Anau and Parkhai as they were homogenous.

• ANAU1 (I4085): Context date of 4000-3000 BCE. Genetically male.

• ANAU2 (I4086): Context date of 4000-3000 BCE. Genetically female.

• ANAU3 (I4087): Context date of 4000-3000 BCE. Genetically male.

S2.3.3.6 Parkhai I and II, Sumbar Valley, Turkmenistan (n=9)
Overview description by Andrey Gromov and Vyacheslav Moiseyev

We present genome-wide data from nine individuals from the sites of Parkhai I and II. The burial mound of Parkhai was discovered in 1977 on a loess hill located on the edge of the village Kara-Kala (modern Mahtumquli) in the Sumbar Valley in southwestern Turkmenistan. Parkhai II is located 800 meters from Parkhai I and defines the earliest known burial mound within southwestern Turkmenistan, dating back to the Eneolithic-Bronze Age, that is, 4000-3000 BCE. Parkhai II is located 1 km to the west of the present-day village and covers an area of 1.5 to 2 hectares and contains more than 290 burial chambers.

Archaeological excavations at Parkhai II were led by I. N. Khlopin, and revealed funerary rites and burial items that were distinctive compared to other burials in the Sumbar Valley in the Eneolithic period. In particular, numerous burials were collective graves (multi-individual), which is atypical for burials in southeastern Turkmenistan and northeastern Iran at the time. Notably, the grave construction at Parkhai II included underground tombs with definable entrances that were separate from the main burial chamber. Parkhai II ceramics are dominated by the presence of Gray Ware (vases, bowls, etc.) as well as fragments of painted vessels. After 14 years of fieldwork, more than 290 burials have been excavated at Parkhai II.

Parkhai_Anau_EN (n=7)

Most Parkhai II individuals (n=6) were genetically assigned to the *Parkhai_Anau_EN* analysis label.

- UZ-PK-003, Parkhai 117/2, 56-74 (I4259): Date of 3307-2928 calBCE (4425±20 BP, PSUAMS-2799). Genetically female. Archaeological period SWT VII (Early Eneolithic). The individuals in chamber 117 were buried in two different layers. The upper level contained three bodies buried in flexed positions, including female I4259, as well as Gray Ware bowls, copper pins and a ring, and beads. The lower level was a collective grave with a larger group of skeletal remains piled in the eastern portion of the chamber, likely the result of clearance for subsequent burials. A ceramic bowl and a copper pin were found in level 1. This individual is genetically detected to be a 2nd to 3rd degree relative of both I4634 and I4635.

- StPet13, site II, grave 168, individual 9 (I4634): Date of 3264-2924 calBCE (4415±25 BP, PSUAMS-3757). Genetically female. Archaeological period SWT VI (Middle Eneolithic). Morphological age and sex could not be determined. Grave 168 is a collective burial made in a pit with unclear borders, which probably was used as a waste repository during the cleaning of nearby graves. The pit contained disarticulated skeletons of ten individuals including three males, four females, and three children. Five pottery vessels were found among the skeletal remains. This individual is genetically detected to be a 2nd to 3rd degree relative of both I4259.

- StPet14, site II, grave 167, individual 4 (I4635): Context date of 3700-3300 BCE. Genetically female. Archaeological period SWT-VI (Middle Eneolithic Age). Morphologically female, 25 to 35 years old. Grave 167 is a collective burial with four skeletons (two females and two children) found in anatomical order in a round grave pit. The similarly positioned skeletons, which lie in parallel to each other with their heads oriented north-to-west, suggests simultaneous burial of all four individuals. One 7-year-old child (individual 1) and one older female (>60 years old, individual 2) lay on their left side, while the skeleton of a child of 8 to 9 years old (individual 3) and a young female 25 to 30 years old (individual 4, I4635) lay on their right side. The legs of all individuals were outstretched or slightly bent. While collective burials are common for early periods (SWT IV-III, for example see description of grave 235 below), in the Sumbar area simultaneous burials are not typical. Individual 1 was accompanied by a pottery altar and amulet made from a vessel wall with a handle. Individual 2 was associated with two bronze pins, one with a double-spiral top, another with a plain top, and a massive bronze wristband. Five small bronze nails were also found lying between the maxilla and mandible of individual 2, and were likely used for piercing the tongue. No artifacts were reported for individual 3. Individual 4 from whom we obtained ancient DNA (I4635) had a bronze ring and wristbands made of cornelian and steatite beads on both arms. This individual is genetically detected to be a 2nd to 3rd degree relative of both I4259.

- StPet26, site II, grave 178, individual 4 (I6670): Context date of 2700-2300 BCE. Genetically male. Archaeological period SWT-IV (Early Bronze Age). The grave consists of two levels of buried individuals.

- StPet27, site II, grave 235, individual 13, Museum ID 7651-108 (I6671): Context date of 3000-2200 BCE. Genetically female. Archaeological period SWT-IV (Early Bronze Age). Morphologically female, 30 to 35 years old. Grave 235 is a collective burial with remains of no less then 20 individuals of different sexes and ages placed in an oval pit. Based on the positions of the skeletal remains, it can be inferred that the burials were made at different times. During these burials, all older remains were moved to the edge of the grave pit to allow the newest body to be placed on a clean space. The individual buried last in this grave was a 20 to 25-year-old female. Such types of burials are characteristic of early periods in the Sumbar area (SWT VII-III). The archaeological finds consisted of only 11 pottery vessels. Individual 13 (I6671) was found among the commingled skeletal remains.

- StPet25, site II, grave 256, individual 6 (I6669): Date of 3082-2909 calBCE (4365±25 BP, PSUAMS-2950). Genetically female. Morphologically female, 45 to 55 years old. Archaeological period SWT-V (Late Copper Age). Similar to grave 235, grave 256 is a collective burial typical of earlier periods and contains the remains of seven individuals. The oval grave pit (2 x 1.75 m) is elongated from north-to-south. The entrance is located on the west side of the grave and was marked with stones. Skeletons of two individuals buried later than the others were found in anatomical order. The skeleton of a 60-year-old female was placed along the southern wall of the grave on her left side with her head turned to the west and face to the north. A second skeleton of a 35 to 45-year-old female lay on her right side in a crouched position. Remains of five other individuals (one male 30 to 35 years old, two females >50 years old, one 6-year-old child, and the remains of an adult individual of unknown sex) were redeposited in a pile. The genetically analyzed sample (ID I6669) was found among the redeposited skeleton remains. Archaeological artifacts found in the grave included four pottery vessels and a bronze nail. A bone pin with a spherical top was also found near the back of the head of the first skeleton.

• StPet33, site II, grave 267, Museum ID 7651-138 (I6674): Date of 2338-2039 calBCE (3775±40 BP, PSUAMS-2951). Genetically and morphologically female, 45 to 55 years old. Archaeological period SWT-II (Late Bronze Age). Grave 267 is an oval catacomb (1.5 x 1.0 m) elongated from north to south with some traces of a dromos. The entrance located in the east wall was closed off by large stones and bricks. The skeleton of a female was placed in crouched position on its right side with face positioned in the opposite direction of the entrance. Two pottery vessels were found near the lower part of the skeleton.

BMAC (n=1)

One Parkhai II individual was genetically assigned to the *BMAC* analysis label based on its time period and the fact that the ancestry of this individual resembled other samples in BA Turan.

• StPet24, site II, grave 265, Museum ID 7651-136 (I6668): Context date of 1600-1000 BCE. Genetically female. Morphologically female, 35 to 40 years old. Archaeological period SWT-II (Late Bronze Age). Grave 265 is a catacomb with an oval grave pit (1.6 x 1.3 m) elongated from northeast to southwest with remains of the lower part of a dromos. Although dromoi were common in the Late Bronze Age of Parkhai II, later soil disturbances made their detection impossible in some cases. Grave 265 is filled with sand and clay flows and pieces of the grave ceiling, which probably collapsed soon after the funerals. A crouched skeleton of a 35 to 40-year-old female lay on its left side with face oriented toward the grave entrance. The archaeological context of the burial includes four pottery vessels and a spinning spindle made of marble.

Parkhai_LBA_o: Parkhai I, Turkmenistan (n=1)

This individual has the youngest radiocarbon date of all individuals we analyzed from the site and is also a genetic outlier and hence we designate it by the analysis label *Parkhai_LBA_o*.

• StPet23, site I, grave 28, Museum ID 7651-1 (I6667): Date of 1497-1413 calBCE (3170±20 BP, PSUAMS-2998). Genetically female. Archaeological period SWT-II (Late Bronze Age). Morphologically female, 30 to 40 years old. Grave 28 consisted of a single burial in an oval pit (2.1 × 1.1 m) with an entrance in the west-to-south wall. Single burials are very characteristic for the Late Bronze Age (SWT-II and I) of Parkhai I and II, in comparison to earlier periods when collective burials are typical. A skeleton of a female lay crouched on its left side with face turned toward the entrance. The central part of the skeleton was damaged by rodents. Archaeological finds include seven pottery vessels, two bronze pins with ornamental tops that were inserted into the ground near the skeleton's feet, three bronze needles inserted into the ground near the head, and one bronze ring and earring found under the skull. Three beads made of different stones, namely lapis lazuli, carnelian, and limestone, were also found in the burial.

Relatedness summary:
• 3-person family: I4259 is a 2nd to 3rd degree relative of both I4634 and I4635.

S2.3.3.7 Tepe Hissar, Damghan Plain, Iran (n=13)
Description by Janet Monge

All these individuals are assigned to the *Tepe_Hissar_C* analysis label.

We present genome-wide ancient DNA data from 13 individuals from the site of Tepe Hissar for which genome-wide ancient DNA data have not been previously reported. Tepe Hissar is a Copper Age-to-Bronze Age urban settlement, located in the fertile central Iranian Plateau, about 1.5 km from the modern city of Damghan, along the "Silk Road" bridging Central Asia and Mesopotamia. The site, a series of mounds and smaller hillocks, was excavated in two 5-month field seasons (1931 and 1932) by Erich Schmidt. Over 1,637 burials were exposed and excavated, approximately 400 of which are today curated at the University of Pennsylvania Museum. The hilltops and adjacent valleys served as both an inhabited village and a necropolis albeit at different time intervals in close to 1,000 years of prehistory. Schmidt (1937) divided the site into a series of occupation periods (I, II, III with subdivisions), a sequencing that was later revised by R.H. Dyson and M. Tosi after completing a 1976 field season to clarify the site stratigraphy (University of Pennsylvania Museum, Iran Center for Archaeological Research, and the University of Turin) (*162*). A ceramic chronology of the site was presented by Gursan-Salzmann (2016) (*163*). Survey and excavation at Tepe Hissar continue under the supervision of ICAR. Both material culture (primarily ceramics) and skeletal biology of the inhabitants of the site indicate heterogeneity and have been interpreted as suggesting possible migration and replacement between the periods of Tepe Hissar. Thus, the samples for genetic analysis were chosen to be representative of samples from all the chronological periods present at the site.

The data from the 13 individuals we analyzed were remarkably homogeneous from a genetic point of view, as we found that the F_{ST} measuring allele frequency differentiation between any group of individuals from the three time periods was not significantly different from zero. Specifically, when we computed *D*-statistics of the form *D(Mbuti.DG, X; Tepe_Hissar_Individual_J, Tepe_Hissar_Individual_K),* varying *X* over all other groups (and *J* and *K* over all possible pairs of Tepe Hissar individuals), we found no significant skews from zero after correcting for multiple hypothesis testing. We conclude that the Tepe Hissar individuals represent a homogenous group with little evidence of migration into it from genetically very different groups over the extended period from which the individuals derive. In addition, we are not able to detect significant genetic drift over time among the individuals from the site suggesting that the population size was substantial over this period. We wanted to increase our sample size (improving sensitivity) and we thus combined all 13 individuals together into a single grouping, *Tepe_Hissar_C*.

- TH16-4, 33-16-4 (CG95 X 7) (I2335): Date of 3639-3385 calBCE (4760±25 BP, PSUAMS-2346). Genetically male. This is a burial from Period 1, which is skeletally an adult male.

- TH23-9, 33-23-9 (DG96 X 49) (I2337): Date of 3641-3519 calBCE (4780±30 BP, PSUAMS-1919). Genetically male. This is a burial from Period IB, which is skeletally a 30 to 40-year-old female.

- TH16-118, 33-16-118 (DF29 X 2) (I2512): Date of 2916-2876 calBCE (4265±25 BP, PSUAMS-1914). Genetically male. This is a burial from Period II/III, which is skeletally an approximately 25 to 35-year-old male.

- TH23-13, 33-23-113 (DH06 X1) (I2513): Date of 2849-2492 calBCE (4070±25 BP, PSUAMS-2229). Genetically female. This is a burial from Period IIIb, which is skeletally a female older than 50 years.

- TH23-73, 33-23-73 (CS31 X 4) (I2514): Date of 2474-2307 calBCE (3915±25 BP, PSUAMS-1918). Genetically male. This is a burial from Period IIIb, which is skeletally a female older than 50 years. The genetic determination of sex disagrees, which plausibly reflects the fact that older females and males can be difficult to distinguish.

- TH16-2, 33-16-2 (CG95 X 4) (I2918): Date of 3702-3536 calBCE (4850±30 BP, PSUAMS-2228). Genetically female. This is a burial from Period I, which is skeletally a female older than 50 years.

- TH16-11, 33-16-11 (DG36 X 16) (I2921): Date of 3656-3526 calBCE (4820±30 BP, PSUAMS-1912). Genetically female. This is a burial from Period I, which is skeletally an adult male which is inconsistent with the genetic determination.

- TH16-12, 33-16-12 (DG36 X 18) (I2922): Date of 2197-2027 calBCE (3705±25 BP, PSUAMS-2227). Genetically female. This is a burial from Period I, which is skeletally a 25 to 35-year-old female.

- TH16-51, 33-16-51 (CH95 X 1) (I2923): Date of 2878-2636 calBCE (4160±25 BP, PSUAMS-1915). Genetically male. This is a burial from Period III, which is skeletally a male older than 55 years.

- TH16-56, 33-16-56 (DF07 X 6) (I2924): Date of 2881-2666 calBCE (4170±25 BP, PSUAMS-2262). Genetically female. This is a burial from Period IIIA, which is skeletally determined to be a 12 to 15-year-old of unclear sex.

- TH16-110, 33-16-110 (DF19 X 55) (I2925): Date of 2881-2666 calBCE (4170±25 BP, PSUAMS-1913). Genetically female. This individual is genetically a 2nd to 3rd degree relative of I2927. This is a burial from Period II/III, which is morphologically determined to be that of a 25 to 35-year-old female.

• TH23-124, 33-23-124 (DF16 X 2) (I2927): Date of 2575-2350 calBCE (3970±30 BP, PSUAMS-1916). Genetically male. This individual is genetically a 2nd to 3rd degree relative of I2925. This is a burial from Period IIIC, which is morphologically determined to be that of a 30 to 40-year-old male.

• TH23-205, 33-23-205 (DG11 X 35) (I2928): Date of 2858-2505 calBCE (4095±25 BP, PSUAMS-1917). Genetically female. This is a burial from Period III, which is morphologically that of a 40 to 50-year-old male, inconsistent with the genetic determination.

Relatedness summary:

- At Tepe Hissar we detected a pair of 2nd to 3rd degree relatives I2925 and I2927 although we cannot discern the exact nature of the relationship.

S2.3.4 Sites from Turan in the BA

S2.3.4.1 Darra-i-Kur Cave, Afghanistan (n=1)
Description by Katerina Douka

This individual is placed in the *Darra_i_kur_MBA* analysis label.

The site of Darra-i-Kur was excavated in 1966 under the umbrella of an American-Afghan research collaboration focusing on the prehistory of Afghanistan and led by L. Dupree. The excavations revealed a sequence of Middle Paleolithic (Mousterian), Neolithic and Iron Age implements, as well as Late Islamic and modern 20th century material. A fragmentary, human right temporal bone consisting of the tympanic and petrous portions, with the mastoid process and paramastoid crest broken at the base, was discovered at the site. Originally thought to be associated with the Mousterian level, following recent direct radiocarbon dating and mitochondrial DNA analyses the bone is now believed to belong to a Neolithic male individual (*80*). Further Neolithic human remains consisting of skull fragments and several long bones of one or two children were discovered in the adjacent upper level of the site as well. Here we report whole genome data from the same ancient DNA library.

• Darra.I.Kur (Darra.I.Kur_d): Previously reported date of 2850-2460 calBCE (3989±31 BP, OxA-31781). Genetically male.

S2.3.4.2 Farkhor, Tajikistan (n=2)

Description by Nadezhda Dubova and James Mallory

The necropolis of Farkhor is located in the Parkhar district of the Khatlon region (South Tajikistan). Excavations were carried out by N. M. Vinogradova (Russia), S. Bobomulloyev (Tajikistan) and G. Lombardo (Italy). The length of the cemetery in the north-south direction is 550-600 m, and in the west-east direction is 300-350 m. Several excavation areas were dug into the slopes of the loess hills, in places free from modern burials. More than 50 burials were excavated. The buried person typically lay on the right or left side, heading facting to the north-northwest. The hands are typically in front of the face, legs bent at the knees. Stone items (marble weights, rods, onyx vessels), clay "house models," long bronze pins with rounded tops, silver bracelets and ceramic vessels were placed in the grave as burial offerings. According to the authors of the excavation, the findings can be dated to the Middle Bronze Age, but it is possible that they also belong to earlier period (judging by such objects as the stone "lamps", decorated with a circle ornament, as well as the stone rods) (*164–167*).

These two individuals are assigned to the *Farkhor_BA* analysis label and split label. Genetically, their ancestry is between the *Western_Steppe_MLBA/Central_Steppe_MLBA* genetic clusters and the *BMAC* main cluster.

• UfaNov17-027, Farkhor 2014 Area 4, tomb 11, Skeleton 1? (I11044): Context date of 3000-2000 BCE. Genetically undetermined sex.

• UfaNov17-024, Farkhor 2017 Area 6A, tomb 43 (I10551): Context date of 3000so -2000 BCE. Genetically male.

S2.3.4.3 Zaman Baba, Uzbekistan (n=1)

This individual is assigned to its own analysis label (*Zaman_Baba_BA*). However, the individual had data from only 16,000 SNPs, barely above our analysis threshold of 16,000

SNPs. At this level it was challenging to analyze without a more detailed analysis focusing specifically on this individual and we did not analyze its genetic ancestry further.

• MOS311, Zaman Baba, 1951, Grave N23, IE-56-1 (I8508): Context date of 2500-1500 BCE. Genetically undetermined sex.

S2.3.4.4 Shahr-i-Sokhta, Seistan, Iran (n=22)
Description by Massimo Vidale

The site of Shahr-i Sokhta (Seistan, Iran), literally "Burnt City" developed near the endorheic delta of the Helmand river and its terminal lakes. It was excavated by the IsMEO Italian Archaeological Mission from 1967 to 1977 under the direction of M. Tosi, and in the final two years by S. Salvatori and Marcello Piperno (*168*, *169*). After a break of 18 years, the excavations were resumed by S. M. Sajjadi for the Iranian Centre for Archaeological Research and are still going on today. According to radiocarbon dating, its sequence stretches from the last centuries of the 4th to the early 2nd millennium BCE. The graveyard seems to contain approximately 30,000 burials, more than 1,000 of which, so far, have been excavated. Extraordinary conservation of organic materials, as well as early evidence of long-distance contacts with southern Central Asia, Afghanistan (especially trade in lapis lazuli) and Baluchistan in South Asia are its main archaeological features.

Genetically we observe two distinct clusters of individuals, *Shahr_I_Sokhta_BA1* and *Shahr_I_Sokhta_BA2*, which we discuss further in **Supplementary Materials S4** and **S5**. The first of the two groupings reflect individuals which appear to be relatively similar to other individuals from Copper Age and Eneolithic Turan, from Geoksyur, Parkhai and Tepe Anau. The second group of individuals have significantly higher proportions of *AHG* related admixture and together with similarly outlying individuals from Gonur lie on what we call the *Indus Periphery Cline.* From the archeological contexts, we observe a chronological trend that distinguishes *Shahr_I_Sokhta_BA1* and *Shahr_I_Sokhta_BA2*. *Shahr_I_Sokhta_BA1* includes 7 individuals who died between approximately 3100 and 2800 BCE, and 4 between approximately 2700 and 2600 BCE. However no individuals so far have been dated later to later than this time from Shahr-i-Sokhta. This chronology for the *Shahr_I_Sokhta_BA1* individuals who bear close genetic affinities to Copper Age and Eneolithic individuals from Turan is consistent with the archaeological evidence of similarities between the ceramic assemblages of Shahr-i Sokhta Period I and Central Asia (*170–172*). However current analytical studies clearly indicate that even vessels that closely resemble those from south Turkmenistan wares were made locally in Seistan.

Shahr_I_Sokhta_BA2, which is the source of the majority of individuals in the *Indus Periphery Cline,* includes two graves that do not have archaeological context as they are completely void of furnishings (31 and 405), four Period I graves whose chronology is indistinguishable from the *Shahr_I_Sokhta_BA1* on archaeological grounds (406, 22e, 109, 201) and four graves from the later Period II (398, 10, 46, 303c). Thus, half of the graves of *Shahr_I_Sokhta_BA2* are consistent with being more recent than those of *Shahr_I_Sokhta_BA1*, dating to the mid or to the beginning of the second half of the 3rd millennium BCE. Of these four later graves, one (10) has a vase whose form closely matches well-known types from south-central Baluchistan, two others (46, 303c) have vessels that typologically might be linked to Baluchistan and/or possibly to the Makran coastal valleys. The archaeological and genetic evidence thus suggest that a flow of migrants from the northwestern borderlands of South Asia was active at the beginning of the local settlement, and that the same flow, different from the earlier one from northwest, intensified in the following centuries. We unfortunately do not have ancient DNA from Period III and the later centuries of the Shahr-i-Sokhta sequence when cultural influence from the Indus Valley Civilization appears to become stronger (*47*, *173*).

We provide detailed descriptions of the individuals from Shahr-i-Sokhta organized by their genetic label below. Here we use the context dates based on the absolute chronology outlined in (*174*), which was later criticized by (*175*) which proposed absolute dates 2-3 centuries earlier for Period III. The true chronology is still contentious and is currently being revisited in the light of the radiocarbon dates generated for this study. However, whatever the true chronology turns out to be will not affect the picture of genetic exchange and cultural interaction between the Helmand basin and the northwestern regions of the Indo-Pakistani subcontinent in the 3rd millennium BC emerging from the present study.

A total of 11 individuals genetically cluster into the *Shahr_I_Sokhta_BA1* analysis label (*Shahri_I_Sokhta_BA1* split label).

• 38, T.38B (I8724): Context date of 3000-2800 BCE [detected as a 2nd to 3rd degree relative of I8725 at 3010-2881 calBCE (4300±25 BP, PSUAMS-3989]. Genetically male. There were three consecutive burials in this circular pit grave, that on the basis of the pottery types, is estimated to have begun in early Period II, phase 6, and ended in Period II, early phase 5. The order of deposition was individual C (morphologically female), B (morphologically male, I8724), and A (morphologically female). Few pots were associated with this grave, but beads made from turquoise, chlorite, and calcite were found. I8724 is genetically detected as being a 2nd or 3rd degree relative of I8725.

• 48, IRL G-48 (I8725): Date of 3010-2881 calBCE (4300±25 BP, PSUAMS-3989). Genetically male. This individual was associated with one of three burials that occurred during late Period I, phase 8/7 in a badly preserved, round catacomb grave. The sex of this individual could not be determined morphologically. The grave was poorly furnished, and the chamber and previous remains were disturbed due to the other individuals being buried at different times. This individual is genetically detected as being a 2nd or 3rd degree relative of I8724.

• 405, IRV Grave 45 (I11461): Context date of 3200-2100 BCE. Genetically male.

• 406, IRV Grave 309 (I11462): Context date of 3000-2900 BCE. Genetically male. Single burial (Period I, phases 10 or 9) in a catacomb grave (vertical shaft at the southeast, lateral chamber at the northwest, closed with a mudbrick wall). The deceased (Individual A, age at death not determined) was placed with flexed lower limbs on the right side. A broken pottery bowl was placed at the bottom of the shaft in front of the wall. A pottery beaker, a banded travertine bowl, and two beads (in turquoise and red limestone, respectively) were found in the chamber near the cranium.

• 412, IRD Grave 307 (I11468): Context date of 3000-2900 BCE. Genetically female. Single burial in a catacomb or pseudo-catacomb grave (Period I or early Period II), with access at the south and a lateral expansion or chamber at the northwest. Individual A (ca. 20-25 years old at death) was placed flexed on the left side. Furnishings included two large woven cylindrical baskets, and inside the larger one a pottery beaker, a calcite and an agate bead, and a cosmetic applicator in banded travertine was found.

- 413, Grave 411 (I11469): Context date of 3200-2900 BCE. Genetically female.

- 418, Grave 303A (I11474): Context date of 2700-2600 BCE. Genetically male. This is a re-opened collective burial, hosting three different individuals (A from which we obtained ancient DNA, approximately 45 years old; B an adult; and C an infant 7-8 months old at death). The last burial events happened during the later Period II. This was a catacomb grave with a vertical shaft at the south, and a chamber at the north; the entrance to the chamber had been filled with mudbricks and blocks of clay. Various ceramic fragments and other finds were found in the final filling of the shaft. The offerings associated with Individual A were two bowls in banded travertine and three ceramic beakers. Individual B was not associated with any objects. Another beaker (stylistically ascribed to a late phase of Period II, probably 5B) was placed near the infant.

- 420, IRV Grave 20 (I11476): Context date of 3200-2100 BCE. Genetically female. This was a simple pit burial of an infant approximately 2 years old at the time of death. The skeleton was flexed on the left side. There were no grave furnishings.

- 422, IRW Grave 7 (I11478): Context date of 3200-1900 BCE. Genetically male. The excavation records mention Grave 7 Sup (superior) as a simple pit burial, and it was described by the excavators as "totally disturbed." It contained a child's cranium and among the ceramics a small jar and a fragmentary beaker. Below this filled-in cavity, Grave 7 Inf. (inferior) was found, a catacomb grave (burial chamber at the north, access at the south) containing at different levels of the fill the partial bones of two individuals, A and B (both 9-10 years old at death). This burial contained two small pottery jars, a turquoise bead, and fragments of cloth. The few recovered objects are undiagnostic, but all three burials might belong to Period I. It is unclear from the archeological record which of the burials is the source of the ancient DNA.

- 423, IRW Grave 4 (I11479): Date of 2911-2761 calBCE (4245±25 BP, PSUAMS-4779). Genetically female. Single burial (Period II, phase 6), in a catacomb grave with vertical shaft at the south and lateral chamber at the north, closed with a mudbrick wall. Individual A, ca. 25 years old at death, was buried with a bone stamp seal worn at the wrist, an agate bead near the cranium, and a polychrome necked jar. Behind the pelvis, a woven bag of vegetal fiber and closed with a copper pin contained a cosmetic set: a bowl in banded travertine, its applicator in the same material and a *Cardium* valve with orange and black pigments.

• 427, Grave 9 (I11483): Context date of 2800-2600 BCE. Genetically female. Double burial in a catacomb grave (Period II, phases 7 or early 6), with a wide vertical shaft at the south and a chamber at the north. Individual A, ca. 25 years old at death, was placed to rest with a line of five woven fiber baskets set along the inner face of the wall. Two objects (a pottery bowl and a small calcite bowl) were found in two of these baskets. Individual B, resting on top (undetermined age), was buried with three additional baskets and a calcite bead.

A total of 10 individuals fall within the *Indus Periphery Cline* analysis label (*Shahri_I_Sokhta_BA2* split label). It is unclear from the archeological record which of the burials is the source of the ancient DNA.

• 398, IRV Grave 18A (I11456): Context date of 2600-2500 BCE. Genetically female. A re-opened collective burial, the first individual (A) was buried in Period II, phases 5A/5B. The grave was re-used, probably twice, in Period III (ca. 2500-2400 BCE or later). The first occupant (Individual A, ca. 25 years old at death, from which we obtained DNA) was buried in a catacomb grave (vertical shaft in a lateral chamber at the west, closed with a mudbrick wall). The original furnishings (including painted pear-shaped beakers, truncated-cone-like bowls, and two small jars) were disturbed when Individual B (ca. 30 years old at death) was buried and Period III unpainted pottery (slender unpainted beakers and hemispherical bowls with S-shaped contour, and tall ogival and cylindrical jars) was placed in the upper levels of the chamber. Other finds were a banded travertine vial with a round cosmetic applicator (Period III), a truncated-cone like bowl in banded travertine, five beads in white-greenish stones and turquoise, a copper pin, and a fragmentary stamp seal in the same material.

• 415, Grave 406 (I11471): Date of 3328-3022 calBCE (4450±20 BP, PSUAMS-4592). Genetically female. Single burial in a catacomb grave (vertical shaft at S and a lateral chamber at N closed with a mudbrick wall). In the chamber, the deceased (30-35 years at death) had the limbs sharply folded with an *Ovis/capra* kid placed between the skeleton and the inner face of the mudbrick wall. Pottery furnishings (a bowl, a large globular jar, small jar, and an alabaster conical bowl with cosmetic applicator in banded travertine) were left in the base of the shaft. On the body, there were 12 agate beads near the cranium and 38 beads (calcite, lapis lazuli and gold foil) that belonged to a necklace worn on chest. Our direct radiocarbon date is consistent with the archaeological ascribed date (Period I, or early Period II), but limited to the last centuries of the 4th millennium BCE.

• 401, IRW Grave 10 (I11459): Context date of 2500-2400 BCE. Genetically male. Simple pit burial (Period II (late), phase 5B) of a single individual (A), ca. 20-25 years old at death. The skeleton was flexed on the left side, and the partial bones of an *Ovis/capra* kid were scattered near it. Most of the grave furnishings were placed near the cranium (a globular jar, two bowls and three beakers). The jar was closed by a Nal-type cylindrical canister vessel, stylistically linked to central-southern Baluchistan. One of the bowls contained a wooden comb and 38 flint bladelets, 45 turquoise and three lapis lazuli beads, and a vegetable fiber basket. Two calcite beads were found near the cranium, a turquoise bead near the hand and two lapis lazuli beads near the legs. The late date of this individual which is contemporaneous with that of the mature Indus Valley Civilization, the clear links in the ceramic style with pottery from further East, and the fact that this particular individual has very high proportions of ancestry related to *AHG* and a H1 Y chromosome haplogroup typical of South Asia (both absent in all the individuals we have from Copper Age and Eneolithic Turan) is significant. It provides a link between the genetic and archeological evidence and suggests that this individual and possibly the other individuals we have from the *Indus Periphery Cline* are migrants from the Indus Valley Civilization or its immediate neighborhood.

• 404, IRQ Grave 31 (I11460): Context date of 3200-2100 BCE. Genetically male.

• 410, IRV Grave 46 (I11466): Context date of 2500-2000 BCE. Genetically female. Simple pit burial of an infant (ca. 1 year old at death). The skeleton was flexed on the left side. A small painted pottery jar (frieze on shoulder with panels of multiple vertical lines and rows of sigmas) possibly related to southwestern Baluchistan. A necklace of 24 turquoise beads was near the neck.

• 424, IRR Grave 22e (Central-E) (I11480): Context date of 2900-2800 BCE. The grave was archaeologically ascribed to Period I/II, phases 8-7. Genetically male.

• 432, IPV Grave 109 (I11488): Context date of 3200-2100 BCE. Genetically female. The furnishings include three bowls in grey ware possibly related to Baluchistan or to the Makran coast and an alabaster bead.

• 400, Grave 405 (I11458): Context date of 3200-2100 BCE. Genetically male.

• 201, Grave 201, LS1 (I8726): Context date of 3100-3000 BCE. Genetically male. This individual was one of two burials, deposited one above the other, in a catacomb grave dated to the early Period I, phases 10/9. The grave had a few pots and was otherwise poor except for the inclusion of an alabaster vessel. Salvatori et al. (*176*) note that a distinctive pottery type among the grave goods belongs to a distinctive cluster of graves at Shahr-i-Sokhta that is "possibly local or northeast oriented (Kandahar area)" (*172*).

• SHAR_303C, T.303C (I8728): Context date of 2550-2450 BCE. Genetically male. This catacomb grave contained three burials and was dated to Period II, phase 5b. Individual I8728 was buried on top, and morphological sex was undetermined. The grave was not rich except for an alabaster vessel.

One individual fell genetically in the *Shahri_I_Sokhta_BA3* analysis label and split label, reflecting additional Anatolian farmer-related ancestry compared to the other samples.

• 416, Grave 408 (I11472): Context date of 3000-2800 BCE. Genetically female. Single burial in a catacomb grave (Period I (late) phase 8), with a vertical shaft at the south and a lateral chamber at the north, closed with a mudbrick wall. Individual A, an adult of undetermined age, was placed on the right side in flexed position. The furnishings—a beaker and a bowl in Common Buff Ware, a small globular jar in Gray Ware, and a cosmetic applicator in banded travertine—were placed below the left leg.

Relatedness summary:

• 2-person family: I8724-I8725 are genetically detected to be 2nd to 3rd degree relatives. Based on the stylistic variations in pottery, the link between the individuals sampled in I8725 and I8724 is likely to be a relationship between a grand-grandfather and grand-nephew.

S2.3.4.5 Gonur Tepe, Bactria Margiana Archaeological Complex (BMAC), Turkmenistan (n=47)

Description by James Mallory and Nadezhda Dubova

We present genome-wide ancient DNA data from 47 individuals from the Bactria-Margiana Archaeological Complex (BMAC) site of Gonur Tepe, dating to 2300-1600 BCE. The BMAC was primarily an urban civilization, and it had geographic and cultural contacts with cattle-breeding populations from the Steppe that often settled close to the towns and whose characteristic pottery is sometimes found within BMAC settlements. The following section lists cemeteries or individual burials at the BMAC site of Gonur Tepe.

Gonur Tepe is situated in southeastern Turkmenistan 85 km north of Bayramali, the ancient city of Merv. It was the largest of the more than 200 settlements in the ancient delta of the Murghab River. The site was discovered in 1972 and excavated until the early 21st century by Prof. Victor Sarianidi. Excavation continues on the site governed by an agreement between the Institute of Ethnology and Anthropology of RAS (Moscow) and the Ministry of Culture of Turkmenistan. Gonur Tepe is characterized by monumental architecture, fine jewelry and pottery, a highly-developed bronze industry, and a complicated social structure and symbolic system. A variety of funeral rituals and types of graves with similar sets of grave goods is known. Many trade and cultural connections, including evidence for migration, have been established by archaeological and physical anthropological data with Near Eastern, southern Trans-Caspian and Indus Valley communities. In addition, there is archaeological and physical anthropological evidence indicating contacts with Eurasian Steppe populations.

The main Gonur cemetery covers an area of 10 hectares and comprises well over 3,000 burials. Burials were routinely inhumed in the flexed position, lying on their right side, with their head oriented to the north or northwest. There were a variety of grave structures. The most frequent (85%) were shaft graves comprising a vertical shaft at the base of which was a large hollowed-out niche (catacomb) in which the deceased was inserted. Simple pit graves constitute about 11% of the burials and these constitute the poorest of the graves in respect to grave gifts (or lack of gifts). Approximately 2% of the burials were placed in cists, brick-built chambers that were usually erected on the surface of the ground. These could be quite rich with respect to grave goods although the overwhelming majority of these tombs were plundered in antiquity. Finally, there are the most elaborate chamber graves where a virtual room (bedroom) was replicated in brick and largely sunk underground. Unlike the cists that were sealed, the chamber tombs could be reopened and permitted successive burials similar to a family mausoleum. Although almost all of these tombs had been plundered, they would appear to have been the graves of the aristocracy.

We observed extensive genetic diversity in the group, with the main cluster of 38 individuals broadly related to farming communities of the Iranian plateau; we assign this group the *BMAC* analysis label (**Supplementary Materials S4**). The data also include 9 outlier individuals (**Supplementary Materials S4**), who we break into 5 additional analysis labels as discussed in detail below:

BMAC analysis label (n=38; main genetic cluster)

- Gonur 2003 Tomb 3012 Sample 1 (I1781): Date of 2009-1772 calBCE (3550±30 BP, PSUAMS-2065). Genetically and skeletally male. This burial of a male, skeletally determined to be 25 to 30 years old at the time of death, was found flexed on the right side, oriented north-northwest, in a two-chambered cist in Room 218 in Area 5. The burial was accompanied by 20 ceramic vessels along the west wall of the cist. This individual is genetically determined to a probable first degree relative of I6117.

- Gonur Tomb 3037 Sample 4 (I6117): Context date of 2500-1600 BCE. Genetically male. This individual is genetically determined to be a probable first degree relative of I1781.

- Gonur 2003 Tomb 3049 sample 3 (I1782): Date of 2288-2142 calBCE (3785±20 BP, PSUAMS-2309). Genetically female. Burial of an individual skeletally identified as a 30 to 40-year-old male was found flexed on the right side, oriented northwest, in a pit grave (80 x 70 cm, 25 cm deep) in Room 45 in Area 6. The individual was accompanied by one ceramic vessel.

- Gonur Tomb 2380 sample 17 (I1784): Date of 2201-2031 calBCE (3720±30 BP, Poz-83485). Genetically male. Nicknamed the 'Tomb of the Warrior,' this was skeletally a male, 40 to 50 years old at the time of his death, flexed supine and oriented north-northwest. The neck of the man was broken, and this was the probable cause of his death. He was buried in a shaft grave on the southeast edge of the large cemetery. This very rich grave was accompanied by one bronze knife, one silver plate, one bronze vessel (diameter 16 cm and height 12 cm), one bronze mace head in the form of a horse head, one bronze mace head with four spikes, one bronze semi-cylindrical artifact near the head, one bronze leaf-shaped arrowhead near the pelvis, and one bronze plate with perforations wrapped in linen cloth near the right shin.

- Gonur Tomb 1315 Sample 65 (I1787): Date of 2139-1981 calBCE (3680±20 BP, PSUAMS-2310). Genetically female. The scattered remains of an individual determined to be female based on skeletal evidence, and skeletally estimated to be 40 to 50 years old at the time of death, was recovered from the fill of a robbed two-chamber cist that also contained the remains of a male estimated to be 30 to 40 years old at the time of death. The bones, some of which revealed traces of bronze oxides, were accompanied by 28 ceramic vessels, some of late Namazga VI type, one gypsum spindle, and one fragment of kaolin.

- Gonur Tomb 1340 Sample 66 (I1788): Date of 2127-1905 calBCE (3630±30 BP, PSUAMS-2066). Genetically female. The sample derived from one of three human remains scattered through the fill of a robbed shaft grave (without a step). All three individuals were skeletally female. The individual sampled was skeletally determined to be 35 to 40 years old. The tomb yielded sherds of four ceramic vessels, and three faience beads. Some bones revealed traces of bronze oxides.

- Gonur 2013 Area 12 pit room 29 N4258 N15m (I1790): Date of 2134-1957 calBCE (3660±20 BP, PSUAMS-2311). Genetically female. The teeth, identified odontologically as belonging to a female 11 to 18 years of age, were recovered from Room 29 in Area 12, southwest of Gonur Tepe and outside the encircling wall. No grave goods were recovered.

- Gonur 2013 Area 12 occasional find N4384 N14m (I1793): Date of 2190-2029 calBCE (3700±20 BP, PSUAMS-2312). Genetically female. Teeth, identified as belonging to a female who was skeletally 14 to 16 years old, were recovered from a pit (not a formal tomb) near the cenotaph in shaft grave 4367 in Area 12 (southwest of Gonur Tepe, outside the encircling wall).

- Gonur Tomb 1506 Sample 32 (I2087): Date of 2196-2034 calBCE (3715±20 BP, PSUAMS-2335). Genetically and skeletally male. This individual, skeletally determined to be 35 to 45 years old at the time of his death, was found flexed on his left side, oriented north-northwest, in a robbed shaft grave with a step in the main cemetery. Sherds of two ceramic vessels were found near his head.

- Gonur 2013 Area 19 cist tomb 4290 (1) N2m (I2116): Date of 2118-1883 calBCE (3605±35 BP, Poz-83490). Genetically female. This individual was skeletally determined to be a female, 5 to 15 years old, recovered from a robbed rectangular cist located north-northwest of Gonur Tepe just outside of the encircling wall. The tomb also contained the remains of an adult female, skeletally, aged 35 to 40. Grave goods included one ceramic goblet near the head of the adult female, five ceramic vessels (and some small sherds) as well as a terracotta female figurine, one gold bead, one gold conical plaque, one gold earring, one gold rosette, and traces of bronze oxides.

- Gonur 2003 Tomb 3007 sample 8 (I2121): Date of 2203-2041 calBCE (3735±20 BP, PSUAMS-2314). Genetically female. Skeletally identified as a male 40 to 50 years in age, was buried in a flexed position on the right side, oriented N. The burial had been placed in a former oven in room 129 in Area 5, north of Gonur Tepe between the citadel and the second retaining wall. There were no grave goods.

- Gonur tomb 2871 sample 15 (I2125): Date of 2136-1977 calBCE (3670±20 BP, PSUAMS-2315). Genetically female. An individual, osteologically male, skeletally determined to be 35 to 40 years old at the time of death, was found in a flexed prone position, oriented north, in a former oven in Area 5, north of Gonur Tepe between the citadel and the second retaining wall. No grave goods were found.

- Gonur Tomb 1311 sample 59 (I2128): Date of 2198-2036 calBCE (3720±20 BP, PSUAMS-2316). Genetically male. Osteologically male, skeletally determined to be 25 to 30 years old at the time of his death, oriented north-northwest in a robbed shaft grave with a step in the main cemetery. Grave goods were sherds from two ceramic vessels.

- Gonur 2003 Area 6 Tomb 3042 sample 7 (I3374): Date of 2125-1945 calBCE (3645±20 BP, PSUAMS-2317). Genetically female. A skeletally determined female, around 9 years old at the time of her death, was found in prone position with hands on face and legs flexed to right side, oriented northwest, in a pit tomb in the “small round altar” in Area 6 (east side of Gonur Tepe, between the citadel and the second retaining wall). No grave goods were found.

- Gonur 2004 Area 8 Cist 3201 N28m (I2085): Date of 2011-1886 calBCE (3580±20 BP, PSUAMS-2313). Genetically male. A male, osteologically, 40 to 60 years old at the time of his death, was buried on the right side near the east wall of a cist between rooms 4 and 6 in South Area 8, south of Gonur Tepe and outside the encircling wall. The grave had been robbed. One ceramic vessel was recovered.

- Gonur 2005 Tomb 3466 Sample 2 (I6119): Date of 2130-1948 calBCE (3650±20BP, PSUAMS-2840). Genetically male. This infant, 3-4 years old, was found lying on its left side, head to the west. Three ceramic vessels were found in the tomb.

- Gonur 2000 Tomb 1646 Sample 39 (I6122): Context date of 2500-1600 BCE. Genetically male. Burial of an individual, anthropologically female, 16 to 17 years old at time of death, buried in the flexed position on the right side in a shaft grave, oriented west-northwest. The burial was accompanied by one ceramic vessel.

- Gonur 2003 Area 5 Cist N3011 N31m (I3365): Context date of 2500-1700 BCE. Genetically female. Burial of an individual, anthropologically female, 40-50 years old, lying supine with head to the north, hands near her face. In front of her hands stood a large conical red-slipped ceramic vessel. Close to it was a miniature vessel with a small spout, also red-slipped. Four more ceramic vessels stood near the woman's legs and near her left shoulder was a stone amulet decorated on both sides. At her feet were two additional vessels, one of which was a miniature.

- Gonur 2003 Tomb 1906 Sample 49 (I6125): Context date of 2500-1600 BCE. Genetically female. Burial of an adult with remains scattered in a shaft grave, oriented north. Burial was accompanied by 4 ceramic vessels and one faience bead.

- Gonur 2003 Tomb 3050 Sample 6 (I6118): Context date of 2500-1600 BCE. Genetically male. The burial was in an oven in Area 7 (room 1) in the southern part of the central area between the citadel and the second enclosing wall. The individual was anthropologically male, around 6 years old, strongly flexed in the supine position with head to the southwest. The knees were near the elbows with the hands in front of the face.

• Gonur 2005 Area 10 pit N3300 (1) N22m (I7102): Context date of 2000-1800 BCE. Genetically female. Two burials of children were in a pit grave in Area 10 (northwest of the central part of Gonur Tepe). The first [3300 (1)], anthropologically female, 8-9 years old, lay on her right side, head to the northwest, legs flexed. The skull was artificially deformed. Her forehead rested on the occipital bone of the second burial [3300 (2)].

• Gonur 2006 Area 10 pit tomb 3300 (2 = individual 2), N23m (I7170): Context date of 2000-1800 BCE. Genetically male. This was the second of two children buried in tomb 3300 (the first child was I102). The individual was anthropologically male, about 6 years old, and lay in the supine position, head to the northwest, legs bent vertically at the knees. Near the loins of this boy were one ceramic vessel and near the top of the head, another vessel.

• Gonur 2005 Tomb 3483 Sample 5 (I6126): Context date of 2500-1600 BCE. Genetically female. This was a burial in a shaft grave in Area 12 (southwest part of the site, beyond the main enclosure wall). The individual was anthropologically female, older than 60 years, lying on the right side, head to the northwest. Her hands were near the face, legs bent at the knees. In the northern part of the shaft stood one globular ceramic vessel. One miniature ceramic vessel and a round bronze item were near the top of the skull.

• Gonur 2006 Area 13 pit 3536 25m (I7171): Context date of 2500-1700 BCE. Genetically female. Burial in a ground pit in Area 13 (southwest of the site, beyond the second enclosure wall). The pit was in a large room with a double-hearth. Anthropologically possibly male, 8-9 years old, lying on the right side, head to the west. The hands were near the face, the heels at the pelvis. There were no grave goods.

• Gonur 2013 Area 22 shaft tomb N4236 N8m (I7173): Context date of 2500-1700 BCE. Genetic sex is unclear. This burial was in a shaft grave in Area 22 (south of the site, to the south of the “Royal Cemetery”, near the group of “dog burials”). The skeleton is anthropologically determined to be that of a 18-20 year old individual possibly a female lying in the shaft on the right side, head to the north. Near the head, in the north part of the shaft there were 6 ceramic vessels. Between the hands and the rib cage were two bronze bracelets. Close to them beads: 1 white and 3 black stone, 2 lapis-lazuli, 1 bronze, and 1 lead. On the 4th finger of the right hand – a simple bronze ring. The shaft was closed by 12 mud bricks. The head of a fetus (about 30 weeks) was close to the elbows of the young adult, and lay in a fetal posture.

• Gonur Tomb 1128 Sample 22 (I6127): Context date of 2500-1600 BCE. Genetically male. Burial of an individual, anthropologically female, 30 to 35 years at time of death, remains scattered in a shaft grave, oriented north-northeast.

• Gonur Tomb 1300 Sample 51 (I6217): Date of 2285-2135 calBCE (3770±20 BP, PSUAMS-2806). Genetically female. Burial of a female (genetically) whose remains were scattered with three other burials in a chamber grave. The goods recovered from the grave included 20 ceramic vessels, one gold bead, 100 gypsum beads, two bone hair-pins, two bone arrowheads, fragments of silver and gold foil, foil-covered objects of gypsum, and one steatite mace head.

• Gonur Tomb 1306 Sample 57 (I6218): Context date of 2500-1600 BCE. Genetically female. Burial of an anthropologically determined female, 50 to 60 years at time of death, whose remains were scattered along with another female of the same age in a shaft grave, oriented north-northwest. Burials were accompanied by two ceramic vessels and one stone bead.

• Gonur Tomb 1307 Sample 62 (I6120): Context date of 2500-1600 BCE. Genetically male. Burial of an individual, anthropologically female, 18 to 20 years at time of death, scattered in a possible pit grave, oriented north-northwest. Burial was accompanied by two ceramic vessels and one zoomorphic figurine.

• Gonur Tomb 1368 Sample 20 (I6310): Context date of 2500-1600 BCE. Genetically female. Burial of an individual, anthropologically female, age 30 to 40, flexed on the right side (possibly) with another burial of similar age in a shaft grave, possibly oriented west-northwest. Burial was accompanied by 4 ceramic vessels.

• Gonur Tomb 1415 Sample 46 (I6312): Context date of 2500-1600 BCE. Genetically male. Burial of an individual, anthropologically female, 35 to 40 years at time of death, remains scattered in a shaft grave, oriented north-northwest.

• Gonur Tomb 1651 Sample 45 (I6318): Context date of 2500-1600 BCE. Genetically female. Burial of a female (genetically), 20 to 25 years at time of death, flexed on the right side in a shaft grave, oriented north-northwest. The burial was accompanied by a female (anthropologically, 60 to 70 years at time of death. Burial included 1 ceramic vessel.

• Gonur Tomb 1466 Sample 25 (I7101): Context date of 2500-1700 BCE. Genetically male. Burial of an anthropologically determined male, 50 to 60 years at time of death, flexed on the right side in a shaft grave, oriented north-northwest.

• Gonur Tomb 3734, UfaNov17-017 (I10410): Context date of 1900-1700 BCE. Genetically female. The burial was in a shaft grave in Area 16 (southwest of the site, close to the east side of the Large Cemetery, room 52). This individual was anthropologically female, 16-18 years old, lying on her right side, head to the north. Her hands were near her face, heels near the pelvis. The west wall of the shaft was destroyed by another tomb. Near the head of the deceased stood three ceramic vessels, and small stone cylindrical and ring beads were near her humerus.

• Gonur skull 3250, UfaNov17-055 (I10411): Context date of 2300-2200 BCE. Genetically male. The burial was in a two-room chamber-tomb in the “Royal Necropolis” (near burial 3220). Anthropologically male, 50-60 years old. The chamber tomb was organized on the outer northwestern part of burial 3220. The “encircling wall” of tomb 3220 was built on the top of the southern part of this chamber. Tomb 3250 was robbed in antiquity. All human remains were concentrated (not in anatomical position) in the northern part of the tomb. In different places of the tomb were found only a few fragments of burnt stone mosaics. All other funeral gifts were stolen.

• Gonur tomb 3699, UfaNov17-002 (I11040): Context date of 2000-1800 BCE. Genetically female. The burial was in a shaft grave in Area 16 (southwest of the site, close to the east side of the Large Cemetery). This tomb is just under pit grave 3703. Anthropologically female, 50-60 years old, lying in the shaft on her right side, head to the north. Four ceramic vessels stood close to the top of the skull in the northern part of the shaft.

- Gonur tomb 3220, skeleton 3, UfaNov17-005 (I11042): Context date of 2300-2250 BCE. The remains derive from the primary burial in Tomb 3220, the central tomb of the "Royal Cemetery". This consisted of an underground "house" with 4 rooms. Rich funeral offerings were deposited: 2 stone scepters, 1 stone disk and 2 miniature columns, gold, silver and bronze items, a stone sculpture of a sheep, an agate bead and 51 arrow heads (44 flint and 7 bone), mosaic panels depicting pipala-leaf compositions and some other items. Interred were three possible primary burials (one man and two women) and 16 servants (8 women and 11 men). The individual was anthropologically male matching the genetic sex determination, and estimated to be 35-40 years old. The main part of the remains were scattered and found in all the rooms of the "house of dead," and the remains from which ancient DNA was extracted belong to the skeleton buried under the floor in the southeastern corner of the tomb. He lay on the right side, head orientated to the northeast. His head was lying on a stone sculpture of a sheep. A wooden box with mosaic decoration (which has been called the "ostensory") stood on top of the buried body. Near the legs of this person there was a treasure (25 large gold, silver and bronze vessels) and a large bronze dagger.

- Gonur tomb 3385, UfaNov17-015 (I12255): Context date of 2000-1700 BCE. Genetically female, The burial was in a pit tomb in Area 12 (southwest part of the site, out of the main enclosure wall; there was a special group of tombs in the north of this Area). This individual was anthropologically female matching the genetic determination, was 40-50 years old, and was lying strongly flexed on her right side, her knees close to the elbows. A miniature ceramic red-slipped round vessel lay between her thighs and forearms. Head to the west. Near it, in the northern part of the tomb, there were three ceramic vessels.

BMAC_o2 (n=2): Individuals with a mixture of ancestry related to the main *BMAC* cluster along with additional ancestry related to *Western_Steppe_MLBA/Central_Steppe_MLBA*.

- Gonur 2013 Area 12 shaft tomb N4329 (I1789): Date of 2277-2030 calBCE (3735±35 BP, Poz-83486). Genetically male. This tomb contained two burials, one inserted after the other, and it is uncertain which of the graves has been analyzed. Both graves were sexed as female on anthropological grounds, although the genetic evidence indicates a male, although all of the grave goods associated with burial 1 are consistent with a female burial: one ceramic vessel near the head, and four vessels near the legs, one of which contained a smaller vessel of the same type and one steatite spindle, one globular and 13 disk shaped stone beads, one gold earring, a bronze mirror with a ceramic sherd and 11 globular white stone beads. Burial 1 belonged to an individual skeletally determined to be 15 to 20 years old at the time of death while the remains of burial 2 were from an individual determined to be 35 to 50 years old.

- Gonur 2005 Area 13 Room 12 tomb 3453 sample 11 (I2122): Date of 2139-1981 calBCE (3680±20 BP, PSUAMS-2152). Genetically female. An individual of undetermined sex by craniology but male by odontology and about 9 years old at the time of death, was found in a destroyed pit in Room 12 in Area 13, southwest of Gonur Tepe between the second retaining wall and encircling wall. No grave goods were found.

BMAC2 (n=1): An individual with a mixture of ancestry related to the main *BMAC* cluster along with additional Anatolian farmer-related ancestry.

- Gonur 2003 tomb 1899 sample 54 (I6124): Date of 2193-2031 calBCE (3705±20 BP, PSUAMS-2804). Genetically female. Burial of an anthropologically determined female, 35 to 40 years old at time of death, with remains scattered in a shaft grave, oriented north-northwest. Burial was accompanied by 15 ceramic vessels and one bone hair-pin.

Gonur1_BA_o (n=2): Individuals with a mixture of ancestry related to the main *BMAC* cluster and additional ancestry related to *Central_Steppe_EMBA/WSHG*.

- Gonur 2005 Tomb 3454 Sample 12 (I1783): Date of 2275-2024 calBCE (3725±35 BP, Poz-83484). Genetically female. An individual identified as a female osteologically, and 20 to 25 years at the time of death, was recovered from a destroyed pit along with some sheep remains in Area 13, southwest of Gonur Tepe, between the second retaining wall and the encircling wall. No grave goods were recovered.

- Gonur 2004 large pit, royal cemetery 3240(9) (I1792): Date of 2458-2202 calBCE (3840±35 BP, Poz-83487). Genetically male. The individual, sexed as female on skeletal grounds, was 50 to 60 years old at the time of death. The individual was buried in a large rectangular pit near tomb 3235 in the “Royal Cemetery” where 17 individuals were found together with a four-wheeled wagon, two camels, two dogs, one calf, and two lambs. The tomb had been robbed and the bones of the deceased were found in the northwest part of the grave at various depths. Grave good associations are unknown.

Gonur4_BA (n=1): An individual with less Anatolian farmer-related ancestry (and relatively more Iranian farmer-related ancestry) than the main *BMAC* cluster.

• Gonur tomb 3230, skeleton 4, UfaNov17-001 (I11039): Date of 2297-2147 calBCE (3805±20 BP, PSUAMS-4602). Genetically female. This grave from Tomb 3230 in the ”Royal Cemetery” was partly destroyed and robbed in antiquity. It was comprised of an underground house with three rooms and an entrance hall. Some stone artifacts--symbols of power (1 miniature column, 1 scepter and 1 disk), many semi-precious and gold beads, gold foil, 10 flint arrowheads, a bronze axe, two bronze seals, a bronze knife and 30 large ceramic vessels--were found in this tomb. The three main individuals (2 men, 1 woman) were buried in a separate room in the southwestern part of the tomb. Remains of 7 servants (5 women and 2 men, including one dwarf) were lying close to the entrance to the tomb. Four other male and four female skeletons were scattered and their bones found in different places of the tomb. The female skeleton (45-55 years old by anthropological data), whose DNA was examined was lying on the right side, head to the south-west, close to the other female remains near the entrance.

Indus Periphery Cline (n=3): This group also has the *Gonur2_BA* split label. Like the individuals in the *Indus Periphery Cline* analysis label from the site of Shahr-i-Sokhta in northeastern Iran, these individuals have a mixture of ancestry related to South Asian hunter-gatherers on the one hand, and on the other hand a distinctive type of West Eurasian-related ancestry with a lower ratio of Anatolian to Iranian farmer-related ancestry than contemporary individuals sampled from Iran or Turan.

• Gonur 2005 tomb 3465 sample 9 (I2123): Date of 2452-2140 calBCE (3815±35 BP, Poz-83491, date suspect because no quality collagen). Genetically female. An infant, identified as male by odontology, 1.5 to 2 years old, was found in a sitting position in a large ceramic vessel in Room 350 in Area 10 (northwest part of Gonur Tepe between second retaining wall and encircling wall). Grave goods were two miniature ceramic vessels.

- Gonur tomb 3225, skeleton 3, UfaNov17-003 (I11041): Date of 2140-1972 calBCE (3675±25 BP, PSUAMS-4599). Genetically male. The Great Pit ("ditch"), Tomb 3225, was situated in the "Royal Cemetery" and contained a four-wheeled wagon. It was situated in the "Royal necropolis" close to Tomb 3230. Two camels (an adult and young) were also buried there. In this "ditch" remains of 10 people (7 male, 1 female and 2 persons whose sex cannot be identified) were found. There were four full skeletons and six separated skulls lying in different places. Skeleton 3 was a complete skeleton who was anthropologically male matching his genetically determined sex, 35-40 years old. He was found near the north wall of the tomb. He was positioned on his knees, slightly falling on his left side.

- Gonur tomb 3225, skeleton 1, UfaNov17-004 (I10409): Date of 2280-2044 calBCE (3755±25 BP, PSUAMS-4603). Genetically male. Skeleton 1 was anthropologically determined to be a male (matching the genetic sex) 45-55 years old, and was lying prone, head to the south on one of the wheels of the wagon.

Relatedness summary:

- I1781-I6117 are genetically detected as likely to be first-degree relatives.

S2.3.4.6 Jarkutan, post-BMAC, Uzbekistan (n=10)

Description by Michael Frachetti

The site of Jarkutan (Dzharkutan) consists of a large area (about 100 hectares) containing both a fortified dispersed settlement and a large necropolis. The burial ground occupies more than 20 hectares and is located on two large hills along the ancient bed of the Sherabad River in the Surhandarya Province of south Uzbekistan. The site was discovered in 1973. The micro-relief of the site shows that Jarkutan 1 (4 hectares) was the settlement's fortress, Jarkutan 2 was the lower settlement, and Jarkutan 3 and 4 (18 and 2.5 hectares respectively) were the locations of the burial mounds.

The main burial excavations were conducted at the Jarkutan 4A burial mound, where about 731 graves with different levels of disturbance were uncovered in an area of 1 hectare. The burial mound consists of pit graves, visible on the modern surface as shallow depressions. Preliminary analysis of the findings from the graves established three chronological stages – the Jarkutan, Kozali and Mullali phases – of the greater Bronze Age “Sappali Culture” of northern Bactria. The main feature of Jarkutan grave construction is niche-type and catacomb pits. Bodies are most commonly placed in a flexed position, where men are laid down on their right sides, while women are placed on their left side. Another feature is the standardization of funeral items and similar burial rituals across the area in each phase.

Funeral rites changed throughout the periods. For instance, the early “Jarkutan Stage” differs by the presence of females buried with numerous items of jewelry and fine ceramic ware, perhaps reflecting a matriarchal society. Later in the Kozali and Mullali Stages, the female burials become poorer, while there is an increased presence of metal in male burials, reflecting a relative rise of male status. Later, metal weapons and labor tools are replaced with imitations and the number of metal jewelry objects decreases. At the Mullali Stage, metal objects were completely replaced with votive objects.

Another specific feature of Jarkutan is the abundant cenotaphs (empty graves) containing rich funeral items. At the Jarkutan Stage the cenotaphs did not include sacrificed animals, which is a common feature of Kozali Stage graves that intensified during the Mullali Stage.

BMAC (n=8): The main group of individuals falls within the *BMAC* analysis label, so-named because of tight genetic clustering with the main cluster at the BMAC site of Gonur. The split label is *Dhzarkutan1_BA.*

- UZ-JAR-001, Jarkutan 4a 1976, Grave 401 (I4312): Date of 1736-1621 calBCE (3370±20 BP, PSUAMS-2516). Genetically female.

- UZ-JAR-004, Jarkutan 4b-85, Grave 60 (I4315): Date of 1609-1465 calBCE (3255±15 BP, PSUAMS-2518). Genetically male.

- UZ-JAR-005 , Jarkutan 4a-86, Grave 106, 44-20 (I4313): Date of 1513-1431 calBCE (3210±20 BP, PSUAMS-2517). Genetically female.

• UZ-JAR-007, Jarkutan 4a 1976, Grave 384 (I7411): Date of 1686-1534 calBCE (3335±20 BP, PSUAMS-3227). Genetically female.

• UZ-JAR-008, Jarkutan 4a 1975, Grave 206 (I4161): Context date of 2100-1800 BCE. Genetically female.

• UZ-JAR-011, Jarkutan 4a 1975, Grave 211 (I4163): Date of 1611-1453 calBCE (3250±25 BP, PSUAMS-2112). Genetically female.

• UZ-JAR-018, Jarkutan 4a 1977, Grave 673, 57-25 (I7412): Date of 1749-1642 calBCE (3405±20 BP, PSUAMS-3228). Genetically female.

• UZ-JAR-003, Jarkutan 4b-137, 44-24 (I4314): Date of 1885-1701 calBCE (3480±25 BP, PSUAMS-2800). Genetically female.

BMAC_o2 (n=2): These individuals have significantly elevated proportions of ancestry related to Central_Steppe_MLBA. We designate them by the BMAC_o2 analysis label and the Dhzarkutan2_BA split label.

• UZ-JAR-021, Jarkutan, XG-1980, Catacomb N. 6, 56-95 (I4901): Context date of 2100-1800 BCE. Genetically female.

• UZ-JAR-022, Jarkutan, Catacomb N3, п. 2, XG-1980, 56-97 (I5608): Context date of 2100-1800 BCE. Genetically female.

S2.3.4.7 Sappali-Tepe, post-BMAC, Uzbekistan (n=14)
Description by Michael Frachetti

The Sappali-Tepe Settlement is located on the bank of springs flowing down from of the Kohitang mountains (southwestern Hissar Range of Pamir system), in Surhandarya Province of southern Uzbekistan. The initial fieldwork at the settlement was carried out in 1969-1971. The site is located on a hill covering roughly 3 hectares that is spread east-to-west and includes a central mound. Excavations at the site have produced rich material characterizing the culture, everyday life, religious beliefs, economy, and social organization of its inhabitants during the Middle and Late Bronze Age (mid-to late- 2nd millennium BCE).

Archaeological investigations in the larger western part of the settlement uncovered numerous dwellings (houses) separated by streets that formed neighborhoods, all surrounded by three parallel rows of fortress walls (separated by two rows of corridors). The excavations also exposed a unique quadrangle settlement with 84.6 m long fortress walls on each side built in a labyrinthine design.

Ancient burials at Sappali-Tepe included 46 undisturbed burials with rich funerary items with the artifact items identical to those in the settlement cultural contexts making it possible to assign them to the same period. These burials were found in the uncovered part of the settlement under the house floors, in the ruins of abandoned houses and streets, and inside the wall fillings and corridors. The burial constructions are divided into three groups: catacomb, niche-type and pit-type graves. There are also some bodies buried in large ceramic vessels, usually child burials, and two animal burials. The pit-type graves usually contained child burials. Common funerary items include ceramic vessels, as well as objects made of metal, stone, leather and wood. Three of the burials were collective burials.

BMAC (n=11): The main group of individuals falls within the *BMAC* analysis label, so-named because of tight genetic clustering with the main cluster at the BMAC site of Gonur. The split label is *Sappali_Tepe_BA.*

• UZ-ST-001, Sappali Tepe (ST), Grave 00-56 (I7419): Date of 1881-1701 calBCE (3475±20 BP, PSUAMS-3229). Genetically male. This individual was classified to be an adolescent 12 to 13 years-old in a pit grave.

• UZ-ST-002, Sappali Tepe (ST), Grave 00-91 (I7421): Date of 1931-1767 calBCE (3525±25 BP, PSUAMS-3120). Genetically male. This was a child burial, 6-8 months old, buried in a Jar.

• UZ-ST-003, Sappali Tepe (ST) 1971, 38, Grave 00-95 (I4285): Date of 1873-1661 calBCE (3430±25 BP, PSUAMS-2536). Genetically male.

- UZ-ST-005, Sappali Tepe (ST), Grave 00-102 (I4286): Date of 1886-1756 calBCE (3500±20 BP, PSUAMS-2165). Genetically female.

- UZ-ST-007, Sappali Tepe (ST), Grave 00-113 (I7414): Date of 2031-1915 calBCE (3615±20 BP, PSUAMS-3106). Genetically female. This individual was a catacomb burial, aged 40 to 45 years old.

- UZ-ST-008, Sappali Tepe (ST) 1971, 33, Grave 00-78 (I7492): Date of 1971-1782 calBCE (3560±20 BP, PSUAMS-3121). Genetically male.

- UZ-ST-009, Sappali Tepe (ST), Grave 00-112 (I7416): Date of 1948-1777 calBCE (3545±20 BP, PSUAMS-3117). Genetically female. This individual was a child 12 to 15 months old burid in a catacomb grave.

- UZ-ST-012, Sappali Tepe (ST), Grave 00-61 (I7420): Context date of 2000-1600 BCE. Genetically male. This individual was 40 to 45 years old buried in a catacomb grave.

- UZ-ST-014, Sappali Tepe (ST), Grave 00-59 (I7542): Date of 1885-1752 calBCE (3495±20 BP, PSUAMS-3231). Genetically female. This individual was a cenotaph, catacomb burial.

- UZ-ST-015, Sappali Tepe (ST), Grave 00-75 (I4288): Context date of 2000-1600 BCE. Genetically female. This was part of a double burial, that is, a paired grave. It was not possible to determine which individual the ancient DNA data was from. One individual 75(1) was 35-40 years old and the second, 75(2), 25 t o 30 years old. These were pit graves.

- UZ-ST-016, Sappali Tepe (ST), Grave 00-57 (I4289): Date of 1931-1767 calBCE (3525±25 BP, PSUAMS-2125). Genetically female. This was a female 25 to 30 years old in a catacomb burial.

BMAC_o2 (n=1): This individual has a significantly elevated proportion of ancestry related to *Central_Steppe_MLBA*. We call this individual by the *BMAC_o2* analysis label and the *Sappali_Tepe_BA_o* split label.

• UZ-ST-010, Sappali Tepe (ST) 1975, 6, Grave 02-05 (I7493): Context date of 2000-1600 BCE. Genetically male.

BMAC2 (n=2): These individuals have additional Anatolian farmer-related ancestry but not *Central_Steppe_MLBA* ancestry when compared to the main cluster of individuals from Sappali Tepe. We call this individual by the *BMAC2* analysis label and the *Sappali_Tepe2_BA* split label.

• UZ-ST-004, Sappali Tepe (ST), 71, 41, Grave 00-118 (I7495): Date of 1971-1782 calBCE (3560±20 BP, PSUAMS-3122). Genetically female.

• UZ-ST-006, Sappali Tepe (ST), 71, 8, zn2, Grave 00-149 (I7494): Date of 2010-1883 calBCE (3575±20 BP, PSUAMS-3230). Genetically male.

S2.3.4.8 <u>Bustan, mostly post-BMAC, Uzbekistan (n=14)</u>
Description by Michael Frachetti

The Bustan burial group is located along the right bank of the now dry bed of the Bustan River in the Sherabad District of Surhandarya Province, southern Uzbekistan. The archaeological and anthropological excavations started here in 1974 and have been continued until the present time with some breaks. Up until to now, more than 400 burials have been exposed and dated to the Mullali-Bustan latest period of the Sappali Culture. Bustan complexes are considered the latest stage of the Sappali Culture, dated to the middle part of the 2nd millennium BCE. The anthropological materials from the Bustan 1-7 sites are synchronous and similar to the Jarkutan materials, and the burial methods and bio-anthropological features of the skeletons do not differ. It is noteworthy that the direct AMS dates from the individuals from Bustan indicate that they are contemporaneous with the individuals from Jarkutan. The Bustan material assemblages have strong similarities to the Jarkutan materials and have been considered as part of a unique ethno-cultural system in the post-BMAC chronology.

