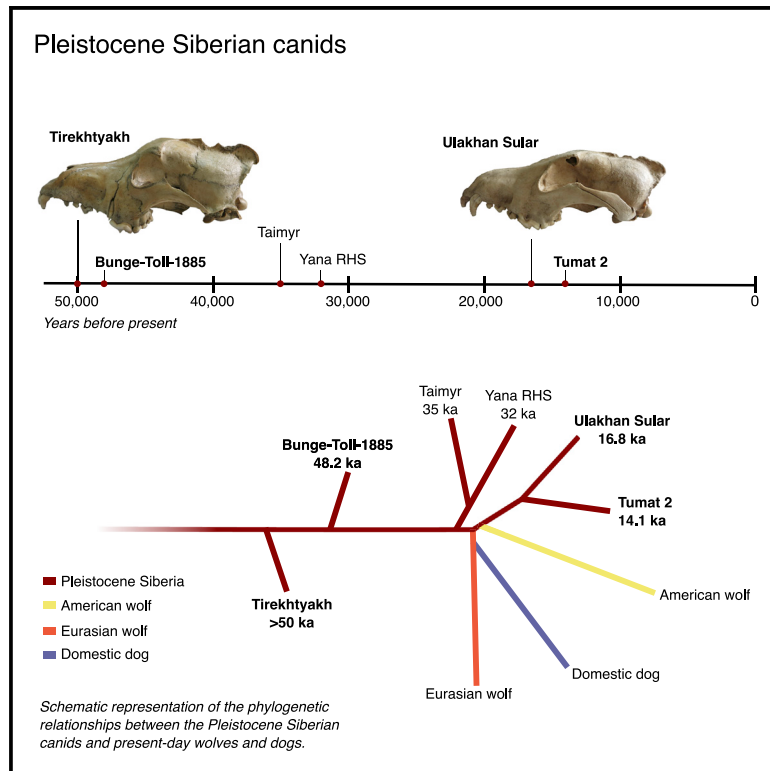# Current Biology

Genomes of Pleistocene Siberian Wolves Uncover Multiple Extinct Wolf Lineages

Graphical Abstract

Authors

Jazmín Ramos-Madrigal,
Mikkel-Holger S. Sinding,
Christian Carøe, ..., Anders J. Hansen,
M. Thomas P. Gilbert,
Shyam Gopalakrishnan

Correspondence

shyam.gopalakrishnan@sund.ku.dk

In Brief

Ramos-Madrigal et al. sequence the genomes of four Pleistocene Siberian wolves, two of which have divergent cranial morphologies. These canids represent multiple extinct lineages that dwelled in Siberia >50 ka ago and at least until 14.1 ka ago and that contributed to the genetic ancestry of arctic dogs and some East Asian wolves.

Highlights

- Pleistocene Siberian wolves represent multiple extinct evolutionary lineages

- Pleistocene wolves share ancestry with arctic dogs and some East Asian wolves

- A Paleolithic dog specimen is genetically similar to other Pleistocene wolves

CellPress

Current Biology

Report

Genomes of Pleistocene Siberian Wolves Uncover Multiple Extinct Wolf Lineages

Jazmín Ramos-Madrigal,[1,2,26] Mikkel-Holger S. Sinding,[1,3,4,5,6,26] Christian Carøe,[1] Sarah S.T. Mak,[1,2] Jonas Niemann,[1] José A. Samaniego Castruita,[1] Sergey Fedorov,[7] Alexander Kandyba,[8] Mietje Germonpré,[9] Hervé Bocherens,[10,11] Tatiana R. Feuerborn,[2,4,12] Vladimir V. Pitulko,[13] Elena Y. Pavlova,[14] Pavel A. Nikolskiy[15]

(Author list continued on next page)

[1]Section for Evolutionary Genomics, The GLOBE Institute, University of Copenhagen, Copenhagen, Denmark
[2]Center for Evolutionary Hologenomics, The GLOBE Institute, University of Copenhagen, Copenhagen, Denmark
[3]Natural History Museum, University of Oslo, Oslo, Norway
[4]The Qimmeq Project, University of Greenland, Nuussuaq, Greenland
[5]Greenland Institute of Natural Resources, Nuuk, Greenland
[6]Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland
[7]Mammoth Museum of North-Eastern Federal University, Yakutsk, Russia
[8]Department of Stone Age Archeology, Institute of Archaeology and Ethnography of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[9]Directorate Earth and History of Life, Royal Belgian Institute of Natural Sciences, Brussels, Belgium
[10]Department of Geosciences, Biogeology, University of Tübingen, Tübingen, Germany
[11]Senckenberg Centre for Human Evolution and Palaeoenvironment, Tübingen, Germany
[12]Section for GeoGenetics, The GLOBE Institute, University of Copenhagen, Copenhagen, Denmark

(Affiliations continued on next page)

SUMMARY

Extant *Canis lupus* genetic diversity can be grouped into three phylogenetically distinct clades: Eurasian and American wolves and domestic dogs.[1] Genetic studies have suggested these groups trace their origins to a wolf population that expanded during the last glacial maximum (LGM)[1–3] and replaced local wolf populations.[4] Moreover, ancient genomes from the Yana basin and the Taimyr peninsula provided evidence of at least one extinct wolf lineage that dwelled in Siberia during the Pleistocene.[3,5] Previous studies have suggested that Pleistocene Siberian canids can be classified into two groups based on cranial morphology. Wolves in the first group are most similar to present-day populations, although those in the second group possess intermediate features between dogs and wolves.[6,7] However whether this morphological classification represents distinct genetic groups remains unknown. To investigate this question and the relationships between Pleistocene canids, present-day wolves, and dogs, we resequenced the genomes of four Pleistocene canids from Northeast Siberia dated between >50 and 14 ka old, including samples from the two morphological categories. We found these specimens cluster with the two previously sequenced Pleistocene wolves, which are genetically more similar to Eurasian wolves. Our results show that, though the four specimens represent extinct wolf lineages, they do not form a monophyletic group. Instead, each Pleistocene Siberian canid branched off the lineage that gave rise to present-day wolves and dogs. Finally, our results suggest the two previously described morphological groups could represent independent lineages similarly related to present-day wolves and dogs.

RESULTS AND DISCUSSION

Sequencing Four Northeast Siberian Pleistocene Canids

We generated genomic data for three ancient canids from Northeast Siberia and increased the sequencing coverage of a fourth previously published canid (Figure 1A),[8] obtaining an average genomic depth between 5.1 and 15.2× (Table S1). The authenticity of the ancient DNA (aDNA) sequencing data were confirmed based on the read length and misincorporation patterns (Figures S1A and S1B). The samples consist of a >50-ka-old canid skull from the Tirekhtyakh River,[7] a 48-ka-old canid humerus found at the Bunge-Toll-1885 site,[9] a 16.8-ka-old canid skull from the Ulakhan Sular site,[7] and a 14.1-ka-old puppy found in association with mammoth bones from possibly an archaeological context,[10] near the Tumat village. The Tirekhtyakh and Ulakhan Sular skulls were previously characterized using multivariate analyses of cranial measurements and were classified as a Pleistocene wolf (morphology similar to

Aleksei K. Kasparov,[13] Varvara V. Ivanova,[16] Greger Larson,[17] Laurent A.F. Frantz,[18,25] Eske Willerslev,[12,19,20,21]
Morten Meldgaard,[4,12] Bent Petersen,[1,22] Thomas Sicheritz-Ponten,[1,22] Lutz Bachmann,[3] Øystein Wiig,[3]
Anders J. Hansen,[2,4,12] M. Thomas P. Gilbert,[1,2,23] and Shyam Gopalakrishnan[1,2,24,27,*]

[13]Institute for the History of Material Culture, Russian Academy of Sciences, St. Petersburg, Russia
[14]Arctic and Antarctic Research Institute, St. Petersburg, Russia
[15]Geological Institute, Russian Academy of Sciences, Moscow, Russia
[16]VNIIOkeangeologia Research Institute (The All-Russian Research Institute of Geology and Mineral Resources of the World Ocean), St. Petersburg, Russia
[17]The Palaeogenomics and Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, University of Oxford, Oxford, UK
[18]School of Biological and Chemical Sciences, Queen Mary University of London, London, UK
[19]Danish Institute for Advanced Study (D-IAS), University of Southern Denmark, Odense, Denmark
[20]Department of Zoology, University of Cambridge, Cambridge, UK
[21]Wellcome Trust Sanger Institute, University of Cambridge, Cambridge, UK
[22]Centre of Excellence for Omics-Driven Computational Biodiscovery (COMBio), Faculty of Applied Sciences, AIMST University, Kedah, Malaysia
[23]University Museum, Norwegian University of Science and Technology, Trondheim, Norway
[24]Bioinformatics, Department of Health Technology, Technical University of Denmark, Lyngby, Denmark
[25]Palaeogenomics Group, Department of Veterinary Sciences, Ludwig Maximilian University, Munich, Germany
[26]These authors contributed equally
[27]Lead Contact
*Correspondence: shyam.gopalakrishnan@sund.ku.dk
https://doi.org/10.1016/j.cub.2020.10.002

