

This is a repository copy of *Origins and genetic legacy of prehistoric dogs*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/168236/>

Version: Accepted Version

Article:

Bergström, Anders, Frantz, Laurent, Schmidt, Ryan et al. (53 more authors) (2020) Origins and genetic legacy of prehistoric dogs. *Science* (New York, N.Y.). pp. 557-564. ISSN 0036-8075

<https://doi.org/10.1126/science.aba9572>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Title: Origins and Genetic Legacy of Prehistoric Dogs

Authors: Anders Bergström^{1*}, Laurent Frantz^{2*}, Ryan Schmidt^{3,4}, Erik Ersmark^{5,6}, Ophelie Lebrasseur^{7,8}, Linus Girdland-Flink^{9,10}, Audrey T. Lin^{7,11,12}, Jan Stora¹³, Karl-Göran Sjögren¹⁴, David Anthony¹⁵, Ekaterina Antipina¹⁶, Sarieh Amiri¹⁷, Guy Bar-Oz¹⁸, Vladimir I. Bazaliiskii¹⁹, Jelena Bulatović²⁰, Dorcas Brown¹⁵, Alberto Carmagnini², Tom Davy¹, Sergey Fedorov²¹, Ivana Fiore^{22,23}, Deidre Fulton²⁴, Mietje Germonpré²⁵, James Haile²⁶, Evan K. Irving-Pease^{7,27}, Alexandra Jamieson⁷, Luc Janssens²⁸, Irina Kirillova²⁹, Liora Kolska Horwitz³⁰, Julka Kuzmanovic-Cvetkovic³¹, Yaroslav Kuzmin^{32,33}, Robert J. Losey³⁴, Daria Ložnjak Dizdar³⁵, Marjan Mashkour^{17,36}, Mario Novak³⁷, Vedat Onar³⁸, David Orton³⁹, Maja Pasarić⁴⁰, Miljana Radivojević⁴¹, Dragana Rajković⁴², Benjamin Roberts⁴³, Hannah Ryan⁷, Mikhail Sablin⁴⁴, Fedor Shidlovskiy²⁹, Ivana Stojanović⁴⁵, Antonio Tagliacozzo²², Katerina Trantalidou⁴⁶, Inga Ullén⁴⁷, Aritza Villaluenga⁴⁸, Paula Wapnish⁴⁹, Keith Dobney^{8,9,50}, Anders Götherström¹⁵, Anna Linderholm⁵¹, Love Dalén^{5,6}, Ron Pinhasi^{52*}, Greger Larson^{7*}, Pontus Skoglund^{1*}

Affiliations:

1. The Francis Crick Institute, London, UK
2. Queen Mary University of London, London, UK
3. School of Archaeology and Earth Institute, University College Dublin, Dublin, Ireland
4. CIBIO - InBIO, University of Porto, Campus de Vairão, Portugal
5. Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden
6. Centre for Palaeogenetics, Svante Arrhenius väg 18C, Stockholm, Sweden
7. The Palaeogenomics and Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, University of Oxford, Oxford, UK
8. Department of Archaeology, Classics and Egyptology, University of Liverpool, Liverpool, UK
9. Department of Archaeology, University of Aberdeen, Aberdeen, UK
10. Liverpool John Moores University, Liverpool, UK
11. Department of Zoology, University of Oxford, Oxford, UK
12. Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA
13. Stockholm University, Stockholm, Sweden
14. Gothenburg University, Gothenburg, Sweden
15. Hartwick College, Oneonta, USA
16. Institute of Archaeology of the Russian Academy of Sciences, Moscow, Russian Federation
17. Bioarchaeology Laboratory, Central Laboratory, University of Tehran, Tehran, Iran
18. University of Haifa, Haifa, Israel
19. Irkutsk State University, Irkutsk, Russian Federation
20. University of Belgrade, Belgrade, Serbia
21. North-Eastern Federal University, Yakutsk, Russian Federation
22. Bioarchaeology Service, Museo delle Civiltà, Rome, Italy
23. Environmental and Evolutionary Biology Doctoral Program, Sapienza University of Rome, Rome, Italy
24. Baylor University, Waco, Texas, USA
25. Royal Belgian Institute of Natural Sciences, Brussels, Belgium
26. University of Copenhagen, Copenhagen, Denmark
27. Lundbeck GeoGenetics Centre, The Globe Institute, 1350 Copenhagen, Denmark
28. University of Leiden, Leiden, Netherlands
29. Ice Age Museum, Moscow, Russian Federation
30. The Hebrew University, Jerusalem, Israel
31. Homeland Museum of Toplica, Prokuplje, Serbia
32. Sobolev Institute of Geology and Mineralogy of the Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russian Federation
33. Tomsk State University, Tomsk, Russian Federation
34. University of Alberta, Edmonton, Canada
35. Institute of Archaeology, Zagreb, Croatia
36. Archéozoologie, Archéobotanique, Sociétés, Pratiques et Environnements, AASPE, Centre National de la Recherche Scientifique, Muséum National d'Histoire Naturelle, Paris, France
37. Centre for Applied Bioanthropology, Institute for Anthropological Research, Zagreb, Croatia
38. Istanbul University-Cerrahpaşa, Istanbul, Turkey
39. University of York, York, UK
40. Institute of Ethnology and Folklore Research, Zagreb, Croatia
41. University College London, London, UK
42. Archaeological Museum Osijek, Osijek, Croatia
43. Durham University, Durham, UK

44. Zoological Institute of the Russian Academy of Sciences, Saint Petersburg, Russian Federation
45. Institute of Archaeology, Belgrade, Serbia
46. Hellenic Ministry of Culture & Sports, Athens, Greece
47. National Historical Museums, Stockholm, Sweden
48. Consolidated Research Group on Prehistory (IT-1223-19), University of the Basque Country (UPV-EHU), Vitoria-Gasteiz, Spain
49. The Pennsylvania State University, University Park, Pennsylvania, USA
50. Department of Archaeology, Simon Fraser University, Burnaby, BC., Canada
51. Texas A&M University, College Station, USA
52. Department of Evolutionary Anthropology, University of Vienna, Vienna, Austria

*Correspondence to: anders.bergstrom@crick.ac.uk (A.B.), l.frantz@qmul.ac.uk (L.F.),
ron.pinhasi@univie.ac.at (R.P.), greger.larson@arch.ox.ac.uk (G.L.),
pontus.skoglund@crick.ac.uk (P.S.)

Abstract: Dogs were the first domestic animal, but little is known about their population history and to what extent it was linked to humans. We sequenced 27 ancient dog genomes and found that all dogs share a common ancestry distinct from present-day wolves, with limited gene flow from wolves since domestication, but substantial dog-to-wolf gene flow. By 11,000 years ago, at least five major ancestry lineages had diversified, demonstrating a deep genetic history of dogs during the Paleolithic. Co-analysis with human genomes reveals aspects of dog population history that mirror humans, including Levant-related ancestry in Africa and early agricultural Europe. Other aspects differ, including the impacts of steppe pastoralist expansions in West- and East Eurasia, and a complete turnover of Neolithic European dog ancestry.

One Sentence Summary: Ancient dog genomes reveal no evidence for multiple origins but an early diversification, followed by a genetic history that both mirrors and differs from humans.

Main text: Wolves were the first animal with which humans formed a mutualistic relationship, eventually giving rise to dogs. While there is little consensus regarding when (1–9), where (2, 8–13), and how many times (1, 8, 9, 14) domestication took place, the archaeological (9, 15) record attests to a long-term and close relationship to humans (9, 16–18). Modern dog genomes have revealed a complex population structure (5, 8, 10, 12, 19, 20), but because only six ancient dog and wolf genomes are currently available (4, 9, 14, 21), the process by which this structure emerged remains largely unknown.

Previous mitochondrial DNA (22–29) and genomic (9, 14, 21) studies have suggested an association between the genetic signatures of dogs and their archeological context. However, dog and human genomes have not been quantitatively co-analyzed to assess the degree to which the population history of dogs was linked to that of humans—or may have been decoupled as a result of trade, human preference for particular types of dogs, variation in infectious disease susceptibility, or dogs moving between human groups.

To reconstruct dog population history we sequenced 27 ancient dog genomes up to 10,900 years old from Europe, the Near East and Siberia (**table S1**) to a median of 1.5-fold coverage (range 0.1–11X) (**Fig. 1A, table S2; (30)**). To test the association with human population history, we compiled 17 sets of human genome-wide data (30) that matched the age, geographic location and cultural contexts of the ancient dogs (**table S4**), and directly compared genetic relationships within the two species.

Global dog population structure has its origins in the Pleistocene

To characterize the global population structure of ancient and modern dogs, we applied principal component analysis (PCA) to a matrix of all possible f_4 -statistics (30), alleviating differences in error rates and missing data. This approach recapitulates a major east-west axis of dog ancestry (PC1) (8, 9, 12), in which the western extreme comprises modern and ancient western Eurasian dogs and modern African dogs (**Fig. 1B**). The eastern extreme is represented by pre-contact North American dogs (21), three 7 ky dogs from Lake Baikal in Siberia, and modern East Asian dogs including New Guinea Singing Dogs and Australian dingoes. Similar results were obtained through standard model-based clustering (**fig. S2**).

All ancient and modern European dogs have greater affinity to eastern dog ancestry than ancient Near Eastern dogs have in f_4 -tests (**fig. S3**), despite the overall east-west axis on PC1. Ancient European dogs are also distributed widely across a genetic cline between the East Eurasian and ancient Near Eastern dogs, which furthermore manifests as a linear cline along the diagonal when contrasting shared genetic drift with Baikal dogs and Levantine (Israel, 7kya) dogs using outgroup- f_3 statistics (**Fig. 1C**). Simulations indicate that this linear, diagonal cline is difficult to explain with long-standing continuous gene flow or a tree-like history, but instead suggest that the history of Mesolithic and Neolithic European dogs was marked by a major admixture episode (**Fig. 1D**) (30).

We modeled the genetic history underlying dog population structure for five populations that represent major ancestries, and tested all 135,285 possible admixture graph models with up to two admixture events (30). One model uniquely fits the data, and features the Mesolithic Karelian dog (10.9 kya) as having received part of its ancestry from a lineage related to eastern dogs, and part from the Levantine lineage (**Fig. 1E**) (two highly similar models nearly fit, **fig. S4**). The model can be extended to feature the earliest Neolithic European dog (7 kya)(14) as a mixture of the Karelian and the Levantine branches without loss of fit (**fig S5**), supporting the dual ancestry model for European dogs suggested by the ancient ancestry cline (**Fig. 1C**). The observed phylogenetic structure implies that all five ancestry lineages (Neolithic Levant, Mesolithic Karelia, Mesolithic Baikal, ancient America, New Guinea Singing dog) must have existed by 10.9kya (the radiocarbon date of the Karelian dog), and thus most likely prior to the transition from the Pleistocene to the Holocene epoch ~11.6 kya.

No detectable evidence for multiple dog origins or extensive gene flow from wild canids

Studies have suggested that wolf populations in Europe (3, 11), the Middle East (12), Central Asia (10), Siberia (31), and East Asia (2, 8), or more than one of these (9), contributed to early dog diversity. One study, however, demonstrated that modern wolves and dogs are reciprocally monophyletic, and suggested bidirectional gene flow (5). We corroborated that gene flow must have occurred by identifying widespread asymmetries between dogs in their affinity to wolves (**Fig. 2A,B, fig S7**). However, the gene flow was likely largely unidirectional from dogs into wolves, since we also identified some gray wolves that are symmetrically related to all modern and ancient dogs (**Fig. 2C**). Past gene flow from wolves into specific dog populations would have manifested as an affinity to any member of the modern gray wolf lineage in these tests, so our results suggest that persistent gene flow into dogs has been so limited as to be undetectable at the current resolution of the data. Furthermore, this result is consistent with a scenario in which all dogs derive from a single ancient, now extinct wolf population, or possibly multiple closely related wolf populations. While it is still possible that other, thus far unsampled ancient wolf populations were independently involved in early domestication (3, 9, 31), our data indicate that they did not contribute substantially to later dogs.

In contrast to the lack of wolf admixture into dogs, we identified dog admixture into almost all analyzed present-day wolves (**Fig. 2B**), with the strongest signals typically coming from dogs into geographically proximate wolf populations in Europe, the Near East and East Asia (**fig S7**). We also replicated affinities between ancient American dogs and Coyotes (21), and between African dogs and African Golden Wolves (32), though the direction of gene flow in both cases is unclear, and the small magnitude is unlikely to impact most analyses of dog relationships (**table**

95 **S5).** We did not find genome-wide evidence for gene flow from Tibetan wolves into Tibetan dogs, despite evidence for wolf ancestry locally around the *EPAS1* gene associated with adaptation to altitude (33, 34). Dogs thus do not show similar evidence of wild introgression that has been found in pigs, goats, horses, sheep and cattle (35–40).

Assessing the relationship between dog and human population histories

100 We next quantitatively compared the population relationships observed in dogs with those of humans. First, using Procrustes rotation to align f_4 -PCA results obtained on dog and human genomes matched in time and space (**Fig. 3A**; (30)), we find that the population structures of the two species resemble each other (Procrustes correlation = 0.48, $p = 0.043$). However, there are also several cases where the matched dogs and humans cluster in different parts of the PCA space. The greatest differences (**Fig. 3B**) are observed for Chalcolithic Iran, in which the human population is different from the Neolithic Levant (41, 42) but the dogs in the two regions are similar. In Neolithic Germany and Ireland, the humans are more shifted towards the Levant (43, 44) but the dogs are shifted towards Northern European hunter-gatherer contexts. In the Bronze Age Steppe and in Corded Ware Germany, the humans are shifted away from the Neolithic European cluster (45, 46) in a manner not seen in dogs.

110 Second, we evaluated if the admixture graph topologies that best fit the data for one species could also explain population relationships of the other. Though we found no graphs that fit the data perfectly for both species, graphs that fit, or nearly fit dogs rank among the 0.8-2.8% top scoring graphs in the human search, and graphs that fit humans rank among the 0.007-1.2% top scoring graphs in the dog search (**Fig. 3C**, **fig. S9**). However, we note that this analysis does not take into account the different time depth of the two species' population histories: the >40kya divergence of human East- and West Eurasian ancestries (47) is significantly older than the earliest appearance of dog morphology in the fossil record, conservatively dated to 14.5kya (48) though older (3, 31), disputed specimens (49, 50), have been claimed.

120 Third, we found that the sign (positive or negative) of f_4 -statistics in dogs match the sign in humans in 71% of 31,878 tests (null expectation 50%) across 24 matched dog-human pairs, although this decreases to 58% when restricted to dogs and humans from Europe. We identified specific f_4 -statistics that exemplify both concordance and discrepancy between the species (**Fig. 3D**). While it is not known what degree of concordance would be expected between the histories of two species based on biogeographical factors alone, the results of these three analyses demonstrate that ancestry relationships in dogs and humans share overall features, but are not identical over space and time, and there are several cases where they must have been decoupled.

Recurrent population histories

130 One notable example of concordance is that both humans and dogs in East Asia are closer to European than to Near Eastern populations, which in both humans (43) and our best-fitting graph (**Fig. 1E**) is best modelled by European ancestry being a mixture of ancestry related to the Near East and East Asia. However, the divergence of Near Eastern 'Basal Eurasian' ancestry in humans was likely >45 kya (43), suggesting that dog population dynamics may have mimicked earlier processes in humans. A second example is that all European dogs have a stronger affinity towards American and Siberian dogs than they have to New Guinea singing dogs, which likely represent a type of unadmixed East Asian dog ancestry, mirroring a circumpolar affinity between

humans in Europe and the Americas (51) (**Fig. 3D**). Human groups at Lake Baikal 24-18kya had western Eurasian origins and contributed to Native American ancestry (51), but were largely replaced by the Holocene (52). Though the dogs at Lake Baikal dated to 7kya constitute a similar link between the Americas and Europe (**Fig. 1C,E**), they do so >10ky later (**Fig. 3D**). Thus, shared circumpolar ancestry through northern Eurasia is an important feature of both human and dog population structures, though this did likely not result from the same migration episodes.

145 **Neolithic expansion into Europe**

Ancient human genomes have revealed a major ancestry transformation associated with the expansion of Neolithic agriculturalists from the Near East into Europe (43, 45, 53), and a study of ancient dog mitochondria suggested they were accompanied by dogs (27). We hypothesized that the genomic ancestry cline we observe across ancient European dogs (**Fig. 1C**) could be, at least in part, due to admixture between dogs associated with Mesolithic hunter-gatherers and incoming Neolithic farmers. Three observations support this: first, the hypothesized hunter-gatherer end of the cline is occupied by the 10.9kBP Mesolithic Karelian dog, and dogs from a 4.8kBP hunter-gatherer Pitted Ware Culture site in Sweden. Second, relative to the Swedish hunter-gatherer dogs, a contemporaneous dog from a Swedish Neolithic agricultural context is shifted towards the Levantine end of the cline, mirroring humans at the same sites (41, 53, 54) (**Fig. 3A,D; fig. S10D**). Third, Neolithic Levantine affinity increases towards the south ($p=0.0196$, linear regression), consistent with a range expansion alongside Neolithic human groups. While dogs clearly associated with Mesolithic continental 'Western hunter-gatherer' (43) human groups have yet to be identified, our results suggest that such dogs would have strong affinity towards the Siberian end of the European cline. Overall, these results indicate that the Neolithic expansion of farmers into Europe was also associated with an ancestry transformation for dogs.

Increased copy number of the *AMY2B* gene, involved in starch digestion, has been linked to dietary adaptations of dogs during the agricultural transition (6, 55, 56). The paralogous *AMY1* gene has been under adaptive evolution in humans (57), though this does not seem clearly linked to agriculture (58). We observe low copy numbers in dogs from human hunter-gatherer contexts (**Fig. 4**), although the Mesolithic Karelian dog may already have possessed an elevated number relative to wolves. Several Neolithic dogs have as many copies as present-day dogs, as early as in 5.8 ky old Iranian and 6 ky old Spanish dogs, but others display low numbers (14, 56), e.g. the 7 ky Levantine individual. These results suggest that selection for increased *AMY2B* copy number did not take place during the early stages of domestication, and in contrast to humans (58) was not advanced in Mesolithic hunter-gatherer contexts, but was variable in early agricultural populations and did not become widespread until several thousand years after the first appearance of starch-rich agricultural lifestyles.

Africa and the Near East

The clustering of modern African dogs with ancient dogs from the Levant and Iran, especially the oldest individual dating to 7 kya, suggests a Near Eastern origin (**Fig. 1B,C, fig. S2**). Western (Anatolia and the Levant) and eastern (Zagros mountains of Iran) human groups in the Fertile Crescent were highly genetically differentiated (41), and the western groups were the primary source of gene flow into Europe and Africa (41, 59) during the Neolithic. A source of African dog ancestry from the Levant (7kya) is a better fit than Iran (5.8kya) (**Fig. 5A**), mirroring the

185 human history, as well as that of cattle (40). In contrast, we are unable to distinguish whether the
Levant or Iran is the better source for Neolithic dog ancestry in Europe. Our results suggest a
single origin of sub-Saharan African dogs from the Levant (**Fig. 5B**), with limited gene flow
from outside the continent until the past few hundred years.

190 In contrast to Africa, the 7kya Neolithic Levantine population does not appear to have
contributed much, if any, ancestry to present-day dogs in the Near East. Instead, 2.3 ky old dogs
in the Levant can be modelled as having 81% Iran-related and 19% Neolithic Europe-related
ancestry (**Data S1**). By this time in the Levant, there was also human gene flow from Iran (41)
and transient gene flow from Europe (60). However, our results suggest a more complete
replacement of dog ancestry in the Levant by 2.3 kya (**Fig. 5B**). Later, modern Near Eastern dogs
195 are best modelled as mixtures of the 2.3 ky Levantine and modern European sources (**Data S1**).

Steppe pastoralist expansions

Expansions of steppe pastoralists associated with the Yamnaya and Corded Ware cultures into
Late Neolithic and Bronze Age Europe transformed the ancestry of human populations (43, 45,
200 46). To test if dog ancestry was similarly affected, we analyzed a 3.8 ky old dog from the eastern
European steppe associated with the Bronze Age Srubnaya culture. While its ancestry resembles
that of western European dogs (**Fig. 1C, fig. S10**), it is an outlier in the center of PC1-PC2 space
(**Fig. 1B**). A Corded Ware-associated dog (4.7kya) from Germany, hypothesized to have steppe
ancestry (14), can be modelled as deriving 51% of its ancestry from a source related to the
205 Srubnaya steppe dog, and the rest from a Neolithic European source ((30); **Data S1**). We obtain
similar results for a Bronze Age Swedish dog (45%; 3.1kya), but not a Bronze Age Italian dog
(4kya).

210 Despite this potential link between the steppe and the Corded Ware dog, most later European
dogs display no particular affinity to the Srubnaya dog. Modern European dogs instead cluster
with Neolithic European dogs (**Fig. 1B**), and do not mirror the lasting ancestry shift seen in
humans after the pastoralist expansion (**Fig. 3A**). While earlier and additional steppe dog
genomes are needed to better understand this process, the relative continuity between Neolithic
and present-day individuals suggests that the arrival of steppe pastoralists did not result in
215 persistent large-scale shifts in the ancestry of European dogs.