The catacomb-type burials at Bustan consist of a shallow, rectangular entrance pit and burial chambers of oval shape with dome-shaped tops. The level of the chamber floor is usually lower than the level of the entrance pit so that the catacomb burials have a stair-type gradient.

After burying the body, the entrance hole was piled with mud-blocks and covered with clay plaster. Catacomb burials were also located under the floors and walls of the residential buildings as well as under the streets. The niche-type burials are similar to the catacombs except niche-type burials were always dug under the house walls, usually very shallow, and lacked entrance pits. The pit-type graves were usually used for adolescent burials.

Archaeological investigations at Bustan Burial Mound have revealed a complex funerary ritual related to the usage of fire. On top of the graves there were piled rocks, showing the influence of Steppe traditions. There were inhumation as well as cremation burials. There was a dedicated chamber for cremation of bodies at Bustan, including multi-usage hearths and altars. The altars were functionally classified into ones used for libations, ones used for meals, and ones used for sacrifices. The funerary rite documented at Bustan, specifically in relation to the role of fire, is not known at this time from any other site Iran, South Asia, or the Central Eurasian Steppes.

BMAC (n=10): The main group of individuals falls within the *BMAC* analysis label, so-named because of tight genetic clustering with the main cluster at the BMAC site of Gonur. The split label is *Bustan_BA*.

- UZ-BST-005, Site 7, Grave 60, 56-63 (I4157): Context date of 1600-1300 BCE. Genetically male.

- UZ-BST-007, Site 4, Grave 12 (I4159): Context date of 1600-1300 BCE. Genetically male.

- UZ-BST-010, Site 7, Grave 84, 56-62 (I5605): Context date of 1600-1300 BCE. Genetically female.

- UZ-BST-014, Site 5, Grave 11 (I4899): Context date of 1600-1300 BCE. Genetically male.

- UZ-BST-004, Site 7, Grave 33 (1), 44-91 (I4156): Context date of 1600-1300 BCE. Genetically male.

- UZ-BST-001, Site 4, Grave 37 (I11025): Date of 1600-1300 BCE. Genetically female.

• UZ-BST-013, Site 4, Grave 20 (I11519): Context date of 1600-1300 BCE. Genetically female. This individual is genetically detected as a 2nd or 3rd degree relative of I5064.

• UZ-BST-011, Site 4, Grave 26, 57-030 (I5604): Date of 1880-1697 calBCE (3465±20 BP, PSUAMS-2774). Genetically male. This individual is genetically detected as a 2nd or 3rd degree relative of I11519.

• UZ-BST-008, Site 7, Grave 84 (2), 58-20 (I11026): Context date of 1600-1300 BCE. Genetically female.

• UZ-BST-002, Site 7, Grave 33 (2), 44-92 (I11027): Date of 1662-1521 calBCE (3315±25 BP, PSUAMS-6198). Genetically male.

BMAC_o (n=1): This individual was an individual of low coverage. We did not perform quantitative analysis of its genetic ancestry in relation to other individual. However, on the PCA plots it appeared to be an outlier and we indicated it as such. We call this individual by the *BMAC_o_LowCov* analysis label and the *Bustan_BA_o_LowCov* split label.

• UZ-BST-006, Site 5, Grave 4, 57-28 (I4158): Context date of 1600-1300 BCE. Genetically male.

Bustan_BA_o1 (n=1): This individual has additional *WSHG* related admixture (probably via a population such as *Central_Steppe_EMBA*. We call this individual by the *Bustan_BA_o1* analysis label.

• UZ-BST-012, Site 4, Grave 41, 57-29 (I11521): Context date of 1600-1300 BCE. Genetically female.

Bustan_BA_o2 (n=1): This individual has additional ancestry related to the hunter-gatherers of South Asia. While the radiocarbon date of this individual places it to 700 years prior to the Swat Proto Historic Grave complex, the genetic ancestry appears to lie along a cline established by these individuals. We assign this individual the *Bustan_BA_o2* analysis label.

• UZ-BST-015, Site 4, Grave 4, 57-27 (I11520): Date of 1613-1509 calBCE (3280±20 BP, PSUAMS-4605). Genetically male.

Bustan_EN (n=1): This individual has greatly reduced *WSHG* and Anatolian farmer-related ancestries when compared to the other individuals from Bustan. Radiocarbon dating shows that this individual was actually from the Copper Age/Eneolithic period, and thus we use the anlaysis label *Bustan_EN*. The genetic ancestry seen in this individual suggests that the admixture that we detect in the later BMAC towns occurred during the Bronze Age.

• UZ-BST-009, Site ?, 58-21, 119-4 (I11028): Date of 3331-2972 calBCE (4445±25 BP, PSUAMS-4780). Genetically male.

Relatedness summary:
• 2-person family: I11519-I5604 are genetically detected as 2nd to 3rd degree relatives.

S2.3.4.9 Sumbar II, Sumbar Valley, Southern Turkmenistan (n=1)
Overview description by Andrey Gromov and Vyacheslav Moiseyev

This individual falls within the *BMAC* analysis label, so-named because of tight genetic clustering with the main cluster at the BMAC site of Gonur. The split label is *Sumbar_LBA.*

The Sumbar Valley together with Namazga was one of the main centers of early agriculture in southern Turkmenistan. Surrounded by the Kopet Dag Mountains, the area was well protected from the dry air from the Kara Kum Desert. The area also had a good water supply from the Sumbar River, which is a tributary of the Atrek River (*177*). Such conditions favored early introduction of agriculture to the area.

Regular excavations in Sumbar Valley began in 1969 by the Sumbar Archaeological Expedition of the Leningrad Sector of the Institute of Archaeology, USSR Academy of Sciences, headed by I. N. Khlopin. The archaeological complex includes the Eneolithic and Bronze Age settlement of Parkhai Tepe and several cemeteries, namely Sumbar I and Sumbar II (described below). Parkhai I and Parkhai II are described above.

• StPet34, grave 16, Museum ID 7651-153 (I6675): Context date of 1600-1000 BCE. Genetically male. Archaeological period SWT-I, Sumbar culture (Late Bronze Age). Morphologically male, 40 to 50 years old. Sumbar II, Grave 16 is oval in form (1.65 x 1.2 m) and elongated from east-to-west with an entrance located on the north side. Grave 16 is located to the east of grave 12. This grave was disturbed twice. The first time was during the construction of grave 18 when the entrance stones from grave 16 were removed and the northern part of the grave was partially destroyed, with vessels from grave 18 being placed on the left hip of the individual buried in grave 16. Later during construction of grave 17, the center of grave 16 was cut along its long axis. Although this resulted in cutting the large vessel in grave 16, the skeleton and other vessels that lie lower remained mostly undisturbed. The crouched skeleton was on its left side oriented from the east to west with face turned towards the entrance. The archaeological finds include six pottery vessels, a bronze ring, and fragments of two other bronze rings.

S2.3.5 Sites from historical period Turan

S2.3.5.1 Kushan empire context individuals from Ksirov, Tajikistan (n=5)
Description by Nadezhda Dubova and James Mallory

The kurgan burial ground of Ksirov in South Tajikistan is located in the middle part of the Vakhsh River in the Dangara district of the Khatlon region, and also spreads out along the river of the same name. It was excavated in 1977-1990 by N. Simakova and E.P. Denisov (*178*). The kurgans are located in groups (Ksirov I - Ksirov IX) between which may be either 150, 300 or 500 m and sometimes as much as 3-4 km. They include, in addition to the mounds themselves, enclosures, stone settings and cenotaphs. Most of the mounds (34 sites) are associated with enclosures. The diameters of the structures over the graves range from 3.5 to 4.5 m. In all the mounds (as well as in most of the enclosures), the graves are arranged in ordinary ground pits, which are filled with stones and loess. In the richest graves, the stones are found only at the very top of the filling. The buried lie mainly in the extended position. Most burials are in a west / southwest orientation, but some (especially female burials) also have a north / south orientation. The most important feature of the Ksirov ceramic complex is the presence of up to 40% of handmade ware, as well as a limited range of other forms. Pots and pitchers predominate. The same forms are available in wheeled pottery. Gold items (earrings, pendants) that in many respects resemble those found at Tillya-tepe in Afghanistan are found in a number of graves. There are also weapons. The material culture similarity of the graves at Ksirov to the monuments of the nomads of the Northeast of Central Asia is evidence. Analysis of the archaeological material and historical sources suggests that nomads who buried their relatives in Ksirov should be identified with the Yuezhi who, according to E.P. Denisov, was identical in the Bactria region to the Tokhars (Kushans). The inventory found in Ksirov and a number of features of the burial rites support this hypothesis.

• UfaNov17-053, Ksirov 1984, kurgan 7, enclosure 3, Ksirov III (I12257): Context date of 200 BCE - 100 CE. Genetically male.

• UfaNov17-049, tomb 26, Ksirov IV (I12260): Context date of 30-380 CE. Context date of 200 BCE - 1 CE (based on being genetically detected as a 1st degree relative of I12292 who has a radiocarbon date of 166-44 calBCE (2075±20 BP, PSUAMS-6159)). Genetically male.

• UfaNov17-047, Ksirov 1979, kurgan 25, Ksirov III (I12292): Date of 166-44 calBCE (2075±20 BP, PSUAMS-6159). Genetically male. This individual is genetically detected as a 1st degree relative of I12260.

• UfaNov17-052, Ksirov 1979, kurgan 8 (SE group), Ksirov III (I12293): Context date of 200 BCE - 100 CE. Based on direct dates on other skeletals at the same site.

• UfaNov17-054, Ksirov 1979, kurgan 9, Ksirov III (I12294): Date of 88 calBCE - 50 calCE (2020±20 BP, PSUAMS-6160). Genetically male.

Relatedness summary:
• 2-person family: I12260-I12292 are genetically detected as 1st degree relatives.

S2.4 Sites from South Asia

S2.4.1 Late Bronze-Iron Age Swat Protohistoric Graves (Gandhara Grave Culture), Swat Valley, Pakistan (n=99)

Descriptions by Luca M. Olivieri, Roberto Micheli, Massimo Vidale, and Muhammad Zahir

These individuals are from 8 sites in the Swat District of the Khyber-Pakhtunkhwa Province of northwestern Pakistan, all within 15 kilometers of each other: Udegram, Loebanr, Katelai, Arkotkila (Arkot-kala), Gogdara, Barikot, Butkara and Aligrama (each described below). They are from a material culture complex of the Late Bronze Age to Iron Age from northern Pakistan that is best known from an archaeological point of view in the Swat Valley, and known as "Swat Protohistoric Graves" or "Gandhara Grave Culture." (We use the former term here because the latter one is viewed by some as encompassing possibly unconnected ancient archaeological cultures.) Most of the individuals with reported ancient DNA data come from the IsMEO (now ISMEO) Italian Archaeological Mission in the 60s and 70s. The archaeological reports on these excavations proposed a now-outdated absolute chronology, although the relative chronology (based on ceramics) is still valid. Excavations (Udegram and Gogdara) were made by the same Mission in the framework of the ACT-Field School Project in 2011 and 2012 (*179*, *180*). These excavations provided more consistent information about grave architecture and funerary practices, as well as a consistent chronology. The summaries below reflect this synthesis.

Fig S 7 Image of the Swat Protohistoric Graves at Udegram

S2.4.1.1 Udegram, Swat Protohistoric Graves, Pakistan (n=16)

These individuals derive from an extensive graveyard in Udegram village in the Swat Valley (the left bank) dating to ~1200-800 BCE. The graveyard of Udegram (32 excavated graves) features two burial phases, encompassing 1400-1100 calBCE and 1000-800 calBCE. One inter-phase documented mainly from a nearby graveyard at Gogdara IV (approximately 1200-900 calBCE) is also represented at Udegram (*179*, *180*). With respect to the graves discussed in what follows, the stratigraphic evidence shows that Graves 3, 28, and 29 are the earliest ones, Grave 6 possibly belongs to the intermediate phase, while Graves 4, 5, 10, 12, 26, and 27 belong to the latest phase (see figs. 226, 229 in ref. (*180*)). The physical succession is not always reflected by the absolute dates (see below). In fact, the graveyard features a tight cluster of graves, sometimes superimposed, and dug in a steep sloping profile.

We generated direct radiocarbon dates on 8 of the 16 individuals with ancient DNA data, and while 5 are consistent with this chronology, there are 3 dates from bones that are stratigraphically in the earliest phase and that have definitively more recent dates than 1400-1100 BCE: Grave 29, Individual 1 (direct date of 1195-978 calBCE (2890±30 BP, Beta-428665)) and Grave 28, Individuals 1 and 2 (respectively: 921-831 calBCE (2740±20 BP, PSUAMS-2798) and 992-830 cal BCE (2760±30 BP, Beta-428664)). In fact, of the total of 22 dates we generated from the Swat Late Bronze-Iron Age individuals (not just from Udegram but also from the other sites), all are clustered in a time range of 1195-789 BCE conservatively considering the union of 95% confidence intervals, and 1087-804 BCE if we consider the range of the means of these dates. Thus, the individuals for which we obtained DNA that we directly radiocarbon dated for this study do not have any evidence of coming from the earliest burial phase (1400-1100 BCE) and it is possible that our analysis is entirely reflecting individuals of the inter-phase and late phase.

Recent archaeological fieldwork (2014-2016) and radiocarbon dating from other Swat sites (Saidu Sharif, a cemetery dated to 500-200 BCE, and the multiphase settlement at Barikot) shows that the youngest date of the Swat Protohistoric Graves is around 800 BCE, which can be considered the most recent extent of the specific burial tradition and the associated material culture. The graveyard of Udegram was excavated by M. Vidale, M. Cupitò, R. Micheli, M. Zahir, and L.M. Olivieri (*179*, *180*).

All but one of the Udegram individuals from the Late Bronze-Iron Age cluster tightly with other individuals of this period from nearby sites. We use the analysis label *SPGT* and the split label *Udegram_IA* for these 15 individuals.

- Grave 1, Individual 1 or 2 in a double burial, DA-UDE0317-024 (I6899): Previously reported dates of 1044-830 calBCE (2785±45 BP, CEDAD-LTL13328A (see Tables 3, 10 and pl. XIV of ref. (*180*), Individual 1) and 901-792 calBCE (2659±40 BP, CEDAD

LTL13328A (see Tables 3, 10 and pl. XIV of ref. (*180*), Individual 2). Genetically male. Grave 1 hosts the remains of two infants, whose remains were each packed into a cloth or a cloth bag. The grave was marked by a wooden structure. The grave furnishing includes unusual finds for an infant's grave such as a female terracotta figurine, a copper/bronze ear-ring, and a gold ear-ring (crescent-shaped with embossed dots) (more information on the grave can be obtained from pages 149-153 of ref. (*180*)).

• Grave 3, Individual 1 in a single burial, DA-UDE0317-026 (I6900): Previously reported date of 1400-1126 calBCE (3018±45 BP, CEDAD-LTL13327A) (see Tables 3, 10 and pl. XIV of ref. (*180*)). Genetically male. The grave features an oblong walled pit. Individual 1, a male 40 to 50 years old, was found on his right side facing north in a flexed position. Eight vessels, and the remains of two wooden boxes, were found in primary deposition near the body (more information on the grave can be obtained from pages 96-105 of ref. (*180*)). Individual 1 (I6900) is a first degree relative of Grave 28, Individual 1 (see below).

• Grave 28, Individual 1 in a double burial, UDG2, DA-UDE0317-059 (I3260): Union of non-overlapping dates: a date of 921-831 calBCE (2740±20 BP, PSUAMS-2798) and a previously reported date of 1416-1214 calBCE (3056±40 BP, CEDAD-LTL13332A) (see Table 10, p. 224 of ref. (*180*)). Genetically female. Individual 1 is >40 years old according to anthropological analysis. Grave 28 contains the remains of two individuals 1 and 2, both female (contrary to the initial morphological analysis in ref. (*180*)), respectively on the left and the right side of the grave. Individual 2 is described below. The bones of Individual 1 were rearranged and piled without a precise order, with the long bones gathered and placed above the others, and the cranium in the upper part of the heap, as seen in other graves at Udegram. The crania were rearranged and intentionally set in a tableau, in front of each other, with the maxillae in direct contact (additional information on the grave can be obtained from pages 158-161 of ref. (*180*)). Individual 1 is a first degree relative of Grave 3, Individual 1 (see above).

• Grave 28, Individual 2, UDG31, DA-UDE0317-050 (I8190): Date of 992-830 calBCE (2760±30 BP, Beta-428664 (note that this date was obtained from a different bone sample than the bone sample from which we obtained DNA, but is genetically a duplicate and we dropped the data from the bone sample that we dated due to lower quality). Genetically female. Individual 1 was 30 to 40 years old according to morphological analysis. For the description of the burial features see above (additional information on the grave can be obtained from pages 158-161 of ref. (*180*))

- Grave 4, Individual 1 in a double burial, UDG60, DA-UDE0317-034 (I6198): Context date of 1200-800 BCE. Genetically female. The grave, flanking Grave 5, was partly cut by the creation of Grave 3. Individual 1, 40 to 50 years old, was found in the left side of the grave, on her left side in a flexed position with the skull facing right (to the south) (more information on the grave can be obtained from pages 109-115 of ref. (*180*)). Individual 2 in Grave 4 is described below.

- Grave 4, Individual 2 in a double burial, UDG58, DA-UDE0317-48 (I6197): Context date of 1200-800 BCE. Genetically male. The skeleton of Individual 2, 35 to 45 years old, was not articulated. Instead, the bones were piled up on the right side of the grave, with the cranium on top. It is possible that the bundle of bones was originally contained in a cloth bag that later decayed (more information on the grave can be obtained from pages 109-115 of ref. (*180*)).

- Grave 5, Individual 1 in a double burial, UDG53, DA-UDE0317-05 (I1799): Union of non-overlapping dates: a date of 1044-922 calBCE (2830±20 BP, PSUAMS-2632) and a previously reported date of 1491-1231 calBCE (3098±45 BP, CEDAD-LTL13335A) (see Tables 3, 10 and pl. XIV of ref. (*180*)). Genetically male. Grave 5 is partly walled. Individual 1 was found in a flexed position on the left side of the grave, on the left flank facing south, while the remains of Individual 2 were piled on the right side (see below). This individual is part of a 4-person family (I1799, I1992, I6194, I3262), and is directly detected as father or son of Individual 1 in Grave 26 (I3262), a 2nd to 3rd degree relative of Individual 2 in the same Grave 5 (I6194), and a 2nd to 3rd degree relative of Individual 1 in Grave 29 (I1992).

- Grave 5, Individual 2 in a double burial, UDG45, DA-UDE0317-055 (I6194): Previously reported date of 1376-1041 calBCE (2969± 45 BP, CEDAD LTL14411A, see Tables 3, 10 and pl. XIV of ref. (*180*)). Genetically male. Grave 5 is partly walled. Remains of the defleshed skeleton of Individual 2, skeletally a young male, were piled on the right side of Individual 1 with the skull placed above and near the cranium of Individual 1 (see above; more information on the grave can be obtained from pages 109-115 of ref. (*180*)). The two individuals in Grave 5 (I1799 and I6194) are genetically detected as 2nd to 3rd degree relatives. This individual is part of a 4-person family (I1799, I1992, I6194, I3262), and directly detected as being a 2nd to 3rd degree relative both of Individual 1 in the same Grave 5 (I1799), and Individual 1 in Grave 26 (I3262).

• Grave 26, Individual 1 in a double burial, UDG5, DA-UDE0317-017 (I3262): Date of 976-832 calBCE (2760±25 BP, PSUAMS-2157). Genetically male. The grave contains two individuals. Individual 1, a 30 to 40-year-old male, is the primary burial; it was found on the left side and is a partly connected skeleton with the cranium facing south. The cranium and the lower part of the skeleton apparently were left in their original position, while the upper part of the skeleton was moved to make space for Individual 2, a secondary interment. Individual 2, a secondary interment, was a >55-60-year-old male (contrary to the results of the anthropological autoptic analysis). It was found as a heap of scattered bones with the cranium facing south (more information can be found in pages 164-169 of ref. (*180*)). This individual is part of a 4-person family (I1799, I1992, I6194, I3262), and directly detected as being father or son of Individual 1 in Grave 5 (I1799), a first degree relative of Individual 1 in Grave 29 (I1992), and a 2nd to 3rd degree relative of Individual 2 in the Grave 5 (I6194).

• Grave 6, Individual 1 in a double burial, DA-UDE0317-032 (I6901): Context date of 1200-800 BCE. Genetically female. This partly walled grave, cut by the creation of Grave 3, contains the remains of Individual 1, 25 to 35 years old, on the left side of the grave, in a flexed position on the left side, facing south. Two copper hairpins associated to Individual 1 were found. Individual 2, a male 20 to 30 years old, was a later secondary interment: a bundle of bones defleshed and disconnected, with a partly preserved skull on top of it (more information on the grave can be obtained from pages 105-109 of ref. (*180*)).

• Grave 7, Individual 1 in a single burial, UDG48 (I6195): Date of 1011-909 calBCE (2810±20 BP, PSUAMS-2841). Genetically female. Grave 7 is a rectangular walled pit. Individual 1, 18 to 25 years old, was in semi-flexed position on her left flank facing south. Four vessels were documented in primary position placed near the deceased (more information on the grave can be obtained from pages 67-71 of ref. (*180*)).

• Grave 10, Individual 1 in a double burial, data from UDG34+UDG39, DA-UDE0317-040 (I1985): Combined date of 1192-824 calBCE using the union of a date of 1192-939 calBCE (2880±30 BP, Beta-428667) and a previously reported date of 1001-824 calBCE (2758±40 BP, CEDAD-LTL13334A, see Table 10, p. 224 of ref. (*180*)). Genetically male. Grave 10 hosts two individuals. The remains of Individual 1 were rearranged in a pile of broken bones with the cranium set vertically on top. Individual 2, the primary burial, is described below (more information on the grave can be obtained from pages 135-142 of ref. (*180*)). This individual is either the father or son of Individual 1 in Grave 27 (I3261).

• Grave 10, Individual 2 in a double burial, UDG38 (I1994): Combined date of 1107-804 calBCE using the union of a dates of 1027-848 calBCE (2800±30 BP, Beta-428666) and a previously reported date of 1107-840 calBCE (2808±45 BP, CEDAD LTL13329A, see Table 10, p. 224 of ref. (*180*)). Genetically female. Individual 2 was deposited later as shown by microstratigraphy of the grave and was found on the left side of the grave with her wrist and hands placed below the cranium in a sleeping tableau. The cranium faces south (more information on the grave can be obtained from pages 135-142 of ref. (*180*)).

• Grave 27, Individual in a double burial, DA-UDE0317-020 (I3261): Context date of 1200-800 BCE. Genetically male. Grave 27 contains the remains of two individuals, 1 and 2, respectively to the right and to the left side of the pit. Both skeletons were manipulated. Individual 1, a male 40 to 50 years old according to anthropological analysis, was lateralized to make space for the remains of Individual 2, a female 35 to 45 years old. The furnishing of the grave is particularly rich: ten vessels for Individual 1 and eight vessels for Individual 2 (more information on the grave can be found in pages 176-184 of ref. (*180*)). Individual 1 is father or son of Grave 10, Individual 1 (I1985).

• Grave 12, Individual 1 in a double burial, DA-UDE0317-003 (I6897): Context date of 1200-800 BCE. Genetically female. The grave was initially recorded in Section N of the site. Individual 1 was found on her the left side on the left flank facing south. A bronze hairpin was found attached to her skull (more information on the grave can be obtained from pages 75-83 of ref. (*180*)).

SPGT_o: Swat Protohistoric Grave Type outlier (n=1). This individual is genetically detected as being part of a four-person family whose other individuals are in the main *SPGT* genetic cluster. However, this individual is lower coverage, a mild genetic outlier from the cluster in PCA, and has some evidence of mitochondrial DNA contamination, so we remove them from the *SPGT* analysis label.

• Grave 29, Individual 1 in a double burial, UDG12 (I1992): Date of 1195-978 calBCE (2890±30 BP, Beta-428665). Genetically male. The grave contains two individuals. Individual 1, a young male of 25 to 35 years as inferred by anthropological analysis, was undisturbed, and found to the right in a flexed position facing south (right). Individual 2 was characterized by a patch of whitish and friable small human bones, collected within a basket, as demonstrated by a single clay lump with wickerwork imprint. These bones, as suggested by anthropological analysis, belonged to a child approximately 10 years old (more information on the grave can be obtained from pages 169-176 of ref. (*180*)). This Individual 1 (I1992) is part of a 4-person family (I1992, I6194, I1799, I3262), and is directly detected as being a father or son of Individual 1 in Grave 5 (I16194), a 2nd to 3rd relative of Individual 1 in Grave 5 (I1799), and a first degree relative of Individual 1 in Grave 26 (I3262).

Relatedness summary:
• 2-person family: I3260-I6900 are 1st degree relatives
• 2-person family: I1985-I3261 are 1st degree relatives
• 4-person family: I1799 (Grave 5), I1992 (Grave 29), I6194 (Grave 5), I3262 (Grave 26) are four males from a 4-person family. I3262 is at the center of the family, being a father or son of I1799 and a 1st degree relative of unknown type of I1992 (I1799 and I1992 are 2nd or 3rd degree relatives of each other). In addition, I6194 is a 2nd or 3rd degree relative of both I1799 (an individual from the same Grave 5) and I3262.

These familial relationships reinforce the hypothesis that Udegram was a cemetery linked to a limited group or number of families (see e.g. page 191 (*180*)).

S2.4.1.2 Gogdara, Swat Protohistoric Graves (n=2)

These two individuals both fall into the *SPGT* analysis label and *Gogdara_IA* split label.

These two individuals are from a small graveyard in Gogdara village in the Swat Valley (left bank, near Udegram) dating to 1300/1200-900 calBCE. The graveyard of Gogdara (3 excavated graves; labeled as Gogdara IV) features three phases: the earliest one (Grave A), a later one (Grave B), which corresponds to the inter-phase represented at Udegram (*179*, *180*), and a slightly later one represented by Grave C. Archaeological fieldwork was conducted in 2011. The graveyard of Gogdara IV was excavated by M. Vidale, Noor Agha Noori, Atif Iqbal, and L.M. Olivieri (*179*, *180*).

- G IV Grave B, Individual 1 in a double burial, DA-UDE0317-071 (I8193): Previously reported date of 1372-1027 calBCE (2964±45 BP, CEDAD-LTL12131A, see Table 1, p. 45 of ref. (*180*)). Genetically female. The grave, a secondary burial and partly eroded, was covered in its latest stage by a mound surrounded by circular wooden fencing. It contains the remains of two individuals both lateralized. Individual 1 according to anthropological analysis was probably a male >45 years old. Furnishings include a stemmed bowl, a restricted bowl with foot, and a globular pot (more information on the grave can be obtained from pages 27-47 of ref. (*180*)).

- G IV Grave C, Individual 1 in a double burial, DA-UDE0317-072 (I8194): Context date of 1100-900 BCE. Genetically male. The grave probably belonged to an adult individual, whose remains were too scanty to be thoroughly analyzed. The excavation revealed a sub-rectangular space – an open-air mortuary or decoration room - fenced by wooden logs protected by a small clay wall to prevent erosion. The furnishing was extremely rich including eleven vessels, a polisher, one copper/bronze pin and three spindle-whorls (more information on the grave can be obtained from pages 27-47 of ref. (*180*)).

S2.4.1.3 <u>Loebanr, Swat Protohistoric Graves, Pakistan (n=38)</u>

All but one of the Loebanr individuals cluster tightly with other individuals of the same Late Bronze-Iron Age period from nearby sites. We use the analysis label *SPGT* and the split label *Loebanr_IA* for these 37 individuals.

The excavations at Loebanr were carried out by the IsMEO Italian Archaeological Mission between 1962 and 1965 in a left side valley of Swat (Jambil) (C. Silvi Antonini and G. Stacul) (*181–184*). The graveyard includes 183 excavated graves, with tombs that have a generally rectangular structure and are covered by stone slabs. The axis of the graves is generally oriented along an uphill-downhill gradient, with an orientation that is mainly north/south. Single and double graves are very common; occasional graves with three individuals are present as well. Funerary practices included primary and secondary interments, and cremation and the co-occurrence of cremation with inhumation rituals. Bodies were placed flexed on one side and were often accompanied by disarticulated skeletal remains grouped in one or more clusters of bones. Grave goods included pottery, bone and ivory objects, and metal items (copper and iron), but few personal ornaments and only rarely weapons. Individuals of both sexes and all ages are represented in the excavated graves (*181, 182, 185*, *186*). Specific body treatments were not entirely restricted to particular phases, but varied over time. Grave architecture, burial features and grave furnishing are similar in style to those documented at Udegram (see above).

- Grave 73, Individual 1 in a double burial (I6554): Date of 831-796 calBCE (2645±20 BP, PSUAMS-2796). Genetically male. The grave contains two individuals. Individual 1 is a primary deposition with an articulated and well-preserved skeleton, with its body flexed and resting on the right side (see pages 232-233 of ref. (*181*)). The chronology of the pottery is associated to the middle/late phase of the Swat Protohistoric Graves assemblages. This individual is genetically detected as being part of a 3-person family (I10000-I6292-I6554): he is the son of Individual 1 in Grave 108 (I10000, his father) and also the son of Individual 1 in Grave 135 (I6292, his mother) (see below for descriptions of these other individuals).

- Grave 108, Individual 1 in a double burial (I10000): Context date of 1000-800 BCE. Genetically male. Primary burial, supine with flexed raised legs, and half raised arms. The grave's furnishings include a decorated jar similar to the one documented at Udegram in Grave 28 (see above). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages (see pages 169-171 of ref. (*181*)). Individual 1 is the father [or son] of Grave 73, Individual 1 (see above). This individual is genetically detected as being part of a 3-person family (I10000-I6292-I6554), and is directly detected as being the father of Individual 1 in Grave 73 (I6554).

- Grave 135, Individual 1 in a double burial (I6292): Date of 895-801 calBCE (2675±20 BP, PSUAMS-2794). Genetically female. The grave contains two individuals. Individual 1 is a primary deposition with a skeleton anatomically connected and fairly well preserved; body flexed and resting on the left side (see pages 198-201 of ref. (*181*)). The chronology of the pottery furnishing is associated to the latest phase of the Swat Protohistoric Graves assemblages. Individual 1 is the mother [?] of Grave 73, Individual 1 (see above). This individual is genetically detected as being part of a 3-person family (I10000-I6292-I6554), and is directly detected as being the mother of Individual 1 in Grave 73 (I6554).

• Grave 76, Individual 1 in a double burial (I8997): Context date of 1000-800 BCE. Genetically male. Primary burial, body flexed and resting on the right side. The grave's furnishings include six vessels, copper/bronze hair-pins and ear-ring. The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages (see pages 137-138 of ref. (*181*)). This individual is a brother of Individual 1 in Grave 78 (I8998) (see below).

• Grave 78, single burial (I8998): Context date of 1000-800 BCE. Genetically male. Primary burial, body flexed and resting on the right side. The grave's furnishings are poor and contain three pots. The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages (see pages 139-140 of ref. (*181*)). This individual is a brother of Grave 76, Individual 1 (see above).

• Grave 77, single burial (I6555): Date of 906-820 calBCE (2720±20 BP, PSUAMS-2797). Genetically male. The grave contains a single burial. It is a primary deposition with a skeleton anatomically connected and fairly well-preserved; body flexed and resting on the right side (see pages 138-139 of ref. (*181*)). The chronology of the pottery is associated to the middle/late phase of the Swat Protohistoric Graves assemblages. This individual is genetically detected as being part of a 3-person family (I6555-I12136-I12459), and is directly detected as being the son of Individual 1 in Grave 85 (I12136, his mother), and a 2nd to 3rd degree relative of the single individual in Grave 107 (I12459).

• Grave 85 Individual 1 in a double burial, 545 (I12136): Context date of 1000-800 BCE. Genetically female. Monumental double grave with side chambers and cruciform plan. Remains refer to secondary, possibly manipulated or disturbed burials. Assemblage includes 18 vessels (see pages 146-148 of ref. (*181*)). The chronology of the pottery is associated to the middle/late phase of the Swat Protohistoric Graves assemblages. This individual is genetically detected as being part of a 3-person family (I6555-I12136-I12459), and is directly detected as the mother of the single individual in Grave 77 (I6555, her son).

At the time that our analysis was performed we did not recognize this individual as a first degree relative of I6555 and thus remove this individual as a first degree relative of another in the dataset. However, as described above, inclusion of two closely related individuals in an analysis is not expected to bias results, and at worst will (conservatively) inflate the standard errors of statistics compared to what would be expected if we were representing the *SPGT* with entirely unrelated individuals.

• Grave 107, single burial, 596 (I12459): Context date of 1000-800 BCE. Genetically male. Single burial, with individual resting flexed on the right side and 2 vessels near the hands (see page 169 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages. This individual is genetically detected as being part of a 3-person family (I6555-I12136-I12459), and is directly detected as a 2nd to 3rd degree relative of the single individual in Grave 77 (I6555).

• Grave 9, single burial, 595 (I12458): Context date of 1000-800 BCE. Genetically male. Single burial, individual resting flexed on the right side with few vessels and a mace-head (see pages 67-68 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

• Grave 80, Individual 1 in a double burial (I8999): Context date of 1000-800 BCE. Genetically female. This is a monumental structure with niches, and contains a primary burial, body flexed and resting on the right side. Individual 2 is a secondary deposition of piled-up bones with the cranium lying on the top. The grave’s furnishings are particularly rich and contains nineteen pots, two spindle-whorls, a bone figurine, and a copper/bronze hair-pin. The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages (see pages 141-145 of ref. (*181*)).

• Grave 112, single burial (I10001): Context date of 1000-800 BCE. Genetically male. Primary burial, body flexed and resting on the right side. The chronology of the pottery is associated to the middle/late [uncertain] phase of the Swat Protohistoric Graves assemblages (see pages 174-175 of ref. (*181*)). This individual is genetically detected as the son of I12985 in Grave 105.

• Grave 105, Individual 2 in a double burial, 714 (I12985): Context date of 1000-800 BCE. Genetically female. These are secondary, possibly manipulated or disturbed burials. The grave furnishings include 8 vessels and 2 copper buckles (?) (see pages 166-167 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages. This individual is genetically detected as the mother of I10001 in Grave 112.

• Grave 135, Individual 1 (?) in a double burial, LOEB_53 (I6553): Date of 971-834 calBCE (2755±20 BP, PSUAMS-2795). Genetically male. The archaeological metadata for this individual is not clear. We reconstructed its label but could have made an error in its assignement. Grave 135 is a double burial (see above), and this sample may instead come from Individual 2. However, we include the data from this sample in our analysis because it forms a tight genetic cluster with other *SPGT* analysis label individuals from this site.

• Grave 63 (?), Individual 3 (C) in a triple burial, T.163C (I6556): Date of 894-798 calBCE (2670±20 BP, PSUAMS-2792). Genetically female. Individual I6556 is marked as being from Grave 163, a very small grave found almost empty (see page 229 of ref. (*181*)). Therefore, we reconstructed the provenance of the sample as possibly from Grave 63. Grave 63 contains three skeletons. Primary burial (Individual A = 1) is a disturbed skeleton flexed and resting on left side. Secondary burials (Individual B = 2 and Individual C = 3) were manipulated: two secondary depositions of piled-up bones on the disturbed skeleton of Individual 1 (see pages 123-125 of ref. (*181*)). The chronology of the pottery is associated to the middle/late phase of the Swat Protohistoric Graves assemblages.

• Grave 63 (?), Individual 1 (A), 592 (I12455). Context date of 1000-800 BCE. Genetically female. This individual is marked as being from Grave 61 but the attribution does not appear to be plausible since the archaeological data report no human remains from this grave (see page 123 of ref. (*181*)) and a best guess is that this individual is a member of the triple burial in Grave 63 (I12455). The documentation of Grave 63 (multiple burial with 3 individuals A-B-C) which is summarized above and that we suspect is the source of individual I12455, I6556 (see above), and I10974 (see below) is presented in detail on pages 124-125 of ref. (*181*).

• Grave 63 (?) (I10974): Date of 1006-904 calBCE (2800±20 BP, PSUAMS-4867). Genetically male. Individual I10974 is marked as being from Grave 163, a very small grave found almost empty (see page 229 of ref. (*181*)). Therefore, we reconstructed the provenance of the sample as possibly from Grave 63. The documentation of Grave 63 (multiple burial with 3 individuals A-B-C) is summarized above. This individual is genetically detected as a 2^{nd} to 3^{rd} degree relative of the single burial in Grave 90 (I12987).

• Grave 90, single burial, 716 (I12987): Context date of 1000-800 BCE. Genetically female. Single burial with individual flexed resting on the left side. Assemblage includes 6 vessels, a spindle-whorl, and a copper hair-pin (see pages 151-152 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages. This individual is genetically detected as a 2^{nd} to 3^{rd} degree relative of individual I10974.

• Grave 170, single burial (I5400): Date of 927-831 calBCE (2745±20 BP, PSUAMS-2793). Genetically male. The grave contains a single burial: a primary deposition with a well-preserved articulated skeleton, body flexed and resting on the right side (see pages 233-234 of ref. (*181*)). The chronology of the pottery is associated to the middle/late phase of the Swat Protohistoric Graves assemblages.

• Grave 169, Individual 1 in a double burial, 931 (I13220): Context date of 1000-800 BCE. Genetically female. Double burial. The remains of Individual 1 are scattered towards the left side of the grave, while Individual 2 is partly connected (flexed) resting on the right side (see page 232-233 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

• Grave 176, Individual 2 in a double burial, 932 (I13221): Context date of 1000-800 BCE. Genetically male. The remains in this double burial correspond to secondary, possibly manipulated burials. The remains of Individual 2 were in partial anatomical articulation flexed resting on the right side, while the remains of Individual 1 were piled up towards the central left side of the grave. The assemblage included 2 vessels and 2 copper earrings near the skull of Individual 1 (see page 239 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

- Grave 182, single burial, 933 (I13222): Context date of 1000-800 BCE. Genetically male. Single burial with individual resting (flexed) on the right side. Assemblage includes a vessel, a copper earring and a mace-head (see pages 243-244 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

- Grave 131, Individual 1 in a double burial, 708 (I12979): Context date of 1000-800 BCE. Genetically female. The remains Individual 1 are scattered towards the left side of the grave, while Individual 2 is flexed resting on the right side. The grave assemblage includes 6 vessels (see pages 194-195 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

- Grave 131, Individual 2 in a double burial, 717 (I12988): Context date of 1000-800 BCE. Genetically male. See above for a description of the grave. This individual is genetically detected as a 2nd to 3rd degree relative of both individuals in Grave 70 (I13223 and I13224).

- Grave 70, Individual 1 in a double secondary burial, 935 (I13224): Context date of 1000-800 BCE. Genetically female. See above for a description of the grave. In this double secondary burial, both skeletons are disarticulated. The assemblage includes 12 vessels (see pages 130-131 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages. This individual is genetically detected as a 2nd to 3rd degree relative of Individual 2 in Grave 131 (I12988) and of Individual 2 in Grave 70 (I13233).

- Grave 70, Individual 2 in a double secondary burial, T70B, 934 (I13223): Context date of 1000-800 BCE. Genetically male. See above for a description of the grave. This individual is genetically detected as a 2nd to 3rd degree relative of Individual 2 in Grave 131 (I12988) and of Individual 2 in Grave 70 (I13233).

- Grave 127, Individual 2 in a double burial, 937 (I13226): Date of 1011-909 calBCE (2810±20 BP, PSUAMS-5503). Genetically female. Individual 2 is resting on the left side, while the remains of individual 1 are scattered towards the right side of the grave. The assemblage includes 7 vessels (see pages 189-190 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

- Grave 178, Individual 1 (B) in a double burial, 711 (I12982) Context date of 1000-800 BCE. Genetically male. The skeletal remains are from secondary deposition of possibly manipulated burials. The remains of Individual 1 are piled up toward the left side of the grave, while Individual 2 is in partial anatomical articulation and flexed resting on the left side. The assemblage includes 8 vessels, a mace-head, and 1 copper hair-pin near the skull of Individual 2 (see pages 240-241 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

- Grave 178, Individual 2 (A) in a double burial, 938 (I13227): Context date of 1000-800 BCE. Genetically female. See above for a description of the grave.

- Grave 168, single burial, 939 (I13228): Context date of 1000-800 BCE. Genetically male. Single burial with individual flexed resting on the right side (2 vessels) (see page 232 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

- Grave 136, Individual 1 (B) in a double burial, 543 (I12134): Context date of 1000-800 BCE. Genetically male. The skeletal remains are from secondary deposition of possibly manipulated burials. The remains of Individual 1 are piled up towards the right side of the grave, while Individual 2 is in partial anatomical articulation, flexed and resting on the left side. The assemblage includes 17 vessels, 1 spindle-whorl (inside vessel no. 8), and 2 hair-pins (iron and copper) near the skull of Individual 2 (see pages 201-203 of ref. (*181*)). The chronology of the pottery is associated to the middle/late phase of the Swat Protohistoric Graves assemblages.

- Grave 136, Individual 2 (A) in a double burial, 546 (I12137): Context date of 1000-800 BCE. Genetically female. See above for a description of the grave.

- Grave 60, Individual 1 (A or B?) in a double burial, 593 (I12456): Context date of 1000-800 BCE. Genetically female. Individual 1 (A?) was partly preserved and resting (flexed) on the left side, while the remains of Individual 2 (B?) were piled near the upper right corner of the pit. The assemblage included 13 vessels (see pages 121-122 of ref. (*181*)). The chronology of the pottery is associated to the middle/late phase of the Swat Protohistoric Graves assemblages.

- Grave 65, Individual 2 (A) in a double burial, 594 (I12457): Context date of 1000-800 BCE. Genetically male. Individual 2 was found resting in flexed position on the left side, while the remains of Individual 1 were piled near the upper right corner of the grave. The assemblage includes a few vessels (see page 126 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

- Grave 145, single burial, 709 (I12980): Context date of 1000-800 BCE. Genetically male. This is a single burial with the individual flexed resting on the right side (one vessel) (see page 212 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

- Grave 174, Individual 2 (B) of a double burial, 710 (I12981): Context date of 1000-800 BCE. Genetically female. The skeletal remains are from secondary deposition of possibly manipulated burials. Individual 2 was found in partial anatomical articulation flexed resting on the right side, while the remains of Individual 1 were piled up towards the upper left side of the grave. Assemblage includes 9 vessels, and 4 copper hair-pins near the skull of both individuals (2 for each) (see pages 237-238 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

- Grave 133, Individual 2 (B) in a double burial, 712 (I12983): Context date of 1000-800 BCE. Genetically female. Individual 2 was found resting (flexed) on the right side, while the remains of Individual 1 were found scattered towards the left side of the grave. The assemblage includes 6 vessels (see pages 196-197 of ref. (*181*)). The chronology of the pottery is associated to the middle/late phase of the Swat Protohistoric Graves assemblages.

- Grave 137, single burial, 713 (I12984): Context date of 1000-800 BCE. Genetically female. Remains refer to a secondary, possibly manipulated single burial (with few vessels and 2 copper hair-pins positioned on the skull) (see pages 203-204 of ref. (*181*)). The chronology of the pottery is associated to the earliest phase of the Swat Protohistoric Graves assemblages.

SPGT_o: Swat Protohistoric Grave Type outlier (n=1). This individual is a mild genetic outlier from the *SPGT* cluster in PCA, and hence we give them a separate analysis label.

• Grave 140, single burial, 547 (I12138): Context date of 1000-800 BCE. Genetically female. Single burial with individual resting (flexed) on the left side. The assemblage includes 20 vessels and miniature vessels, beads of semi-precious stones, an iron pin and a spindle-whorl (see pages 207-208 of ref. (*181*)). The chronology of the pottery is associated to the middle/late phase of the Swat Protohistoric Graves assemblages.

Relatedness summary:

• 3-person family: I6554 (Individual 1 in Grave 73) is the son of both I10000 (Individual 1 in Grave 108, his father) and I6292 (Individual 1 in Grave 135, his mother).
• 3-person family: I6555 (Individual 1 in Grave 77) is the son of I12136 (Individual 1 in Grave 85, his mother) and a 2nd to 3rd degree relative of I12459 (Grave 107).
• 3-person family: I12988 (Individual 2 in Grave 131) is a 2nd to 3rd degree relative of both individuals in Grave 70 (I13223 and I13224).
• 2-person family: I8997 (Individual 1 in Grave 76) and I8998 (Grave 78) are brothers.
• 2-person family: I12985 (Individual 2 in Grave 105) is the mother of I10001 (Grave 112).
• 2-person family: I10974 (Grave 63) and I12987 (Grave 90) are 2nd to 3rd degree relatives.

S2.4.1.4 Katelai, Swat Protohistoric Graves, Pakistan (n=33)

All the Katelai individuals cluster tightly with other individuals of the same Late Bronze-Iron Age period from nearby sites. We use the analysis label *SPGT* and the split label *Katelai_IA.*

The excavations at Katelai were carried out by the IsMEO Italian Archaeological Mission between 1962 and 1965 (C. Silvi Antonini, E. Castaldi and G. Stacul). The graveyard of Katelai is located near the village of the same name at the entrance of the Saidu side valley of Swat (left bank). The graveyard includes 241 excavated graves and two horse burials (*181–184*, *186–188*). The qualitative description of the Katelai graves is similar to that at Loebanr (see above).

• Grave 36, single burial (I5396): Date of 904-817 calBCE (2715±20 BP, PSUAMS-2790). Genetically male. The grave contains a single burial. It is a secondary deposition with the skeleton disarticulated and in a disorderly pile near the southeast side (see pages 284-285 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

• Grave 53, single burial (I5397): Date of 968-833 calBCE (2750±20 BP, PSUAMS-2791). Genetically female. The grave contains a single burial. It is a primary deposition with the skeleton anatomically articulated and well-preserved; the body was flexed and resting on the left side (see pages 492-494 of ref. (*187*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

• Grave 141, single burial (I5398): Context date of 1000-800 BCE. Genetically female. The grave contains a single burial. It is a primary deposition with the skeleton anatomically articulated and well-preserved; the body was flexed and resting on the left side (see page 333 of ref. (*181*)). The chronology of the pottery is associated to the earliest phase of the Swat Protohistoric Graves assemblages.

• Grave 191, single burial (I5399): Context date of 1000-800 BCE. Genetically male. The grave contains a single burial. It is a primary deposition with the skeleton anatomically connected and fairly well-preserved; body flexed and resting on the right side (see page 379 of ref. (*181*)).

• Grave 194, Individual 2 (A) of a double burial, 598 (I12461): Context date of 1000-800 BCE. Genetically female. See above for a description of the grave.

• Grave 153, Individual 1 (B) in a double burial (I10523): Context date of 1000-800 BCE. Genetically male. The grave contains a double burial. Individual 2 (A) rested on the left side, flexed, partly above Individual 1 (resting on the right side) (see more information on pages 343-348 of ref. (*181*)). Near the skull of Individual 2 were two hair-pins (copper and bone), and at the foot of both individuals were 10 vessels. The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages. This individual is genetically detected as a 1st degree relative of I12140 (Grave unrecorded).

• Grave unrecorded, 559 (I12140): Context date of 1000-800 BCE. Genetically male. We don't have the archaeological context information for this individual, but we genetically detected them as a 1st degree relative of Individual 1 in Grave 153 (I10523).

• Grave 204, single burial, 610 (I12472): Context date of 1000-800 BCE. Genetically male. Poorly preserved single burial, with individual resting flexed on the right side, with a single vessel (see pages 388-389 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

• Grave 158, Individual 2 (A) in a double burial, 560 (I12141): Context date of 1000-800 BCE. Genetically female. Individual 2 (A) was an articulated skeleton resting on the left side. The remains of Individual 1 (B) were piled up near the upper right corner of the grave. The assemblage includes 6 vessels, an iron hair-pin, spindle-whorls, beads, pendants and an iron arrow-head (see more information on pages 347-348 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

• Grave 90, single burial, 561 (I12142): Context date of 1000-800 BCE. Genetically female. The skeleton, in poor condition, was lying on its left side inside a simple pit (see page 299 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

• Grave 237, Individual 1 (A) in a double burial, 563 (I12144): Context date of 1000-800 BCE. Genetically female. This is a double burial with two individuals resting on their sides and facing each other. The assemblage includes 9 vessels (see pages 415-416 of ref. (*181*)). The chronology of the pottery is associated to the early/middle phase of the Swat Protohistoric Graves assemblages.

• Grave 237, Individual 2 (B) in a double burial, 562 (I12143): Context date of 1000-800 BCE. Genetically male. This is a double burial with two individuals resting on their sides. See above for the grave description.

• Grave 146, Individual 2 (A) in a double burial, 565 (I12146): Context date of 1000-800 BCE. Genetically female. This is a double burial. Individual 2 (A) is documented in anatomical articulation resting on the left side. The remains of Individual 1(B) were piled up with the cranium on top near the upper right corner of the grave. The assemblage includes 8 vessels, a copper hair-pin, and glass and semi-precious stone beads (see more information on pages 337-339 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

• Grave 145, single burial, 566 (I12147): Date of 1207-1014 calBCE (2910±25 BP, PSUAMS-5426). Genetically male. Genetically male. This grave contained poorly preserved remains of a single burial (individual resting on the left side) with few vessels (see more information on page 337 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages. This individual is a 2nd to 3rd degree relative of the individual in Grave 139 (I12462).

• Grave 122, single burial, 568 (I12149): Context date of 1000-800 BCE. Genetically male. Single burial with partly preserved articulated skeleton (flexed) resting on the right side, with 5 vessels (see more information on pages 319-320 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

• Grave 35, single burial, 581 (I12444): Context date of 1000-800 BCE. Genetically female. The grave contains a single burial. It is a primary deposition with a skeleton partially articulated; body flexed and resting on the left side. The assemblage includes 2 vessels (see pages 283-284 of ref. (*181*)). The chronology of the pottery is associated to the earliest phase of the Swat Protohistoric Graves assemblages.

• Grave 236, single burial, 582 (I12445): Context date of 1000-800 BCE. Genetically male. This is a single burial with the individual resting on the right side and few vessels (see page 415 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

• Grave 233, single burial, 583 (I12446): Context date of 1000-800 BCE. Genetically female. This is a single burial with the individual flexed on the left side, with a few vessels and a copper hair-pin (see pages 412-413 of ref. (*181*)). The chronology of the pottery is associated to the earliest phase of the Swat Protohistoric Graves assemblages.

- Grave 79, single burial, 584 (I12447): Context date of 1000-800 BCE. Genetically male. The skeleton isin poor condition (secondary inhumation) was lying on the left side inside a simple pit (see more information on page 507 of ref. (*187*)). The chronology of the pottery is associated to the latest phase of the Swat Protohistoric Graves assemblages.

- Grave 46, single burial, 585 (I12448): Context date of 1000-800 BCE. Genetically female. Child burial with two superimposed chambers. The burial is notable for the presence of a small niche with a carinated bowl on one of the long sides of the upper chamber (see more information on page 489 of ref. (*187*)). The chronology of the pottery is associated to the latest phase of the Swat Protohistoric Graves assemblages.

- Grave 101, Individual 1 (C) in a triple burial, 590 (I12453): Context date of 1000-800 BCE. Genetically male. In this grave with multiple burials, individual 3 (A) was resting on the left side flexed, partly articulated, the remains of Individual 2 (B) were piled up on the lower left side of the pit, and the remains of Individual 1 (C) were on the right side of the grave. The assemblage includes 12 vessels and a copper hair-pin (near the skull of Individual 3) (see more information on pages 305-306 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

- Grave 131, single burial, 591 (I12454): Context date of 1000-800 BCE. Genetically female. This is a single burial with a partly preserved articulated skeleton (flexed) resting on the left side, with 6 vessels (see more information on pages 325-326 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

- Grave 169, single burial, 597 (I12460): Context date of 1000-800 BCE. Genetically female. This is a single burial with the remains of the individual resting on the left side (flexed), with few vessels (see pages 356-357 of ref. (*181*)). The chronology of the pottery is associated to the earliest phase of the Swat Protohistoric Graves assemblages.

- Grave 139, single burial, 599 (I12462): Context date of 1000-800 BCE. Genetically male. Grave 139 (superimposed on Grave 130) was disturbed by the construction of Grave 123. This is a single burial with partially excavated remains of individual resting on the right side (see more information on page 332 of ref. (*181*)). This individual is a 2nd to 3rd degree relative of the individual in Grave 145 (I12147).

• Grave 30, single burial, 600 (I12463): Context date of 1000-800 BCE. Genetically male. The grave contains a single burial. It is a primary deposition with a skeleton anatomically articulated and fairly well-preserved; body flexed and resting on the right side (see pages 279-280 of ref. (*181*)). The assemblage includes 8 vessels. The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

• Grave 159, Individual 1 (A?) in a double burial, 601 (I12464): Context date of 1000-800 BCE. Genetically female. The grave is a simple pit with two secondary burials. Individual 1 is disconnected and piled above Individual 2 (B?), partly articulated and well-preserved and resting on the right side. No assemblage (see page 349 of ref. (*181*)).

• Grave 186, single burial, 603 (I12465): Context date of 1000-800 BCE. Genetically male. This is a disarticulated secondary burial with 7 vessels, a pendant and a copper knife (see page 372 of ref. (*181*)). The chronology of the pottery is associated to the earliest phase of the Swat Protohistoric Graves assemblages.

• Grave 68, Individual 2 (A) in a double burial, 604 (I12466): Context date of 1000-800 BCE. Genetically male. This is a grave with a double burial. According to the archaeological information, the sample taken for ancient DNA comes from the primary burial (a young adult male flexed on the right side). Disconnected bones and skull of a female individual are grouped on the left side of the grave (see more information on page 500 of ref. (*187*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

• Grave 113 (?) in a single burial, 608 (I12470): Context date of 1000-800 BCE. Genetically male. This grave is a box-urn and a few scattered bones (see on page 312 of ref. (*181*)). The chronology of the pottery is associated to the earliest phase of the Swat Protohistoric Graves assemblages.

• Grave 61, single burial, 609 (I12471): Context date of 1000-800 BCE. Genetically male. This is a large monumental grave covered with a huge single slab. On the upper surface, which corresponds to the lower part of the upper chamber, are some vessels. The individual, anatomically disarticulated, is a secondary burial (see more information on pages 495-496 of ref. (*187*)). The chronology of the pottery is associated to the latest phase of the Swat Protohistoric Graves assemblages.

- Grave 187, Individual 1 (B) in a double burial, 611 (I12473): Context date of 1000-800 BCE. Genetically male. In this double burial, Individual 2 is well preserved and flexed on the left side, while the remains of Individual 1 are piled up on the upper right side of the grave. The assemblage includes 19 vessels, spindle-whorls, ornaments and a terracotta animal figurine (see more information on pages 373-375 of ref. (*181*)). The chronology of the pottery is associated to the latest phase of the Swat Protohistoric Graves assemblages.

- Grave 234, single burial, 613 (I12475): Context date of 1000-800 BCE. Genetically male. This is a single burial with the individual resting on the right side and a few vessels (see pages 413-414 of ref. (*181*)). The chronology of the pottery is associated to the middle phase of the Swat Protohistoric Graves assemblages.

- Grave 209, single burial, 615 (I12477): Context date of 1000-800 BCE. Genetically female. Single burial with individual flexed on the left side with cooper hair-pin on the skull (see page 393 of ref. (*181*)).

Relatedness summary:
- 2-person family: I10523 (Individual 1 (B) in Grave 153) and I12140 (Grave unrecorded) are genetically detected as 1st degree relatives.
- 2-person family: I12147 (the individual in Grave 145) and I12462 (the individual in Grave 139) are genetically detected as 2nd to 3rd degree relatives.

S2.4.1.5 Arkotkila (Arkot-kala), Swat Protohistoric Graves, Pakistan (n=1)

This individual is assigned the *SPGT* analysis label and the *Arkotkila_IA* split label.

This is one of the first graves discovered and excavated in Swat (by G. Tucci in 1956; see more details in ref. (*184*, *189*, *190*)). The site is located on the right bank of River Swat uphill in the Kabbal valley (near Aligrama). Grave architecture, burial feature and grave furnishings are consistent with those excavated at Udegram (see above).

- ARKT_1, Grave 1, Individual 1 (I6557): Context date of 1000-800 BCE. Genetically female. The surviving archaeological description gives no information on the number, position and condition of the individual(s). The grave was monumental, almost double in length compared to a normal grave of the same culture (2.50 m) and covered by five large stone slabs. The chronology of the pottery furnishing is associated to the latest phase of the Swat Protohistoric Graves assemblages (more information on the grave can be obtained from pages 104-105 of ref. (*190*)).

S2.4.1.6 Barikot, Swat Protohistoric Graves, Pakistan (n=3)

These three individuals are assigned the *SPGT* analysis label and the *Barikot_IA* split label.

The site of Barikot is located on the left bank of River Swat, to the west of the modern village of the same name and is surrounded to the north by a crescent-shaped hill (*ghwaṇḍai*, acropolis), bathed by the Swat River and to the south by the Kandak and Karakar streams. Excavations at the protohistoric graveyard were initially carried out by G. Tucci, IsMEO Italian Archaeological Mission (between 1956 and 1958) (see more details in ref. (*184*, *189*, *190*)). Since 1978 the site has been further explored (see more details in ref. (*191*, *192*)) and became the major settlement excavation project for the Italian Archaeological Mission (still in progress; see ref. (*193*–*195*)). Recently, the entire chronological sequence of the protohistoric proto-urban phase, of the Early Historic city, as well of the abandonment of the latter, underwent a drastic revision on the basis of stratigraphically-reliable series of radiocarbon dating (see more details in ref. (*196*)). Grave architecture, burial features and grave furnishing are consistent with those excavated at Udegram (see above).

- Grave 2, Individual 1 in a triple burial, BKOT1 (I6545): Date of 921-831 calBCE (2740±20 BP, PSUAMS-2786). Genetically female. Remains of three individuals were found dispersed inside the grave. Individuals 1 and 2 were found disjointed. Individual 3 is described below. Grave 2 is relatively long (2 m), and housed an extremely rich set of funerary objects comprised of 38 vases (more information on the grave can be obtained from pages 113-118 of ref. (*190*)). The chronology of the pottery assemblage is associated to the latest phase of the Swat Protohistoric Graves assemblages.

- Grave 2, Individual 3 in a triple burial, BKOT2 (I6546): Date of 974-836 calBCE (2760±20 BP, PSUAMS-2787). Genetically female. The skeleton of Individual 3, although not articulated, was found in anatomical position. See above for description of Grave 2 and the other two individuals found in this grave.
- Grave 2, Individual 2 in a triple burial, BKOT4 (I6548): Context date of 1000-800 BCE. Genetically male. See above for description of Grave 2 and the other two individuals found in this grave.

S2.4.1.7 Aligrama, Swat Protohistoric Graves, Pakistan (n=3)

Description by Luca M. Olivieri

The site of Aligrama is located on the right bank of the River Swat. The individuals from this site include three from the "Swat Protohistoric" graveyard associated with the protohistoric settlement of Aligrama (I6888, I8246, I8219) and three from a later sector of the settlement excavated by G. Stacul and S. Tusa (*197*, *198*) (possibly dated to 970-550 calBCE 95%, see table 17.2 in ref. (*199*)). Here we present the first three individuals from this site and describe the individuals from the later period below.

Three individuals are assigned the *Aligrama2_IA* analysis label (which is the same as their split label). Although the three individuals are possibly from three of the 35 protohistoric graves of Aligrama excavated in autumn 1981, these individuals are not grouped with the individuals from the other Swat Protohistoric Grave sites that are assigned the *SPGT* analysis label, because we empirically observe that they have less Steppe pastoralist-related ancestry than the *SPGT*. The archeological metadata related to these samples is uncertain. The provenance of the samples is instead reconstructed on archeological grounds.

- Grave 29, DA-ALI0317-001 (I6888): Date of 818-789 calBCE (2620±20 BP, PSUAMS-4423). Genetically female. No further archaeological information is for the time being available from ISMEO archival records. A study of the graveyard is in progress by the excavators of the site.

• Grave 5, Individual 1 (I8246): Context date of 1000-800 BCE. Genetically male. No further archaeological information is for the time being available from ISMEO archival records. A study of the graveyard is in progress by the excavators of the site.

• Grave 2, Individual 1 in a double burial (I8219): Date of 974-836 calBCE (2760±20 BP, PSUAMS-4853). Genetically female. No further archeological information is available at the present time. A study of the graveyard is in progress by the excavators of the site.

S2.4.1.8 Butkara, Swat Protohistoric Graves, Pakistan (n=5)

The individuals we analyzed from the Butkara Protohistoric Period are all assigned to the *SPGT* analysis label and *Butkara_IA* split label.

The excavations on the hill above the village of Butkara (Butkara II) were directed by M. Taddei, IsMEO Italian Archaeological Mission in 1961 in a side valley of Swat (Jambil) (C. Silvi Antonini and G. Stacul) (*181–183*) (see also ref. (*185*), (*200*), (*201*) and (*202*)). The graveyard includes 48 excavated graves, with tombs that have a generally rectangular structure and are covered by stone slabs. Grave architecture, burial features and grave furnishings are similar in style to those documented at Udegram (see above).

• Grave 26, Individual 3 (A) in a triple burial, 586 (I12449): Context date of 1000-800 BCE. Genetically male. The skeletons of Individual 1 (A) and Individual 2 (B) were disarticulated and piled up on the upper left side of the grave. Individual 3 (C) was found in primary deposition. The assemblage includes 7 vessels (see pages 463-464 of ref. (*181*)). The chronology of the pottery is associated to the middle/late phase of the Swat Protohistoric Graves assemblages. This individual is genetically detected as being father or son of individual I12968 in the same grave (see below).

• Grave 26, Individual 3 (C) in a triple burial, 695 (I12968): Context date of 1000-800 BCE. Genetically male. For the archaeological description, see the previous entry (see also pages 463-464 of ref. (*181*)). This individual is genetically detected as being father or son of individual I12449 in the same grave (see above).

• Grave 26, Individual 2 (B) in a triple burial, 930 (I13219): Context date of 1000-800 BCE. Genetically female. For the archaeological description, see above (see also pages 463-464 of ref. (*181*)).

• Grave unrecorded (25?), Individual 1/?, ID 3542, 587 (I12450): Context date of 1000-800 BCE. Genetically male (for Grave 25 see pages 462-463 ref. (*181*)). The chronology of the pottery of Grave 25 is associated to the middle phase of the Swat Protohistoric Graves assemblages.

• Grave 14, single burial, 588 (I12451): Context date of 1000-800 BCE. Genetically male. The grave contains a single burial. It is a primary deposition with a skeleton anatomically articulated and fairly well-preserved; body flexed and resting on the right side (see pages 448-449 of ref. (*181*)). The wall-lined grave (with a square niche on the east side) contained 13 vessels and iron and copper fragments. The chronology of the pottery is associated to the latest phase of the Swat Protohistoric Graves assemblages.

Relatedness summary:

• 2-person family: In Grave 26 Individual I12968 is genetically detected as the father or son of Individual I12449 from the same grave.

S2.4.2 Individuals from South Asia in the early historic and historic periods

S2.4.2.1 <u>Aligrama, Early Historic Period, Pakistan (n=3)</u>

Descriptions by Luca M. Olivieri

These three individuals come from scattered later burials at the site, which have stylistic features that differ from the "Swat Protohistoric Graves" (see more information on pages 306-308, and on footnote 7, page 306 of ref. (*197*) and on footnote 1, page 207 of ref. (*203*) The individuals are assigned to the *Swat_H* analysis label and *Aligrama_H* split label as they differ genetically from the other samples from the site, and cluster instead with the main group of samples from the site of Saidu Sharif (see below). Both the archaeological metadata and the stratigraphic record related to these samples are uncertain. A chronological framework (c. 970-550 BCE) was initially reconstructed on the basis of the old data (see table 17.2 in ref. (*199*)). The provenance of the samples from a set of 5 later graves excavated in 1972 (*197*, *203*)) is reconstructed here on the basis of the available archeological information including the radiocarbon dates generated for this study.

- Grave 1, Individual 1, ALIG_1_4 (I8218): Date of 50 calBCE - 52 calCE (2010±20 BP, PSUAMS-6195). Genetically female. The data from a sample labeled “ALIG_1_4” and a sample “ALIG_4_1” are genetically identical, and hence the are likely to be two bones from the same individual with confusing sample mislabeling and we merged them for analysis. According to the archaeological records, Individual 1 in Grave 1 (the expected individual and grave number associated with ALIG_1_4) can be referring to a later burial dug in room 4 of the building in Trench B (Phase III). A single globular vessel and an iron object tentatively identified as a hair-pin are associated with the deceased. The presence of a pit or covering was not recorded. The bottom of the burial pit cuts the surface of the latest floor of the unit, sealing off the remains (see below) (see more information on page 307 of ref. (*197*) and pages 208-211 of ref. (*203*)).

- Grave 4, Individual 1 (?), ALIG_4_2 (I8245): Date of 356-121 calBCE (2165±20 BP, PSUAMS-6197). Genetically male. The sample possibly derives from Grave 4, a single burial of a morphologically young individual (<10 years old) found in a crouched position, probably disturbed by the final collapse of the building in Trench B (Phase III) (see more information on page 308 of ref. (*197*) and pages 218-220 of ref. (*203*)). The original labeling has been reconstructed as it was done with I8218 (see above) and I8220 (see below).

- Grave 3, Individual 1, ALIG_3_2 (I8220): Date of 756-430 calBCE (2460±20 BP, PSUAMS-6196). Genetically male. The sample possibly derives from Grave 3. The individual (morphologically a young adult) was found in crouched position, surrounded by stones that sealed off the skeleton from room 3 located above in the same building of the previous individual I8245 (Phase III) (see more information on pages 307-308 of ref. (*197*)), and pages 213-217 of ref. (*203*)). The original labeling has been reconstructed as with I8218 and I8245 (see above).

S2.4.2.2 Saidu Sharif Buddhist monastery, early historic graves in the Swat Valley, Pakistan (n=12)

Descriptions by Luca M. Olivieri

These individuals are from one site in the Swat District of the Khyber-Pakhtunkhwa Province of northwestern Pakistan, dated to 400-200 BCE.

The Buddhist site of Saidu Sharif I was the target of a series of excavations made by the IsMEO Italian Archaeological Mission (with interruptions) between 1963 and 1982 and was directed by D. Faccenna with P. Callieri (and F. Noci) (see more details in ref. (*204–206*)). A total of 17 single graves were found under the monastery, one was found under the Stupa Terrace (Grave 13), another one was found in the area further west, outside the protected archaeological area (Grave 00). The site was explored again by ISMEO in 2011-2015, and 26 more graves were documented west of the Stupa Terrace (see more details in ref. (*207*)). A revised chronology, a reconstruction of the extension of the graveyard, as well as a study of the stratigraphy have been recently proposed after a study by the ISMEO Italian Archaeological Mission/ACT-Field School Project (see more details in ref. (*207*)). Grave architecture and burial features differ from the "Swat Protohistoric Graves" (see above).

Swat_H (n=11):

Eleven individuals are assigned to the *Swat_H* analysis label and cluster with 3 individuals dated to a slightly earlier context from the site of Aligrama (see above; see also table 17.2 in ref. (*199*)). These individuals are assigned to the *Saidu_Sharif_H* split label.

- Grave 00, single burial, DA-SIM0317-052 (I7720): Context date of 400-200 BCE. Genetically female. No printed records survive for this grave. However, in the Italian Mission Archives for Saidu Sharif the discovery of an unnumbered grave to the west of the stupa terrace outside the protected archaeological site is recorded and it is a reasonable guess that the remains come from this grave.

- Grave 1, single burial, DA-SIM0317-066 (I7721): Context date of 400-200 BCE. Genetically male. The grave contained the almost complete remains of a 20 to 30-year-old male, supine. Two flat stones found between the legs were possibly fragments of the collapsed *sema* (more information on the grave can be obtained from page 20 of ref. (*205*)).

• Grave 3, single burial, DA-SIM0317-037 (I7719): Context date of 400-200 BCE. Genetically female. Almost complete remains of a 40 to 50 year old morphologically defined as a male (inconsistent with the genetic sex), lying on the right side with the head turned to the right (more information on the grave can be obtained from page 20 of ref. (*205*)).

• Grave 4, single burial, DA-SIM0317-084 (I7723): Context date of 400-200 BCE. Genetically female. Almost complete remains of a young female 17 to 18 years old, lying on her right side with the head turned to her right (more information on the grave can be obtained from page 20 of ref. (*205*)).