present-day wolves) and a "Paleolithic dog" (morphology intermediate between wolves and dogs), respectively (see Germonpré et al.[7] and Figures S1C and S1D). Some studies have suggested that Paleolithic dogs display a unique morphology that is not part of the morphological diversity found in other Pleistocene Eurasian wolves.[6,7] Morphological characteristics of Paleolithic dogs, such as the size of the skull and the shortening of the muzzle, have been used to suggest they possess features of domestic dogs. However, ancient mtDNA analyses of Paleolithic dogs suggest these were not related to present-day dogs and may instead represent a population that underwent an independent domestication trajectory and did not contribute to modern domestic dogs.[11] A previous study based on mtDNA data from Pleistocene canids spanning the Northern Hemisphere found that most of the oldest specimens from Northeast Siberia, including the >50-ka-old Tirekhtyakh and 48-ka-old Bunge-Toll-1885 canids, fall at the base of present-day wolf and dog diversity (Skoglund et al.[3] and Loog et al.[4]; Figure S3B). In contrast, younger specimens from the same region, including the 16.8-ka-old Ulakhan Sular[3,4] and Tumat 2 (Figure S3B), were placed within the mtDNA diversity of modern wolves and dogs.
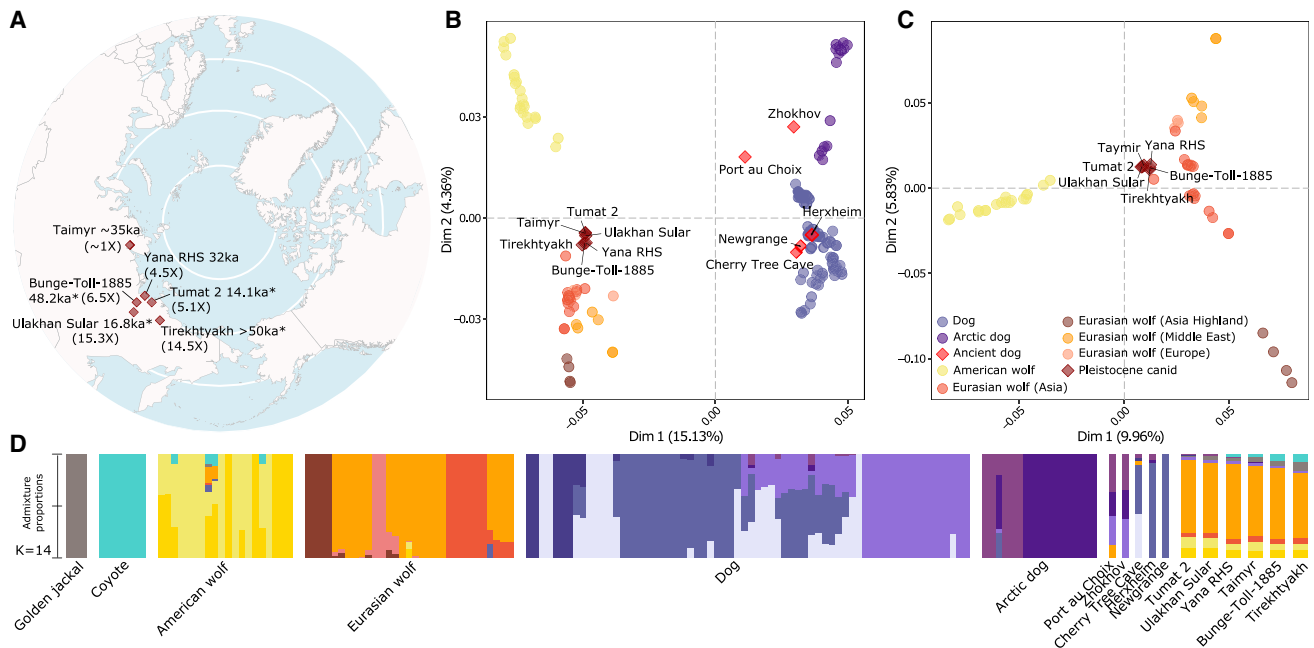
To investigate the genomic relationships between the Siberian Pleistocene and present-day canids, we assembled a dataset consisting of whole-genome data from 31 Eurasian gray wolves (*Canis lupus*), 21 American gray wolves (*Canis lupus*), 88 dogs (*Canis lupus familiaris*; including 5 ancient dogs), 2 previously published ancient wolves,[3,5] 7 coyotes (*Canis latrans*), 3 golden jackals (*Canis aureus*), and an Andean fox (*Lycalopex culpaeus*), which we used as an outgroup (Table S2). Given the varying depth of coverage of the samples, we used two approaches to compare these genomes: we created a dataset with called genotypes for a subset of the samples with a minimum depth of coverage of 10× and a second dataset of haploid data consisting of a consensus sequence for each sample. Hereafter, for simplicity, we refer to the previously published Taimyr[3] and

Yana rhinoceros horn site (RHS)[5,12] wolves and the four samples sequenced in this study including the Ulakhan Sular sample, as Pleistocene canids.

Ancestry Patterns in the Pleistocene Canids

To explore the genomic affinities between the new Pleistocene specimens and ancient and present-day wolves and dogs, we used multidimensional scaling (MDS) on the haploid dataset. The first MDS dimension separates present-day wolves from domesticated dogs and the second dimension separates Eurasian and American wolves, while splitting dogs from different geographic groups (Figure 1B).[13] We find all new Pleistocene samples cluster together with the previously published Yana RHS (32-ka-old)[5] and Taimyr (35-ka-old)[3] wolves, which are in turn closest to present-day Eurasian wolves (Figures 1B and S2A). Similarly, when excluding dogs from the MDS analysis, the Pleistocene canids form a cluster that falls close to, but does not overlap with, present-day Eurasian wolves (Figure 1C).

We then investigated the genomic relationships and structure of ancient and present-day canids using ADMIXTURE on the haploid dataset. In addition to the samples from the MDS analyses, we included 3 golden jackals (which contributed to the ancestry of the Eurasian wolf-domestic dog common ancestor)[2,14,15] and 7 coyote genomes (which contributed to the ancestry of American wolves).[16–18] When estimating 14 ancestry components (K = 14) the dataset was partitioned according to their species or to the groups identified in the MDS analysis (Figure 1D). Moreover, we observed that the ancestry components in the six Pleistocene canids were shared with present-day Eurasian and American wolves and dogs and the Eurasian wolf component was the most prevalent (μ = 73.1% ± 2.8%; Figures 1D and S2B). This ancestry profile is not shared with any modern wolf or dog. In addition, we identified a coyote-related ancestry component in the Pleistocene canids (Figure 1D), which varies with the age of the sample;

Current Biology
Report

CellPress
OPEN ACCESS

Figure 1. Genomic Relationships between Pleistocene Canids and Present-Day Wolves and Dogs
(A) Map showing the approximate geographic location of the Pleistocene canids. The genomic coverage is indicated in parentheses. *Samples sequenced in this study.
(B) MDS plot including new and reference samples. For each sample, we used a consensus sequence at sites with coverage ≥3 (5,057,255 transversion sites were used).
(C) MDS plot excluding dogs. Colors for (B) and (C) are indicated in the (C) legend.
(D) Clustering analysis using ADMIXTURE and assuming 14 ancestry components (2,387,804 transversion sites were used). Horizontal bars show different samples, colors indicate the inferred ancestry components, and the proportion of each color shows the estimated ancestry proportions. We show the six Pleistocene canids as wider bars in the rightmost side, sorted from the youngest to the oldest. See also Tables S1 and S2 and Figures S1 and S2.

the two younger specimens (16.8-ka-old Ulakhan Sular and 14.1-ka-old Tumat 2) possess negligible proportions (<0.001%), but the four older specimens show increased coyote ancestry with age (2%–7%). These results suggest that the four newly sequenced canids are genetically similar to the Taimyr and Yana RHS Pleistocene wolves and belong to the same extinct group of wolves. Furthermore, we find that the ancestry profile of the 16.8-ka-old Ulakhan Sular sample—which possesses Paleolithic dog morphology—is similar to that of the other Pleistocene specimens, in particular to that of the 14.1-ka-old Tumat 2.

Phylogenetic Relationships between Pleistocene and Present-Day Canids

Having characterized the broad genomic ancestry of the Siberian Pleistocene canids, we sought to reconstruct their phylogenetic relationships. First, we used D-statistics to test whether Pleistocene canids formed a monophyletic clade with respect to Eurasian and American wolves and domestic dogs (Figure S3). For each pair of Pleistocene canids, we computed D-statistics of the form D(Pleistocene canid 1, Pleistocene canid 2; H3, Andean fox), where H3 corresponds to dogs and Eurasian and American wolves. If a given pair of Pleistocene canids forms a clade with respect to H3, we expect them to be symmetrically related to H3 (D ∼0). We could not reject that the 14.1-ka-old Tumat 2 and the 16.8-ka-old Ulakhan Sular or the 32-ka-old Yana RHS

and ∼35-ka-old Taimyr canids form a clade to the exclusion of present-day wolves and dogs (Figure S3C).

We then used f-statistics-based admixture graphs to model the population relationships between the Pleistocene and present-day canids. We built a graph including five of the Pleistocene canids (Taimyr wolf was not included due to low coverage [∼1×]), the Eurasian wolf, American wolf, golden jackal, coyote, and dog, and used the Andean fox as an outgroup (Figure 2A shows the best fitting admixture graph and Figure S4 shows alternative graphs with similar fits). In agreement with previous studies, the best-fitting graph models the Eurasian gray wolf and dog (represented by the Portuguese wolf and the boxer, respectively) as a clade whose ancestry is a mixture of the wolf (a lineage that shares a common ancestor with the coyote) and golden jackal.[2,14,15,19] In turn, the American wolf lineage (as represented by the Ellesmere 1 wolf) could best be modeled as a mixture of wolf and coyote ancestries.[16–18] We find that all five Pleistocene canids can be modeled as lineages that sequentially branch off from the wolf lineage that leads to present-day wolves and dogs (Figure 2A orange branch). With the exception of the two youngest specimens, which form a clade (14.1-ka-old Tumat 2 and 16.8-ka-old Ulakhan Sular), the >50-ka-old Tirekhtyakh, 48-ka-old Bunge-Toll-1885, and 32-ka-old Yana RHS specimens are best modeled as independent lineages. Finally, in most models, Ulakhan Sular and Tumat 2 derive from the same lineage that gives rise to the American wolf (Figures 2A and S4). We
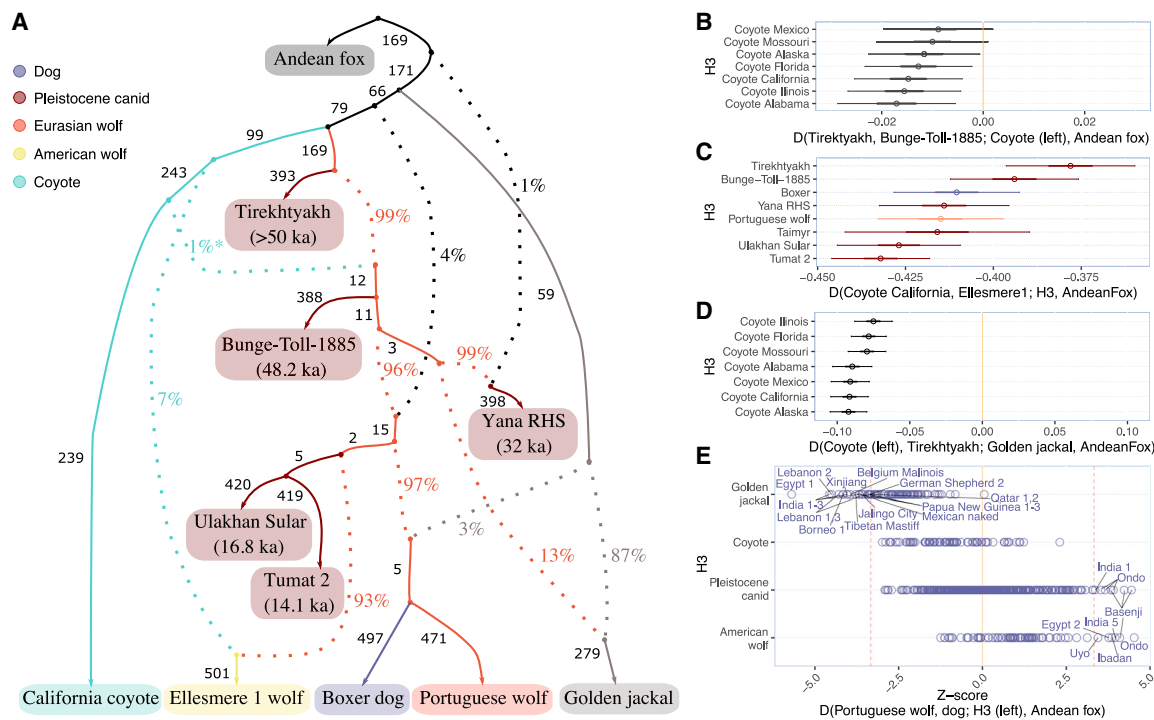
Figure 2. *F*-Statistics-Based Admixture Graph Modeling the Ancestry of Ancient and Present-Day Wolves and Dogs

(A) Schematic representation of the best model, including the Pleistocene canids and representative samples of relevant groups: Eurasian wolf; American wolf; dog; golden jackal; and coyote. Admixture graph was built using the haploid panel and 846,672 transversion sites. Continuous lines show phylogenetic relationships between samples with the numbers at the right side indicating the estimated genetic drift. Dotted lines represent admixture edges with the number at the left side indicating the percentage of ancestry deriving from each lineage. Samples included in the model are shown as color-coded boxes as indicated in the legend. This graph considers the gene flow between golden jackal and Eurasian wolf reported in Freedman et al.[2] *This admixture edge was not recovered when using the diploid dataset.