Although steppe pastoralists also expanded east, they do not appear to have contributed much
ancestry to present-day people in East Asia (46, 52). Many modern Chinese dogs display
unambiguous evidence (negative f_3 tests (30)) of being the product of admixture between a
220 population related to the New Guinea Singing Dog (and the Australian Dingo) and a West
Eurasian-related population (**table S6**). A recent study also found a mitochondrial turnover in
Chinese dogs in the last few thousand years (61). The best-fitting models involve ancestry from
modern European breeds, but also substantial contributions from the 3.8k BP Srubnaya steppe
dog (**Fig. 5A, Data S1**). Some populations, especially those in Siberia, additionally require a
225 fourth source related to the 7ky old Lake Baikal dogs, but no or minimal New Guinea Singing
Dog-related ancestry. Our results thus raise the possibility that the eastward migrations of steppe
pastoralists had a more substantial impact on the ancestry of dogs than humans in East Asia (**Fig.**
5B).

230 **Later homogenization of dog ancestry in Europe**

The extensive range of ancestry diversity among early European dogs is not preserved today, as modern European dogs are all symmetrically related to the ancient dogs in our dataset (**Fig. 1C**, **fig. S13**, **Data S1**, (30)). This suggests little to no contribution of most local Mesolithic and Neolithic populations to present-day diversity in Europe. Instead, we found that a single dog from a Neolithic megalithic context dated to 5 kya at the Frälsegården site in southwestern Sweden can be modelled as a single-source proxy for 90-100% of the ancestry of most modern European dogs, to the exclusion of all other ancient dogs (**fig. S13**, **Data S1**). This implies that a population with ancestry similar to this individual, but not necessarily originating in Scandinavia, replaced other populations and erased the continent-wide genetic cline (**Fig 5B**). This ancestry was in the middle of the cline (**Fig 1C**), such that present-day European dogs can be modelled as about equal proportions of Karelian and Levantine-related ancestries (54% and 46% respectively, for German Shepherd using the admixture graph (**Fig 1E**)).

The Frälsegården dog is also favored as a partial ancestry source for a 4ky old Bronze Age dog from Italy, a 1.5ky old dog from Turkey and Byzantine and Medieval, but not earlier dogs in the Levant (**Data S1**), providing some constraints on the timing of this ancestry expansion. However, the circumstances that initiated or facilitated the homogenization of dog ancestry in Europe from a narrow subset of that present in the European Neolithic, including the phenomenal phenotypic diversity and genetic differentiation of modern breeds (12, 19, 20) (**Fig. 1C**), remain unknown.

More recently, this modern European ancestry has dispersed globally, and today is a major component of most dog populations worldwide (**Fig. 5A**). Our ancestry models, however, reveal that some pre-colonial ancestry does survive in breeds such as the Mexican Chihuahua (~4%) and Xoloitzcuintli (~3%), and the South African Rhodesian Ridgeback (~4%) (**Data S1**).

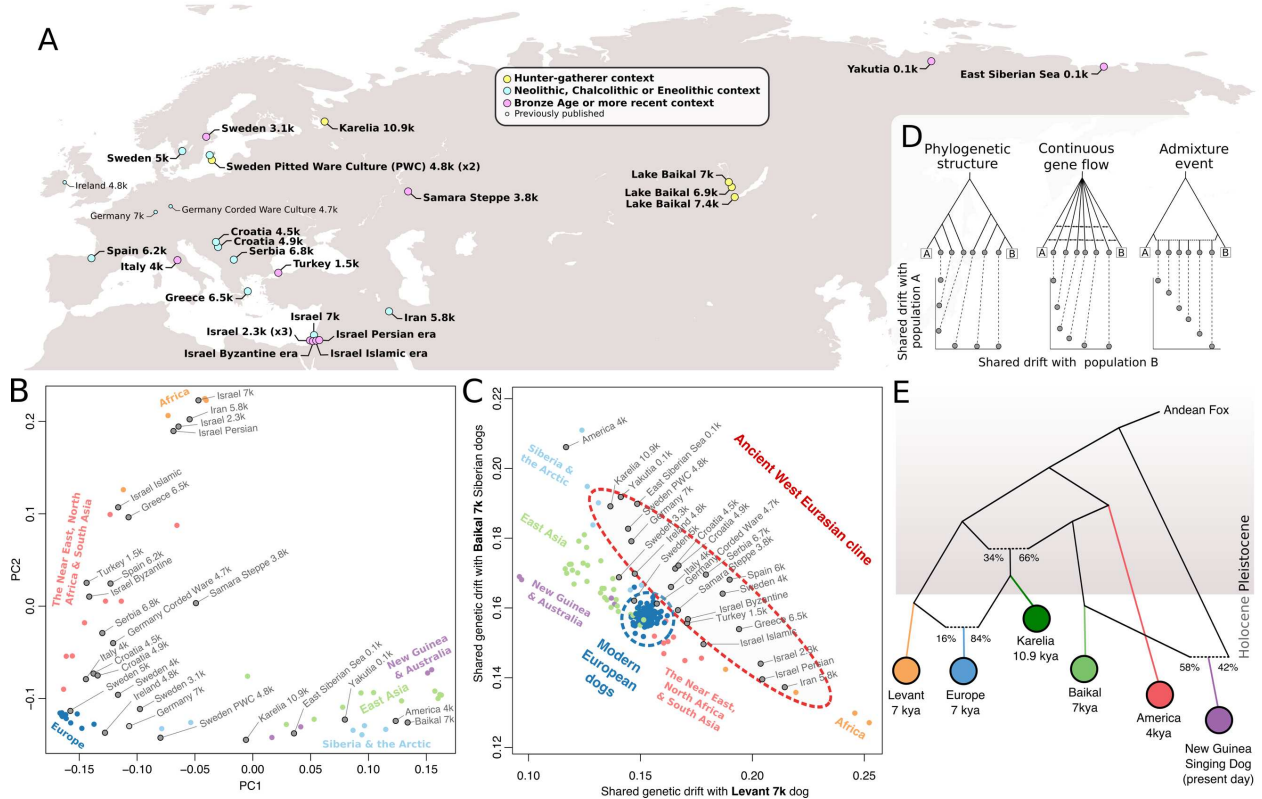
Discussion

The diversification of at least five dog ancestry lineages by the onset of the Holocene was followed by a dynamic population history that in many ways tracked that of humans, likely reflecting how dogs migrated alongside human groups. However, in several instances, these histories do not align, suggesting that humans also dispersed without dogs, dogs moved between human groups, or that dogs were cultural and/or economic trade commodities.

Certain aspects of genetic relationships between dog populations, such as an east-west Eurasian differentiation, circumpolar connections, and possible basal lineages in the Near East, resemble features of human population history that were established before the earliest estimated dates of dog domestication. This superficial mirroring between the species may therefore instead point to recurrent population dynamics, due to biogeographic or anthropological factors that remain to be understood. A key question is how dogs spread across Eurasia and the Americas by the Holocene, since no major human population movements have been identified after the initial out-of-Africa expansion that could have driven this global dispersal.

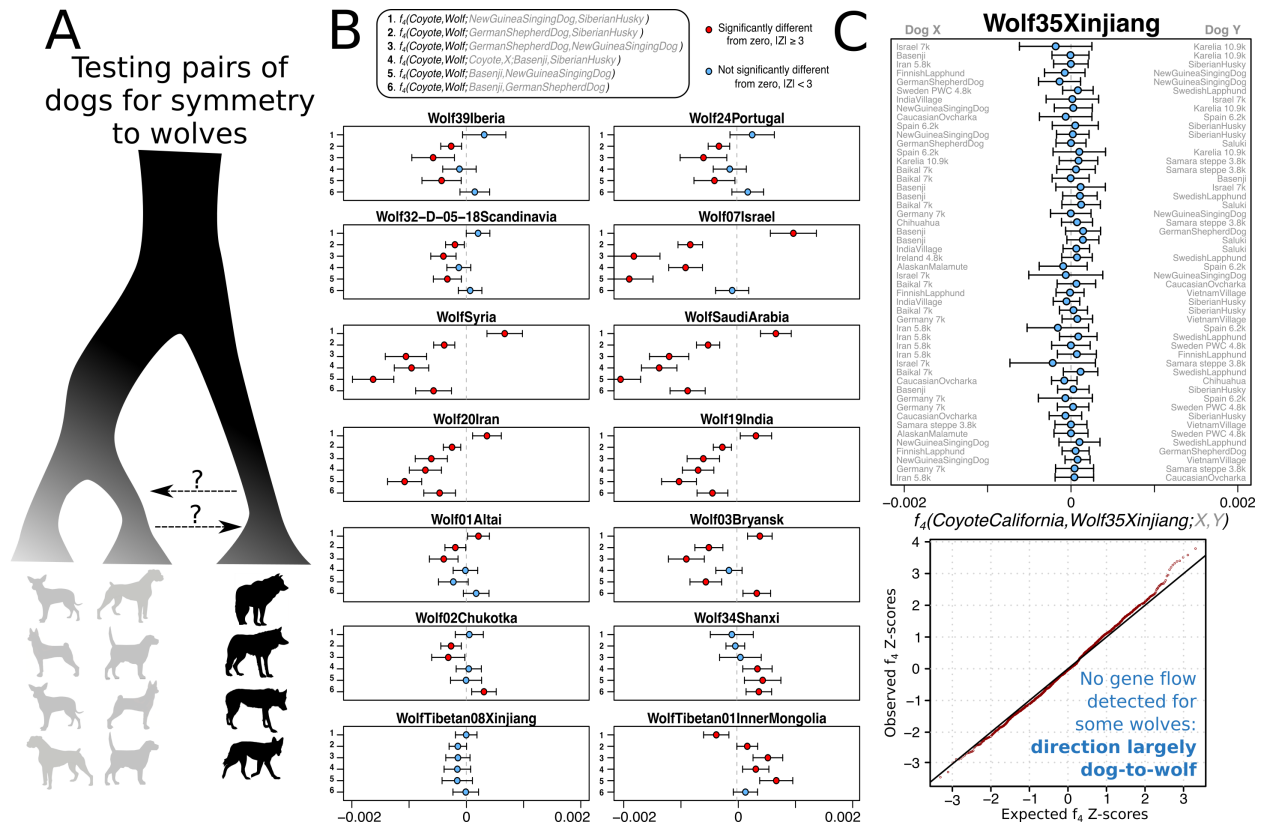
We find that the modern and ancient genomic data are consistent with a single origin for dogs, though a scenario involving multiple closely related wolf populations remains possible. However, in our view, the geographical origin of dogs remains unknown. Previously suggested

280 points of origin based upon present-day patterns of genomic diversity (2, 8, 10) or affinities to modern wolf populations (12) are sensitive to the obscuring effects of more recent population dynamics and gene flow. Ultimately, integrating DNA from dogs and wolves even older than those analyzed here with archaeology, anthropology, ethology and other disciplines, is needed to determine where, and in which environmental and cultural context the first dogs originated.

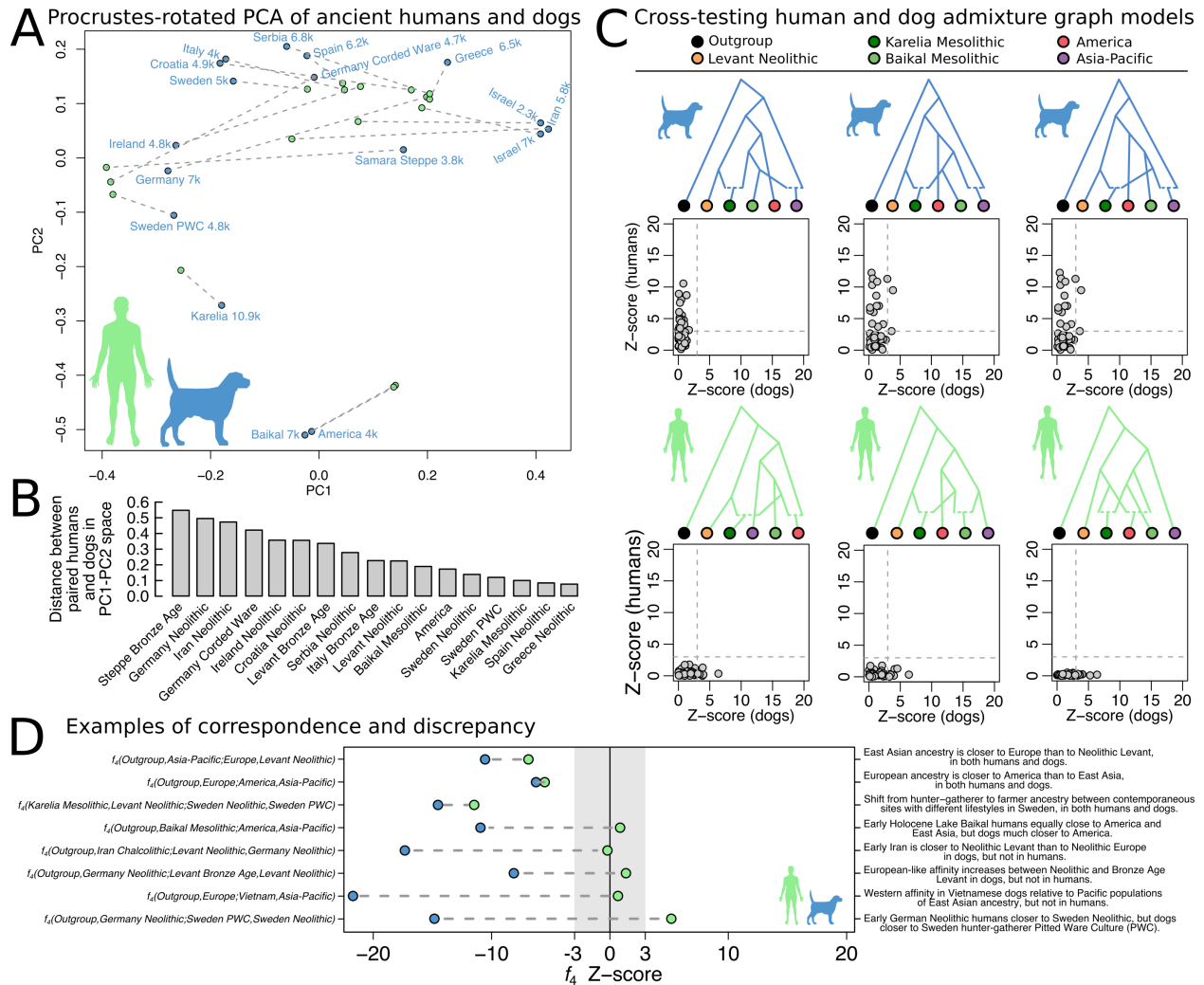


285 **Figure 1. Genomic structure of dogs dates to the Pleistocene.** A) Sampling locations of ancient dogs. B) Principal components analysis on all possible f_4 -statistics among ancient dogs (gray) and a selection of worldwide modern dogs. C) Outgroup f_3 -statistics reveal a cline of Levant versus Baikal (horizontal and vertical axes, respectively) related ancestry across ancient west Eurasian dogs, but not among modern European dogs. D) Coalescent simulations demonstrating that a diagonal f_3 -cline as in panel C is consistent with an admixture event, but less so with continuous gene flow and not with phylogenetic structure alone. E) An admixture graph that fits all f_4 -statistics between major dog lineages. The European dog was grafted onto the graph identified through exhaustive testing.

290

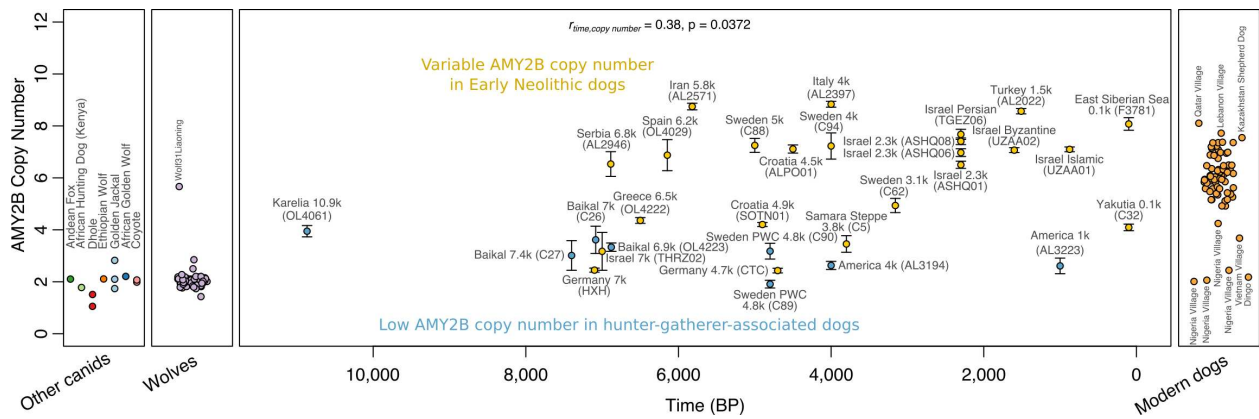


295 **Fig. 2. All detectable gene flow is consistent with being unidirectional from dogs into wolf**
populations. A) Illustration of asymmetry tests (f_4 -statistics) comparing 35 Eurasian gray wolves
 300 to all pairs of 66 ancient and modern dogs. B) Selected results using Coyote as outgroup. C) A
 wolf from Xinjiang, western China, is not closer to some dog populations than to others, as the
 test statistics are consistent with being normally distributed around 0 (the quantile-quantile plot
 includes all 66 dogs). If there was a substantial gene flow from some wolf population into some
 dog population, we would expect all wolf individuals to display asymmetric relationships.

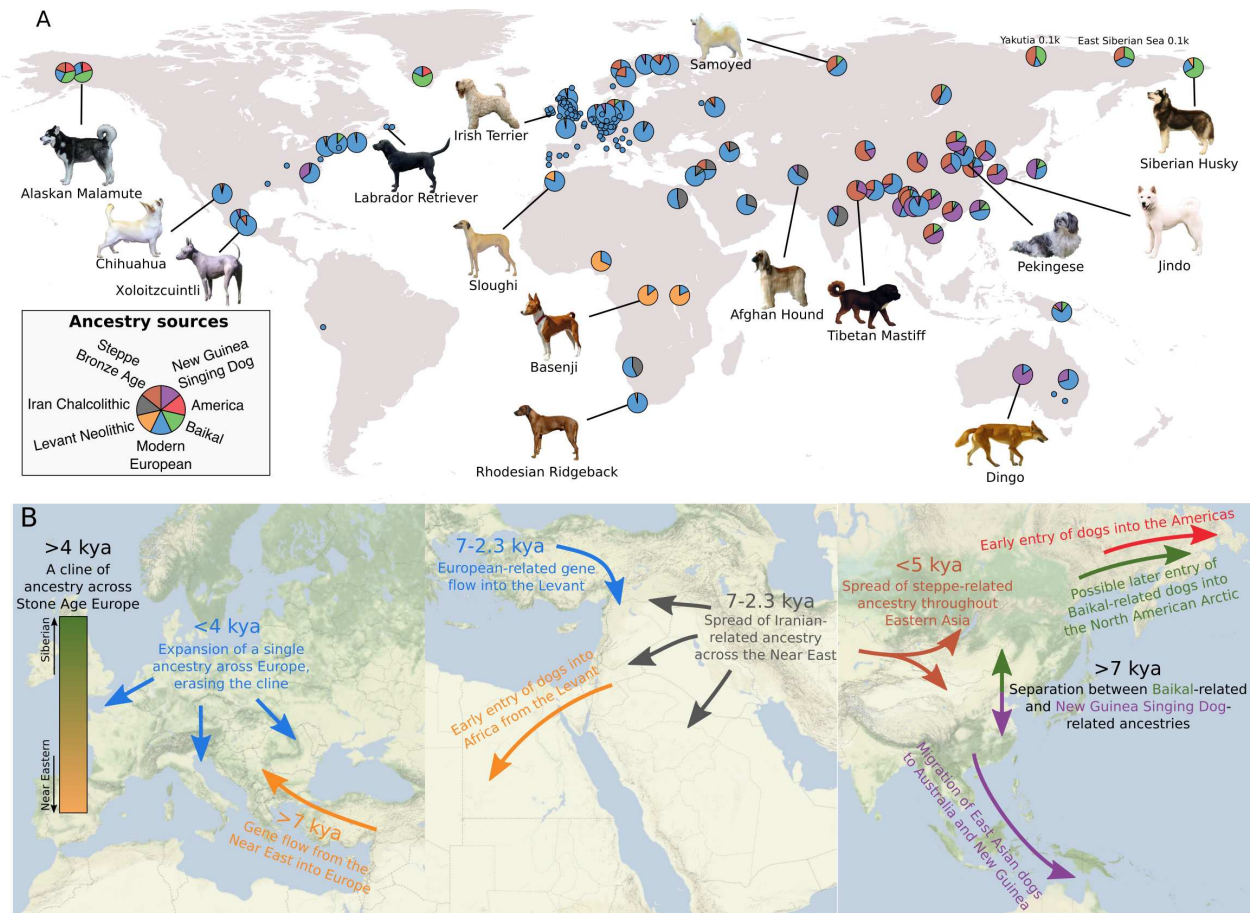


305 A) Principal components analysis on all possible f_4 -statistics on ancient dogs (blue), overlaid through Procrustes transformation by the corresponding analysis performed on ancient humans matched in time, space, and cultural context to the dogs (green). Dashed lines connect each matched pair. B) Euclidian residuals between the Procrustes-rotated human and dog coordinates. C) The three admixture graphs that fit for one species and provide the smallest error for the other. Scatter plots show absolute Z-scores for the difference between observed and predicted f_4 -statistics. D) Examples of f_4 -statistics that reveal similarities and differences between humans and dogs (far right text).