• Grave 5, Individual 1 in a double burial, SSI G5 6 (I2954): Previously reported date of 403-360 calBCE (2296±19 BP, CIRCE-DSH-5624, see Table 1 of ref. (*207*)). Genetically male. Disconnected incomplete remains of a male 30 to 40 years old. Individual 1 in this grave consists of disarticulated incomplete remains of a male 30 to 40 years old, while Individual 2 corresponds to the disarticulated incomplete remains of an adult male (adult). Both individuals were found in a rough pit inside the foundation trench of a wall of the Buddhist monastery. They represent a secondary burial of human remains disturbed during the construction phase of the monastery, which has been interpreted as an act of piety of the Buddhist monks (see page 108-109 of ref. (*205*); more information on the grave can be obtained from page 21 of ref. (*205*)).

• Grave 5, Individual 2 in a double burial, DA-SIM0317-029 (I7718): Context date of 400-200 BCE. Genetically male. See ref. (*205*) for the entire description of the grave including information on the burial context of Individual 2.

• Grave 7, single burial, DA-SIM0317-111 (I6893): Date of 361-203 calBCE (2210±20 BP, PSUAMS-3721). Genetically female. Disconnected incomplete remains of a 30 to 40 year old female, laying on the right flank facing left north (more information on the grave can be obtained from page 21 of ref. (*205*)).

• Grave 9, Individual 1, DA-SIM0317-114 and DA-SIM0317-125 (I6894): Context date of 500-300 BCE. Genetically female. This individual consists of the almost complete remains of a female 17 to 19 years old lying on her back (more information on the grave can be obtained from page 21 of ref. (*205*)).

• Grave 11, single burial, DA-SIM0317-005 (I6891): Union of dates of 431-168 calBCE combining a previously published date of 431-381 calBCE (2346±21 BP, CIRCE-DSH-6526, see Table 1 of ref. (*207*), a date of 361-168 calBCE [361-168 calBCE (2180±30 BP, PSUAMS-3984), and a date of 358-181 calBCE (2185±20 BP, PSUAMS-4424)]. Genetically male. Almost complete remains of male 20 to 30 years old, laying on his back. Grave 11 partly cuts into Grave 12 (see Figure 2 of ref. (*207*)) (more information on the grave can be obtained from pages 21-22 of ref. (*205*)).

• Grave 12, single burial, DA-SIM0317-015 (I7717): Previously reported date of 406-353 calBCE (2292±31 BP, CIRCE-DSH-5625, see Table 1 of ref. (*207*). Genetically female. The grave contains the almost complete remains of female <20 years old, lying on her right side. A large flat stone found askew inside the body was possibly the collapsed *sema* (more information on the grave can be obtained from page 24 of ref. (*205*)). Grave 12 is partly superimposed onto Grave 11, described above (see Figure 2 of ref. (*207*)).

• Grave 16, single burial, DA-SIM0317-129 (I6896): Date of 361-171 calBCE [357-171 calBCE (2175±20 BP, PSUAMS-3761), 361-202 calBCE (2205±20 BP, PSUAMS-4425)]. Genetically female. The grave contained the almost complete remains of female 17 to 19 years old, laying on the right side facing left (north) (more information on the grave can be obtained from page 25 of ref. (*205*)).

Saidu_Sharif_H_o (n=1). A genetic outlier was assigned to the *Saidu_Sharif_H_o* analysis label (we use the same name for the split label).

• Grave 2, single burial, DA-SIM0317-077 (I7722): Context date of 400-200 BCE. Genetically male. Almost complete remains of a 40 to 50 year old male, supine with the head turned to his right (more information on the grave can be obtained from page 20 of ref. (*205*)).

S2.4.2.3 Butkara, Early Historic Period, Swat Valley, Pakistan (n=5)
Descriptions by Luca M. Olivieri

The individuals we analyzed from the Butkara Early Historic Period are all assigned to the *Butkara_H* analysis and split label.

These skeletons excavated by M. Taddei are from a site in the Swat District of the Khyber-Pakhtunkhwa Province of northwestern Pakistan near Butkara village. The graves should not be confused with the nearby "Swat Protohistoric" graveyard by the same name (Butkara II, see above) which was the source of the five individuals listed above. The historical period individuals came from three graves found in 1963 at the foot of the hill of Butkara II near the left bank of the Jambil River, a site that is now labeled Butkara IV. Grave 49 was severely damaged, while Grave 51 and Grave 50 are better preserved. The three adjoining structures show monumental features. Graves 51 and 49 consist of two superimposed chambers: a lower ossuary and an upper chamber built above the ground level. Grave 50 (between the others) features only an upper chamber above ground. Grave 50, 51 and 49 (?) are multiple burials. The three burials were already disturbed at the time of their discovery, and no other information is available in the archaeological records; the imperfect documentation explains why there was some initial uncertainty in the correlation of the samples from which we obtained DNA to the records). The graves were attributed to the Historic Period by the excavator (M. Taddei, field notes 1963, reproduced on pages 23-24 of ref. (*208*)) and can be reasonably associated to Indo-Greek or Saka acculturation period.

- Grave 50, Individual 1 in a multiple burial, T.50a (I6550): New date of 41 calBCE - 57 calCE (1990±20 BP, PSUAMS-2789). Genetically male. This individual is genetically detected as a 2nd to 3rd degree relative of individual I6549 in the same grave.

- Grave 50, Individual 4 in a multiple burial, T.50d (I6552): Context date of 200-0 BCE [mother of directly dated I6549 at 167-46 calBCE (2080±20 BP, PSUAMS-2788)]. Genetically female. This individual is genetically detected as the mother of I6549 (her son).

- Grave 50, Individual 2 in a multiple burial, T.50b (I6549): New date of 167-46 calBCE (2080±20 BP, PSUAMS-2788). Genetically male. This individual is genetically detected as the son of I6552 (his mother), and a 2nd to 3rd degree relative of I6550.

• Grave 51, Individual 8 in a multiple burial, T. 51, upper chamber (I6551): Context date of 200 BCE - 100 CE. Genetically female.

• Grave 49, Individual 2 in a multiple burial, T.49b (I6547): Context date of 200 BCE-100 CE. Genetically female.

Relatedness summary:

• 2-person family: In Grave 51 we find a son (T51.b, I6549) and his mother (T.51d, I6552). In addition, the son (T51.b, I6549) is genetically detected as a 2nd to 3rd degree relative of the third individual we obtained data on from this grave (T.51a, I6550).

S2.4.2.4 Barikot: Historic Period Graves in the Swat Valley, Pakistan (n=4)

Descriptions by Luca M. Olivieri

Fieldwork at Barikot (Bazira/Beira in western classical sources) has provided evidence of a human presence starting from the 2nd millennium BCE, with a major urban phase from 500 BCE to 300 CE (see ref. (*193*)) and a post-urban phase from 600-1200 CE. The site of Barikot is considered, together with Sirkap (Taxila), to be the most important excavation of an Early Historic-to-Historic urban settlement in the northwestern part of South Asia.

Barikot_H early Historic Period graves (n=3): the analysis label is the same as the split label.

• BKG12E, Feature 48, DA-BIR0317-021 (I7714): Date of 47 calBCE - 52 calCE (2005±20 BP, PSUAMS-6207). Genetically female. In the area outside the defensive wall of the ancient city, on the upper filling (*172*) of the outer ditch, an isolated skull was found near a pit-well (*184*). The waste material associated with this individual is consistent with the Saka-Parthian structural phase of the urban defense (c. BCE 50-50 CE) (more information from ref. (*195*, *209*)).

• Grave 1002, BKG11, DA-BIR0317-014 (I6889): Date of 1287-1393 calCE (640±20 BP, PSUAMS-6205). Genetically male. In trench BKG11 (the southwest quarters of the ancient city) a rough burial was dug into the site almost a millenium after the abandonment of the town in the 5th-6th century CE. A fragment of a terracotta potter's wheel, a part of a low terracotta three-legged table, and a large stone rotary quern, were originally placed above the grave to mark its location. The grave mound was formed by a scattered pile of stones. The skeleton, supine, was badly damaged under the weight of the mound. The individual was a young male adult (<30 years?). Both femurs show scraping marks, probably made by rodents scavenging the pit (more information on pages 248-251 of ref. (*193*)).

• BKG11W, Feature 46, DA-BIR0317-028 (I6890): Date of 1021-1154 calCE (960±20 BP, PSUAMS-6206). Genetically female. During the excavation of the layers pertaining to post-urban phases of trench BKG 11 (the southwest quarters of the ancient city) a skull was found on an external surface of the abandoned Street 10 of the ancient city; our radiocarbon dating shows that the skull was from a later burial dug centuries after the end of the urban phase. The skull was poorly preserved. Terracotta sherds, bones (probably broken phalanges) and pebbles had fallen into the cavity. One complete and one partial vertebra (presumably an atlas) were found embedded in the soil around the skull. Individual was a young female adult (<25 years) (more information on page 251 of ref. (*193*)).

Swat_Medieval: Late Historic Period grave (n=1)

One individual from Barikot was classified into the analysis label *Swat_Medieval* (split label *Barikot2_Medieval*) based on genetic clustering patterns and archaeological context information. After carrying out population genetic analysis, we obtained a direct date on this individual, which revealed that they lived ~1250-1400 CE, which is much more recent than the range of dates of ~500-150 BCE for the other 12 individuals assigned to the *Swat_H* analysis label (all from the site of Saidu Sharif), and the other individuals from Barikot (~50 BCE - 600 CE).

• BKG308, Feature 139, DA-BIR0317-030 (I7715): Date of 1276-1390 calCE (670±25 BP, PSUAMS-4427). Genetically male. In the later occupation phases of the area of ancient city, some burials were documented in Trench BKG3 (Phase 4) (see ref. (*210*)). Nine single-burial graves were excavated; and their general features allowed to interpret them as graves of an Islamic cemetery. BKG3 Phase 4 corresponds to Macrophase 9 (Medieval-Premodern), which is documented also in Trenches BKG1 and BKG 6-9 (more information from ref. (*210*, *211*), and also at Udegram (Raja Gira Mosque, Islamic graveyard and settlement, see below).

S2.4.2.5 Medieval graves in Udegram, Swat Valley, Pakistan (n=3)
Descriptions by Luca M. Olivieri and Alessandra Bagnera

Fieldwork at the Ghaznavid Mosque of Udegram (second building phase dated to 1048-1049 CE) was carried out under the direction of U. Scerrato by the Italian Archaeological Mission (with interruptions) between 1985 and 1999. Excavation at the Mosque, oratory and nearby settled area and graveyard provided evidence of the earliest Islamic presence in the north of South Asia, dating back to the conquest of the first Ghaznavid rulers (~1000-1180 CE) as well as the subsequent Islamic history of the site (uninterrupted occupation until the late 13^{th} or early 14^{th} century CE; (see pages 49-61 on ref. (*212*)). An introduction to the fieldwork at the Ghaznavid Mosque of Udegram, can also be found in ref. (*212*).

We assigned the *Swat_Medieval* analysis label to the Udegram Medieval individuals along with other Medieval individuals from Singoor, and Parwak. The split label is *Udegram_Medieval_Ghaznavid*.

• Grave 370, single burial, DA-RAJ0317-011 (I7716): Date of 1169-1250 calCE (835±15 BP, PSUAMS-4428). Genetically male. The single-burial grave is located in the northern part of the graveyard in Room XII of the oratory, where the stratified sequence with three layers of Islamic graves shows that Grave 370 probably belonged to a slightly later phase of the graveyard (see page 119 of ref. (*212*)). The pit is lined with stone slabs and was originally covered by larger slabs. Some stones propped the head of the deceased to the right (facing the *qibla*) (see page 129 of ref. (*212*)).

• Grave 233, Individual 1 in a double burial, UD T219 2+3 (I2960): Context date of 1000-1250 CE [son of I2959 at 1037-1207 calCE (905±30 BP, Poz-83508)]. Genetically male. Grave 233 is located in the oratory at the back of the prayer hall of the Mosque, in Room II (with a *mihrab*), near Grave 219 (the tomb of a pious person, buried along with a bronze amulet and a precious rosary, erected near the *mihrab*). Grave 233 is a secondary burial containing the manipulated remains of two individuals: a young male (I2960), and an adult male (I2959). A reasonable guess is that the two individuals were reburied in the mosque because of the venerability of the place. Grave 233 also includes some animal bones (more information on the grave is provided in pages 117, and 113-135 of ref. (*212*). This individual is genetically detected as the father or son of the other individual in the same grave (I2959).

• Grave 233, Individual 2 in a double burial, UD T219 1 (I2959): Date of 1037-1207 calCE (905±30 BP, Poz-83508). Genetically male. For more information on this grave see the previous entry. This individual is genetically detected as the father or son of the other individual in the same grave (I2960).

Relatedness summary:
• 2-person family: In Grave 233, Individual 1 (I2960) and Individual 2 (I2959) have a father-son relationship (the order is unknown from the genetic data).

S2.4.2.6 Parwak, Chitral District, Khyber Pakhtunkhwa Province, Pakistan (n=3)
Description by Luca M. Olivieri, Roberto Micheli, Massimo Vidale and Muhammad Zahir

These three individuals are assigned the *Swat_Medieval* analysis label and the *Parwak_Medieval* split label.

Parwak is located at Latitude 36°16'31.24" N and Longitude 72°25'8.63" E, around 2235 meters above mean sea level, in the Chitral District of Khyber Pakhtunkhwa Province, Pakistan. Ihsan Ali and Muhammad Zahir excavated the site in 2003-2004, under the auspices of the Directorate of Archaeology and Museums, Government of Khyber Pakhtunkhwa (formerly the Northwest Frontier Province or NWFP) of Pakistan. Brian E. Hemphill of University of Alaska, Fairbanks studied the skeletal materials from the site.

The site consisted of three mounds, designated as Mound A, B and C. Mound C was the lowest and Mound A was the highest. The site measures 121 meters from east to west and 85 meters from north to south. Fourteen trenches, primarily focusing on Mound C, were opened on the site and an area of 321 square meters was excavated.

Eleven shallow graves were exposed on Mound C. Stones placed vertically and horizontally along all four sides marked the graves. These stones were placed close to the bodies and were never very deep. The graves included inhumation, cremation and fractional burials. All burials, regardless of type, contained grave goods placed all around the body. Grave goods included bangles, earrings, finger rings, beads, arrowheads, tools, ceramics and stone objects. Inhumation burials were generally oriented from east-to-west, with the head to the west and the face turned either to south, to the north, or to the northwest. Almost all inhumations were single burials, but in Grave 31 the remains of at least four and perhaps as many as six different individuals were recovered. Fractional burials were encountered in Graves 41 and 42. An elaborate structure designated Grave 4 or Cremation Area, was also unearthed.

Due to their proximity to the surface, typically 20 to 50 cm below the surface, the bones were fragile and difficult to expose. Six individuals recovered were identified as male, 1 individual (Grave 31, Burial 1) was identified as female, and 2 individuals (Grave 1, Grave 2/42) were in such fragmentary condition that sex could not be determined. The male recovered from Grave 51 appears to have died at the youngest age (24-28 years), followed by the males recovered from Grave 31, Burials 2 (20-35 years) and 3 (25-37 years). The male recovered from Grave 31, Burial 4 appears to have died at the most advanced age (38-45 years) of those whose age at death could be fixed with any accuracy (for details see references 115 – 120). Based upon radiocarbon dates, the site can be dated from the 8th century CE to the beginning of the 15th century CE (see reference 118).

- Grave 31, Skeleton 2 in a multiple burial, PARWAK 28 (I11562): Context date of 1-1000 CE. Genetically and anthropologically male. Grave 31 is the largest and the most complex grave found at the Parwak site. The grave was excavated in Trench NC II/1, measured 3 x 3 square meters, and was framed by two rows of stones on all four sides of the grave. It, contained the remains of at least 4 or maximum of 6 individuals. Individuals 2 and 3 were largely complete, while individuals 1 and 4 were incomplete. Individuals 2 and 3 were double burial, with one individual placed between the thighs of the other. Grave 31 contained the richest grave goods at Parwak (see reference 116, pages 191-208 for details). Individual 2, analyzed here, was estimated to be 20-35 years at the time of his death (see reference 116, page 195).

- Grave 31 (?), unassigned skeleton probably in a multiple burial, PARWAK 32 (I11565): Context date of 1-1000 CE. Genetically male. The context information for this burial was lost during museum storage, but it is most likely from Grave 31. See the previous entry for a description of Grave 31.

- Grave 52, single burial, PARWAK30 (I1808): Date of 720-895 calCE (1200±30 BP, Beta-428663). Genetically and anthropologically male. Grave 52 was encountered just 15 cm below the present ground surface and was excavated at boundary separating trenches NC II/1 and C II/1. The grave measured 2.2 x 1.1 square meters and was about 30 centimeters deep. The poorly preserved and fragmentary extended inhumated remains of an adult male were oriented from east-to-west and the head was turned to the south.

S2.4.2.7 Shah Mirandeh Graves, Singoor, Chitral District, Khyber Pakhtunkwha Province, Pakistan (n=1)

Description by Muhammad Zahir

This individual is given the *Swat_Medieval* analysis label and *Parwak_Medieval* split label.

The Shah Mirandeh Graves, Singoor, are located at a latitude 35°53’54.78” N and longitude 71°47’45.60” E about six kilometers north of the Chitral Museum, Chitral. Ihsan Ali and Muhammad Zahir excavated the site in 2005 under the auspices of the Directorate of Archaeology and Museums, Government of Khyber Pakhtunkhwa (formerly the Northwest Frontier Province or NWFP) of Pakistan. Brian E. Hemphill of University of Alaska, Fairbanks actively participated and studied the skeletal materials from the site. The site is located among a cluster of nine other protohistoric/historic grave sites around the village Singoor, including the sites of Chakasht and Gankoreneotek, which were excavated extensively from 2007 onward (see references 115 to 122 for details).

A total of 18 skeletons were encountered during excavations, coming from primary inhumations as well as disturbed secondary inhumations. These included 5 males and 3 females; the sex of the remaining 10 individuals could not be determined based on morphological information. Skeletons were found in burial crypts as single inhumations (Graves 1, 21, and 51), as pairs (Grave 22), and as multiple interments (Grave 3). Ages at death range from infancy to mature adults (see reference 122 for details). Based upon the radiocarbon dates, the site was in use from the 4th century BCE to the 15th century CE.

- Grave 51, single burial, SINGOOR20 (I1805): Date of 1485-1650 calCE (310±30 BP, Beta-428668). Genetically and anthropologically male. Grave 51 represents a primary inhumation of an adult individual buried in a supine position oriented with the head to the northeast and the feet to the southwest. The cranium is resting on the left parietal so that the face is turned toward the southeast. The left arm is fully extended at the shoulder and at the elbow so that the arm lies parallel to the torso with the left hand placed along the left side of the hip. The right arm is semi-adducted at the shoulder, placing the elbow at the level of the top of the head. The right elbow is flexed so that the right hand is placed above the head. Both right and left legs are fully extended at the hip. The right leg is also fully extended at the knee. By contrast, the left leg is semi-flexed at the knee and at the ankle so that this leg is slightly drawn upwards and the foot is slightly turned outward. The age of male at death was about 30-45 years (see reference 122, pages 38 – 51).

S3 Sample classification and nomenclature

S3.1 Classification of ancient samples from South and Central Asia

Our analysis dataset comprised of a total of 523 individuals for which we report genome wide data, along with 19 individuals for which we produced additional libraries to improve coverage. For the analysis in the subsequent sections we chose to classify and cluster our individuals genetically in three phases. In the first phase, we removed individuals that had fewer than 16,000 SNPs covered genome wide on the ~1.2 million SNP targets and assigned them the suffix: *LowCov*. In the second phase, we performed kinship analysis and removed all 1st and 2nd degree relatives. In the third phased, we labeled each of the individuals according to the site and archeological period. To ensure that individuals from the same site and time period were genetically homogenous, we ran *qpAdm* to test whether all the individuals from a particular site were consistent with one another with respect to the *RightAll* outgroup set of highly differentiated populations from the pre-Copper Age period or from other continental regions.

Any individuals that appeared to be heterogenous with respect to the rest of the individuals from that site and time period were given the suffix: *o* for outlier. We had large sample sizes from some sites, and in some of these cases observed multiple individuals that were outliers with respect the majority of the other samples from the site. In such cases we chose to label these individuals with numeric suffixes along with the site name, for example, *Gonur1_BA* and *Gonur2_BA,* with the more numerous individuals given the lower numeric suffix. We call the labels obtained at the end of this process the "Split label" in **Table S 1**. Although we report the data, we did not analyze outlier individuals that were the only representatives of their site and time and for which it was not possible to obtain a radiocarbon date, as in the case of such individuals we could not determine if the nature of their outlier genetic ancestry status was due to them being from a different time period than the rest of the individuals. Finally, to analyze larger groups of individuals, we ran *qpAdm* with the *RightAll* outgroup set once again across sites and report the results grouping individuals with similar ancestry as indicated by the "Analysis labels" column in **Table S 1**. We also identified a special set of individuals from the sites of Gonur and Shahr-i-Sokhta that were genetically heterogeneous but that were united in harboring mixtures in varying proportions of South Asian Hunter-Gatherer (*AASI*)-related ancestry and a unique type of West Eurasian-related ancestry with a lower ratio of Anatolian to Iranian farmer-related ancestry than neighboring groups. We call this set of individuals the *Indus Periphery Cline*, and chose to model them both as a pool of individuals of similar ancestry that we called *Indus_Periphery_Pool*, and using a clinal framework as described in Section S4.

In **Supplementary Materials S4.4**, we report qualitative analysis using the "Split labels" and *qpAdm* modelling using "Analysis labels".

In the analysis that follows, other than when examining models with data on present-day people (discussed below), we analyze samples across ~1.15 million SNPs (the subset of the SNPs on the ~1.24 million SNP enrichment reagent that map to chromosomes 1-22). We do not have thresholds on the call rate of these positions as each of the statistics we compute are carried out on the intersection of the set of SNPs covered for each *f*-statistic. We used the allsnps:YES option for our modeling of populations in *qpAdm* that integrates information across *f*-statistics covered by different sets of SNPs (*33*).

S3.2 Classification of 1,789 individuals from 246 groups in South Asia

Our analysis dataset consists of 1,789 present-day individuals from South Asia categorized into 246 groups, genotyped at 597,573 autosomal single nucleotide polymorphisms (SNP) on the Affymetrix Human Origins array (*33*). We obtained the analysis dataset by filtering a dataset previously published in a study focusing on opportunities for disease gene discovery (*10*). We co-analyzed the data with published data from present-day people outside South Asia (1304 individuals from 88 groups) (*9*, *19*, *32*, *33*) and ancient DNA data (1620 individuals; 523 of these are reported in this study; **Table S 1**) (*9*).

We first performed a step in which we removed individual that were outliers relative to others from the same group in principal component analysis (PCA) (*33*). We then filtered out 24 groups from further analysis for the following reasons:

Ten groups known to have atypical ancestry:
(1) *Balochi* (Northwest group with West Eurasian-related ancestry atypical of South Asia)
(2) *Brahui* (Northwest group with West Eurasian-related ancestry atypical of South Asia)
(3) *Burusho* (Northwest group with West Eurasian-related ancestry atypical of South Asia)
(4) *Jew_AP* (Culture suggests West Eurasian-related ancestry atypical of South Asia)

(5) *Jew_Cochin* (Culture suggests West Eurasian-related ancestry atypical of South Asia)
(6) *Siddi_Gujarat* (Documented evidence of sub-Saharan African-related ancestry (*11*))
(7) *Siddi_Karnataka* (Documented evidence of sub-Saharan African-related ancestry (*11*))
(8) *AHG* (Andaman islanders with deeply divergent ancestry atypical of South Asia)
(9) *Nicobarese* (Island group with ancestry known to be atypical for mainland South Asia)
(10) *Shia_Iranian_Hyderabad* (group with recent descent from Iranian migrants)

Seven additional groups were excluded because of evidence of African admixture based on a significant symmetry test statistic f_4*(Mbuti, Karitiana; Palliyar, X),* which evaluates whether a test population *X* shares ancestry with Mbuti from Central Africa. African admixture occurring within the last thousand years is likely to be a confounder for studies of population history in South Asia in the Bronze Age and Iron Age. Here we show the Z-scores for the test statistic for these seven populations:

(11) *Kamboj*	-5.1
(12) *Muslim_Karnataka*	-4.1
(13) *Gujjar*	-4.0
(14) *Sindhi_Pakistan*	-3.9
(15) *Scheduled_caste_TN*	-3.6
(16) *Muthuliar*	-3.1
(17) *Dawoodi*	-2.9

We excluded one group because of inconsistent position of its two individuals in the PCA:
(18) *Relli* (The two individuals from this group do not cluster with each other)

We excluded 16 groups with only a single individual each:
(19-34) *Gadaba, Gounder, Gowli, Ho_Jharkhand, Kissan, Kondh_AP, Malmi, Mudaliar, Tamta, Agamudayar, Devendrakulathan, DevendrakulathanPall, Kamsali, Padmashali, Sonr, Syed*

We projected 1594 individuals from the remaining 214 groups onto a PCA carried out using *French* to represent West Eurasians, *Han* to represent East Eurasians, and *AHG* to represent an isolated Australasian group without close relatedness to mainland South Asians (*33*). We used the results of the PCA to identify four clusters.

The Indian Cline – a gradient of West Eurasian-relatedness evident in most Indian groups (*11*) – appears clearly in the PCA. To do this we carried out a PCA analysis of the present-day individual from South Asia and projected these onto an axes determined by present-day French, Andamanese Hunter Gatherers and Han Chinese. We manually drew two horizontal lines (**Fig S 8**) that bracketed the great majority of individuals within the Indian Cline. Restricting to groups that met this criterion is in line with the analytical philosophy that we take in this study, which is that we are not attempting to model the full complexities of South Asian population history, and instead to identify models that capture some key patterns.

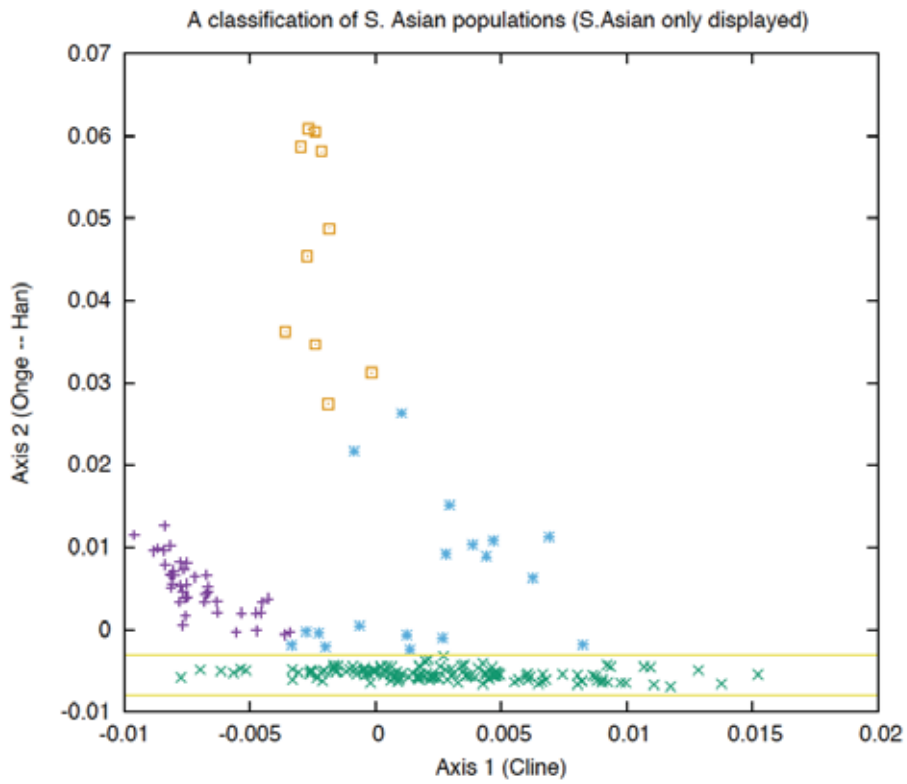

Fig S 8 PCA of Indian cline populations, Axis 1, going left to right, represents increasing West Eurasian (French)-related ancestry. Axis 2 reflects increased East Eurasian-related as represented by excess allele sharing Han Chinese.

S3.2.1 Austroasiatic-Related Cluster (n=41)

In the lower central left of **Fig S 8** there is a cluster of groups that are not on the Indian Cline (we manually colored these groups purple). These are concentrated in a belt in central and eastern India, and many (n=18) such as Munda, Kharia, and Santhal speak languages of the Austroasiatic family. We name the cluster based on the language family that is overrepresented within this group. One possible reason not all of these groups speak

Austroasiatic languages is that some of the ancestors of these groups spoke an Austroasiatic language in the past and have switched to a language of their more numerous neighbors.

S3.2.2 East Asian-Related Cluster (n=10)

In the top left of **Fig S 8** there is a cluster with genetic affinity to Han Chinese (we manually colored these groups with yellow squares). They are situated in the Himalayan region and northeastern India, and geographically proximate to Tibeto-Burman speaking groups outside India, which is unsurprising based on their East Asian-relatedness. We also place the Khasi in this cluster even though they speak an Austroasiatic language because the PCA suggests they have a large East Asian genetic component that is qualitatively different from that in the Austroasiatic-related cluster. The different genetic clustering pattern of the Khasi compared to other Austroasiatic-speaking groups is striking because they also belong to a different Austroasiatic linguistic subfamily (*213*). Taken together, the genetic and linguistic evidence suggest a different demographic history for the Khasi than for the other Austroasiatic speaking groups analyzed in this study, and we will not discuss them further.

S3.2.3 Other groups not included in any analysis cluster (n=21)

A total of 21 groups did not clearly fall within any cluster in the PCA of **Fig S 8** (we manually colored these with blue asterisks). These tended to fall along the geographic fringe of South Asia. Many fall near the Indian Cline but show a shift towards the East Asian-related cluster, with an example being Bengali.

Basu and colleagues (*214*) identify five clusters for South Asia, a result that is in no way contradictory to our findings. One of their clusters corresponds to Andaman Islanders, who we agree are a highly distinctive group. In our study, we do not assign the Andamanese to an analysis cluster, as our analysis focuses on mainland South Asians. Basu and colleagues also chose to define the Indian Cline in terms of admixtures of its two poles – the one enriched in Indo-European speakers and the one enriched in Dravidian speakers – whereas we treat them together.

Table S 5 provides the mean for each group of the x- and y- coordinates of the rotated PCA (**Fig S 8**). It also provides for the 140 Indian Cline groups the principal component of a PCA of those groups by themselves, which is highly correlated to the x-axis of **Fig S 8.**

S4 Population Genetic Characterization of Ancient Individuals

S4.1 Terminology and acronyms

S4.1.1 Time periods

M	Mesolithic
N	Neolithic
EN	Eneolithic
C	Copper Age
BA	Bronze Age
EMBA	Early to Middle Bronze Age
MLBA	Middle to Late Bronze Age
LBA	Late Bronze Age
H	Historical period

S4.1.2 Population labels

WEHG	Western European Hunter Gatherer
EEHG	Eastern European Hunter Gatherer
AHG	Andamanese Hunter Gatherer
WSHG	West Siberian Hunter Gatherer
ESHG	East Siberian Hunter Gatherer

In all the figures and tables in the sections that follow, we present ranges corresponding to ±1 standard errors.

S4.2 Population modeling strategy

In this section, we describe our overall strategy for modeling the admixture history in our set of ancient individuals. Our procedure takes 3 steps in increasing order of complexity, each drawing on an understanding developed in the previous step. Note that while this analysis focuses on individuals for whom we generate data in this study, to make it possible to carry out comparisons of results across studies we also carry out the same analyses on relevant individuals reported in an independent study by Damgaard et al. (*30*), which appeared after the preprint reporting the first version of this analysis was published (*215*).

S4.2.1 First line of analysis – PCA and ADMIXTURE

We examine qualitative differences in ancestry in the ancient individuals by observing the position of the individuals on the West Eurasian and All Eurasian PCA plot and the magnitude and number of major components for each individual on the ADMIXTURE plot. We have chosen the orientation of the West Eurasian and All Eurasian PCA plots to correlate with geography (the genetic patterns mirror geography to an extent), providing an intuitive map of the population structure and to some extent the history.

S4.2.2 Second line of analysis: f_3- and f_4-statistics

Guided by the PCA and ADMIXTURE plots, we compute f_4-statistics (**Methods**) to examine the differences between one ancient population and another with respect to highly divergent pre-Copper Age populations from across Eurasia, and to examine the extent of allele sharing of a single ancient population compared to different pairs of pre-Copper Age populations (using sub-Saharan Africans as an outgroups). These analyses allow us to formalize observations about population transformations based on PCA and ADMIXTURE. As a further test, we compute admixture-f_3 statistics (**Methods**) for each test population with more than two individuals, between all pairs of ancient populations, and comparing to previously published individuals we use as reference populations for this study.

S4.2.3 Third line of analysis - admixture modeling with *qpAdm*

We model the ancestry of the populations in East Eurasia and the various admixture events in their history using the *qpWave* (*216*) and *qpAdm* (*7*) methods, which test formally whether a specified set of *Left* populations can be modeled as a mixture of *N* source populations related differentially to a set of *Right* populations. The methods provide a single chi-squared p-value for fit, appropriately correcting for correlation in ancestry among the analyzed populations. If a model passes, *qpAdm* is able to model a chosen *Left* population as a mixture in different proportions of groups related to the other *Left* populations and to give point estimates and standard errors for mixture proportions (https://github.com/DReichLab/AdmixTools).

S4.2.4 "Distal" modeling of each group from the Copper Age period and onward

Our diverse analyses suggested that populations from the Bronze Age and later can often only be modeled as derived from a minimum of 4 sources differentially related to pre-Copper Age populations. To deal with this complexity, we used a modeling framework in *qpAdm* that differs slightly from approaches used in the past (*7, 9, 217*). The most exhaustive way to model later populations would be to test every possible combination of sources (*Left* populations) and outgroups *(Right* populations), but the number of possible combinations renders this approach impractical. To strike a balance between testing all sources and practicality, we attempt multiple rounds of modeling, with sources that are stratified by time.

In the first round we attempt to model each population as descended from "distal" (pre-Copper Age) sources. Due to the lack of ancient DNA we use two modern populations, *AHG* (Andamanese Hunter Gatherers) to represent groups deeply related to Ancient Ancestral South Indians (*AASI*), a group that we also sometimes call South Asian Hunter-Gatherers. We adopt an approach first used in ref. (*9*). We first define a set of outgroups that help us to parse the ancestry seen in populations from the Copper Age and afterward. These include hunter-gatherers or early farmers from Siberia, the Caucasus and the Near East. We supplement this outgroup set by diverse individuals who are less plausible as direct ancestors for Central and South Asians, but provide leverage for our modeling as they include populations related in different ways to the individuals being studied. This set includes an ancient African (a ~4500-year-old hunter-gatherer from Ethiopia), present-day indigenous Andaman Hunter-Gatherers *(AHG)*, Han Chinese, and Upper Paleolithic Europeans:

RightDistal: Ethiopia_4500BP.SG, Ust_Ishim_HG_published.DG, Tianyuan, ANE, Raqefet_M_Natufian, Villabruna, Belt_Cave_M, AHG, EEHG, WSHG, PPN, WEHG, Anatolia_N, Ganj_Dareh_N, ESHG

To find plausible "Distal" models with a prespecified number of N source populations for each *Test* population, we move all possible subsets of N populations that are not Upper

Paleolithic individuals (in Italics in *RightDistal*) from the *RightDistal* set into the *Left* set, set the *Right* set to be all *RightDistal* populations except for the chosen N, and use *qpAdm* to test whether the *Left* populations are rank=N-1 relative to the *Right* ones (that is, the *Left* populations can be modeled as deriving from N source populations that share alleles in different ways with the *Right* populations). If the *qpAdm* model passes (at a p>0.01 level), we compute mixture proportions with standard errors, and treat models as feasible only if their point estimates and two standard errors of mixture proportions are fully within the interval [0,1]. We use the *allsnps: YES* option of *qpAdm* that allows us to utilize the complete set of SNPs for each f_4 statistic. In our descriptions of working models for each population, we only show models that meet these criteria, an approach that carries with it an implicit assumption that the *Left* set of populations contains within it groups that are descended directly without mixture from the same ancestral populations as the sources of ancestry in the *Test* population.

S4.3 "Proximal" modeling of each group from the Copper Age and onward

After obtaining the full set of working Distal models for each population, we next attempted "Proximal" modeling of the ancestry of each *Test* population from sources close both geographically and in time to it. This also allows us to test specific hypotheses about the establishment of the ancestry of the many Bronze and Iron Age individuals we analyze here.

A first challenge to proximal modeling is uneven sampling both in space and time. We have particular gaps in Kazakhstan prior to the Bronze Age and South Asia prior to the Iron Age.

A second challenge is that by the time of the Copper Age and Bronze Ages, it is clear from multiple lines of analyses (discussed below) that the ancient individuals in our study were the products of extensive mixture from groups related differentially to those from *RightDistal*. This creates a situation where many potential proximal source populations that we are testing contain the necessary ancestry ingredients for the *Test* populations we are attempting to model, so there are many solutions that are a statistical fit to the data.

Thirdly, the allele sharing statistics (f_4-statistics) that drive *qpAdm* are sensitive not just to the types of ancestry but also to their proportions. The *Test* population could reside on a cline, with an ancestry proportion intermediate between two potential source populations we are testing but not matching either of them.

Fourth, with the large number of populations we have available, estimating the covariance matrix of *f*-statistics of a large number of populations (>25) becomes difficult, and we can no longer utilize the procedure we used in our Distal modeling.

To practically address these issues, we carried out Proximal modeling in two stages. In the first, we used an outgroup (*Right*) set *RightAll*. This includes pre-Copper Age populations from both West Eurasia and East Eurasia, thus making our test sensitive to the differences in distal ancestry we observe in source populations from the Copper Age and Bronze Ages.

RightAll: Ethiopia_4500BP.SG, EEHG, WSHG, PPN, WEHG, Anatolia_N, Ganj_Dareh_N, ESHG

As sources (*Left*), we focused on individuals that were of archeological, geographic, temporal and genetic relevance and restricted to sites for which we had at least 2 individuals with at least 100,000 SNPs each to ensure that we had sufficient quality data (**Table S 7**). In situations where there were many sites with similar ancestry as reflected by the PCA, we chose a representative site with the most data. For each populations we report from Central and South Asia, we tested possible models of interactions of up to a maximum of 3 sources.

Forest / Steppe	Europe	Caucasus	Iran / Turan	East Asia	South Asia
Khvalynsk_EN	*Romania_C*	*Armenia_C*	*Seh_Gabi_C*	*Shamanka_EBA.SG*	*SPGT*
Western_Steppe_EMBA	*Baden_LCA*	*Armenia_EBA*	*Tepe_Hissar_C*	*Okunevo_BA.SG*	*Indus_Periphery_Pool*
Central_Steppe_EMBA	*Globular_Amphora*	*Armenia_MBA*	*Parkhai_Anau_EN*		*Saidu_Sharif_o*
Western_Steppe_MLBA	*Tripolye*		*Tepe_Anau_EN*		
Central_Steppe_MLBA			*Sarazm_EN*		
Steppe_MLBA_oBMAC			*Hajji_Firuz_C*		
Steppe_LBA			*BMAC*		
			Namazgha_CA.SG		
			Geoksyur_EN		
			Shahr_I_Sokhta_BA1		
			Aigyrzhal_BA		
			Bustan_EN		

Table S 7 Populations used as Proximal sources for ancestry in our *Test* populations

For the second stage of analysis, we only considered passing models from the first stage (p>0.01) that had confidently non-zero mixture proportions for all proximal sources (95%

confidence intervals wholly contained in the interval [0, 1]). There were often many passing models, and so we performed a "model competition" experiment with pairs of passing models. Specifically, for proximal sources that were used in one model and not the other, we added these additional sources to the right populations and reran *qpAdm* to determine if some models could be rejected. Our approach resembles that of (*9*) in that it begins with a filtering step using a minimal set of outgroups to prune out models that are strongly rejected, and then adds additional outgroups one at a time until one or a few "winners" remain. In contrast to (*9*) we make this choice automatically by choosing as outgroups populations that are sources in competing models. Another fully automated approach as in (*217*) extends the approach in (*9*) to form the union of a set that includes all populations of interest. A concern with this approach is that the use of many correlated populations in the outgroups (in terms of the vectors of *f*-statistics that they produce) increases the number of degrees of freedom in the calculation of significance and therefore may be underpowered to reject certain models.

The principle behind the model competition experiment is two-fold. First, there might be some *Test* populations that have higher quality data and sufficient ancestry from a pre-Copper Age source to allow us to reject models that passed the first stage. Second, if one of the two source populations shares more drift with the test population and is thereby a more proximal source than the other, after the model competition experiment only this model will pass. To minimize situations where the model competition experiment rejects both models—because *Test* populations harbor ancestry that is a mixture of sources related to both proximal sources that are evaluated—we restricted the set of competing models to only allow *Right* groups from a set of populations earlier in time than the *Test* population.

There are several technical considerations raised by the model competition procedure:

(i) Every model competition test is likely to produce subtly different p-values because the estimated f_4 matrices and covariance matrices change with the addition of a *Right* population. As the number of possible combinations of tests is large, we report the p-value and standard errors with respect to the original pre-Copper Age outgroups.

(ii) We were aware that due to the large number of tests we carried out, we might find models that pass simply as a statistical fluctuation even if the truth is that the models do not fit (the problem of multiple hypothesis testing). However, it is important to recognize that due to shared drift between populations, many of our tested models for admixture proportions are not independent, and so the number of tested models is not effectively as large as it seems. In addition, multiple hypothesis testing is not as problematic for our procedure as it might at first seem because our process is a test of acceptance of a model and not rejection, and thus increasing the number of tests should decrease the rate of observing feasible working models at a fixed p-value threshold. We use a p-value at 0.01 as acceptance for the interpretation of working models, a level that is far more conservative than what we would obtain by Bonferroni correction (1×10^{-8})

(iii) As a final caveat, we note that the final proximal models are unlikely to reflect exactly the movements and admixture of populations that have occurred, but are meant to find working solutions that identify the type of genetic sources that could plausibly have participated in the mixtures. In fact, we can be sure that in many instances we do not have access to exactly the right source populations. For example, in East Asia and South Asia, regions where we lack ancient DNA data, we use present-day groups as proxies for ancient populations. We view these not as the true sources but as ones that harbor ancestry that is plausibly derived from an ancestral population shared with the true sources. However it is worth commenting that if the true admixing population *A* has a modern population *A'* that is a descendant of *A* without further admixture into *A'*, then if we use *A'* instead of *A* in *qpAdm* (or *qpWave*) there is no technical concern, as the f_4-statistics used are insensitive to drift from *A* to *A'*.

In what follows, we apply this approach to the ancient individuals from Central and South Asia. In certain situations, we contend with situations where we do not have a model that passes our criteria that we established above. In such cases, we report the model with the highest p-value amongst the ones that were tested.

S4.4 Population modeling

S4.4.1 Iran and Turan

S4.4.1.1 Admixture and movement in the Neolithic and Copper Age Periods

We begin our analysis by examining individuals from major agricultural settlements in the Neolithic and Copper Age periods in Iran and regions in the northeast that have sometimes been called "Turan" due to their evidence of ancient cultural contact with Iran (Turkmenistan, Tajikistan, Uzbekistan and Kyrgyzstan) (**Fig S 9**). We use "Iranian" here not in the linguistic sense, as the languages spoken by the Neolithic and Copper Age periods are not recorded, and of course not in the political sense either as the political borders of today did not exist in prehistory. Instead, we are using it to refer to individuals from the Iranian plateau. We also analyze in this section individuals from Neolithic and Copper Age Anatolia to understand the directionality and timing of gene flow across these two regions.

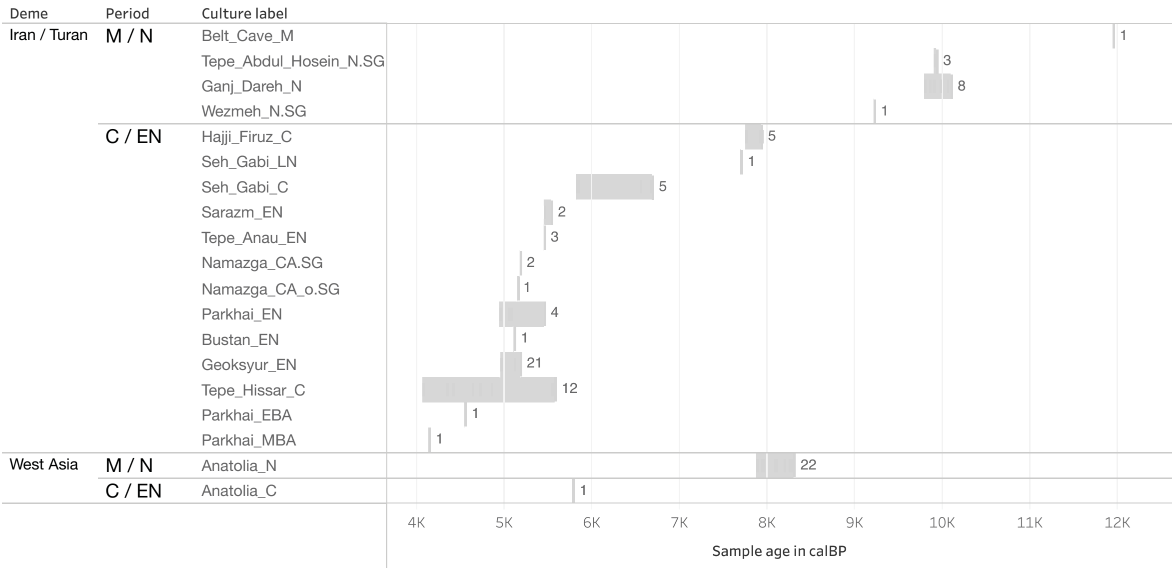

Fig S 9 The number of and age distribution of individuals from sites analyzed in this section. Dates for carbon dated individuals are shown or alternatively we show the archeological context date if no radiocarbon dates are available. Here and in what follows, when we plot each individual we show the midpoint of its associated uncertainty interval (thus, when we are plotting a single individual it is one number and not a range).

We first examined our set of individuals from this period in relation to ancient DNA individuals from diverse pre-Copper Age groups from Eurasia.

We begin by observing that the Neolithic and Copper Age individuals from Iran and Turan occupy the extreme bottom right of the West Eurasian PCA and extreme bottom left of the All Eurasian PCA, consistent with their geographic position with regard to ancient populations of Europe and East Asia (**Fig S 10**, top panels).

Zooming in on each of these plots (**Fig S 10**, second panel), we note that the individuals with the most extreme position on the bottom right of the West Eurasian PCA plot and the bottom left of the All Eurasian PCA plot are all from early Neolithic sites: *Ganj_Dareh_N*, *Tepe_Abdul_Hossien_N.SG*, and *Wezmeh_N.SG* (*31*). These individuals also have only a single component of ancestry as determined by ADMIXTURE and reflect populations from the Neolithic period in Iran that are highly differentiated from other populations in Eurasia (**Fig S 10**, bottom panel) and do not have evidence of admixture with respect to other populations analyzed in this study (via admixture-f_3 statistics). Of particular note is a hunter-gatherer period individual from Belt Cave in Central Iran, who has similar ancestry to the early farmers of the Zagros mountains. These findings show that this type of ancestry extended East of the Zagros mountains at least as anciently as the Mesolithic period.

Following the time of the early Neolithic, individuals from the late Neolithic and Copper Age are shifted away from those of the earlier Neolithic individuals in two directions on the PCA: toward farmers from Neolithic Anatolia and toward HGs from West Siberia (**Fig S 10**, top and second panel). The ADMIXTURE plots confirm that the primary ancestry seen in these later individuals resembles that of Neolithic Iranian individuals, but also contains two additional distinct components in teal and orange, which are maximized in farmers from Neolithic Anatolia and *WSHG*s, respectively. Whilst we cannot interpret the fraction of each of these components in ADMIXTURE as accurate ancestry proportions, we note that there is a west-to-east decrease in the proportion of the Neolithic Anatolian farmer-associated component and a corresponding increase in the proportion of the *WSHG*-associated component, observations that we confirm with formal statistical modeling in the next sub-section. We also observe evidence of bi-directional gene flow, with Iranian farmer-related ancestry appearing in later groups in Anatolia just as Anatolian farmer-related ancestry appears in later groups in Iran, as shown by the shifts in ancestry toward Iranian farmers on the PCA as well as in the ADMIXTURE plots.

We next tested the differences among these populations using *f*-statistics (**Fig S 11).** We use the term “Pre-Copper Age affinity” f_4-statistics” to refer to f_4-statistics that compare each *Test* population to pairs of pre-Copper Age populations. The patterns observed with these statistics confirm our findings from ADMIXTURE and PCA that the Copper Age populations are significantly different in ancestry from the Neolithic ones. Populations from the east (ordered east-to-west as: *Tepe_Hissar_C*, *Parkhai_EN*, *Tepe_Anau_EN*, *Geoksyur_EN*, *Bustan_EN*, *Sarazm_EN*) have significantly higher proportions of ancestry related to *WSHG* and lower proportions of ancestry related to Anatolian farmers compared with those from the west, with significant differences between each neighboring pair.

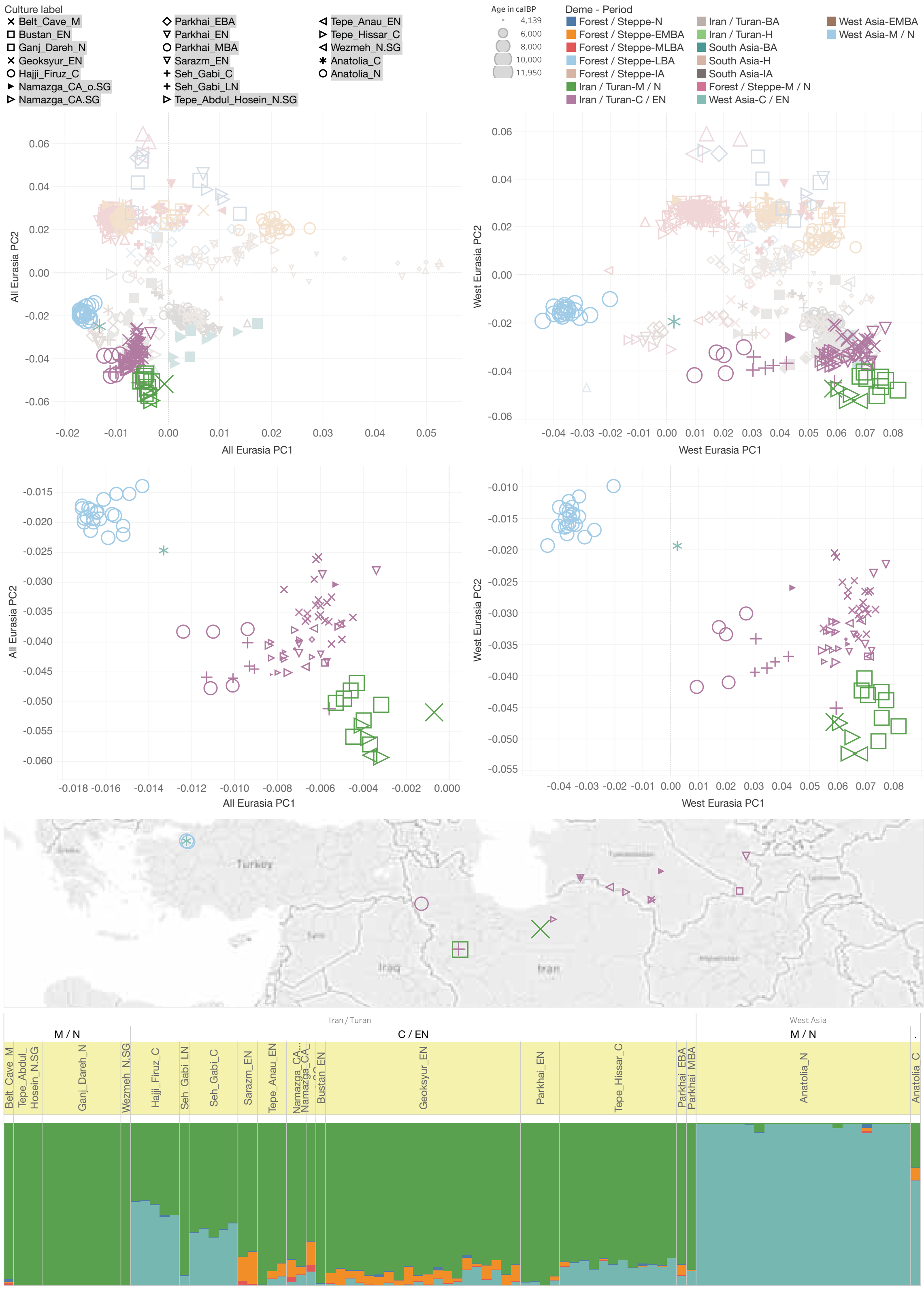

Fig S 10 Top panel: left, All Eurasian PCA; right: West Eurasian PCA. Second panel: left, blow up of All Eurasian PCA; right, blow up of West Eurasian PCA. Third panel: geographic locations of individuals. Fourth panel: ADMIXTURE results, with the colors of green, teal, orange, red and blue representing ancestry maximized in Iranian farmers from the Zagros, farmers from Anatolia, and HGs from West Siberia, South Asia, and Western Europe respectively.

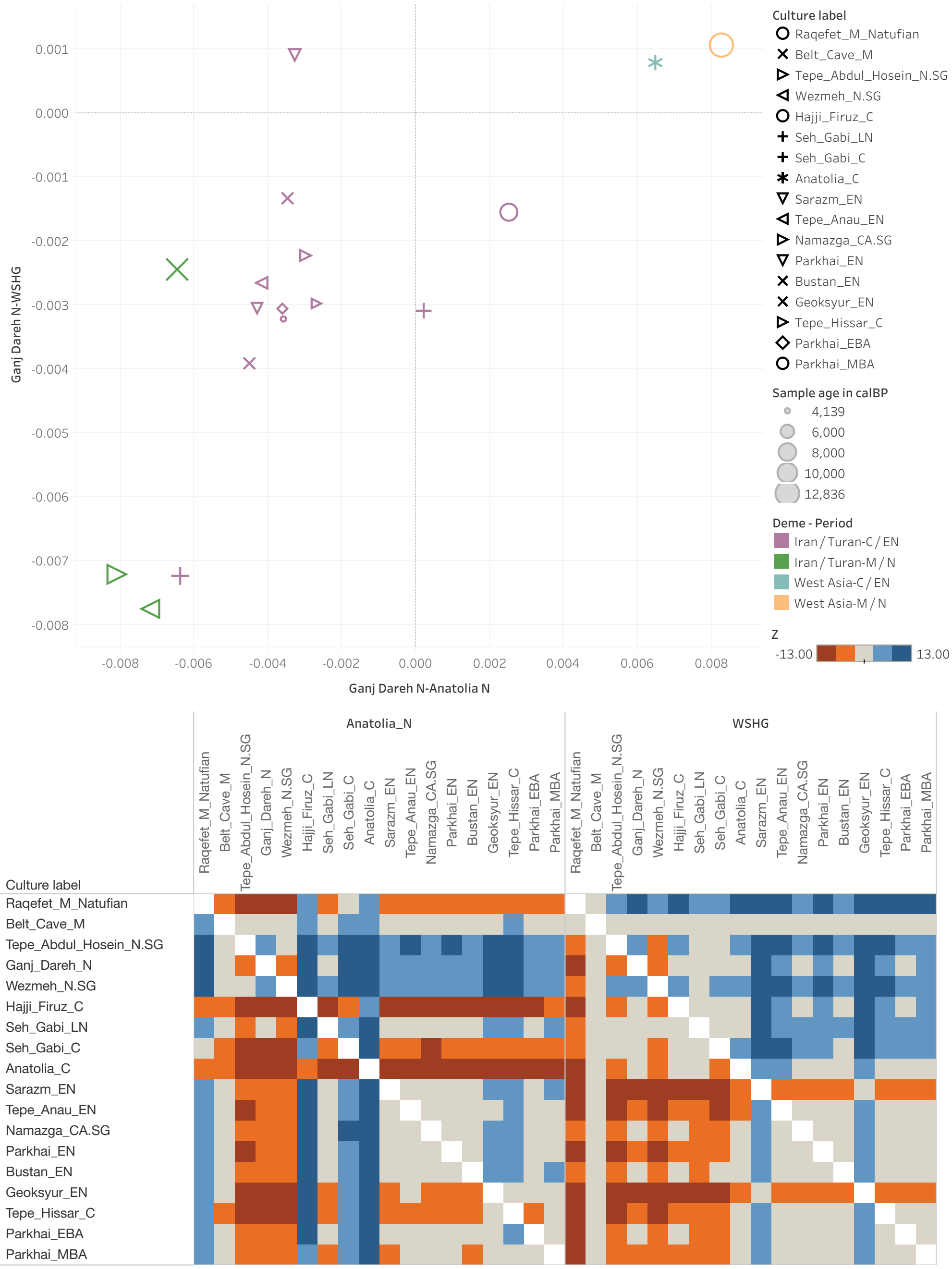

Fig S 11 Top panel: "Pre-Copper Age affinity" f_4-statistics of the form *f_4(Ethiopia_4500BP.SG, Test, Ganj_Dareh_N, Anatolia_N)* against *f_4(Ethiopia_4500BP.SG, Test, Ganj_Dareh_N, WSHG)* for all populations analyzed in this section. Bottom panel: "Two-population comparison" f_4-statistics of the form, *f_4(Ethiopia_4500BP.SG, Anatolia_N, Test1, Test2)* and *f_4(Ethiopia_4500BP.SG, WSHG, Test1, Test2)* for all *Test1* and *Test2* in the set of individuals we analyzed in this section,

showing significant differences in Anatolian farmer-related and West Siberian Hunter-Gatherer-related ancestry between many pairs of populations. Orange squares indicate |Z-scores| > 3 for sharing more alleles with the Test population (*Anatolia_N*, or *WSHG*) compared to a population from the columns, with respect to an outgroup. Blue squares indicate significantly less sharing at the same threshold. Darker orange and blue squares show sharing at |Z-score| > 6.

Finally, we show admixture-f_3 statistics from all possible ancient pairs of populations that are significant at Z < -3. These statistics confirm a broad pattern of admixture from groups related to those in the north and the west after the Neolithic (**Fig S 12)**. We are unable to compute this statistic for populations with only a single individual and therefore do not report these for the late Neolithic individual from Seh Gabi. In cases where there are many source pairs or populations with similar ancestries that contribute to significant tests, we list only the 30 most significant signals.

Our first observation is that there no pairs of populations in our dataset that show significant admixture-f_3 statistics for the two Neolithic populations from Ganj Dareh and Tepe Abdul Hosein, consistent with the hypothesis that these populations are unadmixed with respect to the set of sources available to us (alternatively, these groups could have substantial population-specific drift that masks signals of mixture by this test). This finding is consistent with their position at extremes of the PCA space.

Second, the later individuals from western Iran from the sites of Seh Gabi and Hajji Firuz show significant admixture-f_3 statistics that are maximized when the two sources are a population from Neolithic/Copper Age Iran and Anatolian farmers (**Fig S 12**). From eastern Iran, individuals from Tepe Hissar show significant statistics with the same populations but also with Bronze Age populations that harbor both Anatolian farmer-related and *EEHG*-related ancestry such as Bell Beakers and Corded Ware, consistent with admixture related to hunter-gatherers of West Siberia (**Fig S 13**). Further east for *Tepe_Anau_EN*, *Geoksyur_EN* and *Sarazm_EN*, the top statistics reflect a mixture of both Anatolian farmer- and West Siberian HG-related ancestry (**Fig S 14 - Fig S 16**).

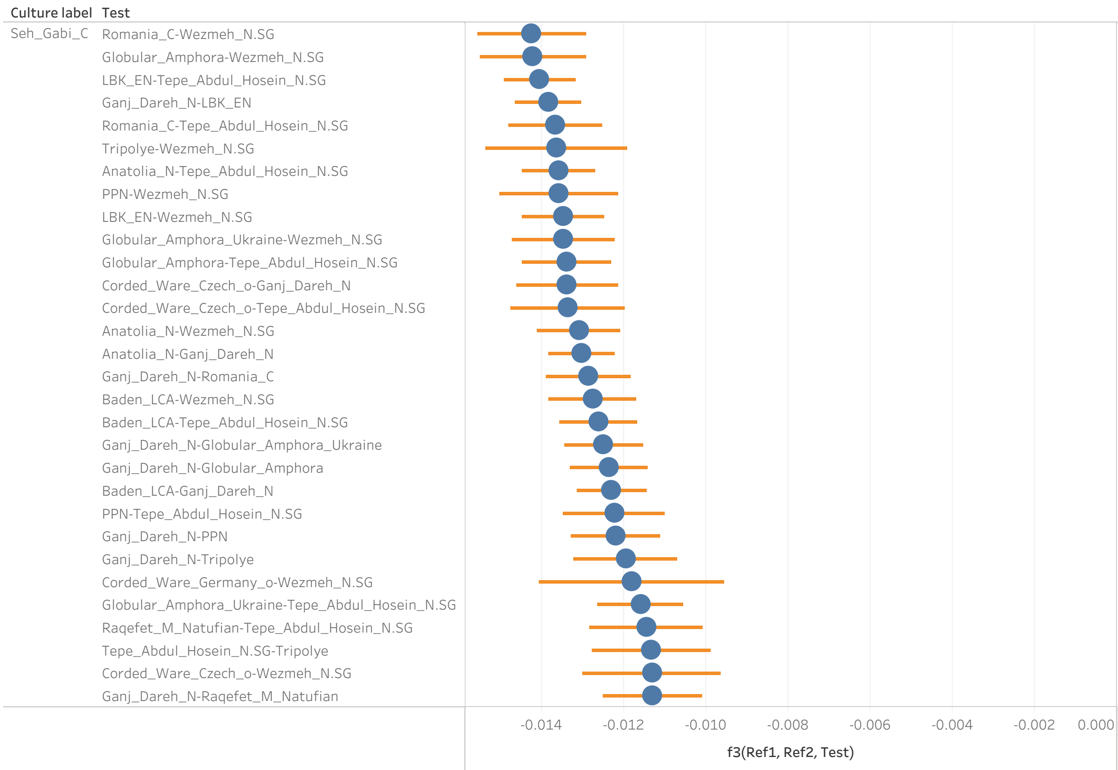

Fig S 12 Top 30 admixture-f_3 pairs for *Seh_Gabi_C*

Fig S 13 Top 30 admixture-f_3 pairs for *Tepe_Hissar_C*

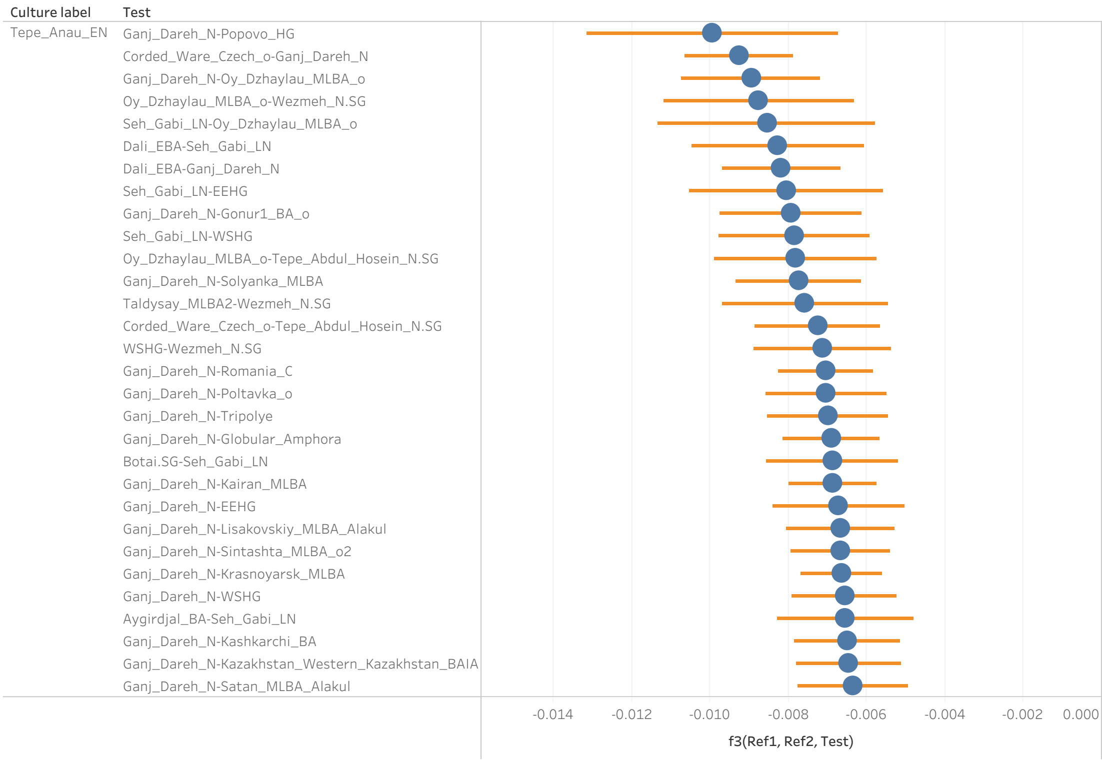

Fig S 14 Top admixture-f_3 pairs for *Tepe_Anau_EN*

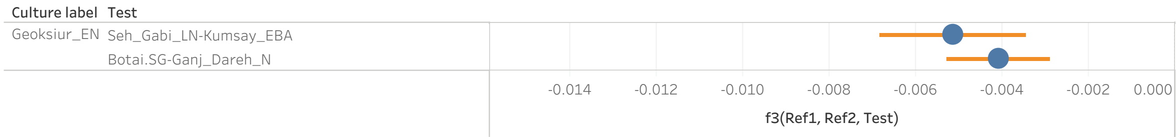

Fig S 15 Top admixture-f_3 pairs for *Geoksyur_EN*

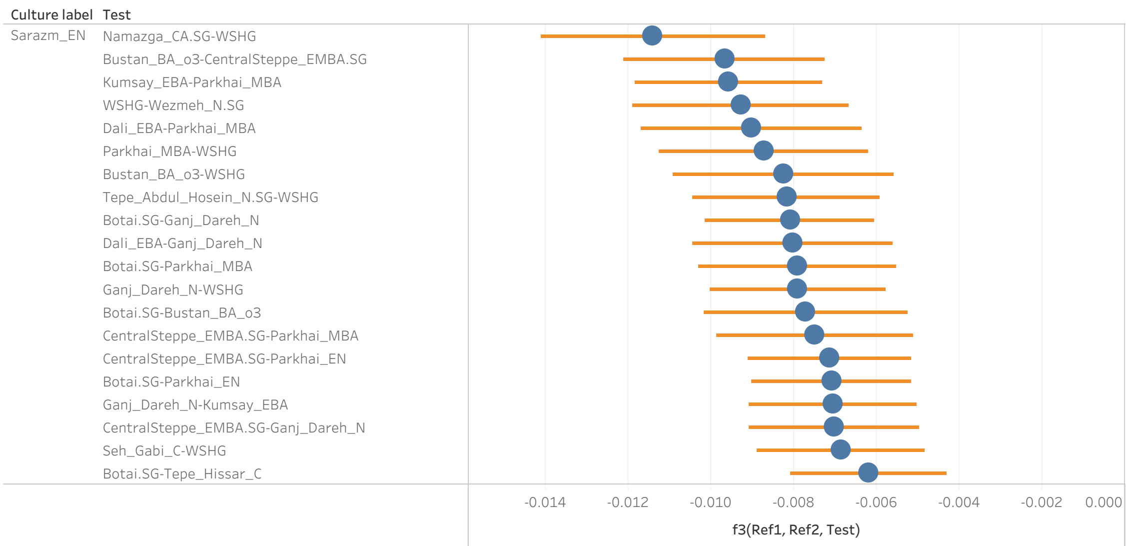

Fig S 16 Top admixture-f_3 pairs for *Sarazm_EN*

In summary, our analysis from the PCAs, ADMIXTURE plots, f_4-statistics, and admixture-f_3 statistics paint a consistent picture of admixture from a population related to farmers from Anatolia, and additional admixture in the eastern set of individuals from a population related to HGs from West Siberia.

In light of these findings, we attempted to model the ancestry of these populations using *qpAdm*.

We show the working models for our distal sources given the thresholds discussed in the previous subsection in **Table S 10 - Table S 27**. The fitting models recapitulate our observations from the analysis described above. The populations from western Iran from Seh Gabi and Hajji Firuz, can be modeled parsimoniously as a mixture of just 2 sources: a population related to that from Neolithic Iran and a population related to the farmers from Anatolia. The populations from eastern Iran and Turan require an additional source of ancestry from a population related to West Siberian Hunter-Gatherers. Consistent with the *f*-statistic patterns, we also observe that Anatolian farmer-related ancestry decreases from west to east while the West Siberian Hunter-Gatherer-related ancestry increases.