(B–E) *D*-statistics estimated using *qpDstat* testing the principal features of the admixture graph in (A). Points indicate the *D* obtained from each test. Horizontal bars indicate 1 (wider lines) and 3.33 (thinner lines) standard errors.

(B) Bunge-Toll-1885 shares more alleles with the coyote than Tirekhtyakh.

(C) The American wolf (Ellesmere 1) shares more alleles with Tumat 2 and Ulakhan Sular canids when compared to the rest of the samples in the graph.

(D) The golden jackal shares more alleles with the wolf lineage (as represented by the >50-ka-old Tirekhtyakh canid) than with coyotes.

(E) The Eurasian wolf (Portuguese wolf) forms a clade with dogs to the exclusion of all other groups in the graph (H3): points indicate the *Z* score obtained from each test, and names are indicated for dogs that yielded a significant *Z* score (|*Z*| > 3.33).

See also Figures S3 and S4.

observed a similar pattern when evaluating the branch length going from the American Ellesmere 1 wolf to all other ancient and present-day wolves using *D*-statistics, which showed Ulakhan Sular and Tumat 2 are the closest samples to Ellesmere 1 (Figure 2C; see also Treemix[20] tree in Figure S3A). These observations suggest the American wolf and the lineage represented by Ulakhan Sular and Tumat 2 might share a common ancestor before they do with the Eurasian wolf-dog clade. However, internal branches connecting the Ulakhan Sular-Tumat 2 lineage and present-day groups have small drift values showing that the relationships between these lineages are difficult to resolve. In fact, most of the internal branches that separate the different lineages contributing to the ancestries of ancient and present-day wolves in the "wolf lineage" have small drift values, and this is consistent when using a consensus sequence or called genotypes (Figures 2A and S4C). Such short branches might indicate that ancestral populations of wolves were either very large, split in a very short period of time from each other, or were constantly mixing, characteristics that have been previously documented among canids.

Previous genetic studies have found that the relationships between the Eurasian wolves, American wolves, and the dog lineages are best modeled as a star-like phylogeny,[19] pointing to a rapid split, and that gene flow is a common feature in the evolutionary history of canids.[2,15–17,19] Furthermore, the archaeological record of wolves in Eurasia suggests wolves were widespread and potentially had large population sizes,[21,22] which is also in agreement with these scenarios.

Gene Flow between the Ancient Samples and Present-Day Wolves and Dogs

Our admixture graph results suggest that, similar to the ~35-ka-old Taimyr and Yana RHS wolves, the four Pleistocene canids sequenced in this study represent extinct wolf lineages that were present in Northeast Siberia from >50 to at least 14.1 ka ago. It was previously found that Taimyr and Yana wolves contributed to the ancestry of the Siberian husky and Greenland indigenous dog breeds.[3,5,19] These Arctic breeds are part of a dog clade that diverged early during dog domestication.[5,13,23]

Current Biology
Report

CellPress
OPEN ACCESS

Additionally, ancient dog genomes that belong to the Arctic dogs' lineage have been found to share more alleles with Taimyr and Yana RHS wolves compared to European dog breeds.[5,13] We used D-statistics to test whether any of the new Pleistocene canids contributed to the ancestry of particular dog breeds. For each of the Pleistocene canids, we computed D-statistics of the form D(boxer dog, H2; Pleistocene canid, Andean fox), where H2 represents all the dogs in the dataset (Figures 3A and 3B). Similar to the patterns observed with Taimyr[3,19] and Yana RHS wolves,[5] we find that some dogs share significantly more alleles with Pleistocene canids than the boxer dog does (Figure 3C). This was the case for Arctic dogs (Greenland indigenous dogs [n = 11], Siberian [n = 3] and Alaskan [n = 2] huskies, and Alaskan malamute [n = 1]),[5,23–25] the ancient American pre-contact dog from Port au Choix[13] and the 9,549-year-old Zhokhov dog.[5] Although the significance of the tests varied, we consistently observe Arctic dogs yield the highest D values, irrespective of which of the Pleistocene canids were tested (Figure 3C). We then investigated whether the Pleistocene canid ancestry identified in the Arctic dogs was most closely related to any of the Pleistocene samples in particular. We computed D-statistics of the form D(dog 1, dog 2; Pleistocene canid 1, Pleistocene canid 2) for every possible pair of dogs and every possible pair of Pleistocene canids. We could not reject that all pairs of dogs are symmetrically related to all pairs of Pleistocene samples (Figure 3D). The latter suggests that the excess allele sharing between the Arctic dogs and the Pleistocene canids derives from a population that is related to all Pleistocene specimens tested. Alternatively, that ancestry could derive from a yet unsampled population that contributed to both Arctic dogs and Pleistocene canids.

A previous study based on ancient wolf mtDNA data suggested present-day wolves originated from a recent expansion out of Beringia—the region connecting Northern North America and Northeast Siberia—that replaced local wolf populations throughout Eurasia.[4] The expanding population could have subsumed local populations of wolves or replaced them entirely. A study based on SNP chip data, including wolves from Northeast, Southwest, and Central Asia; Europe; and North America, did not find Taimyr wolf ancestry in any of the wolf populations tested,[19] suggesting a complete replacement. To test whether any of the new Pleistocene canids contributed to the ancestry of present-day wolves in our dataset, we computed a D-statistic of the form D(Portuguese/Ellesmere1 wolf, H2; Pleistocene canids, Andean fox), where H2 represents all wolves in our dataset (Figure 4). The Portuguese and Ellesmere wolves were used in the comparison because they did not show Pleistocene canid admixture in any of the admixture graphs (Figures 2A and S4). Only the Shanxi wolf 2,[23] from west China, yielded significant results consistent with gene flow from four of the Pleistocene canids (Z = −3.4 to −4.4; Figure 4C). Furthermore, analysis using qpWave[26] confirmed that Shanxi wolf 2 is consistent with a two-way admixture between the Asian wolves and the Pleistocene canids (p = 0.406) and identified potential Pleistocene wolf ancestry in the Inner Mongolia and Chukotka wolves (p = 0.03 and 0.09).[27] Among the Eurasian wolves tested, Chukotka wolf is geographically the closest to the Siberian Pleistocene canid sites (Figure 4A). The fact that it carries Pleistocene canid ancestry suggests that some modern wolf populations from Siberia might still harbor Pleistocene canid ancestry.