310



315 **Figure 4. Expansion of copy number in the *AMY2B* pancreatic amylase gene largely occurred after the transition to agriculture.** Ancient dogs are plotted against their age, with blue color indicating dogs from likely hunter-gatherer human contexts. Bars denote 95% binomial confidence intervals around the ratio of the number of reads mapping to the copy number variable region to those mapping to control regions throughout the genome.



320

Figure 5. Ancestry of global dogs today. A) For each present-day population, the ancestry proportions estimated by the best-fitting *qpAdm* model, restricted to models containing up to four of seven selected sources, are displayed. Populations for which a single component accounts for $\geq 98\%$ of the ancestry are collapsed to smaller circles. B) Illustrations of inferred population histories in three regions of the world.

325

References and notes

1. C. Vilà, P. Savolainen, J. E. Maldonado, I. R. Amorim, J. E. Rice, R. L. Honeycutt, K. A. Crandall, J. Lundeberg, R. K. Wayne, Multiple and ancient origins of the domestic dog. *Science*. **276**, 1687–1689 (1997).
330
2. P. Savolainen, Y. Zhang, J. Luo, J. Lundeberg, T. Leitner, Genetic evidence for an East Asian origin of domestic dogs. *Science*. **298**, 1610–1613 (2002).
3. M. Germonpré, M. V. Sablin, R. E. Stevens, R. E. M. Hedges, M. Hofreiter, M. Stiller, V. R. Després, 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 (2009).
335
4. P. Skoglund, E. Ersmark, E. Palkopoulou, L. Dalén, Ancient Wolf Genome Reveals an Early Divergence of Domestic Dog Ancestors and Admixture into High-Latitude Breeds. *Curr. Biol.* **25**, 1515–1519 (2015).
5. A. H. Freedman, I. Gronau, R. M. Schweizer, D. Ortega-Del Vecchyo, E. Han, P. M. Silva, M. Galaverni, Z. Fan, P. Marx, B. Lorente-Galdos, H. Beale, O. Ramirez, F. Hormozdiari, C. Alkan, C. Vilà, K. Squire, E. Geffen, J. Kusak, A. R. Boyko, H. G. Parker, C. Lee, V. Tadigotla, A. Siepel, C. D. Bustamante, T. T. Harkins, S. F. Nelson, E. A. Ostrander, T. Marques-Bonet, R. K. Wayne, J. Novembre, Genome Sequencing Highlights the Dynamic Early History of Dogs. *PLoS Genet.* **10**, e1004016 (2014).
340
6. E. Axelsson, A. Ratnakumar, M.-L. Arendt, K. Maqbool, M. T. Webster, M. Perloski, O. Liberg, J. M. Arnemo, A. Hedhammar, K. Lindblad-Toh, The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature*. **495**, 360–364 (2013).
345
7. P. Skoglund, A. Götherström, M. Jakobsson, Estimation of population divergence times from non-overlapping genomic sequences: examples from dogs and wolves. *Mol. Biol. Evol.* **28**, 1505–1517 (2011).
350
8. G.-D. Wang, W. Zhai, H.-C. Yang, L. Wang, L. Zhong, Y.-H. Liu, R.-X. Fan, T.-T. Yin, C.-L. Zhu, A. D. Poyarkov, D. M. Irwin, M. K. Hytonen, H. Lohi, C.-I. Wu, P. Savolainen, Y.-P. Zhang, Out of southern East Asia: the natural history of domestic dogs across the world. *Cell Res.* **26**, 21–33 (2016).
9. L. A. F. Frantz, V. E. Mullin, M. Pionnier-Capitan, O. Lebrasseur, M. Ollivier, A. Perri, A. Linderholm, V. Mattiangeli, M. D. Teasdale, E. A. Dimopoulos, A. Tresset, M. Duffraisse, F. McCormick, L. Bartosiewicz, E. Gál, É. A. Nyerges, M. V. Sablin, S. Bréhard, M. Mashkour, A. Bălăşescu, B. Gillet, S. Hughes, O. Chassaing, C. Hitte, J.-D. Vigne, K. Dobney, C. Hänni, D. G. Bradley, G. Larson, Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science*. **352**, 1228–1231 (2016).
355
360
10. L. M. Shannon, R. H. Boyko, M. Castelhano, E. Corey, J. J. Hayward, C. McLean, M. E. White, M. Abi Said, B. A. Anita, N. I. Bondjengo, J. Calero, A. Galov, M. Hedimbi, B. Imam, R. Khalap, D. Lally, A. Masta, K. C. Oliveira, L. Pérez, J. Randall, N. M. Tam, F. J. Trujillo-Cornejo, C. Valeriano, N. B. Sutter, R. J. Todhunter, C. D. Bustamante, A. R.

- 365 Boyko, Genetic structure in village dogs reveals a Central Asian domestication origin. *Proceedings of the National Academy of Sciences*. **112**, 13639–13644 (2015).
11. O. Thalmann, B. Shapiro, P. Cui, V. J. Schuenemann, S. K. Sawyer, D. L. Greenfield, M. B. Germonpré, M. V. Sablin, F. López-Giráldez, X. Domingo-Roura, H. Napierala, H.-P. Uerpmann, D. M. Loponte, A. A. Acosta, L. Giemisch, R. W. Schmitz, B. Worthington, J. E. Buikstra, A. Druzhkova, A. S. Graphodatsky, N. D. Ovodov, N. Wahlberg, A. H. Freedman, R. M. Schweizer, K.-P. Koepfli, J. A. Leonard, M. Meyer, J. Krause, S. Pääbo, R. E. Green, R. K. Wayne, Complete Mitochondrial Genomes of Ancient Canids Suggest a European Origin of Domestic Dogs. *Science*. **342**, 871–874 (2013).
- 370
12. B. M. vonHoldt, J. P. Pollinger, K. E. Lohmueller, E. Han, H. G. Parker, P. Quignon, J. D. Degenhardt, A. R. Boyko, D. A. Earl, A. Auton, A. Reynolds, K. Bryc, A. Brisbin, J. C. Knowles, D. S. Mosher, T. C. Spady, A. Elkahloun, E. Geffen, M. Pilot, W. Jedrzejewski, C. Greco, E. Randi, D. Bannasch, A. Wilton, J. Shearman, M. Musiani, M. Cargill, P. G. Jones, Z. Qian, W. Huang, Z.-L. Ding, Y.-P. Zhang, C. D. Bustamante, E. A. Ostrander, J. Novembre, R. K. Wayne, Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature*. **464**, 898–902 (2010).
- 375
13. J.-F. Pang, C. Kluetsch, X.-J. Zou, A.-B. Zhang, L.-Y. Luo, H. Angleby, A. Ardlan, C. Ekström, A. Sköllermo, J. Lundeberg, S. Matsamura, T. Leitner, Y.-P. Zhang, P. Savolainen, mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. *Mol. Biol. Evol.* **26**, 2849–2864 (2009).
- 380
14. L. R. Botigué, S. Song, A. Scheu, S. Gopalan, A. L. Pendleton, M. Oetjens, A. M. Taravella, T. Seregély, A. Zeeb-Lanz, R.-M. Arbogast, D. Bobo, K. Daly, M. Unterländer, J. Burger, J. M. Kidd, K. R. Veeramah, Ancient European dog genomes reveal continuity since the Early Neolithic. *Nat. Commun.* **8**, 16082 (2017).
- 385
15. D. F. Morey, The Early Evolution of the Domestic Dog. *Am. Sci.* **82**, 336–347 (1994).
- 390
16. J. Clutton-Brock, Man-made dogs. *Science*. **197**, 1340–1342 (1977).
17. S. J. M. Davis, F. R. Valla, Evidence for domestication of the dog 12,000 years ago in the Natufian of Israel. *Nature*. **276**, 608–610 (1978).
18. M. Sablin, G. Khlopachev, The Earliest Ice Age Dogs: Evidence from Eliseevichi 11. *Curr. Anthropol.* **43**, 795–799 (2002).
- 395
19. H. G. Parker, L. V. Kim, N. B. Sutter, S. Carlson, T. D. Lorentzen, T. B. Malek, G. S. Johnson, H. B. DeFrance, E. A. Ostrander, L. Kruglyak, Genetic structure of the purebred domestic dog. *Science*. **304**, 1160–1164 (2004).
- 400
20. H. G. Parker, D. L. Dreger, M. Rimbault, B. W. Davis, A. B. Mullen, G. Carpintero-Ramirez, E. A. Ostrander, Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development. *Cell Rep.* **19**, 697–708 (2017).

21. M. Ní Leathlobhair, A. R. Perri, E. K. Irving-Pease, K. E. Witt, A. Linderholm, J. Haile, O. Lebrasseur, C. Ameen, J. Blick, A. R. Boyko, S. Brace, Y. N. Cortes, S. J. Crockford, A. Devault, E. A. Dimopoulos, M. Eldridge, J. Enk, S. Gopalakrishnan, K. Gori, V. Grimes, E. Guiry, A. J. Hansen, A. Hulme-Beaman, J. Johnson, A. Kitchen, A. K. Kasparov, Y.-M. Kwon, P. A. Nikolskiy, C. P. Lope, A. Manin, T. Martin, M. Meyer, K. N. Myers, M. Omura, J.-M. Rouillard, E. Y. Pavlova, P. Sciulli, M.-H. S. Sinding, A. Strakova, V. V. Ivanova, C. Widga, E. Willerslev, V. V. Pitulko, I. Barnes, M. T. P. Gilbert, K. M. Dobney, R. S. Malhi, E. P. Murchison, G. Larson, L. A. F. Frantz, The evolutionary history of dogs in the Americas. *Science*. **361**, 81–85 (2018).
22. B. van Asch, A.-B. Zhang, M. C. R. Oskarsson, C. F. C. Klütsch, A. Amorim, P. Savolainen, Pre-Columbian origins of Native American dog breeds, with only limited replacement by European dogs, confirmed by mtDNA analysis. *Proceedings of the Royal Society B: Biological Sciences*. **280** (2013).
23. J. A. Leonard, R. K. Wayne, J. Wheeler, R. Valadez, S. Guillen, C. Vila, Ancient DNA evidence for Old World origin of New World dogs. *Science*. **298**, 1613–1616 (2002).
24. S. Castroviejo-Fisher, P. Skoglund, R. Valadez, C. Vila, J. Leonard, Vanishing native American dog lineages. *BMC Evol. Biol.* **11**, 73 (2011).
25. K. Greig, A. Gosling, C. J. Collins, J. Boocock, K. McDonald, D. J. Addison, M. S. Allen, B. David, M. Gibbs, C. F. W. Higham, F. Liu, I. J. McNiven, S. O’Connor, C. H. Tsang, R. Walter, E. Matisoo-Smith, Complex history of dog (*Canis familiaris*) origins and translocations in the Pacific revealed by ancient mitogenomes. *Sci. Rep.* **8**, 9130 (2018).
26. P. Savolainen, T. Leitner, A. N. Wilton, E. Matisoo-Smith, J. Lundeberg, A detailed picture of the origin of the Australian dingo, obtained from the study of mitochondrial DNA. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 12387–12390 (2004).
27. M. Ollivier, A. Tresset, L. A. F. Frantz, S. Bréhard, A. Bălăşescu, M. Mashkour, A. Boroneanţ, M. Pionnier-Capitan, O. Lebrasseur, R.-M. Arbogast, Others, Dogs accompanied humans during the Neolithic expansion into Europe. *Biol. Lett.* **14**, 20180286 (2018).
28. C. Ameen, T. Feuerborn, A. Linderholm, S. Brown, A. Hulme-Beaman, O. Lebrasseur, M.-H. S. Sinding, Z. T. Lounsbury, A. Lin, M. Appelt, Others, Specialised sledge dogs accompanied Inuit dispersal across the North American Arctic. *Proceedings of the Royal Society B: Biological Sciences* (2019) (available at <https://abdn.pure.elsevier.com/en/publications/specialised-sledge-dogs-accompanied-inuit-dispersal-across-the-no>).
29. H. Malmström, C. Vilà, M. T. P. Gilbert, J. Storå, E. Willerslev, G. Holmlund, A. Götherström, Barking up the wrong tree: modern northern European dogs fail to explain their origin. *BMC Evol. Biol.* **8**, 71 (2008).
30. Materials and methods are available as supplementary materials.

- 440 31. N. D. Ovodov, S. J. Crockford, Y. V. Kuzmin, T. F. G. Higham, G. W. L. Hodgins, J. van
der Plicht, A 33,000-Year-Old Incipient Dog from the Altai Mountains of Siberia: Evidence
of the Earliest Domestication Disrupted by the Last Glacial Maximum. *PLoS One*. **6**,
e22821 (2011).
- 445 32. Y.-H. Liu, L. Wang, T. Xu, X. Guo, Y. Li, T.-T. Yin, H.-C. Yang, Y. Hu, A. C. Adeola, O.
J. Sanke, N. O. Otecko, M. Wang, Y. Ma, O. S. Charles, M.-H. S. Sinding, S.
Gopalakrishnan, J. Alfredo Samaniego, A. J. Hansen, C. Fernandes, P. Gaubert, J. Budd, P.
M. Dawuda, E. Knispel Rueness, L. Jiang, W. Zhai, M. T. P. Gilbert, M.-S. Peng, X. Qi, G.-
D. Wang, Y.-P. Zhang, Whole-Genome Sequencing of African Dogs Provides Insights into
Adaptations against Tropical Parasites. *Mol. Biol. Evol.* **35**, 287–298 (2018).
- 450 33. B. Miao, Z. Wang, Y. Li, Genomic Analysis Reveals Hypoxia Adaptation in the Tibetan
Mastiff by Introgression of the Gray Wolf from the Tibetan Plateau. *Mol. Biol. Evol.* **34**,
734–743 (2017).
34. B. vonHoldt, Z. Fan, D. Ortega-Del Vecchyo, R. K. Wayne, EPAS1 variants in high altitude
Tibetan wolves were selectively introgressed into highland dogs. *PeerJ*. **5**, e3522 (2017).
- 455 35. L. A. F. Frantz, J. Haile, A. T. Lin, A. Scheu, Ancient pigs reveal a near-complete genomic
turnover following their introduction to Europe. *Proceedings of the* (2019) (available at
<https://www.pnas.org/content/116/35/17231.short>).
36. K. G. Daly, P. Maisano Delser, V. E. Mullin, A. Scheu, V. Mattiangeli, M. D. Teasdale, A.
J. Hare, J. Burger, M. P. Verdugo, M. J. Collins, R. Kehati, C. M. Ereik, G. Bar-Oz, F.
460 Pompanon, T. Cumer, C. Çakırlar, A. F. Mohaseb, D. Decruyenaere, H. Davoudi, Ö. Çevik,
G. Rollefson, J.-D. Vigne, R. Khazaeli, H. Fathi, S. B. Doost, R. Rahimi Sorkhani, A. A.
Vahdati, E. W. Sauer, H. Azizi Kharanaghi, S. Maziar, B. Gasparian, R. Pinhasi, L. Martin,
D. Orton, B. S. Arbuckle, N. Benecke, A. Manica, L. K. Horwitz, M. Mashkour, D. G.
Bradley, Ancient goat genomes reveal mosaic domestication in the Fertile Crescent.
465 *Science*. **361**, 85–88 (2018).
37. A. Fages, K. Hanghøj, N. Khan, C. Gaunitz, A. Seguin-Orlando, M. Leonardi, C. McCrory
Constantz, C. Gamba, K. A. S. Al-Rasheid, S. Albizuri, A. H. Alfarhan, M. Allentoft, S.
Alquraishi, D. Anthony, N. Baimukhanov, J. H. Barrett, J. Bayarsaikhan, N. Benecke, E.
Bernáldez-Sánchez, L. Berrocal-Rangel, F. Biglari, S. Boessenkool, B. Boldgiv, G. Brem,
470 D. Brown, J. Burger, E. Crubézy, L. Daugnora, H. Davoudi, P. de Barros Damgaard, M. de
Los Ángeles de Chorro Y de Villa-Ceballos, S. Deschler-Erb, C. Detry, N. Dill, M. do Mar
Oom, A. Dohr, S. Ellingvåg, D. Erdenebaatar, H. Fathi, S. Felkel, C. Fernández-Rodríguez,
E. García-Viñas, M. Germonpré, J. D. Granado, J. H. Hallsson, H. Hemmer, M. Hofreiter,
475 A. Kasparov, M. Khasanov, R. Khazaeli, P. Kosintsev, K. Kristiansen, T. Kubatbek, L.
Kuderna, P. Kuznetsov, H. Laleh, J. A. Leonard, J. Lhuillier, C. Liesau von Lettow-
Vorbeck, A. Logvin, L. Lõugas, A. Ludwig, C. Luis, A. M. Arruda, T. Marques-Bonet, R.
Matoso Silva, V. Merz, E. Mijiddorj, B. K. Miller, O. Monchalov, F. A. Mohaseb, A.
Morales, A. Nieto-Espinet, H. Nistelberger, V. Onar, A. H. Pálsdóttir, V. Pitulko, K.
Pitskhelauri, M. Pruvost, P. Rajic Sikanjic, A. Rapan Papeša, N. Roslyakova, A. Sardari, E.
480 Sauer, R. Schafberg, A. Scheu, J. Schibler, A. Schlumbaum, N. Serrand, A. Serres-Armero,

- 485 B. Shapiro, S. Sheikhi Seno, I. Shevnina, S. Shidrang, J. Southon, B. Star, N. Sykes, K. Taheri, W. Taylor, W.-R. Teegen, T. Trbojević Vukičević, S. Trixl, D. Tumen, S. Undrakhbold, E. Usmanova, A. Vahdati, S. Valenzuela-Lamas, C. Viegas, B. Wallner, J. Weinstock, V. Zaibert, B. Clavel, S. Lepetz, M. Mashkour, A. Helgason, K. Stefánsson, E. Barrey, E. Willerslev, A. K. Outram, P. Librado, L. Orlando, Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. *Cell*. **177**, 1419–1435.e31 (2019).
- 490 38. M. Barbato, F. Hailer, P. Orozco-terWengel, J. Kijas, P. Mereu, P. Cabras, R. Mazza, M. Pirastru, M. W. Bruford, Genomic signatures of adaptive introgression from European mouflon into domestic sheep. *Sci. Rep.* **7**, 7623 (2017).
- 495 39. S. D. E. Park, D. A. Magee, P. A. McGettigan, M. D. Teasdale, C. J. Edwards, A. J. Lohan, A. Murphy, M. Braud, M. T. Donoghue, Y. Liu, A. T. Chamberlain, K. Rue-Albrecht, S. Schroeder, C. Spillane, S. Tai, D. G. Bradley, T. S. Sonstegard, B. J. Loftus, D. E. MacHugh, Genome sequencing of the extinct Eurasian wild aurochs, *Bos primigenius*, illuminates the phylogeography and evolution of cattle. *Genome Biol.* **16**, 234 (2015).
- 500 40. M. P. Verdugo, V. E. Mullin, A. Scheu, V. Mattiangeli, K. G. Daly, P. Maisano Delser, A. J. Hare, J. Burger, M. J. Collins, R. Kehati, P. Hesse, D. Fulton, E. W. Sauer, F. A. Mohaseb, H. Davoudi, R. Khazaeli, J. Lhuillier, C. Rapin, S. Ebrahimi, M. Khasanov, S. M. F. Vahidi, D. E. MacHugh, O. Ertuğrul, C. Koukouli-Chrysanthaki, A. Sampson, G. Kazantzis, I. Kontopoulos, J. Bulatovic, I. Stojanović, A. Mikdad, N. Benecke, J. Linstädter, M. Sablin, R. Bendrey, L. Gourichon, B. S. Arbuckle, M. Mashkour, D. Orton, L. K. Horwitz, M. D. Teasdale, D. G. Bradley, Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. *Science*. **365**, 173–176 (2019).
- 505 41. I. Lazaridis, D. Nadel, G. Rollefson, D. C. Merrett, N. Rohland, S. Mallick, D. Fernandes, M. Novak, B. Gamarra, K. Sirak, S. Connell, K. Stewardson, E. Harney, Q. Fu, G. Gonzalez-Fortes, E. R. Jones, S. A. Roodenberg, G. Lengyel, F. Bocquentin, B. Gasparian, J. M. Monge, M. Gregg, V. Eshed, A.-S. Mizrahi, C. Meiklejohn, F. Gerritsen, L. Bejenaru, M. Blüher, A. Campbell, G. Cavalleri, D. Comas, P. Froguel, E. Gilbert, S. M. Kerr, P. Kovacs, J. Krause, D. McGettigan, M. Merrigan, D. A. Merriwether, S. O'Reilly, M. B. Richards, O. Semino, M. Shamoony-Pour, G. Stefanescu, M. Stumvoll, A. Tönjes, A. Torroni, J. F. Wilson, L. Yengo, N. A. Hovhannisyan, N. Patterson, R. Pinhasi, D. Reich, Genomic insights into the origin of farming in the ancient Near East. *Nature*. **536**, 419–424 (2016).
- 510 42. F. Broushaki, M. G. Thomas, V. Link, S. López, L. van Dorp, K. Kirsanow, Z. Hofmanová, Y. Diekmann, L. M. Cassidy, D. Díez-del-Molino, A. Kousathanas, C. Sell, H. K. Robson, R. Martiniano, J. Blöcher, A. Scheu, S. Kreutzer, R. Bollongino, D. Bobo, H. Davoudi, O. Munoz, M. Currat, K. Abdi, F. Biglari, O. E. Craig, D. G. Bradley, S. Shennan, K. R. Veeramah, M. Mashkour, D. Wegmann, G. Hellenthal, J. Burger, Early Neolithic genomes from the eastern Fertile Crescent. *Science*. **353**, 499 (2016).
- 520 43. I. Lazaridis, N. Patterson, A. Mittnik, G. Renaud, S. Mallick, K. Kirsanow, P. H. Sudmant, J. G. Schraiber, S. Castellano, M. Lipson, B. Berger, C. Economou, R. Bollongino, Q. Fu,