			Mixture Proportions		SE	
Source1	**Source2**	**P**	**S1**	**S2**	**S1**	**S2**
PPN	*Ganj_Dareh_N*	0.47	0.15	0.85	0.03	0.03
Anatolia_N	*Ganj_Dareh_N*	0.62	0.13	0.87	0.03	0.03

Table S 8 Distal models for *Seh_Gabi_LN*

			Mixture Proportions		SE	
Source1	**Source2**	**P**	**S1**	**S2**	**S1**	**S2**
Anatolia_N	*Ganj_Dareh_N*	0.14	0.37	0.63	0.02	0.02

Table S 9 Distal models for *Seh_Gabi_C*

			Mixture Proportions		SE	
Source1	**Source2**	**P**	**S1**	**S2**	**S1**	**S2**
Anatolia_N	*Ganj_Dareh_N*	0.02	0.53	0.47	0.02	0.02

Table S 10 Distal models for *Hajji_Firuz_C*

				Mixture Proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
WSHG	*Anatolia_N*	*Ganj_Dareh_N*	0.002	0.05	0.31	0.64	0.01	0.02	0.02

Table S 11 Distal models for *Tepe_Hissar_C*

				Mixture Proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
WSHG	*Anatolia_N*	*Ganj_Dareh_N*	0.01	0.11	0.17	0.73	0.01	0.02	0.02

Table S 12 Distal models for *Parkhai_Anau_EN*

				Mixture Proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
WSHG	*Anatolia_N*	*Ganj_Dareh_N*	0.01	0.17	0.21	0.63	0.01	0.02	0.02

Table S 13 Distal models for *Geoksyur_EN*

				Mixture Proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
EEHG	*PPN*	*Ganj_Dareh_N*	0.01	0.06	0.13	0.81	0.02	0.03	0.04
EEHG	*Anatolia_N*	*Ganj_Dareh_N*	0.07	0.07	0.09	0.85	0.02	0.03	0.03
WSHG	*PPN*	*Ganj_Dareh_N*	0.05	0.1	0.17	0.73	0.02	0.03	0.05
WSHG	*Anatolia_N*	*Ganj_Dareh_N*	0.44	0.11	0.13	0.76	0.02	0.03	0.04

Table S 14 Distal models for *Bustan_EN*

				Mixture Proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
WSHG	*Anatolia_N*	*Ganj_Dareh_N*	0.04	0.25	0.08	0.68	0.02	0.03	0.04

Table S 15 Distal models for Sarazm_EN

				Mixture Proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
EEHG	..	*Ganj_Dareh_N*	0.01	0.06	..	0.94	0.02	..	0.02
EEHG	*PPN*	*Ganj_Dareh_N*	0.06	0.08	0.09	0.83	0.02	0.03	0.04
WSHG	*PPN*	*Ganj_Dareh_N*	0.35	0.14	0.15	0.71	0.02	0.03	0.05
WSHG	*WEHG*	*Ganj_Dareh_N*	0.01	0.08	0.04	0.88	0.02	0.02	0.02
WSHG	*Anatolia_N*	*Ganj_Dareh_N*	0.57	0.14	0.10	0.76	0.02	0.03	0.04

Table S 16 Distal models for *Namazga_CA.SG*

Following this distal modeling, we attempted to find more proximal sources that explain the sources of our ancestry more directly. The final working models for the Proximal sources confirm provide additional insight beyond the distal modeling. The individuals from western Iran (*Hajji_Firuz_C* and *Seh_Gabi_C*), can be modeled as two-way mixtures of populations from the central Zagros with additional ancestry plausibly from the west as their ancestry is consistent with deriving from farmers of Anatolia and the Levant. For *Seh_Gabi_LN*, we do not report any proximal models, as it is an individual from a time period where the only available source populations are the distal sources we use in the outgroups for the proximal modeling framework. The population from Tepe Hissar can be modeled similarly with proximate populations directly to the east and west, but we the only model we were able to obtain reasonable fits for involved negative coefficients with a population with high *WSHG*-related ancestry. The populations from Parkhai, Anau and Geoksyur can be modeled with the population further east from Sarazm along with additional ancestry from the west (**Table S 17 - Table S 21**), though the population from Bustan appeared to be a genetic clade with *Parkhai_Anau_EN* (**Table S 22**). Finally, we show that the populations that are on the eastern extreme of our sampling from Tajikistan can be modeled as being mixtures of populations from Turkmenistan with large amounts of ancestry from the Bronze Age central Steppe (**Table S 23**).

			Mixture proportions		SE	
Source1	**Source2**	**P**	**S1**	**S2**	**S1**	**S2**
Anatolia_C	*Seh_Gabi_C*	0.800	0.378	0.622	0.035	0.035
Baden_LCA	*Seh_Gabi_C*	0.223	0.209	0.791	0.02	0.02

Table S 17 Proximal models for *Hajji_Firuz_C*

			Mixture proportions		SE	
Source1	**Source2**	**P**	**S1**	**S2**	**S1**	**S2**
Bustan_EN	*Hajji_Firuz_C*	0.011	0.343	0.657	0.03	0.03

Table S 18 Proximal models for *Seh_Gabi_C*

				Mixture proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Botai.SG	*Geoksyur_EN*	*Hajji_Firuz_C*	0.312	-0.07	0.936	0.137	0.009	0.031	0.027
Botai.SG	*Geoksyur_EN*	*Seh_Gabi_C*	0.537	-0.06	0.848	0.214	0.009	0.044	0.039

Table S 19 Proximal models for *Tepe_Hissar_C*

				Mixture proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Zagros_N.SG	*Anatolia_C*	*Geoksyur_EN*	0.039	0.245	0.047	0.709	0.034	0.019	0.049
Zagros_N.SG	*Armenia_C*	*Geoksyur_EN*	0.025	0.263	0.071	0.666	0.043	0.032	0.072
Zagros_N.SG	*Geoksyur_EN*	*Hajji_Firuz_C*	0.044	0.225	0.713	0.061	0.029	0.045	0.023
Zagros_N.SG	*Geoksyur_EN*	*Seh_Gabi_C*	0.055	0.206	0.715	0.079	0.026	0.043	0.028
Zagros_N.SG	*Geoksyur_EN*	*Tepe_Hissar_C*	0.076	0.184	0.621	0.195	0.023	0.068	0.065

Table S 20 Proximal models for *Parkhai_Anau_EN*

				Mixture proportions			SE		
Source1	Source2	Source3	P	S1	S2	S3	S1	S2	S3
Aigyrzhal_BA	*Hajji_Firuz_C*	*Parkhai_Anau_EN*	0.019	0.169	0.076	0.755	0.025	0.023	0.035
Aigyrzhal_BA	*Parkhai_Anau_EN*	*Tepe_Hissar_C*	0.044	0.195	0.444	0.36	0.025	0.112	0.101
Botai.SG	*Hajji_Firuz_C*	*Parkhai_Anau_EN*	0.011	0.06	0.094	0.846	0.009	0.024	0.027
Botai.SG	*Parkhai_Anau_EN*	*Tepe_Hissar_C*	0.032	0.074	0.468	0.458	0.01	0.115	0.11
Hajji_Firuz_C	*Parkhai_Anau_EN*	*Central_Steppe_EMBA*	0.075	0.087	0.825	0.089	0.023	0.028	0.013
Seh_Gabi_C	*Parkhai_Anau_EN*	*Central_Steppe_EMBA*	0.036	0.12	0.784	0.096	0.036	0.042	0.013
Parkhai_Anau_EN	*Central_Steppe_EMBA*	*Tepe_Hissar_C*	0.195	0.486	0.104	0.411	0.106	0.013	0.1

Table S 21 Proximal models for *Geoksyur_EN*

Source1	P
Parkhai_Anau_EN	0.017

Table S 22 Proximal models for *Bustan_EN* (clade with *Parkhai_Anau_EN*)

				Mixture proportions			SE		
Source1	Source2	Source3	P	S1	S2	S3	S1	S2	S3
Bustan_BA_o1	..	..	0.089	..	..	..	..	..	..
Aigyrzhal_BA	*Parkhai_Anau_EN*	..	0.015	0.437	0.563	..	0.036	0.036	..
Indus_Periphery_Pool	*Parkhai_Anau_EN*	*Central_Steppe_EMBA*	0.026	0.121	0.687	0.192	0.041	0.036	0.019
Okunevo_BA.SG	*Parkhai_Anau_EN*	*Central_Steppe_EMBA*	0.026	0.064	0.781	0.155	0.022	0.017	0.028
Parkhai_Anau_EN	*Shamanka_EBA.SG*	*Central_Steppe_EMBA*	0.025	0.779	0.029	0.192	0.017	0.01	0.02

Table S 23 Proximal models for *Sarazm_EN*

In summary, the proximate models confirm that from the west of Iran at Hajji Firuz to the east of Turan at Sarazm, most populations can be modeled as mixtures of neighboring populations. The observed gradient of ancestry from Anatolian farmer-related groups and from a population related to the West Siberian Hunter-Gatherer individuals provide clues to the formation of the populations of subsequent sites and places bounds on the spread of ancestry into South Asia. We highlight two key findings:

(1) North Eurasian hunter-gatherer ancestry in Central Asia before the Yamnaya expansion. We observe significant West Siberian Hunter-Gatherer-related ancestry (also related to Ancient North Eurasians (*ANE*) and Eastern European HGs (*EEHG*)) in individuals from Iran and Turan during the early Neolithic period. That most extreme example of this are individuals from Sarazm in Tajikistan, dated to the mid-4th millennium BCE, which can be modeled as having about ~20% of their ancestry attributable to this source. This ancestry

cannot be attributed to admixture with Steppe pastoralist populations such as the Yamnaya and their successors, which appeared in the region only 700 years later. It is tempting to speculate that this ancestry is also characteristic of the hunter-gatherer Kelteminar culture that was spread over this region at this time. Although we do not have direct genetic data from people buried with artifacts of this culture, we do have data from skeletons associated with cultures that were on the central Steppe bearing this type of ancestry, which work formally as sources for this ancestry in Sarazm and the other individuals from Turan. In the next section we analyze and model these individuals from the middle of the 3rd millennium BCE from northern and central Kazakhstan, and show that the ancestry of the individuals from throughout the central and southern Steppe (present-day Kazakhstan) at this time was not derived from groups related to the Yamnaya and instead was derived from people related to those of the Russian forest zone (that is, ancestry related to that in West Siberian Hunter-Gatherers). Taken together, these results imply that West Siberian Hunter-Gatherer-related ancestry was widespread in Kazakhstan and eastern Turan before the spread of the Yamnaya, and thus the Central Asian Steppe was a plausible conduit by which this ancestry might have arisen in these individuals from Turan.

(2) A decrease in Anatolian farmer-related ancestry moving west-to-east in Turan.
Both our proximal and distal models document an admixture cline between Iranian and Anatolian farmers that was established between the Neolithic and Copper Age periods, as evidenced by our data from western Iran where we document the timing of the arrival of Anatolian farmer-related ancestry by radiocarbon dated individuals from a time transect at Seh Gabi. This cline continues eastward into Turan with low to almost no proportion of Anatolian farmer-related ancestry in Sarazm, the population that we have that is at the end of the cline from this period. Importantly, the documentation of the west-to-east ancestry gradient provides insight into the type of West Eurasian ancestry that we might expect to be found in South Asia. We examine this particular question in further detail in a later section modeling outlier individuals from BA Turan that we hypothesize are migrants from populations of South Asia.

S4.4.1.2 The Bronze Age period of Turan and the formation of the BMAC

In this section we model the admixture history of individuals from 4 sites associated with the BMAC; a single individual from Afghanistan (who has fewer than our minimum cutoff of 100,000 SNPs but who we model anyway because of the unique archaeological provenance); Bronze Age individuals from Parkhai (whose Neolithic and Copper Age genetic patterns we described in the previous section); the Bronze Age settlement of Shahr-i-Sokhta close to the border of Iran, Pakistan and Afghanistan; and Bronze Age individuals from Aigyrzhal in Kyrgyzstan. The BMAC and Shahr-i-Sokhta sites represent large urban towns associated with people who adopted irrigated farming techniques. Of particular interest is the archaeological evidence of cultural contact of people at both of these sites with Steppe pastoralist cultures who lived to the north as well as with the Indus Valley Civilization that was spread in the south, with artifacts from the BMAC found in both of these areas and Indus seals and Steppe pottery found within the walled BMAC city of Gonur-Tepe. Considering this archeological evidence, we sought to (a) examine the degree of genetic similarity across the different sites, (b) to determine if we could find evidence for continuity of ancestry from the Copper Age sites in the same region we described in the previous section with individuals from BMAC and Shahr-i-Sokhta, (c) to examine if there are any additional genetic influences from the south or the north and (d) to determine if these influences change with time by comparing to the individuals we have available on the southern Steppe prior to and after around 2000 BCE. The four BMAC sites are spread in time over ~2100-1500 BCE, while the archeological dates from Shahr-i-Sokta as well as the radiocarbon date from Darra-i-kur cave place these individuals around 2600 BCE. The westernmost site of the BMAC, Gonur, is the oldest, dating to a mean of 2062 BCE (the average of the midpoints of the ranges of all individuals for which we have direct radiocarbon dates), while the three other sites (Sappali Tepe, Jarkutan and Bustan), which are very close geographically, have mean dates of 1848 BCE, 1639 BCE, and 1517 BCE, respectively **Fig S 17**.

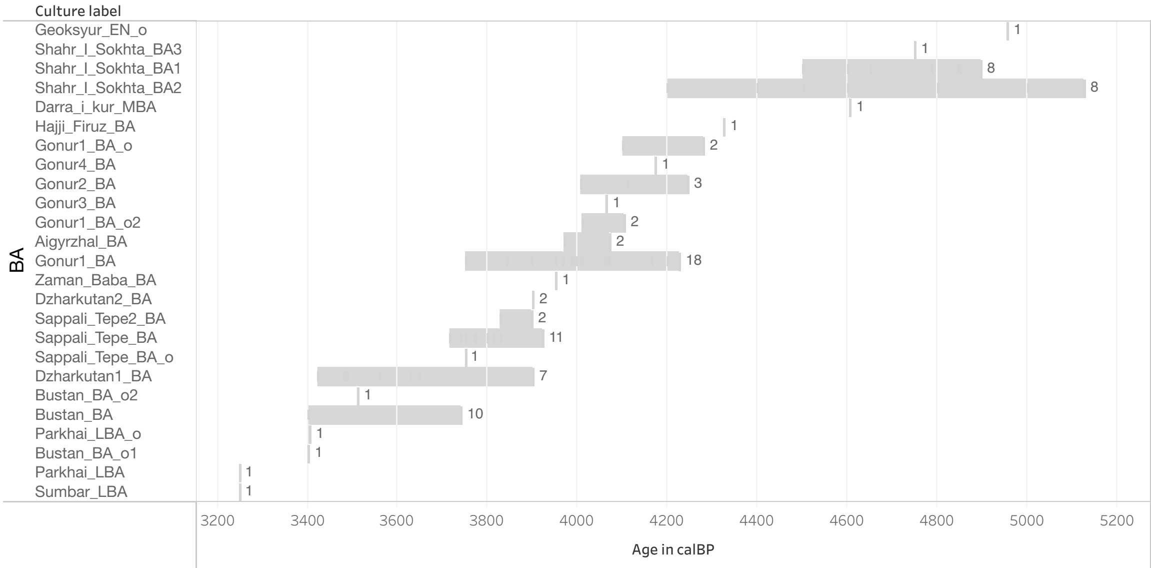

Fig S 17 Number of individuals and distribution of dates for sites analyzed in this section.

We begin with qualitative assessment of our data. The main cluster of BMAC and other BA Turan individuals is situated at the bottom left of the All Eurasian PCA and the bottom right of the West Eurasia PCA plot together with the previously discussed populations from the pre-Copper Age period, suggesting that their primary ancestry is related to that of Neolithic farmers from Iran. Geographically as well as genetically, this main cluster of individuals from these sites lie intermediate between the individuals from Tepe Hissar and Sarazm (**Fig S 18**, top and second panels), consistent with suggestions in the archaeological literature that smaller independent agricultural communities from the east and west coalesced (in a process that included genetic admixture) into larger communities during the Bronze Age (*5*). On the ADMIXTURE plots, we observe that these individuals are mixtures of three main ancestry components related to those in the Neolithic and Copper Age periods of Turan (**Fig S 18**, bottom panel). However, some individuals from each site also harbor trace amounts of a component that is maximized in Andamanese Hunter-Gatherers (*AHG*s) and Dravidian speaking groups in southern India. Particularly revealing is our observation of outlier individuals from several of these sites that are exceptions to these patterns. We hypothesize that these individuals were migrants from South Asia (or descendants of recent migrants) and assign them to a different analysis groupings. We discuss these in more detail at the end of this section as well as in the section focusing on South Asia. There are several individuals that are outliers from the main cluster in other directions and we describe and model these during outlier analysis at the end of the present section.

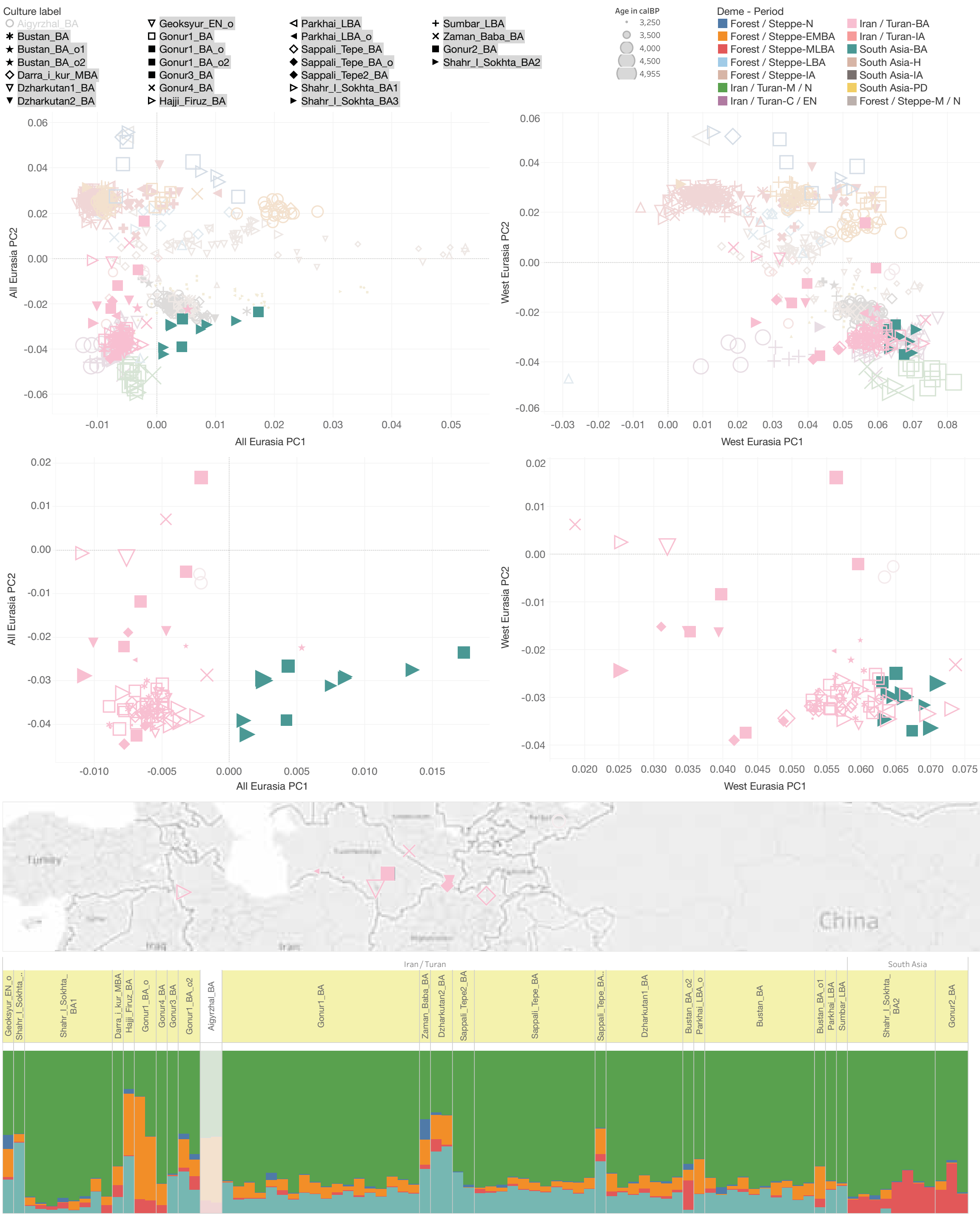

Fig S 18 Top panel: left, All Eurasian PCA; right: West Eurasian PCA. Second panel: left, blow up of All Eurasian PCA; right, blow up of West Eurasian PCA. Third panel: left, geographic locations of individuals. Fourth/bottom panel: ADMIXTURE results, with the colors of green, teal, orange, red and blue representing ancestry maximized in Iranian farmers from the Zagros, farmers from Anatolia, and HGs from West Siberia, South Asia, and Western Europe respectively.

Guided by these results from the PCA and ADMIXTURE analysis, we carried out formal tests of allele sharing. First, we computed “Pre-Copper Age affinity” f_4-statistics within this analyzed subset of populations (**Fig S 19**, left panel). We observe that the BMAC individuals from the main cluster in all 4 sites appear to be homogenous with respect to f_4-statistics as well as on the PCA and ADMIXTURE plots despite a ~500-year difference comparing our earliest to our latest radiocarbon dates, suggesting that after its foundation, the bulk of the population of the BMAC remained stable for a long period of time without substantial exogenous influx and admixture over this period. We caution, however, that the peoples of the BMAC were not completely genetically isolated as we will see in the discussion of outliers below. The main set of individuals from Shahr-i-Sokhta (*Shahr_I_Sokhta_BA1*) resemble the ancestry of the *BMAC*, but with less Anatolian farmer-related and more West Siberian Hunter-Gatherer-related ancestry. In contrast, the individuals from Aigyrzhal appear to be shifted significantly towards *WSHG* but without very much Anatolian farmer ancestry. These individuals extend the cline that we observed in the Copper Age into the mountain corridor east of the Steppe even though they are later in time. Along with an early individual from Dali, they provide insight into the type of ancestry that was present north of Turan, prior to the arrival of Steppe pastoralist ancestry.

Our second observation is that while there are subtle albeit significant differences in the proportion of Anatolian farmer-related and West Siberian Hunter-Gatherer-related ancestry between the *BMAC* genetic cluster and groups of the previous period, there are no statistics showing consistent differences in *WSHG*-affinity comparing Copper Age to the Bronze Age Turan individuals. This suggests that if there was movement from the Steppe into the region during the mature BMAC—mediated for example by the spread of Yamnaya Steppe pastoralists—it impacted most BMAC individuals only minimally.

Fig S 19 Left panel: "Two population comparison" f_4-statistics showing how pairs of groups differ in their affinity to Anatolian farmers (*Anatolia_N*), West Siberian Hunter-Gatherers (*WSHG*), and South Asian-related Hunter-Gatherers (*AHG*). Orange squares indicate |Z-scores| > 3 for sharing more alleles with the test population (*Anatolia_N*, *WSHG*, or *AHG*) compared to a population from the columns, with respect to an outgroup. Blue squares indicate significantly less sharing at the same threshold. Darker orange and blue squares show sharing at |Z-score| > 6. Right panel: "Pre-Copper Age affinity" f_4-statistics measuring the relative affinities of *Test* populations with different pairs of pre-Copper Age populations.

We next examined admixture-f_3 statistics for *Gonur1_BA* and *Shahr_I_Sokhta_BA1* and observed that the strongest signals of admixture are between sources related to Anatolian farmers; West Siberian Hunter-Gatherers; individuals from the southern Steppe that have ancestry derived from both Anatolian farmers and West Siberian Hunter-Gatherers; and Neolithic and Copper Age groups from Iran and Turan (**Fig S 20** - **Fig S 22**).

For the *BMAC* main cluster, we also observe significant (Z<-3) admixture signals with a source from pre-Copper Age Iran and Turan and a source related to present day groups within the Indian subcontinent, a signal that we do not detect in individuals from the earlier period in Turan (**Fig S 21** - **Fig S 23**). This is consistent with the hypothesis that the main *BMAC* cluster harbors a proportion of ancestry from gene flow from the south, plausibly from South Asia. As the four BMAC-associated sites are indistinguishable from each other genetically, we show just admixture-f_3 statistics from Gonur, the BMAC site for which we have the largest number of individuals. We observe that the individuals from Shahr-i-Sokhta, also show significant admixture-f_3 statistics with one source as *AHG*. Taken together with the fact that there are individuals with significantly high proportions of *AHG*-related ancestry at both sites, this suggests that there was gene flow from South Asia out into Turan during the BA. However, we do not observe significant statistics of this sort in *Aigyrzhal_BA*, suggesting a geographic limit to the extent of this gene flow from the South. Similarly, while both main clusters of individuals from Gonur and Shahr-i-Sokhta show highly significant admixture-f_3 statistics with populations with high proportions of Anatolian farmer-related ancestry, *Aigyrzhal_BA* does not, again suggesting that the gene flow of Anatolian farmer ancestry during the Copper Age had minimal impact on people of this site.

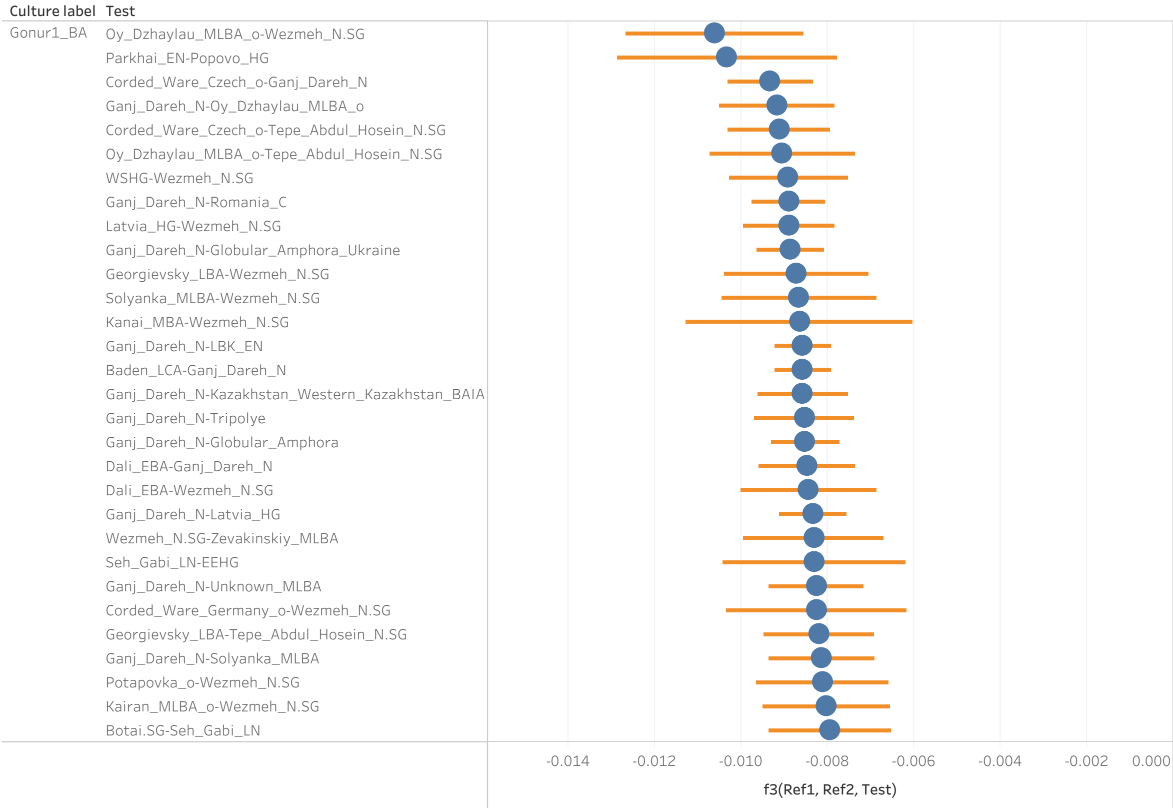

Fig S 20 Top admixture-f_3 pairs for *Gonur1_BA*

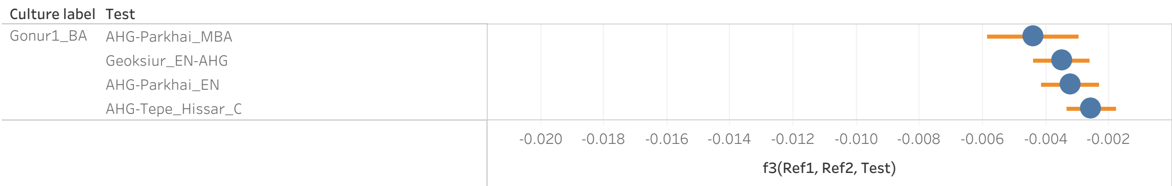

Fig S 21 Additional admixture-f_3 for *Gonur1_BA* with South Asians

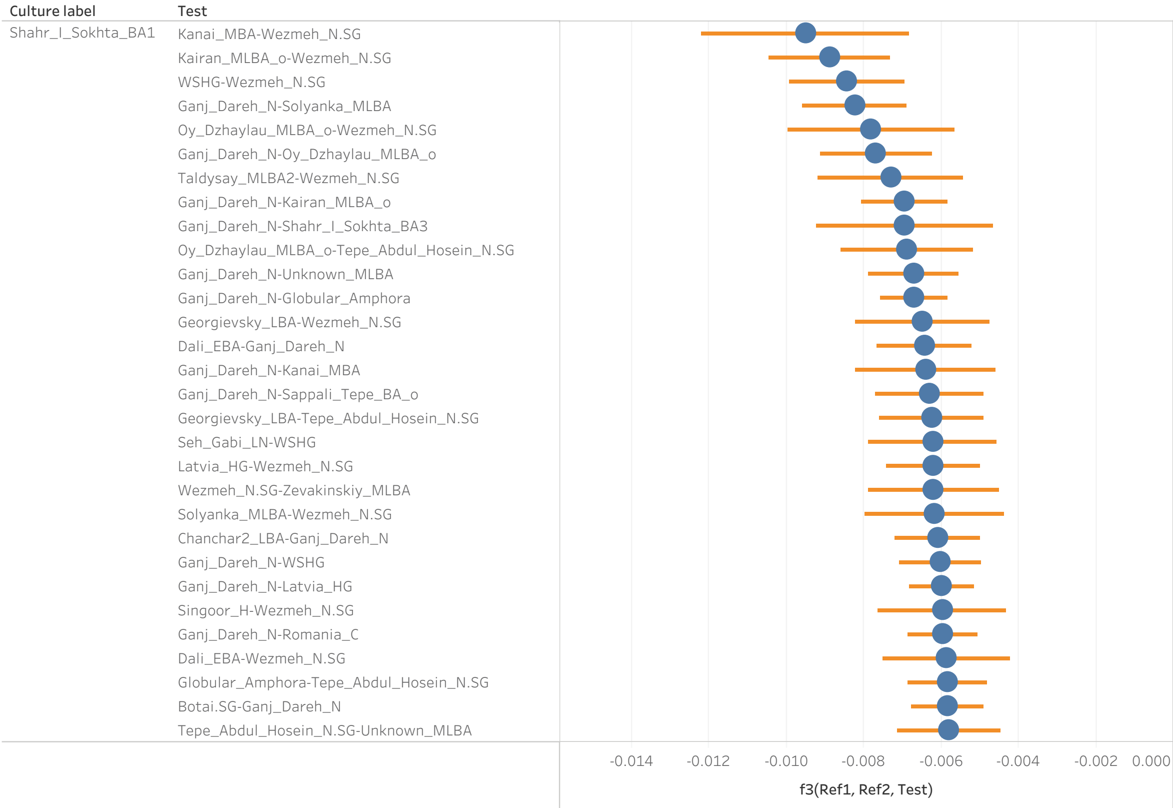

Fig S 22 Top admixture-f_3 pairs for *Shahr_I_Sokhta_BA1*

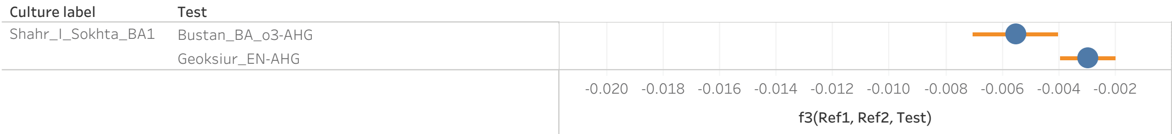

Fig S 23 Additional admixture-f_3 for *Shahr_I_Sokhta_BA1* with South Asians

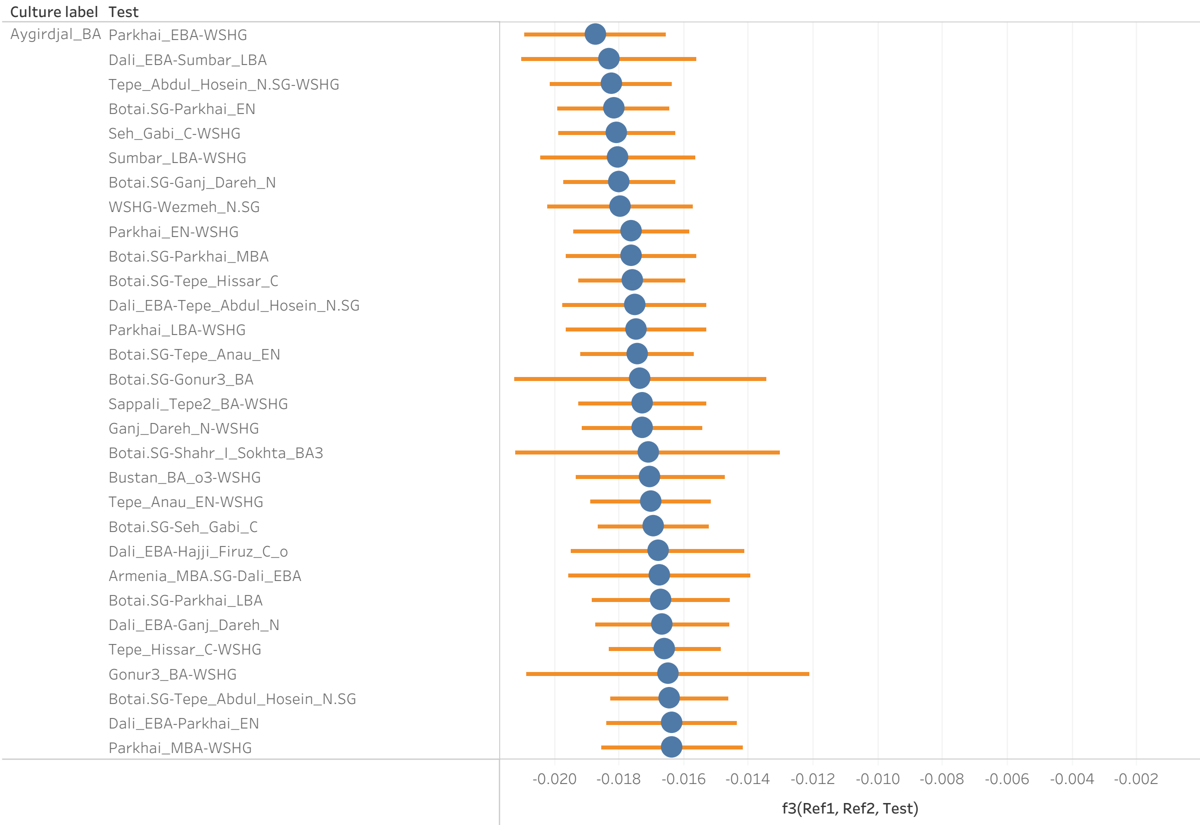

Fig S 24 Top admixture-f_3 for *Aigyrzhal_BA*

To investigate these findings further, we next show our results for our distal and proximal modeling of the BA Turan sites. Consistent with our expectations from the qualitative analysis, formal modeling of the individuals from the Eneolithic and Copper Age period in Turan shows that the majority of the ancestry in all of these sites is related to that of early Iranian farmers, with additional ancestry related to Anatolian farmers, *WSHG*s, and a small proportion related to HGs of South Asia (**Table S 24 - Table S 26**). In addition, the BA individuals from Shahr-i-Sokhta which is geographically closer to South Asia can be modeled as having even lower amounts of Anatolian farmer-related ancestry and higher proportions of *AHG*-related ancestry, providing further evidence that the observed cline in the Copper Age extended into the Bronze Age.

					Mixture Proportions				SE			
Source1	**Source2**	**Source3**	**Source4**	**P**	**S1**	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**
AHG	*WSHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.04	0.02	0.12	0.26	0.59	0.01	0.01	0.02	0.03

Table S 24 Distal models for *BMAC*

Mixture Proportions	SE

Source1	Source2	Source3	P	S1	S2	S3	S1	S2	S3
EEHG	*Ganj_Dareh_N*	..	0.31	0.12	0.88	..	0.04	0.04	..
WSHG	*Ganj_Dareh_N*	..	0.26	0.12	0.89	..	0.05	0.05	..
Ganj_Dareh_N	..	..	0.08	..	..	..	..	..	..

Table S 25 Distal models for Darra-i-kur_MBA

					Mixture Proportions				SE			
Source1	**Source2**	**Source3**	**Source4**	**P**	**S1**	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**
AHG	*WSHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.04	0.04	0.12	0.19	0.66	0.02	0.01	0.02	0.05
WSHG	*Anatolia_N*	*Ganj_Dareh_N*	*ESHG*	0.03	0.11	0.19	0.65	0.06	0.01	0.02	0.04	0.02

Table S 26 Distal models for *Shahr_I_Sokhta_BA1*

				Mixture Proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
WSHG	*Anatolia_N*	*Ganj_Dareh_N*	0.07	0.44	0.08	0.48	0.02	0.03	0.03

Table S 27 Distal models for *Aigyrzhal_BA*

To better understand the population changes associated with these BA settlements we studied proximal sources. We first observe that there are many models that fit under our acceptance criteria, but they are of a similar nature, in the sense that all involve a population from the previously described Copper Age period with additional ancestry related to present-day South Asians. Most working models involve populations from Turan with additional South Asian-related mixture, for example present-day individuals from South Asia with minimal Steppe pastoralist-related ancestry, or outlier individuals from Shahr-i-Sokhta and BMAC sites with high proportions of *AHG*-related ancestry (**Table S 28 - Table S 30**).

				Mixture proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Aigyrzhal_BA	*Seh_Gabi_C*	*Shahr_I_Sokhta_BA1*	0.247	0.084	0.247	0.669	0.021	0.029	0.044
Hajji_Firuz_C	*Shahr_I_Sokhta_BA1*	..	0.183	0.144	0.856	..	0.019	0.019	..
Geoksyur_EN	*Hajji_Firuz_C*	*Indus_Periphery_Pool*	0.013	0.656	0.222	0.121	0.027	0.018	0.017
Geoksyur_EN	*Indus_Periphery_Pool*	*Seh_Gabi_C*	0.087	0.576	0.122	0.302	0.028	0.015	0.02
Seh_Gabi_C	*Khvalynsk_EN*	*Shahr_I_Sokhta_BA1*	0.756	0.23	0.039	0.731	0.027	0.009	0.031
Seh_Gabi_C	*Sarazm_EN*	*Shahr_I_Sokhta_BA1*	0.030	0.258	0.184	0.558	0.041	0.077	0.112
Seh_Gabi_C	*Shahr_I_Sokhta_BA1*	*Central_Steppe_EMBA*	0.330	0.25	0.711	0.04	0.03	0.036	0.01
Seh_Gabi_C	*Shahr_I_Sokhta_BA1*	*Western_Steppe_EMBA*	0.691	0.207	0.745	0.048	0.025	0.029	0.011

Table S 28 Proximal models for BMAC

				Mixture proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Bustan_BA_o1	..	..	0.028	..	..	..	..	..	..

Geoksyur_EN	..	..	0.036	..	..	..	..	..	..
Gonur1_BA_o	..	..	0.259	..	..	..	..	..	..
Namazga_CA.SG	..	..	0.190	..	..	..	..	..	..
Sarazm_EN	..	..	0.060	..	..	..	..	..	..
Aigyrzhal_BA	*BMAC*	..	0.170	0.313	0.687	..	0.108	0.108	..
Shahr_I_Sokhta_BA1	..	..	0.063	..	..	..	..	..	..
Aigyrzhal_BA	*Seh_Gabi_C*	..	0.028	0.576	0.424	..	0.08	0.08	..
Aigyrzhal_BA	*Parkhai_Anau_EN*	..	0.038	0.354	0.646	..	0.119	0.119	..
Aigyrzhal_BA	*Tepe_Hissar_C*	..	0.033	0.408	0.592	..	0.109	0.109	..
BMAC	*Indus_Periphery_Pool*	..	0.036	0.744	0.256	..	0.119	0.119	..
BMAC	*Khvalynsk_EN*	..	0.124	0.839	0.161	..	0.054	0.054	..
BMAC	*Central_Steppe_EMBA*	..	0.294	0.825	0.175	..	0.051	0.051	..
BMAC	*Western_Steppe_EMBA*	..	0.081	0.82	0.18	..	0.065	0.065	..
BMAC	*Steppe_LBA*	..	0.116	0.805	0.195	..	0.066	0.066	..
BMAC	*Steppe_MLBA_oBMAC*	..	0.034	0.841	0.159	..	0.071	0.071	..
BMAC	*Central_Steppe_MLBA*	..	0.027	0.861	0.139	..	0.066	0.066	..
BMAC_o2	*Parkhai_Anau_EN*	..	0.016	0.351	0.649	..	0.13	0.13	..
BMAC_o2	*Shahr_I_Sokhta_BA1*	..	0.216	0.271	0.729	..	0.123	0.123	..
Gonur1_BA_o	*Parkhai_Anau_EN*	..	0.639	0.742	0.258	..	0.124	0.124	..
Gonur1_BA_o	*Shahr_I_Sokhta_BA1*	..	0.892	0.651	0.349	..	0.138	0.138	..
Hajji_Firuz_C	*Sarazm_EN*	..	0.186	0.206	0.794	..	0.092	0.092	..
Khvalynsk_EN	*Parkhai_Anau_EN*	..	0.076	0.196	0.804	..	0.053	0.053	..
Khvalynsk_EN	*Shahr_I_Sokhta_BA1*	..	0.310	0.148	0.852	..	0.056	0.056	..
Khvalynsk_EN	*Tepe_Hissar_C*	..	0.034	0.227	0.773	..	0.052	0.052	..
Parkhai_Anau_EN	*SPGT*	..	0.039	0.587	0.413	..	0.122	0.122	..
Parkhai_Anau_EN	*SPGT_o*	..	0.040	0.482	0.518	..	0.209	0.209	..
Parkhai_Anau_EN	*Central_Steppe_EMBA*	..	0.066	0.811	0.189	..	0.053	0.053	..
Parkhai_Anau_EN	*Western_Steppe_EMBA*	..	0.090	0.762	0.238	..	0.062	0.062	..
Parkhai_Anau_EN	*Steppe_LBA*	..	0.110	0.755	0.245	..	0.064	0.064	..
Parkhai_Anau_EN	*Steppe_MLBA_oBMAC*	..	0.065	0.758	0.242	..	0.067	0.067	..
Parkhai_Anau_EN	*Western_Steppe_MLBA*	..	0.039	0.795	0.205	..	0.059	0.059	..
Parkhai_Anau_EN	*Central_Steppe_MLBA*	..	0.054	0.779	0.221	..	0.062	0.062	..
Indus_Periphery_Pool	*Seh_Gabi_C*	*Khvalynsk_EN*	0.014	0.379	0.433	0.188	0.104	0.077	0.062
Shahr_I_Sokhta_BA1	*Central_Steppe_EMBA*	..	0.272	0.859	0.141	..	0.057	0.057	..
Indus_Periphery_Pool	*Seh_Gabi_C*	*Central_Steppe_EMBA*	0.021	0.267	0.508	0.225	0.128	0.084	0.071
Shahr_I_Sokhta_BA1	*Western_Steppe_EMBA*	..	0.395	0.813	0.187	..	0.066	0.066	..
Indus_Periphery_Pool	*Steppe_MLBA_oBMAC*	*Tepe_Hissar_C*	0.038	0.268	0.205	0.527	0.117	0.07	0.117
Indus_Periphery_Pool	*Western_Steppe_MLBA*	*Tepe_Hissar_C*	0.023	0.293	0.172	0.536	0.115	0.062	0.116
Indus_Periphery_Pool	*Central_Steppe_MLBA*	*Tepe_Hissar_C*	0.025	0.266	0.184	0.551	0.118	0.066	0.114
Central_Steppe_EMBA	*Tepe_Hissar_C*	..	0.063	0.224	0.776	..	0.05	0.05	..
Western_Steppe_EMBA	*Tepe_Hissar_C*	..	0.024	0.264	0.736	..	0.062	0.062	..

Table S 29 Proximal models for *Darra-i-kur_MBA*

				Mixture proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Aigyrzhal_BA	*Indus_Periphery_Pool*	*Tepe_Hissar_C*	0.021	0.071	0.220	0.708	0.027	0.028	0.021

BMAC	*Indus_Periphery_Pool*	*Parkhai_Anau_EN*	0.452	0.451	0.121	0.427	0.084	0.023	0.079
Geoksyur_EN	*Indus_Periphery_Pool*	*Tepe_Hissar_C*	0.658	0.394	0.204	0.402	0.085	0.024	0.073
Hajji_Firuz_C	*Indus_Periphery_Pool*	*Parkhai_Anau_EN*	0.180	0.102	0.187	0.711	0.021	0.023	0.034
Indus_Periphery_Pool	*Seh_Gabi_C*	*Parkhai_Anau_EN*	0.046	0.19	0.132	0.678	0.024	0.031	0.043
Indus_Periphery_Pool	*Namazga_CA.SG*	*Tepe_Hissar_C*	0.067	0.175	0.317	0.507	0.038	0.106	0.075
Indus_Periphery_Pool	*Parkhai_Anau_EN*	*Tepe_Hissar_C*	0.295	0.215	0.373	0.412	0.024	0.097	0.084
Indus_Periphery_Pool	*Sarazm_EN*	*Tepe_Hissar_C*	0.327	0.183	0.221	0.6	0.031	0.056	0.037
Indus_Periphery_Pool	*Central_Steppe_EMBA*	*Tepe_Hissar_C*	0.023	0.231	0.035	0.734	0.025	0.013	0.02

Table S 30 Proximal models for *Shahr_I_Sokhta_BA1*

				Mixture proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Gonur1_BA_o	..	..	0.351	..	..	..	..	..	..
BMAC	*Central_Steppe_EMBA*	..	0.045	0.491	0.509	..	0.018	0.018	..
BMAC	*Indus_Periphery_Pool*	*Central_Steppe_EMBA*	0.529	0.391	0.129	0.48	0.037	0.041	0.02
BMAC	*Sarazm_EN*	*Central_Steppe_EMBA*	0.186	0.251	0.307	0.442	0.1	0.127	0.034
BMAC_o2	*Sarazm_EN*	*Central_Steppe_EMBA*	0.171	0.087	0.535	0.378	0.035	0.042	0.025
Namazga_CA.SG	*Central_Steppe_EMBA*	..	0.070	0.62	0.38	..	0.02	0.02	..
Geoksyur_EN	*Indus_Periphery_Pool*	*Central_Steppe_EMBA*	0.733	0.409	0.155	0.437	0.037	0.037	0.02
Sarazm_EN	*Central_Steppe_EMBA*	..	0.040	0.619	0.381	..	0.026	0.026	..
Hajji_Firuz_C	*Sarazm_EN*	*Central_Steppe_EMBA*	0.126	0.066	0.531	0.403	0.029	0.046	0.027
Shahr_I_Sokhta_BA1	*Central_Steppe_EMBA*	..	0.642	0.506	0.494	..	0.019	0.019	..
Indus_Periphery_Pool	*Namazga_CA.SG*	*Central_Steppe_EMBA*	0.410	0.104	0.513	0.382	0.039	0.044	0.02
Indus_Periphery_Pool	*Parkhai_Anau_EN*	*Central_Steppe_EMBA*	0.384	0.146	0.361	0.493	0.04	0.034	0.02
Indus_Periphery_Pool	*Sarazm_EN*	*Central_Steppe_EMBA*	0.101	0.099	0.516	0.385	0.047	0.055	0.024
Indus_Periphery_Pool	*Central_Steppe_EMBA*	*Tepe_Hissar_C*	0.829	0.169	0.499	0.332	0.038	0.02	0.031
Seh_Gabi_C	*Parkhai_Anau_EN*	*Central_Steppe_EMBA*	0.021	0.089	0.374	0.537	0.044	0.051	0.018
Seh_Gabi_C	*Sarazm_EN*	*Central_Steppe_EMBA*	0.128	0.094	0.488	0.419	0.042	0.064	0.03
Shamanka_EBA.SG	*Central_Steppe_EMBA*	*Tepe_Hissar_C*	0.047	0.03	0.518	0.452	0.01	0.02	0.017

Table S 31 Proximal models for *Aigyrzhal_BA*

With the large number of individuals we report in the BA from several urban settlements, we observed the presence of a significant number of individuals that appeared to be outliers to the main ancestry types that we observed previously in the Copper Age, but also to the majority of individuals in BMAC sites or at Shahr-i-Sokhta. In many cases, we observed outliers of similar ancestry spread across multiple sites. These observations suggest that urban settlements in BA Central Asia were cosmopolitan and harbored migrants who plausibly came to these sites from nearby cultures that have archaeological evidence of cultural contact with the BMAC and Shahr-i-Sokhta.

We identified a total of 28 outlier individuals of 6 different ancestry types spread across a range of sites in Turan. We begin by describing the ancestry of these individuals qualitatively through PCA, ADMIXTURE plots, and f_4 statistics.

The first and most important set of these outliers is comprised of 11 individuals, 3 from the city of Gonur and 8 from the city of Shahr-i-Sokhta, which have significant fractions of ancestry related to AHGs. On the ADMIXTURE plots, these individuals, which we call *Gonur2_BA*, *Shahr_I_Soktha_BA2*. These individuals have a large (but variable) proportion of their ancestry attributed to the red component (**Fig S 18**), which is maximized in present-day people from South Asia. Additionally, these individuals are positioned on the PCA close to a set of individuals from the Late Bronze-Iron Age in the Swat Valley of Pakistan, suggesting that they might be migrants to Turan from South Asia. We will examine these relationships further in the characterization of South Asia IA data, but as shown in the "Pre-Copper Age Affinity" and "Two Population Comparison" f_4-statistics (**Fig S 19**), these individuals harbor significantly higher fractions of South Asian Hunter Gatherer (*AASI*)-related ancestry compared to all previously reported individuals from the Turan region. The presence of multiple individuals with this type of ancestry in Turan during this time suggests that admixture from people with ancestry this may have contributed to the *AHG*-related admixture detected at these sites. The observation of this large number of outlier individuals also provides strong circumstantial evidence that they might reflect a type of ancestry typical of people of the Indus Valley Civilization, the civilization of the south with which people of the BMAC and Shahr-i-Sokhta had strong archaeological evidence of cultural contact.

The second set of outliers is comprised of two individuals from Gonur dated to 2330 BCE and 2150 BCE and one individual from Bustan. All have substantial *WSHG*-related ancestry, as seen in the ADMIXTURE plots and direction of the shift in the PCA from the main *BMAC* cluster. The "Pre-Copper Age Affinity" and "Two Population Comparison" f_4-statistics show that these three individuals have significantly more *WSHG*-related ancestry compared to all previously reported individuals from the Turan region. These individuals have relatively little (and possibly no) ancestry from Anatolian farmers (and therefore were not from groups descended from Yamnaya Steppe pastoralists or their descendants who had large fractions of such ancestry). These individuals thus provide another line of evidence that the populations of Copper Age and Bronze Age Turan were in contact with populations from the central

Steppe prior to the arrival of the Steppe pastoralists from the western Steppe in the 2nd and 3rd millennium BCE.

The third set of outliers is comprised of 5 individuals from Gonur, Jarkutan, Sappali Tepe and Parkhai dated to 2107 BCE, 1950 BCE (2 individuals), 1800 BCE and 1455 BCE, which all show increases in both *EEHG*-related and Anatolian farmer-related ancestry and deviated in the direction of 2nd millennium Steppe pastoralist individuals on the West Eurasian PCA (**Fig S 18**). In "Pre-Copper Age Affinity" f_4-statistics, we see that compared to the main *BMAC* genetic cluster, these have an increased amount of both Anatolian farmer and West Siberian Hunter-Gatherer-related ancestries (**Fig S 19**).

A fourth set of outliers comprises 4 individuals: 2 from Sappali Tepe, 1 from Gonur and 1 from Shahr-i-Sokhta. These individuals on both the PCA as well as the ADMIXTURE plots are shifted toward the individuals from western Iran, and have large proportions of a teal component and small to negligible proportions of an orange component maximized in *WSHG*, suggesting that they were migrants from further to the west. On the "Pre-Copper Age Affinity" f_4-statistics, these individuals have clear genetic affinity to individuals from sites that are further west: *Seh_Gabi_C* and *Tepe_Hissar_C*.

The final set of 2 outliers from the BA period of Turan both come from the site of Bustan.

The first of these outliers, *Bustan_BA_o2*, is an individual that has an increased proportion of South Asian Hunter Gatherer (*AASI*)-related ancestry, but also of *Anatolia_N* ancestry as well as *WEHG*-related ancestry, a type of ancestry we observe only later in time upon admixture with people related to Steppe Pastoralists. On the PCA as well as in ADMIXTURE plots, the ancestry of this individual appears to be very similar to that of populations from the Late Bronze-Iron Age Swat Valley (*SPGT*). "Pre-Copper Age affinity" f_4-statistics confirm that the individuals have increased *AHG*-related ancestry, but also higher proportions of *WSHG* and *Anatolia_N* ancestry, unlike the earlier outliers *Gonur2_BA* and *Shahr_I_Sokhta_BA2*. Thus, despite some shared ancestry, we do not pool *Bustan_BA_o2* with *Gonur2_BA* and *Shahr_I_Sokhta_BA2* in the analyses that follow.

The second outlier from Bustan is from a population that appears to have reduced proportions of *Anatolia_N* or *WSHG*-related ancestry, compared with the main cluster of individuals from the *BMAC* exemplified by *Gonur1_BA*, *Sappali_Tepe_BA*, *Dzharkutan1_BA* and *Bustan_BA* (Z > 2 for all such f_4-statistics). In accord with this pattern of genetic ancestry, when we obtained a direct radiocarbon date of this individual we found it was pre-Bronze Age—average of 3152 BCE—even though it was originally categorized as Bronze Age based on archaeological context. This ancestry type is in accord with the genetic pattern we observe in individuals from the Copper Age and Eneolithic periods from Turan from this time, discussed in the previous section.

To disambiguate the sources of ancestry and to estimate the admixture proportions, we next moved to *qpAdm* modeling and first describe the results of distal modeling.

The first set of outlier individuals with large fractions of *AHG*-related ancestry (the *Indus Periphery Cline*) are heterogenous, and we model these individuals in detail in **Section S4** using a modified set of *Right* outgroups that is optimized for studying South Asian ancestry. A summary of this analysis is that it shows that the *Indus Periphery Cline* individuals have a much lower ratio of Anatolian to Iranian farmer-related ancestry than do the other Turan individuals we sampled. Under the interpretation that these individuals reflect migration from South Asia, this suggest that the Iranian farmer-related ancestry in South Asia during the Bronze Age had a minimal proportion of Anatolian farmer-related ancestry, consistent with the hypothesis that the west-to-east reduction of this ancestry type documented in the previous subsection projected all the way into South Asia even though we do not have any individuals directly sampled from South Asia.

The second set of outlier individuals, *Gonur1_BA_o*, can similarly be modeled without any Anatolian farmer-related ancestry, though models with up to 15% *Anatolia_N* ancestry are also consistent with the data. These individuals who date to before 2000 CBE can be modeled simply as a two-way mixture between early farmers from Iran and a source related to *WSHG*s (**Table S 32**). Almost 40% of the ancestry of these individuals is consistent with deriving from this latter (northern) source, providing direct evidence for contact between BMAC sites and peoples from the Steppe that did not have ancestry related to Yamnaya Steppe pastoralists.

The third class of outliers, from the BMAC sites of Gonur, Jarkutan, and Sappali Tepe—as well as from the non-BMAC site of Parkhai—show significantly higher proportions of both eastern European hunter-gatherer and Anatolian farmer-related ancestry. All working models however, also require ancestry related to that in Europe either via *EEHG* or *WEHG*, suggesting that this signal is due to admixture with a Steppe pastoralist population related to those that we observe widespread in the late 3rd millennium and early 2nd millennium BCE that all had large proportions of ancestry from both of these sources (**Table S 33**).

The fourth set of outliers can be modeled parsimoniously with just two sources—*Anatolia_N* and *Ganj_Dareh_N*—without any admixture from *WSHG*, similar to Copper Age populations from western Iran (**Table S 34**). The proportion of Anatolian early farmer-related ancestry in these individuals is significantly higher than in either the main BMAC or Shahr-i-Sokhta genetic clusters, providing evidence for the hypothesis that these individuals are admixed with groups related to those we have also documented in western Iran.

The fifth outlier we observed requires ancestry from all 4 sources that we observe in BA Turan, namely *AHG, WHSG, Anatolia_N* and *Ganj_Dareh_N*, a feature shared with individuals from the Late Bronze-Iron Age Swat Valley as discussed in a later section (**Table S 35**).

				Mixture Proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
WSHG	..	*Ganj_Dareh_N*	0.12	0.43	..	0.57	0.02	..	0.02
EEHG	*WSHG*	*Ganj_Dareh_N*	0.50	0.13	0.29	0.58	0.04	0.05	0.02
WSHG	*PPN*	*Ganj_Dareh_N*	0.87	0.47	0.16	0.37	0.02	0.04	0.05
WSHG	*WEHG*	*Ganj_Dareh_N*	0.39	0.40	0.06	0.54	0.02	0.02	0.03
WSHG	*Anatolia_N*	*Ganj_Dareh_N*	0.97	0.47	0.14	0.39	0.03	0.04	0.05

Table S 32 Distal models for *Gonur1_BA_o*

					Mixture Proportions					SE			
Source1	**Source2**	**Source3**	**Source4**	**P**	**S1**	**S2**	**S3**	**S4**	**S5**	**S1**	**S2**	**S3**	**S4**
AHG	*EEHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.23	0.10	0.18	0.44	0.28	..	0.02	0.06	0.06	0.03
ESHG	*EEHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.50	0.07	0.16	0.44	0.34	..	0.02	..	..	0.02

Table S 33 Distal models for BMAC_o2

					Mixture Proportions					SE			
Source1	**Source2**	**Source3**	**Source4**	**P**	**S1**	**S2**	**S3**	**S4**	**S5**	**S1**	**S2**	**S3**	**S4**
WSHG	*PPN*	*Anatolia_N*	*Ganj_Dareh_N*	0.83	0.06	0.14	0.15	0.66	..	0.02	0.06	0.06	0.03

..	..	*Anatolia_N*	*Ganj_Dareh_N*	0.30	0.26	..	..	0.75	..	0.02	..	..	0.02
WSHG	..	*Anatolia_N*	*Ganj_Dareh_N*	0.56	0.04	..	0.27	0.69	..	0.02	..	0.02	0.03

Table S 34 Distal models for *BMAC2*

					Mixture Proportions					SE			
Source1	**Source2**	**Source3**	**Source4**	**P**	**S1**	**S2**	**S3**	**S4**	**S5**	**S1**	**S2**	**S3**	**S4**
AHG	*WSHG*	*PPN*	*Ganj_Dareh_N*	0.05	0.22	0.18	0.14	0.47	..	0.03	0.02	0.05	0.08
AHG	*WSHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.55	0.27	0.21	0.15	0.38	..	0.03	0.02	0.04	0.08
WSHG	*Anatolia_N*	*Ganj_Dareh_N*	*ESHG*	0.07	0.14	0.11	0.48	0.27	..	0.02	0.04	0.08	0.04

Table S 35 Distal models for Bustan_BA_o2

As with the distal modeling, we do not examine proximal models for the set of 11 outlier individuals with additional *AHG*-related ancestry, that we assign the analysis label: *Indus Periphery Cline* (these individuals are denoted *Gonur2_BA* and *Shahr_I_Sokhta_BA2* in the "split labels"). However, we deal with these individuals using a clinal analysis leveraging the methodology presented in **Supplementary Materials S5**.

We begin by modeling the outliers from Gonur with high proportions of West Siberian Hunter-Gatherer-related ancestry. We observe that these individuals fit without any South Asian Hunter-Gatherer (*AASI*)-related source but can be modeled as a mixture of the main *BMAC* genetic cluster and BA individuals from the southern Steppe (*Central_Steppe_EMBA*). We also note that Eneolithic populations to the northeast of the BMAC from Aigyrzhal work as a single source to explain the ancestry in these individuals, consistent with an alternative hypothesis in which these were migrants from the northeast (**Table S 36**). All models that include a *Western_Steppe_EMBA* group as a source also could only fit with additional ancestry related to *Central_Steppe_EMBA*. This provides an independent line of evidence that admixture between the *BMAC* main cluster and populations from the southern Steppe occurred prior to the arrival of pastoralists from the western Steppe.

Third, we examine the ancestry of the outliers from the BMAC sites of Gonur, Sappali Tepe and Jarkutan that are contemporaneous with various Steppe groups in southern Kazakhstan. We observe that in these outlier individuals, 15-40% of the ancestry can be attributed to admixture from Middle to Late Bronze Age Steppe pastoralists with the rest of the ancestry attributable to a preceding or contemporary population from Iran or Turan (**Table S 37**). These results provide the first direct genetic evidence of admixture between peoples of Steppe pastoralist ancestry and those of the agricultural zones of Turan, and show that it began to have a widespread impact only in the early 2nd millennium BCE despite its wide prevalence in the region thereafter.

Proximal models for the fourth set of outliers that we call *BMAC2* show that these individuals can be parsimoniously explained by two-source models involving BMAC and groups from western Iran, consistent with an ongoing process of admixture from the west between the Copper Age and Bronze Age (**Table S 38**).

Finally, we turn to last two sets of outliers.

The ancestry of *Bustan_BA_o2* resembles that of populations from the Late Bronze-Iron Age Swat Valley and can be explained parsimoniously as a two-way mixture between a pool of the *Indus Periphery Cline* individuals ("*Indus_Periphery_Pool*") and either *Western_Steppe_MLBA*, *Central_Steppe_MLBA*, or *Steppe_MLBA_oBMAC*. These individuals can also be modeled as a mixture of *SPGT* with an additional ~30% ancestry from the *Indus_Periphery_Pool*, suggesting that this individual's ancestry was similar to that of the *SPGT* albeit with higher proportions of ancestry related to *AHG* (**Table S 39**). In the supplementary section focusing on South Asia, we show that these individuals can be co-fit with individuals of the Swat Valley as part of an ancient cline formed through a mixture of two ancestral populations: *Central_Steppe_MLBA* and a pool of *Indus Periphery Cline*.

				Mixture proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
BMAC	*Central_Steppe_EMBA*	..	0.557	0.449	0.551	..	0.025	0.025	..
Geoksyur_EN	*Central_Steppe_EMBA*	..	0.167	0.49	0.51	..	0.027	0.027	..
Parkhai_Anau_EN	*Central_Steppe_EMBA*	..	0.181	0.427	0.573	..	0.024	0.024	..
Sarazm_EN	*Central_Steppe_EMBA*	..	0.084	0.575	0.425	..	0.034	0.034	..
Shahr_I_Sokhta_BA1	*Central_Steppe_EMBA*	..	0.765	0.462	0.538	..	0.025	0.025	..
Central_Steppe_EMBA	*Tepe_Hissar_C*	..	0.195	0.603	0.397	..	0.022	0.022	..
Aigyrzhal_BA	..	..	0.351	..	..	..	..	..	..