Overall, we show the extinct wolf lineage that was previously characterized[3,5,19] descended from a single large or multiple smaller wolf populations that were present in Siberia and persisted from >50 to as recently as 14.1 ka ago. We find evidence that these wolf lineages contributed to the ancestry of ancient and present-day Arctic dogs, modern wolf populations in western China, and potentially to the Inner Mongolia and Chukotka wolves. Given previous evidence showing extensive gene flow among sympatric canid species (e.g., Gopalakrishnan et al.[15] and vonHoldt et al.[18]), the limited extent of Pleistocene canid ancestry among the populations tested suggests that, by the time the population that gave rise to present-day wolves and dogs dispersed through South Asia and Europe, Pleistocene canids were already on the decline in these regions. Alternatively, our Siberian Pleistocene canid genomes might not represent the genetic diversity of Pleistocene canids in South Asian and Europe. Only as more ancient and present-day wolf genomes become available, especially from Siberia and Central Asia, where the sampling is limited, will we be able to know the extent of the contribution of these Pleistocene wolf lineages to modern wolves. Finally, our results show that, despite the cranial morphological differences, the 16.8-ka-old Ulakhan Sular canid branches off from the same lineage as other Siberian Pleistocene canids, and it is not differentially related to present-day populations. Instead, the fact that the 16.8-ka-old Ulakhan Sular and 14.1-ka-old Tumat 2 and the ~35-ka-old Taimyr and 32-ka-old Yana RHS form a clade (Figures 2A and S3C) suggests that temporal proximity indicates genetic similarity.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Description of the ancient samples and radiocarbon dating
  - Tumat 2
  - Ulakhan Sular
  - Bunge-Toll-1885
  - Tirekhtyakh
  - Morphometric differences between Ulakhan Sular and Tirekhtyakh canid skulls
  - Laboratory work
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data processing
  - Authentication of ancient DNA data
  - Genotype calling
  - mtDNA phylogeny
  - Multidimensional scaling plots
  - Clustering using ADMIXTURE
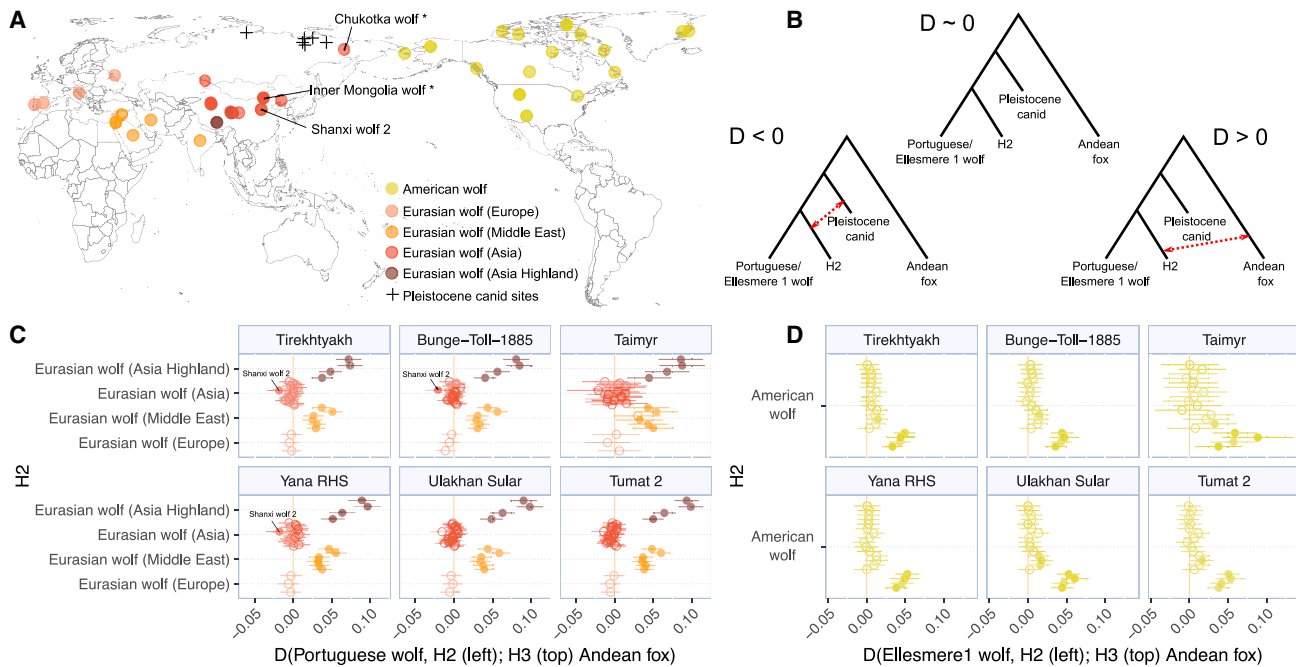 - TreeMix tree
  - Admixture graphs
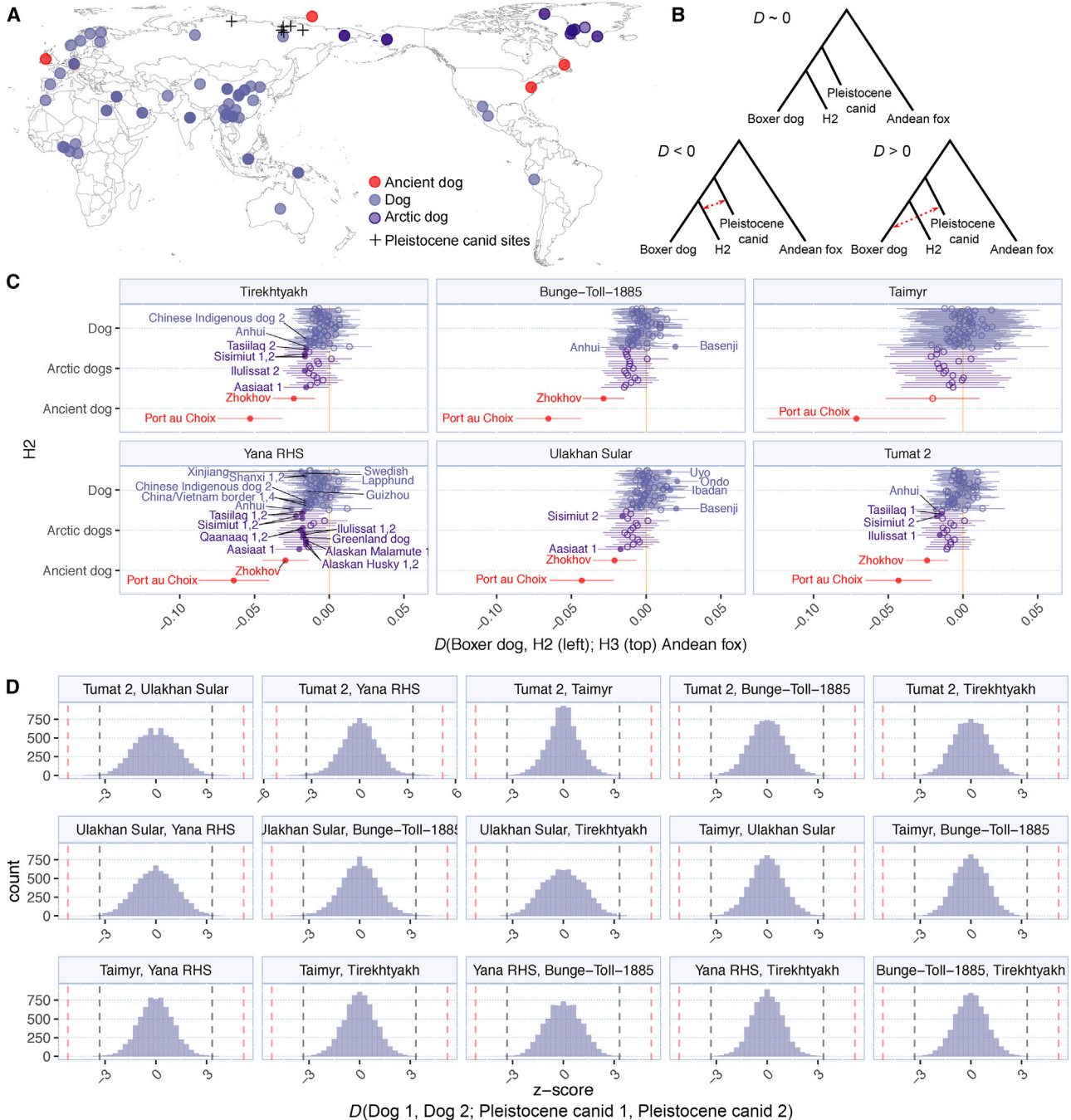
Figure 3. *D*-Statistic Testing for Gene Flow between Dogs and the Pleistocene Canids

(A) Map showing the geographic distribution of the dogs included in the tests.

(B) Graphic representation of the *D*-statistic test in (C).

(C) *D*-statistic test of the form *D*(boxer dog, H2; Pleistocene canid, Andean fox), where H2 corresponds to all dogs in the dataset. Points indicate the *D* obtained from the test. Sample names are shown for tests that yielded significant results ($Z \leq 3.33$), suggesting gene flow between H2 and H3. Horizontal bars indicate 3.33 standard errors.

(D) Histogram of the *Z* scores obtained from a *D*-statistic of the form *D*(dog 1, dog 2; Pleistocene canid 1, Pleistocene canid 2), where dog 1 and 2 are all possible pairs of present-day dogs and Pleistocene canid 1 and 2 are all possible pairs of Pleistocene canids (indicated at the top of each individual histogram). Dotted black lines correspond to a *Z* score of 3.3 (p = ~0.001), and dotted red lines represent a *Z* score of 5.198 (p = ~0.01 after applying a Bonferroni correction for 49,813 comparisons). None of the tests yielded a significant *Z* score after applying the Bonferroni correction.

Current Biology
Report

CellPress
OPEN ACCESS

Figure 4. D-Statistic Testing for Gene Flow between Present-Day Wolves and Pleistocene Canids

(A) Map showing the geographic distribution of the present-day wolves included in the tests.

(B) Graphic representation of the D-statistic tests in (C) and (D). Portuguese and Ellesmere 1 wolves are used in H1 because they do not show Pleistocene wolf ancestry in the admixture graph (Figure 2A).

(C) D-statistic test of the form D(Portuguese wolf, H2; Pleistocene canid, Andean fox), where H2 corresponds to all Eurasian wolves in the dataset. Only Shanxi 2 wolf yielded a Z score consistent with gene flow from 3 of the Pleistocene canids.

(D) D-statistic test of the form D(Ellesmere 1 wolf, H2; ancient wolf, Andean fox), where H2 corresponds to all American wolves in the dataset. American wolves in H2 are sorted according to their proportion of coyote ancestry as identified in Gopalakrishnan et al.[15] from the one with the smallest (top) to the largest (bottom) proportion. Points indicate the D value obtained from each test. Sample names and closed circles are shown for test that yielded significant results ($|Z| \geq 3.33$). Horizontal bars indicate 3.33 standard errors in (C) and (D).

- ○ Gene-flow tests: D-statistics and qpWave
- ○ Eurasian wolves (Europe)
- ○ Eurasian wolves (Middle East and South Asia)
- ○ Eurasian wolves (Asia)
- ○ Eurasian wolf (Asia Highland)
- ○ American wolf

SUPPLEMENTAL INFORMATION

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

M.T.P.G., M.-H.S.S., and S.G. conceived the study. M.-H.S.S., S.S.T.M., and C.C. did the ancient DNA lab work. J.R.-M. and S.G. performed the bioinformatic analysis. S.F., V.V.P., E.Y.P., P.A.N., A.K.K., and V.V.I. contributed with sample collection. S.F., A.K., M.G., H.B., T.R.F., V.V.P., E.Y.P., P.A.N., A.K.K., and V.V.I. provided archaeological context. J.N., J.A.S.C., B.P., and T.S.-P. provided computation expertise. J.R.-M., M.-H.S.S., S.G., and M.T.P.G. supervised the work. J.R.-M., M.-H.S.S., S.G., T.R.F., J.N., G.L., L.A.F.F., E.W., M.M., B.P., T.S.-P., L.B., Ø.W., and A.J.H. interpreted the results. J.R.-M., M.-H.S.S., S.G., and M.T.P.G. wrote the manuscript with input from all authors. All authors read and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

CellPress
OPEN ACCESS

Current Biology
Report

REFERENCES

1. Fan, Z., Silva, P., Gronau, I., Wang, S., Armero, A.S., Schweizer, R.M., Ramirez, O., Pollinger, J., Galaverni, M., Ortega Del-Vecchyo, D., et al. (2016). Worldwide patterns of genomic variation and admixture in gray wolves. Genome Res. 26, 163–173.

2. Freedman, A.H., Gronau, I., Schweizer, R.M., Ortega-Del Vecchyo, D., Han, E., Silva, P.M., Galaverni, M., Fan, Z., Marx, P., Lorente-Galdos, B., et al. (2014). Genome sequencing highlights the dynamic early history of dogs. PLoS Genet. 10, e1004016.

3. Skoglund, P., Ersmark, E., Palkopoulou, E., and Dalén, L. (2015). Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. Curr. Biol. 25, 1515–1519.

4. Loog, L., Thalmann, O., Sinding, M.S., Schuenemann, V.J., Perri, A., Germonpré, M., Bocherens, H., Witt, K.E., Samaniego Castruita, J.A., Velasco, M.S., et al. (2020). Ancient DNA suggests modern wolves trace their origin to a Late Pleistocene expansion from Beringia. Mol. Ecol. 29, 1596–1610.

5. Sinding, M.S., Gopalakrishnan, S., Ramos-Madrigal, J., de Manuel, M., Pitulko, V.V., Kuderna, L., Feuerborn, T.R., Frantz, L.A.F., Vieira, F.G., Niemann, J., et al. (2020). Arctic-adapted dogs emerged at the Pleistocene-Holocene transition. Science 368, 1495–1499.

6. Germonpré, M., Lázničková-Galetová, M., and Sablin, M.V. (2012). Palaeolithic dog skulls at the Gravettian Předmostí site, the Czech Republic. J. Archaeol. Sci. 39, 184–202.

7. Germonpré, M., Fedorov, S., Danilov, P., Galeta, P., Jimenez, E.-L., Sablin, M., and Losey, R.J. (2017). Palaeolithic and prehistoric dogs and Pleistocene wolves from Yakutia: Identification of isolated skulls. J. Archaeol. Sci. 78, 1–19.

8. Mak, S.S.T., Gopalakrishnan, S., Carøe, C., Geng, C., Liu, S., Sinding, M.S., Kuderna, L.F.K., Zhang, W., Fu, S., Vieira, F.G., et al. (2017). Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing. Gigascience 6, 1–13.

9. Pitulko, V.V., Tikhonov, A.N., Pavlova, E.Y., Nikolskiy, P.A., Kuper, K.E., and Polozov, R.N. (2016). Paleoanthropology. Early human presence in the Arctic: Evidence from 45,000-year-old mammoth remains. Science 351, 260–263.