- 525 K. I. Bos, S. Nordenfelt, H. Li, C. de Filippo, K. Prufer, S. Sawyer, C. Posth, W. Haak, F. Hallgren, E. Fornander, N. Rohland, D. Delsate, M. Francken, J.-M. Guinet, J. Wahl, G. Ayodo, H. A. Babiker, G. Bailliet, E. Balanovska, O. Balanovsky, R. Barrantes, G. Bedoya, H. Ben-Ami, J. Bene, F. Berrada, C. M. Bravi, F. Brisighelli, G. B. J. Busby, F. Cali, M. Churnosov, D. E. C. Cole, D. Corach, L. Damba, G. van Driem, S. Dryomov, J.-M. Dugoujon, S. A. Fedorova, I. Gallego Romero, M. Gubina, M. Hammer, B. M. Henn, T. Hervig, U. Hodoglugil, A. R. Jha, S. Karachanak-Yankova, R. Khusainova, E. Khusnutdinova, R. Kittles, T. Kivisild, W. Klitz, V. Kucinkas, A. Kushniarevich, L. Laredj, S. Litvinov, T. Loukidis, R. W. Mahley, B. Melegh, E. Metspalu, J. Molina, J. Mountain, K. Nakkalajarvi, D. Nesheva, T. Nyambo, L. Osipova, J. Parik, F. Platonov, O. Posukh, V. Romano, F. Rothhammer, I. Rudan, R. Ruizbakiev, H. Sahakyan, A. Sajantila, A. Salas, E. B. Starikovskaya, A. Tarekegn, D. Toncheva, S. Turdikulova, I. Uktveryte, O. Utevska, R. Vasquez, M. Villena, M. Voevoda, C. A. Winkler, L. Yepiskoposyan, P. Zalloua, T. Zemunik, A. Cooper, C. Capelli, M. G. Thomas, A. Ruiz-Linares, S. A. Tishkoff, L. Singh, K. Thangaraj, R. Villems, D. Comas, R. Sukernik, M. Metspalu, M. Meyer, E. E. Eichler, J. Burger, M. Slatkin, S. Paabo, J. Kelso, D. Reich, J. Krause, Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*. **513**, 409–413 (2014).
- 540 44. L. M. Cassidy, R. Martiniano, E. M. Murphy, M. D. Teasdale, J. Mallory, B. Hartwell, D. G. Bradley, Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. *Proceedings of the National Academy of Sciences*. **113**, 368–373 (2016).
- 545 45. W. Haak, I. Lazaridis, N. Patterson, N. Rohland, S. Mallick, B. Llamas, G. Brandt, S. Nordenfelt, E. Harney, K. Stewardson, Q. Fu, A. Mittnik, E. Bánffy, C. Economou, M. Francken, S. Friederich, R. G. Pena, F. Hallgren, V. Khartanovich, A. Khokhlov, M. Kunst, P. Kuznetsov, H. Meller, O. Mochalov, V. Moiseyev, N. Nicklisch, S. L. Pichler, R. Risch, M. A. Rojo Guerra, C. Roth, A. Szécsényi-Nagy, J. Wahl, M. Meyer, J. Krause, D. Brown, D. Anthony, A. Cooper, K. W. Alt, D. Reich, Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. **522**, 207 (2015).
- 550 46. M. E. Allentoft, M. Sikora, K.-G. Sjogren, S. Rasmussen, M. Rasmussen, J. Stenderup, P. B. Damgaard, H. Schroeder, T. Ahlstrom, L. Vinner, A.-S. Malaspinas, A. Margaryan, T. Higham, D. Chivall, N. Lynnerup, L. Harvig, J. Baron, P. D. Casa, P. Dabrowski, P. R. Duffy, A. V. Ebel, A. Epimakhov, K. Frei, M. Furmanek, T. Gralak, A. Gromov, S. Gronkiewicz, G. Grupe, T. Hajdu, R. Jarysz, V. Khartanovich, A. Khokhlov, V. Kiss, J. Kolar, A. Kriiska, I. Lasak, C. Longhi, G. McGlynn, A. Merkevcicius, I. Merkyte, M. Metspalu, R. Mkrtychyan, V. Moiseyev, L. Paja, G. Palfi, D. Pokutta, L. Pospieszny, T. D. Price, L. Saag, M. Sablin, N. Shishlina, V. Smrcka, V. I. Soenov, V. Szeverenyi, G. Toth, S. V. Trifanova, L. Varul, M. Vicze, L. Yepiskoposyan, V. Zhitenev, L. Orlando, T. Sicheritz-Ponten, S. Brunak, R. Nielsen, K. Kristiansen, E. Willerslev, Population genomics of Bronze Age Eurasia. *Nature*. **522**, 167–172 (2015).
- 560 47. Q. Fu, M. Meyer, X. Gao, U. Stenzel, H. A. Burbano, J. Kelso, S. Pääbo, DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the National Academy of Sciences*. **110**, 2223–2227 (2013).

- 565 48. L. Janssens, L. Giemsch, R. Schmitz, M. Street, S. Van Dongen, P. Crombé, A new look at an old dog: Bonn-Oberkassel reconsidered. *J. Archaeol. Sci.* **92**, 126–138 (2018).
49. A. Perri, A wolf in dog's clothing: Initial dog domestication and Pleistocene wolf variation. *J. Archaeol. Sci.* **68**, 1–4 (2016).
50. D. F. Morey, In search of Paleolithic dogs: a quest with mixed results. *J. Archaeol. Sci.* **52**, 300–307 (2014).
- 570 51. M. Raghavan, P. Skoglund, K. E. Graf, M. Metspalu, A. Albrechtsen, I. Moltke, S. Rasmussen, T. W. Stafford Jr, L. Orlando, E. Metspalu, M. Karmin, K. Tambets, S. Rootsi, R. Magi, P. F. Campos, E. Balanovska, O. Balanovsky, E. Khusnutdinova, S. Litvinov, L. P. Osipova, S. A. Fedorova, M. I. Voevoda, M. DeGiorgio, T. Sicheritz-Ponten, S. Brunak, S. Demeshchenko, T. Kivisild, R. Villems, R. Nielsen, M. Jakobsson, E. Willerslev, Upper
575 Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature.* **505**, 87–91 (2014).
52. P. de B. Damgaard, N. Marchi, S. Rasmussen, M. Peyrot, G. Renaud, T. Korneliussen, J. V. Moreno-Mayar, M. W. Pedersen, A. Goldberg, E. Usmanova, N. Baimukhanov, V. Loman, L. Hedeager, A. G. Pedersen, K. Nielsen, G. Afanasiev, K. Akmatov, A. Aldashev, A.
580 Alpaslan, G. Baimbetov, V. I. Bazaliiskii, A. Beisenov, B. Boldbaatar, B. Boldgiv, C. Dorzhu, S. Ellingvag, D. Erdenebaatar, R. Dajani, E. Dmitriev, V. Evdokimov, K. M. Frei, A. Gromov, A. Goryachev, H. Hakonarson, T. Hegay, Z. Khachatryan, R. Khaskhanov, E. Kitov, A. Kolbina, T. Kubatbek, A. Kukushkin, I. Kukushkin, N. Lau, A. Margaryan, I. Merkyte, I. V. Mertz, V. K. Mertz, E. Mijiddorj, V. Moiyesev, G. Mukhtarova, B.
585 Nurmukhanbetov, Z. Orozbekova, I. Panyushkina, K. Pieta, V. Smrčka, I. Shevnina, A. Logvin, K.-G. Sjögren, T. Štolcová, A. M. Taravella, K. Tashbaeva, A. Tkachev, T. Tulegenov, D. Voyakin, L. Yepiskoposyan, S. Undrakhbold, V. Varfolomeev, A. Weber, M. A. Wilson Sayres, N. Krادين, M. E. Allentoft, L. Orlando, R. Nielsen, M. Sikora, E. Heyer, K. Kristiansen, E. Willerslev, 137 ancient human genomes from across the Eurasian
590 steppes. *Nature.* **557**, 369–374 (2018).
53. P. Skoglund, H. Malmström, M. Raghavan, J. Storå, P. Hall, E. Willerslev, M. T. P. Gilbert, A. Götherström, M. Jakobsson, Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science.* **336**, 466–469 (2012).
54. P. Skoglund, H. Malmström, A. Omrak, M. Raghavan, C. Valdiosera, T. Günther, P. Hall,
595 K. Tambets, J. Parik, K.-G. Sjögren, J. Apel, E. Willerslev, J. Storå, A. Götherström, M. Jakobsson, Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and Farmers. *Science.* **344**, 747–750 (2014).
55. M. Arendt, K. M. Cairns, J. W. O. Ballard, P. Savolainen, E. Axelsson, Diet adaptation in dog reflects spread of prehistoric agriculture. *Heredity* . **117**, 301–306 (2016).
- 600 56. M. Ollivier, A. Tresset, F. Bastian, L. Lagoutte, E. Axelsson, M.-L. Arendt, A. Bălăşescu, M. Marshour, M. V. Sablin, L. Salanova, J.-D. Vigne, C. Hitte, C. Hänni, Amy2B copy number variation reveals starch diet adaptations in ancient European dogs. *R Soc Open Sci.* **3**, 160449 (2016).

- 605 57. G. H. Perry, N. J. Dominy, K. G. Claw, A. S. Lee, H. Fiegler, R. Redon, J. Werner, F. A. Villanea, J. L. Mountain, R. Misra, N. P. Carter, C. Lee, A. C. Stone, Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* **39**, 1256 (2007).
58. S. Mathieson, I. Mathieson, FADS1 and the Timing of Human Adaptation to Agriculture. *Mol. Biol. Evol.* **35**, 2957–2970 (2018).
- 610 59. P. Skoglund, J. C. Thompson, M. E. Prendergast, A. Mittnik, K. Sirak, M. Hajdinjak, T. Salie, N. Rohland, S. Mallick, A. Peltzer, A. Heinze, I. Olalde, M. Ferry, E. Harney, M. Michel, K. Stewardson, J. I. Cerezo-Román, C. Chiumia, A. Crowther, E. Gomani-Chindebvu, A. O. Gidna, K. M. Grillo, I. T. Helenius, G. Hellenthal, R. Helm, M. Horton, S. López, A. Z. P. Mabulla, J. Parkington, C. Shipton, M. G. Thomas, R. Tibesasa, M. Welling, V. M. Hayes, D. J. Kennett, R. Ramesar, M. Meyer, S. Pääbo, N. Patterson, A. G. 615 Morris, N. Boivin, R. Pinhasi, J. Krause, D. Reich, Reconstructing Prehistoric African Population Structure. *Cell.* **171**, 59–71.e21 (2017).
60. M. Feldman, D. M. Master, R. A. Bianco, M. Burri, P. W. Stockhammer, A. Mittnik, A. J. Aja, C. Jeong, J. Krause, Ancient DNA sheds light on the genetic origins of early Iron Age Philistines. *Sci Adv.* **5**, eaax0061 (2019).
- 620 61. M. Zhang, G. Sun, L. Ren, H. Yuan, G. Dong, L. Zhang, F. Liu, P. Cao, A. M.-S. Ko, M. A. Yang, S. Hu, G.-D. Wang, Q. Fu, Ancient DNA Evidence from China Reveals the Expansion of Pacific Dogs. *Mol. Biol. Evol.* **37**, 1462–1469 (2020).
62. A. Sampson, Ed., *Σκοτεινή Θαρρουνίων. Το σπήλαιο. Ο οικισμός και το νεκροταφείο (Skoteini, Tharrounia. The cave, the settlement and the cemetery)* (Athens, Ephorate of 625 Paleanthropology - Speleology, 1993).
63. E. Kotjabopoulou, K. Trantalidou, Faunal analysis of the Skoteini cave. In: *Sampson A, ed. Skoteini, Tharrounia: the Cave, the Settlement and the Cemetery*, 392–434 (1993).
64. M. Abdollahi, Z. A. Sardari, EASTERN CENTRAL ZAGROS DURING THE NEOLITHIC PERIOD: BASED ON THE EXCAVATION AT TAPPEH QELA GAP. 630 *PAZHOHESH-HA-YE BASTANSHENASI IRAN.* **3**, 117-138 (2013).
65. S. Amiri, M. Mashkour, A. Mohaseb, M. Tengberg, M. Abdolahi, A. Sardari, The Subsistence Economy of Qela Gap; Lorestan, Iran: From the Late Neolithic to the Iron Age. *Archaeology*, 125–132 (2019).
66. S. Amiri, M. Mashkour, A. Mohaseb, M. Tengberg, M. Abdolahi, A. Sardari, in *Proceeding of the International Conference of Young Archaeologists, The Faculty of Litterature and 635 Humanities and the Cultural Division of University of Tehran*, A. K. M. H, M. Hkanipour, R. Naseri, Eds. (2014), pp. 597–626.
67. J. de G. Mazzorin, A. Tagliacozzo, Morphological and osteological changes in the dog from the Neolithic to the Roman period in Italy. *Dogs through time: an archaeological 640 perspective.* Oxford: *Archaeopress*, 141–161 (2000).

68. V. Onar, C. Çakırlar, M. Janeczek, Z. Kızıltan, Skull Typology of Byzantine Dogs from the Theodosius Harbour at Yenikapı, Istanbul. *Anat. Histol. Embryol.* **41**, 341–352 (2012).
69. V. Onar, G. Pazvant, H. Alpak, N. G. Ince, A. Armutak, Z. S. Kiziltan, Animal skeletal remains of the Theodosius harbor: general overview. *Turkish Journal of Veterinary and Animal Sciences.* **37**, 81–85 (2013).
- 645
70. V. Onar, M. Janeczek, G. Pazvant, N. Gezer Ince, H. Alpak, A. Armutak, A. Chrószcz, Z. Kızıltan, Estimating the body weight of Byzantine dogs from the Theodosius Harbour at Yenikapı, Istanbul. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi.* **21**, 55–59 (2015).
71. S. V. Oshibkina, The Mesolithic culture Veret'ye. Chronology and periods. *Rossiiskaya Arkheologiya*, 100–110 (2004).
- 650
72. S. V. Oshibkina, Mezolit Vostochnogo Prionezhya. Kultura Veretye. *Moscow: Institute of Archaeology, Russian Academy of Sciences* (2006).
73. K. G. Sjögren, in *Aspects of Neolithic burial practices. I: K Brink, S. Hydén, K. Jennbert, L. Larsson, D. Olausson (eds.): Neolithic Diversities. Perspectives from a conference in Lund, Sweden. Lund* (2015), pp. 200–212.
- 655
74. I. Ullén, Horse and dog in the Swedish Bronze Age: A close-up study of the relation of horse and dog to man in the Bronze Age settlement of Apalle. *Archäologisches Korrespondenzblatt.* **26**, 145–166 (1996).
75. I. Ullén, Lager och hus. *Bronsåldersboplatsen vid Apalle i Uppland. Arkeologi på väg--undersökningar för E18*, 22–75 (1997).
- 660
76. R. J. Losey, S. Garvie-Lok, J. A. Leonard, M. Anne Katzenberg, M. Germonpré, T. Nomokonova, M. V. Sablin, O. I. Goriunova, N. E. Berdnikova, N. A. Savel'ev, Burying Dogs in Ancient Cis-Baikal, Siberia: Temporal Trends and Relationships with Human Diet and Subsistence Practices. *PLoS One.* **8**, e63740 (2013).
- 665
77. R. J. Losey, V. I. Bazaliiskii, S. Garvie-Lok, M. Germonpré, J. A. Leonard, A. L. Allen, M. Anne Katzenberg, M. V. Sablin, Canids as persons: Early Neolithic dog and wolf burials, Cis-Baikal, Siberia. *Journal of Anthropological Archaeology.* **30**, 174–189 (2011).
78. A. Cava, El depósito arqueológico de la cueva de Marizulo (Guipúzcoa). *Munibe.* **30**, 155–172 (1978).
- 670
79. J. M. Basabe, Restos humanos del yacimiento de Marizulo. *Munibe.* **23**, 104–124 (1971).
80. J. Altuna, *Fauna de mamíferos del yacimiento prehistórico de Marizulo (Urnieta), Guipúzcoa* (Sociedad de Ciencias Naturales Aranzadi, 1967).
81. J. Bulatović, Arheozoološki aspekti društvenih i kulturnih promena na Centralnom Balkanu u petom milenijumu pre nove ere. *PhD thesis, Faculty of Philosophy, University of Belgrade* (2018).
- 675

82. M. Radivojević, B. W. Roberts, M. Marić, J. Kuzmanović-Cvetković, T. Rehren, *The Rise of Metallurgy in Eurasia: The Archaeology of Early Metallurgy and Society in the Central Balkans*. UCL Press, London (2018).
83. M. V. Garašanin, *Hronologija vinčanske grupe*. Ljubljana: Univerza v Ljubljana (1951).
- 680 84. M. Radivojević, J. Kuzmanović-Cvetković, Copper minerals and archaeometallurgical materials from the Vinča culture sites of Belovode and Pločnik: Overview of the evidence and new data. *Starinar*. **2014**, 7–30 (2014).
85. B. Stalio, Pločnik-Prokuplje-naselje. *Arheološki pregled*. **2** (1960).
- 685 86. D. Šljivar, J. Kuzmanović-Cvetković, Pločnik kod Prokuplja, istraživanja u 1997. *Glasnik Srpskog arheološkog društva*. **14**, 79–85 (1998).
87. A. Whittle, A. Bayliss, A. Barclay, B. Gaydarska, E. Bánffy, D. Borić, F. Draşovean, J. Jakucs, M. Marić, D. C. Orton, I. Pantović, W. Schier, N. Tasić, M. Vander Linden, A Vinča potscape: : formal chronological models for the use and development of Vinča ceramics in south-east Europe. *Documenta Praehistorica*, 60 (2016).
- 690 88. D. W. Anthony, D. R. Brown, The dogs of war: A Bronze Age initiation ritual in the Russian steppes. *Journal of Anthropological Archaeology*. **48**, 134–148 (2017).
89. T. Dayan, E. Galili, A preliminary look at some new domesticated dogs from submerged Neolithic sites off the Carmel coast. *BAR INTERNATIONAL SERIES*. **889**, 29–34 (2000).
- 695 90. L. K. Horwitz, S. R. Wolff, S. Ortiz, J. Lev-Tov, A. Gilbert, P. Hesse, The Context and Biometry of Iron Age II and Hellenistic Period Dog “Burials” from Tel Gezer Compared to Those from Other Sites in the Region. *The Wide Lens in Archaeology: Honoring Brian Hesse’s Contributions to Anthropological Archaeology*, 297–333 (2017).
- 700 91. W. G. Dever, H. D. Lance, G. E. Wright, Gezer I: Preliminary Report of the 1964-66 Seasons, vol. 1. *Jerusalem: Hebrew Union College Biblical and Archaeological School in Jerusalem* (1970).
92. S. M. Ortiz, S. R. Wolff, Tel Gezer excavations 2006--2015: the transformation of a border city. *The Shephelah during the Iron Age: Recent Archaeological Studies*, 61–102 (2017).
93. N. Getzov, D. Avshalom Gorni, Y. Gorin-Rosen, E. J. Stern, D. Syon, A. Tatcher, Horbat ‘Uza: The 1991 Excavations, ii. *The Late Periods* (2009).
- 705 94. D. M. Master, J. D. Schloen, L. E. Stager, *Ashkelon I: Introduction and overview (1985-2006)* (Eisenbrauns, 2008).
95. L. E. Stager, Why were hundreds of dogs buried at Ashkelon? *Biblical Archaeology Review*. **17**, 26–42 (1991).
96. P. Wapnish, B. Hesse, Pampered Pooches or Plain Pariahs? The Ashkelon Dog Burials. *The*

- 710 *Biblical Archaeologist*. **56**, 55–80 (1993).
97. H. B. Hansen, P. B. Damgaard, A. Margaryan, J. Stenderup, N. Lynnerup, E. Willerslev, M. E. Allentoft, Comparing Ancient DNA Preservation in Petrous Bone and Tooth Cementum. *PLoS One*. **12**, e0170940 (2017).
- 715 98. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.*, **6**, pdb.prot5448 (2010).
- 720 99. R. Rodríguez-Varela, T. Günther, M. Krzewińska, J. Storå, T. H. Gillingwater, M. MacCallum, J. L. Arsuaga, K. Dobney, C. Valdiosera, M. Jakobsson, A. Götherström, L. Girdland-Flink, Genomic Analyses of Pre-European Conquest Human Remains from the Canary Islands Reveal Close Affinity to Modern North Africans. *Curr. Biol*. **27**, 3396–3402.e5 (2017).
- 725 100. T. Günther, H. Malmström, E. M. Svensson, A. Omrak, F. Sánchez-Quinto, G. M. Kılınç, M. Krzewińska, G. Eriksson, M. Fraser, H. Edlund, A. R. Munters, A. Coutinho, L. G. Simões, M. Vicente, A. Sjölander, B. Jansen Sellevold, R. Jørgensen, P. Claes, M. D. Shriver, C. Valdiosera, M. G. Netea, J. Apel, K. Lidén, B. Skar, J. Storå, A. Götherström, M. Jakobsson, Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biol*. **16**, e2003703 (2018).
101. E. Ersmark, L. Orlando, E. Sandoval-Castellanos, I. Barnes, R. Barnett, A. Stuart, A. Lister, L. Dalén, Population demography and genetic diversity in the Pleistocene cave lion. *Open Quaternary*. **1**, 1–14 (2015).
- 730 102. A. W. Briggs, U. Stenzel, M. Meyer, J. Krause, M. Kircher, S. Pääbo, Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res*. **38**, e87 (2010).
- 735 103. J. Dabney, M. Knapp, I. Glocke, M.-T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N. García, S. Pääbo, J.-L. Arsuaga, M. Meyer, 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).
- 740 104. C. Gamba, E. R. Jones, M. D. Teasdale, R. L. McLaughlin, G. Gonzalez-Fortes, V. Mattiangeli, L. Domboróczki, I. Kővári, I. Pap, A. Anders, A. Whittle, J. Dani, P. Raczky, T. F. G. Higham, M. Hofreiter, D. G. Bradley, R. Pinhasi, Genome flux and stasis in a five millennium transect of European prehistory. *Nature Communications*. **5** (2014), , doi:10.1038/ncomms6257.
105. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. **25**, 1754–1760 (2009).
- 745 106. M. Meyer, M. Kircher, M.-T. Gansauge, H. Li, F. Racimo, S. Mallick, J. G. Schraiber, F. Jay, K. Prüfer, C. de Filippo, P. H. Sudmant, C. Alkan, Q. Fu, R. Do, N. Rohland, A. Tandon, M. Siebauer, R. E. Green, K. Bryc, A. W. Briggs, U. Stenzel, J. Dabney, J. Shendure, J. Kitzman, M. F. Hammer, M. V. Shunkov, A. P. Derevianko, N. Patterson, A.