Table S 36 Proximal models for *Gonur1_BA_o*

				Mixture proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Baden_LCA	*Shahr_I_Sokhta_BA1*	*Steppe_LBA*	0.016	0.267	0.468	0.266	0.012	0.016	0.016
Globular_Amphora	*Seh_Gabi_C*	*Steppe_LBA*	0.406	0.111	0.518	0.371	0.015	0.017	0.015
Hajji_Firuz_C	*Okunevo_BA.SG*	*Western_Steppe_EMBA*	0.099	0.635	0.097	0.268	0.016	0.012	0.022

Hajji_Firuz_C	*Okunevo_BA.SG*	*Steppe_MLBA_oBMAC*	0.073	0.588	0.12	0.291	0.019	0.011	0.024
Hajji_Firuz_C	*Shamanka_EBA.SG*	*Western_Steppe_EMBA*	0.388	0.611	0.054	0.336	0.016	0.007	0.017
Hajji_Firuz_C	*Shamanka_EBA.SG*	*Steppe_MLBA_oBMAC*	0.017	0.549	0.068	0.384	0.019	0.006	0.02
Seh_Gabi_C	*Okunevo_BA.SG*	*Western_Steppe_MLBA*	0.342	0.553	0.106	0.341	0.015	0.01	0.018
Seh_Gabi_C	*Okunevo_BA.SG*	*Central_Steppe_MLBA*	0.088	0.558	0.091	0.351	0.015	0.01	0.019
Seh_Gabi_C	*Shamanka_EBA.SG*	*Steppe_MLBA_oBMAC*	0.047	0.497	0.057	0.446	0.017	0.006	0.018
Seh_Gabi_C	*Shamanka_EBA.SG*	*Western_Steppe_MLBA*	0.230	0.53	0.064	0.406	0.015	0.006	0.016
Seh_Gabi_C	*Shamanka_EBA.SG*	*Central_Steppe_MLBA*	0.419	0.536	0.055	0.409	0.015	0.006	0.016
Seh_Gabi_C	*Steppe_LBA*	*Western_Steppe_MLBA*	0.015	0.544	0.268	0.188	0.016	0.026	0.029

Table S 37 Proximal models for *BMAC_o2*

				Mixture proportions			**SE**		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Anatolia_C	*BMAC*	*Tepe_Hissar_C*	0.012	0.095	0.327	0.578	0.023	0.131	0.132
Anatolia_C	*Shahr_I_Sokhta_BA1*	*Tepe_Hissar_C*	0.021	0.125	0.311	0.564	0.025	0.117	0.129
Anatolia_EBA	*Shahr_I_Sokhta_BA1*	*Tepe_Hissar_C*	0.011	0.121	0.386	0.493	0.025	0.129	0.143
BMAC	*Seh_Gabi_C*	..	0.246	0.629	0.371	..	0.052	0.052	..
Geoksyur_EN	*Seh_Gabi_C*	..	0.129	0.502	0.498	..	0.043	0.043	..
Aigyrzhal_BA	*Seh_Gabi_C*	*Tepe_Hissar_C*	0.041	0.079	0.356	0.565	0.028	0.071	0.086
Hajji_Firuz_C	*Parkhai_Anau_EN*	..	0.107	0.365	0.635	..	0.031	0.031	..
BMAC	*Hajji_Firuz_C*	*Tepe_Hissar_C*	0.031	0.397	0.172	0.431	0.129	0.039	0.138
Seh_Gabi_C	*Namazga_CA.SG*	..	0.602	0.454	0.546	..	0.044	0.044	..
Seh_Gabi_C	*Parkhai_Anau_EN*	..	0.308	0.508	0.492	..	0.041	0.041	..
Seh_Gabi_C	*Shahr_I_Sokhta_BA1*	..	0.540	0.472	0.528	..	0.045	0.045	..
BMAC_o2	*Seh_Gabi_C*	*Tepe_Hissar_C*	0.037	0.111	0.188	0.701	0.04	0.072	0.064
Gonur1_BA_o	*Seh_Gabi_C*	*Parkhai_Anau_EN*	0.612	0.119	0.499	0.382	0.057	0.038	0.065
Gonur1_BA_o	*Seh_Gabi_C*	*Tepe_Hissar_C*	0.125	0.18	0.338	0.482	0.049	0.059	0.087
Hajji_Firuz_C	*Shahr_I_Sokhta_BA1*	*Tepe_Hissar_C*	0.057	0.234	0.383	0.384	0.045	0.121	0.148
Seh_Gabi_C	*Khvalynsk_EN*	*Tepe_Hissar_C*	0.014	0.309	0.033	0.658	0.066	0.014	0.071
Seh_Gabi_C	*Sarazm_EN*	*Tepe_Hissar_C*	0.108	0.42	0.189	0.391	0.079	0.06	0.125
Seh_Gabi_C	*Central_Steppe_EMBA*	*Tepe_Hissar_C*	0.026	0.342	0.039	0.619	0.07	0.014	0.076
Seh_Gabi_C	*Western_Steppe_EMBA*	*Tepe_Hissar_C*	0.013	0.287	0.042	0.671	0.065	0.018	0.069

Table S 38 Proximal models for *BMAC2*

				Mixture proportions			**SE**		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Indus_Periphery_Pool	*SPGT*	..	0.016	0.288	0.712	..	0.117	0.117	..
Indus_Periphery_Pool	*Steppe_LBA*	..	0.192	0.77	0.23	..	0.034	0.034	..
BMAC	*Saidu_Sharif_H_o*	*Steppe_MLBA_oBMAC*	0.011	0.264	0.618	0.119	0.047	0.038	0.036
Indus_Periphery_Pool	*Western_Steppe_MLBA*	..	0.011	0.847	0.153	..	0.025	0.025	..
Indus_Periphery_Pool	*Central_Steppe_MLBA*	..	0.021	0.836	0.164	..	0.026	0.026	..
Indus_Periphery_Pool	*Saidu_Sharif_H_o*	*SPGT_o*	0.152	0.327	0.254	0.419	0.121	0.09	0.069

Indus_Periphery_Pool	*Saidu_Sharif_H_o*	*Steppe_MLBA_oBMAC*	0.063	0.592	0.23	0.178	0.096	0.088	0.027
Indus_Periphery_Pool	*Saidu_Sharif_H_o*	*Western_Steppe_MLBA*	0.043	0.652	0.191	0.157	0.093	0.088	0.024
Indus_Periphery_Pool	*Saidu_Sharif_H_o*	*Central_Steppe_MLBA*	0.067	0.651	0.182	0.167	0.092	0.087	0.026
Parkhai_Anau_EN	*Saidu_Sharif_H_o*	*SPGT_o*	0.055	0.118	0.482	0.4	0.056	0.058	0.093
Parkhai_Anau_EN	*Saidu_Sharif_H_o*	*Western_Steppe_EMBA*	0.017	0.247	0.612	0.141	0.037	0.038	0.033
Parkhai_Anau_EN	*Saidu_Sharif_H_o*	*Steppe_LBA*	0.099	0.285	0.528	0.187	0.033	0.045	0.038
Parkhai_Anau_EN	*Saidu_Sharif_H_o*	*Steppe_MLBA_oBMAC*	0.023	0.231	0.626	0.144	0.039	0.037	0.033
Parkhai_Anau_EN	*Saidu_Sharif_H_o*	*Central_Steppe_MLBA*	0.016	0.248	0.622	0.13	0.037	0.037	0.031
Saidu_Sharif_H_o	*Sarazm_EN*	*SPGT_o*	0.234	0.474	0.234	0.292	0.053	0.079	0.104
Saidu_Sharif_H_o	*Sarazm_EN*	*Steppe_LBA*	0.111	0.511	0.381	0.108	0.047	0.044	0.042
Saidu_Sharif_H_o	*Sarazm_EN*	*Steppe_MLBA_oBMAC*	0.192	0.56	0.338	0.102	0.041	0.052	0.036
Saidu_Sharif_H_o	*Sarazm_EN*	*Western_Steppe_MLBA*	0.194	0.555	0.357	0.088	0.041	0.048	0.031
Saidu_Sharif_H_o	*Sarazm_EN*	*Central_Steppe_MLBA*	0.186	0.552	0.355	0.093	0.041	0.048	0.032
Saidu_Sharif_H_o	*Shahr_I_Sokhta_BA1*	*Steppe_LBA*	0.063	0.49	0.333	0.177	0.048	0.039	0.038
Saidu_Sharif_H_o	*Shahr_I_Sokhta_BA1*	*Steppe_MLBA_oBMAC*	0.014	0.591	0.272	0.136	0.041	0.047	0.033
Saidu_Sharif_H_o	*Steppe_LBA*	*Tepe_Hissar_C*	0.030	0.553	0.178	0.269	0.045	0.038	0.032

Table S 39 Proximal models for *Bustan_BA_o2*

The transition to the Bronze Age from the Copper Age in Iran and Turan coincided with the formation of larger urban settlements. Our analysis supports the hypothesis that these BA urban settlements were formed by the amalgamation of groups with ancestry such as what we observed in the Copper Age. The main *BMAC* genetic cluster as well as the individuals from Shahr-i-Sokhta show considerable admixture, and while the primary source of ancestry remains that related to farmers from Iran, there is also significant input from the west bringing ancestry related to Anatolian farmers; from the north bringing ancestry related to Steppe pastoralists; and from the south bringing ancestry potentially related to (unsampled) populations from the Indus Valley. Indus Valley seals and ivory have been found at Gonur, so contact between BMAC and the IVC is certainly plausible (*218*). Proto-Elamite tablets abundant in sites in southwestern Iran also appear in Shahr-I-Sokhta documenting that there was cultural exchange there with sites from the west (*219*).

The outlier individuals in our study are particularly informative for three reasons. First, the outlier groups have significantly higher proportions of *AHG*, *Anatolia_N* and *WSHG*-related ancestry respectively—all ancestries we observe in different proportions in the majority of individuals from BMAC sites and Shahr-i-Sokhta. This suggests that these urban settlements were cosmopolitan and that admixture between individuals from these sites and from other sites from the south, west and north was ongoing during the time period of our sampling.

Second, other than the last two outliers from Bustan, the genetic ancestry of the outliers is characteristic not just of a single site, but is shared across multiple sites in Turan. This suggests that these outlier individuals are reflecting a process of movement into the region from nearby regions; not site-specific phenomena.

Third, we find that one of the outliers, *Bustan_BA_o2*, is consistent with being admixed between an individual related to people on the *Indus Periphery Cline* and Middle to Late Bronze Age Steppe pastoralists, a type of admixture event we also observe in the Late Bronze-Iron Age Swat Valley that we will examine later, suggesting that the admixture events that led to the formation of the *SPGT* in Pakistan also occurred between outlier individuals at the BMAC and Steppe pastoralists who arrived at the end of the 2nd millennium.

Taken together with the data we have from the Copper Age period, the BMAC sites also suggest that there was not a large-scale influx of ancestry ultimately derived from the Early Bronze Age Yamnaya Steppe pastoralists into Central Asia prior to around 2000 BCE. The clearest evidence for this is that the BMAC sites before this period do not seem to harbor individuals with significant increases in Steppe pastoralist-related ancestry compared with the populations from the pre-Copper Age period. It is possible of course that such ancestry may have arrived earlier and that it is not visible in the urban context of the BMAC sites with its different economic base. However, the outlier individual from Gonur with northern ancestry also documents admixture with a population related to *WSHG*s and but not the Steppe pastoralist cultures of the Yamnaya and Andronovo culture horizons, as these have additional Anatolian farmer-related ancestry and ancestry related to *WEHG* as we describe below. It is only in the outliers from the late to middle 2nd millennium that we observe admixture with Yamnaya- related sources, which corresponds to the archaeological record of the timing of the arrival of Steppe pastoralists in large numbers into this region. We observe this qualitatively also on the West Eurasian PCA plot (**Fig S 10**; top right panel), where *Gonur1_BA_o* individuals appear on an admixture cline between Iranian farmers and *Central_Steppe_EMBA* or *Okunevo_BA.SG*. However, all subsequent outliers with additional ancestry related to *ANE* are drawn to Middle to Late Bronze Age Steppe pastoralists.

Our data from the site of Parkhai—-on the border between present-day Turkmenistan and Iran—is particularly valuable because it allows us to trace a time transect from a single location over the key period: we have genetic clusters of individuals with mean dates of 3550 BCE, 2600 BCE, 2189 BCE, and 1455 BCE. We observe from PCA (**Fig S 10**; top panel) that there are no shifts in Steppe pastoralist-related ancestry until the late Bronze Age (that is, the individual dated to a mean of 1455 BCE). To examine this pattern more broadly over Turan, we plotted the Steppe genetic affinity (as measured by an appropriate f_4-statistic) of all individual from Copper Age and Bronze Age Turan along with their mean dates (**Fig S 25**). This analysis shows that there was little change in Steppe pastoralist-related ancestry until the time of the later BMAC towns, a pattern that reflects the arrival of Steppe pastoralist-related ancestry into Turan at the beginning of the 2nd millennium BCE. In the next section, we carry out similar analyses in the Steppe region itself, showing that the timing of the arrival of Steppe pastoralist-related ancestry in the Turan region corresponds with ancestry changes in the Steppe region at a similar time. Thus, both of these regions were affected around the same time by large-scale gene flow from people with ancestry ultimately deriving from the *Western_Steppe_EMBA* genetic cluster (via successor cultures that carried this ancestry in admixed form like *Western_Steppe_MLBA* and *Central_Steppe_MLBA*).

Fig S 25 Allele sharing between Steppe pastoralists and Copper and Bronze Age individuals from Turan plotted again mean individual age. More negative values indicate greater sharing. We show one standard error bars.

Finally, we examine two outlier individuals from the Bronze Age from Hajji Firuz in western Iran. Our distal modeling suggests that the individuals have about 30% of their ancestry from an *EEHG*-related population (**Table S 40**). With proximal modeling, we found further confirmed that the gene flow from the south into the Steppe was not unidirectional, as this analysis shows that *EEHG* derived ancestry (carried by a group related to *Western_Steppe_EMBA*) also spread south of the Caucasus into Iran (**Table S 41**). Further confirmation of this comes from an intrusive burial in the IA from Hajji Firuz, which also carries ancestry related to *Western_Steppe_EMBA* as well as an R1b1a2 Y-chromosome, found at high frequency in *Western_Steppe_EMBA* populations (**Table S 42** and **Table S 43**). We did not carry out more detailed modeling of this individual due to a lack of data from that time period from neighboring areas.

				Mixture proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
EEHG	*Anatolia_N*	*Ganj_Dareh_N*	0.12	0.30	0.44	0.26	0.02	0.03	0.04

Table S 40 Distal models for *Hajji_Firuz_BA*

				Mixture proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Anatolia_C	*Hajji_Firuz_C*	*Western_Steppe_EMBA*	*0.026*	0.285	0.204	0.511	0.081	0.083	0.03
Baden_LCA	*Hajji_Firuz_C*	*Khvalynsk_EN*	*0.053*	0.126	0.5	0.374	0.033	0.037	0.019
Baden_LCA	*Hajji_Firuz_C*	*Western_Steppe_EMBA*	*0.012*	0.144	0.328	0.528	0.032	0.042	0.028

Table S 41 Proximal models for *Hajji_Firuz_BA*

						Mixture Proportions					SE				
Source1	**Source2**	**Source3**	**Source4**	**Source5**	**P**	**S1**	**S2**	**S3**	**S4**	**S5**	**S1**	**S2**	**S3**	**S4**	**S5**
EEHG	*PPN*	*Anatolia_N*	*Ganj_Dareh_N*	*ESHG*	0.25	0.15	0.23	0.25	0.29	0.10	0.03	0.10	0.11	0.08	0.03
EEHG	*Anatolia_N*	*Ganj_Dareh_N*	..	..	0.01	0.08	0.41	0.51	..	..	0.02	0.04	0.04	..	..
WSHG	*Anatolia_N*	*Ganj_Dareh_N*	..	..	0.05	0.12	0.48	0.41	..	..	0.02	0.04	0.05	..	..

Table S 42 Distal models for *Hajji_Firuz_IA*

				Mixture proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Anatolia_C	*Seh_Gabi_C*	*Western_Steppe_EMBA*	0.069	0.132	0.678	0.189	0.053	0.051	0.029
Anatolia_C	*Seh_Gabi_C*	*Steppe_LBA*	0.028	0.164	0.667	0.169	0.053	0.052	0.027
Anatolia_C	*Western_Steppe_EMBA*	*Tepe_Hissar_C*	0.075	0.381	0.102	0.518	0.038	0.032	0.039
Anatolia_C	*Steppe_LBA*	*Tepe_Hissar_C*	0.236	0.382	0.106	0.512	0.037	0.028	0.038
Anatolia_C	*Steppe_MLBA_oBMAC*	*Tepe_Hissar_C*	0.034	0.356	0.111	0.533	0.044	0.04	0.038
Anatolia_C	*Western_Steppe_MLBA*	*Tepe_Hissar_C*	0.032	0.343	0.105	0.552	0.047	0.037	0.036

Anatolia_C	*Central_Steppe_MLBA*	*Tepe_Hissar_C*	0.052	0.346	0.109	0.545	0.045	0.036	0.036
Anatolia_EBA	*BMAC*	*Western_Steppe_EMBA*	0.404	0.366	0.55	0.085	0.032	0.041	0.032
Anatolia_EBA	*BMAC*	*Steppe_LBA*	0.356	0.372	0.552	0.076	0.031	0.041	0.03
Anatolia_EBA	*BMAC*	*Steppe_MLBA_oBMAC*	0.313	0.345	0.566	0.089	0.035	0.039	0.037
Anatolia_EBA	*BMAC*	*Western_Steppe_MLBA*	0.304	0.337	0.582	0.082	0.036	0.035	0.034
Anatolia_EBA	*BMAC*	*Central_Steppe_MLBA*	0.342	0.341	0.576	0.083	0.035	0.036	0.033
Armenia_EBA	*Seh_Gabi_C*	*Western_Steppe_EMBA*	0.158	0.361	0.503	0.135	0.128	0.103	0.039
Armenia_EBA	*Seh_Gabi_C*	*Steppe_LBA*	0.159	0.422	0.462	0.117	0.113	0.095	0.033
Armenia_EBA	*Shahr_I_Sokhta_BA1*	*Central_Steppe_MLBA*	0.553	0.725	0.275	..	0.067	0.05	0.04
BMAC	*Hajji_Firuz_C*	*Western_Steppe_EMBA*	0.345	0.246	0.619	0.135	0.066	0.057	0.032
BMAC	*Hajji_Firuz_C*	*Steppe_LBA*	0.188	0.248	0.63	0.123	0.067	0.057	0.031
BMAC	*Hajji_Firuz_C*	*Steppe_MLBA_oBMAC*	0.193	0.3	0.558	0.142	0.06	0.059	0.035
BMAC	*Hajji_Firuz_C*	*Western_Steppe_MLBA*	0.175	0.336	0.537	0.127	0.057	0.06	0.032
BMAC	*Hajji_Firuz_C*	*Central_Steppe_MLBA*	0.229	0.32	0.55	0.129	0.058	0.059	0.032
Seh_Gabi_C	*Western_Steppe_EMBA*	..	0.012	0.787	0.213	..	0.027	0.027	..
Seh_Gabi_C	*Steppe_MLBA_oBMAC*	..	0.018	0.752	0.248	..	0.031	0.031	..
Seh_Gabi_C	*Western_Steppe_MLBA*	..	0.017	0.775	0.225	..	0.028	0.028	..
Seh_Gabi_C	*Central_Steppe_MLBA*	..	0.040	0.772	0.228	..	0.028	0.028	..
Hajji_Firuz_C	*Western_Steppe_EMBA*	*Tepe_Hissar_C*	0.281	0.614	0.163	0.223	0.06	0.029	0.062
Hajji_Firuz_C	*Steppe_LBA*	*Tepe_Hissar_C*	0.245	0.612	0.153	0.235	0.06	0.027	0.061
Hajji_Firuz_C	*Steppe_MLBA_oBMAC*	*Tepe_Hissar_C*	0.082	0.544	0.179	0.278	0.064	0.033	0.06
Hajji_Firuz_C	*Western_Steppe_MLBA*	*Tepe_Hissar_C*	0.062	0.513	0.163	0.323	0.067	0.031	0.058
Hajji_Firuz_C	*Central_Steppe_MLBA*	*Tepe_Hissar_C*	0.126	0.526	0.166	0.308	0.065	0.03	0.058

Table S 43 Proximal models for *Hajji_Firuz_IA*

In the next section we report an analysis of the individuals directly north of Turan, and in the final section we examine how contact between people with genetic ancestry typical of each of these communities impacted the Indian subcontinent.

S4.4.2 The Forest Zone and Steppe

In this section we focus on genetic analysis of 207 individuals for whom we generated data in this study. These individuals cover a temporal range from the Early Neolithic to the Late Iron Age and span a geographic range from the region from the present-day central Europe to southeastern Tajikistan. Previous work suggests that large-scale spread of Indo-European languages into Europe likely occurred in the early Bronze Age with the spread of ancestry from pastoralist communities from the Steppe such as Yamnaya (*7*, *8*, *30*). In this section we examine the genetic aspects behind the formation of these communities by sampling sites in the Neolithic forest zone, as well as pastoralist communities during the Early and Late

Bronze Age. We also characterize further population transformations toward the end of this period into the Iron Age (**Fig S 26**). As part of this process, we examine the timing of spread of Steppe pastoralist ancestry southward into Iran and eastward toward the Indian subcontinent, as well as contact between people with Steppe pastoralist ancestry and people of the BMAC.

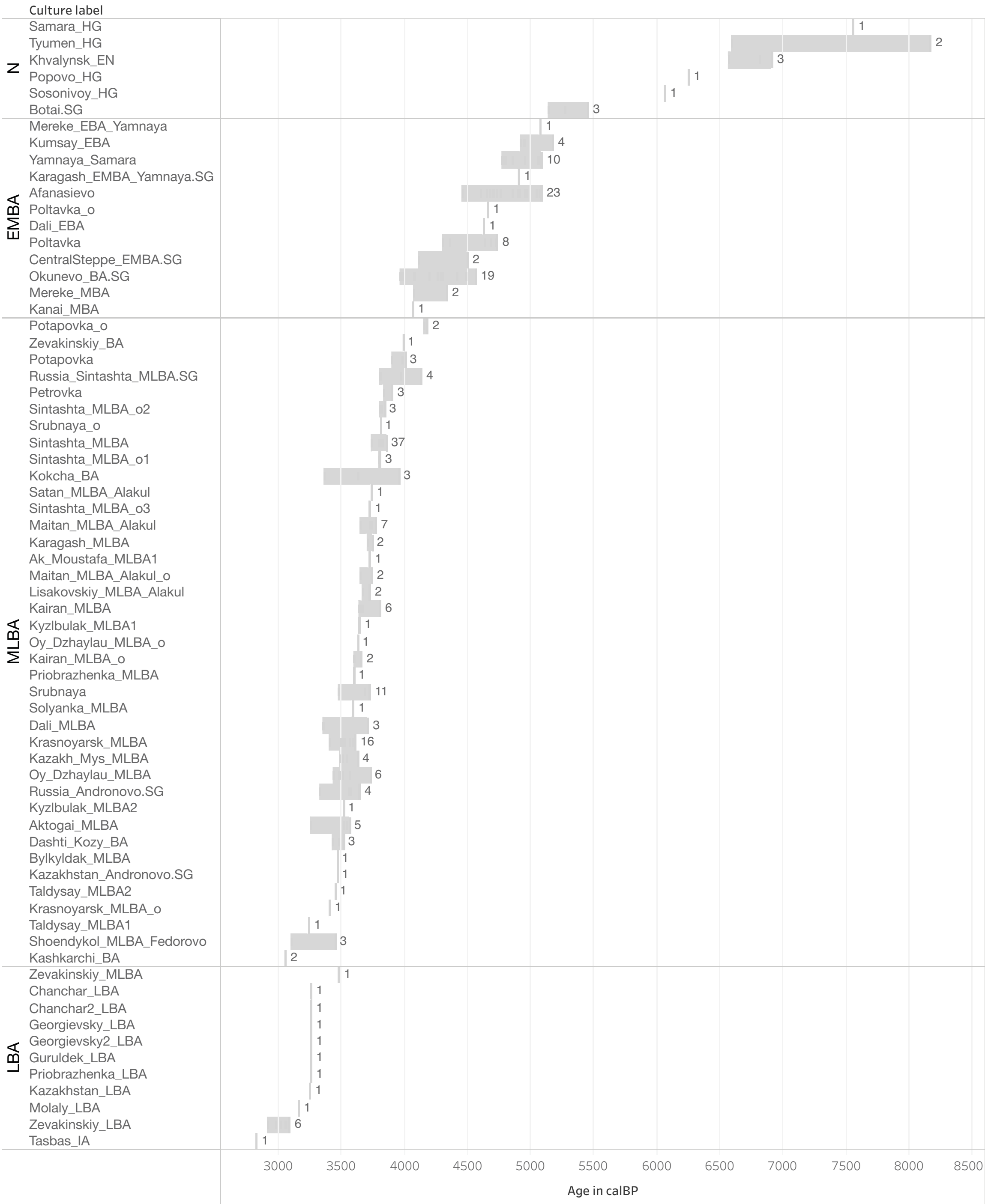

Fig S 26 The number of individuals and distribution of dates from sites analyzed in this section. We report the average of the mean radiocarbon date of individuals from the site, or if not available, the archeological context date.

We divide the genetic history of the samples in this dataset into four temporal phases.

- The first phase is the Neolithic and Eneolithic period from 6000-3000 BCE, for which we have 3 individuals from the Russian forest zone (Forest / Steppe N). We analyze these together with previously published HG individuals from Karelia, Samara and Popovo as well as Eneolithic individuals from Khvalynsk and individuals from individuals associated with the Botai culture of the central Steppe.

- The second phase is the early Bronze Age, dated from 3000-2000 BCE and marked by the spread of the Yamnaya-related Steppe pastoralist cultures (*Western_Steppe_EMBA*). Our data allow us to study how ancestry derived from the Yamnaya or their close genetic relatives spread southward and eastward. We also report data from BA populations from central Kazakhstan from the sites of Kumsay and Mereke. We coanalyze these with individuals from the eastern Steppe and the Altai mountain who have been assigned to the Okunevo archaeological culture.

- The third phase corresponds to the Middle to Late Bronze Age and the so-called Andronovo cultural horizon from around 2200-1500 BCE. We track the spread of Yamnaya-related ancestry (in admixed form) into the regions occupied by the BMAC archeological complex discussed in the previous section. These individuals span the Middle to Late Bronze Age for which we use the abbreviation *Steppe_MLBA* (sometimes preceded by a prefix or suffix to indicate subtypes of ancestry).

- The fourth phase corresponds to the Late Bronze Age and genetic transformation that occurred after 1500 BCE and into the Iron Age. For simplicity we call individuals from this period *Steppe_LBA* (sometimes preceded by a prefix or suffix to indicate subtypes of ancestry).

S4.4.2.1 <u>The Neolithic and Eneolithic periods in the Steppe and Russian Forest Zone</u>

As with the previous analysis of the populations from Iran and Turan, we begin by examining descriptive statistics of the data by means of PCA and ADMIXTURE (**Fig S 27**). All the individuals from this period fall in the top left of the All Eurasian PCA and top right of the

West Eurasian PCA, consistent with geography in each of these plots. Within these regions, we observe multiple ancestry clines that direct themselves toward populations from East Asia, Europe and Iran. On the ADMIXTURE plot, all individuals from the Steppe also have a significant orange component that makes up a large percentage of their characteristic assigned ancestry components.

We first examine the ancestry of three HGs from the Russian forest zone dated to around 5000 BCE from Tyumen Oblast. These individuals, the oldest from the Central Forest Zone for which we have data, occupy the top right position on the West Eurasian PCA and have the maximal amount of the orange component among all the individuals analyzed here, suggesting that all later populations descend at least in part from groups related to these. To understand their ancestry, we analyze them together with previously published HGs from further west in Karelia and Samara and HGs from Lake Baikal (*Shamanka_EN.SG*). On both the West Eurasian and All Eurasian PCA, as well as in ADMIXTURE, these individuals appear on a genetic cline with the individuals from Eastern Europe on the extreme left, East Asians on the extreme right, and *WSHG*s lying between them (**Fig S 27**).

To test these observations more formally, we examine the relationships of all the individuals analyzed here using “Pre-Copper Age Affinity” and “Two Population Comparison” f_4-statistics. As a basis for comparison, we compute statistics for all populations with respect to *AfontovaGora3*, an approximately 15,000 BCE individual from Upper Paleolithic Siberia reflecting deep ancestry from the region. We observe that the Neolithic and Eneolithic individuals give the most extreme values for both statistics, with higher degrees of *AfontovaGora3* affinity when compared with all other individuals (**Fig S 28**). Computing admixture-f_3 statistics, we find no significant signals of admixture from sources related to any ancient groups from whom we have genetic data.

We attempted to model the Neolithic forest zone individuals with respect to other published populations. Since the population we would like to model here, *WSHG*, is a part of the outgroup set of our standard modeling procedure we remove that population as a part of our outgroup set and attempted to model its ancestry with respect to the others. We also wanted to investigate the contribution that Upper Paleolithic populations from Siberia had on *WSHG*s, and therefore we chose to use one of our individuals that make up the *ANE*, *AfontovaGora3* as a source while keeping *MA1_HG.SG* as an outgroup population.

RightWSHG: Ethiopia_4500BP.SG, Ust_Ishim_HG_published.DG, Kostenki14, MA1_HG.SG, AHG, Anatolia_N, Ganj_Dareh_N, WEHG, EEHG, ESHG

We then utilize our procedure for Distal modeling to determine a set of populations that can be moved to the *Left* as sources that provide a good fit. While no model fits according to our acceptance criteria, one model works marginally with p-value 0.003, with the majority of the ancestry derived from *ANE* (**Table S 44**). We note that this is plausibly to be due to a small but minor contribution by an *AfontovaGora3*-related population to *Ganj_Dareh_N*, as models including it pass our p-value threshold but the admixture proportion assigned to it remains consistent with 0, indicative that including it in the outgroup set may be violating the assumption of lack of gene flow from the source populations into the outgroups directly.

Importantly, the *WSHG* population shows no discernable ancestry from groups related to Anatolian farmers or *WEHG*s, suggesting that these inhabitants of the Russian forest zone were not admixed with populations related to those in Western Europe or the Near East. However, a significant component of their ancestry (point estimate of 6%) is from a population related to present-day East Asians. This is different from the pattern in the *Western_Steppe_EMBA* genetic cluster that spread over this region in the Bronze Age who require no additional East Asian related ancestry for modeling (see below). This suggests that the *Western_Steppe_EMBA* had little if any ancestry from the *WSHG* groups that preceded them.

				Mixture Proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
EEHG	*AfontovaGora3*	*ESHG*	0.003	0.20	0.73	0.06	0.02	0.03	0.01

Table S 44 Modeling deep ancestry proportions for *WSHG*

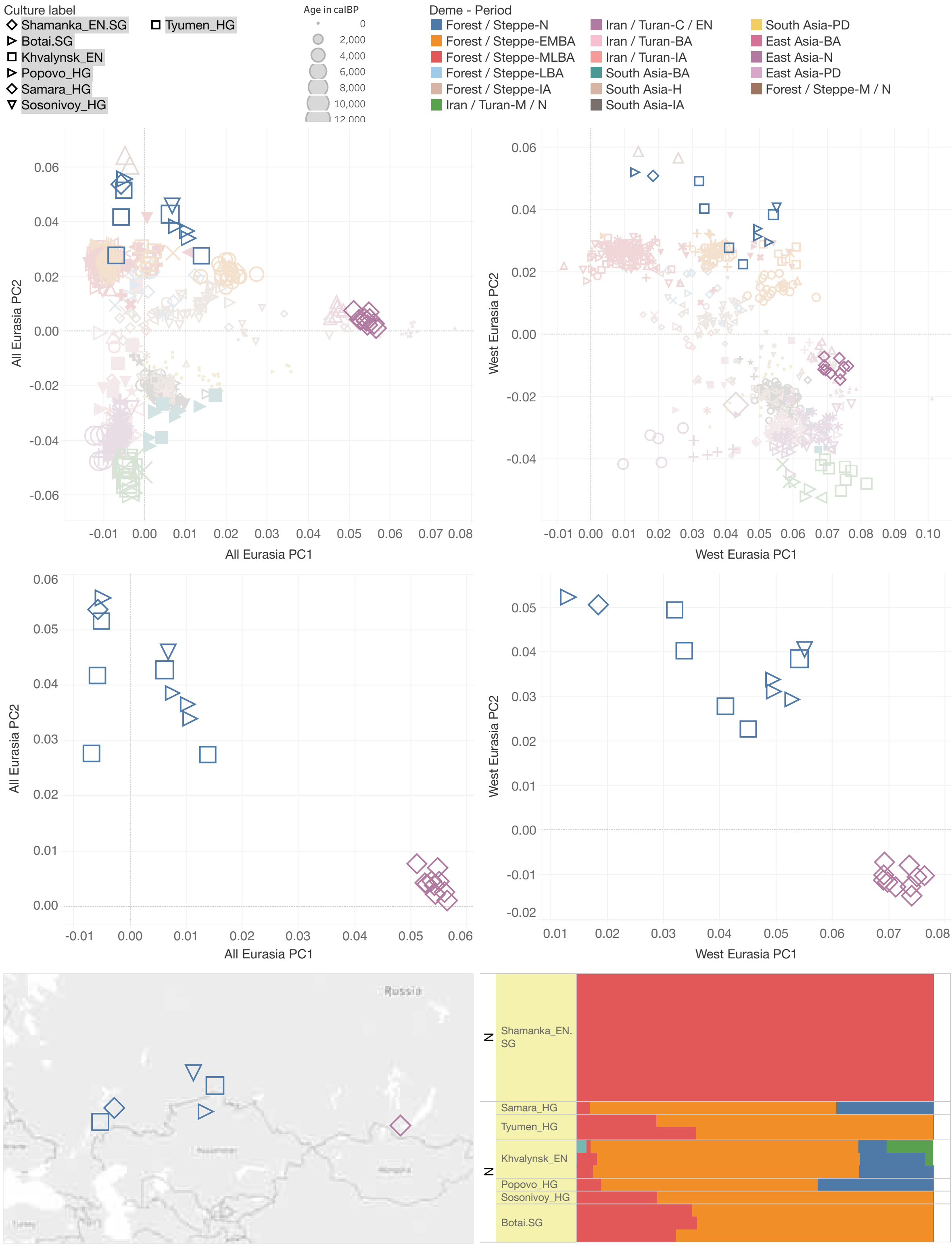

Fig S 27 Top panel: left, West Eurasian PCA; right, All Eurasian PCA. Middle panel: left, blow up of the West Eurasian PCA; right, blow up of the All Eurasian PCA. Bottom panel: left, geographic locations of individuals; right, ADMIXTURE results, with the colors of green, teal, orange, red and blue representing ancestry maximized in

Iranian farmers from the Zagros, farmers from Anatolia, and HGs from West Siberia, East Siberia, and Western Europe respectively.

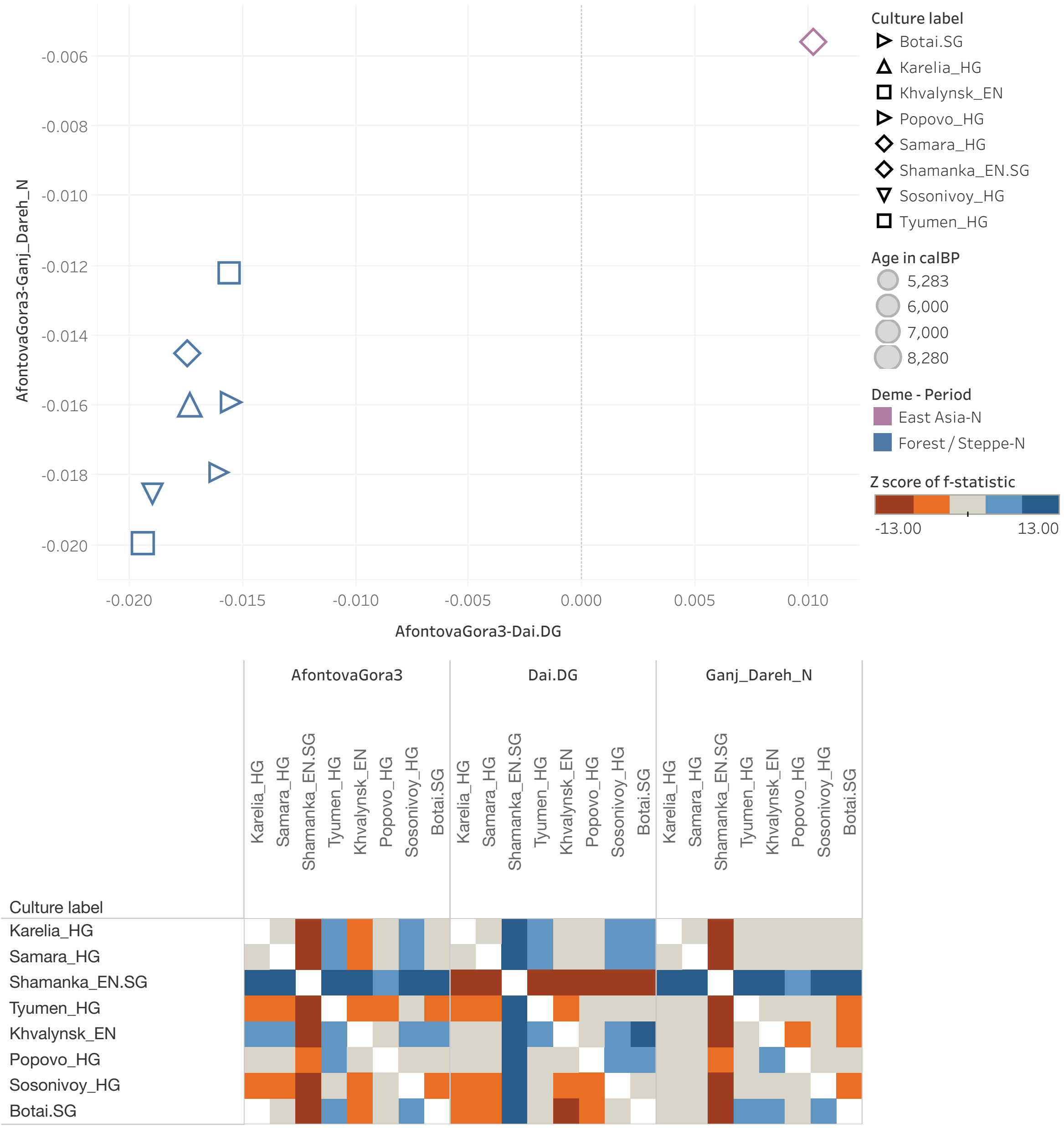

Fig S 28 Top panel: "Pre-Copper Age affinity" *f*$_4$-statistics comparing the statistics *f*$_4$*(Ethiopia_4500BP.SG, Test, AfontovaGora3, Ganj_Dareh_N)* vs. *f*$_4$*(Ethiopia_4500BP.SG, Test, AfontovaGora3, Dai.DG)*. Bottom panel: "Two Population Comparison" *f*$_4$-statistics showing differences in affinity to *ANE*, Eastern Non-African-related and Iranian farmer-related ancestry between populations. Orange squares indicate |Z-scores| > 3 for sharing more alleles with the test population compared to a population from the columns, with respect to an outgroup. Blue squares indicate significantly less sharing at the same threshold. Darker orange and blue squares show sharing at |Z-score| > 6.

To understand population transformations occurring on the western Steppe, we also analyze 3 individuals from the Steppe zone of Samara, *Khvalynsk_EN*, dated to 5th millennium BCE that were published in a previous study (*19*). On the ADMIXTURE plot, the Khvalynsk individuals show additional components that are maximized in Iranian farmers in green, and not seen in either *EEHG*s or *WSHG*s (**Fig S 27**). The "Pre-Copper Age Affinity" f_4-statistics show that the individuals from Khvalynsk share excess ancestry with Iranian populations, which also supports such admixture (**Fig S 28**). We also examined admixture-f_3 statistics for the Eneolithic individuals from Khvalynsk (**Fig S 29**). The Khvalynsk individuals show signals of admixture between groups related to Ancient North Eurasians as represented by *AfontovaGora3*, as well as with hunter-gatherer and farmer groups from Europe and the Caucasus. While these significant admixture-f_3 signals with European farmer groups might appear to be in conflict with our ADMIXTURE results, there is in fact no conflict. Instead, these results reflect the fact that European farmers harbor large proportions of European hunter-gatherer-related ancestry which is likely to be driving the signal. We examine these more closely using *qpAdm*, our formal modeling procedure.

To relate these patterns to those seen in populations from a similar age from the central Steppe, we co-analyzed these individuals with other published individuals from the Botai culture (*30*). These individuals lie between *WSHG*s and *ESHG*s, and on the ADMIXTURE plots show the similar components as *WSHG* albeit with a larger proportion of ancestry assigned to the red component. This suggests that the ancestry of the individuals living in the central Steppe at the beginning of the third millennium BCE was similar to that of HG individuals from the Forest Zone roughly a millennium earlier, although this does not mean that the individuals of the Botai culture are directly descended from them. Using admixture-f_3 statistics we see that individuals from Botai are admixed between the *ANE* and Eastern Non-Africans, suggesting that there was admixture into the central Steppe between the time of the *ANE* and the time of the West Siberian Hunter-Gatherers and the Botai culture.

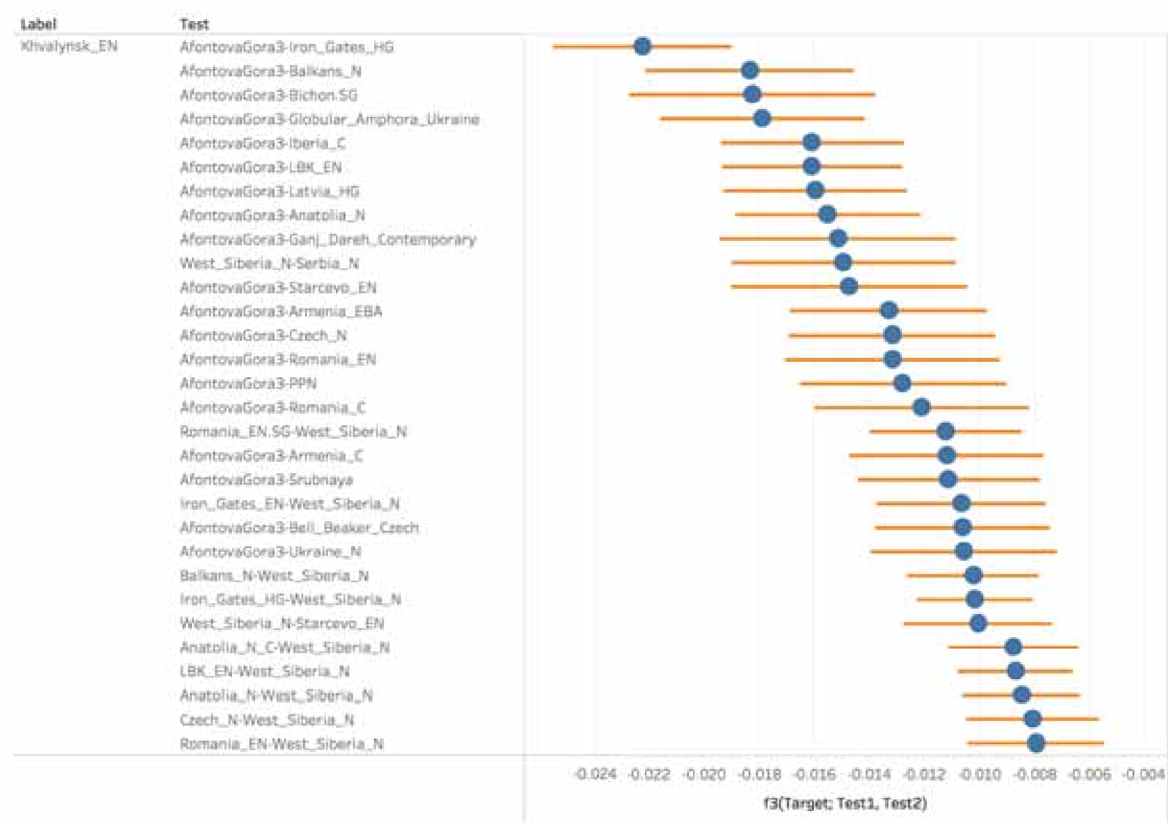

Fig S 29 Top admixture-f_3 pairs for *Khvalynsk_EN*

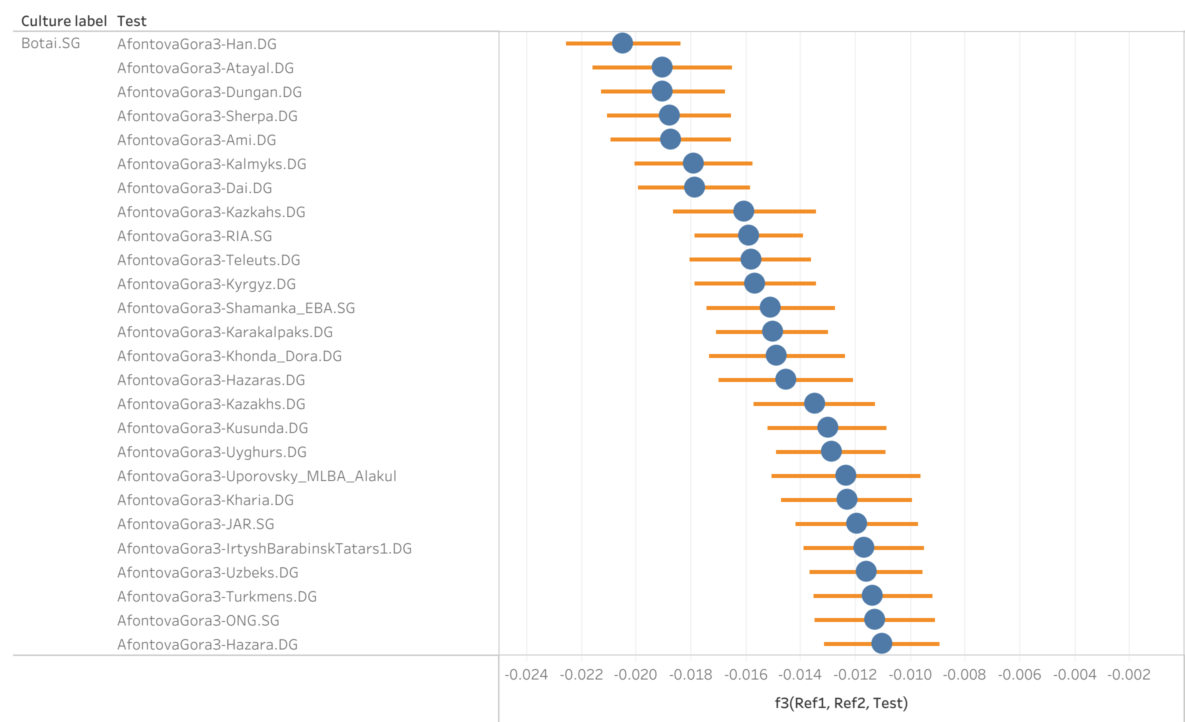

Fig S 30 Top admixture-f_3 pairs for *Botai.SG*

With this understanding, we modeled the ancestry of both of these populations with respect to a set of pre-Copper Age surrogates for the source populations. The final models show that *Khvalynsk_EN* can be modeled as a mixture of ~80% *EEHG*-related and ~20% ancestry related to early Iranian farmers. These results are qualitatively consistent with the observations in refs. (*25*) and (*7*), where a different *qpAdm* model-fitting protocol and different sets of outgroups were used (**Table S 45**). We do not have a model for *Botai.SG* that works given the thresholds for our modeling procedure. This could be due to the differential processing of the Botai genome (non-UDG-treated shotgun genomes) compared to the data we generated (UDG-treated 1.24-million SNP capture data). Such differences in processing can easily cause artifacts in *f*-statistics. However, the best fitting model reflects our observations of bring primarily comprised of *WSHG*, with an additional 10% ancestry related to *ESHG*, a model that is similar to that in the original report of data from these individuals (**Table S 46**).

				Mixture Proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
EEHG	..	*Ganj_Dareh_N*	0.03	0.81	..	0.19	0.02	..	0.02
EEHG	*PPN*	*Ganj_Dareh_N*	0.20	0.10	0.08	0.11	0.02	0.03	0.03

Table S 45 Distal models for *Khvalynsk_EN*

			Mixture Proportions		SE	
Source1	**Source2**	**P**	**S1**	**S2**	**S1**	**S2**
WSHG	*ESHG*	0.001	0.9	0.1	0.01	0.01

Table S 46 Distal models for *Botai.SG*

S4.4.2.2 The Early Bronze Age: Formation and decline of Steppe EMBA groups

We now examine the second temporal phase of genetic transformation on the Steppe dating from 3300-2500 BCE. The PCA and ADMIXTURE plots show that all the individuals from the EMBA fall into 3 main clusters that correspond roughly to geographic locations on the western Steppe (*Yamnaya_Samara, Potapovka and Poltavka*), central Steppe (*Kumsay_EBA, Mereke_EBA* and *Dali_EBA* and *Karagash_EMBA_Yamnaya.SG*), and eastern Steppe (*Afanasievo, Okunevo_BA.SG* and *Kanai_MBA*). A first observation is that these groups show increasing evidence of allele sharing with Eastern Non-Africans further to the east. This correlates with what we observed in the Neolithic, with an eastward cline of increasing East Asian-related ancestry beginning with *EEHG* and extending to *ESHG*. A second observation is that the individuals from the EMBA are shifted “southward” on both the All Eurasian PCA and West Eurasian PCA toward Anatolian and Iranian farmer populations (*7*). This observation is also consistent with the ADMIXTURE plots which show additional ancestry components in teal and green in individuals from the western and central Steppe, suggesting that during the BA in contrast with the Neolithic, populations from across the Steppe were admixing with communities from the south.

We begin our PCA and ADMIXTURE analysis by examining the genetic ancestry typical of the archaeological cultures of the western Steppe and document a series of genetic transformations. From the HG period to the Eneolithic, we observe that the individuals from Khvalynsk have an increase in ancestry related to that of Iranian farmers. From the Eneolithic into the BA, we observe that *Yamnaya_Samara*, *Poltavka* and *Potapovka* individuals are all genetically similar to each other and compared to the Eneolithic Khvalynsk individuals have additional ancestry related to both Iranian and Anatolian farmers (**Fig S 31**). The ADMIXTURE ancestry components are mostly similar across individuals from the same archaeological culture, with notable exceptions. The Poltavka outlier, directly radiocarbon dated to a mean of 2708 BCE, is of particular interest, as it shows significant ancestry related to Anatolian farmers, implying that the movement of this ancestry onto the steppe had already begun by this time (this is the oldest documented evidence of ancestry similar to that of the Corded Ware Complex for which we have both a radiocarbon date as well as ancient DNA confirming the ancestry proportions.). The individuals from the later Potapovka culture also appear to have additional ancestry related to Anatolian farmers based on the shift toward Western Europe on the PCA as well as an increased proportion of teal colored ancestry (European-associated) on the ADMIXTURE plot. This suggests that the populations of the Steppe were impacted by gene flow into the Steppe over an extended period of up to 4000 years. Late Neolithic western Iranian individuals (for instance from the sites of Hajji Firuz and Seh Gabi) have substantial Anatolian farmer-related ancestry as we have already shown (**Table S 17** - **Table S 18**), but we are uncertain if the Anatolian farmer-related flow into the *Western_Steppe_EMBA* arrives via Iran and the South Caucasus or via flow from Eastern Europe and the Ukraine.

In the central Steppe, we analyzed data from multiple locations across Kazakhstan: from the far west on the border with Russia, to Mereke, to Kumsay, to Dali in the southeast. When compared with the HGs from slightly north of Kazakhstan, and individuals from the Botai culture, we see that these individuals have increased allele sharing with Iranian farmers, as reflected in their shift toward the south on both PCA plots and the presence of a green component on the ADMIXTURE plot that is absent in the previous time period. This suggests that peoples of the central Steppe—like those of the western Steppe—experienced gene flow from agricultural communities from the south. Consistent with its position in the far east end of Kazakhstan, the individual from Dali appears to have the highest proportion of ancestry related to *ESHG* and is shifted slightly to the right of the other individuals from western and central Kazakhstan on the PCA plots and has a higher proportion of the red component (East Asian-associated) in ADMIXTURE analysis. Of particular interest is an individual located geographically in the central Steppe *Karagash_EMBA_Yamnaya.SG*—that appears on the PCA and ADMIXTURE plots to be genetically similar to individuals from the Yamnaya and Poltavka cultures of the western Steppe.

In the far east in the Altai region, we examine a set of individuals from the northeast of Kazakhstan: *Central_Steppe_EMBA.SG*, a single individual from Kanai, individuals from the BA Okunevo culture and *Afanasievo*. Based on the PCA and ADMIXTURE plots, we observe that the individuals from *Kanai_MBA*, *Okunevo_BA.AG*, and *Central_Steppe_EMBA.SG* are genetically homogenous, and as expected based on the geography, more closely related to *ESHG*s than to *WSHG*s. In contrast to these individuals, individuals from the *Afanasievo* culture appear to be genetically similar to those from the western Steppe, consistent with the hypothesis of population movement from the west to the east that leapfrogged the intervening groups with primarily ancestry related to *Central_Steppe_EMBA* (*8*)

We next examine these individuals using "Pre-Copper Age Affinity" f_4-statistics. We recapitulate our findings from PCA and ADMIXTURE separating the various groups into 3 main clusters. The individuals from the eastern Steppe appear to have significantly higher ancestry related to *ESHG* when compared to the individuals from further west, while the individuals from the west have additional affinity to *EEHG* and *WEHG*. Another notable pattern is that compared to the individuals from before the EMBA (*Khvalynsk_EN* and *Botai.SG*), the later individuals from both the western and central Steppe appear to have considerably higher proportions of ancestry related to *Ganj_Dareh_N*, confirming our observations on the PCA and ADMIXTURE plots (**Fig S 32**) and in agreement with previously reported studies (*30*).

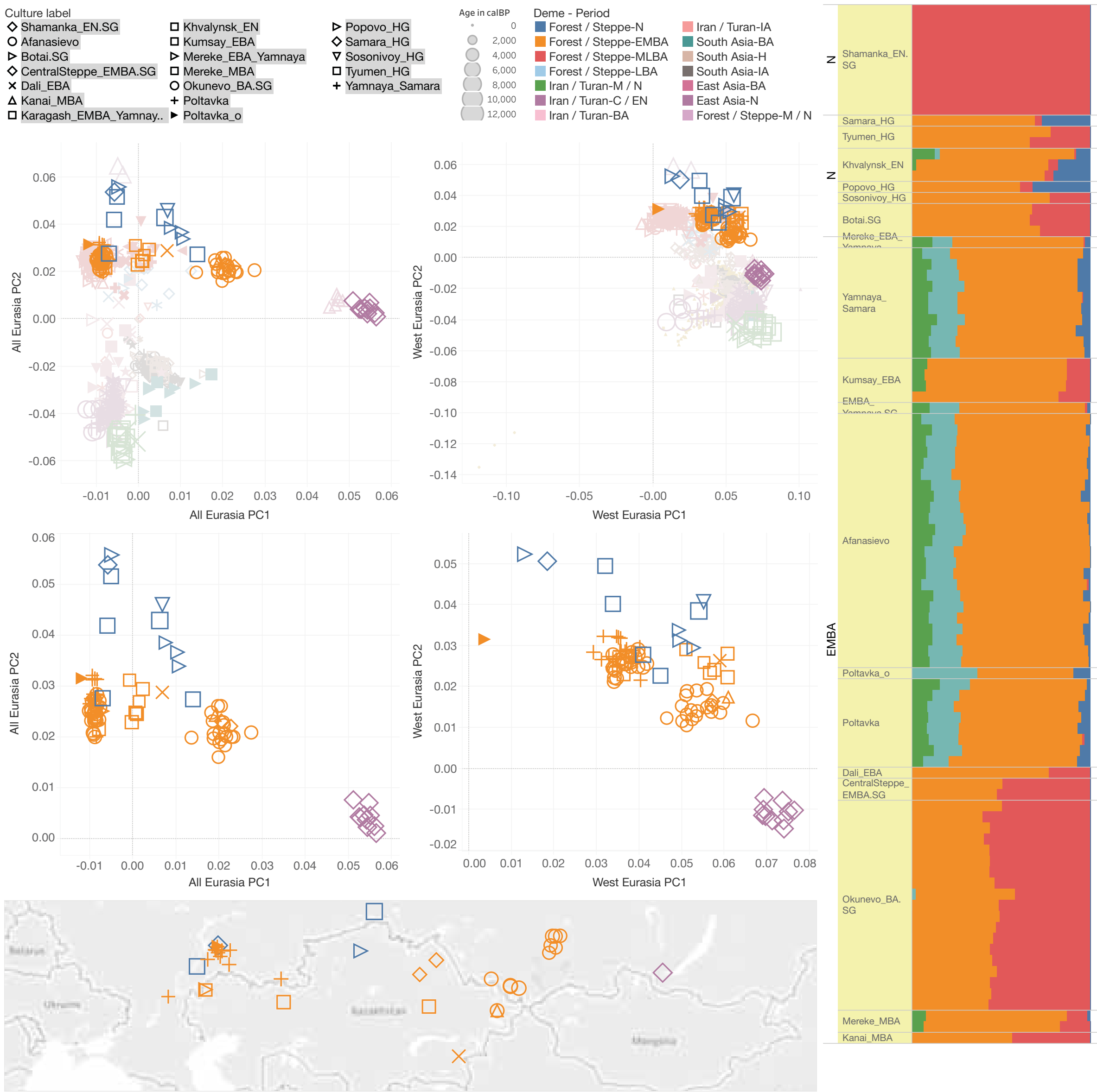

Fig S 31 Top panel: left, West Eurasian PCA; right, All Eurasian PCA. Middle panel: left, blow up of the West Eurasian PCA; right, blow up of the All Eurasian PCA. Bottom panel: geographic locations of individuals. Right panel: ADMIXTURE results, with the colors of green, teal, orange, red and blue representing ancestry maximized in Iranian farmers from the Zagros, farmers from Anatolia, and HGs from West Siberia, East Siberia, and Western Europe respectively.

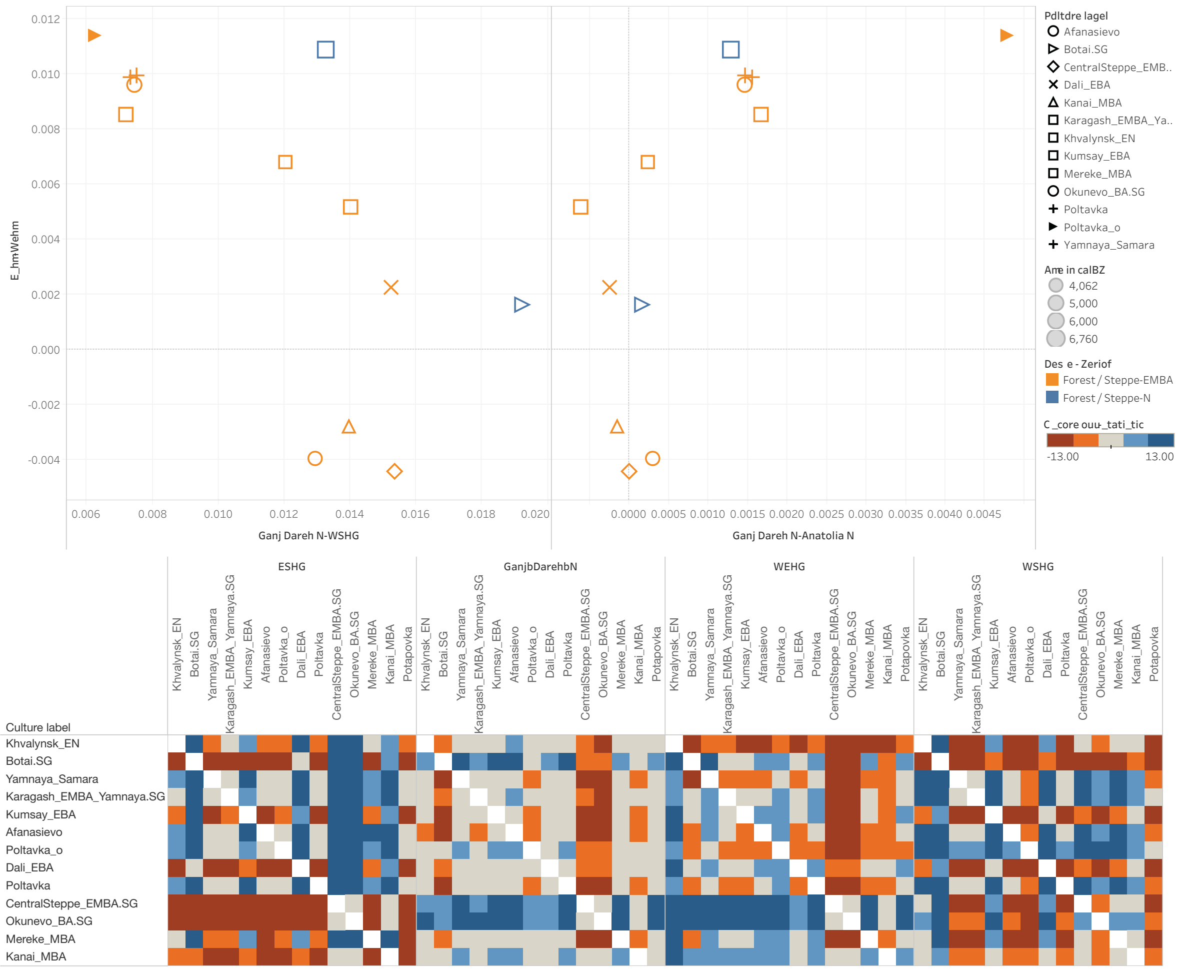

Fig S 32 Top panel: left; "Pre-Copper Age Affinity" f_4-statistics comparing the statistics *f_4(Ethiopia_4500BP.SG, Test, ESHG, WEHG)* vs. *f_4(Ethiopia_4500BP.SG, Test, Ganj_Dareh_N, WSHG)*. Bottom panel: "Two Population Comparison" f_4-statistics showing differences in affinity to *ANE*, Eastern Non-Africans and early Iranian farmers between populations. Orange squares indicate |Z-scores| > 3 for sharing more alleles with the test population compared to a population from the columns, with respect to an outgroup. Blue squares indicate significantly less sharing at the same threshold. Darker orange and blue squares show sharing at |Z-score| > 6.

Prior to the qpAdm modeling of the populations we show admixture-f_3 statistics confirming our findings for each of the 3 sets of analysis populations. We see that the Poltavka and other *Western_Steppe_EMBA* populations appear to be admixed between Eastern European hunter-gatherers or populations descending from Eastern European hunter-gatherers on the Steppe and Iranian and Anatolian farmer populations from south of the Steppe (**Fig S 33**). Similarly, *Mereke_MBA* and other *Central_Steppe_EMBA* populations appear to be admixed between groups descended both from groups related to West Siberian Hunter-Gatherers and from groups with Iranian farmer-related ancestry from Turan (**Fig S 34**).

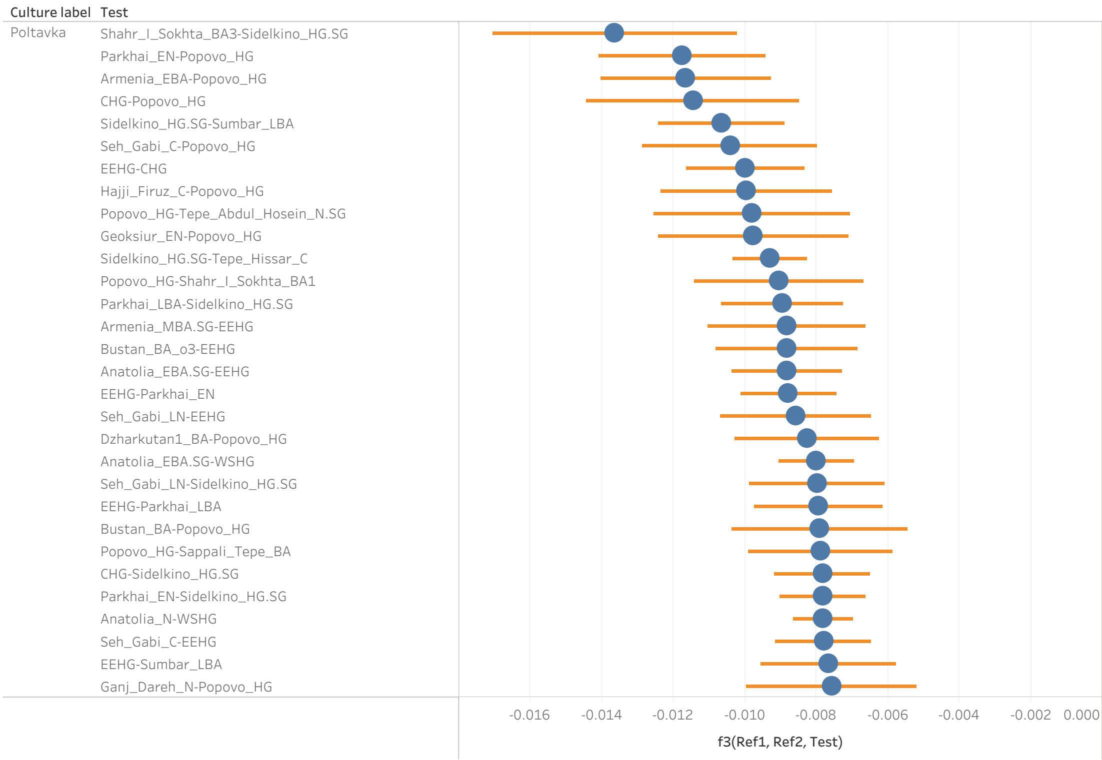

Fig S 33 Top admixture-f_3 statistics for *Poltavka* (to represent other *Western_Steppe_EMBA* groups)

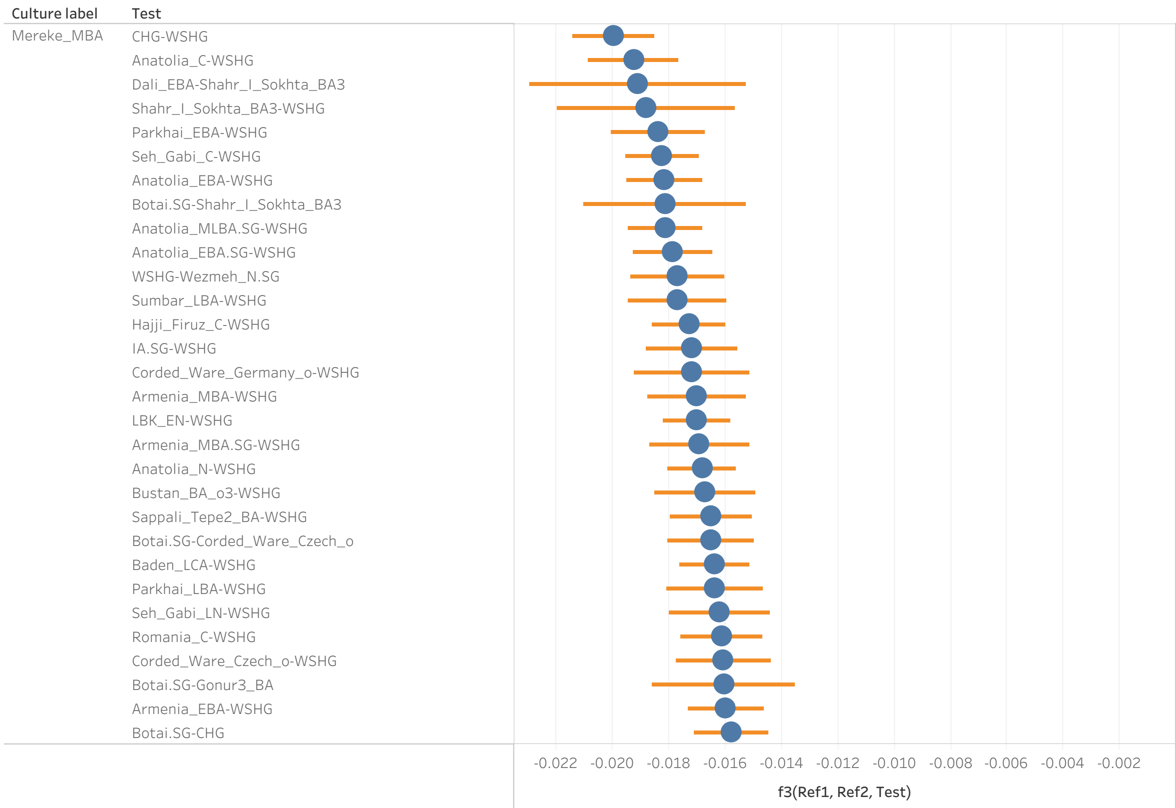

Fig S 34 Top admixture-f_3 statistics for *Mereke_MBA* (to represent other *Central_Steppe_EMBA*) groups

Our distal modeling analysis confirms the finding of ref. (*9*) that *Western_Steppe_EMBA* had about half its ancestry from earlier groups in the Steppe / Forest Zone similar (*EEHG*-related ancestry), with the remainder of their ancestry consistent with being primarily derived from a mixture of Iranian and Anatolian farmers (**Table S 47**). It also suggests that there was continued gene flow after the Neolithic period, as the amount of Iranian farmer- as well as Anatolian farmer-related ancestry increased significantly between the *Khvalynsk_EN* and *Western_Steppe_EMBA*, as previously reported (*49*). We are, however, unable to produce a fitting proximal model for *Western_Steppe_EMBA* using any combination of our source populations. Thus, at present, without additional data from the Caucasus as well as the Steppe from time periods closer to that of the *Western_Steppe_EMBA* samples, we are unable to evaluate if the additional Anatolian farmer-related ancestry in addition to the ancestry seen in the samples from Khvalysnk arrived from an estward migration from Europe as suggested in ref. (*49*) or if alternatively, there was no substantial eastward gene flow from Europe into the ancestors of the Yamnaya, and instead the ancestry was from a population South of the Caucasus with the correct mixture of Anatolian and Iranian farmer-related ancestry that has not yet been sampled (and thus is not available for our proximal modeling). In addition, our distal modeling suggests that *Western_Steppe_EMBA* had 4% of their ancestry derived from *WEHG*, suggesting that a source related to the farmers of Europe, was a part of *Western_Steppe_EMBA* ancestry. However this could be due to the fact the *Western_Steppe_EMBA* populations were from later in time, closer to the MLBA when this type of ancestry spread into the Steppe. Indeed, the Afanasievo culture thought to be derived from the Yamnaya culture, one of the groups part of our *Western_Steppe_EMBA* analysis label/grouping requires no additional *WEHG* admixture despite having similar or greater sample size. Regardless, we view the process of mixture that formed *Western_Steppe_EMBA* and indeed the Yamnaya steppe pastoralists who served as a vector for the spread of this characteristic ancestry across large areas of Eurasia as an unsolved and important question that should be addressed in future studies.

By examining an outlier individual assigned to the Poltavka culture, we also begin to see indications about the type of genetic transformation that occurs on the Steppe during the EMBA leading to the formation of populations that become widespread in the MLBA. In addition to the 3 sources of distal ancestry required to explain the ancestry in *Western_Steppe_EMBA*, distal modeling for these sets of individuals shows that additional ancestry from *WEHG* is required to obtain good fits, suggesting that there was additional gene flow onto the Steppe from the west by this time (**Table S 48**). Proximal models show the Poltavka outlier is a clade with a set of individuals from the MLBA which we describe in more detail in the next section (**Table S 49**).

Distal modeling of the *Central_Steppe_EMBA* individuals shows that almost four-fifths of the ancestry is *WSHG* with additional input from people related to Anatolian and Iranian farmers (**Table S 52**). However, once again our limited sampling makes it difficult to produce proximal models for the ancestry in *Central_Steppe_EMBA* and we do not obtain suitable fits for any combination of source populations.

Distal modeling the ancestry of the *Afanasievo* individuals shows that they are very similar in ancestry to *Western_Steppe_EMBA* (**Table S 50**), consistent with previous reports (*8*). However, as described above, *Western_Steppe_EMBA* and *Afanasievo* are not a genetic clade, and we can show that this is due to an additional 4% ancestry related *WEHG* in our sample of *Western_Steppe_EMBA* individuals as compared with *Afanasievo*. This could reflect a small amount of admixture as Yamnaya pastoralists moved into the central and eastern Steppe, a process in which they would have encountered people with *Central_Steppe_EMBA* ancestry characterized by high percentages of *WSHG* and lower percentages of *WEHG*. While our results here prove that there was significant genetic differentiation between peoples of the Afanasievo and Yamnaya culture which makes up a part of the *Western_Steppe_EMBA* analysis label, it remains the case that the ancestry of people of the Afanasievo culture in the Altai mountain region can largely be explained by a large-scale and long-distance eastward movement of people of Yamnaya ancestry. This finding adds to the work by Damgaard et al. who sampled an individual from a Yamnaya archeological and genetic context at Karagash in Central Kazakhstan which shows a similar pattern (*30*). Unlike *Western_Steppe_EMBA*, we are able to produce a fitting proximal model for the Afanasievo (**Table S 51**) involving Eneolithic samples from the steppe, an Iranian farmer population, and a European farmer population.