10. Kandyba, A.V., Fedorov, S.E., Dmitriev, A.I., and Protodiakonov, K.I. (2015). A new late Neo-Pleistocene archaeological object Syalakh site in the Russian Arctic region. In Problems of Archaeology, Ethnography, Anthropology of Siberia and Neighboring Territories, A.P. Derevyanko, ed. (IAET SB RAS), pp. 90–93.

11. Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V.J., Sawyer, S.K., Greenfield, D.L., Germonpré, M.B., Sablin, M.V., López-Giráldez, F., Domingo-Roura, X., et al. (2013). Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. Science 342, 871–874.

12. Nikolskiy, P.A., Sotnikova, M.V., Nikol'skii, A.A., and Pitulko, V.V. (2018). Predomestication wolf-human relationships in the northern East Siberia of 30 000 years ago - evidence from the Yana Paleolithic site, arctic Siberia. Stratum Plus 1, 231–262.

13. Ní Leathlobhair, M., Perri, A.R., Irving-Pease, E.K., Witt, K.E., Linderholm, A., Haile, J., Lebrasseur, O., Ameen, C., Blick, J., Boyko, A.R., et al. (2018). The evolutionary history of dogs in the Americas. Science 361, 81–85.

14. Moura, A.E., Tsingarska, E., Dąbrowski, M.J., Czarnomska, S.D., Jędrzejewska, B., and Pilot, M. (2014). Unregulated hunting and genetic recovery from a severe population decline: the cautionary case of Bulgarian wolves. Conserv. Genet. 15, 405–417.

15. Gopalakrishnan, S., Sinding, M.-H.S., Ramos-Madrigal, J., Niemann, J., Samaniego Castruita, J.A., Vieira, F.G., Carøe, C., Montero, M.M., Kuderna, L., Serres, A., et al. (2018). Interspecific gene flow shaped the evolution of the genus Canis. Curr. Biol. 28, 3441–3449.e5.

16. Sinding, M.S., Gopalakrishan, S., Vieira, F.G., Samaniego Castruita, J.A., Raundrup, K., Heide Jørgensen, M.P., Meldgaard, M., Petersen, B., Sicheritz-Ponten, T., Mikkelsen, J.B., et al. (2018). Population genomics of grey wolves and wolf-like canids in North America. PLoS Genet. 14, e1007745.

17. vonHoldt, B.M., Pollinger, J.P., Earl, D.A., Knowles, J.C., Boyko, A.R., Parker, H., Geffen, E., Pilot, M., Jedrzejewski, W., Jedrzejewska, B., et al. (2011). A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. Genome Res. 21, 1294–1305.

18. vonHoldt, B.M., Cahill, J.A., Fan, Z., Gronau, I., Robinson, J., Pollinger, J.P., Shapiro, B., Wall, J., and Wayne, R.K. (2016). Whole-genome sequence analysis shows that two endemic species of North American wolf are admixtures of the coyote and gray wolf. Sci. Adv. 2, e1501714.

19. Pilot, M., Moura, A.E., Okhlopkov, I.M., Mamaev, N.V., Alagaili, A.N., Mohammed, O.B., Yavruyan, E.G., Manaseryan, N.H., Hayrapetyan, V., Kopaliani, N., et al. (2019). Global phylogeographic and admixture patterns in grey wolves and genetic legacy of an ancient Siberian lineage. Sci. Rep. 9, 17328.

20. Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8, e1002967.

21. Leonard, J.A., Vilà, C., Fox-Dobbs, K., Koch, P.L., Wayne, R.K., and Van Valkenburgh, B. (2007). Megafaunal extinctions and the disappearance of a specialized wolf ecomorph. Curr. Biol. 17, 1146–1150.

22. Sommer, R., and Benecke, N. (2005). Late-Pleistocene and early Holocene history of the canid fauna of Europe (Canidae). Mamm. Biol. 70, 227–241.

23. Wang, G.-D., Zhai, W., Yang, H.-C., Wang, L., Zhong, L., Liu, Y.-H., Fan, R.-X., Yin, T.-T., Zhu, C.-L., Poyarkov, A.D., et al. (2016). Out of southern East Asia: the natural history of domestic dogs across the world. Cell Res. 26, 21–33.

24. Wiedmer, M., Oevermann, A., Borer-Germann, S.E., Gorgas, D., Shelton, G.D., Drögemüller, M., Jagannathan, V., Henke, D., and Leeb, T. (2015). A RAB3GAP1 SINE insertion in Alaskan huskies with polyneuropathy, ocular abnormalities, and neuronal vacuolation (POANV) resembling human Warburg micro syndrome 1 (WARBM1). G3 (Bethesda) 6, 255–262.

25. Decker, B., Davis, B.W., Rimbault, M., Long, A.H., Karlins, E., Jagannathan, V., Reiman, R., Parker, H.G., Drögemüller, C., Corneveaux, J.J., et al. (2015). Comparison against 186 canid whole-genome sequences reveals survival strategies of an ancient clonally transmissible canine tumor. Genome Res. 25, 1646–1655.

26. Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M.V., Rojas, W., Duque, C., Mesa, N., et al. (2012). Reconstructing Native American population history. Nature 488, 370–374.

27. Wang, G.-D., Zhai, W., Yang, H.-C., Fan, R.-X., Cao, X., Zhong, L., Wang, L., Liu, F., Wu, H., Cheng, L.-G., et al. (2013). The genomics of selection in dogs and the parallel evolution between dogs and humans. Nat. Commun. 4, 1860.

28. Botigué, L.R., Song, S., Scheu, A., Gopalan, S., Pendleton, A.L., Oetjens, M., Taravella, A.M., Seregély, T., Zeeb-Lanz, A., Arbogast, R.-M., et al. (2017). Ancient European dog genomes reveal continuity since the Early Neolithic. Nat. Commun. 8, 16082.

29. Frantz, L.A.F., Mullin, V.E., Pionnier-Capitan, M., Lebrasseur, O., Ollivier, M., Perri, A., Linderholm, A., Mattiangeli, V., Teasdale, M.D., Dimopoulos, E.A., et al. (2016). Genomic and archaeological evidence suggest a dual origin of domestic dogs. Science 352, 1228–1231.

30. Auton, A., Rui Li, Y., Kidd, J., Oliveira, K., Nadel, J., Holloway, J.K., Hayward, J.J., Cohen, P.E., Greally, J.M., Wang, J., et al. (2013). Genetic recombination is targeted towards gene promoter regions in dogs. PLoS Genet. 9, e1003984.

Current Biology
Report

CellPress
OPEN ACCESS

31. Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas, E.J., 3rd, Zody, M.C., et al. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature *438*, 803–819.

32. Zhang, W., Fan, Z., Han, E., Hou, R., Zhang, L., Galaverni, M., Huang, J., Liu, H., Silva, P., Li, P., et al. (2014). Hypoxia adaptations in the grey wolf (*Canis lupus chanco*) from Qinghai-Tibet Plateau. PLoS Genet. *10*, e1004466.

33. Gopalakrishnan, S., Samaniego Castruita, J.A., Sinding, M.-H.S., Kuderna, L.F.K., Räikkönen, J., Petersen, B., Sicheritz-Ponten, T., Larson, G., Orlando, L., Marques-Bonet, T., et al. (2017). The wolf reference genome sequence (*Canis lupus lupus*) and its implications for *Canis* spp. population genomics. BMC Genomics *18*, 495.

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

35. Schubert, M., Ermini, L., Der Sarkissian, C., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M.D., Fernández, R., Kircher, M., McCue, M., et al. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nat. Protoc. *9*, 1056–1082.

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

37. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

38. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.

39. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics *15*, 356.

40. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

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

42. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics *29*, 1682–1684.

43. Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–3066.

44. Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321.

45. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.

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

47. Leppälä, K., Nielsen, S.V., and Mailund, T. (2017). admixturegraph: an R package for admixture graph manipulation and fitting. Bioinformatics *33*, 1738–1740.

48. Ramsey, C.B., Scott, E.M., and van der Plicht, J. (2013). Calibration for archaeological and environmental terrestrial samples in the time range 26–50 ka cal BP. Radiocarbon *55*, 2021–2027.

49. Pitulko, V., Pavlova, E., and Nikolskiy, P. (2017). Revising the archaeological record of the Upper Pleistocene Arctic Siberia: human dispersal and adaptations in MIS 3 and 2. Quat. Sci. Rev. *165*, 127–148.

50. Galeta, P., Lázničková-Galetová, M., Sablin, M., and Germonpré, M. (2020). Morphological evidence for early dog domestication in the European Pleistocene: new evidence from a randomization approach to group differences. Anat. Rec. Published online August 31, 2020. https://doi.org/10.1002/ar.24500.

51. Germonpré, M., Sablin, M.V., Stevens, R.E., Hedges, R.E.M., Hofreiter, M., Stiller, M., and Després, V.R. (2009). Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: osteometry, ancient DNA and stable isotopes. J. Archaeol. Sci. *36*, 473–490.

52. Germonpré, M., Bocherens, H., Lázničková-Galetová, M., and Sablin, M.V. Could incipient dogs have enhanced differential access to resources among Upper Palaeolithic hunter-gatherers in Europe? In Social Inequality before Farming. Multidisciplinary Approaches to the Study of Social Organisation in Prehistoric and Extant Hunter-Gatherer Societies, L. Moreau, ed. (McDonald Institute Conversations).

53. Campana, M.G., Parker, L.D., Hawkins, M.T.R., Young, H.S., Helgen, K.M., Szykman Gunther, M., Woodroffe, R., Maldonado, J.E., and Fleischer, R.C. (2016). Genome sequence, population history, and pelage genetics of the endangered African wild dog (*Lycaon pictus*). BMC Genomics *17*, 1013.

54. Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., and Pääbo, S. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. Proc. Natl. Acad. Sci. USA *104*, 14616–14621.

55. Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., et al. (2013). Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. Nature *499*, 74–78.

56. Lipson, M., and Reich, D. (2017). A working model of the deep relationships of diverse modern human genetic lineages outside of Africa. Mol. Biol. Evol. *34*, 889–902.

