RESEARCH ARTICLE SUMMARY

HUMAN EVOLUTION

The formation of human populations in South and Central Asia

Vagheesh M. Narasimhan *et al.*

RATIONALE: To elucidate the extent to which the major cultural transformations of farming, pastoralism, and shifts in the distribution of languages in Eurasia were accompanied by movement of people, we report genome-wide ancient DNA data from 523 individuals spanning the last 8000 years, mostly from Central Asia and northernmost South Asia.

RESULTS: The movement of people following the advent of farming resulted in genetic gradients across Eurasia that can be modeled as mixtures of seven deeply divergent populations. A key gradient formed in southwestern Asia beginning in the Neolithic and continuing into the Bronze Age, with more Anatolian farmer–related ancestry in the west and more Iranian farmer–related ancestry in the east. This cline extended to the desert oases of Central Asia and was the primary source of ancestry in peoples of the Bronze Age Bactria Margiana Archaeological Complex (BMAC). This supports the idea that the archaeologically documented dispersal of domesticates was accompanied by the spread of people from multiple centers of domestication.

The main population of the BMAC carried no ancestry from Steppe pastoralists and did not contribute substantially to later South Asians. However, Steppe pastoralist ancestry appeared in outlier individuals at BMAC sites by the turn of the second millennium BCE around the same time as it appeared on the southern Steppe. Using data from ancient individuals from the Swat Valley of northernmost South Asia, we show that Steppe ancestry then integrated further south in the first half of the second millennium BCE, contributing up to 30% of the ancestry of modern groups in South Asia. The Steppe ancestry in South Asia has the same profile as that in Bronze Age Eastern Europe, tracking a movement of people that affected both regions and that likely spread the unique features shared between Indo-Iranian and Balto-Slavic languages.

The primary ancestral population of modern South Asians is a mixture of people related to early Holocene populations of Iran and South Asia that we detect in outlier individuals from two sites in cultural contact with the Indus Valley Civilization (IVC), making it plausible that it was characteristic of the IVC. After the IVC's decline, this population mixed with northwestern groups with Steppe ancestry to form the "Ancestral North Indians" (ANI) and also mixed with southeastern groups to form the "Ancestral South Indians" (ASI), whose direct descendants today live in tribal groups in southern India. Mixtures of these two post-IVC groups—the ANI and ASI—drive the main gradient of genetic variation in South Asia today.

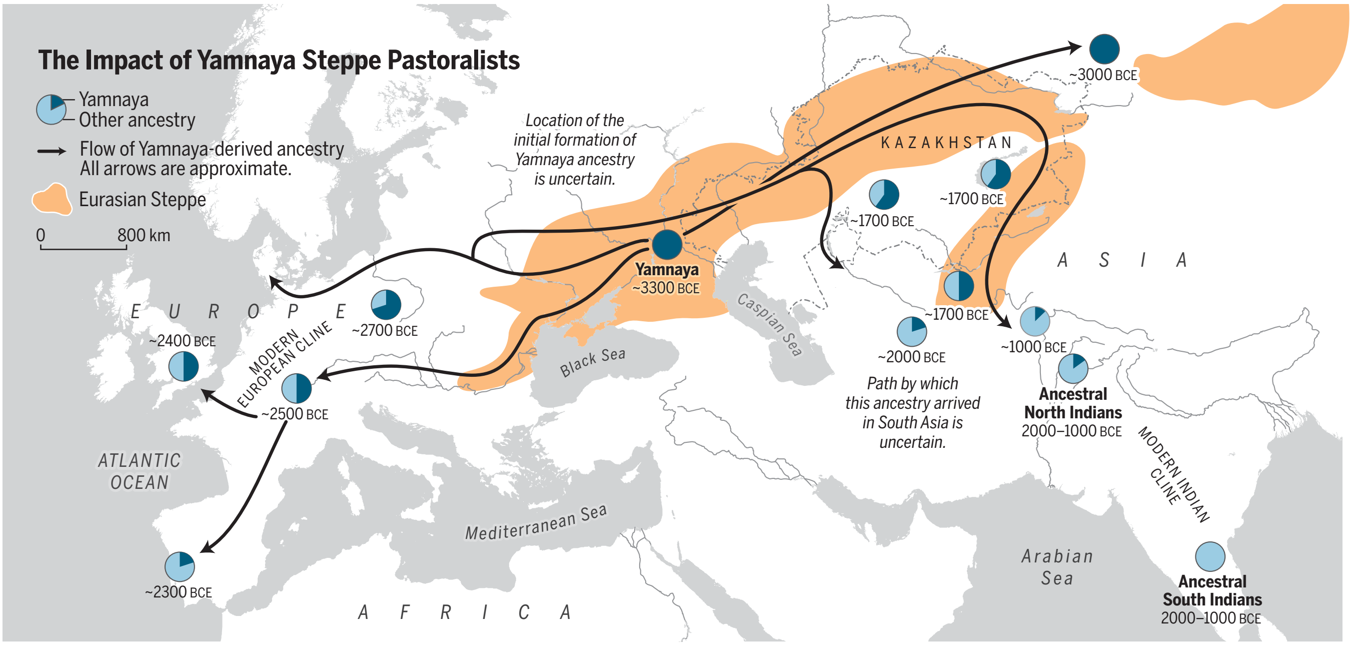
CONCLUSION: Earlier work recorded massive population movement from the Eurasian Steppe into Europe early in the third millennium BCE, likely spreading Indo-European languages. We reveal a parallel series of events leading to the spread of Steppe ancestry to South Asia, thereby documenting movements of people that were likely conduits for the spread of Indo-European languages. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat7487

The list of authors and affiliations is available in the full article online.
Corresponding authors: Vagheesh M. Narasimhan (vagheesh@mail.harvard.edu); Nick Patterson (nickp@broadinstitute.org); Michael Frachetti (frachetti@wustl.edu); Ron Pinhasi (ron.pinhasi@univie.ac.at); David Reich (reich@genetics.med.harvard.edu)

Cite this article as V. M. Narasimhan *et al.*, *Science* **365**, eaat7487 (2019). DOI: 10.1126/science.aat7487



The Bronze Age spread of Yamnaya Steppe pastoralist ancestry into two subcontinents—Europe and South Asia. Pie charts reflect the proportion of Yamnaya ancestry, and dates reflect the earliest available ancient DNA with Yamnaya ancestry in each region. Ancient DNA has not yet been found for the ANI and ASI, so for these the range is inferred statistically.

RESEARCH ARTICLE

HUMAN EVOLUTION

The formation of human populations in South and Central Asia

Vagheesh M. Narasimhan[1]*†, Nick Patterson[2,3]*†, Priya Moorjani[4,5]‡, Nadin Rohland[1,2]‡, Rebecca Bernardos[1], Swapan Mallick[1,2,6]‡, Iosif Lazaridis[1], Nathan Nakatsuka[1,7], Iñigo Olalde[1], Mark Lipson[1], Alexander M. Kim[1,8], Luca M. Olivieri[9], Alfredo Coppa[10], Massimo Vidale[9,11], James Mallory[12], Vyacheslav Moiseyev[13], Egor Kitov[14,15,16], Janet Monge[17], Nicole Adamski[1,6], Neel Alex[18], Nasreen Broomandkhoshbacht[1,6]§, Francesca Candilio[19,20], Kimberly Callan[1,6], Olivia Cheronet[19,21,22], Brendan J. Culleton[23], Matthew Ferry[1,6], Daniel Fernandes[19,21,22,24], Suzanne Freilich[22], Beatriz Gamarra[19,21,25]||¶, Daniel Gaudio[19,21], Mateja Hajdinjak[26], Éadaoin Harney[1,6,27], Thomas K. Harper[28], Denise Keating[19], Ann Marie Lawson[1,6], Matthew Mah[1,2,6], Kirsten Mandl[22], Megan Michel[1,6]#, Mario Novak[19,29], Jonas Oppenheimer[1,6]††, Niraj Rai[30,31], Kendra Sirak[1,19,32], Viviane Slon[26], Kristin Stewardson[1,6], Fatma Zalzala[1,6], Zhao Zhang[1], Gaziz Akhatov[15], Anatoly N. Bagashev[33], Alessandra Bagnera[9], Bauryzhan Baitanayev[15], Julio Bendezu-Sarmiento[34], Arman A. Bissembaev[15,35], Gian Luca Bonora[36], Temirlan T. Chargynov[37], Tatiana Chikisheva[38], Petr K. Dashkovskiy[39], Anatoly Derevianko[38], Miroslav Dobeš[40], Katerina Douka[41,42], Nadezhda Dubova[14], Meiram N. Duisengali[35], Dmitry Enshin[33], Andrey Epimakhov[43,44], Alexey V. Fribus[45], Dorian Fuller[46,47], Alexander Goryachev[33], Andrey Gromov[13], Sergey P. Grushin[48], Bryan Hanks[49], Margaret Judd[49], Erlan Kazizov[15], Aleksander Khokhlov[50], Aleksander P. Krygin[51], Elena Kupriyanova[52], Pavel Kuznetsov[50], Donata Luiselli[53], Farhod Maksudov[54], Aslan M. Mamedov[55], Talgat B. Mamirov[15], Christopher Meiklejohn[56], Deborah C. Merrett[57], Roberto Micheli[9,58], Oleg Mochalov[50], Samariddin Mustafokulov[54,59], Ayushi Nayak[41], Davide Pettener[60], Richard Potts[61], Dmitry Razhev[33], Marina Rykun[62], Stefania Sarno[60], Tatyana M. Savenkova[63], Kulyan Sikhymbaeva[64], Sergey M. Slepchenko[33], Oroz A. Soltobaev[37], Nadezhda Stepanova[38], Svetlana Svyatko[13,65], Kubatbek Tabaldiev[66], Maria Teschler-Nicola[22,67], Alexey A. Tishkin[68], Vitaly V. Tkachev[69], Sergey Vasilyev[14,70], Petr Velemínský[71], Dmitriy Voyakin[15,72], Antonina Yermolayeva[15], Muhammad Zahir[41,73], Valery S. Zubkov[74], Alisa Zubova[13], Vasant S. Shinde[75], Carles Lalueza-Fox[76], Matthias Meyer[26], David Anthony[77], Nicole Boivin[41]‡, Kumarasamy Thangaraj[30]‡, Douglas J. Kennett[23,28,78]‡, Michael Frachetti[79,80]†, Ron Pinhasi[19,22]†, David Reich[1,2,6,81]†**

By sequencing 523 ancient humans, we show that the primary source of ancestry in modern South Asians is a prehistoric genetic gradient between people related to early hunter-gatherers of Iran and Southeast Asia. After the Indus Valley Civilization's decline, its people mixed with individuals in the southeast to form one of the two main ancestral populations of South Asia, whose direct descendants live in southern India. Simultaneously, they mixed with descendants of Steppe pastoralists who, starting around 4000 years ago, spread via Central Asia to form the other main ancestral population. The Steppe ancestry in South Asia has the same profile as that in Bronze Age Eastern Europe, tracking a movement of people that affected both regions and that likely spread the distinctive features shared between Indo-Iranian and Balto-Slavic languages.

The past 10,000 years have witnessed profound economic changes driven by the transition from foraging to food production, as well as major changes in cultural practice that are evident from archaeology, the distribution of languages, and the written record. The extent to which these changes were associated with movements of people has been a mystery in Central Asia and South Asia, in part because of a paucity of ancient DNA. We report genome-wide data from 523 individuals from Central Asia and northernmost South Asia from the Mesolithic period onward (*1*), which we coanalyze with previously published ancient DNA from across Eurasia and with data from diverse present-day people.

In Central Asia, we studied the extent to which the spread of farming and herding practices from the Iranian plateau to the desert oases south of the Eurasian Steppe was accompanied by movements of people or adoption of ideas from neighboring groups (*2*–*4*). For the urban communities of the Bactria Margiana Archaeological Complex (BMAC) in the Bronze Age, we assessed whether the people buried in its cemeteries descended directly from earlier smaller-scale food producers, and we also documented their genetic heterogeneity (*5*). Farther to the north and east, we showed that the Early Bronze Age spread of crops and domesticated animals between Southwest Asia and eastern Eurasia along the Inner Asian Mountain Corridor (*6*) was accompanied by movements of people. Finally, we examined when descendants of the Yamnaya, who spread across the Eurasian Steppe beginning around 3300 BCE (*7*–*9*), began to appear in Central Asia south of the Steppe.

In northernmost South Asia, we report a time transect of >100 individuals beginning ~1200 BCE, which we coanalyze along with modern data from hundreds of present-day South Asian groups, as well as ancient DNA from neighboring regions (*10*). Previous analyses place the majority of present-day South Asians along a genetic cline (*11*) that can be modeled as having arisen from mixture of two highly divergent populations after around 4000 years ago: the Ancestral North Indians (*ANI*), who harbor large proportions of ancestry related to West Eurasians, and the Ancestral South Indians (*ASI*), who are much less closely related to West Eurasians (*12*). We leveraged ancient DNA to place constraints on the genetic structure of the *ANI* and *ASI* and, in conjunction with other lines of evidence, to make inferences about when and where they formed. By modeling modern South Asians along with ancient individuals from sites in cultural contact with the Indus Valley Civilization (IVC), we inferred a likely genetic signature for people of the IVC that reached its maturity in northwestern South Asia between 2600 and 1900 BCE. We also examined when Steppe pastoralist–derived ancestry (*9*) mixed into groups in South Asia, and placed constraints on whether Steppe-related ancestry or Iranian-related ancestry is more plausibly associated with the spread of Indo-European languages in South Asia.

Dataset and analysis strategy

We generated whole-genome ancient DNA data from 523 previously unsampled ancient individuals and increased the quality of data from 19 previously sequenced individuals. The individuals derive from three broad geographical regions: 182 from Iran and the southern part of Central Asia that we call Turan (present-day Turkmenistan, Uzbekistan, Tajikistan, Afghanistan, and Kyrgyzstan), 209 from the Steppe and northern forest zone mostly in present-day Kazakhstan and Russia, and 132 from northern Pakistan. The ancient individuals are from (i) Mesolithic, Copper, Bronze, and Iron Age Iran and Turan (12,000 to 1 BCE, from 19 sites) including the BMAC; (ii) early ceramic-using hunter-gatherers from the western Siberian forest zone, who we show represent a point along an early Holocene cline of North Eurasians and who emerge as a valuable source population for modeling the ancestry of Central and South Asians (6400 to 3900 BCE from 2 sites); (iii) Copper and Bronze Age pastoralists from the Central Steppe, including from Bronze Age Kazakhstan (3400 to 800 BCE from 56 sites); and (iv) northernmost South Asia, specifically Late Bronze Age, Iron Age, and historical settlements in the Swat and Chitral districts of present-day Pakistan (~1200 BCE to 1700 CE from 12 sites) (Fig. 1 and table S1) (*1*, *13*). We prepared samples in dedicated clean rooms, extracted DNA (*14*, *15*), and constructed libraries for Illumina sequencing (*16*, *17*). We enriched the libraries for DNA overlapping around 1.2 million single-nucleotide polymorphisms (SNPs), sequenced the products on Illumina instruments, and performed quality control (table S2) (*7*, *18*, *19*). Our final dataset after merging with previously reported data (*7*–*9*, *16*, *18*, *20*, *21*–*31*) spans 837 ancient individuals who passed all our analysis filters. These filters included restricting to the 92% of individuals who were represented by at least 15,000 of the targeted SNPs (the number at which we began to be able to reliably estimate proportions of the deeply divergent ancestry sources) (table S1). These filters also included removing individuals determined genetically to be first-degree relatives of other higher-coverage individuals (table S3). The median number of SNPs analyzed per individual was ~617,000. We also merged with previously reported whole-genome sequencing data from 686 present-day individuals (table S1) and coanalyzed with 1789 present-day people from 246 ethnographically distinct groups in South Asia genotyped at ~600,000 SNPs (table S5) (*10*, *13*, *27*, *32*).