Our analysis of a single individual from northwestern Kazakhstan from the site of Mereke, *Mereke_EBA_Yamnaya* is also significant in suggesting that admixture between *Central_Steppe_EMBA* and *Western_Steppe_EMBA* related groups was occurring as early as the end of the 4th millennium or early 3rd millennium BCE (**Table S 53 - Table S 54**).

We also model individuals from the eastern Steppe, *Okunevo_BA.SG*, and observe that about 30% of their ancestry can be attributed to ancestry related to *ESHG* and another 15% to Iranian farmers (**Table S 55**). As with the individuals from the central Steppe, the presence of people with Iranian farmer-related ancestry in the Altai region and Minusinsk Basin suggests that the contact with farmer populations from the south is a feature all across the Steppe, albeit with different source populations from the Steppe and Iran/Turan. However, due to the lack of additional individuals with high proportions of ancestry related to *ESHG,* we were unable to obtain proximal models for *Okunevo_BA.SG*

We end the modeling of individuals from this temporal phase by examining a single individual *Dali_EBA* from southeastern Kazakhstan. Distal modeling suggests that this individual has ancestry related to *WSHGs* with additional ancestry from groups related to early Iranian farmers, while proximal modeling that this individual can be modeled as a mixture between *Botai.SG*-related ancestry and people related to Copper and Bronze Age farmers of Turan (**Table S 56 - Table S 57**). Our observation of this type of ancestry in southeastern Kazakhstan at this time provides an important insight. We know from the geographic distribution of *Afanasievo* and the genetically similar *Western_Steppe_EMBA* groups that people of this ancestry were highly mobile and capable of large-scale geographic movement. However, the lack of evidence for substantial *Western_Steppe_EMBA* admixture in the late Copper Age sites in Turan (which are roughly contemporaneous with the *Afanasievo* culture) as well as in the great majority of individuals we analyzed from BMAC sites shows that their spread had little demographic impact on agricultural settlements to the south in the Early to Middle Bronze Age. The fact that *Dali_EBA*, which was from a site due south of the *Afanasievo* on an ancient migratory route between Turan and Central Asia (*6*) (the Inner Asian Mountain Corridor), thus shows that *Western_Steppe_EMBA* ancestry was not only scarce in the agricultural settlements of Turan but in some of the contemporary hunter-gatherer and pastoralist cultures to its north. The wide spread of ancestry related to *Central_Steppe_EMBA* but not ancestry related to *Western_Steppe_EMBA* across this region in the Early and Middle Bronze Age is also consistent with our observation of an outlier individual from the BMAC site of Gonur which did not have any ancestry related to *Western_Steppe_EMBA* but instead could be modeled as being admixed with ancestry related to *Central_Steppe_EMBA* (**Table S 32**).

					Mixture Proportions				SE			
Source1	**Source2**	**Source3**	**Source4**	**P**	**S1**	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**
EEHG	*WEHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.01	0.51	0.04	0.15	0.30	0.02	0.01	0.02	0.03

Table S 47 Distal models for Western_Steppe_EMBA

					Mixture Proportions				SE			
Source1	**Source2**	**Source3**	**Source4**	**P**	**S1**	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**
EEHG	*WEHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.27	0.27	0.21	0.26	0.26	0.04	0.03	0.04	0.05
WSHG	*WEHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.25	0.25	0.26	0.32	0.17	0.03	0.02	0.04	0.05

Table S 48 Distal models for *Poltavka_o*

Source1	P
Western_Steppe_MLBA	0.060

Table S 49 Proximal models for *Poltavka_o*

				Mixture Proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
EEHG	*Anatolia_N*	*Ganj_Dareh_N*	0.03	0.53	0.18	0.29	0.01	0.02	0.02

Table S 50 Distal models for *Afanasievo*

				Mixture Proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Khvalysnk_EN	*Seh_Gabi_C*	*Globular_Amphora*	0.05	0.55	0.12	0.33	0.02	0.03	0.04

Table S 51 Proximal models for Afanasievo

					Mixture Proportions				SE			
Source1	**Source2**	**Source3**	**Source4**	**P**	**S1**	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**
EEHG	*WSHG*	*PPN*	*Ganj_Dareh_N*	0.44	0.09	0.67	0.09	0.15	0.03	0.03	0.02	0.03
WSHG	*PPN*	*WEHG*	*Ganj_Dareh_N*	0.35	0.73	0.09	0.03	0.14	0.02	0.02	0.01	0.03
WSHG	*WEHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.47	0.76	0.04	0.05	0.15	0.02	0.02	0.03	0.03
WSHG	*PPN*	*Ganj_Dareh_N*	..	0.03	0.76	0.11	0.12	..	0.01	0.02	0.03	..
WSHG	*Anatolia_N*	*Ganj_Dareh_N*	..	0.09	0.79	0.1	0.11	..	0.02	0.02	0.03	..

Table S 52 Distal models for *Central_Steppe_EMBA*

				Mixture Proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
EEHG	*PPN*	*Ganj_Dareh_N*	0.05	0.55	0.12	0.33	0.02	0.03	0.04
EEHG	*WEHG*	*Ganj_Dareh_N*	0.02	0.54	0.05	0.41	0.03	0.03	0.03
EEHG	*Anatolia_N*	*Ganj_Dareh_N*	0.40	0.57	0.07	0.36	0.02	0.03	0.04

Table S 53 Distal models for *Mereke_EBA_Yamnaya*

Mixture Proportions	SE

Source1	Source2	P	S1	S2	S1	S2
Central_Steppe_EMBA	*Western_Steppe_EMBA*	0.020	0.199	0.801	0.044	0.044

Table S 54 Proximal models for *Mereke_EBA_Yamnaya*

				Mixture Proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
WSHG	*Ganj_Dareh_N*	*ESHG*	0.08	0.52	0.15	0.33	0.01	0.01	0.01

Table S 55 Distal models for *Okunevo_BA.SG*

				Mixture Proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
WSHG	*Ganj_Dareh_N*	..	0.05	0.77	0.23	..	0.02	0.03	..

Table S 56 Distal models for *Dali_EBA*

			Mixture Proportions		SE	
Source1	**Source2**	**P**	**S1**	**S2**	**S1**	**S2**
Botai.SG	*Aigyrzhal_BA*	0.054	0.674	0.326	0.03	0.03
Botai.SG	*BMAC*	0.121	0.802	0.198	0.018	0.018
Botai.SG	*Geoksyur_EN*	0.142	0.778	0.222	0.02	0.02
Botai.SG	*Sarazm_EN*	0.019	0.767	0.233	0.021	0.021
Botai.SG	*Shahr_I_Sokhta_BA1*	0.051	0.798	0.202	0.018	0.018

Table S 57 Proximal models for *Dali_EBA*

S4.4.2.3 The Middle and Late Bronze Age of the Western and Central Steppe

In this next section, we analyze close to 1000 years of population transformations in the Steppe following the EMBA, corresponding roughly to 2000-1000 BCE. We coanalyze the individuals for whom we generated data together with individuals from the Corded Ware Complex of Eastern and Central Europe (**Fig S 26**). From previous work, we know that the movement of the *Yamnaya*-related ancestry from the Steppe into central and western Europe was involved in the spread of the Bronze Age Bell Beaker and Corded Ware Complexes (*7*). Additional work documented genetic similarity of people of the Corded Ware Complex to those of both the Sintashta and Srubnaya archaeological cultures of the western Steppe (*8*). While the impact of these spreads westward into Europe and their contribution to modern Europeans has been characterized, in this study we wished to examine the spread of the Sintashta and Andronovo cultural horizon eastward and southward into Central and South Asia.

We begin by studying the ADMIXTURE and PCA plots of all the individuals from this period (**Fig S 26**) and make a series of observations that we confirm later with *qpAdm*. Our first and most significant observation is that the Middle to Late Bronze Age cultures of the Steppe as well as from Central and Eastern Europe appear homogenous on the PCA (**Fig S 35**; top left panels) despite being separated by thousands of miles and in some cases close to a millennium. The second observation is that despite the homogeneity, these individuals appear highly admixed. On the ADMIXTURE plot (**Fig S 35**; right panel), they have up to four different components, reflecting deep ancestry similar to that in Neolithic Iran, Anatolia, Europe and the Steppe. The third observation is that despite the apparent homogeneity, there are subtle distinctions between the various populations suggesting that ADMIXTURE with local groups was happening albeit not at high levels. Finally, leveraging to our large sample sizes, we detected the presence of a significant number of outliers which showed additional admixture related to *WSHG* and people related to Iranian farmers, a process that became a ubiquitous by the LBA (**Fig S 35**). Of particular note here is the presence of an outlier from Poltavka during the EMBA, which appears to be genetically identical to the later populations from the Sintashta and Andronovo horizon. This individual has a mean radiocarbon date of 2658 BCE, and thus is the earliest individual for which we observe this characteristic mix of ancestries, thus extending the time period during which this ancestry was established in the region into the middle of the 3rd millennium BCE. We discuss the results of our outlier analysis after examining the ancestry observed in the two major groups of individuals from this period, *Western_Steppe_MLBA* and *Central_Steppe_MLBA*, which account for the large visually homogenous cluster on the PCA plots.

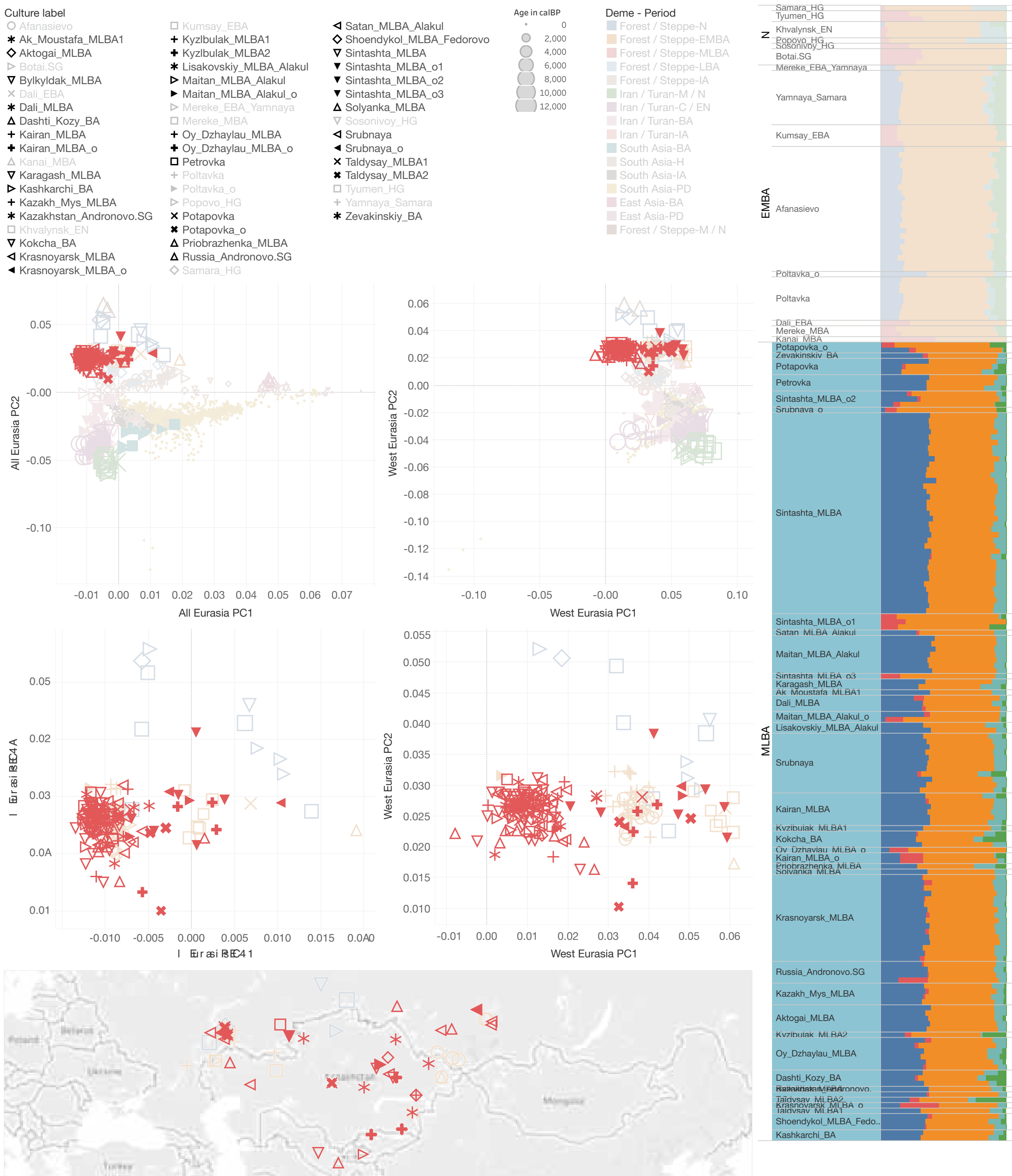

Fig S 35 Top panel: left, West Eurasian PCA; right, All Eurasian PCA. Middle panel: left, blow up of the West Eurasian PCA; right, blow up of the All Eurasian PCA. Bottom panel: geographic locations of individuals. Right panel: ADMIXTURE results, with the colors of green, teal, orange, red and blue representing ancestry maximized in early Iranian farmers from the Zagros, farmers from Anatolia, and HGs from West Siberia, East Siberia, and Western Europe respectively.

The *f*-statistic analysis confirms the pattern seen qualitatively in ADMIXTURE and PCA in which the proportions of ancestry related to *EEHG* are relatively homogenous across a wide geographical area ranging from central Europe to the central Steppe. However, moving from west-to-east geographically, we see that there are subtle differences in the proportion of Anatolian farmer ancestry and ancestry related to *WSHG* across the various MLBA individuals (**Fig S 35**). On the "Pre-Copper Age Affinity" *f*-statistic plot, (**Fig S 36**; Top right panel), we observe this ancestry cline with more clarity, with individuals of decreasing Anatolian farmer-related ancestry and increasing ancestry related to *WSHG* being drawn toward the outlier individuals as well as to *Central_Steppe_EMBA* individuals.

The fact that there are outlier individuals with considerably higher proportions of ancestry related to *WSHG* and that these outlier individuals lie close to the *Central_Steppe_EMBA* individuals adds further support to the theory that as these pastoralist populations moved into the central Steppe, they experienced admixture from local individuals inhabiting that area at that time. In addition we observe some outlier individuals that have considerably higher affinity to *ESHG*s, and some outliers that have considerably higher affinity to Iranian farmers, suggesting that these populations also admixed with populations from further east and south of the Steppe respectively.

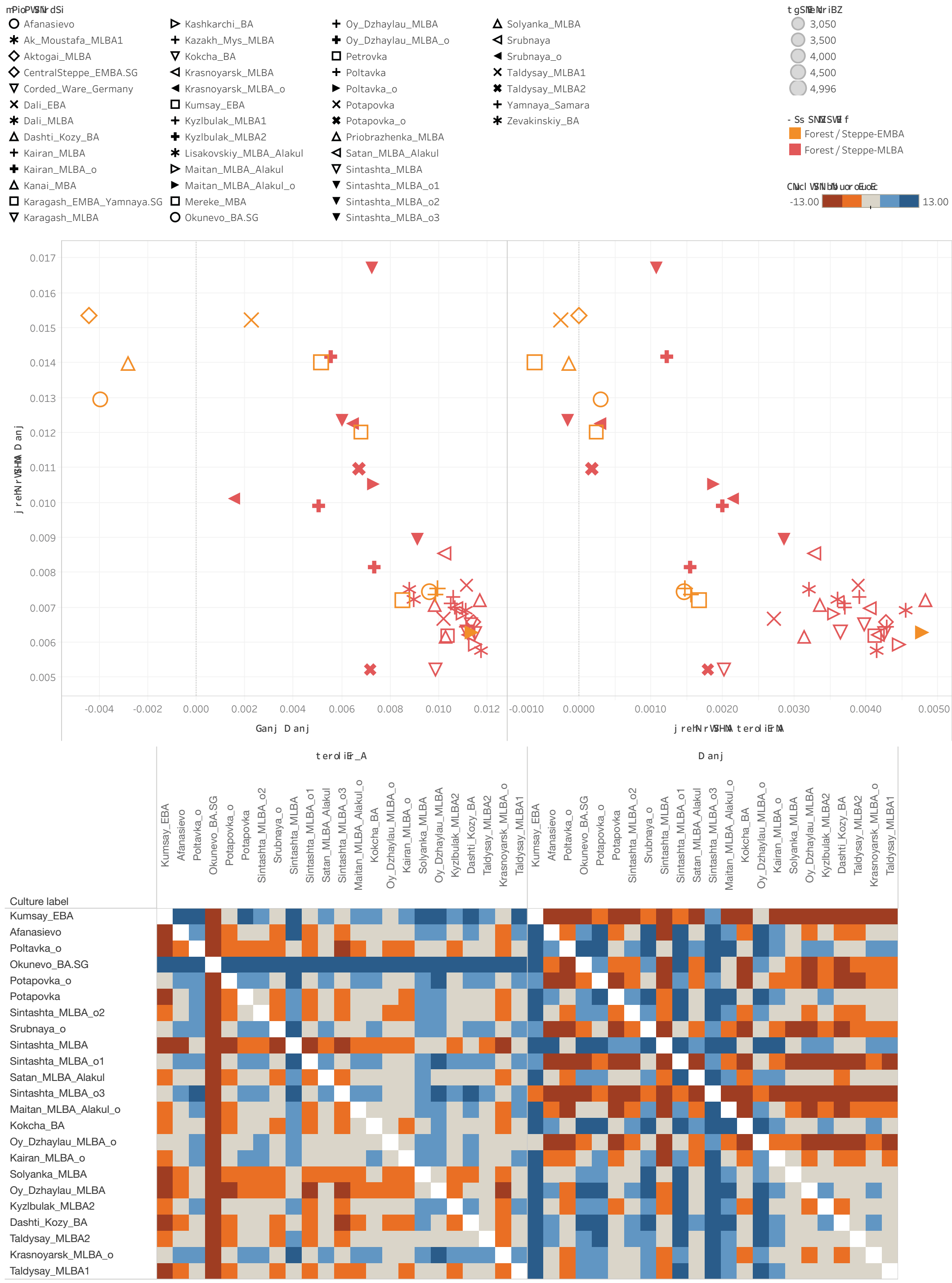

Fig S 36 Top panel: left; "Pre-Copper Age Affinity" f_4-statistics comparing f_4(Ethiopia_4500BP.SG, Test, ESHG, WEHG) vs. f_4(Ethiopia_4500BP.SG, Test, Ganj_Dareh_N, WSHG) and f_4(Ethiopia_4500BP.SG, Test, ESHG, WEHG) vs. Bottom panel: "Two Population Comparison" f_4-statistics showing differences in ancestry related to Anatolia_N and ancestry related to WSHG between populations. Orange squares indicate |Z-scores| > 3 for sharing more alleles with the test population compared to a population from the columns, with respect to an outgroup. Blue squares indicate significantly less sharing at the same threshold. Darker orange and blue squares show sharing at |Z-score| > 6.

Using admixture-f_3 statistics, we observe that MLBA Steppe populations appear to be admixed between one source related to the *ANE* (either *WSHG* or *EEHG*), and another source related to European farmers, as exemplified by *Globular_Amphora* or *Romania_C*. We illustrate this using *Sintashta_MLBA*, the *Western_Steppe_MLBA* population for which we had the most data (**Fig S 37**). We note that a single statistic involving *Shahr_I_Sokhta_BA3* and *Sidelkino_HG.SG* is much more negative than the other pairs, this is due to low overlap between the sets of SNPs from these two individuals as reflected in the large standard errors.

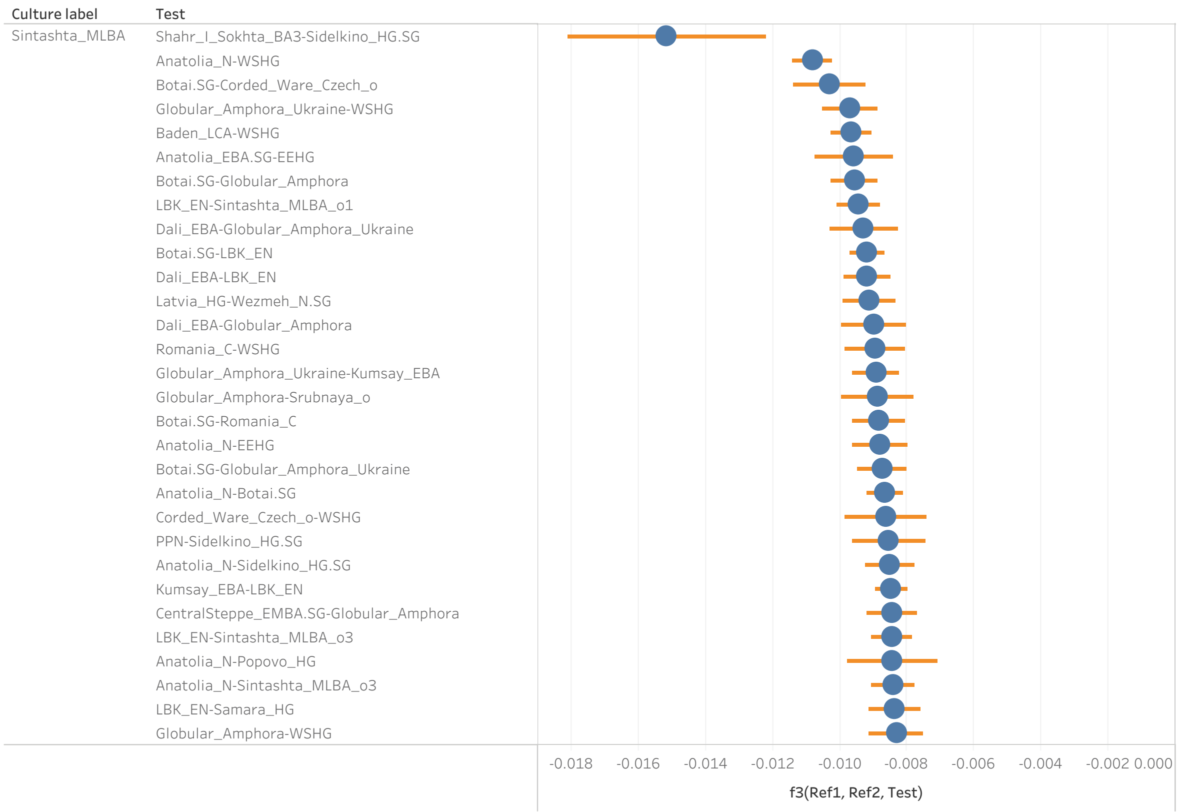

Fig S 37 Top 30 admixture-f_3 pairs for *Western_Steppe_MLBA*

We confirm these qualitative findings in *qpAdm* modeling. In our clustering analysis to create genetically homogenous analysis labels across sites, the individuals from the MLBA split into two genetically distinct clusters that correlate strongly to geography. The first group of individuals were genetically identical to Corded Ware Complex-associated individuals from Central and Europe as well as individuals associated with the Sintashta, Srubnaya and Petrovka cultures of the western Steppe. The second group comprises individuals from the Andronovo cultural horizon further east in the central Steppe and along the Inner Asian Mountain Corridor (*6*). We show this geographic separation in (**Fig S 38**).

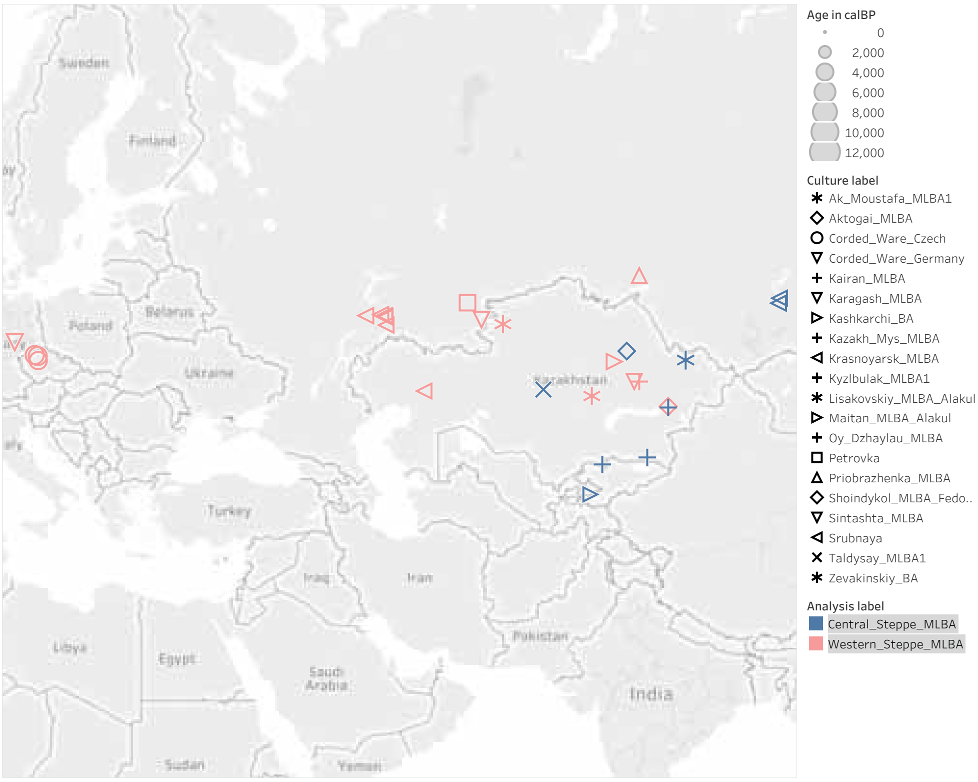

Fig S 38 Geographic distribution of individuals that fall in the *Western_Steppe_MLBA* and *Central_Steppe_MLBA* genetic clusters tracks genetic differences. The most plausible explanation for these patterns is a small amount of admixture of groups with *Western_Steppe_MLBA* ancestry with local groups they encountered east of the Urals.

In distal modeling, we observe that both *Western_Steppe_MLBA* and *Central_Steppe_MLBA* require four different West Eurasian-related source populations: three in common with *Western_Steppe_EMBA*, plus additional ancestry from *WEHG*s of Europe. However, *Central_Steppe_MLBA* and *Western_Steppe_MLBA* differ in their ancestry proportions from these sources. The *Central_Steppe_MLBA* individuals have significantly higher ancestry related to *WSHG* and significantly less ancestry related to Anatolian farmers and *WEHG* when compared to *Western_Steppe_MLBA*. We confirm through proximal modeling that this is plausibly due to admixture with local *Central_Steppe_EMBA* populations (**Table S 58** - **Table S 59**).

In proximal modeling, we observe only a few working models that are good fits for the ancestry in both these large groups of individuals. We find that a simple two way fit, involving *Western_Steppe_EMBA* and a European farmer-related population (*Globular_Amphora*) in the ratio of about 2/3 to 1/3 serves as a good fit to the data for

Western_Steppe_MLBA. However, *Central_Steppe_MLBA* groups require additional admixture of about 13-14% from a population related to *Central_Steppe_EMBA* along with the original two sources needed for *Western_Steppe_MLBA*, providing support for the theory that Steppe pastoralists moving through Kazakhstan experienced additional admixture from a local *WSHG*-derived population. We can formally model the *Central_Steppe_MLBA* populations as mixtures of the *Western_Steppe_MLBA* populations with about 9% additional ancestry from *Central_Steppe_EMBA* (**Table S 60 - Table S 61**). It is important to note that in the modeling of *Western_Steppe_MLBA* and *Central_Steppe_MLBA,* we treat samples across a large geographic and temporal range as a homogenous unit, whereas in fact of course they comprise individual populations that are separated over space and time. Thus, while our models use *Globular Amphora* and *Western_Steppe_EMBA* as proximal models for a genetic analysis label that encompasses all of these populations ranging from Corded Ware in the West to Andronovo in the East, these two populations we use for our proximal models are not actually reflecting the nature of the population transformation that occurred leading to the formation of these populations, and individual populations we analyzed were likely be sources for each other. We do not have information at present to decipher where exactly this population with this homogenous ancestry first formed, but do know that it occupied a large homogenous genetic group across this vast area by the turn of the 2nd millennium BCE.

					Mixture Proportions				SE			
Source1	**Source2**	**Source3**	**Source4**	**P**	**S1**	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**
EEHG	*WEHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.03	0.35	0.13	0.34	0.18	0.01	0.01	0.01	0.02

Table S 58 Distal models for *Western_Steppe_MLBA*

					Mixture Proportions				SE			
Source1	**Source2**	**Source3**	**Source4**	**P**	**S1**	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**
EEHG	*WEHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.1	0.42	0.10	0.32	0.17	0.01	0.01	0.02	0.02

Table S 59 Distal models for Central_Steppe_MLBA

			Mixture Proportions		SE	
Source1	**Source2**	**P**	**S1**	**S2**	**S1**	**S2**
Globular_Amphora	*Western_Steppe_EMBA*	*0.01*	0.325	0.675	0.007	0.007

Table S 60 Proximal models for *Western_Steppe_MLBA*

			Mixture Proportions		SE	
Source1	**Source2**	**P**	**S1**	**S2**	**S1**	**S2**
Central_Steppe_EMBA	*Western_Steppe_MLBA*	0.021	0.084	0.916	0.007	0.007

Table S 61 Proximal models for *Central_Steppe_MLBA*

We now move to outlier analysis where we observe that there are groups of individuals from different sites that are shifted away from the main cluster in three main directions: toward *Central_Steppe_EMBA*, toward *BMAC,* and toward *ESHG*s. These individuals serve as key additional lines of evidence for the inferred admixture events as determined by our proximal models. We use "Pre-Copper Age Affinity" and "Two Population Comparison" f_4-statistics to parse the ancestry of each of these outliers with respect to the main clusters of individuals in *Western_Steppe_MLBA* and *Central_Steppe_MLBA* (**Fig S 36**, top panel) and report the individuals in each outlying cluster below.

A complexity in these analyses is that the extent and source of the admixture in different outlier individuals is variable, as expected during a period of ongoing admixture. We therefore attempted estimates of average mixture proportions across all individuals. **Online Table 1** presents our assignment of analysis labels for these groups after our clustering procedure described in the beginning of this section. We observe that—perhaps unsurprisingly—the outlier individuals from each of the three groups derive from locations on the Steppe relatively close to other populations rich in the admixing ancestry type (**Fig S 39**). Groups with particularly large sample sizes, for example the dozens of individuals from the Kamennyi Ambar 5 Sintashta culture cemetery with high quality genetic data, are particularly valuable as like other large sample size sites in this study (e.g. Gonur from the BMAC) they include enough individuals to show outliers of multiple types that we have radiocarbon dated to overlapping periods.

Sintashta_MLBA_o3: Individual shifted toward *WSHG*, with lower fractions of Anatolian farmer as well as Iranian farmer-related ancestry.

Steppe_MLBA_oWSHG: Individuals with varying proportions of ancestry related to *Western_Steppe_MLBA* and *Central_Steppe_EMBA* (*Potapovka_o, Sintashta_MLBA_o1, Sintashta_MLBA_o2, Srubnaya_o, Maitan_BA_Alakul_o, Kairan_MLBA_o, Oy_Dzhaylau_MLBA_o, Dali_MLBA)*

Steppe_MLBA_oBMAC: Individuals shifted in the direction of farmers from Turan (*Dashti_Kozy_BA, Kokcha_BA, Kyzlbulak_MLBA2, Taldysay_MLBA2*)

Steppe_MLBA_oESHG: Individuals shifted in the direction of *ESHG*s (*Krasnoyarsk_MLBA_o)*

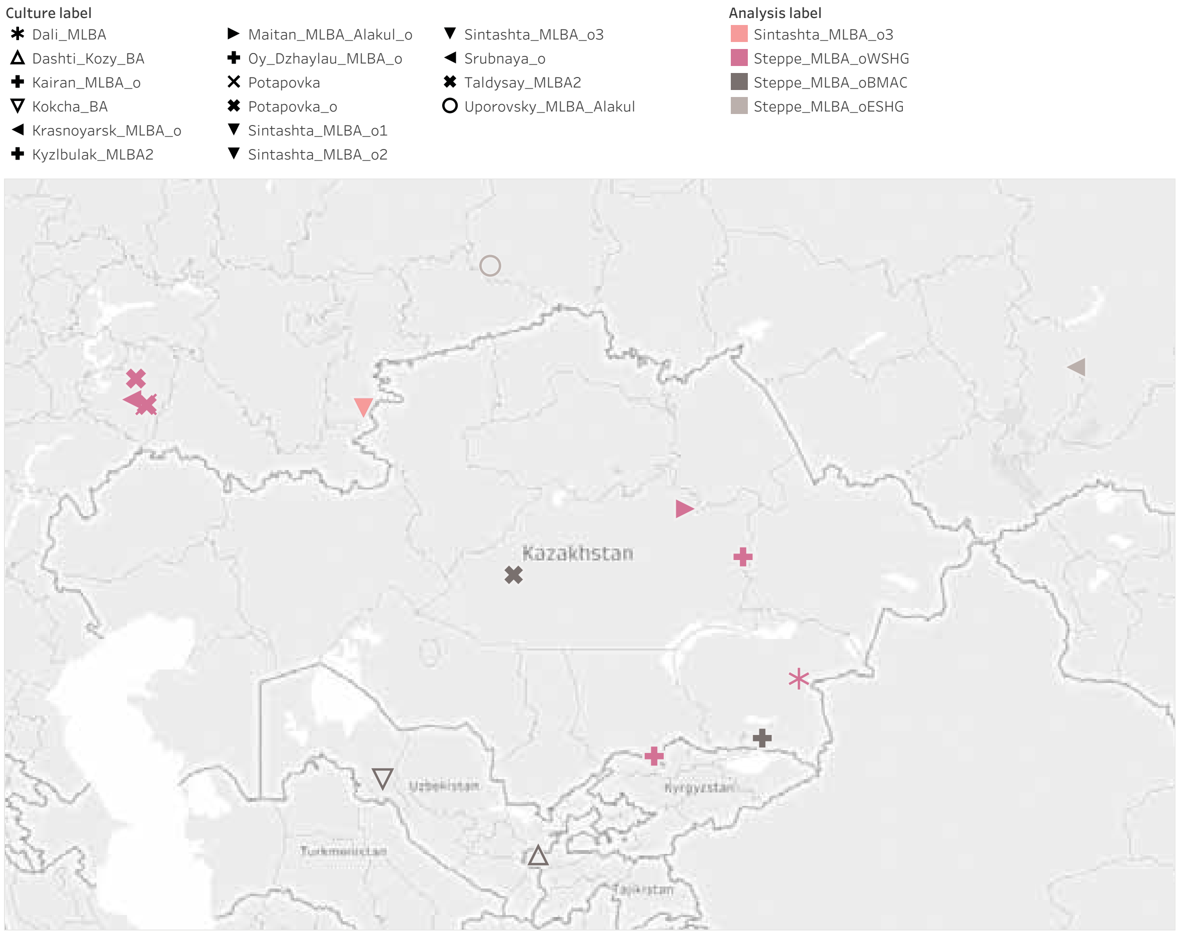

Fig S 39 Geographic locations of outlier individuals from the MLBA. We see that outliers with *WSHG* ancestry are distributed across the western and central Steppe, while outliers related to Iranian farmers and *ESHG*s lie on the southern and eastern edges of the Steppe respectively.

We start by examining the outliers with *WSHG* ancestry. These individuals are found in a variety of different sites and across a range of time periods. While they all share a general trend of having less European farmer-related and ancestry related to *WEHG*, they are heterogeneous in their proportions. On the PCA and ADMIXTURE plots, most of the outliers appear to have considerable admixture with populations related to *Western_Steppe_MLBA*, and a single set of individuals, *Sintashta_MLBA_o3*, appears to have little if any ancestry related to Anatolian or Iranian farmers (green or teal on the ADMIXTURE plots), and appears on the PCA close to populations on the Steppe prior to the EMBA (**Fig S 35**). Aside from this set of individuals, the other outliers related to *ANE* appear to be mixtures in different proportions between *Western_Steppe_MLBA* and *Central_Steppe_EMBA*, whose admixture we have seen also gave rise to the *Central_Steppe_MLBA* grouping.

One possible explanation for these observations could be that we are obtaining individuals from different time periods buried in the same location, but we directly radiocarbon dated at least one individual from each of these outlying groups and showed them to be contemporaneous with the main clusters of individuals. A second argument against temporal differences as an explanation comes from the observation of genetically similar outlier individuals spread across multiple sites. This suggests that we are observing at multiple sites instances of a general interaction and admixture process.

Despite their heterogeneity, we studied the average ancestry proportion for all individuals in each outlier cluster by grouping them under a larger analysis label. For *Steppe_MLBA_oWSHG*, we observe in distal modeling a requirement for 4 to 5 sources with large proportions of ancestry related to *WSHG* (**Table S 64**). Proximally, they can be modeled parsimoniously using simple 2-way models of *Western_Steppe_MLBA* or *Central_Steppe_MLBA* in about equal ratios (**Table S 65**).

The second set of outliers from this time period related to Turan (*Steppe_MLBA_oBMAC*) are drawn toward the BMAC on the PCA (**Fig S 35**) and have substantially higher proportions of Iranian farmer-related affinity as reflected both in the ADMIXTURE plot and in f_4-statistics (**Fig S 36**, top panel). These individuals are at the southern end of our geographic sampling, suggesting that the individuals of the Andronovo material culture horizon came into contact with the people of the BMAC by at least 2000 BCE. As above, we examine whether these individuals form a clade with each other, and we find that while they are not identical, they are related enough to group them and we provide distal fits that have the same components as *Central_Steppe_MLBA* but with significantly larger proportions of Iranian farmer-related ancestry (**Table S 66**). In proximal modeling, these individuals can be fit as having 70% of their ancestry derived from a *Central_Steppe_MLBA* source and 30% of their ancestry from a source related to the main *BMAC* genetic cluster (**Table S 67**).

The presence of these outliers is particularly striking as similar outliers with ancestry admixed between the main *BMAC* and the main *Central_Steppe_MLBA* cluster are also present in the later BMAC sites, suggesting that contact between the two was common and involved bidirectional exchange.

We finally examine the outliers related to *ESHG* (*Steppe_MLBA_oESHG*). We observed both on PCA (**Fig S 35**) and *f*-statistics (**Fig S 36**) that these individuals are shifted significantly away from the other individuals, including ones with ancestry related to *WSHG*. Geographically, these individuals also tend to be distributed in the northeast of the Steppe regions analyzed in this study. Distal models show that these individuals harbor almost 40% ancestry related to *ESHG* (**Table S 68**) Proximal models suggest that these individuals require additional ancestry related to *Shamanka_EBA.SG*, who are individuals from the Bronze Age in Lake Baikal with high *ESHG* proportions (**Table S 69**).

				Mixture proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
WSHG	*WEHG*	..	0.02	0.84	0.16	..	0.02	0.02	..
EEHG	*WSHG*	*PPN*	0.39	0.30	0.61	0.07	0.04	0.04	0.01
EEHG	*WSHG*	*Anatolia_N*	0.45	0.35	0.59	0.07	0.04	0.04	0.02
EEHG	*WSHG*	*Ganj_Dareh_N*	0.67	0.41	0.50	0.09	0.04	0.04	0.02
WSHG	*PPN*	*WEHG*	0.45	0.82	0.07	0.11	0.02	0.02	0.02
WSHG	*WEHG*	*Anatolia_N*	0.18	0.84	0.12	0.04	0.02	0.02	0.02

Table S 62 Distal models for *Sintashta_MLBA_o3*

			Mixture proportions		SE	
Source1	**Source2**	**P**	**S1**	**S2**	**S1**	**S2**
Botai.SG	*Khvalynsk_EN*	0.86	0.401	0.599	0.034	0.034

Table S 63 Proximal models for *Sintashta_MLBA_o3*

					Mixture proportions				SE			
Source1	**Source2**	**Source3**	**Source4**	**P**	**S1**	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**
WSHG	*WEHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.07	0.53	0.12	0.20	0.16	0.01	0.01	0.02	0.02

Table S 64 Distal models for *Steppe_MLBA_oWSHG*

				Mixture proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Okunevo_BA.SG	*Central_Steppe_EMBA*	*Western_Steppe_MLBA*	0.44	0.049	0.45	0.501	0.011	0.017	0.011
Okunevo_BA.SG	*Central_Steppe_EMBA*	*Central_Steppe_MLBA*	0.054	0.05	0.45	0.5	0.01	0.017	0.012

Table S 65 Proximal models for Steppe_MLBA_oWSHG

					Mixture proportions				SE			
Source1	**Source2**	**Source3**	**Source4**	**P**	**S1**	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**

EEHG	*WEHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.94	0.43	0.07	0.27	0.23	0.02	0.01	0.02	0.02

Table S 66 Distal models for *Steppe_MLBA_oBMAC*

				Mixture proportions			**SE**		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
BMAC	*Central_Steppe_EMBA*	*Western_Steppe_MLBA*	0.240986738	0.125	0.142	0.733	0.012	0.013	0.017
BMAC	*Central_Steppe_EMBA*	*Central_Steppe_MLBA*	0.052567681	0.122	0.077	0.801	0.013	0.014	0.019
BMAC	*Western_Steppe_EMBA*	*Central_Steppe_MLBA*	0.305603322	0.112	0.257	0.631	0.013	0.043	0.043
Sarazm_EN	*Central_Steppe_EMBA*	*Western_Steppe_MLBA*	0.536143277	0.15	0.099	0.751	0.015	0.014	0.015
Sarazm_EN	*Central_Steppe_EMBA*	*Central_Steppe_MLBA*	0.190976203	0.148	0.033	0.819	0.015	0.015	0.018

Table S 67 Proximal models for *Steppe_MLBA_oBMAC*

					Mixture proportions				**SE**			
Source1	**Source2**	**Source3**	**Source4**	**P**	**S1**	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**
EEHG	*WSHG*	*Anatolia_N*	*ESHG*	0.03	0.24	0.24	0.20	0.27	0.05	0.05	0.02	0.02
WSHG	*WEHG*	*Anatolia_N*	*ESHG*	0.05	0.44	0.10	0.21	0.26	0.02	0.02	0.02	0.02

Table S 68 Distal models for *Steppe_MLBA_oESHG*

				Mixture proportions			**SE**		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Okunevo_BA.SG	*Shamanka_EBA.SG*	*Western_Steppe_MLBA*	0.031539852	0.453	0.118	0.43	0.064	0.04	0.028
Okunevo_BA.SG	*Shamanka_EBA.SG*	*Central_Steppe_MLBA*	0.022864739	0.391	0.142	0.467	0.068	0.042	0
Shamanka_EBA.SG	*Central_Steppe_EMBA*	*Western_Steppe_MLBA*	0.021833255	0.318	0.25	0.432	0.016	0.034	0.027
Shamanka_EBA.SG	*Central_Steppe_EMBA*	*Central_Steppe_MLBA*	0.013252558	0.315	0.214	0.471	0.016	0.037	0

Table S 69 Proximal models for *Steppe_MLBA_oESHG*

In addition to these autosomal ancestry estimates, additional information about admixture comes from the Y chromosome data. With the exception of a few individuals, all of the males in the *Western_Steppe_MLBA* and *Central_Steppe_MLBA* groups carry a Y chromosome from the R1a haplogroup, whose earliest observations in our dataset are in individuals from the western Steppe. However, most of the outlier individuals with ancestry related to *WSHG* do not carry this haplogroup, but instead carry haplogroups seen in *Botai.SG, WSHG, Central_Steppe_EMBA* and *Okunevo_BA.SG*. Individuals with Iranian farmer-related ancestry tended to carry the J Y haplogroup that appears to be nearly absent in Steppe pastoralists but predominant in BA Turan, providing further evidence of admixture.

S4.4.2.4 The Late Bronze Age and Early Iron Age: Large scale admixture of East and West

We complete our genetic characterization of the Steppe individuals by examining population transformation at the end of the 2nd millennium BCE and into the early Iron Age. In the previous subsection, we showed that outlier individuals from the Bronze Age in the early part of the 2nd millennium harbored additional ancestry from *Central_Steppe_EMBA* and Iranian farmer-related populations. Moving into later times, this admixture of Steppe pastoralists arriving from the western Steppe with local groups continued with additional Iranian farmer-related ancestry and ancestry related to *WSHG* in almost all the individuals we analyze. On the PCA and ADMIXTURE plots, these individuals are shifted both in the direction of *ESHG* and in the direction of ancient farmers from Iran / Turan. In the "Pre-Copper Age Affinity" f_4-statistics, we see that these individuals shift toward *ESHG* and away from the *Western_Steppe_MLBA* populations, but in comparison to outlier individuals that are shifted toward *WSHGs* from the MLBA, these LBA individuals appear to have additional Iranian farmer-related ancestry (**Fig S 40 - Fig S 41**). To support these observations, we show admixture-f_3 statistics yielding negative values with both admixing sources (**Fig S 42**).

Fig S 40 Top panel: left, West Eurasian PCA; right, All Eurasian PCA. Middle panel: left, blow up of the West Eurasian PCA; right, blow up of the All Eurasian PCA. Bottom panel: geographic locations of individuals. Right panel: ADMIXTURE results, with the colors of green, teal, orange, red and blue representing ancestry maximized in Iranian farmers from the Zagros, farmers from Anatolia, and HGs from West Siberia, East Siberia, and Western Europe respectively.

Fig S 41 Top panel: "Pre-Copper Age Affinity" f_4-statistics comparing the statistics *f_4(Ethiopia_4500BP.SG, Test, Ganj_Dareh_N, WSHG)* vs. *f_4(Ethiopia_4500BP.SG, Test, ESHG, WEHG)*. Bottom panel: "Two Population Comparison" f_4-statistics showing differences in ancestry related to *ESHG* and *WSHG* between populations. Orange squares indicate |Z-scores| > 3 for sharing more alleles with the test population compared to a population from the columns, with respect to an outgroup. Blue squares indicate significantly less sharing at the same threshold. Darker orange and blue squares show sharing at |Z-score| > 6.

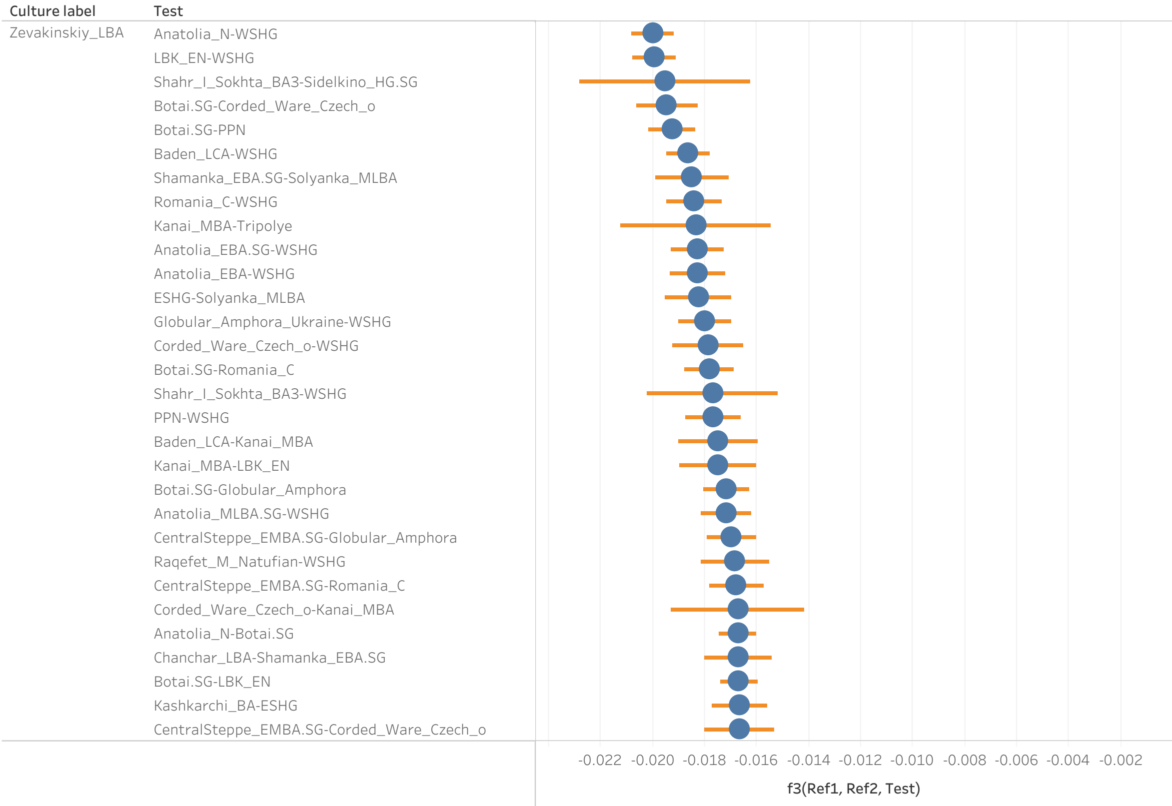

Fig S 42 Most negative admixture-f_3 statistics for East Asians for *Zevakinskiy_LBA*

We then show distal and proximal models for these populations and observe that the proportion of ancestry related to *ESHG* and *Ganj_Dareh_N* both increase significantly (**Table S 70**). Proximal models suggest that these LBA individuals can be modeled with *Central_Steppe_MLBA* populations along with additional ancestry from Turan as well as from the eastern Steppe (**Table S 71**)

						Mixture Proportions					SE				
Source1	**Source2**	**Source3**	**Source4**	**Source5**	**P**	**S1**	**S2**	**S3**	**S4**	**S5**	**S1**	**S2**	**S3**	**S4**	**S5**
EEHG	*WSHG*	*Anatolia_N*	*Ganj_Dareh_N*	*ESHG*	0.02	0.33	0.16	0.28	0.06	0.17	0.03	0.04	0.03	0.04	0.02

Table S 70 Distal models for *Steppe_LBA*

				Mixture proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Aigyrzhal_BA	*Okunevo_BA.SG*	*Central_Steppe_MLBA*	0.000986	0.126	0.311	0.564	0.023	0.012	0.018

Table S 71 Proximal models for *Steppe_LBA*

S4.4.3 The historical period in Turan and the Steppe

Prior to modeling the ancestry of South Asians, we examined 3 individuals from the historical period (30-380 CE) from Turan and the Steppe that were archaeologically associated with the Kushan empire, known from historical records to have had a political impact on South Asia. These individuals were not included in the originally submitted version of this manuscript (*215*) and we therefore do not fully co-analyze these individuals in our admixture modeling framework. However, here we report some key analyses on these individual to shed light on the question of whether the Kushans could have been the source of a substantial proportion of the Steppe pastoralist-related ancestry in the subcontinent.

Our PCA results (left middle panel of **Fig S 43**) show that the Kushan individuals while genetically heterogenous are all intermediate in ancestry between individuals from East Asia, the BA Steppe, and BA Turan. They lie on a cline of ancestry that was established at the start of the LBA between people related to *ESHG* and people related to MLBA Steppe pastoralists and Turan farmers. In *qpAdm*, our distal modeling shows that the Kushan individuals like many LBA Steppe groups fit as a mixture of 5 different ancestry types, albeit with slightly higher proportions of Iranian farmer-related ancestry than LBA Steppe groups. This is in line with the theory that the Kushans were derived from an LBA Steppe pastoralist group that admixed with later IA populations in eastern Asia and Turan. Proximal modeling further shows that the 3 Kushan individuals can be modeled as mixtures of *Steppe_LBA* and the *BMAC*, with additional ancestry from a source rich in Anatolian farmer-related ancestry that plausibly increased in proportion in Turan in the two millennia after the *BMAC*.

With respect to South Asia, our key finding is that people with ancestry like the Kushan individuals can be excluded as important sources of the Steppe pastoralist-related ancestry that is widespread in South Asia today. In particular, the East Asian-related admixture (via *Steppe_LBA* ancestors) that characterized the Kushan individuals is nearly absent in South Asia. We formally confirmed this inference through *qpAdm* modeling that excludes the Kushan individuals, as well as nearly all the other Iron Age and historical period individuals from other cultural contexts that were published in two recent studies (*29*, *30*) as plausible sources for the Steppe pastoralist-related ancestry in South Asia (**Fig S 50**).

						Mixture Proportions					SE				
Source1	**Source2**	**Source3**	**Source4**	**Source5**	**P**	**S1**	**S2**	**S3**	**S4**	**S5**	**S1**	**S2**	**S3**	**S4**	**S5**
EEHG	*WSHG*	*Anatolia_N*	*Ganj_Dareh_N*	*ESHG*	0.042	0.13	0.12	0.37	0.27	0.12	0.05	0.06	0.04	0.06	0.03
AHG	*EEHG*	*WSHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.29	0.14	0.13	0.15	0.39	0.20	0.03	0.04	0.05	0.04	0.06
AHG	*WSHG*	*WEHG*	*Anatolia_N*	*Ganj_Dareh*	0.20	0.14	0.21	0.05	0.39	0.21	0.02	0.02	0.02	0.04	0.07

Table S 72 Distal models for *Ksirov_H_Kushan*

				Mixture proportions			SE		
Source1	**Source2**	**Source3**	**P**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
BMAC	*Hajji_Firuz_C*	*Steppe_LBA*	0.09	0.168	0.397	0.435	0.05	0.041	0.026

Table S 73 Proximal models for *Ksirov_H_Kushan*

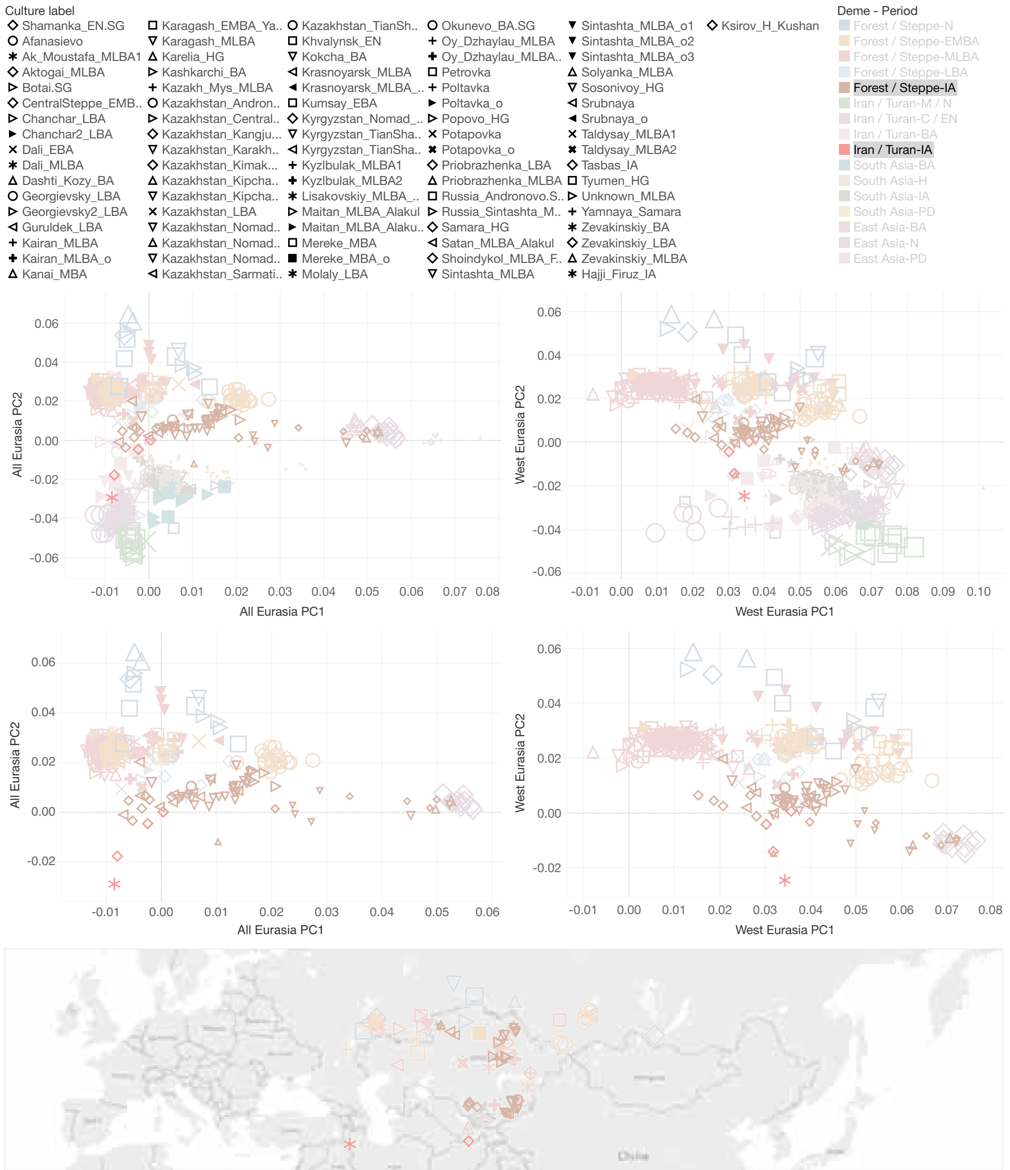

Fig S 43 PCA and geographical location of Kushan individuals with respect to other published individuals from Historical and Iron Age Iran/Turan and the Steppe.

S4.4.4 South Asia

In **Supplementary Materials S4**, we provide a detailed analysis of all individuals from South Asia, including the relationship of ancient individuals that were outliers from Turan, and 140 modern populations from South Asia. Here, we describe qualitative statistics of the ancient individuals mainly from the Swat Valley in the Late Bronze-Iron Age and Historical periods, which will be further analyzed in S4.

S4.4.4.1 <u>The Late Bronze and Iron Age and Historical Period in South Asia</u>

We describe the ancestry of individuals from 8 Late Bronze-Iron Age and 7 historical period sites from the Swat Valley and Chitral in Pakistan, which to our knowledge represents the first whole genome ancient DNA data from South Asia. The 8 Late Bronze-Iron Age sites are dated to ~1000 BCE (Katelai, Loebanr, Barikot, Butkara, Arkotkila, Aligrama, Udegram and Godgara), and the historical era sites date from around 350 BCE (Saidu Sharif), to approximately 500 CE (Singoor) (
Fig S 44). Given their geographic position within northern South Asia, these individuals allow us to examine directly the relationships of (a limited part of) South Asia to Iran, Turan and the Steppe. We were particularly interested in understanding the relationship of these Late Bronze-Iron Age individuals from South Asia to the *Indus Periphery Cline* individuals from BA Turan from the sites of Shahr-i-Sokhta and Gonur. For qualitative analysis we co-analyze these individuals together with whole genome sequencing data from additional present-day South Asian individuals from the Simons Genome Diversity Project (*55*).

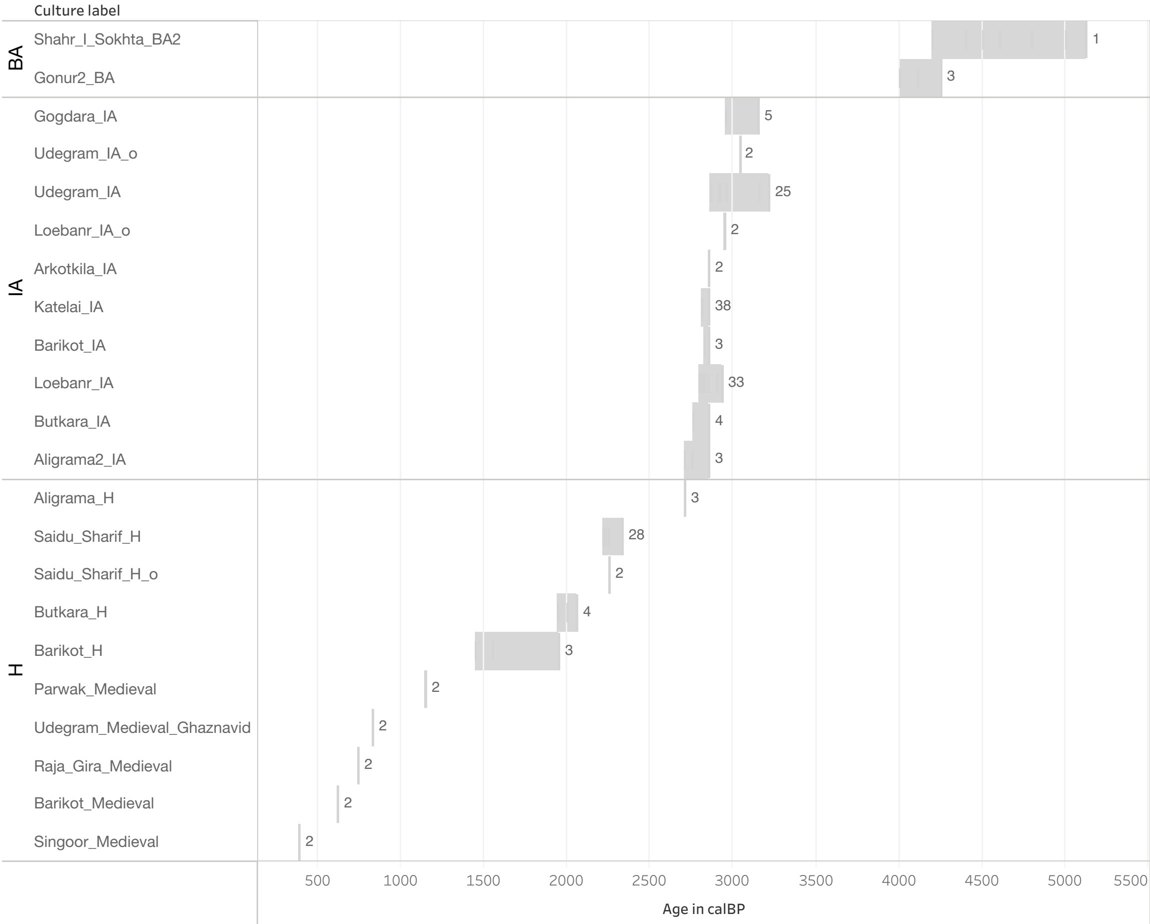

Fig S 44 The number of individuals and distribution of individual dates of individuals analyzed in this section

In the All Eurasian PCA, we observe that these individuals lie between the Pre-Copper Age ancestry type maximized in Iranian farmers, Steppe pastoralists, and present-day South Asians (**Fig S 45;** top and middle panel). This suggests that a significant portion of their ancestry is derived from the pre-agricultural populations of South Asia (at present unsampled with ancient DNA). However, when compared to present-day individuals from India, these individuals appear to be shifted toward early Iranian farmers, suggesting that they have higher proportions of Iranian farmer-related ancestry, and that their geographic position in the northwestern part of South Asia was associated with genetic connections to groups to the west and north. Another striking feature of the genetic ancestry of these individuals is that their ancestry is roughly similar (but with an important exception described below) to that of *Indus Periphery Cline* individuals from Gonur and Shahr-i-Sokhta: large settlements to the northwest of South Asia (described above). The individuals from these sites are dated to almost a millennium prior to the Swat Protohistoric Grave Type (*SPGT*) individuals, a term we use to refer to the main cluster of ancestry from the Swat Late Bronze-Iron Age, but they have substantial proportions of South Asian Hunter-Gatherer (*AASI*)-related admixture including some individuals that have much higher proportions of ancestry related to *AHG*. Examining this more carefully, we observe on the West Eurasian PCA which is sensitive to changes in ancestry from West Eurasia but not East Eurasia (**Fig S 45**; top and middle right panel) that the Late Bronze-Iron Age and historical era individuals from South Asia lie on a cline, with the *Indus Periphery Cline* individuals on one end, and *Western_Steppe_MLBA* and *Central_Steppe_MLBA* populations on the other (however, *Western_Steppe_EMBA* and *Central_Steppe_EMBA* do not align). The cline tends points toward Steppe MLBA populations rather than toward Steppe EMBA populations, providing is another qualitative line of evidence about the nature of the Steppe ancestry in South Asia as well as its time of entry into the region. On the ADMIXTURE plots, we observe that the Late Bronze-Iron Age individuals from South Asia are assigned five ancestry components, with the major elements maximized in South Asia and Iran, plus additional ancestry maximized in *EEHG*, *WEHG*, and early Anatolian farmers (**Fig S 45**; bottom panel). These patterns contrast with the *Indus Periphery Cline* individuals which only have two components, maximized in Iranian farmers and *WSHG*s, again suggesting admixture in South Asia in the Late Bronze Age and early Iron Age including an arrival of Steppe pastoralist-related ancestry.

The individuals from Aligrama, which are contemporaneous to the Late Bronze-Iron Age populations, appear to have less Steppe pastoralist-related admixture than the individuals from other sites, based on the extent of the shift from the *Indus Periphery Cline* individuals on the PCA plot.

From two sites, Loebanr and Udegram, we find two outliers with more Steppe pastoralist-related admixture. Both observations on PCA are confirmed by ADMIXTURE (**Fig S 45**).

Finally, individuals from the historical periods are shifted toward populations from southern India on the All Eurasian PCA plot, and shifted toward Steppe pastoralists on the West Eurasian PCA plot, suggesting that both of these ancestry types increased in their contribution to peoples of the Swat Valley through the Late Bronze-Iron Age and into historical time.

Finally, an outlier individual from Bustan in Turkmenistan that we described in the section on BA Turan appears qualitatively to be indistinguishable genetically from other *SPGT* individuals and directly dated to 1511 BCE (the midpoint of its 95% confidence interval).

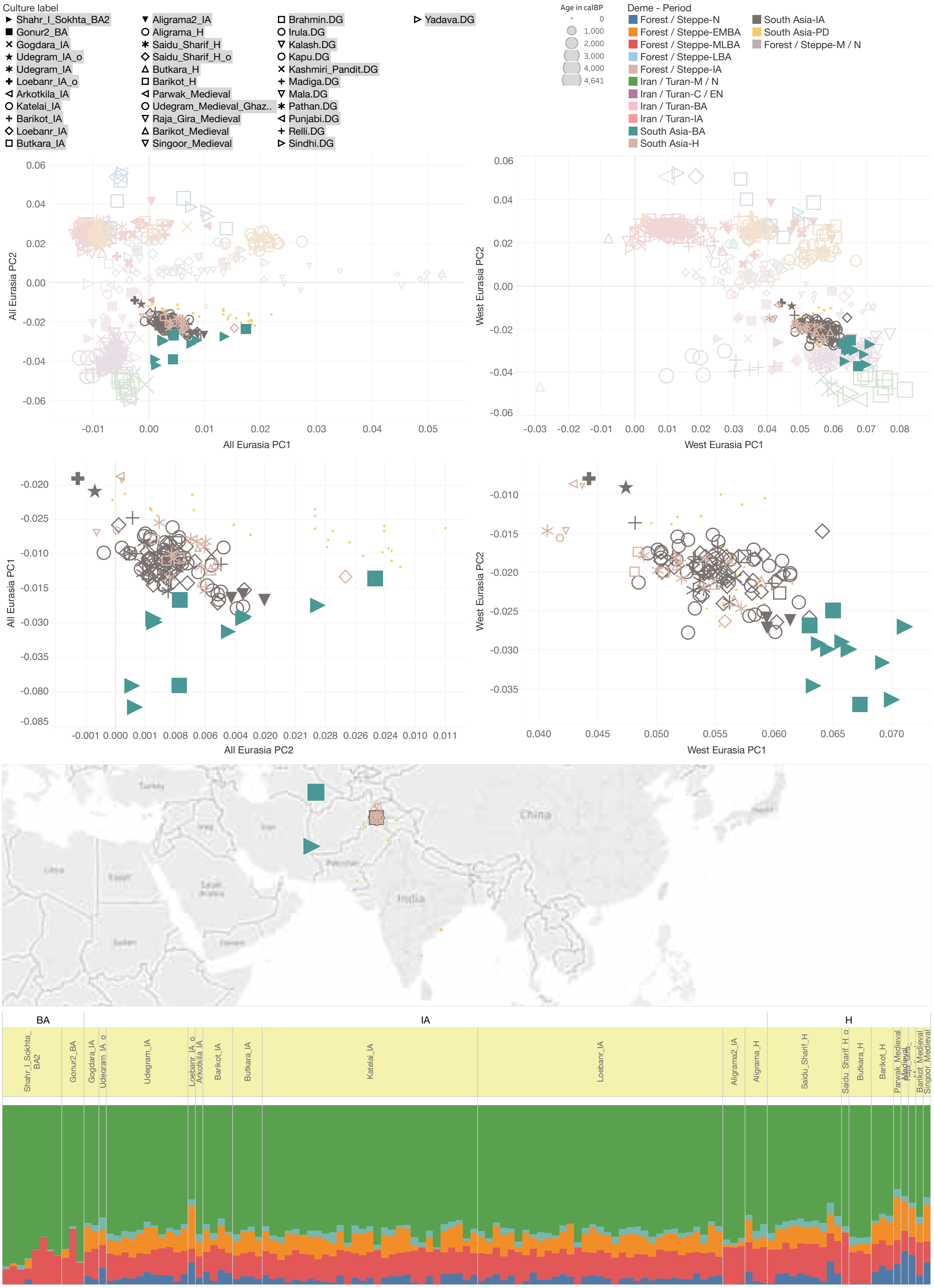

Fig S 45 Top panel: left, All Eurasian PCA; right, West Eurasian PCA. Second panel: left, blow-up of West Eurasian PCA; right, blow-up of All Eurasian PCA. Third panel: geographic locations of individuals. Bottom panel: ADMIXTURE results, with the colors of green, teal, orange, red and blue representing ancestry maximized in Iranian farmers from the Zagros, farmers from Anatolia, and HGs from West Siberia, South Asia, and Western Europe respectively.