CellPress
OPEN ACCESS

Current Biology
Report

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Canid bone paleontological remains (3 individuals) and mummified soft tissue (1 individual)	This paper	Tirekhtyakh (CGG32), Bunge-Toll-1885 (CGG29), Ulakhan Sular (CGG33), Tumat 2
Chemicals, Peptides, and Recombinant Proteins		
Proteinase K	Sigma-Aldrich	Cat#3115844001
Critical Commercial Assays		
MinElute PCR Purification Kit	QIAGEN	Cat#28006
PfuTurbo Cx Hotstart DNA Polymerase	Agilent	Cat#600414
T4 DNA ligase	New England Biolabs Inc.	Cat#M0202L
T4 Polynucleotide Kinase	New England Biolabs Inc.	Cat#M0201L
T4 DNA Polymerase	New England Biolabs Inc.	Cat#M0203S
BSt 2,0 warmstart polymerase	New England Biolabs Inc.	Cat#M0538S
Deposited Data		
Tirekhtyakh (CGG32), Bunge-Toll-1885 (CGG29), Ulakhan Sular (CGG33), Tumat 2 sequencing data	This study	Table S1; ENA: PRJEB39580, https://www.ebi.ac.uk/ena/browser/view/PRJEB39580
Tumat 2 sequencing data	Mak et al.[8]	Table S1; SRA: PRJEB21089, https://www.ncbi.nlm.nih.gov/bioproject/PRJEB21089
Yana RHS (CGG23), Zhokhov dog (CGG6) and Greenland dogs sequencing data	Sinding et al.[5]	Table S1; SRA: PRJNA608847, https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA608847
Taimyr wolf sequencing data	Skolund et al.[3]	Table S1; SRA: PRJEB7788, https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJEB7788
Port au Choix (AL3194) sequencing data	Ní Leathlobhair et al.[13]	Table S1; SRA: PRJEB22148, https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJEB22148
Cherry Tree Cave and Herxheim dogs sequencing data	Botigué et al.[28]	Table S1; SRA: PRJNA319283, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA319283
Newgrange dog sequencing data	Frantz et al.[29]	Table S1; SRA: PRJEB13070, https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJEB13070
Andean fox sequencing data	Auton et al.[30]	Table S1; SRA: PRJNA232497, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA232497
Sequencing data for: 3 coyotes, 2 golden jackals and 5 wolves	Gopalakrishnan et al.[15]	Table S1; SRA: PRJNA494815, https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA494815
Sequencing data for 3 coyotes and Alaskan wolf 2	von Holdt et al.[18]	Table S1; SRA: PRJNA494719, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA494719
Sequencing data for 8 wolves and 41 dogs	Wang et al.[23]	Table S1; ENA: SRA307300, https://www.ebi.ac.uk/ena/browser/view/SRA307300
Sequencing data for 2 dogs, Israel golden jackal and 4 wolves	Freedman et al.[2]	Table S1; SRA: PRJNA274504, https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNP274504
Boxer dog sequencing data	Lindblad-Toh et al.[31]	Table S1
Sequencing data for 6 dogs and 4 wolves	Wang et al.[27]	Table S1; SRA: SRA068869, https://www.ncbi.nlm.nih.gov/sra/?term=SRA068869
Sequencing data for Siberian Husky 2 (SYXX)	Wiedmar et al.[24]	Table S1; SRA: PRJEB9590, PRJEB9591, PRJEB10823 https://www.ncbi.nlm.nih.gov/sra/?term=PRJEB9590 https://www.ncbi.nlm.nih.gov/sra/?term=PRJEB9591 https://www.ncbi.nlm.nih.gov/sra/?term=PRJEB10823
Sequencing data for Siberian Husky 3	Decker et al.[25]	Table S1; SRA: PRJNA288568, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA288568

(Continued on next page)

Current Biology
Report

CellPress
OPEN ACCESS

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Sequencing data for 13 American wolves	Sinding et al.[16]	Table S1; SRA: PRJNA496590, https://www.ncbi.nlm.nih.gov/bioproject/496590
Sequencing data for 7 wolves	Fan et al.[1]	Table S1; SRA: PRJNA255370, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA255370
Sequencing data for 10 wolves	Zhang et al.[32]	Table S1; SRA: PRJNA266585, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA266585; SRA: PRJNA448733, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA448733
Complementary admixture graphs	This study	https://figshare.com/articles/Pleistocene_canids_admixture_graphs/12681863
Grey wolf reference genome	Gopalakrishnan et al.[33]	https://sid.erda.dk/wsgi-bin/ls.py?share_id=f1ppDgUPQG
Oligonucleotides		
Illumina-compatible adapters	Meyer et al.[34]	N/A
Software and Algorithms		
PALEOMIX v1.2.13.1	Schubert et al.[35]	https://github.com/MikkelSchubert/paleomix
AdapterRemoval2	Schubert et al.[36]	https://github.com/MikkelSchubert/adapterremoval
bwa-*backtrack* v0.7.15	Li et al.[37]	http://bio-bwa.sourceforge.net/
Picard v1.128	N/A	https://broadinstitute.github.io/picard
GATK v3.4	DePristo et al.[38]	https://gatk.broadinstitute.org
GATK v3.6	DePristo et al.[38]	https://gatk.broadinstitute.org
ANGSD 0.614	Korneliussen et al.[39]	https://github.com/ANGSD/angsd
samtools v1.9	Li et al.[40]	http://samtools.sourceforge.net/
ADMIXTOOLS	Patterson et al.[41]	https://github.com/DReichLab/AdmixTools
TreeMix	Pickrell et al.[20]	https://bitbucket.org/nygcresearch/treemix/wiki/Home
mapDamage 2.0	Jónsson et al.[42]	https://github.com/ginolhac/mapDamage
mafft v7.205	Katoh et al.[43]	https://mafft.cbrc.jp/alignment/software/
phyml v3	Guindon et al.[44]	http://www.atgc-montpellier.fr/phyml/binaries.php
Plink 2.0	Chang et al.[45]	https://www.cog-genomics.org/plink/2.0/
ADMIXTURE	Alexander et al.[46]	http://dalexander.github.io/admixture/download.html
Admixture_graph	Lepälä et al.[47]	https://github.com/mailund/admixture_graph

RESOURCE AVAILABILITY

Lead Contact
Further information and requests of resources and reagents should be directed to and will be fulfilled by the Lead Contact Shyam Gopalakrishnan (shyam.gopalakrishnan@sund.ku.dk).

Materials Availability
This study did not generate new unique reagents.

Data and Code Availability
The accession number for the sequence reported in this paper is ENA: PRJEB39580. Admixture graphs complementary to those presented in the supplemental information are available in the figshare repository at: https://figshare.com/articles/Pleistocene_canids_admixture_graphs/12681863.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

We extracted DNA and generated sequencing data from three skeletal canid remains and soft tissue of one mummified canid that were excavated across four sites in Northeast Siberia. We provide a description of the sites and specimens in the Method Details.

METHOD DETAILS

Description of the ancient samples and radiocarbon dating

All samples are from different sites in Northeast Siberia (see Figure 1A for an approximate geographic location). The samples are named according to their location of origin.

Tumat 2

The Tumat 2 sample corresponds to a well preserved mummified large canid found in the permafrost near the village Tumat in East Siberia, Russia. The specimen belongs to the collections of the Mammoth Museum in Yakutsk Russia and has been directly dated to 12.223 ± 34 ^{14}C years BP (ETH-73412 calibrated to ca. 14.122 years BP).[8] Calibration was made using OxCal v4.2.4[48]. This sample was found in an archaeological context.

Ulakhan Sular

The Ulakhan Sular (C_1346) ((CGG33)) corresponds to the skull of a large canid found on the bank of the lower reach of the Adycha River, a tributary of the Yana River in East Siberia, Russia. The specimen belongs to the collections of the Mammoth Museum in Yakutsk Russia and has been directly dated to 13.925 ± 70 ^{14}C years BP (OxA-31946 calibrated to ca. 16.863,5 years BP).[7] Calibration was made using OxCal v4.2.4.[48] A previous study of this sample based on cranial measurements identified this sample as a potential 'Paleolithic dog'[7].

Bunge-Toll-1885

The Bunge-Toll-1885 sample (BT-1885) ((CGG29)) corresponds to a humerus of a large canid excavated from permafrost deposits from the Bunge-Toll-1885 site, East Siberia, Russia. The specimen belongs to the collections of the Russian Academy of Sciences and has been directly dated to 44.650 ± 825 ^{14}C years BP (GrA-57022 calibrated to ca. 48.209,5 years BP)[49]. Calibration was made using OxCal v4.2.4.[48]

Tirekhtyakh

The Tirekhtyakh sample (C_1216) ((CGG32)) corresponds to a skull of a large canid found near the junction of the Tirekhtyakh tributary with the Indigirka river, East Siberia, Russia. The specimen belongs to the collections of the Mammoth Museum in Yakutsk Russia and has been directly dated to > 50.300 ^{14}C years BP (OxA-31945 going beyond possible radiocarbon calibration). A previous study based on cranial measurements classified this sample as a 'Pleistocene wolf'[7].

Morphometric differences between Ulakhan Sular and Tirekhtyakh canid skulls

The osteometry of Ulakhan Sular (US17K) and Tirekhtyakh (TT50K) skulls (Figures S1C and S1D) was characterized in Germonpré et al.[7] In brief their skulls were compared to four reference groups based on their cranial morphology. The first group (recent Northern dog group) consisted of skulls of native dogs from Yakutia, Chukotka, Sakhalin Island, and Greenland from the 19th and 20th centuries. The second group (recent Northern wolf group) consisted of skulls of Palaearctic wolves from Belgium, Sweden, and several regions in Russia (Russian Plain, Yamal, Yakutia, Kamchatka, Far East) that lived in the 19th, 20th or 21st century. The third group (Pleistocene wolf morpho-population) consisted of seven specimens, most of which date to the Pleniglacial and were discovered in Belgium (Trou des Nutons), France (Maldidier), the Czech Republic (Předmostí), Ukraine (Mezin) and Russia (Yakutia and Russian Plain). Finally, the fourth group (the Palaeolithic dog morpho-population) consisted of eight skulls from a number of major Upper Palaeolithic sites in three European regions: Western-Europe dating to the Aurignacian (n = 1; Goyet, Belgium, calibrated age: c. 35,500 years BP), Central-Europe from the Czech Předmostí site dating to the Gravettian (n = 3; calibrated age: c. 28,500 years BP), and Eastern-Europe from the Epigravettian sites located in the Russian Plain (n = 4; Mezherich and Mezin, Ukraine, estimated calibrated age: 18,000 years BP, and Eliseevichi, Russia, calibrated age: c. 16,500 years BP)[7,50]. Several of these Palaeolithic dog skulls from the fourth reference group were handled postmortem by prehistoric people and exhibit cultural modifications[6,51,52]. The Palaeolithic dog reference group can be morphometrically distinguished from the Pleistocene and recent northern wolf reference sets by its unique combination of a relatively short skull and a relatively wide palate and braincase.[6,7,50,51] According to Germonpré et al.[7,52] and Galeta et al.,[50] the most parsimonious way to describe the members of this Palaeolithic dog morpho-population is as domestic canids which could be related or unrelated to the ancestors of the extant domestic dogs; nevertheless, they likely played specific roles in some European Upper Palaeolithic societies. It is worth noting that the Siberian Ulakhan Sular and Tirekhtyakh skulls were discovered in natural, fluvial deposits in the Sakha Republic. This is in contrast with the locations where the European Palaeolithic dogs were found, which are all key Upper Palaeolithic sites.

Figures S1E and S1F show scatterplots of the first two discriminant scores of two Discriminant Function Analyses (DFA), based on five ln-transformed measurements (DFA$_{raw}$) and size-adjusted data (DFA$_{size-adjusted}$), adapted from Germonpré et al.[7] (Figures 10 and 11 and Tables 10 and 11 of that publication). The group centroids of all groups except those of the two wolf sets, are well separated in both biplots. In the DFA$_{raw}$ biplot (Figure S1E), the Tirekhtyakh skull lies near the convex hulls of both wolf groups and is characterized by a long skull and a long P4. The Ulakhan Sular skull lies between the convex hulls of the recent northern dogs and the Palaeolithic dogs; its osteometric features include a rather short skull, a short carnassial and a wide palate. In the DFA$_{size-adjusted}$ biplot (Figure S1F), the Tirekhtyakh skull is situated near the convex hull of the Pleistocene wolves, as it has a relatively long skull and a relatively

small braincase and slender palate. The Ulakhan Sular skull lies above the convex hull of the Palaeolithic dogs and is characterized by a relatively short skull, and a relatively wide braincase and palate (see Germonpré et al.[7] for more information).