We grouped individuals on the basis of archaeological and chronological information, taking advantage of 269 direct radiocarbon dates on skeletal material that we generated specifically for this study (table S4). We further clustered individuals who were genetically indistinguishable within these groupings and labeled outliers with ancestry that was significantly different from that of others at the same site and time period (*13*). For our primary analyses, we did not include individuals who were the sole representatives of their ancestry profiles, thereby reducing the chance that our conclusions were being driven by single individuals with contaminated DNA or misattributed archaeological context. This also ensured that each major analysis grouping was represented by many more SNPs than our minimum cutoff of 15,000 per individual. Thus, all but one analysis cluster included at least one individual covered by >200,000 SNPs, which is sufficient to support high-resolution analysis of population history (*18*) (the exception is a pair of genetically similar outliers from the site of Gonur that are not the focus of any main analyses). We use italic font to refer to genetic groupings and nonitalic font to indicate archaeological cultures or sites.

To make inferences about population structure, we began by carrying out principal components analysis (PCA) projecting ancient individuals onto the patterns of genetic variation in present-day Eurasians, a procedure that allowed us to obtain meaningful constraints even on ancestry of ancient individuals with limited coverage because each SNP from each individual can be compared to a large reference dataset (*33*–*35*). This revealed three major clusters strongly correlating to the geographic regions of the Forest Zone/Steppe, Iran/Turan, and South Asia (Fig. 1), a pattern we replicate in ADMIXTURE unsupervised clustering (*36*). To test if groups of ancient individuals were heterogeneous in their ancestry, we used f_4-statistics to measure whether different partitions of these groups into two subgroups differed in their degree of allele sharing to a third group (using a distantly related outgroup as a baseline). We also used f_3-statistics to test for admixture (*32*). To model the ancestry of each group, we

---

[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [2]Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA. [3]Radcliffe Institute for Advanced Study, Harvard University, Cambridge, MA 02138, USA. [4]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA. [5]Center for Computational Biology, University of California, Berkeley, CA 94720, USA. [6]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA. [7]Harvard-MIT Division of Health Sciences and Technology, Harvard Medical School, Boston, MA 02115, USA. [8]Department of Anthropology, Harvard University, Cambridge, MA 02138, USA. [9]ISMEO - International Association of Mediterranean and Oriental Studies, Italian Archaeological Mission in Pakistan, 19200 Saidu Sharif (Swat), Pakistan. [10]Department of Environmental Biology, Sapienza University, Rome 00185, Italy. [11]Department of Cultural Heritage: Archaeology and History of Art, Cinema and Music, University of Padua, Padua 35139, Italy. [12]School of Natural and Built Environment, Queen's University Belfast, Belfast BT7 1NN, Northern Ireland, UK. [13]Peter the Great Museum of Anthropology and Ethnography (Kunstkamera), Russian Academy of Science, St. Petersburg 199034, Russia. [14]Center of Physical Anthropology, Institute of Ethnology and Anthropology, Russian Academy of Sciences, Moscow 119991, Russia. [15]A.Kh. Margulan Institute of Archaeology, Almaty 050010, Kazakhstan. [16]Al-Farabi Kazakh National University, Almaty 050040, Kazakhstan. [17]University of Pennsylvania Museum of Archaeology and Anthropology, Philadelphia, PA 19104, USA. [18]Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720, USA. [19]Earth Institute, University College Dublin, Dublin 4, Ireland. [20]Soprintendenza Archeologia, Belle Arti e Paesaggio per la Città Metropolitana di Cagliari e le Province di Oristano e Sud Sardegna, Cagliari 09124, Italy. [21]School of Archaeology, University College Dublin, Dublin 4, Ireland. [22]Department of Evolutionary Anthropology, University of Vienna, 1090 Vienna, Austria. [23]Institutes of Energy and the Environment, Pennsylvania State University, University Park, PA 16802, USA. [24]CIAS, Department of Life Sciences, University of Coimbra, Coimbra 3000-456, Portugal. [25]Catalan Institute of Human Paleoecology and Social Evolution (IPHES), Tarragona 43007, Spain. [26]Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany. [27]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. [28]Department of Anthropology, Pennsylvania State University, University Park, PA 16802, USA. [29]Institute for Anthropological Research, Zagreb 10000, Croatia. [30]CSIR-Centre for Cellular and Molecular Biology, Hyderabad 500 007, India. [31]Birbal Sahni Institute of Palaeosciences, Lucknow 226007, India. [32]Department of Anthropology, Emory University, Atlanta, GA 30322, USA. [33]Tyumen Scientific Centre SB RAS, Institute of the Problems of Northern Development, Tyumen 625003, Russia. [34]CNRS-EXT500, Directeur de la Delegation Archaologique Francaise en Afghanistan (DAFA), Embassy of France in Kabul, Afghanistan. [35]Aktobe Regional Historical Museum, Aktobe 030006, Kazakhstan. [36]Archaeology of Asia Department, ISMEO - International Association of Mediterranean and Oriental Studies, Rome RM00186, Italy. [37]Kyrgyz National University, 720033 Bishkek, Kyrgyzstan. [38]Institute of Archaeology and Ethnography, Siberian Branch, Russian Academy of Sciences, Novosibirsk 630090, Russia. [39]Department of Political History, National and State-Confessional Relations, Altai State University, Barnaul 656049, Russia. [40]Institute of Archaeology, Czech Academy of Sciences, Prague 118 01, Czech Republic. [41]Department of Archaeology, Max Planck Institute for the Science of Human History, Jena 07745, Germany. [42]Oxford Radiocarbon Accelerator Unit, Research Laboratory for Archaeology and the History of Art, University of Oxford, Oxford OX1 3QY, UK. [43]Institute of History and Archaeology, Ural Branch RAS, Yekaterinburg 620990, Russia. [44]South Ural State University, Chelyabinsk 454080, Russia. [45]Department of Archaeology, Kemerovo State University, Kemerovo 650043, Russia. [46]Institute of Archaeology, University College London, London WC1H 0PY, UK. [47]School of Cultural Heritage, Northwest University, Shanxi, 710069, China. [48]Department of Archaeology, Ethnography and Museology, Altai State University, Barnaul 656049, Russia. [49]Department of Anthropology, University of Pittsburgh, Pittsburgh, PA 15260, USA. [50]Samara State University of Social Sciences and Education, Samara 443099, Russia. [51]West Kazakhstan Regional Center for History and Archaeology, Uralsk 090000, Kazakhstan. [52]Scientific and Educational Center of Study on the Problem of Nature and Man, Chelyabinsk State University, Chelyabinsk 454021, Russia. [53]Department of Cultural Heritage, University of Bologna, 48121 Ravenna, Italy. [54]Institute for Archaeological Research, Uzbekistan Academy of Sciences, Samarkand 140151, Uzbekistan. [55]Center for Research, Restoration and Protection of Historical and Cultural Heritage of Aktobe Region, Aktobe 030007, Kazakhstan. [56]Department of Anthropology, University of Winnipeg, Winnipeg, MB R3B 2E9, Canada. [57]Department of Archaeology, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. [58]MiBAC – Ministero per i Beni e le Attività Culturali - Soprintendenza Archeologia, belle arti e paesaggio del Friuli Venezia Giulia, 34135 Trieste, Italy. [59]Afrosiab Museum, Samarkand 140151, Uzbekistan. [60]Department of Biological, Geological and Environmental Sciences, Alma Mater Studiorum – University of Bologna, Bologna 40126, Italy. [61]Human Origins Program, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013, USA. [62]National Research Tomsk State University, Tomsk 634050, Russia. [63]F. Voino-Yasenetsky Krasnoyarsk State Medical University, Krasnoyarsk 660022, Russia. [64]Central State Museum Republic of Kazakhstan, Samal-1 Microdistrict, Almaty 050010, Kazakhstan. [65]CHRONO Centre for Climate, the Environment, and Chronology, Queen's University of Belfast, Belfast BT7 1NN, Northern Ireland, UK. [66]Kyrgyz-Turkish Manas University, Bishkek, Kyrgyzstan. [67]Department of Anthropology, Natural History Museum Vienna, 1010 Vienna, Austria. [68]Department of Archaeology, Ethnography and Museology, The Laboratory of Interdisciplinary Studies in Archaeology of Western Siberia and Altai, Altai State University, Barnaul 656049, Russia. [69]Institute of Steppe, Ural Branch RAS, Orenburg 460000, Russia. [70]Center for Egyptological Studies RAS, Moscow 119991, Russia. [71]Department of Anthropology, National Museum, Prague 115 79, Czech Republic. [72]Archaeological Expertise LLP, Almaty 050060, Kazakhstan. [73]Department of Archaeology, Hazara University, Mansehra 21300, Pakistan. [74]N.F. Katanov Khakassia State University, Abakan 655017, Russia. [75]Department of Archaeology, Deccan College Post-Graduate and Research Institute, Pune 411006, India. [76]Institute of Evolutionary Biology, CSIC-Universitat Pompeu Fabra, Barcelona 08003, Spain. [77]Anthropology Department, Hartwick College, Oneonta, NY 13820, USA. [78]Department of Anthropology, University of California, Santa Barbara, CA 93106, USA. [79]Department of Anthropology, Washington University in St. Louis, St. Louis, MO 63112, USA. [80]Spatial Analysis, Interpretation, and Exploration Laboratory, Washington University in St. Louis, St. Louis, MO 63112, USA. [81]Max Planck-Harvard Research Center for the Archaeoscience of the Ancient Mediterranean, Cambridge, MA 02138, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: vagheesh@mail.harvard.edu (V.M.N.); nickp@broadinstitute.org (N.P.); frachetti@wustl.edu (M.Fr.); ron.pinhasi@univie.ac.at (R.Pi.); reich@genetics.med.harvard.edu (D.Re.)
‡These authors contributed equally to this work. §Present address: Department of Anthropology, University of California, Santa Cruz, CA 95064, USA. ||Present address: Institut Català de Paleoecologia Humana i Evolució Social (IPHES), 43007 Tarragona, Spain. ¶Present address: Universitat Rovira i Virgili (URV), Àrea de Prehistòria, 43002 Tarragona, Spain. #Present address: Max Planck-Harvard Research Center for the Archaeoscience of the Ancient Mediterranean, Cambridge, MA 02138, USA. **Present address: Department of Human Evolutionary Biology, Harvard University, Cambridge MA, 02138, USA. ††Present address: Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064, USA.

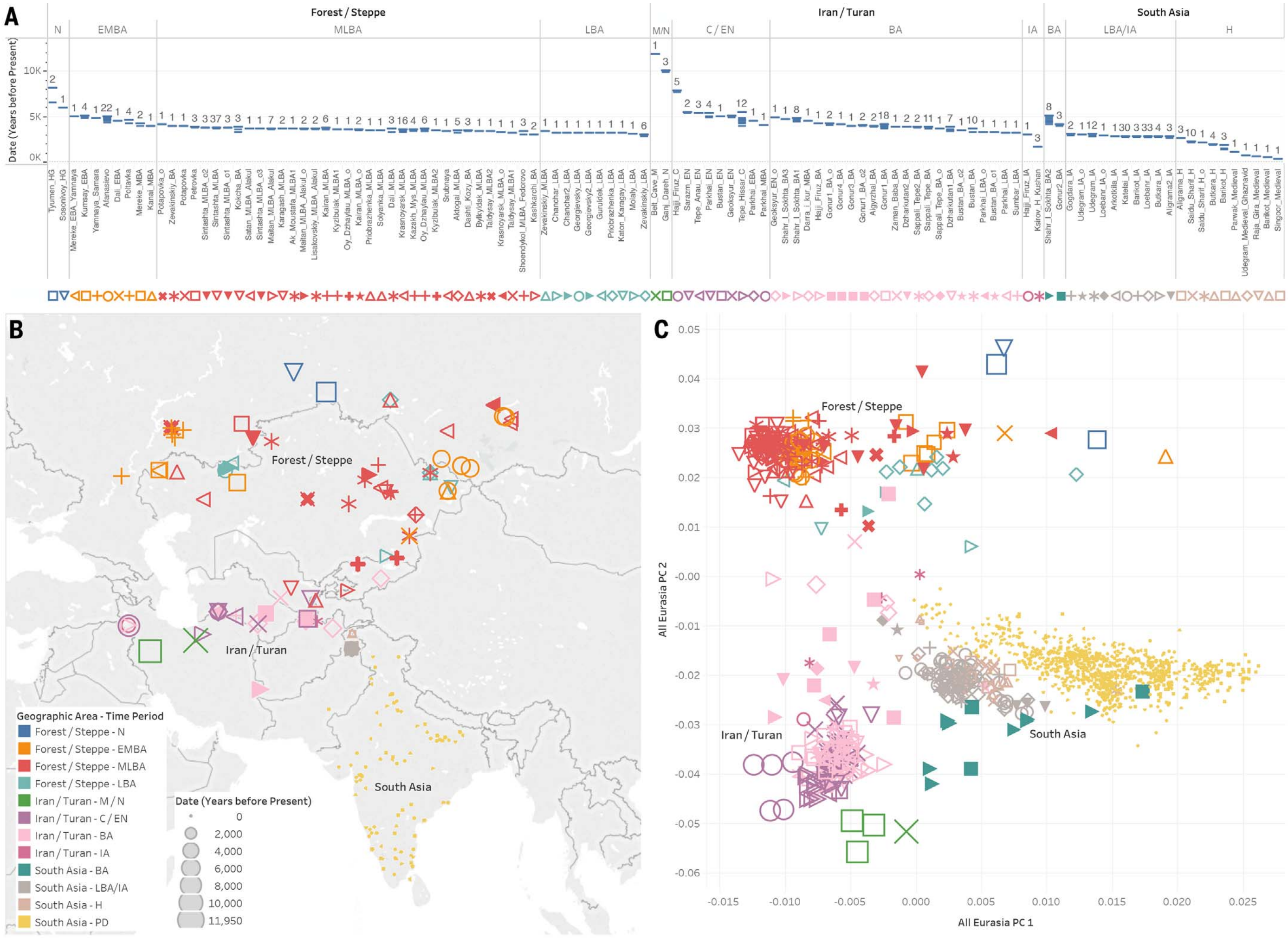


Fig. 1. Overview of ancient DNA data. (**A**) Distribution of sites and associated archeological or radiocarbon dates along with the number of individuals meeting our analysis thresholds from each site. (**B**) Locations of ancient individuals for whom we generated ancient DNA that passed our analysis thresholds along with the locations of individuals from 140 groups from present-day South Asia that we analyzed as forming the *Modern Indian Cline*. Shapes distinguish the individuals from different sites. Data from 106 South Asian groups that do not fit along the *Modern Indian Cline* as well as *AHG* are not shown. (**C**) PCA analysis of ancient and modern individuals projected onto a basis formed by 1340 present-day Eurasians reflects clustering of individuals that mirrors their geographical relationships. An interactive version of this figure is presented in our online data visualizer (*1*).

used *qpAdm*, which evaluates whether a tested group is consistent with deriving from a prespecified number of source populations (relative to a set of outgroups) and, if so, estimates proportions of ancestry (*7*). We first used *qpAdm* to attempt to model groups from the Copper Age and afterward as a mixture of seven "distal" sources, using as surrogates for them six pre–Copper Age populations and one modern Andamanese hunter-gatherer population (Box 1). (The model assumes that each true ancestral population is a clade with the population we use as a surrogate for it in the sense of descending from the same ancestral population, possibly deeply in time.) In this paper, we use the term "farmers" to refer to people who cultivated crops, herded animals, or both; this definition covers not only large settled communities but also smaller and probably less sedentary communities like the early herders of the Zagros Mountains of western Iran from the site of Ganj Dareh. The latter kept domesticated animals but did not cultivate crops and are a key reference population for this study, as they had a distinctive ancestry profile that spread widely after the Neolithic (*9*, *28*, *37*). We also identified proximal models for each group as mixtures of temporally preceding groups. We implemented an algorithm called DATES for estimating the age of the population mixtures (*13*), which is related to previous methods that translate the average size of ancestry blocks into time since mixture by leveraging precise measurements of meiotic recombination rate in humans (*32*, *38*, *39*). DATES has the specific advantage that it is optimized relative to previous methods in being able to work with ancient DNA as well as with single genomes (*13*). In Box 2, we summarize the findings of these analyses (we use the same headings in Box 2 and the main text to allow cross-referencing), whereas our online data visualizer (*1*) allows an interactive exploration of the data.