M. Andrés, E. E. Eichler, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science*. **338**, 222–226 (2012).

- 750 107. P. Skoglund, B. H. Northoff, M. V. Shunkov, Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proceedings of the* (2014) (available at <https://www.pnas.org/content/111/6/2229/>).
- 755 108. P. Pečnerová, D. Díez-Del-Molino, N. Dussex, T. Feuerborn, J. von Seth, J. van der Plicht, P. Nikolskiy, A. Tikhonov, S. Vartanyan, L. Dalén, Genome-Based Sexing Provides Clues about Behavior and Social Structure in the Woolly Mammoth. *Curr. Biol.* **27**, 3505–3510.e3 (2017).
- 760 109. G. Gower, L. E. Fenderson, A. T. Salis, K. M. Helgen, A. L. van Loenen, H. Heiniger, E. Hofman-Kamińska, R. Kowalczyk, K. J. Mitchell, B. Llamas, A. Cooper, Widespread male sex bias in mammal fossil and museum collections. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 19019–19024 (2019).
110. N.-N. Shi, L. Fan, Y.-G. Yao, M.-S. Peng, Y.-P. Zhang, Mitochondrial genomes of domestic animals need scrutiny. *Mol. Ecol.* **23**, 5393–5397 (2014).
- 765 111. J. Krause, A. W. Briggs, M. Kircher, T. Maricic, N. Zwyns, A. Derevianko, S. Pääbo, A Complete mtDNA Genome of an Early Modern Human from Kostenki, Russia. *Curr. Biol.* **20**, 231–236 (2010).
112. J. Plassais, J. Kim, B. W. Davis, D. M. Karyadi, A. N. Hogan, A. C. Harris, B. Decker, H. G. Parker, E. A. Ostrander, Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat. Commun.* **10**, 1489 (2019).
- 770 113. G.-D. Wang, W. Zhai, H.-C. Yang, R.-X. Fan, X. Cao, L. Zhong, L. Wang, F. Liu, H. Wu, L.-G. Cheng, A. D. Poyarkov, N. A. Poyarkov Jr, S.-S. Tang, W.-M. Zhao, Y. Gao, X.-M. Lv, D. M. Irwin, P. Savolainen, C.-I. Wu, Y.-P. Zhang, The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat. Commun.* **4**, 1860 (2013).
- 775 114. M. Kardos, M. Åkesson, T. Fountain, Ø. Flagstad, O. Liberg, P. Olason, H. Sand, P. Wabakken, C. Wikenros, H. Ellegren, Genomic consequences of intensive inbreeding in an isolated wolf population. *Nat Ecol Evol.* **2**, 124–131 (2018).
- 780 115. M.-H. S. Sinding, S. Gopalakrishnan, F. G. Vieira, J. A. Samaniego Castruita, K. Raundrup, M. P. Heide Jørgensen, M. Meldgaard, B. Petersen, T. Sicheritz-Ponten, J. B. Mikkelsen, U. Marquard-Petersen, R. Dietz, C. Sonne, L. Dalén, L. Bachmann, Ø. Wiig, A. J. Hansen, M. T. P. Gilbert, Population genomics of grey wolves and wolf-like canids in North America. *PLoS Genet.* **14**, e1007745 (2018).
- 785 116. S. Gopalakrishnan, M.-H. S. Sinding, J. Ramos-Madrigal, J. Niemann, J. A. Samaniego Castruita, F. G. Vieira, C. Carøe, M. de M. Montero, L. Kuderna, A. Serres, V. M. González-Basallote, Y.-H. Liu, G.-D. Wang, T. Marques-Bonet, S. Mirarab, C. Fernandes, P. Gaubert, K.-P. Koepfli, J. Budd, E. K. Rueness, M. P. Heide-Jørgensen, B. Petersen, T.

- Sicheritz-Ponten, L. Bachmann, Ø. Wiig, A. J. Hansen, M. T. P. Gilbert, Interspecific Gene Flow Shaped the Evolution of the Genus *Canis*. *Curr. Biol.* **28**, 3441–3449.e5 (2018).
117. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013), (available at <http://arxiv.org/abs/1303.3997>).
- 790 118. M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernysky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
- 795 119. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, D. Reich, Ancient admixture in human history. *Genetics.* **192**, 1065–1093 (2012).
- 800 120. P. de Barros Damgaard, R. Martiniano, J. Kamm, J. V. Moreno-Mayar, G. Kroonen, M. Peyrot, G. Barjamovic, S. Rasmussen, C. Zacho, N. Baimukhanov, V. Zaibert, V. Merz, A. Biddanda, I. Merz, V. Loman, V. Evdokimov, E. Usmanova, B. Hemphill, A. Seguin-Orlando, F. E. Yediay, I. Ullah, K.-G. Sjögren, K. H. Iversen, J. Choin, C. de la Fuente, M. Ilardo, H. Schroeder, V. Moiseyev, A. Gromov, A. Polyakov, S. Omura, S. Y. Senyurt, H. Ahmad, C. McKenzie, A. Margaryan, A. Hameed, A. Samad, N. Gul, M. H. Khokhar, O. I. Goriunova, V. I. Bazaliiskii, J. Novembre, A. W. Weber, L. Orlando, M. E. Allentoft, R. Nielsen, K. Kristiansen, M. Sikora, A. K. Outram, R. Durbin, E. Willerslev, The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science.* **360** (2018), doi:10.1126/science.aar7711.
- 805 121. M. Haber, C. Doumet-Serhal, C. Scheib, Y. Xue, P. Danecek, M. Mezzavilla, S. Youhanna, R. Martiniano, J. Prado-Martinez, M. Szpak, E. Matisoo-Smith, H. Schutkowski, R. Mikulski, P. Zalloua, T. Kivisild, C. Tyler-Smith, Continuity and Admixture in the Last Five Millennia of Levantine History from Ancient Canaanite and Present-Day Lebanese Genome Sequences. *Am. J. Hum. Genet.* **101**, 274–282 (2017).
- 810 122. C. L. Scheib, H. Li, T. Desai, V. Link, C. Kendall, G. Dewar, P. W. Griffith, A. Mörseburg, J. R. Johnson, A. Potter, S. L. Kerr, P. Endicott, J. Lindo, M. Haber, Y. Xue, C. Tyler-Smith, M. S. Sandhu, J. G. Lorenz, T. D. Randall, Z. Faltyskova, L. Pagani, P. Danecek, T. C. O’Connell, P. Martz, A. S. Boraas, B. F. Byrd, A. Leventhal, R. Cambra, R. Williamson, L. Lesage, B. Holguin, E. Ygnacio-De Soto, J. Rosas, M. Metspalu, J. T. Stock, A. Manica, A. Scally, D. Wegmann, R. S. Malhi, T. Kivisild, Ancient human parallel lineages within North America contributed to a coastal expansion. *Science.* **360**, 1024–1027 (2018).
- 815 123. I. Mathieson, I. Lazaridis, N. Rohland, S. Mallick, N. Patterson, S. A. Roodenberg, E. Harney, K. Stewardson, D. Fernandes, M. Novak, K. Sirak, C. Gamba, E. R. Jones, B. Llamas, S. Dryomov, J. Pickrell, J. L. Arsuaga, J. M. B. de Castro, E. Carbonell, F. Gerritsen, A. Khokhlov, P. Kuznetsov, M. Lozano, H. Meller, O. Mochalov, V. Moiseyev, M. A. R. Guerra, J. Roodenberg, J. M. Vergès, J. Krause, A. Cooper, K. W. Alt, D. Brown, 825 D. Anthony, C. Lalueza-Fox, W. Haak, R. Pinhasi, D. Reich, Genome-wide patterns of

selection in 230 ancient Eurasians. *Nature*. **528**, 499–503 (2015).

- 830 124. I. Mathieson, S. Alpaslan-Roodenberg, C. Posth, A. Szécsényi-Nagy, N. Rohland, S. Mallick, I. Olalde, N. Broomandkhoshbacht, F. Candilio, O. Cheronet, D. Fernandes, M. Ferry, B. Gamarra, G. G. Fortes, W. Haak, E. Harney, E. Jones, D. Keating, B. Krause-Kyora, I. Kucukkalipci, M. Michel, A. Mittnik, K. Nägele, M. Novak, J. Oppenheimer, N. Patterson, S. Pfrengle, K. Sirak, K. Stewardson, S. Vai, S. Alexandrov, K. W. Alt, R. Andreescu, D. Antonović, A. Ash, N. Atanassova, K. Bacvarov, M. B. Gusztáv, H. Bocherens, M. Bolus, A. Boroneanț, Y. Boyadzhiev, A. Budnik, J. Burmaz, S. Chohadzhiev, N. J. Conard, R. Cottiaux, M. Čuka, C. Cupillard, D. G. Drucker, N. Elenski, 835 M. Francken, B. Galabova, G. Ganetsovski, B. Gély, T. Hajdu, V. Handzhyiska, K. Harvati, T. Higham, S. Iliev, I. Janković, I. Karavanić, D. J. Kennett, D. Komšo, A. Kozak, D. Labuda, M. Lari, C. Lazar, M. Leppek, K. Leshtakov, D. L. Vetro, D. Los, I. Lozanov, M. Malina, F. Martini, K. McSweeney, H. Meller, M. Mendušić, P. Mirea, V. Moiseyev, V. Petrova, T. D. Price, A. Simalcsik, L. Sineo, M. Šlaus, V. Slavchev, P. Stanev, A. Starović, 840 T. Szeniczey, S. Talamo, M. Teschler-Nicola, C. Thevenet, I. Valchev, F. Valentin, S. Vasilyev, F. Veljanovska, S. Venelinova, E. Veselovskaya, B. Viola, C. Virag, J. Zaninović, S. Zäuner, P. W. Stockhammer, G. Catalano, R. Krauß, D. Caramelli, G. Zariņa, B. Gaydarska, M. Lillie, A. G. Nikitin, I. Potekhina, A. Papatnasiou, D. Borić, C. Bonsall, J. Krause, R. Pinhasi, D. Reich, The genomic history of southeastern Europe. *Nature*. **555**, 845 197–203 (2018).
125. Z. Hofmanová, S. Kreutzer, G. Hellenthal, C. Sell, Y. Diekmann, D. Díez-Del-Molino, L. van Dorp, S. López, A. Kousathanas, V. Link, K. Kirsanow, L. M. Cassidy, R. Martiniano, M. Strobel, A. Scheu, K. Kotsakis, P. Halstead, S. Triantaphyllou, N. Kyparissi-Apostolika, D. Urem-Kotsou, C. Ziota, F. Adaktylou, S. Gopalan, D. M. Bobo, L. Winkelbach, J. 850 Blöcher, M. Unterländer, C. Leuenberger, Ç. Çilingiroğlu, B. Horejs, F. Gerritsen, S. J. Shennan, D. G. Bradley, M. Currat, K. R. Veeramah, D. Wegmann, M. G. Thomas, C. Papageorgopoulou, J. Burger, Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 6886–6891 (2016).
- 855 126. I. Olalde, S. Brace, M. E. Allentoft, I. Armit, K. Kristiansen, T. Booth, N. Rohland, S. Mallick, A. Szécsényi-Nagy, A. Mittnik, E. Altena, M. Lipson, I. Lazaridis, T. K. Harper, N. Patterson, N. Broomandkhoshbacht, Y. Diekmann, Z. Faltyskova, D. Fernandes, M. Ferry, E. Harney, P. de Knijff, M. Michel, J. Oppenheimer, K. Stewardson, A. Barclay, K. W. Alt, C. Liesau, P. Ríos, C. Blasco, J. V. Miguel, R. M. García, A. A. Fernández, E. Bánffy, M. Bernabò-Brea, D. Billoin, C. Bonsall, L. Bonsall, T. Allen, L. Büster, S. Carver, 860 L. C. Navarro, O. E. Craig, G. T. Cook, B. Cunliffe, A. Denaire, K. E. Dinwiddy, N. Dodwell, M. Ernée, C. Evans, M. Kuchařík, J. F. Farré, C. Fowler, M. Gazenbeek, R. G. Pena, M. Haber-Uriarte, E. Haduch, G. Hey, N. Jowett, T. Knowles, K. Massy, S. Pfrengle, P. Lefranc, O. Lemerrier, A. Lefebvre, C. H. Martínez, V. G. Olmo, A. B. Ramírez, J. L. Maurandi, T. Majó, J. I. McKinley, K. McSweeney, B. G. Mende, A. Modi, G. Kulcsár, V. Kiss, A. Czene, R. Patay, A. Endrődi, K. Köhler, T. Hajdu, T. Szeniczey, J. Dani, Z. 865 Bernert, M. Hoole, O. Cheronet, D. Keating, P. Velemínský, M. Dobeš, F. Candilio, F. Brown, R. F. Fernández, A.-M. Herrero-Corral, S. Tusa, E. Carnieri, L. Lentini, A. Valenti, A. Zanini, C. Waddington, G. Delibes, E. Guerra-Doce, B. Neil, M. Brittain, M. Luke, R. Mortimer, J. Desideri, M. Besse, G. Brücken, M. Furmanek, A. Hałuszko, M. Mackiewicz,

- 870 A. Rapiński, S. Leach, I. Soriano, K. T. Lillios, J. L. Cardoso, M. P. Pearson, P.
Włodarczak, T. D. Price, P. Prieto, P.-J. Rey, R. Risch, M. A. Rojo Guerra, A. Schmitt, J.
Serralongue, A. M. Silva, V. Smrčka, L. Vergnaud, J. Zilhão, D. Caramelli, T. Higham, M.
G. Thomas, D. J. Kennett, H. Fokkens, V. Heyd, A. Sheridan, K.-G. Sjögren, P. W.
875 Stockhammer, J. Krause, R. Pinhasi, W. Haak, I. Barnes, C. Lalueza-Fox, D. Reich, The
Beaker phenomenon and the genomic transformation of northwest Europe. *Nature*. **555**,
190–196 (2018).
127. S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, M. Zhao, N.
Chennagiri, S. Nordenfelt, A. Tandon, P. Skoglund, I. Lazaridis, S. Sankararaman, Q. Fu,
N. Rohland, G. Renaud, Y. Erlich, T. Willems, C. Gallo, J. P. Spence, Y. S. Song, G.
880 Poletti, F. Balloux, G. van Driem, P. de Knijff, I. G. Romero, A. R. Jha, D. M. Behar, C. M.
Bravi, C. Capelli, T. Hervig, A. Moreno-Estrada, O. L. Posukh, E. Balanovska, O.
Balanovsky, S. Karachanak-Yankova, H. Sahakyan, D. Toncheva, L. Yepiskoposyan, C.
Tyler-Smith, Y. Xue, M. S. Abdullah, A. Ruiz-Linares, C. M. Beall, A. Di Rienzo, C.
885 Jeong, E. B. Starikovskaya, E. Metspalu, J. Parik, R. Villems, B. M. Henn, U. Hodoglugil,
R. Mahley, A. Sajantila, G. Stamatoyannopoulos, J. T. S. Wee, R. Khusainova, E.
Khusnutdinova, S. Litvinov, G. Ayodo, D. Comas, M. F. Hammer, T. Kivisild, W. Klitz, C.
A. Winkler, D. Labuda, M. Bamshad, L. B. Jorde, S. A. Tishkoff, W. S. Watkins, M.
Metspalu, S. Dryomov, R. Sukernik, L. Singh, K. Thangaraj, S. Pääbo, J. Kelso, N.
890 Patterson, D. Reich, The Simons Genome Diversity Project: 300 genomes from 142 diverse
populations. *Nature*. **538**, 201–206 (2016).
128. J. Oksanen, F. G. Blanchet, R. Kindt, P. Legendre, R. B. O’hara, G. L. Simpson, P.
Solyomos, M. H. H. Stevens, H. Wagner, Vegan: community ecology package. R package
version 1.17-4. <https://cran.r-project.org/> **23**, 2010 (2010).
129. H. R. Kunsch, The jackknife and the bootstrap for general stationary observations. *Ann.*
895 *Stat.* **17**, 1217–1241 (1989).
130. I. Lazaridis, A. Belfer-Cohen, S. Mallick, N. Patterson, Paleolithic DNA from the
Caucasus reveals core of West Eurasian ancestry. *bioRxiv* (2018) (available at
<https://www.biorxiv.org/content/10.1101/423079v1.abstract>).
131. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in
900 unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
132. K. Leppälä, S. V. Nielsen, T. Mailund, admixturegraph: an R package for admixture
graph manipulation and fitting. *Bioinformatics*. **33**, 1738–1740 (2017).
133. D. Reich, K. Thangaraj, N. Patterson, A. L. Price, L. Singh, Reconstructing Indian
population history. *Nature*. **461**, 489–494 (2009).
- 905 134. R. R. Hudson, Generating samples under a Wright–Fisher neutral model of genetic
variation. *Bioinformatics*. **18**, 337–338 (2002).
135. P. Skoglund, S. Mallick, M. C. Bortolini, N. Chennagiri, T. Hünemeier, M. L. Petzl-
Erler, F. M. Salzano, N. Patterson, D. Reich, Genetic evidence for two founding populations

of the Americas. *Nature* (2015).

- 910 136. I. Baranowska, K. H. Jäderlund, I. Nennesmo, E. Holmqvist, N. Heidrich, N.-G. Larsson,
G. Andersson, E. G. H. Wagner, Å. Hedhammar, R. Wibom, L. Andersson, Sensory Ataxic
Neuropathy in Golden Retriever Dogs Is Caused by a Deletion in the Mitochondrial
tRNATyr Gene. *PLoS Genetics*. **5** (2009), p. e1000499.
- 915 137. S. Björnerfeldt, M. T. Webster, C. Vilà, Relaxation of selective constraint on dog
mitochondrial DNA following domestication. *Genome Res.* **16**, 990–994 (2006).
138. K. S. Kim, S. E. Lee, H. W. Jeong, J. H. Ha, The complete nucleotide sequence of the
domestic dog (*Canis familiaris*) mitochondrial genome. *Mol. Phylogenet. Evol.* **10**, 210–220
(1998).
- 920 139. A. Strakova, M. Ní Leathlobhair, G.-D. Wang, T.-T. Yin, I. Airikkala-Otter, J. L. Allen,
K. M. Allum, L. Bansse-Issa, J. L. Bisson, A. Castillo Domracheva, K. F. de Castro, A. M.
Corrigan, H. R. Cran, J. T. Crawford, S. M. Cutter, L. Delgadillo Keenan, E. M. Donelan, I.
A. Faramade, E. Flores Reynoso, E. Fotopoulou, S. N. Fruean, F. Gallardo-Arrieta, O.
Glebova, R. F. Häfelin Manrique, J. J. Henriques, N. Ignatenko, D. Koenig, M. Lanza-
Perea, R. Lobetti, A. M. Lopez Quintana, T. Losfelt, G. Marino, I. Martincorena, S.
925 Martínez Castañeda, M. F. Martínez-López, M. Meyer, B. Nakanwagi, A. B. De Nardi, W.
Neunzig, S. J. Nixon, M. M. Onsare, A. Ortega-Pacheco, M. C. Peleteiro, R. J. Pye, J. F.
Reece, J. Rojas Gutierrez, H. Sadia, S. K. Schmeling, O. Shamanova, R. K. Ssuna, A. E.
Steenland-Smit, A. Svitich, I. Thoya Ngoka, B. A. Vițálaru, A. P. de Vos, J. P. de Vos, O.
Walkinton, D. C. Wedge, A. S. Wehrle-Martinez, M. G. van der Wel, S. A. Widdowson, E.
930 P. Murchison, Mitochondrial genetic diversity, selection and recombination in a canine
transmissible cancer. *Elife*. **5** (2016), doi:10.7554/eLife.14552.
140. K. M. Webb, M. W. Allard, Mitochondrial genome DNA analysis of the domestic dog:
identifying informative SNPs outside of the control region. *J. Forensic Sci.* **54**, 275–288
(2009).
- 935 141. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high
throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
142. A. Larsson, AliView: a fast and lightweight alignment viewer and editor for large
datasets. *Bioinformatics*. **30**, 3276–3278 (2014).
- 940 143. R. Bouckaert, J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A.
Rambaut, A. J. Drummond, BEAST 2: a software platform for Bayesian evolutionary
analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).
144. M. Molak, M. A. Suchard, S. Y. W. Ho, D. W. Beilman, B. Shapiro, Empirical calibrated
radiocarbon sampler: a tool for incorporating radiocarbon-date and calibration error into
Bayesian phylogenetic analyses of ancient DNA. *Molecular Ecology Resources*. **15** (2015),
945 pp. 81–86.
145. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior

Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).