We next examined "Pre-Copper Age Affinity" and "Two Population Comparison" f_4-statistics (**Fig S 46**). The main cluster of Late Bronze-Iron Age individuals appear relatively homogenous and similar in ancestry. As observed from the PCA and ADMIXTURE plots, we see that compared to the *Indus Periphery Cline* individuals, the *SPGT* individuals share more alleles with *WEHG* and fewer with *AHG*. The individuals from Aligrama and the other historical era individuals from 350-50 BCE appear to have additional *AHG*-related ancestry. Finally, the two *SPGT* outliers, in accordance with their position on the PCA and the ADMIXTURE plots, also share more alleles with *WEHG*. We confirm all of these findings and attempt to model the ancestry seen in all of these individuals using *qpAdm*.

Based on the PCA and ADMIXTURE plots as well as f_4-statistics, the individuals from the Late Bronze-Iron Age Swat Valley other than those from Aligrama, appear homogenous to within our resolution and we grouped them for the modeling framework under the label *SPGT* (Swat Protohistoric Grave Type). Similarly, individuals from the early historical, and medieval periods appeared relatively similar on *f*-statistics and on the PCA, ADMIXTURE plots and we grouped these into various subgroups (see **Table S 1**) for further analysis.

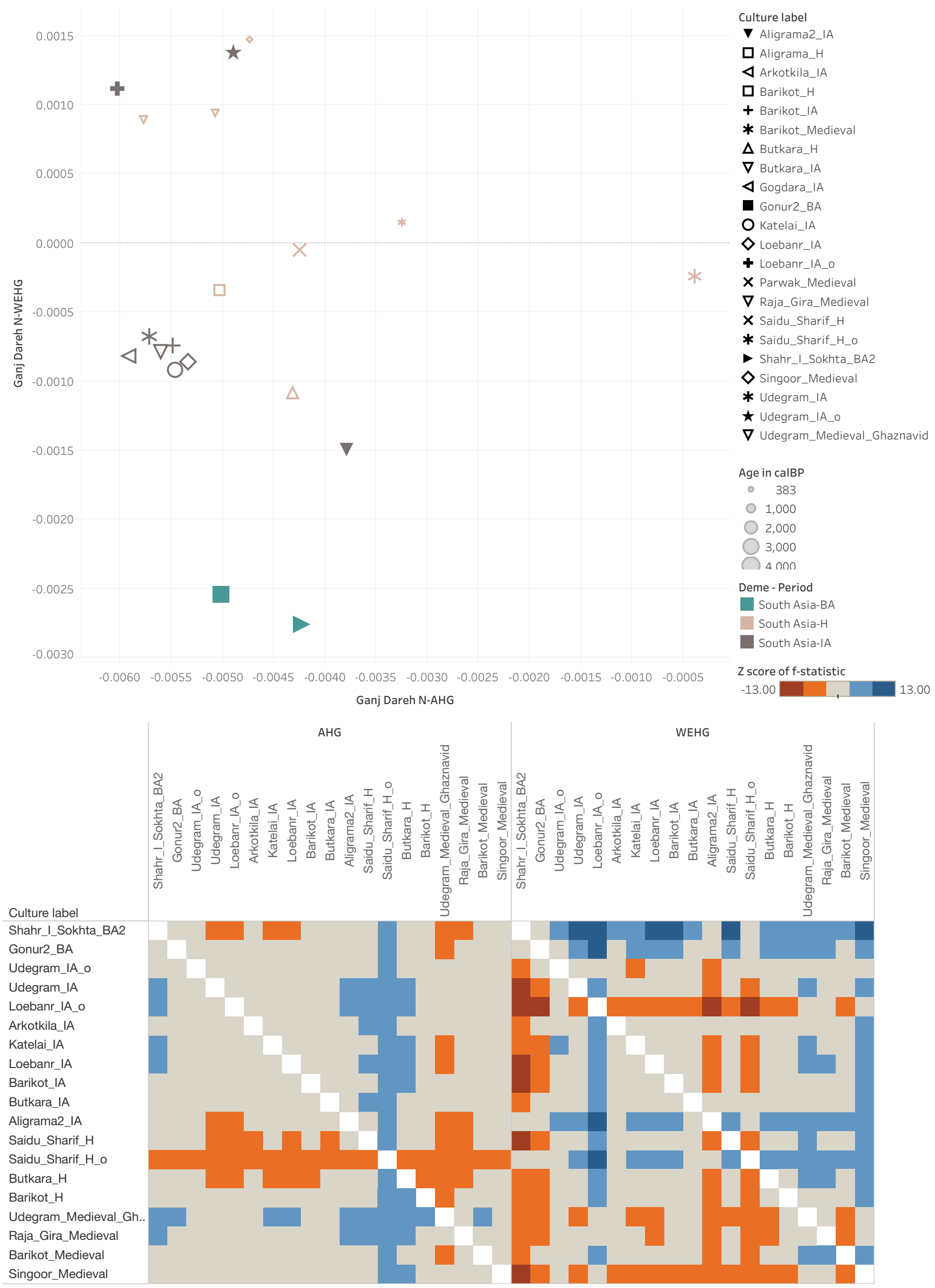

Fig S 46 Top panel: "Pre-Copper Age Affinity" f_4-statistics comparing the statistics f_4(*Ethiopia_4500BP.SG, Test, Ganj_Dareh_N, WSHG*) vs. f_4(*Ethiopia_4500BP.SG, Test, Ganj_Dareh_N, AHG*). Bottom panel: "Two Population Comparison" f_4-statistics showing differences in *AHG*-related and *WEHG*-related ancestry between populations. Orange squares indicate |Z-scores| > 3 for sharing more alleles with the test population compared to a population from the columns, with respect to an outgroup. Blue squares indicate significantly less sharing at the same threshold. Darker orange and blue squares show sharing at |Z-score| > 6.

In **Fig S 47** and **Fig S 48** we show admixture f_3-statistics for *SPGT*, which offer evidence of admixture between groups related to South Asian Hunter-Gatherers, ancient Iranian farmers, and Ancestral North Eurasians (including Steppe pastoralist) groups, suggesting that admixture with all of these regions had occurred in South Asia by this time.

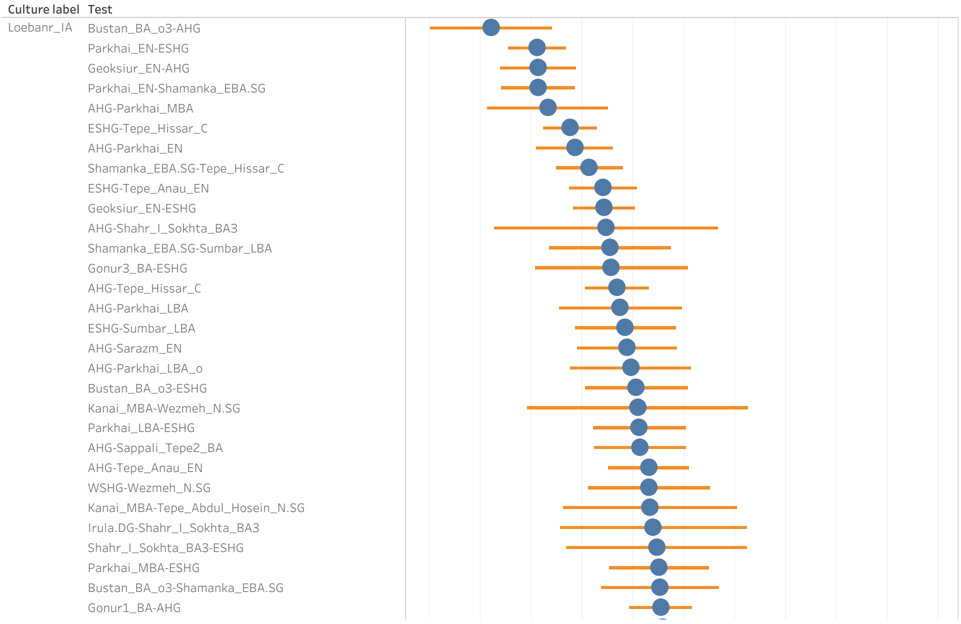

Fig S 47 Top admixture f_3 statistics for *Loebanr_IA*

Fig S 48 Additional negative f_3 statistics for *Loebanr_IA* with *Indus Periphery Cline* and Steppe Middle to Late Bronze Age individuals as sources.

We now begin to examine distal models for the South Asian individuals. They require 5 ancestral components to obtain a good fit to our data for the Late Bronze-Iron Age and historical era sites (**Table S 74 - Table S 81**) consistent with the PCA and ADMIXTURE analyses, the later sites are estimated to have about 10% more *AHG*-related ancestry compared with earlier Late Bronze-Iron Age sites, suggesting that there was additional admixture from the southeast (where we expect to find populations with higher proportions of such ancestry).

						Mixture Proportions					SE				
Source1	**Source2**	**Source3**	**Source4**	**Source5**	**P**	**S1**	**S2**	**S3**	**S4**	**S5**	**S1**	**S2**	**S3**	**S4**	**S5**
AHG	*WSHG*	*Anatolia_N*	*Ganj_Dareh_N*	..	0.03	0.21	0.21	0.23	0.36	..	0.01	9.00	0.02	0.03	..
AHG	*WSHG*	*Anatolia_N*	*WEHG*	*Ganj_Dareh_N*	0.05	0.21	0.20	0.21	0.02	0.36	0.01	0.01	0.02	0.01	0.03

Table S 74 Distal models for *SPGT*

						Mixture Proportions					SE				
Source1	**Source2**	**Source3**	**Source4**	**Source5**	**P**	**S1**	**S2**	**S3**	**S4**	**S5**	**S1**	**S2**	**S3**	**S4**	**S5**
AHG	*EEHG*	*WSHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.07	0.13	0.13	0.17	0.29	0.29	0.04	0.06	0.06	0.05	0.08
AHG	*WSHG*	*WEHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.32	0.13	0.23	0.08	0.28	0.29	0.03	0.06	0.02	0.05	0.08

Table S 75 Distal models for *SPGT_o*

					Mixture Proportions				SE			
Source1	**Source2**	**Source3**	**Source4**	**P**	**S1**	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**
AHG	*WSHG*	*PPN*	*Ganj_Dareh_N*	0.06	0.28	0.15	0.13	0.45	0.02	0.01	0.03	0.05
AHG	*WSHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.17	0.32	0.16	0.13	0.40	0.02	0.02	0.03	0.05
WSHG	*PPN*	*Ganj_Dareh_N*	*ESHG*	0.02	0.09	0.11	0.49	0.32	0.01	0.03	0.05	0.02
WSHG	*Anatolia_N*	*Ganj_Dareh_N*	*ESHG*	0.03	0.09	0.11	0.46	0.34	0.02	0.03	0.06	0.03

Table S 76 Distal models for *Aligrama2_IA*

					Mixture Proportions				SE			
Source1	**Source2**	**Source3**	**Source4**	**P**	**S1**	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**
AHG	*WSHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.01	0.28	0.19	0.24	0.30	0.02	0.02	0.03	0.05

Table S 77 Distal models for *Butkara_H*

						Mixture Proportions					SE				
Source1	**Source2**	**Source3**	**Source4**	**Source5**	**P**	**S1**	**S2**	**S3**	**S4**	**S5**	**S1**	**S2**	**S3**	**S4**	**S5**
AHG	*EEHG*	*WSHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.47	0.27	0.06	0.17	0.2	0.3	0.01	0.02	0.03	0.02	0.03
AHG	*WSHG*	*WEHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.33	0.27	0.21	0.02	0.2	0.3	0.01	0.01	0.01	0.02	0.03

Table S 78 Distal models for *Swat_H*

						Mixture Proportions					SE				
Source1	**Source2**	**Source3**	**Source4**	**Source5**	**P**	**S1**	**S2**	**S3**	**S4**	**S5**	**S1**	**S2**	**S3**	**S4**	**S5**
AHG	*WSHG*	*Anatolia_N*	*Ganj_Dareh_N*	..	0.06	0.21	0.22	0.21	0.36	..	0.02	0.02	0.03	0.05	..
AHG	*WSHG*	*WEHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.34	0.22	0.19	0.04	0.18	0.37	0.02	0.02	0.01	0.03	0.05

Table S 79 Distal models for *Barikot_H*

						Mixture Proportions					SE				
Source1	**Source2**	**Source3**	**Source4**	**Source5**	**P**	**S1**	**S2**	**S3**	**S4**	**S5**	**S1**	**S2**	**S3**	**S4**	**S5**
AHG	*WSHG*	*WEHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.05	0.18	0.22	0.02	0.25	0.33	0.02	0.02	0.01	0.03	0.05

Table S 80 Distal models for *Swat_Medieval*

						Mixture Proportions					SE				
Source1	**Source2**	**Source3**	**Source4**	**Source5**	**P**	**S1**	**S2**	**S3**	**S4**	**S5**	**S1**	**S2**	**S3**	**S4**	**S5**
AHG	*WSHG*	*Anatolia_N*	*Ganj_Dareh_N*	..	0.13	0.45	0.12	0.12	0.30	..	0.02	0.02	0.03	0.05	..
AHG	*WSHG*	*WEHG*	*Anatolia_N*	*Ganj_Dareh_N*	0.03	0.47	0.10	0.05	0.12	0.26	0.02	0.02	0.01	0.03	0.05

Table S 81 Distal models for *Saidu_Sharif_H_o*

We are unable to provide proximal models using our current framework as we do not have data from South Asia prior to the Late Bronze-Iron Age, and we have therefore used the *AHG*, which we hypothesize are deeply relate to the ancestry seen in the HG populations of South Asia, as an outgroup. In the next section we present a modeling scheme for South Asia, taking explicit advantage of the clinal nature of human variation in that region.

S5 The Genomic formation of South Asia in light of ancient DNA

S5.1 The clinal nature of ancient and present-day individuals from South Asia and the rationale for our modeling framework

From the PCA and admixture analysis in **Supplementary Materials S4**, we noted that the ancient individuals from South Asia were heterogenous in their ancestry and appeared to lie on several clines that were driven by admixture of highly differentiated source populations. To investigate this further, we began by plotting individuals from South and Central Asia with respect to two simple f_4-statistics measuring their degree of allele sharing with ancient Steppe pastoralists, ancient Iranian farmers, and present-day *Andamanese Hunter Gatherers* (*AHG*). We observe that the populations from South Asia lie on three clines that succeeded each other in time, with each appearing to be a mixture of two ancient populations (**Fig S 49)**.

The *Indus Periphery Cline*, the earliest of the three in time, is represented by a set of ancient individuals from the sites of Shahr-I-Sokhta and Gonur that are genetically distinct from all the other individuals from Iran and Turan, and that are characterized qualitatively (as reflected for example in ADMIXTURE and PCA analysis) by a high proportion of *AHG*-related admixture on the one hand, and on the other hand a West Eurasian-related ancestry type that has a ratio of Iranian to Anatolian farmer-related ancestry different from that in other groups in Turan. Despite all being well described as having ancestry from these two sources, the *Indus Periphery Cline* individuals are far from homogeneous, and instead lie on a gradient of variable proportions of the source ancestries (**Fig S 49)**.

A second observation is that all other ancient and modern individuals from South Asia appear to be admixed with *Central_Steppe_MLBA*-related populations as compared with the *Indus Periphery Cline* individuals. In particular, ancient individuals from northern Pakistan and an outlier individual from Bustan with a date of ~3500 BP appear on another cline that we call the *Steppe Cline*. The *Modern Indian Cline* appears to be formed between a population on the *Indian Periphery Cline* and one on the *Steppe Cline*, as we confirm in what follows.

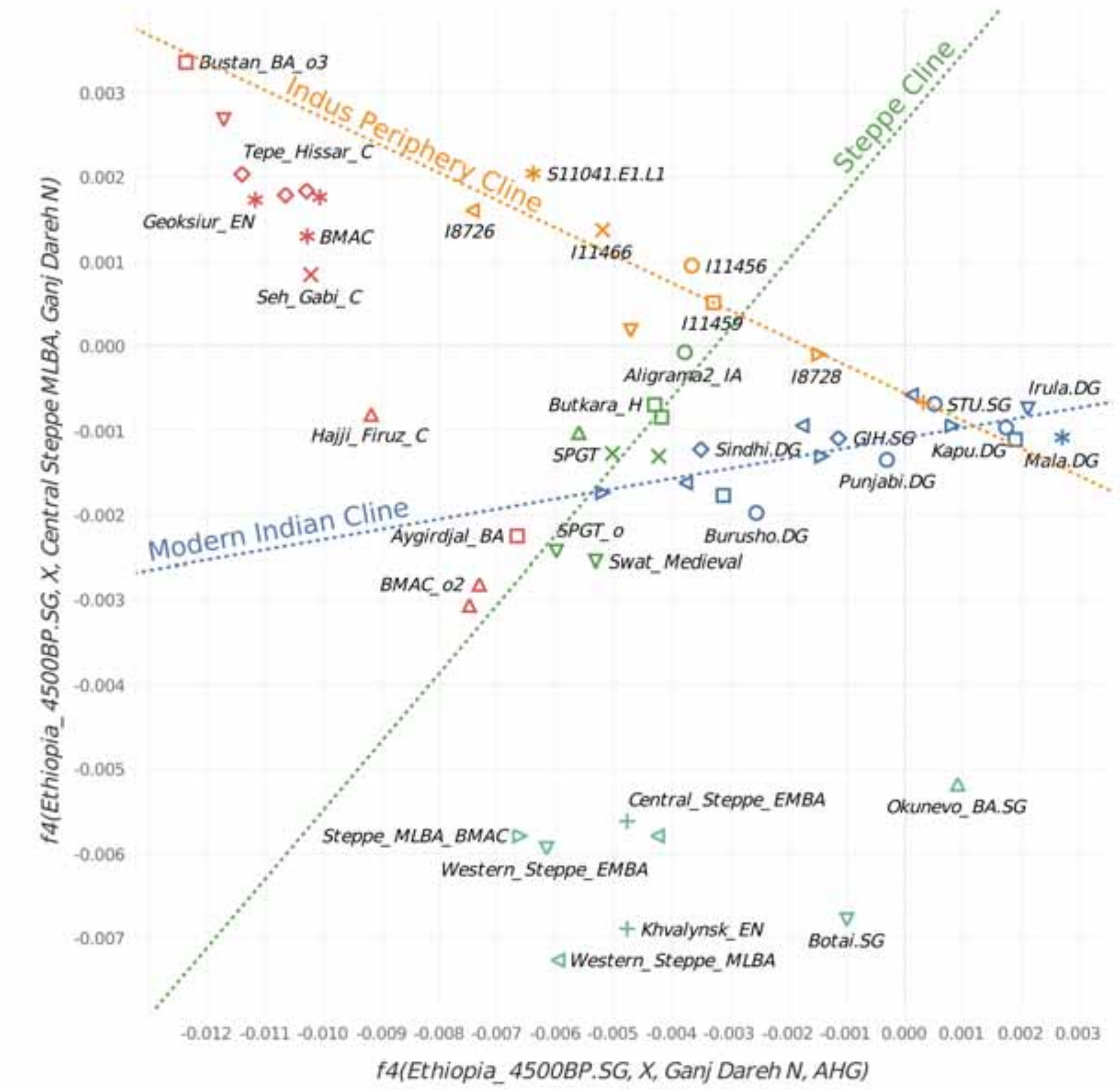

Fig S 49 A plot of f_4-statistics sensitive to *AHG*-related and *Steppe_MLBA*-related ancestry, showing the ancient *Indus Periphery Cline,* an ancient cline defined by the Swat Valley individuals that we call the *Steppe Cline,* and the *Modern Indian Cline*. This plot is the same as Fig. 4A in the main text except that we also plot some non-South Asian groups delineating the *Southwest Asian Cline* and the *Steppe Bronze Age Cline*.

Most population genetic analysis tools focus on modeling single populations in relation to a set of reference populations, and do not provide a facility for jointly modeling multiple populations that lie along admixture clines. However as these clines are so clearly an important part of South Asian population history, we developed a methodology for jointly analyzing populations on two-way admixture clines within a single model.

We begin by establishing a modified set of outgroups for modeling. We need to use a different set of outgroups because the *AHG* (*Andamanese Hunter-Gatherers*) are valuable as a source population for South Asians. Since we need to use the *AHG* as a source in *qpAdm*, we cannot use them in the *Right* outgroup set (**Test 0)**, and compensate for this loss by replacing them in the *Right* outgroup set by *Dai.DG*, a southern Chinese population. We call the modified *Right* outgroup set *Right South Asia* and it contains the following groups:

Right South Asia: Ethiopia_4500BP.SG, WEHG, EEHG, Ganj_Dareh_N, Anatolia_N, WSHG, ESHG, Dai.DG

An overview of our approach is as follows. We begin by observing that individuals from South Asia (both modern and ancient) appear visually to lie on three successive ancestry clines. We used *qpAdm* with the *Right South Asia* outgroup set to examine all possible sources that could be used to fit the ancestry of a large number of representative populations along each cline. For all three clines we identified a unique triple of sources consistent with nearly all groups on the cline. We then co-analyzed all groups on each cline under the theory that they are derived from mixtures in different proportions of two populations admixed more anciently from the three. We use this hierarchical model to extrapolate to the ancestry of the sources, to test if populations were consistent (or not) with lying on more than one of the clines, and to detect individual groups or populations that were significant outliers from the ancestry gradient established by the other individuals along the cline.

S5.2 Hierarchical modeling of ancient and modern clines

S5.2.1 Developing a hierarchical model for estimating ancestry along admixture clines

In this section we detail our hierarchical framework that allowed us to model multiple groups along a two-way admixture cline under a single generative model. We do this in 4 steps:

Step 1: Identify a plausible set of three sources common to a set of individuals.

Prior to modeling a series of individuals or populations jointly, we first determine whether we can obtain a set of populations that work as sources for each of the populations along the cline or at least a substantial majority. We do this using the same approach that we employed in **Supplementary Materials S4** with the exception that we are now choosing to use the outgroup set that we defined above as *Right South Asia.* When applying this method to real data from the *Modern Indian Cline* with over 140 groups to be tested, we chose as test populations 10 representative groups roughly distributed across the ancestry gradient, using individuals for the groups for whom we had genome-wide shotgun data as we did not wish to restrict to the approximately ~600,000 SNPs on the Affymetrix Human Origins array.

Step 2: Estimate admixture proportions (both point estimates and covariances) using *qpAdm.*

Having decided on a set of populations to use as sources for every test population X on the cline, we inferred a vector of ancestry estimates under *qpAdm*:

$$c(X) = (c_A(X), c_B(X), c_C(X)) \qquad \textbf{(Equation 5-1)}$$

where $c_j(X)$ are admixture coefficients for populations A, B and C respectively. We also inferred an associated covariance matrix $V(X)$ using a weighted block jackknife (*220*). We are not assuming here admixture directly from populations A, B and C but instead are postulating admixture from populations that are phylogenetic clades with these groups with respect to the *Right* groups. While these individual estimates provide us with estimates of admixture proportions, they do not provide us with information about the relationship between the various test populations along the cline.

Step 3: Fit these per-group estimates within a single joint clinal model

To enable us to do this, we introduce a hierarchical model, which co-estimates the ancestry proportions for all groups along the cline simultaneously under the assumption of a mixture of two ancestral populations. Suppose that for cline group X the true admixture coefficients for the 3 source populations are $a(X)$ (in contrast to $c(X)$, our *qpAdm* point estimates). Then we take the probability of our data for X, conditional on $a(X)$, to be:

$$P(D|a(X)) = MVN(c(X); a(X), V(X)) \qquad \textbf{(Equation 5-2)}$$

where $MVN(u; m, W)$ is the multivariate normal density for u with mean m and covariance W, and $V(X)$ is an error covariance, estimated by *qpAdm* from the block jackknife.

In our generative model, at the top level we introduce hyper-parameters: *m*, a mean vector invariant across groups, and a covariance matrix *T*. Our generative model is:

(1) For each group *X*, generate a(*X*)**,** normal with mean *m* and covariance *T*. (We think of a(*X*) as the 'true' admixing coefficients for *X*.)

(2) Next, generate *c(X)***,** a random variable, normal with mean *a(X)*, covariance *V(X).*

A technical issue here is that the elements of *a(X)* are constrained to sum to 1, so that the covariance matrices are singular. We solve this by dropping the first coordinate. Thus:

$$a(X) = (a_A(X), a_B(X), a_C(X)) \qquad \textbf{(Equation 5-3)}$$
$$a'(X) = (a_B(X), a_C(X))$$

where $a_A(X)$ can be recovered from $a'(X)$ as $1 - a_B(X) - a_C(X)$. We have a similar definition for *m'* and *c'(X)* with corresponding covariance matrices *T'* and *V'(x).*

Here, *m'* and *T'* are group-independent hyper-parameters, and *a'(X)* and *c'(X)* are the true and estimated proportions of ancestry, respectively, for a particular cline population.

Step 4: Use the obtained maximum likelihood estimates (MLE) for the clinal model to provide conditional estimates for the end of the cline in order to understand the nature of the source populations leading to admixture.

Maximum (composite) likelihood estimates of *m'* and *T'* are readily obtained by an Expectation Maximization (EM) algorithm. We do not have a true likelihood, as distinct populations on the cline will often have correlated genetic drift. We note that the mean estimates for a pair of source populations *A* and *B* are of little interest. For instance, we would shift the mean *m'* by sampling more populations with lower ancestry related to one end of the cline. Of more interest is the distribution of *a'(X)*, the distribution of one population, say *A*, conditional on *B* and *C*, which are in turn also normally distributed, with means and variances that can be computed from the estimates of *m'* and *T'*.

Step 5: Use the Maximum Likelihood Estimate of *m* and *T*, along with the error covariance *V(X)* estimated for each group *X*, to produce a posterior estimates for the ancestry proportions of each source population. We define our hierarchical model as follows:

$$P(c(X|a(X),V(X)) = N(a(X),V(X))$$

$$P(a(X)) = N(m,T) \quad \textbf{(Equation 4-5-4)}$$

Since we have a conjugate prior, the posterior distribution, *P(a(X)|c(X))* is also multivariate normal, $N(\mu_n, \Sigma_n)$ with mean and variance that we can compute analytically.

$$\mu_n = T\left(T + \frac{1}{n}V(X)\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} c(X)_i\right) + \frac{1}{n}V(X)\left(T + \frac{1}{n}(V(X))\right)^{-1} m \quad \textbf{(Equation 4-5)}$$

From this, given observed estimates for a particular cline group *c(X)*, we can obtain posterior estimates for μ.

In the next sections, we apply this modeling framework to 3 different ancestry clines and report the results along with posterior estimates for the ancestry proportions.

S5.2.2 *The Indus Periphery Cline*: An Early to Middle Bronze Age ancestry cline

In **Supplementary Materials S4**, we described the 11 individuals from the Bronze Age sites of Gonur and Shahr-i-Sokhta that define the *Indus Periphery Cline* and that have sufficiently high-quality genetic data to analyze with precision. However, we did not co-model these individuals as they were genetically heterogenous, and they require a different outgroup set in proximal modeling than we were using for the Central Asian analysis (which used *AHG* among the outgroups).

We begin the clinal analysis by running our distal modeling framework on the *Indus Periphery Cline* individuals. Only a single model produces fits above our acceptance threshold for all 11 individuals and this model involves 3 sources: *AHG*, *Ganj_Dareh_N* and *WSHG*. This sort of ancestry is unique in Chalcolithic and Bronze Age Turan, as other individuals including those from sites from which these individuals were found cannot be modeled without additional ancestry related to Anatolian farmers. In **Table S 82**, we list each of the *Indus Periphery Cline* individuals, as well as individuals from Iran and Turan, along with the p-values for the fits of the 3-source model. In **Section S3**, we show working models for the populations from Turan, all of which require Anatolian farmer-related ancestry. This suggests that not only do the *Indus Periphery Cline* individuals contain additional admixture from an *AHG*-related source, but also that they are distinct in their West Eurasian ancestry in that they harbor little to no Anatolian farmer-related ancestry unlike the other individuals from Turan.

Individual	p-value
Indus_Periphery1	0.19
Indus_Periphery2	0.52
Indus_Periphery3	0.43
Indus_Periphery4	0.86
Indus_Periphery5	0.14
Indus_Periphery6	0.31
Indus_Periphery7	0.03
Indus_Periphery8	0.22
Indus_Periphery9	0.03
Indus_Periphery10	0.33
Indus_Periphery11	0.88
BMAC	3.3×10^{-18}
Shahr_I_Sokhta_BA1	1.1×10^{-8}
Sarazm_EN	1.7×10^{-4}
Parkhai_Anau_EN	3.1×10^{-6}
Tepe_Hissar_C	2.4×10^{-15}
Geoksyur_EN	4.0×10^{-8}

Table S 82 *qpAdm* p-values for modeling each of the *Indus Periphery Cline* individuals as well as other individuals from Turan as a mixture of *AHG, Ganj_Dareh_N*, and *WSHG*-related ancestry.

After showing that this 3-way model produces sufficiently good fits for each of the individuals along the cline, we used the individual estimates for each of the individual *qpAdm* runs and fit them jointly to obtain an MLE estimate for the generative model that produces the observed cline, under the assumption that the mixture proportions of *AHG*, *Ganj_Dareh_N* and *WSHG*-related ancestry sum to 1. The mean *AHG* and *Ganj_Dareh_N*-related ancestries and covariance matrix across all individuals are estimated as:

$\boldsymbol{m_{Indus}'}$ = **mean:** $(\mathbf{0.385, 0.0.616})$ **Equation 5-5**

$\boldsymbol{T_{Indus}'}$ = **covariance:** $\begin{pmatrix} \mathbf{0.016} & \mathbf{-0.016} \\ \mathbf{-0.016} & \mathbf{0.016} \end{pmatrix}$ **Equation 5-6**

This provides us with a model along which the *Indus Periphery Cline* individuals lie.

We were interested in the distribution of the conditional ancestries, that is, the proportion of *Ganj_Dareh_N*- and *WSHG*-related ancestry given the proportion of *AHG*-related ancestry. We particularly focused on the case where the proportion of *AHG*-related ancestry was zero, in which case the proportions of *Ganj_Dareh_N*- and *WSHG*-related ancestry are 89 ± 10% and 11 ± 10% respectively. Similarly, on the other end of the cline, the proportions of *AHG*- and *WSHG*-related ancestry (setting *Ganj_Dareh_N*-related ancestry to zero) are 97 ± 10% and 3 ± 10% respectively. While the confidence intervals for *WSHG* ancestry at both ends of the cline overlap 0%, this does not imply that the *Indus Periphery Cline* individual themselves had no *WSHG*-related ancestry (to the contrary, the *qpAdm* analysis proves that they did). What these numbers instead show is that we cannot determine what end of the cline contributed the *WSHG*-related ancestry. Taken together, these analyses suggest that the *Indus Periphery Cline* individuals lie on an admixture gradient between individuals that are of predominantly West and East Eurasian-related ancestry. In addition, the *WSHG*-related admixture proves that ancestry derived ultimately from the *Ancestral North Eurasians* contributed to Turan and South Asia prior to the arrival of Steppe pastoralist-related ancestry

Motivated by these results, we searched for proximal models that fit the *Indus Periphery Cline* individuals, testing all possible sources of ancestries. There is only a single model that provided a consistent fit across all of the individuals that we tested, involving *AHG*, *Parkhai_Anau_EN* and *Sarazm_EN* (**Table S 83**). However, a problem with this as a proximal model for South Asia is that none of these groups were actually from South Asia and the *AHG* is not temporally appropriate either. Moreover, our *qpAdm* modeling using this trio of sources is not able to assign ancestry proportions to each of the modeled groups with small standard errors, because of the ancestry similarity of *Parkhai_Anau_EN* and *Sarazm* (both predominantly Iranian farmer-related ancestry) which means that there is strong negative covariance between the ancestry proportions inferred for each. Thus, for the *Indus Periphery Cline* we use the fitting distal model rather than the fitting proximal one.

We also examined if the *Indus Periphery Cline* individuals could fit as mixtures between an unadmixed East Eurasian-related source (we use *AHG*) and the main clusters of individuals from the same sites (*BMAC* and *Shahr_I_Sokhta_BA1*). **Table S 83** shows the results of two-way fits of this type. Such models all fail, providing further evidence that the *Indus Periphery Cline* individuals were descendants of recent migrants to Turan, plausibly from the Indus Valley or surrounding regions based on archaeological evidence of cultural contacts and their *AHG*-related ancestry.

Individual	*AHG + Parkhai_Anau_EN + Sarazm_EN*	*AHG + BMAC*	*AHG + Shahr_I_Sokhta_BA1*
Indus_Periphery_Pool	0.67	2×10^{-23}	1.5×10^{-8}
Indus_Periphery1	0.65	0.14	0.019
Indus_Periphery2	0.30	2.0×10^{-3}	0.028
Indus_Periphery3	0.64	1×10^{-12}	0.14
Indus_Periphery4	0.089	0.73	0.62
Indus_Periphery5	0.55	0.07	0.17
Indus_Periphery6	0.52	8.6×10^{-7}	7.4×10^{-4}
Indus_Periphery7	0.091	6.8×10^{-5}	0.080
Indus_Periphery8	0.74	2.0×10^{-4}	0.64
Indus_Periphery9	0.90	6.0×10^{-4}	0.65
Indus_Periphery10	0.17	3.7×10^{-4}	1.1×10^{-3}
Indus_Periphery11	0.50	4.9×10^{-4}	4.9×10^{-4}

Table S 83 *qpAdm* p-values using the *RightSouthAsia* set for modeling each of the *Indus Periphery Cline* individuals and different mixtures of sources.

We also evaluated if we could model the ancestry of the *Indus Periphery Cline* individuals with less West Eurasian-related ancestry as a mixture of *AHG* and the *Indus Periphery Cline* individual with the maximal West Eurasian-related ancestry *(Indus_Periphery_West)*. All such models were feasible with p>0.01. We also tested if this were true even when *BMAC* or *Shahr_I_Sokhta_BA1* were in the outgroups, thereby testing if ancestry from the main cluster of *BMAC* or Shahr-i-Sokhta samples contributed directly to the ancestry seen in the Indus Periphery samples and found that the fit remained good. These observations, coupled with the fact that we date the admixture between the East Eurasian source and the West Eurasian source to several millennia earlier than the date of these samples (see **Supplementary Materials S4**), and the fact that the nature of their West Eurasian-related ancestry is different from all other samples from Iran and Turan, suggests that these individuals were indeed migrants to Turan who already had substantial proportions of West Eurasian-related ancestry and not individuals with entirely East Eurasian admixture mixing with local residents at BMAC and Shahr-i-Sokhta.

Having developed our model for the *Indus Periphery Cline* as derived distally from *AHG-*, *WSHG-*, and *Ganj_Dareh_N*-related groups, we next examined whether each of the 140 present day populations from the *Modern Indian Cline* are compatible with lying on this cline. We initially tested whether individual groups could be modeled with the same set of three sources with respect to the outgroups (based on individual *qpAdm* models). We then carried out a second test, estimating proportions for *AHG-*, *WSHG-* and *Ganj_Dareh_N-*related ancestries, and evaluating whether these were consistent with the proportions determined by the joint modeling of the 11 *Indus Periphery Cline* individuals.

Out of the 140 population groups, 55 could be modeled (at p>0.01) using just the 3 sources of ancestry that were plausible sources for the *Indus Periphery Cline.* We then tested each of these 55 groups to determine whether their estimated ancestry proportions were consistent with that from our generative model (that is, derived from the two source populations mixed more anciently from the three). Only 38 groups fit after this additional round of testing. This group was heavily enriched in Dravidian speaking tribal groups from southern India (**Table S 5**), suggesting that their ancestry is close to lying on the *Indus Periphery Cline.*

S5.2.3 The *Steppe Cline:* A South Asian Ancestry Gradient formed by a Mixture of a point on the *Indus Periphery Cline* and MLBA Steppe Pastoralists

We next examine sources for ancient populations in our dataset that are geographically within South Asia. We focused on 7 sets of individuals from the Swat Valley and Chitral. These individuals date to ~1000 BCE (*SPGT*, 85 individuals), ~1000 BCE (*SPGT_o*, 2 individuals with significantly larger proportions of Steppe pastoralist-related ancestry), ~750 BCE (*Aligrama2_IA*, 3 individuals), ~350 BCE (*Swat_H*, 13 individuals), ~50 CE (*Butkara_H*, 4 individuals), 500 CE (*Barikot_H*, 3 individuals), and ~1200 CE (*Swat_Medieval*, 5 individuals).

Motivated by our work in previous sections showing that the Swat Valley individuals could be modeled as a complex mixture of 5 distal ancestry sources (*AHG*, *Ganj_Dareh_N*, *Anatolia_N*, *WEHG*, *EEHG*), we employed our proximal modeling with our *Right South Asia* outgroup set to search for a model that fit all of the populations along this cline. **Fig S 50** shows that at a p>0.01 threshold, the standard we used in our proximal modeling, there was no model that worked for all the populations. We then lowered the threshold to 0.005 to allow for additional fitting models. An acceptance criterion at this threshold is reasonable given the number of tests we are carrying out.

As we saw with the modeling of the *Indus Periphery Cline,* at this level of significance, the *BMAC* and other populations from Turan with high Iranian and Anatolian farmer-related ancestry are not plausible sources for the ancestry we see in the Late Bronze-Iron Age and Historical Swat Valley. The only possible source populations that provide a fit to most of the populations along the *Steppe Cline* are a combination of *AHG*, *Indus_Periphery_Pool*, and a population from a Steppe pastoralist-related source from the Middle to Late Bronze Age (*Central_Steppe_MLBA* or *Western_Steppe_MLBA*). However, groups later in time with substantial East Asian-related ancestry (*Steppe_LBA*) do not provide a good fit. In other words, these results show that Steppe Middle to Late Bronze Age populations are a component in the genetic ancestry of the *Steppe Cline* for all fitting models. Furthermore there is no evidence of such ancestry in the Indus Periphery Cline and we conjecture that such ancestry will not be found in individuals from the Indus Valley Civilization itself (IVC) at least for dates earlier than around 2000 BCE. Given the East Asian ancestry pervasive on the Steppe in the LBA. this leaves only a rather narrow time window for the arrival of Steppe ancestry into South Asia, that is, between around 2000-1500 BCE.

We note that the population from the Medieval period, *Swat_Medieval*, is from an archeological context related to that of the Ghaznavid dynasty, of historically known Turkic Mamluk origin, and can be modeled only with *BMAC*-related ancestry but not with that of *Indus Periphery Cline* individuals. This plausibly reflects admixture that occurred during the historical period, and we remove this individual from the set of individuals we refer to as the *Steppe Cline*.

An additional plausible set of models that fit are those involving a *Shahr_I_Sokhta_BA1*-related population from eastern Iran mixing with a *Steppe_LBA* population. While this fits 4 out of the 6 populations that define the cline, we excluded *Shahr_I_Sokhta_BA1* as a plausible source in further analysis for two reasons. First, when we modeled the main *SPGT* group of Late Bronze-Iron Age Swat Valley individuals adding *Shahr_I_Sokhta_BA1* as a source along with *Indus_Periphery_Pool*, *AHG* and *Central_Steppe_MLBA*, not only is *Shahr_I_Sokhta_BA1* not required to produce a good fit to the data, but its inferred ancestry contribution is consistent with zero (5±11%). Second, for modeling *SPGT*, we also ran our model competition procedure on the two classes of models that fit in **Fig S 50**. We found that adding *Shahr_I_Sokhta_BA1* into the outgroup set for the model with *Indus_Periphery_Pool* as a source does not induce a bad fit (p=0.01 for *SPGT* as the target population). However, adding *Indus_Periphery_Pool* into the model involving *Shahr_I_Sokhta_BA1* as a source causes clear failure (p=1.5 x 10^{-31}). Third, the mean date of the *Shahr_I_Sokhta_BA1* individuals and that of *Steppe_LBA* are almost two millennia apart, making them unlikely to be the combination combination of sources for South Asia.

	Steppe pastoralist-related	SPGT	SPGT_o	Aligrama2_IA	Swat3_H	Swat_H	Barikot_H	Swat_Medieval
Aygirdjal_BA	Botai.SG	0.012	0.000	0.163	0.163	0.000	0.003	0.000
	Central_Steppe_EMBA	0.034	0.000	0.787	0.407	0.000	0.004	0.000
	Central_Steppe_MLBA	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Khvalynsk_EN	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Okunevo_BA.SG	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Steppe_LBA	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Steppe_MLBA_oBMAC	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Western_Steppe_EMBA	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Western_Steppe_MLBA	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BMAC	Botai.SG	0.000	0.000	0.024	0.014	0.000	0.000	0.000
	Central_Steppe_EMBA	0.000	0.000	0.017	0.033	0.000	0.000	0.000
	Central_Steppe_MLBA	0.000	0.129	0.000	0.000	0.000	0.000	0.027
	Khvalynsk_EN	0.000	0.000	0.000	0.000	0.000	0.005	0.000
	Okunevo_BA.SG	0.000	0.000	0.001	0.002	0.000	0.000	0.000
	Steppe_LBA	0.000	0.777	0.000	0.000	0.000	0.006	0.000
	Steppe_MLBA_oBMAC	0.000	0.507	0.000	0.000	0.000	0.002	0.038
	Western_Steppe_EMBA	0.000	0.227	0.000	0.000	0.005	0.005	0.000
	Western_Steppe_MLBA	0.000	0.009	0.000	0.000	0.000	0.000	0.001
Geoksyur_EN	Botai.SG	0.000	0.000	0.728	0.000	0.000	0.000	0.000
	Central_Steppe_EMBA	0.000	0.000	0.768	0.001	0.000	0.000	0.000
	Central_Steppe_MLBA	0.001	0.874	0.073	0.031	0.039	0.014	0.000
	Khvalynsk_EN	0.000	0.000	0.610	0.004	0.000	0.000	0.000
	Okunevo_BA.SG	0.000	0.000	0.470	0.000	0.000	0.000	0.000
	Steppe_LBA	0.000	0.071	0.296	0.285	0.000	0.000	0.000
	Steppe_MLBA_oBMAC	0.010	0.655	0.125	0.032	0.050	0.031	0.000
	Western_Steppe_EMBA	0.000	0.001	0.330	0.016	0.000	0.000	0.000
	Western_Steppe_MLBA	0.000	0.518	0.043	0.010	0.004	0.005	0.001
Hajji_Firuz_C	Botai.SG	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Central_Steppe_EMBA	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Central_Steppe_MLBA	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Khvalynsk_EN	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Okunevo_BA.SG	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Steppe_LBA	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Steppe_MLBA_oBMAC	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Western_Steppe_EMBA	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Western_Steppe_MLBA	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Indus_Periphery_Pool	Botai.SG	0.000	0.000	0.021	0.000	0.000	0.000	0.000
	Central_Steppe_EMBA	0.000	0.000	0.046	0.000	0.000	0.000	0.000
	Central_Steppe_MLBA	0.006	0.563	0.823	0.019	0.021	0.016	0.000
	Khvalynsk_EN	0.000	0.000	0.283	0.000	0.000	0.000	0.000
	Okunevo_BA.SG	0.000	0.000	0.023	0.000	0.000	0.000	0.000
	Steppe_LBA	0.000	0.003	0.852	0.005	0.000	0.000	0.000
	Steppe_MLBA_oBMAC	0.001	0.175	0.845	0.005	0.000	0.003	0.000
	Western_Steppe_EMBA	0.000	0.000	0.722	0.000	0.000	0.000	0.000
	Western_Steppe_MLBA	0.003	0.512	0.783	0.020	0.058	0.027	0.000
Namazga_CA.SG	Botai.SG	0.000	0.000	0.149	0.000	0.000	0.000	0.000
	Central_Steppe_EMBA	0.000	0.000	0.183	0.000	0.000	0.000	0.000
	Central_Steppe_MLBA	0.000	0.284	0.000	0.000	0.000	0.000	0.000
	Khvalynsk_EN	0.000	0.000	0.011	0.000	0.000	0.000	0.000
	Okunevo_BA.SG	0.000	0.000	0.094	0.000	0.000	0.000	0.000
	Steppe_LBA	0.000	0.094	0.015	0.000	0.011	0.000	0.000
	Steppe_MLBA_oBMAC	0.000	0.229	0.001	0.000	0.001	0.001	0.000
	Western_Steppe_EMBA	0.000	0.000	0.003	0.000	0.000	0.000	0.000
	Western_Steppe_MLBA	0.000	0.052	0.000	0.000	0.000	0.000	0.000
Parkhai_Anau_EN	Botai.SG	0.000	0.000	0.353	0.000	0.000	0.000	0.000
	Central_Steppe_EMBA	0.000	0.000	0.692	0.000	0.000	0.000	0.000
	Central_Steppe_MLBA	0.000	0.555	0.001	0.000	0.002	0.000	0.000
	Khvalynsk_EN	0.000	0.000	0.149	0.000	0.000	0.000	0.000
	Okunevo_BA.SG	0.000	0.000	0.052	0.000	0.000	0.000	0.000
	Steppe_LBA	0.000	0.222	0.091	0.006	0.002	0.001	0.000
	Steppe_MLBA_oBMAC	0.000	0.541	0.003	0.000	0.032	0.002	0.000
	Western_Steppe_EMBA	0.000	0.000	0.030	0.000	0.000	0.000	0.000
	Western_Steppe_MLBA	0.000	0.087	0.000	0.000	0.000	0.000	0.000
Sarazm_EN	Botai.SG	0.000	0.000	0.210	0.000	0.000	0.000	0.000
	Central_Steppe_EMBA	0.000	0.000	0.148	0.000	0.000	0.000	0.000
	Central_Steppe_MLBA	0.000	0.085	0.154	0.000	0.000	0.001	0.000
	Khvalynsk_EN	0.000	0.000	0.075	0.000	0.000	0.000	0.000
	Okunevo_BA.SG	0.000	0.000	0.139	0.000	0.000	0.000	0.000
	Steppe_LBA	0.000	0.000	0.089	0.000	0.000	0.000	0.000
	Steppe_MLBA_oBMAC	0.000	0.014	0.131	0.000	0.000	0.000	0.000
	Western_Steppe_EMBA	0.000	0.000	0.086	0.000	0.000	0.000	0.000
	Western_Steppe_MLBA	0.000	0.322	0.192	0.001	0.000	0.006	0.000
Seh_Gabi_C	Botai.SG	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Central_Steppe_EMBA	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Central_Steppe_MLBA	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Khvalynsk_EN	0.000	0.019	0.000	0.000	0.000	0.000	0.021
	Okunevo_BA.SG	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Steppe_LBA	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Steppe_MLBA_oBMAC	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Western_Steppe_EMBA	0.000	0.012	0.000	0.000	0.000	0.000	0.000
	Western_Steppe_MLBA	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Shahr_I_Sokhta_BA1	Botai.SG	0.000	0.000	0.280	0.000	0.000	0.000	0.000
	Central_Steppe_EMBA	0.000	0.000	0.355	0.001	0.000	0.000	0.000
	Central_Steppe_MLBA	0.000	0.488	0.000	0.000	0.000	0.002	0.008
	Khvalynsk_EN	0.000	0.000	0.090	0.000	0.000	0.000	0.000
	Okunevo_BA.SG	0.000	0.000	0.070	0.000	0.000	0.000	0.000
	Steppe_LBA	0.010	0.385	0.003	0.013	0.001	0.006	0.000
	Steppe_MLBA_oBMAC	0.000	0.682	0.000	0.000	0.004	0.019	0.001
	Western_Steppe_EMBA	0.000	0.011	0.005	0.000	0.000	0.000	0.000
	Western_Steppe_MLBA	0.000	0.092	0.000	0.000	0.000	0.000	0.002
SPGT	Botai.SG		0.000	0.000	0.016	0.000	0.091	0.000
	Central_Steppe_EMBA		0.000	0.000	0.018	0.000	0.094	0.000
	Central_Steppe_MLBA		0.795	0.149	0.187	0.014	0.386	0.000
	Khvalynsk_EN		0.000	0.000	0.055	0.000	0.149	0.000
	Okunevo_BA.SG		0.000	0.000	0.013	0.000	0.082	0.000
	Steppe_LBA		0.203	0.010	0.063	0.001	0.221	0.000
	Steppe_MLBA_oBMAC		0.620	0.058	0.213	0.009	0.390	0.000
	Western_Steppe_EMBA		0.064	0.002	0.112	0.003	0.213	0.000
	Western_Steppe_MLBA		0.787	0.213	0.219	0.021	0.411	0.000
Tepe_Hissar_C	Botai.SG	0.000	0.000	0.030	0.023	0.000	0.000	0.000
	Central_Steppe_EMBA	0.000	0.000	0.022	0.116	0.000	0.000	0.000
	Central_Steppe_MLBA	0.000	0.027	0.000	0.000	0.000	0.000	0.003
	Khvalynsk_EN	0.000	0.001	0.000	0.001	0.000	0.001	0.000
	Okunevo_BA.SG	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Steppe_LBA	0.000	0.754	0.000	0.000	0.000	0.000	0.000
	Steppe_MLBA_oBMAC	0.000	0.370	0.000	0.000	0.000	0.000	0.048
	Western_Steppe_EMBA	0.000	0.330	0.000	0.000	0.000	0.000	0.000
	Western_Steppe_MLBA	0.000	0.001	0.000	0.000	0.000	0.000	0.000

Fig S 50 Testing different sources for the Swat Valley populations from ancient northwest South Asia. In green are fits according to the acceptance criterion (p>0.001).

After establishing a set of source populations that provide fits, we proceed to our hierarchical modeling. We obtained a joint MLE estimate for the populations along the cline and then estimated the conditional ancestries.

Extrapolating to the expectation for 0% *AHG*-related ancestry, we estimate proportions of 77 ± 7% *Indus_Periphery_Pool*-related and 23 ± 7% *Central_Steppe_MLBA*-related ancestry, showing that the Steppe pastoralists entering South Asia were already admixed with individuals local to the area prior to the subsequent mixture that formed the *Steppe Cline*. In other words, an *ANI* population that had only West Eurasian-related ancestry and was a proximal source of the *Modern Indian Cline* may never have existed.

Extrapolating to the expectation for 0% *Central_Steppe_MLBA* ancestry, we estimate the proportions of *AHG*-related and *Indus_Periphery_Pool*-related ancestries at 10±4% and 90±4% respectively. Thus, the more South Asian hunter-gatherer-related end of the *Steppe Cline* has proportions of ancestry fairly similar to that of the *Indus_Periphery_Pool*.

Having obtained these estimates for the ends of the cline, we wanted to ascertain if any of the *Modern Indian Cline* groups are consistent with lying on the *Steppe Cline.* We determined that not a single group on the *Modern Indian Cline* is compatible with lying on the *Steppe Cline,* in the sense that all individuals on the *Steppe Cline* have too low a proportion of Steppe pastoralist-related ancestry given their overall proportion of West Eurasian-related ancestry to be consistent with those on the *Modern India Cline*. This suggests that the present-day populations of South Asia had input from a Steppe pastoralist source to a far greater extent than that of the populations we sampled from the ancient Swat Valley.

S5.2.4 The Modern Indian Cline: A Mixture Between a Point on the Indus Periphery Cline (ASI) and a point on the Steppe Cline (ANI)

In the modeling of the *Steppe Cline* we noted that our only working model for the Swat Valley Time transect individuals involved a mixture of people related to individuals from the *Indus Periphery Cline,* and peoples related to Middle to Late Bronze Age Steppe pastoralists. As we had done for the other two clines, we wanted to examine all possible sources for a fit to the *Modern Indian Cline*. Prior to testing all 140 populations on the cline, we returned to the 10 populations along the cline for which we had data at 1.2 million positions.

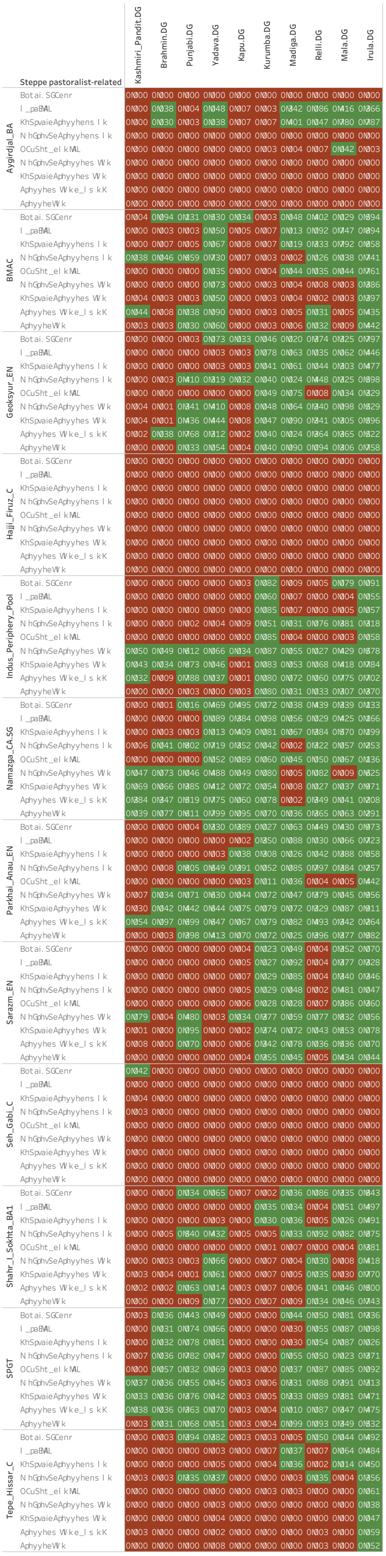

Fig S 51 *qpAdm* p-values for different combinations of sources for 10 populations along the *Modern Indian Cline,* in all cases involving *AHG* and two ancient groups.

We report our results in **Fig S 51**, detailing different populations that could fit as sources for the Steppe pastoralist-related ancestry in the *Modern Indian Cline*, always including *AHG* in the model as a surrogate for the unsampled *AASI* source. All the plausible models we tested that fit all 10 populations involved pairs of source populations including *Western_Steppe_MLBA* or *Central_Steppe_MLBA* on the one hand, and *Indus_Periphery_Pool* on the other. We also obtained passable fits for many of the *Modern Indian Cline* groups when using one as the source of Steppe pastoralist-related ancestry and either *BMAC, Namazga_CA.SG*, or *Pakhai_Anau_EN* as the other. However, there are several problems with these models. First, only the model involving *Indus_Periphery_Pool* also works for the *Steppe Cline* groups; thus, if we apply a parsimony rule search for a single set of three source populations that works both for the *Steppe Cline* and the *Modern Indian Cline, Indus_Periphery_Pool* is required. Second, models involving the *Indus_Periphery_Pool* as a source are the only ones that make sense based on all of genetic, archaeological and temporal evidence. Specifically, our modeling of population movement in Turan in **Supplementary Materials S4**, as well as the timing of the admixture between Steppe pastoralist-related ancestry and South Asian ancestry in **Supplementary Materials S4**, both show that the only temporally and geographically plausible sources for the Steppe ancestry in South Asia are Middle to Late Bronze Age Steppe pastoralists, Our genetic analyses show that these admixed with individuals in Turan and South Asia after the turn of the second millennium BCE. Taking these lines of evidence together, we focus on the trio of sources *Central_Steppe_MLBA* (the geographically more proximate Steppe pastoralist-related group), *Indus_Periphery_Pool*, and *AHG* in what follows.

As South Asia is known to have been impacted by additional historical migrations, we also assessed if later groups we sampled from the late Iron Age and historical periods from the Steppe and Turan made genetic impacts on the *Modern Indian Cline.* By co-analyzing our data with additional Iron Age samples from (*29*), we show that the Sakas, Scythians, Kushans and Huns which are known to have had cultural impacts on South Asia in the Iron Age and afterward, had little genetic impact on most present-day South Asians **Fig S 52**. This provides evidence (along with admixture dating of the ancient Swat Valley and modern Indian individuals) that the Steppe pastoralist-related ancestry that is widespread in South Asia today arrived in the Late Bronze Age. The most probable source is through *Central_Steppe_MLBA* migrants who we show in **Supplementary Materials S4** began to impact Turan by the turn of the 2^{nd} millennium BCE as reflected by outliers at *BMAC* sites mixed with this ancestry.

We were unable to reject a single Iron Age population, *Kazakhstan_Kangju.SG* as a source, though their time period , ~200-300 CE, is much too late for them to be a viable source based on the fact that Late Bronze-Iron Age populations from South Asia from almost a millennia earlier already have substantial amounts of Steppe ancestry, and the fact that our admixture timing estimates on the modern individuals provide dates for the admixture of Steppe pastoralist-related and Iranian farmer-related ancestry in South Asia in the 2nd millennium BCE.

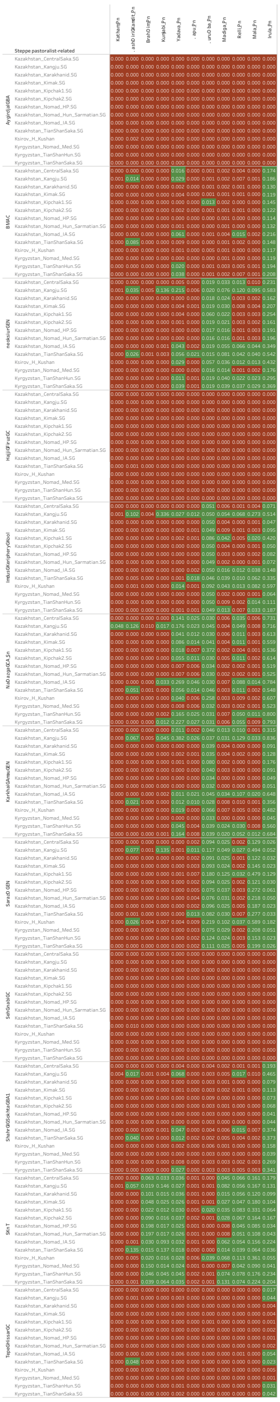

Fig S 52 Tests for fits of populations with Steppe pastoralist-related ancestry from different periods along with individuals from two points of the *Indus Periphery Cline.*

Using the trio of sources *AHG*, *Indus_Periphery_Pool* and *Central_Steppe_MLBA* that we established as described above, we carried out formal clinal analysis of these individuals. To do this we extended our modeling to a much larger number of present-day Indian groups using data from 140 groups from South Asia genotyped on the Affymetrix Human Origins array and curated to be consistent with the *Modern Indian Cline* as described in

Supplementary Materials S3. While these individuals had data on fewer than half the number of SNPs analyzed above, the dataset has the advantage of allowing us to explore much larger sample sizes and also to explore the full extent of ancestry variability along the *Modern Indian Cline*.

We find that the proportions of *Indus_Periphery_Pool*-related and *Central_Steppe_MLBA*-related ancestries are 70±4% and 30±4% respectively for the extrapolated population with 0% *AHG*-related ancestry.

We find that the proportions of *AHG*- and *Indus_Periphery_Pool*-related ancestries are 53±7% and 47±7% respectively for the extrapolated population with 0% *Central_Steppe_MLBA*-related ancestry.

We give the standard errors under the jackknife, but systematic errors are certainly larger as the value depends on the exact source of populations that contributed Iranian farmer-related ancestry (here chosen to be *Indus_Periphery_Pool*) and Steppe pastoralist-related ancestry (here chosen to be *Central_Steppe_MLBA)*, as well as our manual curation procedure that we used to identify a set of 140 present-day populations to represent the *Modern Indian Cline*.

Given the assumption that these 140 groups all lie on a two-way admixture cline (derived from two source populations admixed more anciently from the three), we were able to obtain *Maximum A Posteriori* estimates of *AHG*-related, *Indus_Periphery_Pool*-related, and *Central_Steppe_MLBA* ancestry in all of the groups. When doing so, we were also able to ascertain if any of the groups were consistent with sitting at the theoretical extremes of the cline, as well as to identify any groups that were significant outliers from the cline.

We find that several tribal groups from southern India speaking Dravidian languages were consistent with a proportion of ancestry that was on the *ASI* extreme of the cline (defining the *ASI* or *Ancestral South Indians* as the point on the cline with the maximum allowed percentage of *AHG*-related ancestry). Notably, these groups were also determined to be consistent with being on the *Indus Periphery Cline* as described above. Putting these two observations together shows that the *ASI*, one of the two founding populations of the *Modern Indian Cline*, derives from a combination of ancestries that drove the *Indus Periphery Cline*.

Eight *Modern Indian Cline* groups (all tribal groups from southern India speaking Dravidian languages) in **Table S 5** that are consistent with the proportion expected for an entirely *ASI* ancestry group in the sense that *qpAdm* analysis of these groups as part of the *Indus Periphery Cline* confirms that these groups fit without any Steppe pastoralist-related ancestry (**Table S 84**). These results show that the *ASI* component of the *Modern Indian Clin*e has substantial Iranian farmer-related ancestry (via the *Indus Periphery Cline*), and that a subset of groups derive the great majority of their ancestry from the ancient *ASI*, which as described above we estimate had 53±7% *AHG*-related ancestry.

Group	**3 Sources: *AHG, Indus_Periphery_Pool, Central_Steppe_MLBA***		**2 Sources: *AHG, Indus_Periphery_Pool***	
	p-value	*AHG*-related ancestry	p-value	*AHG*-related ancestry
Palliyar	0.95	0.57	0.44	0.62
Yanidi	0.73	0.47	0.27	0.46
Ulladan	0.35	0.62	0.18	0.61
Gugavellalar	0.91	0.49	0.09	0.47
Irula	0.22	0.57	0.09	0.56
Pulliyar	0.63	0.57	0.08	0.55
Adiyan	0.30	0.66	0.06	0.64
Malayan	0.39	0.59	0.04	0.57

Table S 84 *qpAdm* estimates in *Modern Indian Cline* groups with greatest *AHG* affinity (all the standard errors are ±0.02).

On the other end of the cline, we find that the *Kalash*, estimated to have 66% ± 2% *Indus_Periphery_Pool*-related ancestry, harbor close to the theoretical maximum of 70±4%. This group can be fit under our model with a p-value of 0.0001 (not an egregious outlier given a Bonferroni correction for testing the same working model against 140 different populations). However, the *Kalash* are confidently estimated to have non-negligible *AHG*-related ancestry (4±1%), and thus while it is possible that they had the same mixture of ancestries as the *ANI*, it is also possible that the *ANI* was an unsampled population that existed in the past that was the true source of the Steppe pastoralist-related ancestry in the *Modern Indian Cline* and that had even less *AHG*-related ancestry.

While our formulation of the clinal model as specified by these 3 sources provides a good fit, one drawback of our framework could be that we are using the *AHG* as a proxy for the *AASI* even though the two populations are deeply diverged from one another in time (**Fig S 56**). We decided to address this directly by reformulating our clinal model, separating the *Indus*

Periphery Cline individuals into two groups *Indus_PeripheryA* (6 individuals from the *Indus Periphery Cline* with the highest proportions of *AHG*-related ancestry) and *Indus_PeripheryB* (5 individuals from the *Indus Periphery Cline* with the lowest proportions of *AHG*-related ancestry). This allowed us to represent the *Indus Periphery Cline* individuals as two points that lie on the *Indus Periphery Cline*. We can then use these two points directly as sources for the *Steppe Cline* as well as the *Modern Indian Cline* (along with *Central_Steppe_MLBA*). This procedure avoids the use of the *AHG* entirely in our modeling.

Using this reformulation, we find that we are able to model almost all of the populations with improved fits (**Table S 5**) which provides further confidence in our model. A virtue of this procedure is that it is not at all clear that *AHG* and *AASI* are a true clade. With such long time separations, there is plenty of scope for there to have been further admixtures into the ancestors of the *AASI* or into the ancestors of Andaman islanders that we have not detected, although by choosing outgroup sets that only include *ESHG* and *Dai.DG* as East Asian-related populations--both of which are only very anciently related to either the *AHG* or the *AASI*--we avoid having our analyses be very sensitive to admixture among East Asian-related lineages While we report the results of the analyses using this formulation in **Table S 5**, in the main text we primarily discuss the *AHG* formulation because of ease of interpretation.

S5.3 The *Modern Indian cline* is not a perfect mixture of two ancestral populations

We have based much of the preceding argument on the hypothesis that the *Modern Indian Cline* groups can be approximately modeled as a mixture of two ancestral populations, the *ANI* and the *ASI*, each of which has a characteristic ratio of ancestry. As we did with the *Indus Periphery Cline* and the *Steppe Cline*, we now demonstrate that this model is not perfect, and that deviations from the model are interesting.

The first evidence that the model is not perfect comes from inspecting the individual *qpAdm* fits for the 140 Indian Cline groups in **Table S 5**, as discussed above. In **Table S 85** show all outliers that are particularly poor fits at p<0.0001, the threshold below Bonferroni correction for testing 140 populations.

Population	***Modern Indian Cline qpAdm* p-value**
Pathan	8.0×10^{-14}

Khatri	7.9×10^{-7}
Panta_Kapu	8.6×10^{-7}
Brahmin_Nepal	1.4×10^{-6}
Dogra	2.0×10^{-6}
Muslim_Kashmiri	3.0×10^{-6}
Sikh_Jatt	4.5×10^{-6}
Brahmin_Catholic	8.3×10^{-6}
Gaud_Karnataka	4.2×10^{-5}

Table S 85 *qpAdm* fits from Table S 5 for groups with p<0.0001

Population	**Proportion of *Central_Steppe_MLBA*-related ancestry**	**Ratio of *Central_Steppe_MLBA* to *Indus_Periphery_Pool*-related ancestry**	***Modern Indian Cline* pool fit Z-score**
Brahmin_Tiwari	0.27	0.55	-7.9
Bhumihar_Bihar	0.28	0.56	-7.0
Brahmin_UP	0.26	0.54	-6.4
Sikh_Jatt	0.26	0.48	-6.2
Brahmin_Nepal	0.26	0.54	-5.4
Bhumihar_UP	0.25	0.48	-5.0
Yadav_Pondicherry	0.066	0.11	5.0
Gaud_Karnataka	0.061	0.10	5.1
Coorghi	0.098	0.15	5.2
Nadar	0.046	0.08	5.8
Kallar	0.050	0.08	6.6
Vysya	0.040	0.07	7.1
Panta_Kapu	0.064	0.10	8.8

Table S 86 *Modern Indian Cline* groups that are particularly poorly fit by a model of mixture of two ancestral populations (*ANI* and *ASI*). We list all groups with deviations significant at |Z|>4.5; negative values indicate a ratio of Steppe pastoralist- to Iranian farmer-related ancestry that is larger than that in the rest of the *Modern Indian Cline* after adjusting for *AHG* ancestry proportion. Positive values indicate a ratio of Steppe- to Iranian farmer-related ancestry that is significantly smaller than that in the rest of the *Modern Indian Cline*. We also include our estimates of *Central_Steppe_MLBA*-related ancestry along with estimates of the ratio of *Central_Steppe_MLBA*- to *Indus_Periphery_Pool*-related ancestry for each group.

As a second measure of fit, we can examine outliers from our clinal modeling, using our conditional estimates of whether the *Central_Steppe_MLBA*-to-*Indus_Periphery_Pool* ancestry ratio is consistent with the expectation of our clinal model. In **Table S 86**, we show Z-scores for deviations from the expectation of the clinal 2-way model. Negative, Z-scores have excess Steppe pastoralist-related ancestry compared to the expectation from the model, whereas groups with positive Z-scores have excess Iranian farmer-related ancestry.

S5.4 Evidence for excess Steppe pastoralist-related ancestry in groups of traditionally priestly status

A striking observation is that the majority of groups with excess Steppe pastoralist-related ancestry compared to the expectation based on their proportion of West Eurasian-related ancestry are groups of traditionally priestly status (Brahmins) or that traditionally see themselves as being of priestly status (Bhumihars). In particular, we find that on the Indian Cline, 5 of the 6 groups with $Z < -4.5$ are in this category, even though only 10 of the 140 groups are in this category overall (**Table S 5** and **Table S 86**). If we were to treat the individual groups as independent statistical observations (unrelated to each other since they were formed by admixture) this would correspond to $p=1\times10^{-4}$ by a χ^2 test.