In Germonpré et al.,[7] the results of the two DFAs allowed the assignment of the two Pleistocene Siberian large canid skulls to one of the above-detailed reference groups. The Tirekhtyakh skull was assigned both in the DFA$_{raw}$ and the DFA$_{size-adjusted}$, to the Pleistocene wolf group, based on its high posterior (resp. $P_p = 0.97$; $P_p = 0.96$) and typicality probabilities (resp. $P_t = 0.95$; $P_t = 0.79$) for this group. Germonpré et al.[7] considered this specimen as a Pleistocene wolf characterized by a long skull and a long carnassial and a relatively slender braincase and palate. The Ulakhan Sular skull showed in both the DFA$_{raw}$ and the DFA$_{size-adjusted}$ a close affinity to the Palaeolithic dog group although with low typicality probabilities (DFA$_{raw}$: $P_p = 0.99$, $P_t = 0.07$; DFA$_{size-adjusted}$: $P_p = 0.99$, $P_t = 0.03$) (Germonpré et al.[7] Table 12). Its probabilities belonging to one of the other reference groups ($P_p = 0.00$, $P_t < 0.01$) indicate that it is an outlier for these three groups. It is thus not possible to assign this skull, based on its dimensions, to either of the wolf reference groups nor to the recent northern dog group (Germonpré et al.[7]; Table 12). The Ulakhan Sular skull differs morphometrically from the Tirekhtyakh skull: it is shorter and has a relatively wider braincase and palate. It was tentatively described as an Asian Palaeolithic dog.[7]

Laboratory work
DNA extractions and library constructions were done by following the same protocols as described in.[8,34] Libraries were sequenced on an Illumina HiSeq 2500 platform in 80 bp single read mode at The Danish National High-Throughput DNA Sequencing Centre. For the Tumat 2 specimen previously published data[8] (NCBI; PRJNA497993) was included and merged with the new data. Information on the data generated can be found in Table S1.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data processing
Sequencing reads were mapped to the wolf reference genome[33] using *Paleomix* v1.2.13.1[35] and aDNA standard settings. In brief adaptor sequences were removed using *AdapterRemoval* 2.0[36] and reads with less than 25 bp after removing sequencing adapters were discarded. Trimmed reads were then mapped to the reference wolf genome using BWA-*backtrack* v0.7.15 algorithm (BWA seed was disabled).[37] Reads with a mapping quality lower than 30 were discarded. PCR duplicates were identified and removed using *picard-tools* v1.128 (http://broadinstitute.github.io/picard/) according to their sequencing library (Table S1). Reads were realigned to the reference genome using *GATK* 3.4[38] and the MD-tag recalculated using *samtools calmd*.[40] Sequencing reads for all reference samples (Table S2) were obtained from public databases and mapped with the same parameters as the ancient samples[1–3,5,8,13,15,16,18,23,27–32,53]. Finally, we restricted analyses to scaffolds of at least 1Mb (66.5% of the wolf genome).

Authentication of ancient DNA data
We used *mapDamage* 2.0[42] to assess the substitution patterns in the sequencing data. *mapDamage* was run on the ancient samples on the reads with a minimum mapping quality of 30. We observed an increase in C to T and G to A substitutions in the 3′ and 5′ ends respectively[54] which is characteristic of data derived from ancient samples (Figure S1A). Additionally, we estimated type-specific error rates in the ancient samples using ANGSD 0.614 as described in[55]) which showed an increase in the C to T and G to A substitutions similar to the mapDamage results (Figure S1B).

Genotype calling
Genotype calling was done on a subset of the samples (Andean fox, California coyote, Portuguese wolf, Ellesmere wolf 1, Boxer dog, Tirekhtyakh, Bunge-Toll-1885 and Ulakhan Sular) in order to build admixture graphs using diploid data. We performed genotype calling on the mapped and filtered reads for each individual sample using *HaplotypeCaller* and *UnifiedGenotyper* from *GATK* 3.6.[38] We ran each algorithm independently on sites with a minimum depth of coverage of 10. Then we obtained the intersection of both algorithms (sites that were called in both). Finally, we set to missing heterozygous sites where the less frequent base was present in less than 20% of the reads.

mtDNA phylogeny
We built an mtDNA based phylogenetic tree with all six Pleistocene canids from this study and representatives from ancient and present-day wolves and dogs. For Tumat 2 and Yana RHS, we build a majority count consensus sequence for mtDNA using ANGSD. With the exception of Tumat 2, the mtDNA sequences of all Pleistocene canids sequenced here have been included in previous studies.[3,4] We combined the mtDNA sequences of the 135 present-day and ancient wolves from[4] 31 dogs (10 randomly sampled individuals from dog clades A and B) and 4 coyotes from[11] Tumat 2 and Yana RHS and align them using *mafft* with default parameters.[43] A maximum-likelihood tree was built using *phyml* v3[44] (Figure S3B). We used GTR substitution model (-m GTR) optimized the tree topologies, branch lengths and substitution rates (-o tlr), and obtained maximum likelihood estimates for the gamma distribution shape parameter (-a e). We chose to start from a random tree and for each optimization step, we kept the best tree between those generated through the 'nearest neighbor interchange' and 'subtree prune and regraft' routines (-s BEST). Finally, support for each bipartition was obtained based on 100 non-parametric bootstrap replicates (-b 100).

Multidimensional scaling plots

We created a SNP panel for all the samples in our dataset (Table S2). In order to include low coverage samples such as the Taimyr wolf (~1 ×) and some of the ancient dogs, we used a consensus sequence for each sample instead of called genotypes. ANGSD v0.614[39] was used to build a majority count consensus sequence (-dofasta 2) for each of the samples on sites with a minimum coverage of 3 reads. Transition sites were discarded in order to minimize aDNA damage included in the ancient samples. The final dataset consisted of 156 samples and a total of 5,496,534 transversion sites. We refer to this panel as the haploid dataset, and it was used (or a subset of its samples/sites) for various analyses. We generated a multidimensional scaling (MDS) plot by estimating the pairwise distances between samples using Plink 2.0[45] and then used the *cmdscale* function in R to estimate the MDS.

We build an MDS plot including all samples with the exception of the outgroup (Figure S2A), one including ancient and present-day wolves and dogs (146 individuals and 5,057,255 transversion sites; Figure 1B), and one including present-day wolves and the Pleistocene canids (53 individuals and 4,262,872 transversion sites; Figure 1C).

Clustering using ADMIXTURE

We used ADMIXTURE[46] to estimate ancestry components on the haploid dataset comprising whole-genome data for all samples in Table S2. The haploid panel described in the MDS analysis section was used. Samples with more than 85% missing data were excluded except for Taimyr wolf, which was included despite having 92.4% missing data. Transition sites and sites with more than 50% missing data were discarded. Additionally, we only included sites with a minor allele frequency of 0.05 or greater. The average per-sample and per-site missingness in the final dataset was 12.17% and 12%, respectively. The final dataset consisted of 2,387,804 sites and 155 individuals. ADMIXTURE was run assuming 2 to 20 admixture components (K = 2.20), and for each value of K, we ran 100 independent replicates. Figures 1D and S2B show the replicate with the best likelihood for each value of K.

TreeMix tree

We used TreeMix[20] and the haploid dataset to build a tree that allowed us to have a general idea of the phylogenetic relationships between the Pleistocene canids and present-day wolves and dogs. We included wolves coyotes and golden jackals with less than 80% missingness in the haploid panel, five Pleistocene canids (Taimyr wolf was excluded due to high missingness) and selected 32 dogs. After restricting to sites without missing data, the final dataset consisted of 209,094 transversion sites and 88 samples. TreeMix was run assuming no migration edges and using Andean fox to root the tree (Figure S3A). Additionally, we explored whether allowing for up to 10 migration edges affected the topology of the tree. In particular in relation to the higher relationships between the Dog, Eurasian wolf, American Wolf and Pleistocene canids. The first migration edge inferred indicates admixture from the Toronto American wolf to the common ancestor of coyote (migration weight of 0.27). After this admixture edge is added, the American wolf is moved to the ingroup to form a clade with Ulakhan Sular and Tumat 2 samples, consistent with that obtained in the admixture graph (Figure 1B). The overall topology of these clades remains constant after adding up to 10 migration edges. None of the inferred migration edges involves the Pleistocene canids.

Admixture graphs

F-statistics based admixture graphs as implemented in *qpGraph*[41] were used to investigate the phylogenetic relationships between ancient wolves and present-day canids. Admixture graphs were built using two datasets: the haploid panel consisting of a majority count consensus sequence generated with ANGSD v0.614[39] (as described in the MDS methods section) and a diploid panel with called genotypes for the high coverage samples (see above for the description of the genotype calling). In both cases we restricted the analysis to transversion sites to decrease the aDNA derived error. We indicate the number of sites used for each graph in their corresponding caption. Our dataset included the six Pleistocene wolves and reference samples of relevant groups: golden jackal (Israel), American wolf (Ellesmere1 wolf), Eurasian wolf (Portuguese wolf), dog (boxer dog), coyote (California) and Andean fox (as outgroup). The golden jackal and coyote were included since these species have been shown to contribute to the ancestry of wolves and dogs.[2,14,16,17] Finally we ran *qpGraph* using the following parameters: lsqmode = NO, diag = 0.0001, blgsize = 0.01 (1Mb).

We started by building individual admixture graphs for each of the ancient samples (4 newly sequenced and 2 reference samples[3,5]) together with present-day individuals to get a general idea of the relationships of each of the samples individually and to maximize the number of sites used. Then we built an admixture graph including 5 of the Pleistocene canids (Taimyr was excluded due to low coverage) and present-day individuals. Since we found consistent results using either a consensus sequence or called genotypes, for graphs including a single sample, we only show the results obtained from the consensus sequences panel.

For each graph, we started from a base graph consisting of 3 leaves (Andean fox as outgroup, coyote, and a third sample). It is worth noting that individuals included in the base graph are assumed to be non-admixed (or not allowed to have more admixture edges directly connecting to them, than the ones already present in the base graph), for that reason we explored different starting base graphs. We then used a sequential approach to fit each of the remaining samples both as a single leaf and as an admixed leaf whose ancestry derives from two nodes in the graph. Resulting graphs were evaluated based on the fit score and the *Z*-score of the

Current Biology
Report

worst *D*-statistic. After adding a new leaf, we selected the graph(s) with the best fit (we considered that two graphs had an equally good fit if they had a score difference of 3 (pval = 0.05)[56] and the expected Z-score of the worst *D*-statistic was $|Z| \leq 3.33$) and kept them for the next step. The *Admixture graph* R[47] package was used to create the admixture graphs figures. We explored different base graphs and different orders in which the samples were included as described below.