146. A. J. Drummond, A. Rambaut, BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
- 950 147. F. D. Ciccarelli, Toward Automatic Reconstruction of a Highly Resolved Tree of Life. *Science*. **311** (2006), pp. 1283–1287.
148. A. D. Yates, P. Achuthan, W. Akanni, J. Allen, J. Allen, J. Alvarez-Jarreta, M. R. Amode, I. M. Armean, A. G. Azov, R. Bennett, J. Bhai, K. Billis, S. Boddu, J. C. Marugán, C. Cummins, C. Davidson, K. Dodiya, R. Fatima, A. Gall, C. G. Giron, L. Gil, T. Grego, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, M. Kay, I. Lavidas, T. Le, D. Lemos, J. G. Martinez, T. Maurel, M. McDowall, A. McMahon, S. Mohanan, B. Moore, M. Nuhn, D. N. Oheh, A. Parker, A. Parton, M. Patricio, M. P. Sakthivel, A. I. Abdul Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, M. Sycheva, M. Szuba, K. Taylor, A. Thormann, G. Threadgold, A. Vullo, B. Walts, A. Winterbottom, 955 A. Zadissa, M. Chakiachvili, B. Flint, A. Frankish, S. E. Hunt, G. Iisley, M. Kostadima, N. Langridge, J. E. Loveland, F. J. Martin, J. Morales, J. M. Mudge, M. Muffato, E. Perry, M. Ruffier, S. J. Trevanion, F. Cunningham, K. L. Howe, D. R. Zerbino, P. Flicek, Ensembl 2020. *Nucleic Acids Res.* (2019), doi:10.1093/nar/gkz966.
- 960 149. B. Bai, W.-M. Zhao, B.-X. Tang, Y.-Q. Wang, L. Wang, Z. Zhang, H.-C. Yang, Y.-H. Liu, J.-W. Zhu, D. M. Irwin, G.-D. Wang, Y.-P. Zhang, DoGSD: the dog and wolf genome SNP database. *Nucleic Acids Res.* **43**, D777–83 (2015).
- 965 150. P. Skoglund, J. Storå, A. Götherström, M. Jakobsson, Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40**, 4477–4482 (2013).

970

Acknowledgements: We thank S. Charlton, I. Lazaridis, A. Manin and I. Mathieson for comments on the manuscript, G.-D. Wang and C. Marsden for help with data access, and GORDAILUA (the Gipuzkoa Centre for Heritage Collections), S. San José, C. Olaetxea, M. Urteaga, A. Sampson, A.R. Sardari Zarchi and M. Abdollahi (ICHHTO, Iran) for facilitating sample access. **Funding:** Ancient genome sequencing was supported by SciLifeLab National Projects and the Erik Philip Sörensen Foundation (to P.S.). A.B., T.D., and P.S. were supported by the Francis Crick Institute core funding (FC001595) from Cancer Research UK, the UK Medical Research Council, and the Wellcome Trust. P.S. was also supported by the European Research Council (grant no. 852558) and Wellcome Trust Investigator award (217223/Z/19/Z). 975 R.L. was supported by the Social Sciences and Humanities Research Council of Canada (#SSHRC IG 435-2014-0075). Y. K. was supported by State Assignment of the Sobolev Institute of Geology and Mineralogy. M.S. was supported by ZIN RAS (state assignment no. AAAA-A19-119032590102-7). A.T.L. was supported by the Smithsonian’s Peter Buck Postdoctoral Fellowship. Computations were supported by SNIC-UPPMAX (b2016004), and the UOXF ARC 980 facility. L.A.F.F. was supported by the Wellcome Trust (Grant 210119/Z/18/Z) and by Wolfson College (University of Oxford). G.L. was supported by the ERC (Grant ERC-2013-StG-337574-UNDEAD), and Natural Environmental Research Council (Grants NE/K005243/1 and 985

NE/K003259/1). Dating was supported by the NERC Radiocarbon Facility (NF/2016/2/4).

990 **Author contributions:** GL and PS initiated the study. JS, K-GS, DA, EA, SA, GB-O, VIB, JB,
DB, SF, IF, DF, MG, LH, LJ, JK-C, YK, RJJ, DLD, MM, MN, VO, DO, MP, MR, DR, BR,
MS, IS, AT, KT, IU, AV, PW, AG, and LD contributed material and archaeological information.
995 RS, EE, OL, LG-F, JH, AJ, HR and AL did ancient DNA molecular work, supervised by AG,
LD, RP, GL and PS. AB, LF, AC, TD, EKI-P and PS processed the genome sequence data,
supervised by LF and PS. AB did population genomic analyses, supervised by PS. ATL did
mtDNA analyses, supervised by GL. AB, LF, GL and PS wrote the paper with input from RP,
KD and all other authors. **Competing interests:** Authors declare no competing interests. **Data
and materials availability:** The generated DNA sequencing data will be made available in the
European Nucleotide Archive (ENA) under study accession PRJEB38079.

1000 **Supplementary Materials:**

Materials and Methods

Figs. S1 to S13

Tables S1 to S6

Data S1

1005 References (62 – 150)

Supplementary Materials for **Origins and Genetic Legacy of Prehistoric Dogs**

Anders Bergström*, Laurent Frantz*, Ryan Schmidt, Erik Ersmark, Ophelie Lebrasseur, Linus Girdland-Flink, Audrey T. Lin, Jan Storå, Karl-Göran Sjögren, David Anthony, Ekaterina Antipina, Sarieh Amiri, Guy Bar-Oz, Vladimir I. Bazaliiskii, Jelena Bulatović, Dorcas Brown, Alberto Carmagnini, Tom Davy, Sergey Fedorov, Ivana Fiore, Deidre Fulton, Mietje Germonpré, James Haile, Evan K. Irving-Pease, Alexandra Jamieson, Luc Janssens, Irina Kirillova, Liora Kolska Horwitz, Julka Kuzmanovic-Cvetkovic, Yaroslav Kuzmin, Robert J. Losey, Daria Ložnjak Dizdar, Marjan Mashkour, Mario Novak, Vedat Onar, David Orton, Maja Pasarić, Miljana Radivojević, Dragana Rajković, Benjamin Roberts, Hannah Ryan, Mikhail Sablin, Fedor Shidlovskiy, Ivana Stojanović, Antonio Tagliacozzo, Katerina Trantalidou, Inga Ullén, Aritza Villaluenga, Paula Wapnish, Keith Dobney, Anders Götherström, Anna Linderholm, Love Dalén, Ron Pinhasi*, Greger Larson*, Pontus Skoglund*

*Correspondence to: anders.bergstrom@crick.ac.uk (A.B.), l.frantz@qmul.ac.uk (L.F.), ron.pinhasi@univie.ac.at (R.P.), greger.larson@arch.ox.ac.uk (G.L.), pontus.skoglund@crick.ac.uk (P.S.)

This PDF file includes:

Materials and Methods
Figs. S1 to S13
Tables S1 to S6
Captions for Data S1
References

Other Supplementary Materials for this manuscript include the following:

Data S1

Materials and Methods

Archaeological samples and context

Greece (sample ID: OL4222). Of the 9 bone and tooth specimens selected from the faunal assemblages of the caves Pighes Koromilias in Western Macedonia, Oinoi IV in Attica, Aghia Triada in Euboea, Kouveleiki I in the Peloponnese, Nymphs at Vonitsa and tell Imbrou Pighadi both in central Greece, only a left mandibular fragment found at the sediments of the Skoteini Cave (A12, T6) in central Euboea Island had more than 25% endogenous DNA which was enough to sequence more deeply in order to generate a whole genome.

The Skoteini cave was occupied from the beginning of the Late Neolithic I until the 3rd century AD, though a coin pointed to the 6th c. AD. However, the most intensive use of the cavity occurred for 2000 years in the Late Neolithic cultural sequence (5300/5200-3300/3200 BC) as evidenced by several earth-beaten floors in close succession and the systematic practice of storage, clearly shown by the hundreds of sherds from pithoi and the great number of carbonized seeds (62). The dog analyzed in this study (OL4222) lived between 4689 and 4500 (probability 95.4%) cal BC (calibration: courtesy Y. Facorellis, University of West Attica), a span of time which, according to the excavator, corresponds to the LN Ib phase.

The skeletal remains of domestic dog were very scanty in the cave assemblages (total NISP = 72 out of 17,927 total bone remains), most of the bones were from adult individuals and a total of five bones bore cut marks. The richest deposit has been found in Late Neolithic Ib layers: there were at least 4 dogs, represented by 4 left ulnae (63). The size of the dogs, as estimated from wither's height of individuals from Sitagroi, east Macedonia (during the 4600-3500 phase) can be considered medium, with an average of 43.20cm.

Iran (sample ID: AL2571). Tepe Qela Gap (also known as Ghala Gap) in Lurestan Province was excavated in 2009 by M. Abdollahi and A. Sardari, aiming to establish the chronological sequence for the Azna Plain located in the eastern part of Central Zagros Mountain which had been scarcely studied archaeologically until now (64–66). The excavation revealed a sequence of occupations from Late Neolithic, Chalcolithic, Bronze, Iron Age and Parthian periods. Considering the ecological diversity surrounding the plain, Tepe Qela Gap seems to have been an ideal place for the settlement of permanent villages but could also have been suitable for nomadic and semi-nomadic people. The faunal spectra of Qela Gap from different periods (approximately 6500 specimens), indicates that domesticated sheep/goat and cattle were the major source of animal resources. Among the carnivore remains (26 specimens, 0.4%), a complete cranium of *Canis familiaris* along with some pieces of cervical vertebra from the Middle chalcolithic period (1388_RN1328) was discovered (sample AL2571). On the basis of archaeological evidence, in particular the absence of permanent architecture, with ovens, fireplaces and plastered or paved floors that are present during the Bronze Age, a mobile type of settlement cannot be ruled out for this site during the Neolithic and Chalcolithic. This is also confirmed by the absence of the very young animals in the kill-off pattern for these periods.

Italy (sample ID: AL2397). The sampled specimen is an almost complete right hemimandible that is missing a small portion of the incisor area and is referable to as an adult individual. It

comes from excavations carried out in the first half of the past century in the caves of Belverde di Cetona and is preserved in the collections of the laboratory of Bioarchaeology of the Museum of Civilizations in Rome (67). It dates back to the Bronze Age, which in central-northern Italy lasted from roughly 2,300 to 1,700 BC, (~4,300-3700 BP). The specimen is part of a collection of skulls and mandibles that has attested to the presence at Belverde, in the Italian Bronze Age, of dogs that varied greatly in size, ranging from small to medium to large. The mandibles in this collection have a range of variability between 112-148 mm; the hemimandible sampled in this study falls into the medium-large dog category.

Turkey (sample ID: AL2022) dog. The Turkish dog analyzed dates to the Byzantine period and was found at the Yenikapı site in Istanbul. Remains from a large number of animals of a variety of species have been recovered at this site, which was likely used as a harbor (68–70).

Karelia / Veretye (sample ID: OL4061). The Veretye (Веретьё) (a.k.a. Veret'ye) culture derives its name from the Veretye 1 site. It is located in Arkhangelsk Province, northern European Russia, on the bank of Kinema River ca. 1 km from its mouth; the river flows to the Lake Lacha. The excavated area is ca. 1470 m². The site has a single cultural component associated with the Mesolithic epoch (71, 72). Planigraphically, there are remains of dwellings and other structures (of everyday life use), and scattered artefacts. Cultural material is located in oxygen-free peat layers (wetland site), and the preservation of bones, antlers and other perishable materials (wood, birch bark, and plant fibres) is generally very good. Material culture is represented by a large set of stone tools used for making items for hunting, fishing, and woodworking. Numerous tools made of bone, antler, and wood are also present. No pottery has been found.

The chronology of the Veretye 1 site is based on ¹⁴C dates obtained on different materials from the cultural layer: a) charcoal: 9600 ± 80 BP (Le-1469), 9050 ± 80 BP (GIN-4031), 8560 ± 120 BP (GIN-2452), 8520 ± 80 BP (GIN-4030), 8270 ± 100 BP (Le-1470), and 7960 ± 100 BP (Le-1471); b) antler: 9370 ± 80 BP (GIN-4833) and 8340 ± 120 BP (GIN-4832); and c) wood: 8750 ± 70 BP (Le-1472), 8550 ± 130 BP (GIN-2452), and 7700 ± 80 BP (Le-1773) (72). According to ref. (71) the ¹⁴C dates on charcoal and worked wood of ca. 9600–8550 BP are the most closely associated with the cultural component.

Bones and skulls of 42 dogs were found at the Veretye 1 site, and it constitutes 12.6% of total animal bones from this site (72). The direct ¹⁴C date of the dog tooth from the Veretye 1 site used for DNA extraction (sample OL4061) is 9575 ± 50 BP (OxA-36900), corresponding to calendar age of 10,780–11,080 cal BP (with ±1 sigma; using IntCal13 dataset); the median value is ca. 10,930 cal BP.

Sweden (sample IDs: C62, C88, C89, C90, C94). The sample C88 (archaeological ID FNR139434) is from the site of Frälssegården, Gökhem, Västergötland, Sweden, with a Funnelbeaker culture context. The sample is a tooth bead made from an I3 tooth, found in a passage grave used by the Funnelbeaker culture. It is contextually dated to ca 5000 cal BP based on human bone dates (73).

The samples C89 and C90 are from the Ajvide site, Gotland, Sweden, with a Pitted Ware Culture context. Both samples are mandibles found in the cultural layer of the Pitted Ware Culture settlement (and burial ground) site, and contextually dated to c. 4900-4500 cal BP. Further details on C89: archaeological ID AJCAN6 (alternative ID GMM, No AJV-1), cultural layer, FID31524, X2.73; y-136.4; z12.25, layer III, recovered in 1996, mandible from left side was present, canine tooth was sampled. Further details on C90: archaeological ID AJCAN4 (alternative ID GMM No. AJV-4), cultural layer, X10.65; y-145.9; z 12.29. layer II, recovered in 1994, both mandibles were found together but no cranium was found in the same context, canine tooth was sampled.

The sample C94 (archaeological ID SKJH5) is from the cave site Stora Förvar on the island Stora Karlsö, Gotland, Sweden, with a late Neolithic context. The sample is a humerus (upper arm bone) recovered in layer 5 of section H (H5), a layer containing finds from the late stone age, but without a strict chronological separation of finds from younger periods. The bone was radiocarbon dated to 3680 ± 30 , Beta-440527, $\delta^{13}\text{C} -20,3$, $\delta^{15}\text{N} 8.3$. Calibrated age 4140-3925 cal BP.

The sample C62 (archaeological ID F91916) is a tooth sample (Canine, mandible, dx) from a Bronze Age settlement site at Apalle, Övergran, Uppland, Sweden. The jaw was found just outside the northern wall of one of the largest longhouses at the settlement. Possibly it was deposited in the wall during the construction of the house. Both the jaw and the house have been radiocarbon dated to the period between 3339-2885 cal BP (2 sigma), Ua 8826, 2945 ± 75 BP. $\delta^{13}\text{C}$, -19,6 (74, 75).

Lake Baikal (sample IDs: OL4223, C26, C27). The Pad' Kalashnikova site is located on the Angara River downstream of Lake Baikal in Siberia. The first specimen analyzed for this study (sample C26), a rib, was taken from a complete dog skeleton excavated in the 1950s from a dog burial (a grave) in pit #1. A portion of the rib has been directly AMS radiocarbon dated to 6122 ± 31 (Ox23910) (76). The dog is assigned to the Early Neolithic cultural period, which in Eastern Siberia refers to hunter-gatherer populations that utilized pottery and ground stone implements. To be consistent with other archaeological terminology used here, it is referred to in the text as a Mesolithic dog. The second Pad' Kalashnikova sample (sample OL4223) came from whole dog skeleton buried (a grave) in pit #2, also excavated in the 1950s. A rib from the skeleton was directly AMS radiocarbon dated to 6075 ± 32 (Ox23911), also placing it in the Early Neolithic (in this study Mesolithic) period (76). Stable carbon and nitrogen isotope analysis revealed that both dogs had diets rich in freshwater protein, likely Angara River fish (76).

Shamanka II is an Early Neolithic (i.e. Mesolithic) hunter-gatherer cemetery on the south shore of Lake Baikal. A whole dog skeleton found in 2005 within grave #26 was sampled for this study (sample C27). The grave also contained disarticulated human remains. A vertebra from the dog skeleton was directly AMS radiocarbon dated to 6430 ± 35 (Ox20561) (77). Stable carbon and nitrogen isotope analysis indicates that the dog had a diet rich in protein from freshwater sources, most likely from Lake Baikal seal and fish (77).

Spain (sample ID: OL4029). Marizulo cave is located in Urnieta (Gipuzkoa, Basque Country). In 1961 M. Laborde explored the cave and the same year J.M. Merino made a test pit, then between 1962 and 1967 J.M. Barandiarán carried out extensive excavations. At the base of Layer I, in the square 11C, a human burial was found. It contained a young male adult, along with an almost fully preserved adult middle-sized dog (*Canis familiaris*) a 3 month's old lamb (*Ovis aries*) and several Eneolithic cultural remains (78, 79). This Layer contained 271 archaeozoological remains (NISP) of wild fauna (*Cervus elaphus*, *Capreolus capreolus*, *Sus scrofa*, *Capra pyrenaica*, *Meles meles*, *Martes martes* and *Felis sylvestris*) and 107 remains (NISP) of domestic animals (*Canis familiaris*, *Ovis aries* and *Ovicapridae*). Most of them, except for the bones associated with the burial, show traces of butchery and burning (80). The grave context was dated by ¹⁴C (5285 ± 65uncalBP GrN 5992), and a recent re-dating (present study) specifically dated dog remains (*Canis familiaris*) (5390 ± 34uncalBP OxA-36895 D012576).

Serbia (sample ID: AL2946). Pločnik is a Late Neolithic site located 19 km west of the town of Prokuplje in southern Serbia. Based on its unique and abundant ceramic finds the late phase of the Vinča culture was named after Pločnik. Besides numerous copper artefacts, traces of early metallurgical activity are found across the settlement deposit, as well as the unique find for the central Balkans of an in situ metallurgical workshop for casting and/or repair of metal tools. This settlement was occupied for c. 600 years, between 5200 and 4650/4600 cal BCE. The estimated settlement size is c. 30 ha (which refers to the size of the uppermost cultural layer), with a c. 3.6 m thick cultural layer. However, a total of only c. 0.2 ha was explored so far, even though the site has very long tradition of archaeological excavations. It has been excavated in several campaigns since its discovery in 1926 (81–87). The latest excavations were conducted in 2012–2013 as a part of the UK's AHRC-funded project "Rise of Metallurgy in Eurasia", and the main goal was to understand the role of early metallurgy in the everyday life of Pločnik inhabitants, and beyond, within the Vinča culture. During the most recent excavation campaigns a large amount of animal remains (TNF=12868) was discovered, recorded and analyzed. Domestic cattle remains dominate the assemblage (NISP=4000) with sheep and goat (taken together) the second most numerous taxon, followed by domestic pig, while red deer was the most abundant hunted species. Dog remains comprise 1.5% of the Pločnik assemblage (NISP=61). The dog specimen, a left mandible (AL2946), analyzed in this study was found in feature 22 (daub concentration – possible part of collapsed oven), layer 16, at the relative depth of 1.8 m. The feature belongs to the third horizon (Vinča C/Gradac I) of the Pločnik site. The Pločnik dog specimen was directly dated to 6930–6747 cal BP.

Srubnaya steppe (sample ID: C5). The Bronze Age steppe dog sequenced here (sample C5) comes from a site known as Krasnosamarskoe, located close to the city of Samara, and associated with the Srubnaya culture. A large number of canid remains, mostly from male individuals, have been found at this site. The remains show evidence of unusual butchering patterns and likely consumption by humans, suggesting they might have been ritually sacrificed (88). The site has been dated to between 1900 and 1700 BCE based on radiocarbon dating of multiple human and animal bones (88).

Tel Hreiz, Israel (sample ID: THRZ02). Tel Hreiz is located off the Carmel coast of northern Israel. It is one of fifteen sedentary agro-pastoral-fishing villages, dated to the late Pottery Neolithic (PN), 8th millennium BP, Wadi Rabah culture. These settlements were inundated by

post-glacial sea level rise and currently lie at depths of 0-5 m. Finds from Tel Hreiz include architectural remains—partially preserved stone-built rectangular and rounded structures, concentrations of wooden poles (possibly huts), and recently, a submerged seawall built to protect the village from the rising sea. In addition, abundant flint, wood and bone artefacts were found as well as ground stone vessels, pottery, waterlogged plant remains including hundreds of olive pits, human burials, and fauna - primarily representing domestic herd animals, as well as the dogs sampled for this study. The biometry of the dogs has been described by Dayan and Galili (89) and more recently by Horwitz *et al.* (90) who noted their rounded skull and relatively short nasal area, making them similar to local pariah dogs rather than sighthounds like the Saluki.