Iran and Turan

A west-to-east cline of decreasing Anatolian farmer–related ancestry

We studied the genetic transformations accompanying the spread of agriculture eastward from Iran beginning in the 7th millennium BCE (*3*, *40*, *41*). We replicate previous findings that 9th to 8th millennium BCE herders from the Zagros Mountains of western Iran harbored a distinctive West Eurasian–related ancestry profile (*9*, *31*), whereas later groups across a broad region were admixed between this ancestry and that related to early Anatolian farmers. Our analysis reveals a west-to-east cline of decreasing Anatolian farmer–related admixture in the Copper and Bronze Ages ranging from ~70% in Anatolia to ~31% in eastern Iran to ~7% in far eastern Turan (Fig. 1,

Box 1. Seven source populations used for distal ancestry modeling.

Anatolia_N, Anatolian farmer–related: Represented by seventh millennium BCE western Anatolian farmers (*18*).

Ganj_Dareh_N, Iranian early farmer–related: Represented by eighth millennium BCE early goat herders from the Zagros Mountains of Iran (*9*, *24*).

WEHG, Western European hunter-gatherer–related: Represented by ninth millennium BCE Western Europeans (*7*, *18*, *27*, *91*). (*WEHG* and *EEHG* discussed below were denoted *WHG* and *EHG* in previous studies, but as we coanalyze them with hunter-gatherers from Asia, we modify the names to specify a European origin.)

EEHG, Eastern European hunter-gatherer–related: Represented by sixth millennium BCE hunter-gatherers from Eastern Europe (*18*, *27*).

WSHG, West Siberian hunter-gatherer–related: A previously undescribed deep source of Eurasian ancestry represented in this study by three individuals from the Forest Zone of Central Russia dated to the sixth millennium BCE.

ESHG, East Siberian hunter-gatherer–related: Represented by sixth millennium BCE hunter-gatherers from the Lake Baikal region with ancestry deeply related to East Asians (*26*).

AHG, Andamanese hunter-gatherer–related: Represented by present-day indigenous Andaman Islanders (*53*) who we hypothesize are related to unsampled indigenous South Asians (*AASI*, *Ancient Ancestral South Indians*).

fig. S10, and tables S8 to S16) (*13*). This suggests that the archaeologically documented spread of a shared package of plants and animal domesticates from diverse locations across this region was accompanied by bidirectional spread of people and mixture with the local groups they encountered (*3*, *40*, *42*, *43*). We call this the *Southwest Asian Cline*. In the far east of the *Southwest Asian Cline* (eastern Iran and Turan) in individuals from the third millennium BCE, we detect not only the smallest proportions of Anatolian farmer–related admixture but also admixture related to West Siberian Hunter Gatherers (*WSHG*), plausibly reflecting admixture from unsampled hunter-gatherer groups that inhabited this region before the spread of Iranian farmer–related ancestry into it. This shows that North Eurasian–related ancestry affected Turan well before the spread of descendants of Yamnaya Steppe pastoralists into the region. We can exclude the possibility that the Yamnaya were the source of this North Eurasian–related ancestry, as they had more Eastern European Hunter Gatherer (*EEHG*)–related than *WSHG*-related ancestry, and they also carried high frequencies of mitochondrial DNA haplogroup type U5a as well as Y chromosome haplogroup types R1b or R1a that are absent in ancient DNA sampled from Iran and Turan in this period (tables S93 and S94) (*13*).

People of the BMAC were not a major source of ancestry for South Asians

From Bronze Age Iran and Turan, we obtained genome-wide data for 84 ancient individuals (3000 to 1400 BCE) who lived in four urban sites of the BMAC and its immediate successors. The great majority of these individuals fall in a cluster genetically similar to the preceding groups in Turan, which is consistent with the hypothesis that the BMAC coalesced from preceding pre-urban populations (*5*). We infer three primary genetic sources: early Iranian farmer–related ancestry (~60 to 65%) and smaller proportions of Anatolian farmer–related ancestry (~20 to 25%) and *WSHG*-related ancestry (~10%). Unlike preceding Copper Age individuals from Turan, people of the *BMAC* cluster also harbored an additional ~2 to 5% ancestry related (deeply in time) to Andamanese hunter-gatherers (*AHG*). This evidence of south-to-north gene flow from South Asia is consistent with the archaeological evidence of cultural contacts between the IVC and the BMAC and the existence of an IVC trading colony in northern Afghanistan (although we lack ancient DNA from that site) (*44*) and stands in contrast to our *qpAdm* analyses showing that a reciprocal north-to-south spread is undetectable. Specifically, our analyses reject the *BMAC* and the people who lived before them in Turan as plausible major sources of ancestry for diverse ancient and modern South Asians by showing that their ratio of Anatolian farmer–related to Iranian farmer–related ancestry is too high for them to be a plausible source for South Asians [$P < 0.0001$, χ^2 test; (*13*)] (figs. S50 and S51). A previous study (*30*) fit a model in which a population from Copper Age Turan was used as a source of the Iranian farmer–related ancestry in present-day South Asians, thus raising the possibility that the people of the BMAC whom the authors correctly hypothesized were primarily derived from the groups that preceded them in Turan were a major source population for South Asians. However, that study only had access to two samples from this period compared with the 36 we analyze in this study, and it lacked ancient DNA from individuals from the BMAC period or from any ancient South Asians. With additional samples, we have the resolution to show that none of the large number of Bronze and Copper Age populations from Turan for which we have ancient DNA fit as a source for the Iranian farmer–related ancestry in South Asia.

Steppe pastoralist–derived ancestry arrived in Turan by 2100 BCE

Our large sample sizes from Central Asia, including individuals from BMAC sites, are a particular strength of this study, allowing us to detect outlier individuals whose ancestry differs from that of those living at the same time and place and revealing cultural contacts that would be otherwise difficult to appreciate (Fig. 2). Around 2300 BCE, we observe three outliers in BMAC-associated sites carrying *WSHG*-related ancestry and we report data from the third millennium BCE from three sites in Kazakhstan and one in Kyrgyzstan that fit as sources for them [related ancestry has been found in ~3500-BCE Botai culture individuals (*30*)]. *Yamnaya*-derived ancestry arrived by 2100 BCE, because from 2100 to 1700 BCE we observe outliers from three BMAC-associated sites carrying ancestry ultimately derived from *Western_Steppe_EMBA* pastoralists, in the distinctive admixed form typically carried by many Middle to Late Bronze Age Steppe groups (with roughly two-thirds of the ancestry being of *Western_Steppe_EMBA* origin, and the rest consistent with deriving from European farmers). Thus, our data document a southward movement of ancestry ultimately descended from Yamnaya Steppe pastoralists who spread into Central Asia by the turn of the second millennium BCE.

An ancestry profile widespread during the Indus Valley Civilization

We document 11 outliers—three with radiocarbon dates between 2500 and 2000 BCE from the BMAC site of Gonur and eight with radiocarbon dates or archaeological-context dates between 3300 and 2000 BCE from the eastern Iranian site of Shahr-i-Sokhta—that harbored elevated proportions of *AHG*-related ancestry (range: ~11 to 50%) and the remainder from a distinctive mixture of Iranian farmer– and *WSHG*-related ancestry (~50 to 89%). These outliers had no detectable Anatolian farmer–related ancestry, in contrast with the main *BMAC* (~20 to 25% Anatolian-related) and Shahr-i-Sokhta (~16 to 21%) clusters, allowing us to reject both the main *BMAC* and Shahr-i-Sokhta clusters as sources for the outliers [$P < 10^{-7}$, χ^2 test; (*13*)] (table S83). Without ancient DNA from individuals buried in IVC cultural contexts, we cannot make a definitive statement that the genetic gradient represented by these 11 outlier individuals, which we call the *Indus Periphery Cline*, was also an ancestry profile common in the IVC. Nevertheless, our result provides six circumstantial lines of evidence for this hypothesis. (i) These individuals had no detectable Anatolian farmer–related ancestry, suggesting they descend from groups farther east along the Anatolia-to-Iran cline of decreasing Anatolian farmer–related ancestry than any individuals we sampled from this period. (ii) All 11 outliers had elevated proportions of *AHG*-related ancestry, and two carried Y chromosome haplogroup H1a1d2, which today is primarily found in southern India. (iii) At both Gonur and Shahr-i-Sokhta there is archaeological evidence of exchange with the IVC (*45*, *46*), and all the outlier individuals we dated directly fall within the time frame of the mature IVC. (iv) Several outliers at Shahr-i-Sokhta were buried with artifacts stylistically linked to Baluchistan in South Asia, whereas burials

Box 2. Summary of key findings.

Iran and Turan

1. *A west-to-east cline of decreasing Anatolian farmer–related ancestry.* There was a west-to-east gradient of ancestry across Eurasia in the Copper and Bronze Ages—the *Southwest Asian Cline*—with more Anatolian farmer–related ancestry in the west and more *WSHG*- or *AASI*-related ancestry in the east, superimposed on primary ancestry related to early Iranian farmers. The establishment of this gradient correlates in time to the spread of plant-based agriculture across this region, raising the possibility that people of Anatolian ancestry spread this technology east just as they helped spread it west into Europe.

2. *People of the BMAC were not a major source of ancestry for South Asians.* The primary *BMAC* population largely derived from preceding local Copper Age peoples who were, in turn, closely related to people from the Iranian plateau and had little of the Steppe ancestry that is ubiquitous in South Asia today.

3. *Steppe pastoralist–derived ancestry arrived in Turan by 2100 BCE.* We find no evidence of Steppe pastoralist–derived ancestry in groups at BMAC sites before 2100 BCE, but multiple outlier individuals buried at these sites show that by ~2100 to 1700 BCE, BMAC communities were regularly interacting with peoples carrying such ancestry.

4. *An ancestry profile widespread during the Indus Valley Civilization.* We document a distinctive ancestry profile—~45 to 82% Iranian farmer–related and ~11 to 50% *AASI* (with negligible Anatolian farmer–related admixture)—present at two sites in cultural contact with the Indus Valley Culture (IVC). Combined with our detection of this same ancestry profile (in mixed form) about a millennium later in the post-IVC Swat Valley, this documents an *Indus Periphery Cline* during the flourishing of the IVC. Ancestors of this group formed by admixture ~5400 to 3700 BCE.

The Steppe and Forest Zone

1. *Ancestry clines in North Eurasia established after the advent of farming.* Before the spread of farmers and herders, northern Eurasia was characterized by a west-to-east gradient of very divergent hunter-gatherer populations with increasing proportions of relatedness to present-day East Asians: from Western European hunter-gatherers (*WEHG*), to Eastern European hunter-gatherers (*EEHG*), to West Siberian hunter-gatherers (*WSHG*), to East Siberian hunter-gatherers (*ESHG*). Mixture of people along this ancestry gradient and its counterpart to the south formed five later clines after the advent of farming, of which the three northern ones are the *European Cline*, the *Caucasus Cline*, and the *Central Asian Cline*.

2. *A distinctive ancestry profile stretching from Eastern Europe to Kazakhstan in the Bronze Age.* We add >100 samples from the previously described *Western_Steppe_MLBA* genetic cluster, including individuals associated with the Corded Ware, Srubnaya, Petrovka, and Sintashta archaeological complexes, and characterized by a mixture of about two-thirds ancestry related to Yamnaya Steppe pastoralists (from the *Caucasus Cline*) and European farmers (from the *European Cline*), suggesting that this population formed at the geographic interface of these two groups in Eastern Europe. Our analysis suggests that in the Central Steppe and Minusinsk Basin in the Middle to Late Bronze Age, *Western_Steppe_MLBA* ancestry mixed with ~9% ancestry from previously established people from the region carrying *WSHG*-related ancestry to form a distinctive *Central_Steppe_MLBA* cluster that was the primary conduit for spreading Yamnaya Steppe pastoralist–derived ancestry to South Asia.

3. *Bidirectional mobility along the Inner Asian Mountain Corridor.* Beginning in the third millennium BCE and intensifying in the second millennium BCE, we observe multiple individuals in the Central Steppe who lived along the Inner Asian Mountain Corridor and who harbored admixture from Turan, documenting northward movement into the Steppe in this period. By the end of the second millennium BCE, these people were joined by numerous outlier individuals with East Asian–related admixture that became ubiquitous in the region by the Iron Age (*29*, *52*). This East Asian–related admixture is also seen in later groups with known cultural impacts on South Asia, including Huns, Kushans, and Sakas, and is hardly present in the two primary ancestral populations of South Asia, suggesting that the Steppe ancestry widespread in South Asia derived from pre–Iron Age Central Asians.

South Asia

1. *Three ancestry clines that succeeded each other in time in South Asia.* We identify a distinctive trio of source populations that fits geographically and temporally diverse South Asians since the Bronze Age: a mixture of *AASI*, an *Indus Periphery Cline* group with predominantly Iranian farmer–related ancestry, and *Central_Steppe_MLBA*. Two-way clines that are well modeled as mixtures of pairs of populations that are themselves formed of these three sources succeeded each other in time: before 2000 BCE, the *Indus Periphery Cline* had no detectable Steppe ancestry, beginning after 2000 BCE the *Steppe Cline*, and finally the *Modern Indian Cline*.

2. *The ASI and ANI arose as* Indus Periphery Cline *people mixed with groups to the north and east.* An ancestry gradient of which the *Indus Periphery Cline* individuals were a part played a pivotal role in the formation of both the two proximal sources of ancestry in South Asia: a minimum of ~55% *Indus Periphery Cline* ancestry for the *ASI* and ~70% for the *ANI*. Today there are groups in South Asia with very similar ancestry to the statistically reconstructed *ASI*, suggesting that they have essentially direct descendants today. Much of the formation of both the *ASI* and *ANI* occurred in the second millennium BCE. Thus, the events that formed both the *ASI* and *ANI* overlapped the time of the decline of the IVC.

3. *Steppe ancestry in modern South Asians is primarily from males and disproportionately high in Brahmin and Bhumihar groups.* Most of the Steppe ancestry in South Asia derives from males, pointing to asymmetric social interaction between descendants of Steppe pastoralists and peoples of the *Indus Periphery Cline*. Groups that view themselves as being of traditionally priestly status, including Brahmins who are traditional custodians of liturgical texts in the early Indo-European language Sanskrit, tend (with exceptions) to have more Steppe ancestry than expected on the basis of *ANI-ASI* mixture, providing an independent line of evidence for a Steppe origin for South Asia's Indo-European languages.

associated with the other ancestries did not have these linkages (*13*). (v) In our modeling, the 11 outliers fit as a primary source of ancestry for 86 ancient individuals from post-IVC cultures living near the headwaters of the Indus River ~1200 to 800 BCE as well as diverse present-day South Asians, whereas no other ancient genetic clusters from Turan fit as sources for all these groups (*13*) (fig. S50). (vi) The estimated date of admixture between Iranian farmer–related and *AHG*-related ancestry in the outliers is several millennia before the time they lived (71 ± 15 generations, corresponding to a 95% confidence interval of ~5400 to 3700 BCE assuming 28 years per generation) (*13*, *47*). Thus, *AHG*- and Iranian farmer–related groups were in contact well before the time of the mature IVC at ~2600 to 1900 BCE, as might be expected if the ancestry gradient was a major feature of a group that was living in the Indus Valley during the IVC.