For the individual (single ancient sample) admixture graphs, we explored the following base graphs and paths:

1) We started from a base graph consisting of the Andean fox, golden jackal and coyote. To that, we added each of the ancient samples independently, first as admixed (a leaf that derived from two different nodes) and then as non-admixed leaves (a leaf deriving from a single node). In all cases, the ancient sample was significantly better modeled as a mixture of both the coyote and golden jackal branches. Then, we added the Portuguese wolf, which was modeled as a mixture similar to that of the ancient sample, but with different admixture proportions. We then added Ellesmere wolf which was, in most cases, modeled as a mixture of wolf and coyote, or of wolf and golden jackal. Finally, we added the boxer dog, which was modeled as a non-admixed leaf forming a clade with the Portuguese wolf. This path assumes that Andean fox, coyote, and golden jackal are non-admixed groups.

2) We started from a base graph consisting of the Andean fox, coyote, and the ancient sample in each case. To that, we added the golden jackal, which was modeled as a mixture of golden jackal and wolf (as represented by the ancient sample in each case). Next, we added the Portuguese wolf that was modeled as a mixture of wolf and golden jackal. By allowing both the Portuguese wolf and golden jackal to be incorporated as admixed leaves, we tried to recover previously described relationships between these two species.[2] We then added the boxer dog which was modeled as a non-admixed leaf forming a clade with the Portuguese wolf. Finally, we added the Ellesmere wolf which in most cases was modeled as a mixture of wolf and coyote. This path assumes that the ancient sample, coyote and golden jackal are non-admixed lineages.

3) We started from a base graph consisting of extant groups (the Andean fox, golden jackal, coyote, Portuguese wolf and boxer dog) except for the Ellesmere wolf. To that, we added the ancient sample first as a non-admixed leaf and then as an admixed leaf. In all cases, the ancient sample fitted significantly better as an admixed leaf. We then added Ellesmere1 wolf, which was modeled as a mixture of the wolf (in most cases deriving from the same lineage as the ancient sample) and coyote lineages. This path assumes that the Andean fox, coyote, and golden jackal are non-admixed groups and the Portuguese wolf and dog are formed as a mixture of wolf and golden jackal.

4) We started from four base graphs with similar support consisting of all extant groups (Andean fox, golden jackal, coyote, Portuguese wolf, boxer dog, Ellesmere1 wolf; the topology of these graphs can be found in the complementary figures described in the Data and Code Availability section). These graphs consider the American (Ellesmere) wolf as a mixture of the wolf and coyote lineages, and the Portuguese wolf and dog as a mixture of the wolf and golden jackal lineages. In two of them, the wolf component of the Ellesmere wolf is already a mixture of golden jackal and wolf. To each of the four graphs, we added the ancient sample first as a non-admixed leaf and then as an admixed leaf. In all cases, the ancient sample was modeled significantly better as an admixed leaf with the exception of the Taimyr wolf. The Taimyr wolf was placed as an outgroup to the clade formed by the Eurasian wolf and dog. In this path, the Tumat, Ulakhan Sular and Yana RHS samples were modeled as a mixture of the Eurasian and American wolf lineages, while the Bunge-Toll-1885 and Tirekhtyakh samples were modeled as a mixture of the golden jackal and wolf lineages.

Since models derived from 3) and 4) assume present-day groups were formed by the time the ancient sample in the graphs existed, we consider models derived from 1) and 2) yield more reasonable graphs despite 3) and 4) reached slightly better scores, especially for graphs with the oldest samples. These graphs can be found in FigShare repository under the https://doi.org/10.6084/m9.figshare.12681863.

For the admixture graph that included the five Pleistocene canids we explored the following paths:

1) We started from a base graph consisting of the Andean fox, golden jackal and coyote, and assuming none of those groups is admixed. To that, we added the ancient samples sequentially and starting from the oldest to the youngest (Tirekhtyakh, Bunge-Toll-1885, Yana RHS and Ulakhan Sular). Then, we added the Portuguese wolf, boxer dog and Ellesmere wolf. Finally, we added the Tumat 2 sample. Final graphs are shown in Figure S4B.

2) We started with a base graph consisting of the Andean fox, Tirekhtyakh and coyote. To that, we added the remaining samples in the following order: Bunge-Toll-1885, Yana RHS, golden jackal, Ulakhan Sular, Portuguese wolf, boxer dog, Ellesmere wolf and Tumat 2. We built this graph using both the haploid and diploid datasets. In the case of the diploid dataset, we did not include Yana and Tumat 2 in order to maximize the number of sites available. Overall, we obtain consistent results with both datasets. The only difference is that, when we consider the haploid dataset, the Bunge-Toll-1885 sample is modeled as a mixture of 99% from the wolf lineage and 1% of the coyote lineage. In contrast, when we consider the diploid (called genotypes) dataset the Bunge-Toll-1885 sample is modeled as a non-admixed leaf deriving from the same branch as Tirekhtiakh. A possible explanation for this difference is the smaller number of sites available in the called genotypes dataset, which might not be enough to recover this admixture (846,672 sites in the haploid dataset versus 71,493 sites in the diploid dataset). Graphs resulting from each step are shown in Figures 2A, S4A, and S4C.

We considered the first graph to better recapitulate the admixture history of these groups as it takes into account the bi-directional gene-flow that has been documented between the wolf and golden jackal.[2,14,15] Individual graphs and graphs showing each of the steps followed to reach the final models can be found in the figShare repository under the https://doi.org/10.6084/m9.figshare.12681863.

Gene-flow tests: D-statistics and qpWave

We used D-statistics to evaluate the possibility of gene-flow between the Pleistocene canids and present-day groups and to validate the admixture graphs obtained from the graph fitting approach. D-statistics were estimated using ADMIXTOOLS[41] and the haploid dataset. For a given test of the form D(H1 H2; H3, Outgroup), a test resulting in $D < 0$ indicates that samples in H2 and H3, or H1 and the outgroup share more alleles than the expected under the null hypothesis, conversely, if $D > 0$ then H1 and H3, or H2 and the outgroup share more alleles than the expected under the null hypothesis. A weighted block jackknife procedure over blocks of 1Mb was used to estimate the significance of each test.

We used $qpWave$[26] to further test the possibility of gene-flow between present-day American or Eurasian wolves and the Pleistocene specimens. The haploid dataset and the following parameters were used: allsnps = YES (we found the results to be consistent when using allspns = NO) and blgsize = 0.01 (1Mb). For two groups of samples qpWave tests whether the first group (samples in the left side of the test) derives from at least N independent 'migration streams' from the samples in the second group (right side of the test). For a given test where a group of wolves are in the left side of the test, and the Pleistocene canids and Andean fox are in the right side of the test (pairs of wolves; Pleistocene canids, Andean fox), we expect them to be consistent with a single migration stream if they do not carry Pleistocene canid gene flow. For the left side of the test, we used the 4 geographic groups of Eurasian wolves and the American wolves, and for the right side of the test we used the Andean fox and the 5 high coverage Pleistocene canids (only Taimyr was excluded due to high missingness). We started by testing if each group of wolves was consistent with a single migration stream, and for those that were not consistent we then sequentially removed the individual with the highest residuals until the remaining samples were consistent with a single migration stream. Finally, we evaluated models including and excluding the Andean fox among the population on the right side of the test. A description of the procedure followed for each group is found below.

Eurasian wolves (Europe)

For the European wolves (n = 4) we could not reject the model with a single migration stream in both cases considering or excluding the Andean fox from the population in the right side of the test (p value of 0.219 and 0.230, respectively).

Eurasian wolves (Middle East and South Asia)

For this group (n = 5) we were able to reject the single migration stream model (p value = 9.42E-08) but we could not reject the model with two migration streams (p value = 0.82), when using the Andean fox as outgroup among the populations in the right side of the test. However, when excluding the Andean fox, we cannot reject a single migration stream (p value = 0.775). The latter suggests that some of the Middle East wolves carry ancestry from a population related to the outgroup. To confirm this, we estimated all possible tests of the form D(Portuguese wolf, Middle East wolf; Pleistocene canids, Andean fox) (Figure 4C) and find significant results suggesting gene flow from a population related to the outgroup and all Middle East wolves ($Z > 3.5$). These results are consistent with previous studies showing admixture between African canids the Saudi and Syrian wolves[15]).

Eurasian wolves (Asia)

For the Asian wolves (n = 17) we could reject the models with one (p value = 1.62E-08) and two (p value = 0.018) streams of migration, but we could not reject the model with three (p value = 0.508). We then removed the Shanxi wolf 2[23] from the population on the left side of the test since it yielded the highest residuals. For the remaining samples, we could reject a model with one migration stream (p value = 0.0002), but not the one with two migration streams (p value = 0.064). We next removed the Chukotka wolf,[27] after which we could reject the single migration model (p value = 0.016) but not the two migrations model (p value = 0.391). We next removed the Inner Mongolia wolf 1,[27] and for the remaining samples we again rejected the model with a single migration (p value = 0.048), but could not reject the one with two migrations (p value = 0.308). Finally, we removed the Inner Mongolia wolf 3,[32] after which we could not reject the single migration model (p value = 0.08). We confirmed that each of the wolves that were removed (potentially admixed) were consistent with Pleistocene canid gene-flow by testing the non-admixed set of Asian wolves and each of the admixed wolves. In every case we were able to reject the models with a single migration but not the one with two. The result for the Shanxi wolf 2 is consistent with the D-statistic test shown in Figures 4B and 4C ($Z \leq -3.38$), in contrast, the Inner Mongolia 1 and 3 and Chukotka wolves did not yield significant z-scores in the D-statistic tests ($Z \leq -1.2$, $Z \leq -0.27$, and $Z \leq -0.18$ respectively; Figure 4C).

Eurasian wolf (Asia Highland)

For the Asian Highland wolves (n = 4) we are able to reject the model with one migration (p value = 2.09E-42), but not the model with two migration streams (p value = 0.318). We were not able to fit a model with one migration for any of the combinations of 3 or 2 wolves from this group. Given the results from the tests in Figure 4C, showing all Asian Highland wolves share more alleles with the outgroup than with the Pleistocene canids, we tested whether Asian Highland wolves carry gene-flow from an outgroup related source. To do so, we used the > 50 ka Tirekhtyakh sample as an outgroup instead of the Andean fox, but keeping the latter among the populations in the right side of the test. In this case we find that the Asian Highland wolves are consistent with two migration waves

Current Biology
Report

 CellPress
OPEN ACCESS

(p value = 9.78E-43 (one migration), p value = 0.565 (two migrations)) with the Andean fox sharing residuals with all the Asian Highland wolves.

American wolf

For the American wolves (n = 19) we were able to reject the model with one migration (p value = 0), but the one with two migration streams (p value = 0.0677). We followed a similar procedure as the one described for the Asian wolves, where we removed the sample with the highest residuals from the populations in the left and reran the analysis until the populations fit as a single migration stream. We found that after removing the Alaskan, Yellowstone and Mexican wolves, we cannot reject that the remaining samples derive from a single migration stream (p value = 0.119). American wolves have been shown to carry gene-flow from coyotes,[16–18] in particular the Alaskan, Yellowstone and Mexican wolves were previously found among the wolves with the highest coyote ancestry[16] (see also Figure 4C).