Tel Gezer, Israel (sample ID: TGEZ06). The site is an ancient mound located in the foothills of the Judean hills, overlooking the coastal plain of central Israel (Shephelah) some 8 km southeast of the modern city of Ramla. Excavations at the site have been undertaken by several different teams since 1902 (91), the most recent led by Steven Ortiz and Samuel Wolff (92). The mound has yielded a rich corpus of remains spanning the Late Chalcolithic to the Roman period with peak occupations in the Middle Bronze, Late Bronze and Iron Ages. Five of the eight dogs recovered during the recent excavations were sampled for this study. With the exception of one dog, dated to the Iron Age II which was found in a fill beneath a large 9th century BC wall, all derive from a three-roomed Hellenistic building in an area within the structure measuring 10 x 5m. None of the dogs were accompanied by burial goods or associated with any specific architectural features, and all were interred directly in undistinguished fills or simple pits dug into the earth and so resemble other clusters of “dog burials” reported from numerous Iron Age II, Hellenistic and Persian period sites from the southern Levant. Numerous hypotheses have been offered to account for these burials, since they are usually complete, in anatomical articulation, lie on their side with flexed limbs, include both sexes and various age groups, but lack signs of disease or butchery. In terms of their morphology, as described by Horwitz *et al.* (90), the Gezer dogs most closely resemble local pariah dogs like the local Cnaani breed.

Horbat ‘Uza, Israel (sample IDs: UZAA01, UZAA02). Horbat ‘Uza is a site that lies at the southern foot of a hill situated in the middle of the ‘Akko Plain, less than 10 km inland from the town of Acco (Acre) on the Mediterranean Sea. Salvage excavations were undertaken at the site in 1963, and more recently in 1991, on behalf of the Israel Antiquities Authority (93). The excavators identified an extensive occupation history with four major stratigraphic units spanning the 7th millennium BC Pre-Pottery Neolithic through to the Mamluk/Ottoman periods (15th/16th centuries AD). Multiple phases were recognised in each unit although exposure of the earliest levels (Neolithic through Chalcolithic) was limited in extent. In all phases, domestic and/or industrial architectural remains were found associated with remains of material culture (lithics, pottery etc.) and fauna. The dogs analysed here, derive from salvage excavations carried out in 1991 under the direction of Nimrod Getzov. One specimen dates to the Early Byzantine period (340-410 AD) while another dates to the Crusader period (first half of the 12th century AD). Neither were found in special contexts. The two dogs differ in their cranial proportions but were intermediate in size between the smaller Persian/Hellenistic dogs (such as those from Tel Gezer) and the larger Chalcolithic-Neolithic dogs (such as those from Tel Hreiz) (90).

Tell Ashkelon, Israel (sample IDs: ASHQ01, ASHQ06, ASHQ08). This site is a large port on the southern Mediterranean coastal plain of Israel, about 60 km south of Tel Aviv and 15 km north of Gaza. The 150-acre site encompasses an occupational history of more than 4,000 years, from the 4th mill. BC through to the 13 cent. AD. During most of these periods the site functioned as a thriving commercial center. Major excavations at the site were carried out since the mid-1980's until 2016 by the Leon-Levy Expedition headed by Lawrence Stager and Daniel Master (94). The faunal material retrieved included thousands of animal bones from various domestic and public contexts. Among the most important zoo-archaeological finds was the discovery of in excess of 1400 dog burials of complete or partial skeletons in the southwest to central section of the city at the tell's edge, most concentrated in grid 50, west of an earlier warehouse (95, 96). These date mainly to the Persian period but the practice extended into early Hellenistic levels (mid to late 6th to early 4th century BC). Many of the dogs were found in articulation, both sexes were buried and many are young puppies (>60%). None of the dog skeletons showed evidence of trauma or fatal injuries that could explain the cause of death. Each dog was buried separately in a discrete pit dug into the ground. The burials were not oriented in any consistent direction, they were not marked and lacked any grave goods. Most dogs were buried on their side with tails arranged between the feet. Bone measurements indicate that the dogs varied in their size and conform to the appearance of local, unmanaged, free-roaming urban dogs.

Croatia (sample IDs: ALPO01, SOTN01). These two samples are from around the villages of Aljmaš and Sotin, respectively, both along the Danube River. Both dogs are from Eneolithic contexts.

Aljmaš is a village located in Eastern Croatia, on the right bank of the Danube River some 25 km east of the city of Osijek. The Podunavlje site is situated in the centre of the village on a wide plateau rising steeply above the Danube. Rescue archaeological excavations carried out between 2000 and 2005 by the Museum of Slavonia in Osijek revealed a multi-layered site dated to the Neolithic, Eneolithic, Middle Bronze Age and the Early Modern period. During the excavations conducted in 2001 a pit of irregular form (SU 59/60) was uncovered in the eastern section of the excavated area among Baden culture settlement features. The filling of the pit contained a huge amount of pottery, lithic material, artefacts made of animal bones and numerous fragments of unprocessed animal bone. At the bottom of the pit an almost complete bovine skeleton was unearthed. Typological characteristics of the pottery date the use of the pit to the Eneolithic, more precisely, the Baden culture, while radiocarbon analysis of the left bovine ulna dates the burial to 4445 ± 105 BP (IRB Z-3106). Directly below the cattle skeleton at the same depth, a skeleton of a small canid was uncovered. An almost complete skeleton of the several months old individual was lying on its right side and was oriented east-west with the head toward the east. The animal's hind limbs were crouched and crossing each other while the fore limbs were laid up together towards the rest of the body. A pebble slightly larger than the dog's head was placed between the mandible and the chest. The bones were nicely preserved without any traces of pathological lesions.

Sotin is a multi-layered archaeological site on the Danube River in eastern Croatia with a continuity of settlement from the Neolithic to the present day. The site is located on a high loess plateau on the right bank of the Danube in a landscape where loess plateaus and deep gullies

alternate, representing a natural communication towards the river. Between 2008 and 2018, during the excavation of the Late Bronze Age and Early Iron Age cemeteries at the Sotin - Srednje polje site, a Copper Age settlement was also discovered. In 2008, at the northern edge of the Eneolithic settlement in trench 2, a round pit 1.2 m in diameter with a dog burial was discovered. The pit was filled with brown soil in which several fragments of Copper Age pottery were found. The Sotin dog was directly dated to 4970-4830 cal BP / 3020-2880 cal BC (Beta-307577) which indicates the period of Baden and/or Kostolac Copper Age cultures in eastern Croatia. A complete skeleton of a young dog, around 1 year of age, was placed on its right side with all four legs folded together in front of the body.

Siberia historical (sample IDs: F3781, C32). These two samples were found in Siberia by local mammoth tusk collectors, and are both likely historical (i.e. less than 100 years old). The C32 skull was found in 2014, 55 km north-west of Tumat village in the Yana-Indigirka lowland, and was donated to the Mammoth Museum in Yakutsk. This area is rich in Pleistocene animal remains, but no archaeological finds have been made. The F3781 skull was found in 2013 on the coast of the East Siberian Sea, 100 km east of the Ambarchik settlement. No archaeological finds have been made in the nearby area.

Ancient DNA extraction and library preparation

Samples C62, C88, C89, C90 and C94 were processed in dedicated ancient DNA facilities in the Archaeological Research Laboratory at Stockholm University, Sweden. We targeted the cementum-rich root tip of teeth as previous studies show that cementum preserves DNA better than most other types of bone (97). We first wiped the teeth thoroughly with 1% sodium hypochlorite, and then ddH₂O, and 70% ethanol, followed by UV-irradiation (254 nm at <10 cm distance) for 20 min/side, and obtained ca. 75 mg tooth powder per sample using a Dremel multitool drill at the lowest possible rpm. DNA extractions were carried out in batches of five plus one extraction blank. DNA was extracted by incubating the bone powder for 24 hrs at 37°C in 1.5 mL of digestion buffer (0.45 M EDTA pH 8.0 and 0.25 mg/ml of proteinase K). Supernatant was then concentrated to 100 µL on Amicon Ultra-4 (30kDa MWCO) filter columns (MerckMillipore) and purified on Qiagen MinElute columns following manufacturer's recommendations but with an extra wash step. Purified DNA was eluted in 65 µL Qiagen Elution Buffer. DNA from all samples but C62 were initially constructed into dsDNA Illumina sequencing libraries following the protocol outlined in Meyer and Kircher (98) with modifications (99). Once we verified that the initial libraries contained canid DNA with cytosine deamination patterns expected for ancient DNA (54) we constructed uracil DNA glycosylase (UDG)-treated dsDNA Illumina libraries on all DNA extracts (including C62). The UDG-treated libraries were constructed and prepared for sequencing as described above but with an extra incubation step during blunt-end repair with USER enzyme (NEB) following the protocol outlined in (100).

Samples C26, C27, C32, C5 and F3781 were processed at the dedicated ancient DNA facility at the Swedish Museum of Natural History in Stockholm, Sweden. Compact bone or teeth were targeted for sampling. After thorough cleaning and physical removal of the outermost surface layer, approximately 75 mg of homogenized material was extracted using a Dremel multitool drill. DNA was extracted using a silica-based method with concentration on Vivaspin filters (Sartorius), according to a protocol optimized for recovery of ancient DNA (101). 20 µl of DNA extract was used to construct dsDNA sequencing libraries for the Illumina platform following

Meyer and Kircher (98). Libraries were treated with USER enzyme (New England Biolabs) during the blunt-end repair step, to remove uracils deriving from cytosine deamination (102). Indexing and amplifying was performed with high-fidelity polymerase AccuPrime Pfx (Life technologies). Magnetic beads (Agencourt AMPure XP, Beckman Coulter) were used for cleaning and libraries were checked for concentration and fragment length on the Bioanalyzer 2100 (Agilent).

Samples AL2022, AL2397, AL2571, AL2946, OL4029, OL4061, OL4222 and OL4223 were processed at the dedicated ancient DNA facility at the PalaeoBARN laboratory at the University of Oxford. Illumina libraries were built following (98), but with the addition of a six base-pair barcode added to the IS1_adapter.P5 and IS3_adapter.P5+P7 adapter pair. The libraries were then amplified on an Applied Biosystems StepOnePlus Real-Time PCR system to check that library building was successful, and to determine the minimum number of cycles to use during the indexing amplification PCR reaction. A six base-pair barcode was used during the indexing amplification reaction resulting in each library being double-barcoded with an “internal adapter” directly adjacent to the ancient DNA strand and forming the first bases sequenced, with a traditional external barcode sequenced during Illumina barcode sequencing.

Samples ALPO01, SOTN01, ASHQ01, ASHQ06, ASHQ08, THRZ02, UZAA01, UZAA02 and TGEZ06 were processed at the dedicated ancient DNA facility at University College Dublin (UCD). We sampled the petrous portion for all samples. All facilities at UCD used for processing ancient samples are physically located from other molecular biology laboratories, and measures are taken to minimize contamination of ancient individuals, including head-to-toe suits, face masks, hair nets, multiple layers of gloves, bleaching of all surfaces and UV decontamination of all (non-sensitive) reagents. All laboratory tools used to process samples were decontaminated using bleach (1:5 concentration) and UV irradiated in a cross-linker. The final step of library preparation (amplification) was performed outside the ancient DNA laboratory. We included extraction negative controls (no powder) and library negative controls (extract was supplemented by water) in every batch of samples processed and carried them through the entire wet laboratory processing to test for reagent contamination.

Bone samples were UV irradiated in a cross-linker for 15 minutes on each side. A sandblaster (Renfert) was then used to isolate the cochlea from the petrous pyramid. A MixerMill (Rensch) was then used to powder the sample, resulting in 50-75 mg of bone powder for extraction. All samples were then extracted according to the silica-based protocol of Dabney *et al.* (103) with the modification of the binding apparatus using Qiagen MinElute columns being replaced with the Roche Large Volume Viral Extender tubes. For each sample, 50-75 mg of powder was digested in 1 mL of extraction buffer containing 0.45 M EDTA (pH 8.0) and 0.25 mg/mL of proteinase K. After around 18h incubation, samples were centrifuged at 13,000 rpm for 2 m. The supernatant was then added to 13 mL of binding buffer consisting of 5M Guanidine hydrochloride (MW 95.53), 40% Isopropanol, 0.05% Tween-20, and 9 mM Sodium Acetate in the Roche Extender columns. The 50-mL columns were centrifuged for 4 m at 1,500 x g. The columns were then washed twice in 650 mL of PE buffer (Qiagen), centrifuging at 6,000 rpm between washes. After the second wash, samples were centrifuged at 13,000 rpm to remove any residual ethanol. Samples were eluted two times in 25 µL of EBT after 10 m incubation at 37°C and centrifuged at 13,000 rpm for 30 sec, for a total of 50 µL extract. Extraction batches included negative controls (no DNA template).

Illumina blunt-end libraries were built following the protocol of Meyer and Kircher (98), with modifications (104). Libraries were prepared using 12.5-25 μL of DNA extract in a 70 μL reaction volume for the blunt end repair step and 40 μL reaction volumes for the adapter ligation and adapter fill-in steps. Blunt end repair was performed using NEB End Repair module. Samples were incubated for 15 m at 25°C followed by 5 m at 12°C. Ligation was performed using T4 DNA ligase buffer (10X), PEG-4000 (50%), adapter mix following Meyer and Kircher (98), and T4 DNA ligase (5U/ μL). Samples were incubated for 30 m at 22°C. Adapter fill-in was performed using Thermopol Reaction Buffer (10X) dNTPs (10mM each), and Bst large fragment (8U/ μL). Samples were incubated for 30 m at 37°C and enzyme inactivation was completed by incubating for 20 m at 80°C. Sample cleanup between blunt end repair and ligation, and between ligation and adapter fill-in, was performed using the MinElute PCR Purification kit (Qiagen) according to manufacturer's instructions. Every batch of libraries contained a negative control using nuclease-free water.

The library amplification step following adapter fill-in was set up using Accuprime pfx Supermix, primer IS4 (10 μM), a specific indexing primer (0.2 μM), and 3 μL of library template, for a total of 25 μL reaction mixture. Amplification took place under the following thermal cycling conditions: 5 min at 95°C; 12 cycles of 15 sec at 95°C, 30 sec at 60°C and 30 sec at 68°C; and a final extension step of 5 min at 68°C. The resulting PCR product was purified using the MinElute PCR Purification kit (Qiagen) as described previously. Quantification and quality assessment of the amplified libraries was performed on an Agilent 2100 Bioanalyzer, using a DNA-1000 chip, and a Qubit 2.0 Fluorometer, following manufacturer's protocols. Amplification steps included a blank negative control.

Sequencing

We sequenced all 19 samples processed in Stockholm and Dublin on 1 HiSeq X lane each to obtain ~400 million 2x125 paired end sequences, without any multiplexing, to avoid issues with crosstalk among adaptors. An exception was the Sotin Croatian dog (SOTN01) which was sequenced on two lanes.

For the 8 samples processed in Oxford, up to 200 libraries with unique barcode combinations were pooled at equimolar levels (as determined by an Agilent Technologies 2200 TapeStation) and an 80bp run was carried out on an Illumina 2500 sequencer. Following alignment to the reference genome promising-looking samples were subsequently sequenced to greater depth on multiple Hi-Seq 2500/4000 lanes.

Genome sequence data processing

All sequence read data resulting from paired-end sequencing was merged using seqprep requiring an 11bp overlap minimum. We aligned the resulting merged and adaptor-trimmed sequences to the dog canFam3.1 genome using BWA (105) with permissive parameters including disabled seed (106) (-l 16500 -n 0.01 -o 2). Duplicates were removed by keeping only a single sequence among any set of sequences that had the same orientation, length, and start-end coordinates. Five previously published ancient genomes (9, 14, 21) were also incorporated into the dataset (table S3). We then generated pseudo-haploid genotypes for each sample by sampling a random allele using htsbox pileup (<https://github.com/lh3/htsbox>), restricting to reads

with a length of at least 35 base pairs, a mapping quality of at least 20 and to bases with a base quality of at least 30. Post-mortem damage was quantified using *PMDtools*(107) with the “--first” and “--CpG” arguments.

Sex determination and male overrepresentation

We determined the biological sex of each individual, including previously published ancient dog genomes, by calculating the ratio of coverage on the X chromosome to coverage on the autosomes, and found that 19 of the sequenced individuals were male and 8 were female (**table S2**). Including five previously published genomes (Newgrange (9): male; Cherry Tree Cave (14): male; Herxheim (14): male, AL3194 Port au Choix (21); male, AL3223 Weyanoke (21): female) (**table S3**), we find an overrepresentation of males: 23 males (71.9%), 9 females (28.1%), $p = 0.0201$, Binomial test. This mirrors observations of male overrepresentation among recovered remains in other large mammals (108, 109), but as most of the ancient dogs analyzed here were found in human archaeological contexts the bias in dogs might not necessarily be driven by the same factors as in other species.

Contamination estimation for domesticates and their wild progenitors

Current tools to estimate mitochondrial genome contamination are developed for use on human ancient DNA. Since contamination (and cross-contamination) could also be an issue for domesticates and other animals, we used a software tool—*CALICO* (<https://github.com/pontusssk/calico>)—that allows contamination estimation for domestic animals using databases of their mitochondrial DNA variation. We used a data set of dogs and wolves (n=788), cattle (n=281), sheep (n=23,609), pigs (n=311), horse (n=334), goat (n=51) and chicken (n=67) mitochondrial genomes assembled and curated by Shi *et al.* (110).

Following a previous approach (111) we obtain mitochondrial DNA contamination estimates by identifying informative sites in the form of private or near-private (*e.g.* <1% in the 788 canid mtDNAs) consensus alleles in each ancient individual. Any contamination at these sites can reasonably be assumed to carry the alternative allele that is observed in >99% of modern dogs and wolves. We obtain a point estimate c of mtDNA contamination by tallying the counts of the consensus and the alternative bases, assuming independence, where $\hat{c} = N_{\text{alternative}} / (N_{\text{consensus}} + N_{\text{alternative}})$. If no alternative allele was found, we estimate an upper confidence limit on contamination using a binomial distribution:

$$P = \binom{N}{k} c^k (1 - c)^{N-k}$$

Where $N = N_{\text{consensus}} + N_{\text{alternative}}$, and $k=0$. The upper confidence limit is set to the value of c at $P=0.05$. When alternative alleles were observed, a 95% confidence interval are computed using a binomial approximation:

$$95\% \text{ CI} = c \pm 1.96 \sqrt{\frac{c(1 - c)}{N}}$$

To validate the approach we simulated a total of 100,000 sequence reads 70 bp long and with an error rate of 0.001 from two randomly chosen mtDNA genomes from each species, and varied

the mixture proportion of one genome to the other from 0 to 40% (**fig. S1**). We then aligned the sequence reads to the reference mtDNA using BWA and used CALICO to estimate contamination using default parameters for each mixture proportion (filtering reads with mapping quality less than 30 and samtools BAQ recalibrated base quality less than 30). Since some pairs of mtDNA genomes are near-identical and do not contain any informative sites, we performed this simulation procedure for three randomly drawn pairs and selected the one with the most informative sites (note that increasing the number of informative sites should not decrease any potential bias in the test). The number of informative sites for the simulations plotted in **fig. S1** are cattle=2, sheep=1, dog=1, horse=3, goat=1, elephantid=6, pig=7, chicken=5, human=4.

Merging with modern genomes

We obtained the “722g” VCF file containing genotypes called from whole-genome sequencing data from 722 dogs, wolves and other related canid species compiled by the NHGRI Dog Genome Project (*112*) (NCBI BioProject accession PRJNA448733) from previous studies (*5, 8, 32, 113*).

We also processed and incorporated whole genome sequencing data from a few additional studies: four African Golden wolves and 15 Nigerian dogs (Genome Sequence Archive (<http://gsa.big.ac.cn/>) accession number PRJCA000335) (*32*), 12 Scandinavian wolves (European Nucleotide Archive accession number PRJEB20635) (*114*), 9 North American wolves and coyotes (European Nucleotide Archive accession number PRJNA496590) (*115*) and 8 other canids (African Hunting Dog, Dhole, Ethiopian Wolf, Golden Jackal, Middle Eastern gray wolves) (European Nucleotide Archive accession number PRJNA494815) (*116*). Reads from these were mapped to CanFam3.1 using bwa mem version 0.7.15 (*117*). Duplicate reads were marked using the MarkDuplicates tool from Picard Tools version 2.18.12 (<http://broadinstitute.github.io/picard/>). They were then genotyped only on the sites that were present in 722g VCF file using GATK HaplotypeCaller (*118*) version 3.6 with the “-gt_mode GENOTYPE_GIVEN_ALLELES” argument, and merged into the 722g VCF using bcftools merge (<http://www.htslib.org/>).

The merged VCF was then filtered as follows:

- Sites displaying any allele with excess heterozygosity (“ExcHet” annotation p-value < $1 \cdot 10^{-6}$, as computed using the bcftools fill-tags plugin) were removed.
- Indel alleles were removed by setting the genotype of any individual carrying such an allele to missing (thereby retaining any overlapped SNP alleles).
- The genotypes at any site with a depth lower than one third of the genome-wide average for the given sample or lower than 5 or higher than twice the genome-wide average for the given sample were set to missing (using the sum of the “AD” field, rather than the “DP” field, as the latter was found to be inaccurate for certain sites).
- The genotypes for any sample carrying any allele other than the two most common alleles were set to missing (thereby retaining only the two most common alleles at multi-allelic sites).
- Allele representation was normalized using bcftools norm.
- Sites where 130 or more samples had missing genotypes after these filters were removed.