The Steppe and Forest Zone

Ancestry clines in Eurasia established after the advent of farming

The late hunter-gatherer individuals from northern Eurasia lie along a west-to-east hunter-gatherer gradient of increasing relatedness to East Asians

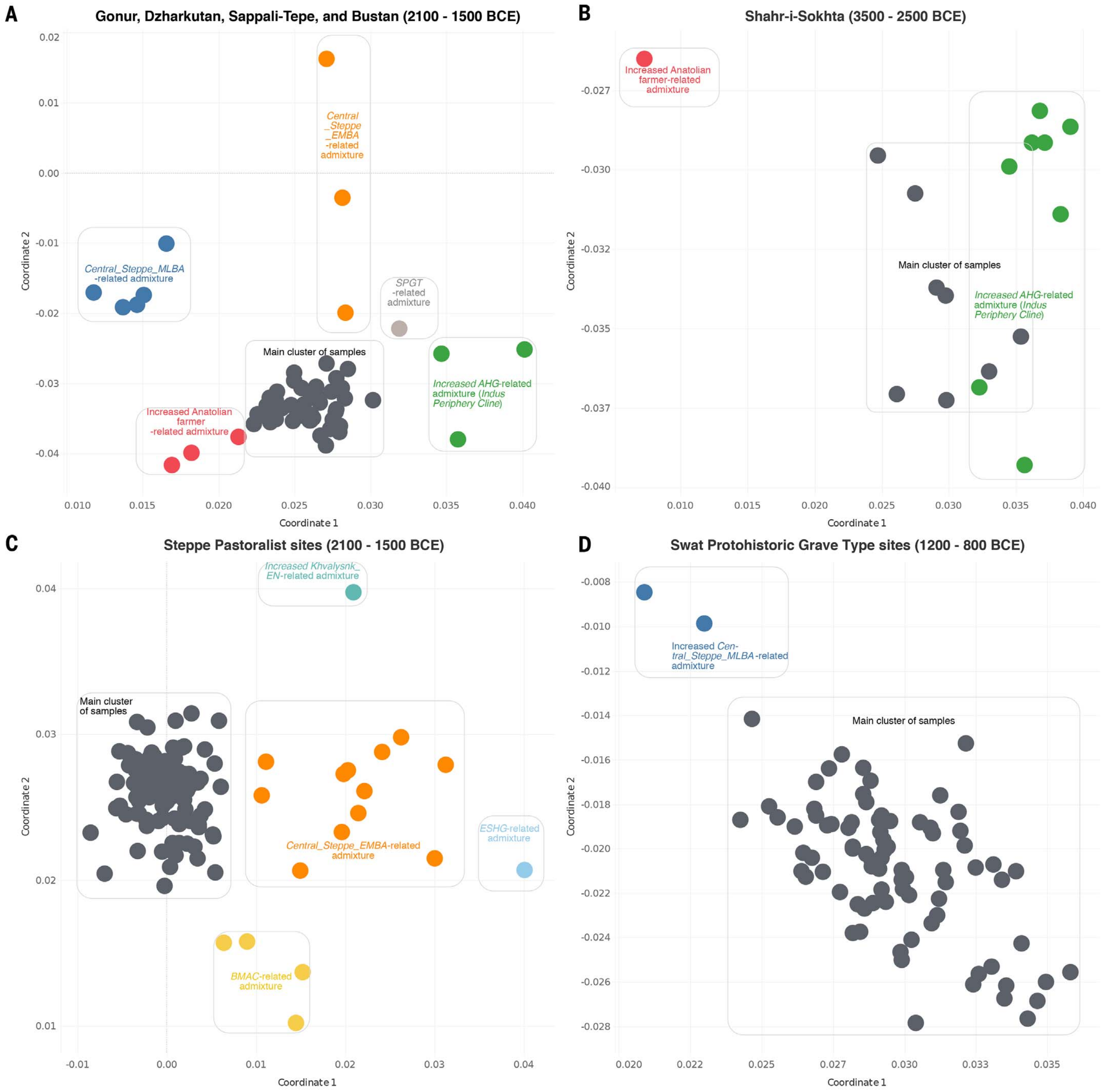


Fig. 2. Outlier analysis reveals ancient contacts between sites. We plot the average of principal component 1 (*x* axis) and principal component 2 (*y* axis) for the West Eurasian and All Eurasian PCA plots, as we found that this aids visual separation of the ancestry profiles. (**A**) In individuals of the BMAC and successor sites, we observe a main cluster as well as numerous outliers: outliers >2000 BCE with admixture related to *WSHG*, outliers >2000 BCE on the *Indus Periphery Cline* (with an ancestry similar to the outliers at Shahr-i-Sokhta), and outliers after 2000 BCE that reveal how *Central_Steppe_MLBA* ancestry had arrived. (**B**) At Shahr-i-Sokhta in eastern Iran, there are two primary groupings: one with ~20% Anatolian farmer–related ancestry and no detectable *AHG*-related ancestry, and the other with ~0% Anatolian farmer–related ancestry and substantial *AHG*-related ancestry (*Indus Periphery Cline*). (**C**) In the Middle to Late Bronze Age Steppe, we observe, in addition to the *Western_Steppe_MLBA* and *Central_Steppe_MLBA* clusters (indistinguishable in this projection), outliers admixed with other ancestries. The *BMAC*-related admixture in Kazakhstan documents northward gene flow onto the Steppe and confirms the Inner Asian Mountain Corridor as a conduit for movement of people. (**D**) In the Late Bronze Age and Iron Age of northernmost South Asia, we observe a main cluster consistent with admixture between peoples of the *Indus Periphery Cline* and *Central_Steppe_MLBA* and variable Steppe pastoralist–related admixture.

(Fig. 3). In the Neolithic and Copper Ages, hunter-gatherers at different points along this cline mixed with people with ancestry at different points along a southern cline to form five later clines, two of which were in the south (the *Southwest Asian Cline* and the *Indus Periphery Cline* described in the previous section) and three of which were in northern Eurasia (Fig. 3). Furthest to the west in the Steppe and Forest Zone there was the *European Cline*, established by the spread of farmers from Anatolia after ~7000 BCE and mixture with Western European hunter-gatherers (*18*). In far eastern Europe at latitudes spanning the Black and Caspian Seas there was the *Caucasus Cline,* consisting of a mixture of Eastern European hunter-gatherers and Iranian farmer–related ancestry with additional Anatolian farmer–related ancestry in some groups (*48*). East of the Urals, we detect a *Central Asian Cline,* with *WSHG* individuals at one extreme and Copper Age and Early Bronze Age individuals from Turan at the other.

A distinctive ancestry profile stretching from Eastern Europe to Kazakhstan in the Bronze Age

Beginning around 3000 BCE, the ancestry profiles of many groups in Eurasia were transformed by the spread of Yamnaya Steppe pastoralist–related ancestry (*Western_Steppe_EMBA*) from its source in the *Caucasus Cline* (*9*, *48*) to a vast region stretching from Hungary in the west to the Altai mountains in the east (*7*, *8*) (Fig. 3). Over the next two millennia, this ancestry spread further while admixing with local groups, eventually reaching the Atlantic shores of Europe in the west and South Asia in the southeast. The source of the *Western_Steppe_EMBA* ancestry that eventually reached Central and South Asia was not the initial eastward expansion but instead a secondary expansion that involved a group that had ~67% *Western_Steppe_EMBA* ancestry and ~33% ancestry from a point on the *European Cline* (*8*) (Fig. 3). We replicate previous findings that this group included people of the Corded Ware, Srubnaya, Petrovka, and Sintashta archaeological complexes spreading over a vast region from the border of Eastern Europe to northwestern Kazakhstan (*8*, *18*, *30*), and our dataset adds more than one hundred individuals from this *Western_Steppe_MLBA* cluster. We also detect an additional cluster, *Central_Steppe_MLBA*, which is differentiated from *Western_Steppe_MLBA* ($P = 7 \times 10^{-6}$ by *qpAdm*) because it carries ~9% additional ancestry derived from Bronze Age pastoralists of the Central Steppe of primarily *WSHG*-related ancestry (*Central_Steppe_EMBA*). Thus, individuals with *Western_Steppe_MLBA* ancestry admixed with local populations as they integrated eastward and southward.

Bidirectional mobility along the Inner Asian Mountain Corridor

As in Iran/Turan, the outlier individuals provide crucial information about human interaction.

Our analysis of 50 individuals from the Sintashta culture cemetery of Kamennyi Ambar 5 reveals multiple groups of outliers that we directly radiocarbon dated to be contemporaries of the main cluster but that were also genetically distinctive, indicating that this was a cosmopolitan site (Fig. 2). One set of outliers had elevated proportions of *Central_Steppe_EMBA* (largely *WSHG*-related) ancestry, another had elevated *Western_Steppe_EMBA* (*Yamnaya*-related), and a third had elevated *EEHG*-related ancestry.

In the Central Steppe (present-day Kazakhstan), an individual from one site dated to between 2800 and 2500 BCE, and individuals from three sites dated to between ~1600 and 1500 BCE, show significant admixture from Iranian farmer–related populations that is well-fitted by the main *BMAC* cluster, demonstrating northward gene flow from Turan into the Steppe at approximately the same time as the southward movement of *Central_Steppe_MLBA*-related ancestry through Turan to South Asia. Thus, the archaeologically documented spread of material culture and technology both north and south along the Inner Asian Mountain Corridor (*3*, *49*, *50*, *51*), which began as early as the middle of the third millennium BCE, was associated with substantial movements of people (Fig. 2).

We also observe individuals from Steppe sites (Krasnoyarsk) dated to between ~1700 and 1500 BCE that derive up to ~25% ancestry from a source related to East Asians (well-modeled as *ESHG*), with the remainder best modeled as *Western_Steppe_MLBA*. By the Late Bronze Age, *ESHG*-related admixture became ubiquitous, as documented by our time transect from Kazakhstan and ancient DNA data from the Iron Age and from later periods in Turan and the Central Steppe, including Scythians, Sarmatians, Kushans, and Huns (*29*, *52*). Thus, these first millennium BCE to first millennium CE archaeological cultures with documented cultural and political impacts on South Asia cannot be important sources for the Steppe pastoralist–related ancestry widespread in South Asia today (because present-day South Asians have too little East Asian–related ancestry to be consistent with deriving from these groups), providing an example of how genetic data can rule out scenarios that are plausible on the basis of the archaeological and historical evidence alone (*13*) (fig. S52). Instead, our analysis shows that the only plausible source for the Steppe ancestry is Steppe Middle to Late Bronze Age groups, who not only fit as a source for South Asia but who we also document as having spread into Turan and mixed with *BMAC*-related individuals at sites in Kazakhstan in this period. Taken together, these results identify a narrow time window (first half of the second millennium BCE) when the Steppe ancestry that is widespread today in South Asia must have arrived.

The genomic formation of human populations in South Asia

Three ancestry clines that succeeded each other in time in South Asia

Previous work has shown that South Asians harbor ancestry from peoples related to ancient groups in northern Eurasia and Iran, East Asians, and Australasians (*9*). Here we document the process through which these deep sources of ancestry mixed to form later groups.

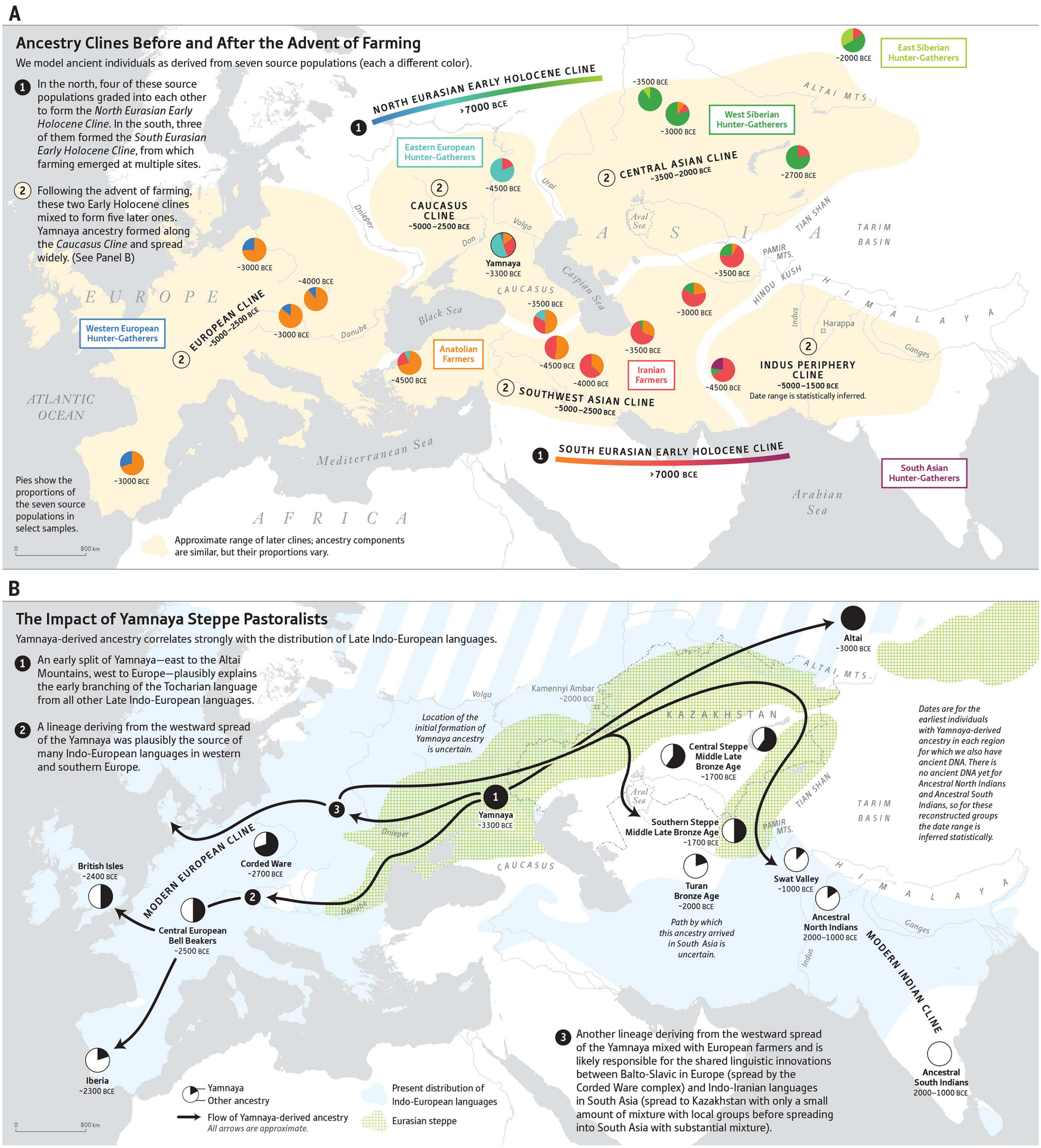
We begin with the pre-2000-BCE *Indus Periphery Cline,* described in an earlier section and detected in 11 outliers from two sites in cultural contact with the IVC (Fig. 4). We can jointly model all individuals in this cline as a mixture of two source populations: One end of the cline is consistent with being entirely *AHG*-related, and the other is consistent with being ~90% Iranian farmer–related and ~10% *WSHG*-related (Fig. 4) (*13*). People fitting on the *Indus Periphery Cline* constitute the majority of the ancestors of present-day South Asians. Through formal modeling, we demonstrate that it is this contribution of *Indus Periphery Cline* people to later South Asians, rather than westward gene flow bringing an ancestry unique to South Asia onto the Iranian plateau, that explains the high degree of shared ancestry between present-day South Asians and early Holocene Iranians (*9*, *13*).

We next characterized the post-2000-BCE *Steppe Cline*, represented in our analysis by 117 individuals dating to between 1400 BCE and 1700 CE from the Swat and Chitral districts of northernmost South Asia (Figs. 2 and 4). We found that we could jointly model all individuals on the *Steppe Cline* as a mixture of two sources, albeit different from the two sources in the earlier cline. One end is consistent with a point along the *Indus Periphery Cline*. The other end is consistent with a mixture of ~41% *Central_Steppe_MLBA* ancestry and ~59% from a subgroup of the *Indus Periphery Cline* with relatively high Iranian farmer–related ancestry (*13*) (fig. S50).

To understand the formation of the *Modern Indian Cline*, we searched for triples of populations that could fit as sources for diverse present-day South Asian groups as well as peoples of the *Steppe Cline*. All fitting models include as sources *Central_Steppe_MLBA* (or a group with a similar ancestry profile), a group of *Indus Periphery Cline* individuals, and either *AHG* or a subgroup of *Indus Periphery Cline* individuals with relatively high *AHG*-related ancestry (*13*) (fig. S51). Coanalyzing 140 diverse South Asian groups (*10*) that fall on a gradient in PCA (*13*), we show that while there are three deep sources, just as in the case of the earlier two clines the great majority of groups on the *Modern Indian Cline* can be jointly modeled as a mixture of two populations that are mixed from the earlier three. Although we do not have ancient DNA data from either of the two statistically reconstructed source populations for the *Modern Indian Cline,* the *ASI* or the *ANI*, in what follows, we coanalyze our ancient DNA data in conjunction with modern data to characterize the exact ancestry of the *ASI* and to provide constraints on the *ANI*.