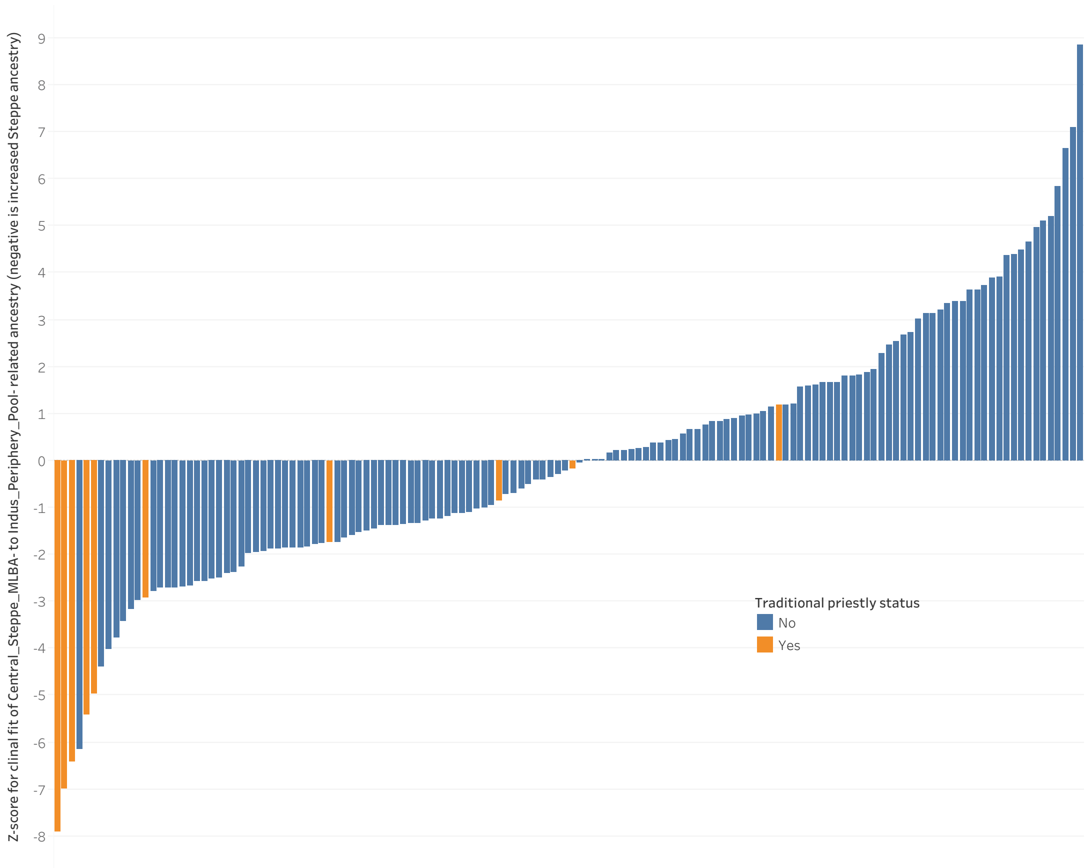

Fig S 53 Distribution of Z-scores for the ratio of Steppe pastoralist-related to *Indus_Periphery_Pool*-related ancestry for Brahmins and Bhumihars of traditionaly priestly status and other groups. This is a version of Fig. 4C in the main text.

The pattern shown in **Fig S 53** suggests shows that some groups that view themselves as being of priestly status in the traditional caste system have significant excess Steppe ancestry.

These results are intriguing in light of the fact that Sanskrit—the language of the ancient Vedic texts—is Indo-European. Brahmins in India are traditionally custodians of texts written in Sanskrit, so these results point to an additional line of evidence—beyond the evidence for a correlation of Indo-European languages to *Western_Steppe_EMBA*-derived admixture both in Europe and in South Asia (*9*)—for an origin of at least some aspects of Indo-European culture in South Asia in the Steppe.

A concern with this analysis is that we are not matching group that see themselves as being of priestly status in this analysis with ones of similar West Eurasian-related ancestry proportion. To address this, we examined the distribution of Z-scores for each grouping (priestly status or non-priestly status). Specifically, we compared the mean rank of these groups with that of other populations that were matched to be within one standard deviation of the mean *AHG*-related ancestry by subsampling different permutations of these matched populations 1000 times. When compared with the null distribution generated by this subsampling, with mean rank (69) and standard deviation of the rank (12), we show that the observed mean rank of the groups seeing themselves as being of traditionally priestly status (31) is an outlier at $p=8\times10^{-6}$ for all South Asians.

One possibility for such a significant deviation in the groups that see themselves as being of traditionally priestly status is that these groups are related to one another more recently than other population groups in India, and therefore we are not actually capturing independent signals from different Brahmin groups. However, previous work using the same dataset has not detected Brahmin groups as sharing IBD segments between each other at rates higher than average for all pairwise IBD calculations computed for all 140 groups (*10*), providing some line of evidence against this hypothesis. In any case, the p-values are sufficiently significant that they plausibly reflect real signal even though there is surely some amount of shared history among these groups since admixture (unknown but likely small).

S5.5 Additional groups with poor fits to the 2-way admixture model

The groups with the most positive Z-scores in **Table S 86** are also interesting, and suggests that there are many groups in India that have excess *Indus_Periphery_Pool*-related compared to Steppe pastoralist-related ancestry (consistent with excess ancestry from *Indus_Periphery_Pool*-related groups).

Taken together, the poor fits at both extremes of the Indian Cline imply that the Indian Cline does not represent a simple mix of two homogeneous ancestral populations, *ANI* and *ASI*. Instead, in the Middle to Late Bronze Age both of these groups were themselves part of metapopulations—relatively well represented by the *Steppe Cline* and the *Indus Periphery Cline*—that were not completely homogenized at the time they met and mixed. Most groups in India today can be represented as mixtures of average points along the *Steppe Cline* (we show below that the *ANI* fit along the *Steppe Cline*) and the *Indus Periphery Cline* (the *ASI*) but there are deviations from this simple model that contribute to the observed patterns.

S5.6 A single joint model for the 3 clines determines the ASI and the ANI

In the previous subsections, we modeled ancestry proportions along each of the 3 clines we first observed qualitatively in **Fig S 49**. We established that groups along the 3 clines were generally distinct from each other, with only a few sets of individuals that appear to lie on more than one. In our modeling approach we treated each of the 3 clines separately. We next sought to develop a single 3-source model that could fit the data from all 3 clines.

In the analyses described above, we showed that individuals from the *Indus Periphery Cline* fit as a primary source for people on the *Modern Indian Cline* as well as the *Steppe Cline,* using both models that pooled all these individuals together (as *Indus_Periphery_Pool* that we co-analyzed with *AHG*) or that considered two pools of *Indus Periphery Cline* individuals (*Indus_PeripheryA* and *Indus_PeripheryB*) to represent different average proportions of *AHG* ancestry along this cline. Next, to develop a model that works for all three clines—the *Indus Periphery Cline*, the *Steppe Cline* and the *Modern Indian Cline*—we used as a source a single *Indus Periphery Cline* individual (I8726) that had high coverage and the highest proportion of Iranian farmer-related ancestry. We call this individual *Indus_Periphery_West* (under the assumption that the Iranian farmer-related ancestry that is maximized in this individual derived from the western part of South Asia closest to Iran).

We modeled individuals on all three clines as mixtures of *AHG, Indus_Periphery_West,* and *Central_Steppe_MLBA.* **Figure 4** shows the results of this modeling analysis, and suggests that the *ANI* and the *ASI* can be viewed as points that lie on the intersections of 2 of the 3 clines: *ANI* as a point at the intersection of the *Steppe Cline* and the *Modern Indian Cline*, and *ASI* as a point at the interaction of the *Indus Periphery Cline* and the *Modern Indian Cline*. As we saw above, eight Dravidian speaking tribal populations from southern India fit as unadmixed descendants of the *ASI* under this definition, but we have no modern groups that fit on the *Steppe Cline* suggesting that descendants of it no longer exist in unmixed form.

S5.7 Simple statistics confirm Iranian farmer-related ancestry in the *ASI*

A remarkable implication of the arguments above is that the *ASI* had admixture from an Iranian farmer-related group (an estimated 47% ± 7% ancestry from *Indus_Periphery_Pool*, which in turn is 80% ± 3% *Ganj_Dareh_N*-related**)**. This suggests that rather than the *Modern Indian Cline* being a mixture of *AASI* (*AHG*-related) and a more West Eurasian-related group as was previously implied by work including by some of the authors of the present study (*11*) it was instead driven by mixtures of descendants of Steppe pastoralists and local South Asian farmers who already had West Eurasian-related (primarily Iranian-related) ancestry.

To obtain simpler and more direct evidence for Iranian farmer-related ancestry in the *ASI*, we computed statistics of the form f_4*(Lithuanian, Iranian; A, B)*. For the two groups on the left side of this statistic, we used present-day *Lithuanian* because they have high quality data and because we know they have extensive ancestry from ancient Steppe groups, and we used present-day *Iranians* because we know they have substantial ancestry related to ancient Iranian populations. For the two populations on the right side of this statistic, we picked *A* to be *AHG*, and *B* to be all 140 Indian Cline groups.

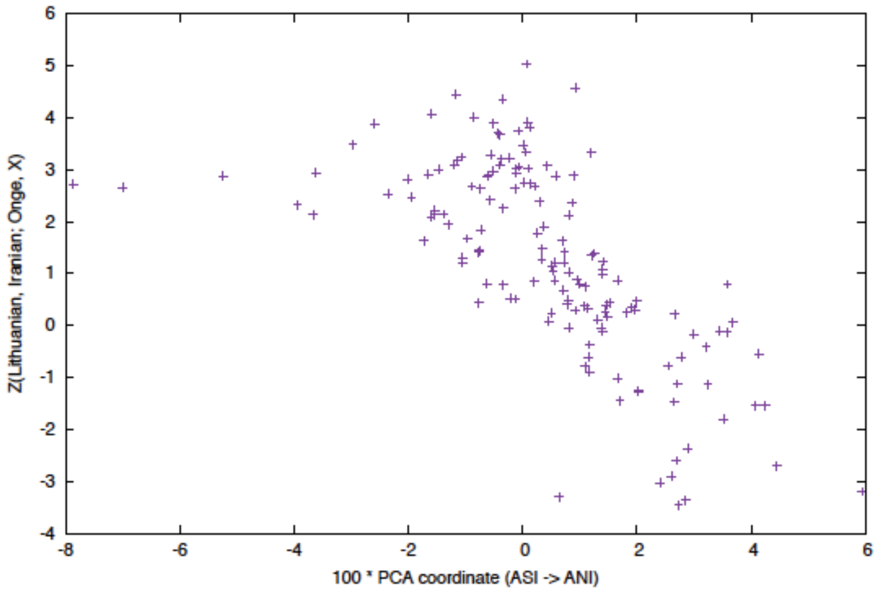

Fig S 54 Statistics of the form *f_4(Lithuanian, Iranian; AHG, X)* show a sign change comparing the high *ANI* to the high *ASI* end of the *Modern Indian Cline*. Under the null hypothesis that the *AHG* are a clade with the *ASI* (no West Eurasian-related admixture), we expect this statistic to have a value of *αf_4(Lithuanian, Iranian; AHG, ANI)* (where α is the proportion of *ANI* ancestry), and so this statistic should always have the same sign. However, we observe a sign change with increasing *ANI* ancestry. Thus, we can conclude that not all the West Eurasian-related ancestry along the *Modern Indian Cline* is coming from the *ANI*.

We observe a sign change from the high-*ANI* to the high-*ASI* extreme of the cline (as measured by the coordinate in the principal component analysis of **Fig S 54** that is proportional to *ANI* ancestry: for example, Z= 3.1 standard errors above zero for *Mala* (a group with a high proportion of *ASI* ancestry) and Z = -3.5 standard errors below zero for *Brahmin_Tiwari* (a group with a high proportion of *ANI* ancestry). The sign change is not expected under the hypothesis that the *ASI* had no West Eurasian-related ancestry, as if so, *f_4(Lithuanian, Iranian; AHG, ASI)* would be consistent with 0, and thus *f_4(Lithuanian, Iranian; AHG, X)*, where *X* is a *Modern Indian Cline* group, would be expected to be proportional to the *ANI* ancestry in *X* and would always have the same sign.

As a second way of verifying the finding of Iranian farmer-related ancestry in the *ASI*, we computed two statistics:

f_4(Mbuti, Kostenki14; AHG, X) **Equation 5-7**

f_4(Ancient_Steppe_Pool, Ancient_Iranian_Pool; AHG, X) **Equation 5-8**

Ancient_Steppe_Pool is a pool of *Afanasievo, Yamnaya_Samara,* and *Poltavka. Ancient_Iranian_Pool* is a pool of *Ganj_Dareh_N, BMAC* and *Tepe_Hissar_C.* **Table S 87** shows the values of these statistics for a variety of East Asian or Australasian groups *X*.

X	f_4(Mbuti, Kostenki14; AHG, X)		f_4(Ancient_Steppe_Pool, Ancient_Iranian_Pool; AHG, X)	
	f_4	Z	f_4	Z
Ami	-0.000034	-0.1	-0.000289	-2.0
Atayal	-0.000112	-0.3	-0.000355	-2.3
Dai	-0.000004	0.0	-0.000248	-1.7
Han	-0.000091	-0.3	-0.000383	-2.8
Nicobarese	-0.000405	1.0	-0.000296	-1.7

Table S 87 Empirical values of two f_4-statistics for groups of almost entirely East Asian or Australasian-related ancestry, which we used to establish a baseline expectation for a *Modern Indian Cline* group with entirely East Asian or Australasian-related ancestry if such a group existed (our analysis shows that it could not).

The values taken by **Equation 5-7** reveal that the ~35,000 year old early European hunter-gatherer population *Kostenki14* on the European hunter-gatherer lineage (*20, 221*) is consistent with sharing alleles at an equal rate with *AHG* and East Asians—that is, East Asians are consistent with being a clade related to *AHG* relative to ancient European hunter-gatherers. This means that the value of this f_4-statistic is expected to be proportional to West Eurasian-related ancestry in a *Modern Indian Cline* group *X*.

The expected value of **Equation 5-8** can be decomposed into a proportion α of West Eurasian-related ancestry and a proportion of (1 - α) of East Asian or Australasian ancestry. The empirical results in **Table S 87** suggest that the East Asian component is a fixed slightly negative value (-0.0003 averaging over the five groups). Thus, under the null hypothesis that the *ASI* have no West Eurasian-related ancestry, and defining K = f_4*(Ancient_Steppe_Pool, Ancient_Iranian_Pool; AHG, ANI)*, the expected value of **Equation 5-8** can be computed as:

$$E[f_4(\text{Ancient\_Steppe\_Pool, Ancient\_Iranian\_Pool; AHG, X})] = \alpha - (1-\alpha)0.0003+K \quad \textbf{Equation 5-9}$$

In **Fig S 55** we show a scatterplot of these two statistics against each other for the 140 groups in the Indian Cline. We also show a fitted least squares regression line. The data are noisy, but it appears impossible that the noise-free value where the cline crosses the Y-axis (corresponding to no West Eurasian-related ancestry) could be negative. Indeed, a block jackknife (deleting each autosomal chromosome in turn) gives a Z-score of a significantly positive Y-intercept for the fitted regression line (Z=6.2), providing strong evidence against the null hypothesis of entirely East Asian or Australasian-related ancestry in the *ASI*.

Fig S 55 also reveals a significantly negative slope (|Z| = 7.7), implying that the *ANI* share more drift with the Iranian-related pool of individuals than with the Steppe pastoralist-related pool of individuals, a result that suggests a component of Iranian-related ancestry in the *ANI* as well as in the *ASI*, consistent with our *qpAdm* results.

We finally observe that the *SPGT* individuals and the *Central_Steppe_MLBA* individuals are off-cline. The former have excess Iranian farmer-relatedness, while the latter have excess Steppe pastoralist-relatedness (compared to the proportion of West Eurasian-related ancestry). This is consistent with our modeling showing that the *Steppe Cline* is not included within the *Modern Indian Cline.* Instead, the *Modern Indian Cline* is a mixture between a point on the *Indus Periphery Cline* (the *ASI)* and a ghost population that once existed on the *Steppe Cline,* which we haven't directly sampled but which we hypothesize existed in the Late Bronze Age and Iron Age in northwestern South Asia (but not in the Swat Valley which is the only place that we have extensively sampled). We predict that individuals from this population will be found in future ancient DNA studies.

Fig S 55 The statistic *f_4(Ancient_Steppe_Pool, Ancient_Iranian_Pool; AHG, X)* is expected to be negative for a population of entirely *ASI* ancestry under the null hypothesis that the *ASI* had no West Eurasian-related ancestry. However, we observe that it is significantly positive.

S5.8 No significant evidence for *AHG*-related gene flow into Iranian farmers

One of our important lines of evidence for admixture into South Asia is based on statistics of the form *f_4(Mbuti, Iranian farmer-related; AHG, High ASI)* where "Iranian farmer-related" is represented by an ancient Iranian farmer-related group, and *"High ASI"* is a population along the *Modern Indian Cline*. For instance, we computed all possible statistics of the form *D(Mbuti, Iranian farmer-related; AHG, Palliyar)*. The largest Z-scores come from ancient groups with Iranian farmer-related ancestry, four present-day groups from Iran, *Georgians* from the Caucasus who have strong genetic similarity to present-day groups from Iran, and *Druze*, which may reflect ancient gene flow between the Levant and Iran (*9*) (**Table S 88**). The fact that the single highest score is seen for *Zoroastrian* is consistent with a previous finding of a particularly close genetic similarity of this group to ancestry in present-day South Asians (*31*).

Iranian farmer-related	**Z-score**
Zoroastrian	11.081
Iranian_Bandari	10.879
Tepe_Hissar_C	10.301
Iranian	10.176
Fars	9.865
Druze	9.771
Seh_Gabi_C	9.614
Ganj_Dareh_N	9.611
Georgian	9.527

Table S 88 Most significant statistics of the form *D(Mbuti, X; AHG, Palliyar)*

However, f_4-statistics do not determine direction of flow. Could gene flow from South Asia into Iran be driving these signals? We show here that the evidence for Iranian farmer-related ancestry across the entire *Modern Indian Cline* is unambiguous, whereas there is no clear evidence of *AHG*-related ancestry—ubiquitous in present-day South Asians—in ancient Iranians, except for individuals from Bronze Age Iran and Turan in the BMAC and Shahr-i-Sokhta (both in the main cluster of individuals as well as at significantly elevated levels in the outliers) as described above.

We begin by analyzing statistics of the form f_3*(Test; Source1, Source2)*, which if negative provide unambiguous evidence of mixture in a *Test* population of ancestry related (perhaps distantly) to the two proposed source populations. We computed this statistic for groups on both the high *ANI* and high *ASI* extremes of the *Modern Indian Cline* that do not have much group-specific drift, using *Tepe_Hissar_C* as the Iranian farmer-related source and *Juang* as the *AASI*-related source. (The *Juang* are an Austroasiatic-speaking group in India with less evidence of West Eurasian-related ancestry than any group on the Indian Cline, which we believe is the reason the absolute value of the f_3-statistic is maximized when using the *Juang* as the *AASI*-related source, rather than the *AHG*, who are only distantly related to the *AASI*.) The statistic is negative all along the cline as shown in **Table S 89**.

Population	**Z-score**
Brahmin_Tiwari	-28.9
Kashmiri_Pandit	-13.1
Mala	-16.7

Table S 89 f_3*(X; Juang, Tepe_Hissar_C)* is negative for populations representing the extreme ends of the *Modern Indian Cline*

We next investigated admixture using *qpAdm*. As shown in **Table S 5** and **Supplementary Materials S4**, and also discussed earlier, this method provides clear evidence of Iranian farmer-related admixture (via an *Indus Periphery Cline* group) into *Palliyar*, a southern Indian group at the extreme-*ASI* end of the *Modern Indian Cline.*

We carried out *qpAdm* analysis using as a target the Iranian farmer-related group *Tepe_Hissar_C*. We fit a model for *Tepe_Hissar_C* in which the *Left* groups were (*Tepe_Hissar_C, Ganj_Dareh_N*, *Afanasievo, Anatolia_N*, *AHG*) and the *Right* groups were (*Mbuti, Chukchi, Karitiana, Ust_Ishim_Published.DG, Kostenki14, MA1_HG.SG, Ganj_Dareh_N* and *Dai*). The fit is good (P=0.50), with a coefficient for *AHG* of 0.6% ± 1.9%, which is not significantly different from zero, providing no evidence of *AHG*-related ancestry in *Tepe_Hissar_C* from eastern Iran. The fit of the model is also good deleting *AHG* from the “Left” set (P=0.65).

We also carried out this analysis using as a target the ancient Iranian Chalcolithic group *Seh_Gabi_C* from western Iran. This analysis produced similar results of no evidence for *AHG*-related ancestry, with a p-value of 0.29 and a coefficient for *AHG*-related ancestry that is not significantly different from zero at 2.0% ± 2.1%. The absence of evidence for South Asian-related ancestry in Iranian Chalcolithic and Bronze Age groups—which is consistent with the exhaustive *qpAdm* analyses of **Supplementary Materials S4**—suggests that gene flow from *AASI* into ancient Iranians, while plausible geographically, is not supported based on the genetic data, and indeed we can put an upper bound on the proportion of *AASI* ancestry in ancient Iranians. Given the standard errors on the *AHG*-related ancestry, it is implausible that *AASI* admixture in *Tepe_Hissar_C* was more than 6%. Note that these statements are limited to the region of present-day Iran and points further to the West; as discussed above, further to the East in Bronze Age Turan there was a low level of *AHG*-related admixture reflecting gene flow from South Asia into this region between the Chalcolithic and Bronze Age.

Jaydeepsinh Rathod (personal communication) has pointed out to us two references—Figure 2B of (*222*) and Figure 1C of (*223*)—both reporting ADMIXTURE automated clustering plots (*37*) suggestive of South Asian-related ancestry in present-day Iranians. This is interesting and plausible, but even if the admixture is real in light of the arguments above, there is no evidence that such admixture affected population groups much further West.

While it maybe true that there is a lack of *AHG*-related ancestry in Iran and Turan, this does not mean that there was a spread of Iranian farmers from the Zagros into South Asia. Another interpretation of the data could be that a broad landscape of hunter-gatherer landscape existed that ranged from the Zagros mountains through the northwest portions of South Asia, and the admixture we are observing between the Iranian farmer-related ancestry and the *AHG*-related ancestry is a mixture of two groups of populations that occupied the northwest and the southeast of South Asia. Ancient DNA data from hunter-gatherers and early farmers from South Asia and eastern Iran could help resolve these issues.

S5.9 Admixture graph relating South Asians to non-South Asians

We show in **Fig S 56** an admixture graph (a phylogenetic tree incorporating admixture events (*33*)) constructed based on allele sharing patterns among 13 groups (ancient and modern) for which we have genome-wide data, including traditionally high and low caste populations, and the Austroasiatic-speaking *Juang*.

The overall fit obtained using the *qpGraph* software (*33*) is good, with the worst *f*-statistic (out of 2871 we computed) being f_4*(Papuan, Dai; BMAC, Rajput)*, which is Z = 2.9 standard errors different between model and expectation. We highlight two historically interesting features of our fitted model.

(1) The *AASI* are tens of thousands of years diverged from other East Asians and Australasians including *AHG*. It is natural to hypothesize that the *AASI* represent a lineage derived from indigenous southern Asian foragers, which diverged from other Eurasian groups very anciently.

(2) In our fitted admixture graph, *AASI*, *AHG*, and *Papuan_anc* (a hypothesized ancestral population to modern *Papuans*, prior to Denisovan admixture) are a clade with respect to indigenous Chinese groups and *Nicobarese* (a group representing the East Eurasian ancestry that plausibly dispersed with the Austroasiatic language expansion). The split between *AASI*, *AHG*, *Papuan_anc* is modeled as nearly a trifurcation. It seems probable based on geographic considerations that the split between *AASI* and *Papuan_anc* occurred prior to modern humans reaching Sahul (the ancient continent uniting Australia and New Guinea). Radiocarbon dating

shows this is unlikely to be much more recently than 47,000 years before present (*224*).

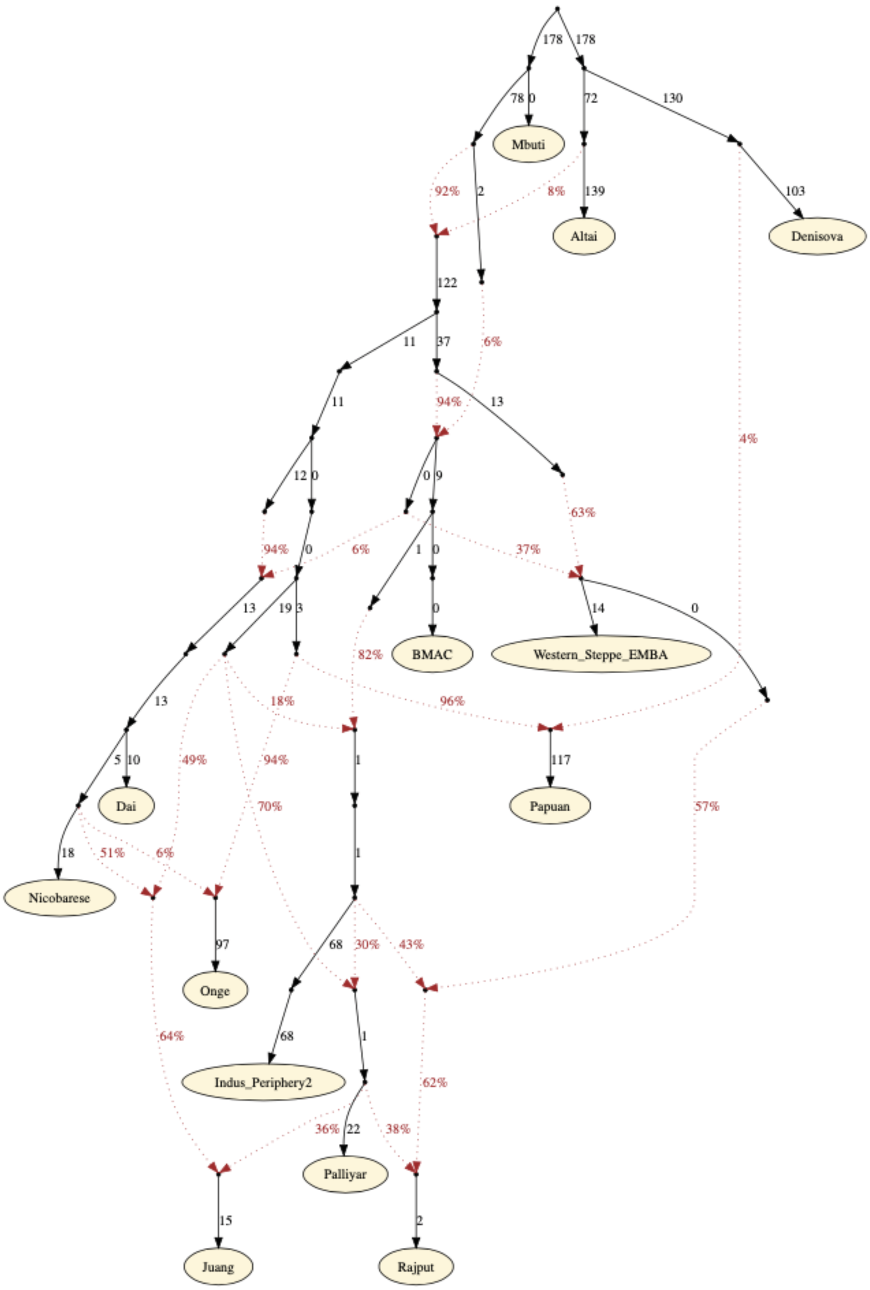

Fig S 56 A fitting admixture graph for South Asia, showing the relationship of the AASI with other East Eurasian populations. This is a version of Fig. 5 in the main text

S5.10 High proportions of *AASI* ancestry in present-day Austroasiatic speakers

The *Juang* are an Austroasiatic speaking group in India whose PCA coordinates (reported in **Table S 5**) suggest have an especially low proportion of West Eurasian-related ancestry. We were unable to model *Juang* as a mixture of *ASI* and a source that was a clade with *Nicobarese* (isolated Austroasiatic speakers from the Nicobar Islands). However *qpGraph* obtains an excellent fit by adding a substantial component of *AASI* ancestry to *Juang* (**Fig S 56**). Thus, the *Juang* have too much *AASI*-related ancestry relative to ancient Iranian farmers to be a simple two-way mixture of a *Nicobarese*-related population and *ASI*. These results suggest that Austroasiatic-speaking groups were in peninsular India at a time when there were still populations that had little if any Iranian farmer-related admixture.

S5.11 Sex bias in the spread of Steppe ancestry into South Asia

Taking advantage of our large sample sizes, we tested if the admixture of *Central_Steppe_MLBA* ancestry with *Indus_Periphery_Pool* ancestry that formed the Swat Valley Late Bronze-Iron Age (*SPGT*) group was sex-biased, and then tested whether the admixture of *Central_Steppe_MLBA* into the ancestors of present-day South Asians was sex-biased.

We began by using *qpAdm* to estimate the proportion of *Central_Steppe_MLBA* ancestry in the *SPGT* both on the autosomes and on chromosome X. Under the assumption that the admixture was driven by equal proportions of *Central_Steppe_MLBA* individuals from each sex, the proportions on the autosomes (equally reflecting male and female demographic history) should be consistent with those on chromosome X (primarily reflecting female history). **Table S 90** shows no significant difference between these two compartments of the genome although standard errors of the X chromosome estimates are so large that the failure to detect a significant difference does not exclude the possibility of substantial bias.

	Indus_Periphery_Pool	***Central_Steppe_MLBA***	**Standard Error**
Autosomes	0.806	0.194	0.009
X	0.850	0.150	0.051

Table S 90 *SPGT* estimates of *Indus_Periphery_Pool* and *Central_Steppe_MLBA* ancestries on chromosome X and on the autosomes along with the standard error.

We also computed estimates of Steppe pastoralist-related ancestry on the Y chromosome, taking advantage of the extremely high (100%) frequency of the R1a haplogroup in the *Central_Steppe_MLBA* individuals and its complete absence in BA Turan. This nearly fixed difference in Y chromosome haplogroup frequencies allows us to treat the proportion of R1a chromosomes as an estimate of the proportion of entirely male-lineage ancestry in the *SPGT* from the *Central_Steppe_MLBA*. We observe only 2 R1a Y chromosomes among the 44 *SPGT* males in whom we could confidently determine a Y chromosome, corresponding to a ninety-five percent binomial confidence interval of 0.4-16% for the Y chromosome ancestry proportions derived from *Central_Steppe_MLBA*. In comparison, the ninety-five percent confidence interval for the estimate on the autosomes is 18-21%. The ninety-five percent confidence intervals are larger on the autosomes than on chromosome Y and do not overlap, thereby showing that while the X-chromosome estimates are too noisy to be useful here, the admixture into the *SPGT* was definitively female-biased.

We also examined if we could detect sex bias on the *Modern Indian Cline,* and here detected the reverse pattern. We began by pooling 4 South Asian populations (*GIH, ITU, STU* and *PJL*) from the 1000 Genomes Project for this analysis (*78*), and computed ancestry proportions on chromosome X and the autosomes as before. The results do not show any significant difference, although the data are also consistent with a substantial difference given the large standard errors on chromosome X (**Table S 91**).

	AHG	***Indus_Periphery_Pool***	***Central_Steppe_MLBA***	***Central_Steppe_MLBA* standard error**
Autosomes	0.293	0.591	0.116	0.011
X	0.34	0.523	0.137	0.052

Table S 91 Proportions of *AHG, Indus_Periphery_Pool* and *Central_Steppe_MLBA* ancestries on chromosome X and on the autosomes along with the standard error of the estimates for *Central_Steppe_MLBA* for the 1000 Genomes Project South Asians.

Using previously reported calls on 1000 Genomes Project Y chromosomes (*225*), we observe that 62 out of the 221 South Asian males have an R1a Y chromosome corresponding to a ninety-five percent binomial confidence interval of 22-34% for Steppe MLBA ancestry on the entirely male line, which is significantly higher than the ninety-five percent confidence interval of 9-14% on the autosomes in the same set of individuals. These results shows the process of admixture of *Central_Steppe_MLBA* into the ancestors of the *ANI* was male-biased, and reveal that the directionality of sex bias was opposite to the pattern observed for the contribution of *Central_Steppe_MLBA* to *SPGT.*

S6 Dating of population mixture events in present-day and ancient individuals

In order to understand the timescale of population mixture events in South Asia, we use ancestry covariance-based statistics to date the admixtures. To this end, we use two main methods: ALDER (*39*) for dating admixture in present-day individuals, and *DATES* (a method we introduce here) for ancient individuals. Below we describe the details of the methods and analyses.

S6.1 Admixture timing estimates on the *Modern Indian Cline*

To date the West Eurasian admixture in present-day groups on the *Modern Indian Cline*, we used ALDER (*39*), which measures the decay of admixture linkage disequilibrium to infer the time elapsed since mixture. ALDER estimates a weighted covariance statistic (*a(d)*) across pairs of markers (*x, y*) as follows:

$$\boldsymbol{a(d)} = \frac{\sum_{\boldsymbol{S(d)}} \boldsymbol{z(x,y)w(x,y)}}{|\boldsymbol{S(d)}|} \qquad \textbf{Equation 6-1}$$

Here, *S(d)* contains markers that are at a genetic distance less than *d* cM, *z(x,y)* is the covariance between the genotypes at markers *x* and *y* in the target populations, and *w(x,y)* is a weight function which represents the allele frequency differences between the ancestral populations. In practice, *w(x,y)* is estimated by using the allele frequency differences in two reference populations that are considered to be close surrogates of the ancestral mixing populations (*39*). For South Asians, good reference data for the ancestral mixing groups is not available. Instead, we have access to admixed groups with variable ancestry proportions, which can be used to extrapolate the allele frequency differences in the ancestral populations by using PCA-based SNP loadings computed using West Eurasians and various *Modern Indian Cline* groups (excluding the target group) to infer *w(x,y).* This idea was first suggested in ref. (*12*), which showed that the use of PCA loadings can be used to reliably date admixture in the absence of close surrogates for the reference populations, without biasing the dates of admixture (*12*).

Using this approach, we infer the dates of admixture for groups on the *Modern Indian Cline* that have a sample size of at least five (the requirement of a minimum sample size is important for measuring LD with precision). Results are shown in **Table S 5**. For the eight groups consistent with having entirely *ASI* ancestry (*Adiyan, Ulladan, Palliyar, Malayan, Yanidi, Irula, Gugavellalar, Pulliyar*) and having extremely low or no Steppe pastoralist-related ancestry as shown in **Supplementary Materials S5**, the inferred date would reflect the timing of the mixture between the *AASI* (Andamanese Hunter-Gatherer (*AHG*)-related) and *Indus Periphery Cline* groups (bearing Iranian farmer-related admixture). In **Fig S 57**, we show the decay of admixture in one of the populations, *Palliyar*, that is consistent with direct descent from the *ASI*. We obtain a date of mixture of between the *AASI* and *Indus Periphery Cline* of 107 ± 11 generations, corresponding to a 95% confidence interval of 1700-400 BCE assuming 28 years per generation (*53*). This date is similar to (but more precise) than the date of 116 ± 29 generations obtained by running ALDER with two reference populations (Iranian farmer- (*Ganj_Dareh_N*) and an Andamanese Hunter-Gatherer *(AHG*)-related group).

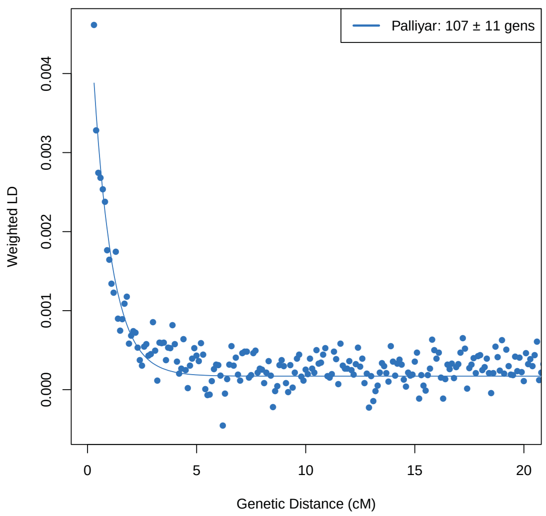

Fig S 57 Decay of admixture linkage disequilibrium in *Palliyar*. We ran ALDER with *Palliyar* as the target population and using weights from PCA-based SNP loadings computed using 141 present-day populations (two West Eurasian groups (Iranian and Georgians) and 139 other *Modern Indian Cline* groups). We used genetic distances from (*40*)**. Following the recommendation by ALDER, we start the fit at genetic distance (d)>0.3cM.**

However, for groups with both Iranian and Steppe ancestry, our dates reflect an average of the two mixture events, complicating interpretation of the inferred dates. We note that here and in what follows, we model admixture as an instantaneous process (a single episode of gene flow), even though in fact it could have occurred at multiple time points or gradually over a long period of time. Thus, one way to interpret the dates reported below is that they reflect the timing of major mixture of the ancestral sources in the history of a tested population and represent an upper bound on the time of the last gene flow as well as lower bound on the time of the earliest gene flow.

S6.2 Ancient Individuals

A limitation of admixture LD-based methods like ALDER (*39*) and *Rolloff* (*33*, *226*) is that they require high quality data from individuals from the admixed population, whereas in practice for ancient DNA studies we often do not have this. To overcome this challenge, we introduce a method for dating admixture that we call *DATES (Distribution of Ancestry Tracts of Evolutionary Signals)* that leverages ancestry covariance patterns that can be measured in a single individual (instead of admixture LD that requires multiple individuals). This method extends the idea introduced in ref. (*53*) to be applicable to dating admixture events in contemporary human populations, without using an ascertainment scheme (*53*). Briefly the method works as follows.

S6.2.1 *DATES*: Theory

Assume we have a single admixed individual *s* who has ancestry from populations, P_1 and P_2, with ancestry proportion $\propto$ and (1-$\propto$) respectively. Our central idea is to model the genotypes of *s* as a linear mix of allele frequencies of population P_1 and P_2, by solving a simple regression. At SNP *i*, let the genotype of *s* be *g(i)* and allele frequency in P_1 and P_2 be $p_1(i)$ and $p_2(i)$. We can then fit the mixing fraction $\propto$ from population, P_1 by minimizing

$$\boldsymbol{S = \sum_i \left(g(i) - \left(\propto p_1(i) + (1-\propto) p_2(i)\right)\right)^2} \qquad \textbf{Equation 6-2}$$

and set the residual *r(i)* as:

$$\boldsymbol{r(i) = g(i) - (\propto p_1(i) + (1-\propto) p_2(i))} \qquad \textbf{Equation 6-3}$$

Under a model of admixture, $x(i) = r(i)(p_1(i) - p_2(i))$ will correlate with the proportion of P_1 mixture at SNP *i*. We can compute a covariance coefficient of $x(i)$ and $x(j)$, where *(i, j)* are SNPs separated by genetic distance *d*, and just as in *Rolloff* and ALDER, we expect this to decay exponentially with genetic distance, with a decay rate depending on the time since admixture. When we have multiple individuals, we simply compute ancestry covariance separately for each individual and then combine the data, which is in principle similar to running *Rolloff* or ALDER. The value of this setup is that we are able to leverage all available data, rather than limiting to the subset of markers with almost complete data (as required by ALDER(*39*)). We note that in principle the ancestry covariance signal can be confounded by drift in the target group post-admixture. However, in practice this is often not a problem. In order to make *DATES* computationally tractable, we implemented the fast Fourier transform (FFT) for computing ancestry covariance as described in ALDER. The FFT reduces the compute time from hours to seconds.

S6.2.2 *DATES:* Software / Web Resources

The software implementing *DATES* is available for download from the following URL. Details and README are provided on GitHub (https://github.com/priyamoorjani/DATES) as well as on Zenodo (https://zenodo.org/record/3263997#.XRnebJNKj6A, DOI: 10.5281/zenodo.3263997).

S6.2.3 Testing the performance of *DATES*

Earlier studies have dated the admixture of East Asian and West Eurasian ancestry in Uygurs and have inferred the average date across ten *Uygur* individuals as 23 ± 3 generations ago using ROLLOFF and 22 ± 1 generations ago using GLOBETROTTER (*227*). To infer the timing of mixture for each individual, we ran DATES for *Uyghurs*, considering each *Uygur* individual in the Human Genome Diversity Panel as the target and using French and Han as the surrogates for the ancestral West Eurasian and East Asian population respectively (**Fig S 58**). The combined analysis with all individuals provides consistent results (22 ± 2 generations ago) with the previous studies (*33*, *227*) but unlike previous studies we have single sample resolution (**Fig S 58**). We note that some differences are expected given that we use a different dataset, genetic map, parameter settings (e.g. minimum and maximum genetic distance, mesh-size for FFT computation, etc.) and reference ancestral populations (French vs. Iranians and Han vs. Mongolians) in *DATES* compared to the earlier studies.

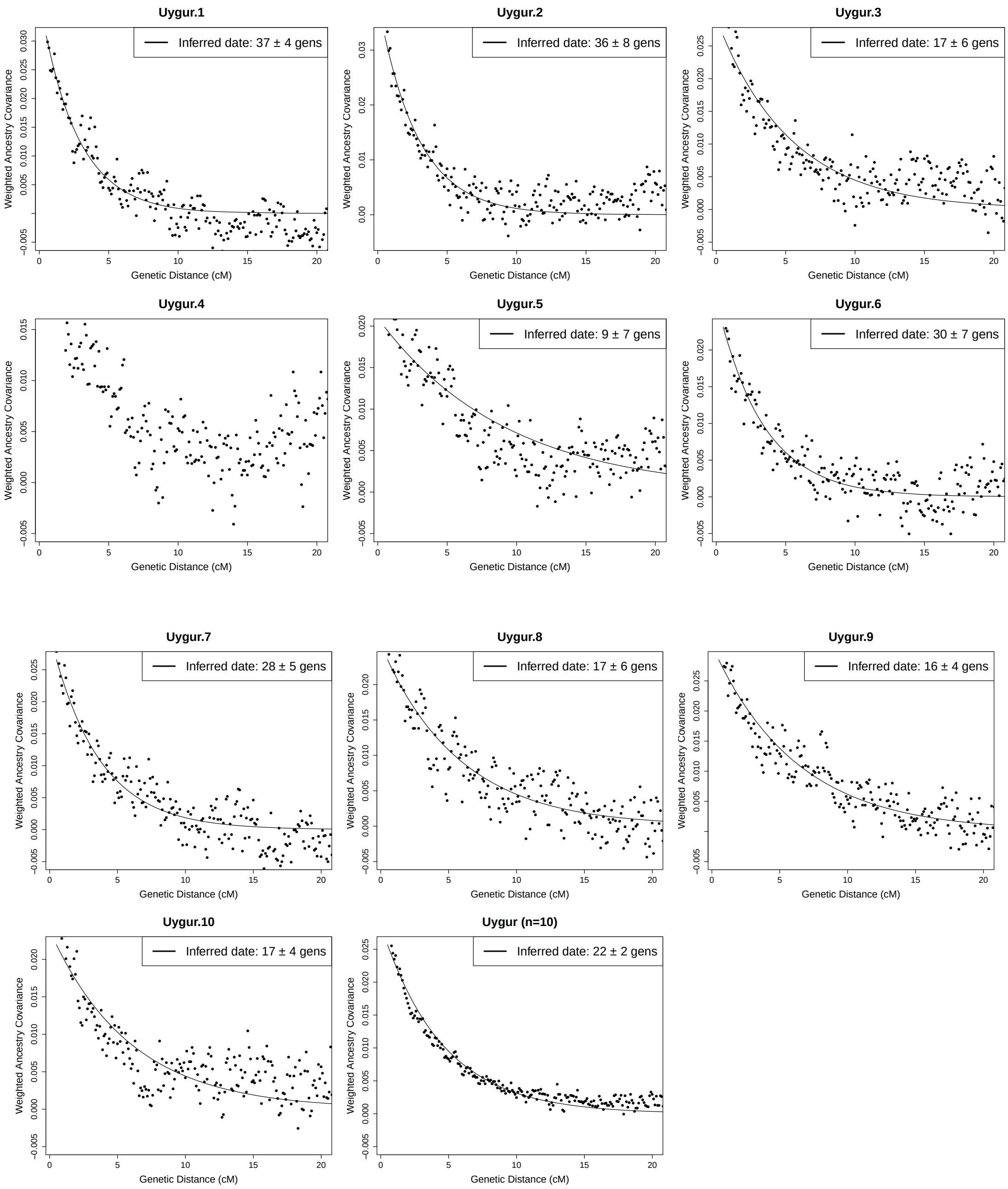

Fig S 58 Ancestry covariance in *Uyghurs*. We ran *DATES* to infer the timing of admixture in *Uyghurs*, considering each *Uygur* in HGDP as the target and using French and Han as the surrogates for the ancestral populations. *DATES* does not infer the start distance automatically and hence we use d>0.5cM, which is conservative and minimizes confounding due to background LD. We estimate a standard error by performing a weighted block jackknife where we remove one chromosome in each run. Note the signal in one individual (Uygur.4) is very noisy and so we do not estimate a date for this individual but show only the raw output. The last plot displays the result of the combined run including 10 individuals.

S6.2.4 Application of *DATES*

We applied DATES to infer the timing of admixture events impacting ancient South Asians. To this end, we applied the analysis to the *Indus Periphery Cline* and the early Late Bronze-Iron Age Swat Valley individuals *(Swat Protohistoric Grave Type - SPGT).* To maximize power, we used our in-solution enrichment data with 1.2 million SNPs and where appropriate combine individuals with the same ancestry, regardless of the age or 14C dates. Unlike our approach for present-day groups, we do not use the PCA-based SNP loadings here, as we do not have all the representative groups in the larger capture array dataset. We note that we currently do not model multiple admixture events in *DATES* and this in principle leads to confounding of the admixture timing. To mitigate the effects of the confounding, we focus on the most recent admixture event only, for which we have the highest power and minimal confounding.

S6.2.4.1 *Indus_Periphery_Pool*

We applied *DATES* to infer the timing of the Iranian farmer-related ancestry and *AASI* ancestry in the 11 *Indus_Periphery_Pool* individuals (3 outliers from the site of Gonur pooled with 8 from the eastern Iranian site of Shahr-i-Sokhta). For this analysis, we use the following two reference populations: *group1* containing individuals of *Iranian farmer-related ancestry (Aigyrzhal_BA, Sarazm_EN, Geoksyur_EN, Parkhai_Anau_EN), and group2* containing surrogate populations harboring *AASI* ancestry (South Asian individuals from the 1000 Genomes Project (Phase 3) including Sri Lankan Tamil from the UK *(STU.SG)* and Indian Telugu from the UK *(ITU.SG*), as well as *BIR.SG)*. We inferred that the *AASI-related* admixture occurred 71 ± 15 generations before the average sampling time of our *Indus_Periphery_Pool* individuals, corresponding to a 95% confidence interval of 5377 - 3697 BCE, assuming an average sampling time of 2549 BCE (range: 3175-2056 BCE) (**Fig S 59**). As the sampling time of our individuals from Gonur and Shahr-i-Sokhta sites differ substantially (by ~1000 years), we also ran separate analysis for the 3 individuals from the Gonur site and 8 individuals from Shahr-i-Sokhta sites. The dates obtained are 95 ± 41 generations and 52 ± 6 generations ago respectively, obtaining dates of which 7127 - 2535

BCE and 4483 - 3811 BCE respectively. The results are noisier in Gonur but consistent with the combined dates reported above.

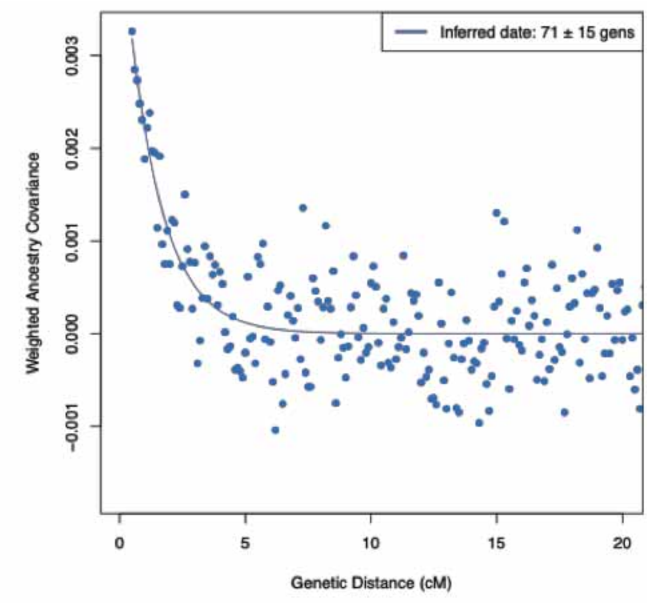

Fig S 59 Ancestry covariance in *Indus_Periphery_Pool*. We ran DATES to infer the admixture date in 11 *Indus_Periphery_Pool* individuals, using two ancestral groups. *group1* contains individuals of Iranian farmer-related ancestry (*Aigyrzhal_BA*, *Sarazm_EN*, *Geoksyur_EN*, *Parkhai_Anau_EN*), and *group2* contains populations harboring *AASI* ancestry (South Asian from the 1000 Genomes Project (Phase 3) including Sri Lankan Tamil from the UK (*STU.SG*) and Indian Telugu from the UK (*ITU.SG*), as well as *BIR.SG*. We start the fit at d>0.5cM, and estimate a standard error by performing a weighted block jackknife removing one chromosome in each run.

S6.2.4.2 South Asia: Early Late Bronze-Iron Age Swat Valley individuals

Our *qpAdm* and admixture graph analysis suggests that the Late Bronze-Iron Age South Asian individuals (Swat Protohistoric Grave Type - *SPGT*) harbor *AASI*, Iranian farmer-related and Steppe pastoralist-related ancestry. To characterize the most recent event of mixing of the *AASI* and Steppe pastoralist-related ancestry, we ran DATES using surrogates of AASI ancestry *AASI* ancestry (South Asian individuals from the 1000 Genomes Project (Phase 3) including Sri Lankan Tamil from the UK (*STU.SG*), Indian Telugu from the UK (*ITU.SG*) and *BIR.SG*) as one reference and *Central_Steppe_MLBA* as the other reference. Using this approach, we infer that the *Steppe Pastoralist-related* admixture in *SPGT* occurred 26 ± 3 generations before the average sampling time of our *SPGT* individuals (919 BCE, range: 1263 - 808 BCE), corresponding to a 95% confidence interval of 1815 - 1479 BCE (**Fig S 60**). This is similar to, but more precise than, the dates we obtain for *ANI* ancestry in *Kalash* in our study (**Table S 5)** or reported in earlier work (*56*).

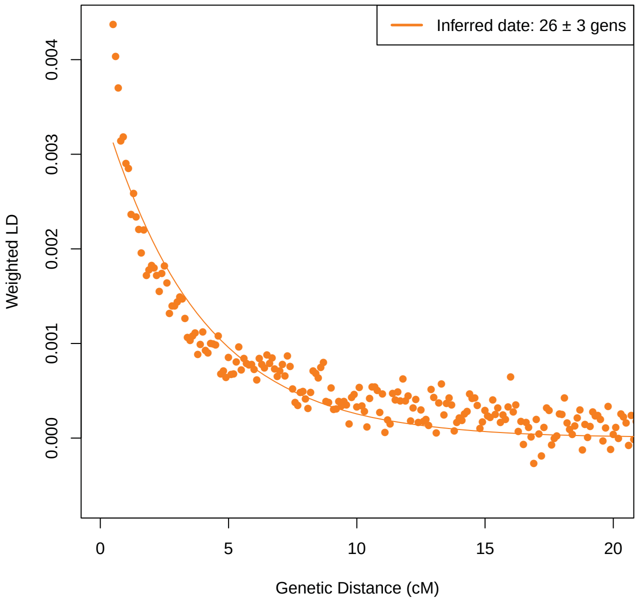

Fig S 60 Dates of admixture in *SPGT*. We ran *DATES* to infer the timing of mixtures of *AASI* and *Steppe pastoralist-related* ancestry in the 86 *SPGT* individuals using surrogate populations harboring *AASI* ancestry (South Asian individuals from the 1000 Genomes Project (Phase 3) including Sri Lankan Tamil from the UK (*STU.SG*), Indian Telugu from the UK (*ITU.SG*) and *BIR*) as one reference and individuals of *Central_Steppe_MLBA* population as the other reference. To minimize confounding with background LD, we start the fit at d>0.5cM. We estimate a standard error by performing a weighted block jackknife removing one chromosome in each run.

S7 Uniparental marker analysis

S7.1 Y-chromosome haplogroup determination

Characteristic ancient DNA damage, low coverage, and contamination make accurate assignment of Y chromosome haplogroups challenging in ancient DNA data. To address this problem, we modified an existing tool for calling Y chromosome haplogroups from sequence data, yHaplo (*225*) (https://github.com/23andMe/yhaplo), to work specifically with ancient DNA.

The original implementation of yHaplo carries out a breadth-first search as it traverses the nodes of the Y chromosome phylogeny and chooses a set of likely search paths based on the number of derived and ancestral alleles observed in the descending branches. The data in our study are often under 1-fold coverage and thus will be missing a significant portion of informative sites. In particular, mutations that define the next possible traversal path will often be missing. To adjust for this, we traverse all possible paths through the tree and compute scores of mismatches against the various haplogroup references. We weight C-to-T mismatches at ⅓ that of other mutational types, as we assume that these are more likely to be errors due to damage in ancient DNA, in which case they are misleading about the true haplogroup. Where the individual is of extremely poor coverage and highly degraded, this might allow paths with erroneous data to cause mis-assignment of the haplogroup. However, these damage patterns are not expected to occur in a systematically biased fashion with respect to the topology of the tree. To ensure that the calls are confident and conservative with respect to the tree, we check to see if the path taken to assign a haplogroup has bypassed an informative position without any derived alleles, and ensure that each haplogroup assignment is determined by two derived alleles on different tiers of the topology.

In order to assess the performance of our calling procedure, we examined a set of gold standard haplogroup calls that were produced using both automated and manual curation on high coverage whole genome sequencing data from Phase 3 of the 1000 Genomes Project (*228*). On the same set of individuals, we simulated data that would emulate the characteristics of ancient DNA. We first produced random allele calls on the 32,670 positions on the Y chromosome that were included as part of the ~1.2 million SNP enrichment reagent. We simulated low coverage by randomly setting 0 to 80% of the genotypes to missing, in 10% increments. To simulate the effect of ancient DNA degradation and damage, we introduced C-to-T errors on the remaining non-missing genotypes for each test case, randomly at different percentages from 0-10%. Finally, we simulated contamination by switching a random set of genotype calls from 0-10% with those from a random European ancestry individual from the CEU cohort from the 1000 Genomes Project. As we have simulated error rates and coverage reductions on ~1.2 million SNP enrichment positions that were ascertained specifically for polymorphism on the Y chromosome phylogeny, this makes the process of recovering the correct haplogroup considerably more challenging than if we were to simulate damage and missingness on random portions of the Y chromosome.

For each level of coverage reduction, damage and contamination on the simulated dataset, we produced haplogroup assignment calls and compared them to the actual haplogroup assigned by the gold standard calling procedure. With the extent of simulated data reduction, it was not theoretically possible to obtain resolution on the haplogroups to the finest granularity, particularly in cases when more than 50% of the data was missing, and thus we compared the haplogroups at the level of 3 alphanumeric significant figures (R1b, C2a, etc.). We considered this to be a level of haplogroup resolution that was informative about large scale population transformation such as that involved in the arrival of Steppe Pastoralists from the Steppe into Europe during the Bronze Age (*8*, *18*)

We report the error rates of haplogroup assignment under the various perturbations in **Fig S 61**. The conservative UDG treatment and trimming at the end of reads as part of our sequencing protocol is expected to reduce the effect of C-to-T errors on our genotype calls to under 1%. Under these settings, the simulations suggest that our error rate in the assignment of haplogroups will be 1% or less. In order to further verify this, we used our calling pipeline to assign haplogroups to an independent ancient DNA dataset from (*18*) that was manually curated and found that the calls differed in only 2 out of 99 individuals.

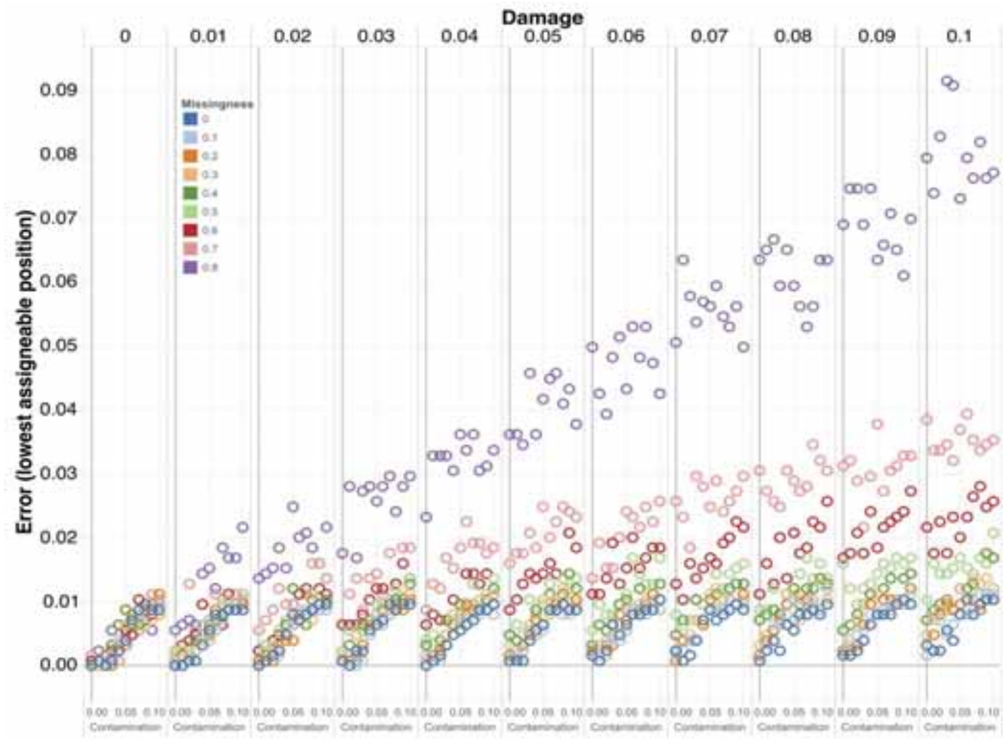

Fig S 61 Performance of Y chromosome caller on simulated Y-chromosome data at different levels of contamination, damage and missingness.

After evaluating the performance of our pipeline on simulated and real data from datasets for which the haplogroup assignments were manually curated, we determined Y-chromosome haplogroups using the International Society of Genetic Genealogy (ISOGG) database version 11.110. We report the Y-chromosome calls for all 280 male individuals for whom we generated data in **Table S 1**. We note however that this automatic calling pipeline may be unreliable for individuals for which we have data on fewer than 100,000 SNPs overall, as the haplogroup assignment is made from few informative positions, and we urge caution in the interpretation of the output in such settings.

S7.2 Interpreting the Y-chromosome data in light of our autosomal admixture models

In **Supplementary Materials S4** and **S5**, we examined evidence for population movement and admixture in South and Central Asia by mainly examining the autosomal DNA markers. Here, we briefly describe our analysis of the results of our Y chromosomal calling procedure in light of the evidence about admixture that we obtained from the autosomal analysis.

We begin by reporting the distribution of haplogroups by region and time period amongst the male individuals from Central and South Asia that have been published previously as well as generated for this study (**Table S 92**).

	Forest / Steppe					Iran / Turan				South Asia			Grand
Y haplogroup	N	EMBA	MLBA	LBA	IA	M / N	C / EN	BA	IA	BA	IA	H	Total
E1b					1			1			6		8
G1							1						1
G2a					1		1	1			1		4
H1a										2	4		6
I2a					1						2		3
J							6				1		7
J1							2	1				3	6
J2							1				1		2
J2a			1		4		1	12		2	3	1	24
J2b							2				1		3
L1											1		1
L1a					1			3			11	1	16
L2							1						1
N1a		2											2
N1c	1				1								2
Q							1						1
Q1a		10	5	1	7			1			1		25
Q1b								1			2	1	4
R			2							1	1		4
R1			5			1							6
R1a	1	1	55	6	14			1	1		2	3	84
R1b	2	26	5	1	3			1	1				39
R2									1		1		2
R2a						3	2	3		1	7		16
T					1		1	1					3
T1a							1						1
Grand Total	4	39	73	8	34	4	20	26	3	6	45	9	271

Table S 92 Y haplogroup counts by broad geographic area and time period. Low coverage individuals (<100,000 SNPs) and related individuals were excluded.

We find that the Y haplogroups of populations cluster both by geography and by time period. We highlight some key observations.

First, farmers from the Bronze Age and earlier from Turan and populations from the Steppe are distinctly different in their Y chromosome distributions. In Bronze Age Iran/Turan and South Asia, we find that haplogroups J, T, and R2 are spread widely from the border of present-day Anatolia to eastern Uzbekistan. This is in line with our observation that Iranian farmer-related ancestry makes up the majority of the ancestry of the individuals in this broad region. Individuals that we hypothesize descend from recent migrants from South Asia also share some of these haplogroups but in small proportions, and harbor additional haplogroups, L and H that appear at higher percentages. We note that two individuals along the

Indus_Periphery_Cline (1 from Gonur and 1 from Shahr-i-Sokhta) are in a South Asia specific haplogroup, H1a, providing additional evidence that these individuals, though found in Turan, actually are descendants of migrants from South Asia. Haplogroup L1a is also found in a small number of individuals from the BMAC and Shahr-i-Sokhta but not elsewhere in Turan, consistent with our autosomal results that peoples at these sites have additional ancestry from South Asia that was not present in Copper Age Turan. In addition, the fact that BA and Iron Age populations from South Asia have only 16/51 observed haplogroups from the J, T and R2 lineages that are shared with the majority of individuals from Turan, despite having more than 70% of their autosomal ancestry related to Iranian farmers, provides evidence along with a characteristically low ratio of Anatolian farmer-ancestry to Iranian farmer-related ancestry that the split between these populations of Iranian farmer-related ancestry is likely old.

We next examine individuals from the Steppe, and find that the Y chromosomal distribution in the region changes strikingly over time. The majority of the individuals we have from the Steppe prior to the MLBA are from the R1b haplogroup. Individuals from this haplogroup extend all across the western and central Steppe and into the Altai mountains (in individuals assigned to the Afanasievo culture), consistent with previous results (*8*). The *Afanasievo* individuals also contained multiple individuals from the Q1a haplogroup that we observe also in outlier individuals with additional *WSHG* ancestry in the MLBA, consistent with our autosomal estimates of admixture with *Central_Steppe_EMBA* related populations (who have high proportions of ancestry related to *WSHG*) as these populations moved eastward through the steppe. In the MLBA, however, there is a sudden transition in the Y chromosome modal haplogroup from R1b to R1a, with the R1a haplogroup occurring at 99% frequency in males in individuals of the *Western_Steppe_MLBA* and *Central_Steppe_MLBA* analysis labels. Further to the south, the R1a haplogroup is not observed in any of our sampled ancient DNA from Iran and Turan until after ~2000 BCE. Most notably, we observe R1a in 2 individuals from Late Bronze-Iron Age South Asia whose autosomal ancestry is typical of the *SPGT* genetic cluster, consistent with genome-wide estimates of Steppe pastoralist-related admixture into this group between 2000-1500 BCE (**Fig S 60**). R1a makes up a significant fraction of the Y haplogroups in present-day day Iran, Turan and South Asia, showing that once it arrived in this region in the MLBA and afterward, it had a long term impact.

S7.3 Interpreting mitochondrial DNA (mtDNA) assignments in light of our autosomal admixture models

In this section we describe the mtDNA haplogroup assignments in the ancient individuals in light of other findings in this study. We restricted our analysis to mtDNA haplogroups for individuals with at least 2-fold average mtDNA coverage. Unlike Y chromosome haplogroups which are highly structured by region and time period, the mtDNA haplogroup frequencies are much less differentiated. Nevertheless, there is some notable structure.

In the Steppe, we observe that the mtDNA haplogroup with the highest frequency in the Bronze Age is U5a (~25% of the sequenced individuals). This haplogroup is found at extremely low frequency in Bronze Age Turan (1.7%), consistent with our findings that individuals with *Western_Steppe_EMBA* related ancestry had not reached Turan by this time. In the Middle to Late Bronze Age, 13.6% of Steppe individuals we sequenced carry the T1a haplogroup which is found at 3.5% in the Late Bronze-Iron Age samples from Swat Valley, but is found only at extremely low frequency in BA Turan, suggesting that this haplogroup could provide a maternally transmitted uniparental marker link between the Steppe and South Asia (**Table S 93**).

We also analyzed subtypes of haplogroup M that are largely unique to South Asia today. The M3a, M4a, M5a, M30, M35 and M65 haplogroups are found at frequencies of around 1-5% in the Late Bronze-Iron Age and historical period samples from Swat Valley, but not in Turan or the Steppe. It is reasonable to link this mtDNA haplogroup to the *AASI* admixture detected in the Swat Valley samples but not in the Turan individuals.

	Forest / Steppe				Iran / Turan			South Asia		
	N	EMBA	MLBA	LBA	M / N	BA	C / EN	BA	IA	H
HV1						0.068	0.116		0.070	
M65									0.070	
M30								0.111	0.058	0.136
W3a				0.067		0.017	0.023	0.222	0.058	
M5a									0.047	0.045
U2b									0.047	0.045
U2c								0.111	0.047	
H							0.023		0.035	
R30						0.017			0.035	0.045
T		0.028	0.008							
T2a		0.083	0.008						0.012	
H13			0.008				0.023		0.012	0.091
C4a			0.008							0.045
G2a			0.008							
H5a			0.008							
H6b			0.008							
H7b			0.008							
HV6			0.008							0.045
K2a			0.008		0.250					
W1			0.008							
Z1			0.008							
R1b		0.028	0.016	0.067		0.017				
H6a		0.056	0.016							
N1a			0.016	0.067					0.023	
H3g			0.016							
H27			0.016							
W1c			0.016							
H2a		0.028	0.024			0.051			0.012	
U4b		0.111	0.024							
J1b			0.024			0.068	0.023		0.012	0.045
H1			0.024							
J2b			0.032	0.067						
T2e			0.032							
K1a			0.040			0.034	0.047		0.012	
H2b			0.040							0.045
J1c			0.040							
U4a		0.028	0.056	0.067						
T2b			0.056							
U5b		0.028	0.064	0.200			0.023			
U5a	0.333	0.250	0.096	0.067		0.017				
U2e	0.333		0.120			0.017			0.023	
T1a		0.028	0.136	0.067		0.017			0.035	
H15		0.028								0.045
I1b		0.028				0.017	0.070			
K1b		0.028					0.023			0.045
U4		0.028								
V1a		0.028								
U4d		0.056							0.012	
C5c		0.056								
J2a		0.083								
U7a					0.250	0.068	0.023		0.035	
I1c						0.017	0.047		0.023	
M									0.023	
T2g							0.023		0.023	
U1a						0.068	0.047	0.111	0.023	
U	0.333						0.023			
U7b									0.023	
U8b									0.023	
A+1									0.012	
D4j									0.012	
H14						0.017			0.012	
H3a									0.012	
HV2						0.068	0.023		0.012	
J1d							0.093		0.012	
M35									0.012	
M3a									0.012	0.091
M4									0.012	
M4a									0.012	
M70									0.012	
U2									0.012	
U2a				0.067					0.012	0.045
U3b									0.012	
U4c									0.012	
W						0.017			0.012	
W3b						0.017	0.070		0.012	
Z3a									0.012	
A1a				0.067						
A8a				0.067						
H2						0.017				
H29						0.017				
H+1						0.034				0.045
HV						0.034	0.093			
I1						0.017	0.047			
I2				0.067						
I4a						0.017				
J							0.023			
J1						0.034				
K1				0.067						
M33						0.017				
M49										0.045
M52										0.045
N1b						0.017				
R0						0.017				
R2					0.500			0.111		
R2+						0.017	0.023			
R5a						0.017		0.111		0.045
R6a						0.017				
R6b										0.045
R7								0.111		
T1						0.017				
T2c						0.051				
T2d						0.017	0.023			
U3a						0.017				
U7							0.023			
W1+							0.023			
W6						0.017		0.111		
X2p							0.023			

Table S 93 mtDNA haplogroup assignment by broad geographic region and time period