The *ASI* and *ANI* arose as *Indus Periphery Cline* people mixed with groups to the north and east

To gain insight into the formation of the *ASI,* we extrapolated to the smallest possible proportion of West Eurasian–related ancestry on the *Modern Indian Cline* by setting the *Central_Steppe_MLBA* ancestry proportion to zero in our model. We estimate a minimum of ~55% ancestry from people



A

Ancestry Clines Before and After the Advent of Farming

We model ancient individuals as derived from seven source populations (each a different color).

① In the north, four of these source populations graded into each other to form the *North Eurasian Early Holocene Cline*. In the south, three of them formed the *South Eurasian Early Holocene Cline*, from which farming emerged at multiple sites.

② Following the advent of farming, these two Early Holocene clines mixed to form five later ones. Yamnaya ancestry formed along the *Caucasus Cline* and spread widely. (See Panel B)

Pies show the proportions of the seven source populations in select samples.

Approximate range of later clines; ancestry components are similar, but their proportions vary.

B

The Impact of Yamnaya Steppe Pastoralists

Yamnaya-derived ancestry correlates strongly with the distribution of Late Indo-European languages.

① An early split of Yamnaya—east to the Altai Mountains, west to Europe—plausibly explains the early branching of the Tocharian language from all other Late Indo-European languages.

② A lineage deriving from the westward spread of the Yamnaya was plausibly the source of many Indo-European languages in western and southern Europe.

③ Another lineage deriving from the westward spread of the Yamnaya mixed with European farmers and is likely responsible for the shared linguistic innovations between Balto-Slavic in Europe (spread by the Corded Ware complex) and Indo-Iranian languages in South Asia (spread to Kazakhstan with only a small amount of mixture with local groups before spreading into South Asia with substantial mixture).

Location of the initial formation of Yamnaya ancestry is uncertain.

Path by which this ancestry arrived in South Asia is uncertain.

Dates are for the earliest individuals with Yamnaya-derived ancestry in each region for which we also have ancient DNA. There is no ancient DNA yet for Ancestral North Indians and Ancestral South Indians, so for these reconstructed groups the date range is inferred statistically.

◐ Yamnaya / Other ancestry

Flow of Yamnaya-derived ancestry *All arrows are approximate.*

Present distribution of Indo-European languages

Eurasian steppe

Fig. 3. Ancestry transformations in Holocene Eurasia. (**A**) Ancestry clines before and after the advent of farming. We document a *South Eurasian Early Holocene Cline* of increasing Iranian farmer– and West Siberian hunter-gatherer–related ancestry moving west-to-east from Anatolia to Iran, as well as a *North Eurasian Early Holocene Cline* of increasing relatedness to East Asians moving west-to-east from Europe to Siberia. Mixtures of peoples along these two clines following the spread of farming formed five later gradients (shaded): moving west-to-east: the *European Cline*, the *Caucasus Cline* from which the Yamnaya formed, the *Central Asian Cline* that characterized much of Central Asia in the Copper and Bronze Ages, the *Southwest Asian Cline* established by spreads of farmers in multiple directions from several loci of domestication, and the *Indus Periphery Cline*. (**B**) Following the appearance of the Yamnaya Steppe pastoralists, *Western_Steppe_EMBA* (*Yamnaya*-like) ancestry then spread across this vast region. We use arrows to show plausible directions of spread of increasingly diluted ancestry (the arrows are not meant as exact routes, which we do not have enough sampling to determine at present). Rough estimates of the timing of the arrival of this ancestry and estimated ancestry proportions are shown.

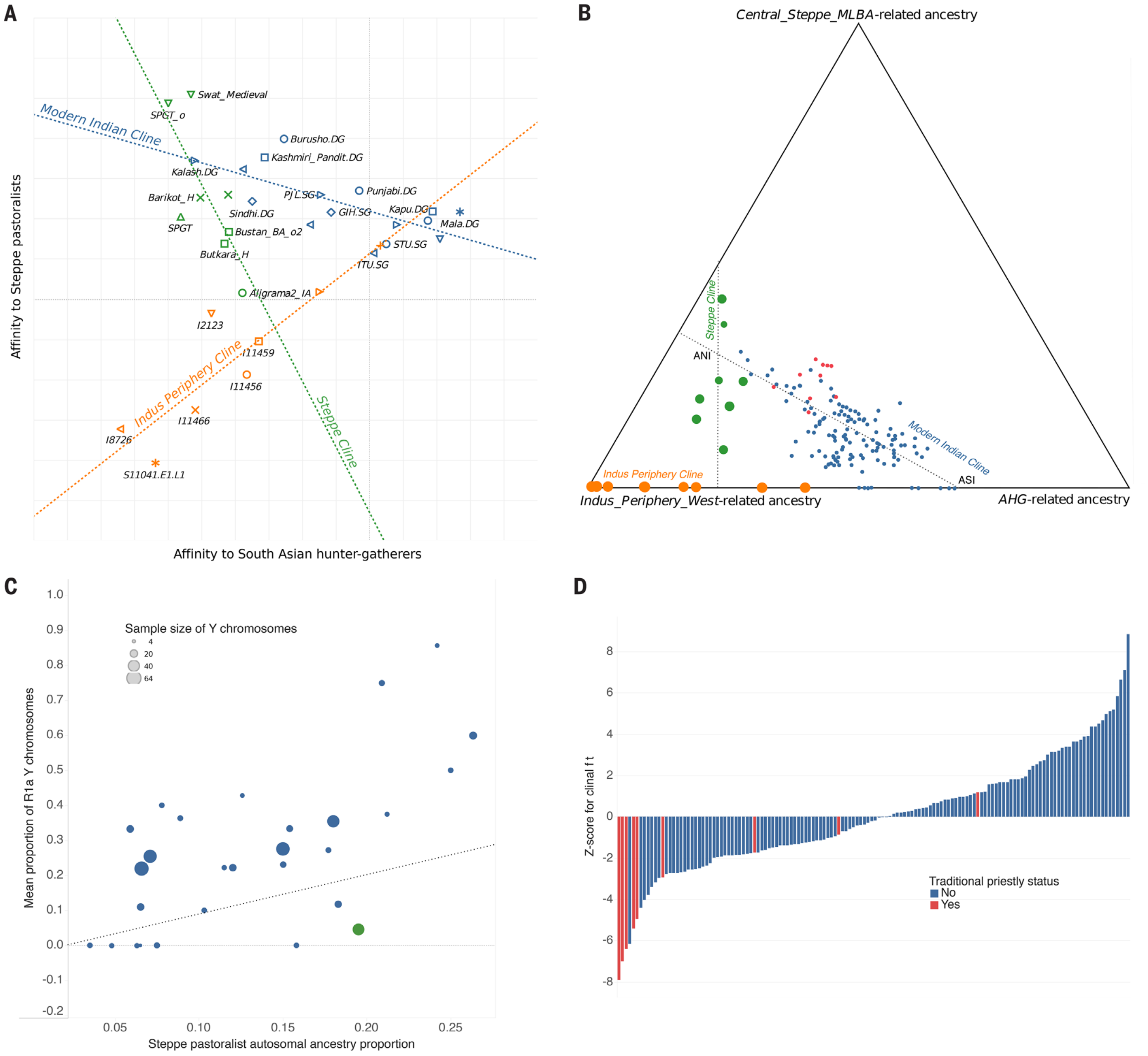


Fig. 4. The genomic formation of South Asia. (**A**) The degree of allele sharing with southern Asian hunter-gatherers (*AASI*) measured by f_4(*Ethiopia_4500BP, X; Ganj_Dareh_N, AHG*) and with Steppe pastoralists measured by f_4(*Ethiopia_4500BP, X; Ganj_Dareh_N, Central_Steppe_MLBA*) reveals three ancestry clines that succeeded each other in time: the *Indus Periphery Cline* before ~2000 BCE, the *Steppe Cline* represented by northern South Asian individuals after ~2000 BCE, and the *Modern Indian Cline.* (**B**) Modeling South Asians as a mixture of *Central_Steppe_MLBA, AHG* (as a proxy for *AASI*) and *Indus_Periphery_West* (the individual from the *Indus Periphery Cline* with the least *AASI* ancestry). Groups along the edges of the triangle fit a two-way model, and in the interior only fit a three-way model. The 140 present-day South Asian groups on the *Modern Indian Cline* are shown as small dots. (**C**) Plot of the proportion of *Central_Steppe_MLBA* ancestry on the autosomes (*x* axis) and the Y chromosome (*y* axis) shows that the source of this ancestry is primarily from females in Late Bronze Age and Iron Age individuals from the Swat District of northernmost South Asia, and primarily from males in most present-day South Asians. (**D**) Groups that traditionally view themselves as being of priestly status (Brahmin, Pandit, and Bhumihar, but excluding Catholic Brahmins) tend to have a significantly higher ratio of *Central_Steppe_MLBA* to *Indus_Periphery_Cline* ancestry than other groups and are labeled in red in this panel and in (B).

on the *Indus Periphery Cline* [representing the *Indus Periphery Cline* by the individual on it with the most Iranian farmer–related ancestry, which we call *Indus_Periphery_West*, and modeling the reminder of the ancestry as deriving from an *AHG*-related group (*13*)]. We find that several tribal groups from southern India are consistent with having no *Central_Steppe_MLBA* ancestry (*13*). The fact that these individuals match the most extreme possible position for the *ASI* reveals that nearly direct descendants of the *ASI* live in South Asia today and furthermore allows us to make a precise statement about the ancestry profile of the *ASI*. In particular, the fact that they harbor substantial Iranian farmer–related ancestry (via the *Indus Periphery Cline*) disproves earlier suggestions that the *ASI* might not have any ancestry related to West Eurasians (*11*). Using the *DATES* software, we estimate an average of 107 ± 11 generations since admixture of the Iranian farmer–related and *AHG*-related groups in one of these groups, *Palliyar*. This corresponds to a 95% confidence interval of 1700 to 400 BCE, assuming 28 years per generation (*47*). Thus, the *ASI* were not fully formed at the time of the IVC and instead must have continued to form through mixture after its decline as material culture typical of the IVC spread eastward (*53*) and *Indus Periphery Cline* ancestry mixed with people of less West Eurasian relatedness.

We also obtained additional evidence for a late (Bronze Age) formation of the *ASI* by building an admixture graph using *qpGraph,* comodeling *Palliyar* and *Juang* (an Austroasiatic-speaking group in India with low West Eurasian relatedness) (Fig. 5). The graph fits the component of South Asian ancestry with no West Eurasian relatedness (Ancestral Ancient South Asians, *AASI*) as an Asian lineage that split off around the time that East Asian, Andaman Islander, and Papuan ancestors separated from each other, consistent with the hypothesis that eastern and southern Asian lineages derive from an eastward spread that in a short span gave rise to lineages leading to *AASI*, East Asians, Andamanese hunter-gatherers, and Papuans (*54*) (Fig. 5). The *Juang* cannot be fit through a mixture of *ASI* ancestry and ancestry related to Austroasiatic language speakers and instead can only be fit by modeling additional ancestry from *AASI*, showing that at the time Austroasiatic groups formed in South Asia, groups with less Iranian farmer–related ancestry than in the *ASI* were also present. Austroasiatic languages are hypothesized to have spread into South Asia in the third millennium BCE [on the basis of hill cultivation systems hypothesized to be associated with the spread of Austroasiatic languages (*41*)], and thus the ancestry profile of the Juang provides an independent line of evidence for a late formation of the *ASI* (in the Bronze Age and plausibly after the decline of the IVC).

To shed light on the formation of the statistically reconstructed *ANI*, we return to the Swat Valley time transect that formed the *Steppe Cline* after 2000 BCE. The *Modern Indian Cline* intersects the *Steppe Cline* at a position close to the

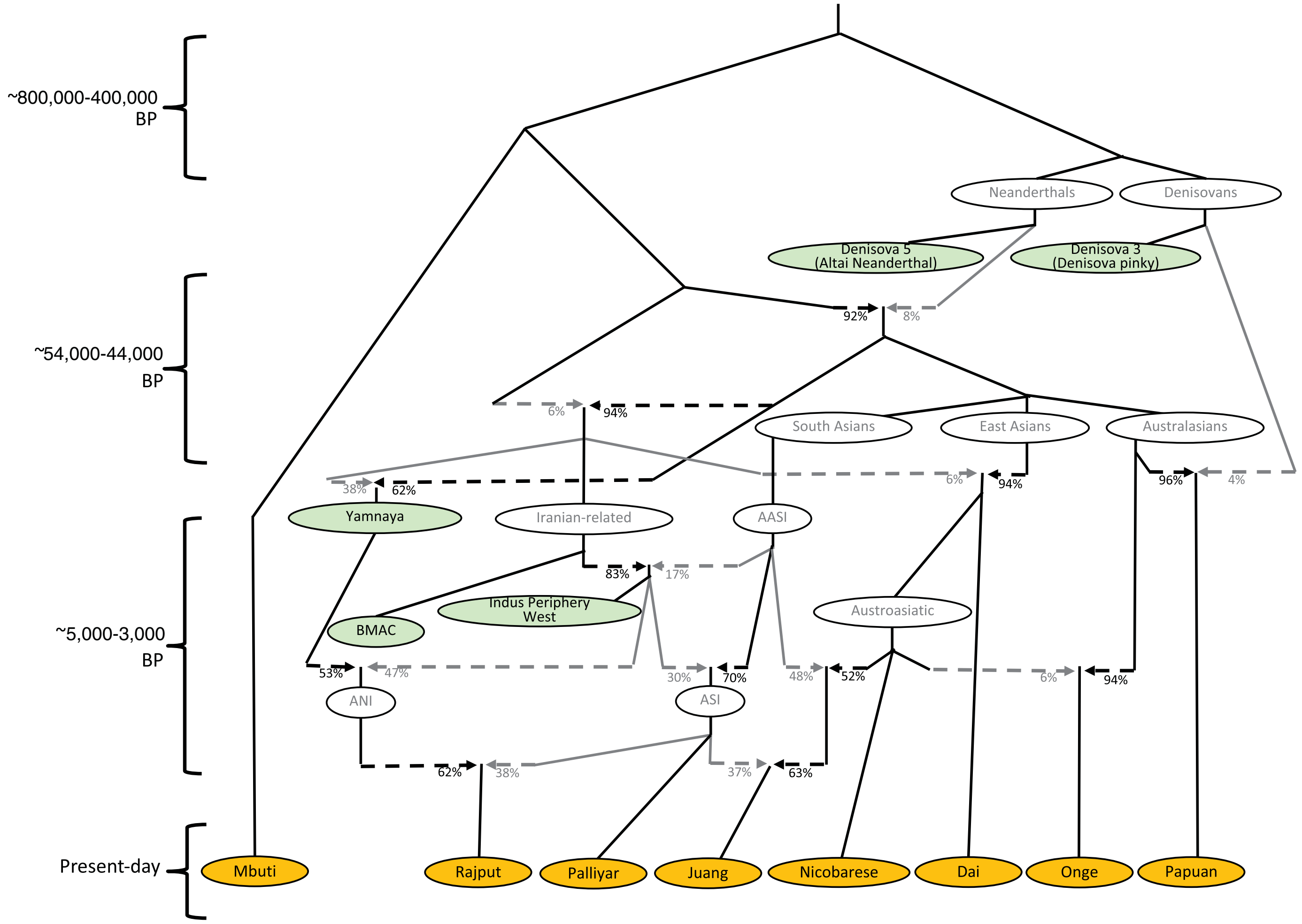


Fig. 5. Admixture graph model. The largest deviation between empirical and theoretical *f*-statistics is |*Z*|= 2.9, indicating a good fit considering the large number of *f*-statistics analyzed. Admixture events are shown as dotted lines labeled by proportions, with the minor ancestry in gray. The present-day groups are shown in orange ovals, the ancient ones in blue, and unsampled groups in white. (The ovals and admixture events are positioned according to guesses about their relative dates to help in visualization, although the dates are in no way meant to be exact.) In this graph, we do not attempt to model the contribution of *WSHG* and Anatolian farmer–related ancestry and thus cannot model *Central_Steppe_EMBA,* the proximal source of Steppe ancestry in South Asia (instead we model the Steppe ancestry in South Asia through the more distally related *Yamnaya*). However, the admixture graph does highlight several key findings of the study, including the deep separation of the *AASI* from other Eurasian lineages and the fact that some Austroasiatic-speaking groups in South Asia (e.g., *Juang*) harbor ancestry from a South Asian group with a higher ratio of *AASI*-related to Iranian farmer–related ancestry than any groups on the *Modern Indian Cline*, thus revealing that groups with substantial Iranian farmer–related ancestry were not ubiquitous in peninsular South Asia in the third millennium BCE when Austroasiatic languages likely spread across the subcontinent.

position of the *Kalash,* the group in northwest South Asia with the highest *ANI* ancestry proportion (55) (Fig. 4). The published estimate of admixture in the *Kalash* is 110 ± 12 generations (55), suggesting a post-IVC date of formation of the *ANI* paralleling the post-IVC date of formation of the *ASI*. Further evidence for a post-IVC integration of Steppe ancestry into South Asia comes from ancient individuals on the *Steppe Cline* (along which the *ANI* could theoretically have formed) whose admixture date for Steppe ancestry is also post-IVC. Specifically, we estimate the date of admixture into the Late Bronze Age and Iron Age individuals from the Swat District of northernmost South Asia to be, on average, 26 generations before the date that they lived, corresponding to a 95% confidence interval of ~1900 to 1500 BCE. This time scale for the arrival of Steppe ancestry in the region is consistent with our observation of six outlier individuals in Turan who lived between ~2000 and 1500 BCE and carry this ancestry in mixed form (Fig. 2) and is also consistent with our finding that the R1a Y chromosome associated with *Central_Steppe_MLBA* ancestry in South Asia is also present in the Swat District Late Bronze and Iron Age individuals (two copies).

Taken together, these results show that neither of the two primary source populations of the *Modern Indian Cline*, the *ANI* or *ASI*, was fully formed before the turn of the second millennium BCE.

Steppe ancestry in modern South Asians is primarily from males and disproportionately high in Brahmin and Bhumihar groups

In the Late Bronze Age and Iron Age individuals of the Swat Valley, we detect a significantly lower proportion of Steppe admixture on the Y chromosome (only 5% of the 44 Y chromosomes of the R1a-Z93 subtype that occurs at 100% frequency in the *Central_Steppe_MLBA* males) compared with ~20% on the autosomes ($Z = -3.9$ for a deficiency from males under the simplifying assumption that all the Y chromosomes are unrelated to each other since admixture and thus are statistically independent), documenting how Steppe ancestry was incorporated into these groups largely through females (Fig. 4). However, sex bias varied in different parts of South Asia, as in present-day South Asians we observe a reverse pattern of excess *Central_Steppe_MLBA*–related ancestry on the Y chromosome compared with the autosomes ($Z = 2.7$ for an excess from males) (13, 56) (Fig. 4). Thus, the introduction of lineages from Steppe pastoralists into the ancestors of present-day South Asians was mediated mostly by males. This bias is similar in direction to what has been documented for the introduction of Steppe ancestry into Iberia in far western Europe, although it is less extreme than the bias reported in that case (57).

Our analysis of Steppe ancestry also identified six groups with a highly elevated ratio of *Central_Steppe_MLBA*– to *Indus_Periphery_West*–related ancestry compared with the expectation for the model at the $Z < -4.5$ level (Fig. 4). The strongest two signals were in *Brahmin_Tiwari* ($Z = -7.9$) and *Bhumihar_Bihar* ($Z = -7.0$). More generally, there is a notable enrichment in groups that consider themselves to be of traditionally priestly status: five of the six groups with $Z < -4.5$ were *Brahmins* or *Bhumihars* even though they make up only 7 to 11% of the 140 groups analyzed ($P < 10^{-12}$ by a χ^2 test, assuming all the groups evolved independently). We caution that this is not a formal test, as there is an unknown degree of shared ancestry among groups since they formed by mixture and because our decisions about which groups to include in the analysis were not made in a blinded way; for example, we excluded four "Catholic Brahmin" groups with strong evidence of substantial shared ancestry in the past millennium (10), which makes them not statistically independent (Fig. 4 and table S5) (13). In addition, the classification of groups as Brahmin may have changed over time, weakening the correlation to genetics. Nevertheless, the fact that traditional custodians of liturgy in Sanskrit (Brahmins) tend to have more Steppe ancestry than is predicted by a simple *ASI-ANI* mixture model provides an independent line of evidence—beyond the distinctive ancestry profile shared between South Asia and Bronze Eastern Europe mirroring the shared features of Indo-Iranian and Balto-Slavic languages (58)—for a Bronze Age Steppe origin for South Asia's Indo-European languages.

Discussion

Our analysis reveals that the ancestry of the greater South Asian region in the Holocene was characterized by at least three genetic gradients. Before ~2000 BCE, there was the *Indus Periphery Cline* consisting of people with different proportions of Iranian farmer– and *AASI*-related ancestry, which we hypothesize was a characteristic feature of many IVC people. The *ASI* formed after 2000 BCE as a mixture of a point along this cline with South Asians with higher proportions of *AASI*-related ancestry. Between ~2000 and 1000 BCE, people of largely *Central_Steppe_MLBA* ancestry expanded toward South Asia, mixing with people along the *Indus Periphery Cline* to form the *Steppe Cline.* Multiple points along the *Steppe Cline* are represented by individuals of the Swat Valley time transect, and statistically we find that the *ANI*, one of the two primary source populations of South Asia, can fit along the *Steppe Cline*. After ~2000 BCE, mixtures of heterogeneous populations—the *ASI* and *ANI*—combined to form the *Modern Indian Cline,* which is represented today in diverse groups in South Asia (Fig. 4).

Our finding, based on the sizes of blocks of ancestry (13) (fig. S59), that the mixture that formed the *Indus Periphery Cline* occurred by ~5400 to 3700 BCE—at least a millennium before the formation of the mature IVC—raises two possibilities. One is that Iranian farmer–related ancestry in this group was characteristic of the Indus Valley hunter-gatherers in the same way as it was characteristic of northern Caucasus and Iranian plateau hunter-gatherers. The presence of such ancestry in hunter-gatherers from Belt and Hotu Caves in northeastern Iran increases the plausibility that this ancestry could have existed in hunter-gatherers farther east. An alternative is that this ancestry reflects movement into South Asia from the Iranian plateau of people accompanying the eastward spread of wheat and barley agriculture and goat and sheep herding as early as the seventh millennium BCE and forming early farmer settlements, such as those at Mehrgarh in the hills flanking the Indus Valley (59, 60). However, this is in tension with the observation that the *Indus Periphery Cline* people had little if any Anatolian farmer–related ancestry, which is strongly correlated with the eastward spread of crop-based agriculture in our dataset. Thus, although our analysis supports the idea that eastward spread of Anatolian farmer–related ancestry was associated with the spread of farming to the Iranian plateau and Turan, our results do not support large-scale eastward movements of ancestry from western Asia into South Asia after ~6000 BCE (the time after which all ancient individuals from Iran in our data have substantial Anatolian farmer–related ancestry, in contrast to South Asians who have very little). Languages in pre-state societies usually spread through movements of people (61), and thus the absence of much Anatolian farmer–related ancestry in the *Indus Periphery Cline* suggests that it is unlikely that the Indo-European languages spoken in South Asia today originate from the spread of farming from West Asia.

Our results not only provide evidence against an Iranian plateau origin for Indo-European languages in South Asia but also evidence for the theory that these languages spread from the Steppe. Although ancient DNA has documented westward movements of Steppe pastoralist ancestry providing a likely conduit for the spread of many Indo-European languages to Europe (7, 8), the chain of transmission into South Asia has been unclear because of a lack of relevant ancient DNA. Our observation of the spread of *Central_Steppe_MLBA* ancestry into South Asia in the first half of the second millennium BCE provides this evidence, which is particularly notable because it provides a plausible genetic explanation for the linguistic similarities between the Balto-Slavic and Indo-Iranian subfamilies of Indo-European languages, which despite their vast geographic separation share the "satem" innovation and "ruki" sound laws (62). If the spread of people from the Steppe in this period was a conduit for the spread of South Asian Indo-European languages, then it is striking that there are so few material culture similarities between the Central Steppe and South Asia in the Middle to Late Bronze Age (i.e., after the middle of the second millennium BCE). Indeed, the material culture differences are so substantial that some archaeologists report no evidence of a connection. However, lack of material culture connections does not provide evidence against spread of genes, as has been demonstrated in the case of the Beaker Complex, which originated largely in western Europe but in Central Europe was associated with skeletons that harbored ~50% ancestry related to Yamnaya Steppe pastoralists (20). Thus, in Europe we have an unambiguous example of people with ancestry from the Steppe making profound demographic impacts on the regions into which they spread while adopting important aspects of local material culture. Our findings document a similar phenomenon in South Asia, with the locally acculturated population harboring up to ~20% *Western_Steppe_EMBA*–derived ancestry according to our modeling (via the up to ~30% ancestry contributed by *Central_Steppe_MLBA* groups) (Fig. 3). Our analysis also provides a second line of evidence for a linkage between Steppe ancestry and Indo-European languages. Steppe ancestry enrichment in groups that view themselves as being of traditionally priestly status is notable, as some of these groups, including Brahmins, are traditional custodians of literature composed in early Sanskrit. A possible explanation is that the influx of *Central_Steppe_MLBA* ancestry into South Asia in the middle of the second millennium BCE created a metapopulation with varied proportions of Steppe ancestry, with people of greater Steppe ancestry (or admixing less with *Indus Periphery Cline* groups) tending to be more strongly associated with Indo-European culture. Because of strong endogamy, which kept groups generally isolated from neighbors for thousands of years (7), some of this population substructure persists in South Asia among present-day custodians of Indo-European texts.

Our findings also shed light on the origin of the second-largest language group in South Asia, Dravidian. The strong correlation between *ASI* ancestry and present-day Dravidian languages suggests that the *ASI*, which we have shown formed as groups with ancestry typical of the *Indus Periphery Cline* moved south and east after the decline of the IVC to mix with groups with more *AASI* ancestry, most likely spoke an early Dravidian language. A possible scenario combining genetic data with archaeology and linguistics is that proto-Dravidian was spread by peoples of the IVC along with the *Indus Periphery Cline* ancestry component of the *ASI*. Nongenetic support for an IVC origin of Dravidian languages includes the present-day geographic distribution of these languages (in southern India and southwestern Pakistan) and a suggestion that some symbols on ancient Indus Valley seals denote Dravidian words or names (63, 64). An alternative possibility is that proto-Dravidian was spread by the half of the *ASI*'s ancestry that was not from the *Indus Periphery Cline* and instead derived from the south and the east (peninsular South Asia). The southern scenario is consistent with reconstructions of Proto-Dravidian terms for flora and fauna unique to peninsular India (65, 66).

Finally, we highlight a remarkable parallel between the prehistory of South Asia and Europe. In both subcontinents of Eurasia, there were exchanges between people related to Southwest Asians and peninsular hunter-gatherers; mixtures of these groups led to the *Indus Periphery Cline* in South Asia and the *European Cline* in Europe. In both subcontinents, people arriving in the second and third millennia BCE who descended from mixtures of people related to Yamnaya Steppe pastoralists and European farmers mixed further with local populations: in South Asia forming the *ANI* and in Europe forming groups like that of the Beaker Complex. In both cases, mixtures of these heterogeneous populations—those with Steppe pastoralist–related admixture and those without—drive the modern ancestry clines in both regions (Fig. 3). However, there are also profound differences between the Bronze Age and Neolithic spreads of ancestry across the two subcontinents. One is that the maximum proportion of peninsular hunter-gatherer ancestry is higher in South Asia (*AASI* ancestry of up to ~60%) than Europe (*WEHG* ancestry of up to ~30%) (7), which could reflect stronger ecological or cultural barriers to the spread of people in South Asia than in Europe, allowing the previously established groups more time to adapt and mix with incoming groups. A second difference is the smaller proportion of Steppe pastoralist–related ancestry in South Asia compared with Europe, its later arrival by ~500 to 1000 years, and a lower (albeit still significant) male sex bias in the admixture, factors that help to explain the continued persistence of a large fraction of non–Indo-European speakers amongst people of present-day South Asia today. The situation in South Asia is somewhat reminiscent of Mediterranean Europe, where the proportion of Steppe ancestry is considerably lower than that of Northern and Central Europe (Fig. 3) and where many non–Indo-European languages are attested in classical times (67). Further studies of ancient DNA from South Asia and the linguistically related Iranian world will extend and add nuance to the model presented here.

Materials and methods

Ancient DNA laboratory work

For the skeletal elements that we were not able to transport from field sites, we drilled directly into bone, for the most part focusing on inner ear portions of petrous bones using a method for sampling from the cranial base (CBD) (68). The great majority of skeletal elements were prepared in dedicated ancient DNA clean rooms at Harvard Medical School, University College Dublin, the University of Vienna, or the Max Planck Institute for Evolutionary Anthropology in Leipzig either by drilling, or by sandblasting to isolate a bone piece followed by milling (tables S1 and S2).

All the molecular work except for that on a single individual (Darra-i-Kur) was carried out at Harvard Medical School. We extracted DNA using a method that is optimized to retain small DNA fragments. We implemented this method either using a manual method based on silica spin columns (565 libraries) (14, 15), or with the assistance of robotic liquid handlers using silica coated magnetic beads and Buffer D (149 libraries) (69). We converted the DNA into a form that could be sequenced using a double-stranded library preparation protocol (711 libraries) (17) and a single stranded library preparation protocol (3 libraries) (70). For all but four of the double stranded libraries, we pre-treated with a mixture of the enzymes Uracil-DNA Glycosylase (UDG) and Endo VIII (USER, New England Biolabs) to greatly reduce the cytosine-to-thymine damage characteristic of ancient DNA sequences while retaining damage in both terminal bases (17). The remaining four libraries were not pre-treated with USER (71). The three single-stranded libraries were also pre-treated with USER in a way that results in a similar damage pattern (70). We prepared most double stranded libraries (n = 524) with the assistance of a robotic liquid handler, substituting the MinElute columns used for cleaning up reactions in manual processing with silica coated magnetic beads in robotic processing, and the MinElute column-based PCR cleanup at the end of library preparation with SPRI beads (72, 73). We enriched all libraries both for sequences overlapping mitochondrial DNA (74), and for sequences overlapping about 1.2 million nuclear targets (7, 18, 19) (table s2). After indexing the enrichment products in a way that assigned a unique index combination to each library (75), we sequenced the enriched products on an Illumina NextSeq500 instrument using v.2 150 cycle kits for 2 × 76 cycles and 2 × 7 cycles (2 × 8 for single-stranded libraries), and sequenced up to the point that the expected number of additional SNPs covered per 100 additional read pairs sequenced was less than about 1. We also shotgun-sequenced libraries to assess the fraction of sequences that mapped to the human genome.

To analyze the data, we began by sorting the read pairs by searching for the expected identification indices and barcodes for each library, allowing up to one mismatch from the expected sequence in each case. We removed adapters and merged together sequences requiring a 15 base pair overlap (allowing up to one mismatch), taking the highest quality base in the merged segment to represent the allele. We mapped the resulting sequences to the *hg19* human reference [GRCh37, the version used for the 1000 Genomes project (76)] using the *samse* command of BWA (77) (version 0.6.1). We removed duplicate sequences (mapping to the same position in the genome and having the same barcode pair), and merged libraries corresponding to the same sample (merging across samples that the genetic data revealed were from the same individual). For each individual, we restricted to sequences passing filters (not overlapping known insertion/deletion polymorphisms, and having a minimum mapping quality 10), and trimmed two nucleotides from the end of each sequence to reduce deamination artifacts. We also further restricted to sequence data with a minimum base quality of 20. To represent each individual at each SNP position, we randomly selected a single sequence (if at least one was available).

For Darra-i-Kur, we analyzed a single-stranded DNA library (L5082) at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, generated as part of a previous study (78). The previous study only analyzed mitochondrial DNA, and for the current study, we enriched the library for sequences overlapping the same panel of about 1.2 million nuclear targets using two rounds of hybridization capture (*7*, *18*, *19*). We sequenced the enriched libraries on two lanes of an Illumina HiSeq2500 platform in a double index configuration (2x76 cycles) (*75*), and we determined alleles using *FreeIbis* (*79*). We merged overlapping paired-end and trimmed using *leeHom* (*80*). We used BWA to align the sequences to the human reference genome *hg19* (GRCh37) (*77*). We retained sequences showing a perfect match to the expected index combination for downstream analyses.

We assessed evidence for ancient DNA authenticity by measuring the rate of damage in the first nucleotide, flagging individuals as potentially contaminated if they had less than a 3% cytosine-to-thymine substitution rate in the first nucleotide for a UDG-treated library and less than a 10% substitution rate for a non-UDG-treated library. We used *contamMix* to test for contamination based on polymorphism in mitochondrial DNA (*81*), and ANGSD to test for contamination based on polymorphism on the X chromosome in males (*82*).

Radiocarbon dating

We generated 269 radiocarbon (^{14}C) dates on bone using accelerator mass spectrometry (AMS) (table S3). Most of these (*n* = 242) were generated at the Pennsylvania State University (PSU) Radiocarbon Laboratory, and here we excerpt a description of the sample preparation methodology at PSU (the methods used at the other laboratories are publicly available and we refer readers to the literature for those methodologies). Possible contaminants (conservants and adhesives) were removed by sonicating all bone samples in successive washes of ACS grade methanol, acetone, and dichloromethane for 30 min each at room temperature, followed by three washes in Nanopure water to rinse. Bone collagen for ^{14}C was extracted and purified using a modified Longin method with ultrafiltration [>30 kDa gelatin; (*83*)]. If collagen yields were low and amino acids poorly preserved we used a modified XAD process [XAD Amino Acids; (*84*)]. For quality assurance we measured carbon and nitrogen concentrations and C/N ratios of all extracted and purified collagen/amino acid samples with a Costech elemental analyzer (ECS 4010). We evaluated quality based on % crude gelatin yield, %C, %N and C/N ratios before AMS ^{14}C dating. C/N ratios for all directly radiocarbon dated samples fell between 2.9 and 3.4, indicating excellent preservation (*85*). Collagen/amino acid samples (~2.1 mg) were then combusted for 3 hours at 900°C in vacuum-sealed quartz tubes with CuO and Ag wires. Sample CO$_2$ was reduced to graphite at 550°C using H$_2$ and a Fe catalyst, with reaction water drawn off with Mg(ClO$_4$)$_2$ (*86*). All ^{14}C measurements were made on a modified National Electronics Corporation compact spectrometer with a 0.5 MV accelerator (NEC 1.5SDH-1). The ^{14}C ages were corrected for mass-dependent fractionation with measured δ^{13}C values (*87*) and compared with samples of Pleistocene whale bone (backgrounds, 48,000 cal BP), late Holocene bison bone (~1850 cal BP), late 1800s CE cow bone, and OX-2 oxalic acid standards. All calibrated ^{14}C ages were calculated using OxCal version 4.3 (Ramsey and Lee 2013) using the IntCal13 northern hemisphere curve (*88*), and we quote 95% confidence intervals (2-sigma ranges).

Principal components analysis (PCA)

We carried out PCA using the *smartpca* package of *EIGENSOFT* 7.2.1 (*35*). We used default parameters and added two options (lsqproject:YES and numoutlieriter:0) to project the ancient individuals onto the PCA space. We used two basis sets for the projection: the first based on 1340 present-day Eurasians genotyped on the Affymetrix Human Origins array, and the second based on a subset of 991 present-day West Eurasians (*7*, *27*, *32*). These projections are shown repeatedly in (*13*) and are used in the Online Data Visualizer. We also computed F_{ST} between groups using the parameters inbreed:YES and fstonly: YES. We restricted these analyses to the dataset obtained by merging our ancient DNA data with the modern DNA data on the Human Origins array and restricting to 597,573 SNPs. We treated positions where we did not have sequence data as missing genotypes.

ADMIXTURE clustering

Using PLINK2 (*89*), we first pruned our dataset using the –geno 0.7 option to ensure that we only performed our analysis on sites where at least 70% of individuals were covered by at least one sequence. This resulted in 892,613 SNPs. Individuals without coverage on specific SNPs were assigned missing data at those sites. We ran ADMIXTURE (*36*) with 10 replicates, reporting the replicate with the highest likelihood. We show results for K = 5 in (*13*), as we found that this provides good resolution for disambiguating the sources of pre-Copper Age ancestry in the ancient individuals.

f-statistics

We used the *qp3pop* and *qpDstat* packages in ADMIXTOOLS to compute f_3-statistics and f_4-statistics. We used the inbreed:YES parameter to compute f_3-statistics as a test for admixture with an ancient population as a target, with all ancient genomes as sources. Using the f4Mode: YES parameter in *qpDstat,* we also computed two sets of f_4-symmetry statistics to evaluate if pairs of populations are consistent with forming a clade relative to a comparison population. The first is a "Two-population comparison" statistic where we compare all possible pairs of ancient groups (the *Test* populations) to a panel of populations that encompasses diverse pre-Copper Age and more widespread genetic variation. Thus, we compute a statistic of the form f_4(*Test 1, Test 2; Pre-Copper Age, Mbuti*). The second is a "Pre-Copper Age affinity" statistic that compares each ancient group in turn against diverse pairs of Pre-Copper Age populations, using statistics of the form f_4(*Pre-Copper Age 1*, *Pre-Copper Age 2*; *Test*, *Mbuti*).

Modeling admixture history

We used *qpAdm* (*32*) in the ADMIXTOOLS software package to estimate the proportions of ancestry in a *Test* population deriving from a mixture of *N* "reference" populations by leveraging (but not explicitly modeling) shared genetic drift with a set of "Outgroup" populations. We set the details:YES parameter, which reports a normally distributed Z-score for the goodness of fit of the model (estimated with a Block Jackknife).

Hierarchical modeling

For each group on a proposed cline, we used *qpAdm* to obtain estimates for the proportion of ancestry from hypothesized source populations, along with the covariance matrix across groups. We jointly modeled these estimates using a bivariate normal model (forcing the three proportions to sum to 100%) and estimated the mean and covariance of the two parameters using maximum likelihood. With this inferred matrix, we tested whether the cline could be modeled by a mixture of two primary source populations. First, we tested if the covariance matrix is consistent with being singular, implying that knowledge of the proportion of ancestry from one of the mixing components was consistent with being fully predictive of the other two, as expected for two-way mixture. Second, if we were able to establish that this was the case, we examined the difference between the expected and observed ratios of the ancestry proportions of the analyzed groups within this generative model by fitting all the groups simultaneously. This resulted in a handful of groups deviating from expectation.

Method for dating admixture events

To understand the time scale 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 (*38*) for dating admixture in present-day individuals, and DATES (Distribution of Ancestry Tracts of Evolutionary Signals, a new method we introduce here) for ancient individuals. DATES leverages ancestry covariance patterns that can be measured in a single individual (instead of admixture LD that requires multiple individuals). Full details of the approach and simulations documenting its efficacy in modern as well as ancient data are presented in (*13*). The software implementing DATES is available at Zenodo (*90*).

REFERENCES AND NOTES

1. V. M. Narasimhan, The formation of human populations in South and Central Asia, Online Data Visualizer; https://public.tableau.com/profile/vagheesh#!/vizhome/TheFormationofHumanPopulationsinSouthandCentralAsia/AncientDNA.
2. D. Q. Fuller, L. Lucas, in *Human Dispersal and Species Movement*, N. Boivin, R. Crassard, M. Petraglia, Eds. (Cambridge Univ. Press, 2017), pp. 304–331.
3. C. J. Stevens *et al*., Between China and South Asia: A Middle Asian corridor of crop dispersal and agricultural innovation in the Bronze Age. *Holocene* **26**, 1541–1555 (2016). doi: 10.1177/0959683616650268; pmid: 27942165

4. R. G. Allaby, C. Stevens, L. Lucas, O. Maeda, D. Q. Fuller, Geographic mosaics and changing rates of cereal domestication. *Philos. Trans. R. Soc. London Ser. B* **372**, 20160429 (2017). doi: 10.1098/rstb.2016.0429; pmid: 29061901
5. A. H. Dani *et al*., *History of Civilizations of Central Asia: The Development of Sedentary and Nomadic Civilizations, 700 B.C. to A.D. 250*, J. Harmatta, B. N. Puri, G. F. Etemadi, Eds. (UNESCO Publishing, 1994).
6. M. D. Frachetti, C. E. Smith, C. M. Traub, T. Williams, Nomadic ecology shaped the highland geography of Asia's Silk Roads. *Nature* **543**, 193–198 (2017). doi: 10.1038/nature21696; pmid: 28277506
7. W. Haak *et al*., Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015). doi: 10.1038/nature14317; pmid: 25731166
8. M. E. Allentoft *et al*., Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015). doi: 10.1038/nature14507; pmid: 26062507
9. I. Lazaridis *et al*., Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016). doi: 10.1038/nature19310; pmid: 27459054
10. N. Nakatsuka *et al*., The promise of discovering population-specific disease-associated genes in South Asia. *Nat. Genet.* **49**, 1403–1407 (2017). doi: 10.1038/ng.3917; pmid: 28714977
11. D. Reich, K. Thangaraj, N. Patterson, A. L. Price, L. Singh, Reconstructing Indian population history. *Nature* **461**, 489–494 (2009). doi: 10.1038/nature08365; pmid: 19779445
12. P. Moorjani *et al*., Genetic evidence for recent population mixture in India. *Am. J. Hum. Genet.* **93**, 422–438 (2013). doi: 10.1016/j.ajhg.2013.07.006; pmid: 23932107
13. See supplementary materials.
14. J. Dabney *et al*., Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15758–15763 (2013). doi: 10.1073/pnas.1314445110; pmid: 24019490
15. P. Korlević *et al*., Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015). doi: 10.2144/000114320; pmid: 26260087
16. M. Meyer *et al*., A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012). doi: 10.1126/science.1224344; pmid: 22936568
17. N. Rohland, E. Harney, S. Mallick, S. Nordenfelt, D. Reich, Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. London Ser. B* **370**, 20130624 (2015). doi: 10.1098/rstb.2013.0624; pmid: 25487342
18. I. Mathieson *et al*., Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015). doi: 10.1038/nature16152; pmid: 26595274
19. Q. Fu *et al*., The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016). doi: 10.1038/nature17993; pmid: 27135931
20. I. Olalde *et al*., The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**, 190–196 (2018). doi: 10.1038/nature25738; pmid: 29466337
21. M. A. Yang *et al*., 40,000-year-old individual from Asia provides insight into early population structure in Eurasia. *Curr. Biol.* **27**, 3202–3208.e9 (2017). doi: 10.1016/j.cub.2017.09.030; pmid: 29033327
22. L. Saag *et al*., Extensive farming in Estonia started through a sex-biased migration from the Steppe. *Curr. Biol.* **27**, 2185–2193.e6 (2017). doi: 10.1016/j.cub.2017.06.022; pmid: 28712569
23. M. Raghavan *et al*., Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014). doi: 10.1038/nature12736; pmid: 24256729
24. A. Mittnik *et al*., The genetic prehistory of the Baltic Sea region. *Nat. Commun.* **9**, 442 (2018). doi: 10.1038/s41467-018-02825-9; pmid: 29382937
25. I. Mathieson *et al*., The genomic history of southeastern Europe. *Nature* **555**, 197–203 (2018). doi: 10.1038/nature25778; pmid: 29466330
26. M. Lipson *et al*., Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* **551**, 368–372 (2017). doi: 10.1038/nature24476; pmid: 29144465
27. I. Lazaridis *et al*., Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014). doi: 10.1038/nature13673; pmid: 25230663
28. M. Gallego-Llorente *et al*., The genetics of an early Neolithic pastoralist from the Zagros, Iran. *Sci. Rep.* **6**, 31326 (2016). doi: 10.1038/srep31326; pmid: 27502179
29. P. B. Damgaard *et al*., 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, 369–374 (2018). doi: 10.1038/s41586-018-0094-2; pmid: 29743675
30. P. de Barros Damgaard *et al*., The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* **360**, eaar7711 (2018). doi: 10.1126/science.aar7711; pmid: 29743352
31. F. Broushaki *et al*., Early Neolithic genomes from the eastern Fertile Crescent. *Science* **353**, 499–503 (2016). doi: 10.1126/science.aaf7943; pmid: 27417496
32. N. Patterson *et al*., Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012). doi: 10.1534/genetics.112.145037; pmid: 22960212
33. P. Skoglund *et al*., Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336**, 466–469 (2012). doi: 10.1126/science.1216304; pmid: 22539720
34. N. Patterson, A. L. Price, D. Reich, Population structure and eigenanalysis. *PLOS Genet.* **2**, e190 (2006). doi: 10.1371/journal.pgen.0020190; pmid: 17194218
35. K. J. Galinsky *et al*., Fast principal-component analysis reveals convergent evolution of ADH1B in Europe and East Asia. *Am. J. Hum. Genet.* **98**, 456–472 (2016). doi: 10.1016/j.ajhg.2015.12.022; pmid: 26924531
36. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009). doi: 10.1101/gr.094052.109; pmid: 19648217
37. J. McNeill, K. Pomeranz, Eds., *The Cambridge World History* (Cambridge Univ. Press, 2015).
38. P.-R. Loh *et al*., Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **193**, 1233–1254 (2013). doi: 10.1534/genetics.112.147330; pmid: 23410830
39. A. G. Hinch *et al*., The landscape of recombination in African Americans. *Nature* **476**, 170–175 (2011). doi: 10.1038/nature10336; pmid: 21775986
40. M. D. Petraglia, B. Allchin, in *The Evolution and History of Human Populations in South Asia* (Springer Netherlands, 2007), pp. 1–20.
41. P. S. Bellwood, C. Renfrew, Eds. *Examining the Farming/Language Dispersal Hypothesis* (McDonald Institute for Archaeological Research, University of Cambridge, 2002).
42. K. G. Daly *et al*., Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science* **361**, 85–88 (2018). doi: 10.1126/science.aas9411; pmid: 29976826
43. A. J. Ammerman, L. L. Cavalli-Sforza, *The Neolithic Transition and the Genetics of Populations in Europe* (Princeton Univ. Press, 1984).
44. L. Dupree, *Notes on Shortugai: An Harappan site in northern Afghanistan* (Centre for the Study of the Civilization of Central Asia, Quaid-i-Azam University, 1981).
45. C. Minniti, S. M. S. Sajjadi, New data on non-human primates from the ancient Near East: The recent discovery of a rhesus macaque burial at Shahr-i Sokhta (Iran). *Int. J. Osteoarchaeol.* oa.2750 (2019). doi: 10.1002/oa.2750
46. M. Vidale, A "Priest King" at Shahr-i Sokhta? *Archaeol. Res. Asia* **15**, 110–115 (2018). doi: 10.1016/j.ara.2017.12.001
47. P. Moorjani *et al*., A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5652–5657 (2016). doi: 10.1073/pnas.1514696113; pmid: 27140627
48. C.-C. Wang *et al*., Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. *Nat. Commun.* **10**, 590 (2019). doi: 10.1038/s41467-018-08220-8; pmid: 30713341
49. M. D. Frachetti, Multiregional emergence of mobile pastoralism and nonuniform institutional complexity across Eurasia. *Curr. Anthropol.* **53**, 2–38 (2012). doi: 10.1086/663692
50. M. D. Frachetti, *Pastoralist Landscapes and Social Interaction in Bronze Age Eurasia* (University of California Press, 2008).
51. T. R. Hermes *et al*., Early integration of pastoralism and millet cultivation in Bronze Age Eurasia. *Proc. R. Soc. B* 20191273 (2019). doi: 10.1098/rspb.2019.1273
52. M. Unterländer *et al*., Ancestry and demography and descendants of Iron Age nomads of the Eurasian Steppe. *Nat. Commun.* **8**, 14615 (2017). doi: 10.1038/ncomms14615; pmid: 28256537
53. L. Giosan *et al*., Fluvial landscapes of the Harappan civilization. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E1688–E1694 (2012). doi: 10.1073/pnas.1112743109; pmid: 22645375
54. S. Mallick *et al*., The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016). doi: 10.1038/nature18964; pmid: 27654912
55. G. Hellenthal *et al*., The Kalash genetic isolate? The evidence for recent admixture. *Am. J. Hum. Genet.* **98**, 396–397 (2016). doi: 10.1016/j.ajhg.2015.12.025; pmid: 26849116
56. M. Silva *et al*., A genetic chronology for the Indian Subcontinent points to heavily sex-biased dispersals. *BMC Evol. Biol.* **17**, 88 (2017). doi: 10.1186/s12862-017-0936-9; pmid: 28335724
57. I. Olalde *et al*., The genomic history of the Iberian Peninsula over the past 8000 years. *Science* **363**, 1230–1234 (2019). doi: 10.1126/science.aav4040; pmid: 30872528
58. D. Ringe, T. Warnow, A. Taylor, Indo-European and Computational Cladistics. *Trans. Philol. Soc.* **100**, 59–129 (2002). doi: 10.1111/1467-968X.00091
59. D. L. Lister *et al*., Barley heads east: Genetic analyses reveal routes of spread through diverse Eurasian landscapes. *PLOS ONE* **13**, e0196652 (2018). doi: 10.1371/journal.pone.0196652; pmid: 30020920
60. L. Costantini, The first farmers in Western Pakistan: The evidence of the Neolithic agropastoral settlement of Mehrgarh. *Pragdhara* **18**, 167–178 (2008).
61. P. Bellwood, in *The Encyclopedia of Global Human Migration* (Blackwell, 2013).
62. B. W. Fortson, *Indo-European Language and Culture: An Introduction* (Wiley, 2011).
63. A. Parpola, *The Roots of Hinduism: The Early Aryans and the Indus Civilization* (Oxford Univ. Press, 2015).
64. I. Mahadevan, M. V. Bhaskar, in *Walking with the Unicorn: Social Organization and Material Culture in Ancient South Asia*, D. Frenez, G. Jamison, R. Law, M. Vidale, R. H. Meadow, Eds. (Archeopress, 2018), pp. 359–376.
65. F. Southworth, in *7th ESCA Harvard-Kyoto Roundtable* (2005).
66. B. Krishnamurti, *The Dravidian Languages* (Cambridge Univ. Press, 2003).
67. D. W. Anthony, D. Ringe, The Indo-European homeland from linguistic and archaeological perspectives. *Annu. Rev. Linguist.* **1**, 199–219 (2015). doi: 10.1146/annurev-linguist-030514-124812
68. K. A. Sirak *et al*., A minimally-invasive method for sampling human petrous bones from the cranial base for ancient DNA analysis. *Biotechniques* **62**, 283–289 (2017). doi: 10.2144/000114558; pmid: 28625158
69. N. Rohland, I. Glocke, A. Aximu-Petri, M. Meyer, Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nat. Protoc.* **13**, 2447–2461 (2018). doi: 10.1038/s41596-018-0050-5; pmid: 30323185
70. M.-T. Gansauge *et al*., Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.* **45**, e79 (2017). pmid: 28119419
71. A. W. Briggs, P. Heyn, Preparation of next-generation sequencing libraries from damaged DNA. *Methods Mol. Biol.* **840**, 143–154 (2012). doi: 10.1007/978-1-61779-516-9_18; pmid: 22237532
72. N. Rohland, D. Reich, Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **22**, 939–946 (2012). doi: 10.1101/gr.128124.111; pmid: 22267522
73. M. M. DeAngelis, D. G. Wang, T. L. Hawkins, Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Res.* **23**, 4742–4743 (1995). doi: 10.1093/nar/23.22.4742; pmid: 8524672
74. T. Maricic, M. Whitten, S. Pääbo, Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLOS ONE* **5**, e14004 (2010). doi: 10.1371/journal.pone.0014004; pmid: 21103372
75. M. Kircher, S. Sawyer, M. Meyer, Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012). doi: 10.1093/nar/gkr771; pmid: 22021376
76. 1000 Genomes Project Consortium, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015). doi: 10.1038/nature15393; pmid: 26432245
77. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010). doi: 10.1093/bioinformatics/btp698; pmid: 20080505
78. K. Douka *et al*., Direct radiocarbon dating and DNA analysis of the Darra-i-Kur (Afghanistan) human temporal bone. *J. Hum. Evol.* **107**, 86–93 (2017). doi: 10.1016/j.jhevol.2017.03.003; pmid: 28526291
79. G. Renaud, M. Kircher, U. Stenzel, J. Kelso, freeIbis: An efficient basecaller with calibrated quality scores for Illumina sequencers. *Bioinformatics* **29**, 1208–1209 (2013). doi: 10.1093/bioinformatics/btt117; pmid: 23471300
80. G. Renaud, U. Stenzel, J. Kelso, leeHom: Adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42**, e141 (2014). doi: 10.1093/nar/gku699; pmid: 25100869
81. Q. Fu *et al*., A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013). doi: 10.1016/j.cub.2013.02.044; pmid: 23523248
82. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014). doi: 10.1186/s12859-014-0356-4; pmid: 25420514

83. S. B. McClure, O. G. Puchol, B. J. Culleton, Ams dating of human bone from Cova De La Pastora: New evidence of ritual continuity in the prehistory of eastern Spain. *Radiocarbon* **52**, 25–32 (2010). doi: 10.1017/S0033822200045008
84. J. C. Lohse, B. J. Culleton, S. L. Black, D. J. Kennett, A precise chronology of Middle to Late Holocene bison exploitation in the far southern Great Plains. *J. Texas Archeol. Hist.* **1**, 94–126 (2014).
85. G. J. van Klinken, Bone collagen quality indicators for palaeodietary and radiocarbon measurements. *J. Archaeol. Sci.* **26**, 687–695 (1999). doi: 10.1006/jasc.1998.0385
86. G. M. Santos, J. R. Southon, K. C. Druffel-Rodriguez, S. Griffin, M. Mazon, Magnesium perchlorate as an alternative water trap in AMS graphite sample preparation: A report on sample preparation at KCCAMS at the University of California, Irvine. *Radiocarbon* **46**, 165–173 (2004). doi: 10.1017/S0033822200039485
87. M. Stuiver, H. A. Polach, Discussion reporting of ^{14}C Data. *Radiocarbon* **19**, 355–363 (1977). doi: 10.1017/S0033822200003672
88. P. J. Reimer *et al*., IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869–1887 (2013). doi: 10.2458/azu_js_rc.55.16947
89. C. C. Chang *et al*., Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015). doi: 10.1186/s13742-015-0047-8; pmid: 25722852
90. Moorjani lab, priyamoorjani/DATES: DATES, Version v753, Zenodo (2019). doi: 10.5281/zenodo.3263997
91. I. Olalde *et al*., Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225–228 (2014). doi: 10.1038/nature12960; pmid: 24463515

ACKNOWLEDGMENTS

We acknowledge the people past and present whose samples we analyzed in this study. We thank O. Uberti for the design of Fig. 3 and the print summary figure. We thank the Minusinsk Regional Museum of N. M. Martyanov for sharing skeletal samples. We thank O. Ismagulov, A. Ismagulova, and A. Kunbolot Toktonsunovich for facilitating access to skeletal material. We thank the Department of Archaeology and Museums, Government of Pakistan, the Directorate of Archaeology and Museums, Government of Khyber-Pakhtunkhwa Province (Pakistan), and the 'S. Sergi' Museum of the Department of Environmental Biology, Sapienza University (Rome) for facilitating access to the materials from Swat excavated by the Italian Archaeological Mission (now ISMEO). **Funding:** N.P. carried out this work while a fellow at the Radcliffe Institute for Advanced Study at Harvard University. P.M. was supported by a Burroughs Wellcome Fund CASI award. N.N. is supported by a NIGMS (GM007753) fellowship. T.C. and A.D. were supported by the Russian Science Foundation (project 14-50-00036). T.M.S. was supported by the Russian Foundation for Basic Research (grant 18-09-00779) "Anthropological and archaeological aspects of ethnogenesis of the population of the southern part of Western and Central Siberia in the Neolithic and Early Bronze Age." D.P., S.S., and D.L. were supported by European Research Council ERC-2011-AdG 295733 grant (Langelin). O.M. was supported by a grant from the Ministry of Education and Sciences of the Russian Federation No. 33.1907, 2017/П4 "Traditional and innovational models of a development of ancient Volga population". A.E. was supported by a grant from the Ministry of Education and Sciences of the Russian Federation No. 33.5494, 2017/BP "Borderlands of cultural worlds (Southern Urals from Antiquity to Early Modern period)." Radiocarbon dating work supported by the NSF Archaeometry program BCS-1460369 to D.Ken. and B.J.C. and by the NSF Archaeology program BCS-1725067 to D.Ken. K.Th. was supported by NCP fund (MLP0117) of the Council of Scientific and Industrial Research (CSIR), Government of India, New Delhi. N.Bo., A.N., and M.Z. were supported by the Max Planck Society. D.Re. is an Investigator of the Howard Hughes Medical Institute, and his ancient DNA laboratory work was supported by National Science Foundation HOMINID grant BCS-1032255, by National Institutes of Health grant GM100233, by an Allen Discovery Center grant, and by grant 61220 from the John Templeton Foundation. **Author contributions:** N.P., P.M., N.Ro., M.Me., N.Bo., K.Th., D.Ken., M.Fr., R.Pi., and D.Re. supervised the study. A.Ki., L.O., A.C., M.V., J.Ma., V.M., E.Ki., J.Mo., G.A., A.Baga., A.Bagn., B.B., J.B., A.Biss., G.B., T.Cha., T.Chi., P.D., A.D., M.Do., K.D., N.D., M.Du., D.E., A.E., A.F., D.Fu., A.Go., A.Gr., S.G., B.H., M.J., E.Ka., A.Kh., A.Kr., E.Ku., P.K., D.L., F.M., A. M., T.M., C.M., D.M., R.M., O.M., S.Mu., A.N., D.P., R.Po., D.Ra., M.R., S.Sa., T.S., K.Sik., S.Sl., O.S., N.S., S.Sv., K.Ta., M.T., A.T., V.T., S.V., P.V., D.V., A.Y., M.Z., V.Z, A.Z., V.Sh., C.L., D.A., N.Bo., M.Fr., and R.Pi. provided samples and assembled archaeological and anthropological information. V.N., N.P., P.M., N.Ro., R.B., S.Ma., I.L., N.N., I.O., M.L., N.Ad., N.A., N.Br., F.C., O.C., B.J.C., M.Fe., D.Fe., S.F., B.Ga., D.G., M.H., E.H., T.H., D.Kea., A.L., M.Ma., K.M., M.Mi., M.N., J.O., N.Ra., K.Sir., V.Sl., K.Ste., Z.Z., M.Me., and D.Re. performed ancient DNA laboratory work or radiocarbon laboratory work or data processing work. V.N., N.P., P.M., I.O., N.Al., S.M., and D.R. analyzed genetic data. V.N., N.P., and D.R. wrote the manuscript with input from all coauthors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All newly reported sequencing data are available from the European Nucleotide Archive, accession number PRJEB32466, and the software for dating admixture events in ancient samples is available from Zenodo (*90*). Our online data visualizer is available at https://public.tableau.com/profile/vagheesh#!/vizhome/TheFormationofHumanPopulationsinSouthandCentralAsia/AncientDNA.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/365/6457/eaat7487/suppl/DC1
Materials and Methods
Figs. S1 to S61
Tables S1 to S93
Genotypes for Newly Reported Individuals
References (*92*–*226*)

30 March 2018; resubmitted 19 February 2019
Accepted 30 July 2019
10.1126/science.aat7487

Science

The formation of human populations in South and Central Asia

Vagheesh M. Narasimhan, Nick Patterson, Priya Moorjani, Nadin Rohland, Rebecca Bernardos, Swapan Mallick, Iosif Lazaridis, Nathan Nakatsuka, Iñigo Olalde, Mark Lipson, Alexander M. Kim, Luca M. Olivieri, Alfredo Coppa, Massimo Vidale, James Mallory, Vyacheslav Moiseyev, Egor Kitov, Janet Monge, Nicole Adamski, Neel Alex, Nasreen Broomandkhoshbacht, Francesca Candilio, Kimberly Callan, Olivia Cheronet, Brendan J. Culleton, Matthew Ferry, Daniel Fernandes, Suzanne Freilich, Beatriz Gamarra, Daniel Gaudio, Mateja Hajdinjak, Éadaoin Harney, Thomas K. Harper, Denise Keating, Ann Marie Lawson, Matthew Mah, Kirsten Mandl, Megan Michel, Mario Novak, Jonas Oppenheimer, Niraj Rai, Kendra Sirak, Viviane Slon, Kristin Stewardson, Fatma Zalzala, Zhao Zhang, Gaziz Akhatov, Anatoly N. Bagashev, Alessandra Bagnera, Bauryzhan Baitanayev, Julio Bendezu-Sarmiento, Arman A. Bissembaev, Gian Luca Bonora, Temirlan T. Chargynov, Tatiana Chikisheva, Petr K. Dashkovskiy, Anatoly Derevianko, Miroslav Dobes, Katerina Douka, Nadezhda Dubova, Meiram N. Duisengali, Dmitry Enshin, Andrey Epimakhov, Alexey V. Fribus, Dorian Fuller, Alexander Goryachev, Andrey Gromov, Sergey P. Grushin, Bryan Hanks, Margaret Judd, Erlan Kazizov, Aleksander Khokhlov, Aleksander P. Krygin, Elena Kupriyanova, Pavel Kuznetsov, Donata Luiselli, Farhod Maksudov, Aslan M. Mamedov, Talgat B. Mamirov, Christopher Meiklejohn, Deborah C. Merrett, Roberto Micheli, Oleg Mochalov, Samariddin Mustafokulov, Ayushi Nayak, Davide Pettener, Richard Potts, Dmitry Razhev, Marina Rykun, Stefania Sarno, Tatyana M. Savenkova, Kulyan Sikhymbaeva, Sergey M. Slepchenko, Oroz A. Soltobaev, Nadezhda Stepanova, Svetlana Svyatko, Kubatbek Tabaldiev, Maria Teschler-Nicola, Alexey A. Tishkin, Vitaly V. Tkachev, Sergey Vasilyev, Petr Velemínský, Dmitriy Voyakin, Antonina Yermolayeva, Muhammad Zahir, Valery S. Zubkov, Alisa Zubova, Vasant S. Shinde, Carles Lalueza-Fox, Matthias Meyer, David Anthony, Nicole Boivin, Kumarasamy Thangaraj, Douglas J. Kennett, Michael Frachetti, Ron Pinhasi and David Reich

Science **365** (6457), eaat7487.
DOI: 10.1126/science.aat7487

Ancient human movements through Asia

Ancient DNA has allowed us to begin tracing the history of human movements across the globe. Narasimhan *et al.* identify a complex pattern of human migrations and admixture events in South and Central Asia by performing genetic analysis of more than 500 people who lived over the past 8000 years (see the Perspective by Schaefer and Shapiro). They establish key phases in the population prehistory of Eurasia, including the spread of farming peoples from the Near East, with movements both westward and eastward. The people known as the Yamnaya in the Bronze Age also moved both westward and eastward from a focal area located north of the Black Sea. The overall patterns of genetic clines reflect similar and parallel patterns in South Asia and Europe.

Science, this issue p. eaat7487; see also p. 981

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.