After these filtering steps, a total of 67.8 million SNPs remained. The genotypes of ancient samples were then added onto this VCF, setting the genotype of any ancient sample carrying an allele not already present in the VCF to missing.

We restricted all analyses to transversions only, leaving 23.1 million SNPs. For certain analyses we wished to further reduce the number of SNPs so as to lower the computational burden. We did this by ascertaining variants that have a heterozygous genotype in a single Coyote individual from California, resulting in approximately 1 million transversion SNPs (the exact number depending on the set of modern dog, wolf and other canid genomes that were retained for a given analysis). This form of ascertainment in an outgroup population has been shown to provide unbiased f -statistics (119).

We largely retained the population labels/descriptions of modern genomes from the 722g VCF metadata, just slightly modifying some labels for consistency. We manually classified modern dog populations into six broad regional groups on the basis of their geographical origin: Africa, Siberia & the Arctic, East Asia, Middle East & North Africa & South Asia, Sahul and Europe. Because unbalanced representations of ancestries can distort the results of analyses such as principal component analysis and model-based clustering, for such analyses we restricted to a hand-picked global selection of 44 modern dog populations (for all regions other than Europe and East Asia, from which much larger numbers of modern dog genomes are available, these represent all the populations available in our dataset with likely origins in those regions):

- **Africa:** AfricaVillageUnknown, Basenji, NamibiaVillage, NigeriaVillage
- **Siberia & the Arctic:** AlaskanHusky, AlaskanMalamute, EastSiberianLaika, GreenlandDog, Samoyed, SiberianHusky
- **East Asia:** BorneoVillage, ChinaVillageDiqing, ChinaVillageGuangdong, ChinaVillageHebei, ChinaVillageLiaoning, ChinaVillageLijiang, ChinaVillageShanxi, ChinaVillageXinjiang, ChowChow, TibetanMastiff, VietnamVillage
- **Middle East & North Africa & South Asia:** AfghanHound, CaucasianOvcharka, EgyptVillage, IndiaVillage, LebanonVillage, QatarVillage, Saluki, Sloughi
- **Sahul:** AustraliaVillage, Dingo, NewGuineaSingingDog, PapuaNewGuineaVillage
- **Europe:** Boxer, Chihuahua, Dalmatian, EnglishCockerSpaniel, FinnishLapphund, GermanShepherdDog, LagottoRomagnolo, PeruvianIncaOrchid, PortugueseWaterDog, SwedishLapphund, Xoloitzcuintli

Analyses of human ancient DNA data

For as many of the ancient dog genomes as possible, we searched the human ancient literature for samples that were close to the dog in terms of geographical location, sample date and cultural context (41, 43, 44, 46, 54, 120–126). We compiled and merged this data across the identified studies (**table S4**), using as many individuals from a given labelled population or site as available. We also similarly included data from a few modern human populations (127) to serve as matches for modern dog populations. In cases when genotypes were not available from the studies themselves, we called pseudo-haploid genotypes from the published read alignments as described above for the dog ancient DNA data (except using a minimum mapping quality threshold of 30, following that most widely used in human ancient DNA studies). As most of the identified human data had been generated using targeted capture, we restricted analyses to sites

that were on the 1240k capture array, and further restricted to transversions only, for a total of 208,893 SNPs.

We then performed corresponding population genetic analyses on the human dataset in the same way as we did for the dog dataset. When fitting admixture graphs to the human data, we used high-coverage modern Native American genomes (the Karitiana group) from the Simons Genome Diversity Project dataset (127) in place of the ancient American genomes that we had matched to the ancient dog samples, to increase the statistical power to reject graphs.

To align dog and human PCA results, we used the “procrustes” function from the R package *vegan* (128), and the associated “protest” function to calculate the statistical significance of the Procrustes correlation through a randomization test.

***f*-statistics analyses**

We calculated f_3 - and f_4 -statistics using AdmixTools (119) v5.0. To overcome excessive memory usage when calculating very large numbers of f_4 -statistics, we split the genome into 220 chunks of size 10 Mb, calculated the statistics with AdmixTools on each one of these chunks independently and then performed our own weighted block jackknifing (129) across the chunks to obtain standard errors.

When calculating f -statistics we pooled certain individuals under shared population labels, if they were close in space and time and were consistent with being a clade genetically in exploratory ancestry analyses: C26, C27 and OL4223 were pooled into a “Lake Baikal” population, ASHQ01, ASHQ06, ASHQ08 were pooled into an “Ashkelon” population, C89 and C90 were pooled into a “Sweden PWC” population and the previously published AL3194 and AL3223 were pooled into an “America” population.

We calculated f_3 statistics to test for evidence of admixture in the history of dog populations. None of the major lineages that we use for admixture graph fitting display any negative f_3 statistics (indicating allele frequencies on average being intermediate between those of two sources), which would constitute unambiguous evidence of mixed ancestry. For example, our admixture graph and f_3 cline results imply that European dogs likely are a mixture of Levant-related and Siberian-related ancestries, but they do not display negative f_3 statistics when such populations are used as sources. The failure of tree-like admixture graphs without admixture events, however, suggests that the lack of negative f_3 -statistics among the major lineages must reflect a lack of power, likely due to drift after the admixture events.

In outgroup f_3 analyses, we excluded the Boxer breed from the plots as it showed an increased affinity to ancient genomes relative to other European dogs. This likely reflects reference bias in the ancient genomes, as the canFam3.1 reference genome was constructed from a Boxer individual. Reference bias would thus manifest itself as an artifactually high affinity to individuals of Boxer ancestry.

In the observed outgroup f_3 cline across ancient European dogs (**Fig 1C**), the tight cluster of modern European dogs appear slightly shifted down on both axes relative to the ancient cline. It is possible that this could reflect the recent addition into modern European dogs of some novel

ancestry, but if so this would need to be an unsampled ancestry that is basal to both the ancient Levantine and the ancient Baikal dogs. Instead, we believe the most likely explanation of this apparent downward shift is technical and reflects ancient attraction, i.e. due to correlated ancient DNA damage patterns, between the Levantine and Baikal dogs and the other ancient genomes in the dataset. The f_3 values for the ancient dogs are thus likely slightly over-estimated.

f_4 -based multivariate analysis

To study ancient dog population structure, we applied a new approach based on computing all possible f_4 -statistics (119) involving a given set of populations and then performing principal component analysis on the resulting matrix, essentially treating each statistic $f_4(X,A;B,C)$ as an independent data point for population X. We set all statistics of the form $f_4(A,A;B,C)$ to 0. This is similar to but slightly different from a previous approach that used statistics where populations A,B and C are outgroups to X (130), while we here use all possible populations within a given analysis. This approach allows us to study population structure among ancient genomes without having to project them onto axes of differentiation defined by modern genomes.

Clustering analyses

We used the ADMIXTURE program (131), to explore population structure in an unsupervised fashion, restricting analyses to variants with a minor allele frequency of 5% or higher (**fig. S2**).

Admixture graph inference

We sought to find explicit population history models that can explain the relationships among ancient and modern dog populations. We selected a number of individuals or populations that we think represent key ancestries on the basis of f_4 -statistics, principal component analyses and model-based clustering analyses as well age and geographical location: the 7 ky old Levantine dog, the 7 ky Lake Baikal dogs, the 10.9 ky old Karelian dog, ancient American dogs and the modern New Guinea Singing Dog (as a representative of South-Eastern Eurasian dog ancestry, displaying less western affinity than all analyzed dogs from China and Vietnam (**Fig 1C**, **Fig 3D**, **Fig 5A**, **table S6**), and most likely a recent arrival in Sahul from South-East Asia (26)). To serve as an outgroup we included an Andean fox individual (while for other analyses we have used the Californian Coyote as an outgroup, for these admixture graph analyses we chose the Andean Fox instead to avoid any confounding effects of possible gene flow between Coyotes and the American dogs included in these analyses(21)).

We then used the *admixturegraph* R package (132) to exhaustively fit to the observed f_4 -statistics all 135,285 possible admixture graphs (133) that relate these six populations with up to two admixture events. Each graph was fitted to the data five times, retaining the best scoring fit out of these. No graphs without admixture events provided good fits between predicted and observed f_4 -statistics, demonstrating that admixture between distinct lineages must have occurred during the early formation of dog ancestries. Only one graph, involving two admixture events, fits the data without outlier statistics ($|Z| > 3$) (**fig. S4**). Two other graphs, with very similar structure to the best fitting graph, both nearly fit the data with two minor outlier f_4 -statistics. After these three, the next best fitting graphs are a set of three with a somewhat different structure from the best fitting graph, but these provide considerably worse fits with five outlier statistics each. Given the existence of fitting graphs with two admixture events, we did not attempt to fit less parsimonious graphs with three or more admixture events (the number of all possible such

graphs is also too large to be tested exhaustively). For the graph displayed in **Fig. 1E**, we manually grafted on the European Neolithic dog, which was not part of the exhaustive search, and this extended graph fit without any outlier statistics. We also repeated the exhaustive admixture graph search using f_4 -statistics computed on variants ascertained to be heterozygous in the Californian Coyote individual, and found that the fitting graph from the above analysis remained the only graph to fit with 0 outlier statistics.

We performed an exactly analogous admixture graph analysis on human data, using the human populations matched to the dog samples with regards to archaeological age, geographical location and cultural context (**table S4**). We found six graphs that fit the human data without outlier statistics. We then determined how well the graph topologies that were found to fit the dog data could explain the human data, and vice versa, using “best_error” score computed for each graph. The three graphs that fit, or nearly fit (with two minor outliers), the dog data rank among the 0.82 to 2.8% top scoring topologies in the human search, and the six graphs that fit the human data rank among the 0.0074 to 1.2% top scoring topologies in the dog search (**fig. S9**). Results are similar if ranking on the number of outlier f_4 -statistics rather than the error score: the three dog graphs rank among the 0.41 to 0.54% of topologies with the lowest number of outliers in the human search, and the six human graphs rank among the top 0.0059 to 0.53% of topologies with the lowest number of outliers in the dog search. Across the full set of 135,285 possible graphs, the relative rank of the error scores achieved by each graph are strongly correlated between the two species: Spearman’s rank correlation = 0.714, $p < 2.2 \cdot 10^{-16}$ (though we note that the magnitude of this p-value is likely overestimated as the graphs are not independent).

Genetic cline simulations

We performed coalescent simulations using *ms* (134) to help distinguish between possible population histories underlying the observed diagonal cline of outgroup f_3 values across ancient European dogs. We considered three possible scenarios: 1) phylogenetic structure, in which different lineages diversify without any admixture between them, but they experience different amounts of genetic drift, 2) continuous gene flow, in which neighbouring populations experience low but constant rates of bidirectional gene flow between each other, 3) admixture, in which two differentiated source lineages admix to give rise to admixed populations with varying degrees of ancestry from each of the two sources. The command lines were: split model <ms 450 1000000 -s 1 -I 9 50 50 50 50 50 50 50 50 50 0 -ej 0.2 1 2 -ej 0.05 9 2 -ej 0.01 3 2 -ej 0.02 4 2 -ej 0.03 5 2 -ej 0.01 8 9 -ej 0.02 7 9 -ej 0.03 6 9>, continuous gene flow model <ms 450 1000000 -s 1 -I 9 50 50 50 50 50 50 50 0 -m 2 3 10 -m 3 2 10 -m 3 4 10 -m 4 3 10 -m 4 5 10 -m 5 4 10 -m 6 7 10 -m 7 6 10 -m 7 8 10 -m 8 7 10 -m 8 9 10 -m 9 8 10 -ej 0.05 3 2 -ej 0.05 4 2 -ej 0.05 5 2 -ej 0.05 6 2 -ej 0.05 7 2 -ej 0.05 8 2 -ej 0.05 9 2 -ej 0.2 1 2> admixture cline model <ms 450 1000000 -s 1 -I 9 50 50 50 50 50 50 50 50 50 0 -ej 0.2 1 2 -ej 0.05 9 2 -es 0.01 3 0.8 -ej 0.01 3 2 -ej 0.01 10 9 -es 0.01 4 0.7 -ej 0.01 4 2 -ej 0.01 11 9 -es 0.01 5 0.6 -ej 0.01 5 2 -ej 0.01 12 9 -es 0.01 6 0.5 -ej 0.01 6 2 -ej 0.01 13 9 -es 0.01 7 0.4 -ej 0.01 7 2 -ej 0.01 14 9 -es 0.01 8 0.3 -ej 0.01 8 2 -ej 0.01 15 9>. We then conditioned on observing each variant in population 1, and estimated shared genetic drift using outgroup f_3 -statistics (51, 133) with population 1 as an outgroup using POPSTATS (135).

Genetic identity of pre-Neolithic dogs in Europe

As described in “Neolithic expansion into Europe” section of the main text, we hypothesize that the ancestry cline we observe across early European dogs, at least in part, reflects admixture between dogs associated with Mesolithic hunter-gatherer groups and dogs associated with incoming Neolithic farming groups from the Near East. While our dataset includes the 10.9 kya Karelian dog and the 4.8 kya dogs from a Swedish Pitted Ware Culture (PWC) hunter-gatherer site, it does not include any unambiguous Mesolithic dogs from continental Europe associated with ‘Western Hunter-Gatherer’ (WHG) (43) human groups, and we are therefore unable to determine what the ancestry of such dogs was. We are also unable to unambiguously estimate the proportion of ancestry in later European dogs that derives from admixture from incoming Near Eastern dogs, as we cannot confidently determine what the baseline Near Eastern affinity was like prior to the Neolithic.

We speculate however, on the basis of the Siberian-shifted ancestry of the Karelian and PWC dogs and the general nature of the European cline, that WHG dogs might also have been similarly Siberian-shifted. However, other scenarios are also possible. The PCA (**Fig 1B**) and admixture graph (**Fig. 1E**) results show that the Karelian dog is already Levant-shifted relative to the dogs at Lake Baikal, such that it could also be viewed as falling onto the European ancestry cline. There might thus already have been a cline across Europe prior to the arrival of dogs from the Near East during the Neolithic, such that the WHG dogs might have been more Levant-shifted than the Karelian dog, and that admixture with such European dogs was what caused the Levant shift in the latter. One possibility is that WHG-related dogs may be represented by the lineage leading to the Levant that contributes ~34% of the Karelian dog's ancestry in the admixture graph (**Fig. 1E**), which would be similar to the relationship of ‘Eastern Hunter-Gatherer’ (EHG) Karelian hunter-gatherers to WHG groups (45). Genomes from Mesolithic continental European dogs will be needed to resolve the identity of the hunter-gatherer dogs of Europe and their contribution to later populations.

Mitochondrial DNA analyses

In addition to the novel mitochondrial genome sequences generated in this study, we also downloaded and analysed publically available modern and ancient dog and wolf mitochondrial genome sequences from a number of publications (9, 11, 13, 21, 28, 136–140) and additional GenBank submissions (accession numbers JF342812, JF342877, KJ139384, KJ139385, KJ139386, KJ139387, KJ139388, KM061481, KM061493, KM061497, KM061534, KM061535, KM061540, KM061592, KX379528, KX379529). Sequence alignments were performed using MUSCLE (141) integrated into Aliview (142).

Evolutionary rates were estimated by using the strict molecular clock in BEAST v2.4.7 (143). For the prior on clock rates, we chose a lognormal distribution with a mean in real space of $1.0 \cdot 10^{-8}$, upper bound of $1.0 \cdot 10^{-4}$ substitutions/site/year, lower bound of $1.0 \cdot 10^{-10}$ substitutions/site/year (these bounds are actually part of a separate uniform prior and are not part of the lognormal distribution itself), and standard deviation of 1.25. The HKY+ Γ substitution model was used, with four rate categories for gamma-distributed rates across sites. We used an exponential prior for kappa and a lognormal prior for the gamma shape prior, with default parameters for both. It has been shown (144) that accounting for age uncertainty had negligible or minimal impacts on the resulting estimates in BEAST, therefore we used the mean date estimates for all of the mitochondrial genome sequences for the analysis. A constant-size

coalescent model was used as the tree prior. Posterior distributions of parameters were estimated by Markov chain Monte Carlo (MCMC) sampling. Two independent runs of chains were performed, and samples were drawn every 5,000 steps over a total of at least 50 million steps, with the first 20% of samples discarded as burn-in. Sampling was considered sufficient when the effective sample size of each parameter exceeded 100. When required, we ran additional MCMC analyses to achieve sufficient sampling. The trace files were assessed using *Tracer* (145) and the samples from the independent runs were merged using LogCombiner (146). Phylogenetic trees were visualized using iTOL v5.4 (147). On the basis of the obtained phylogenetic tree, we also classified each ancient mitochondrial sequence into the four high-level haplogroups A, B, C and D, as well as the recently suggested haplogroup X (28).

Previous studies have found that mitochondrial haplogroup C was the dominant haplogroup in Mesolithic European dogs, while haplogroup D became common during the Neolithic (27). We tested if ancient European dogs carrying these haplogroups were on average positioned differently along the European cline of Levant versus Baikal-related autosomal affinity, as quantified using the f_4 -statistic $f_4(\text{CoyoteCalifornia}, X; \text{Baikal } 7\text{kBP}, \text{Levant } 7\text{kBP})$. The two Swedish Pitted Ware Culture dog individuals had been pooled into a single population before calculating the f_4 -statistic, and we therefore assigned that same f_4 value to both individuals before performing the test. We found no difference in mean f_4 values between dogs carrying these two different haplogroup (mean f_4 haplogroup C: -0.0006591, mean f_4 haplogroup D: 0.001710, t-test p-value for difference in means: 0.1502). However, we note that with only 8 ancient European dogs with haplogroup C and 3 with haplogroup D, we have very little statistical power to detect a difference.

***qpAdm* ancestry modelling**

We used the *qpAdm* framework (45, 132) to test ancestry models and estimate ancestry proportions for dog populations. For a given set of potential source populations, we exhaustively tested all possible models with up to four sources and all remaining populations in the set as outgroups. We additionally included the Californian Coyote individual in the outgroup list in all models. For models with only one source we ran the *qpWave* program rather than *qpAdm*. For a given target we then ranked the models on the basis of their p-values, but prioritizing simpler models (meaning models with a lower number of sources) as follows: we set a threshold at $p=0.01$, and only ranked a more complex model over a simpler one if the p-value of the former was higher and they were both on the same side of this threshold. We also discarded any model in which any inferred ancestry proportion was larger than 1.1 or smaller than -0.1, considering these as failed models. Unless noted, *qpAdm* analyses were performed on SNPs ascertained to be heterozygous in the Californian Coyote individual, and *qpAdm* was run with the “allsnps: YES” argument.

Using this strategy, we ran two separate sets of analyses to model worldwide dog ancestries. The first was with a “comprehensive” set of possible outgroup and source populations, consisting of all ancient dog populations older than 2 ky, plus the New Guinea Singing Dog to represent a divergent Eastern Eurasian ancestry and the German Shepherd to represent modern European ancestry, for a total of 22 populations. The second was restricted to a manually selected “core” set of key populations: ancient America, Lake Baikal 7k, Levant Neolithic 7k, Iran Chalcolithic 5.8k, Samara steppe 3.8k, Karelia 10.9k, Sweden Neolithic 5k, New Guinea Singing Dog and

German Shepherd, for a total of 9 populations. The motivation for also running analyses with this core set was to obtain models that are more interpretable in terms of broader, more ultimate sources. We ran the same analyses with ancient genomes as targets, but in these cases excluding German Shepherd from the lists of sources and outgroups.

To obtain the ancestry estimates displayed on the world map in **Fig. 5A**, without having to include 22 possible source populations in the pie charts, we simplified the results of the comprehensive set of analyses as follows. We identified seven source populations that tended to be part of top ranked models, and restricted the results only to models that feature any combination of these seven populations in the sources. The outgroup lists for this subset of models are unchanged and thus retain all remaining populations in the set of 22. For each target we then picked the top ranking model from this subset of models and tidied up the ancestry proportions by setting any negative proportions to 0 and expanding the others proportionally to sum to 1. This approach produced results that are easier to visualize and interpret, but means that the p-values of the chosen model are sometimes not very high, as in some cases better sources than any of the seven populations used will be present in the outgroup list. Nonetheless, we inspected these simplified results and found that for the vast majority of targets they were appropriately reflecting the full results. Dog pictures were obtained from Wikimedia user Desaix83 under the Creative Commons 3.0 license (<https://commons.wikimedia.org/w/index.php?title=Special:Contributions/Desaix83>).

We excluded from *qpAdm* analyses a small number of modern dog populations that displayed strong reference bias, as quantified using the statistic $f_4(\text{CoyoteCalifornia}, X; \text{Boxer}, \text{canFam3.1})$ which takes advantage of the fact that the canFam3.1 reference genome is derived from a single Boxer individual and thus has an expected value of 0 in the absence of reference bias. The excluded populations were: Tornjak, QingchuanDog, ChinaVillageAnhui, XiasiDog, ChongqingDog, SharPei, ChinaVillage and BorneoVillage.

We found that when including the German Shepherd to represent modern European dog ancestry in these *qpAdm* analyses, most other European dogs can be modelled as being in a clade with German Shepherd to the exclusion of all ancient dogs (i.e. a single-source German Shepherd *qpWave* model). This is the primary line of evidence for the homogenization of European dog ancestry and the lack of continuity to most local Neolithic and earlier populations. We note that the choice of German Shepherd to represent modern European ancestry is arbitrary and a large number of other dogs would be able to serve this role equally well. Further to this, we found that when no modern European dog is included in *qpAdm* analyses, the Swedish dog C88 turns out to behave in a manner that is highly similar to a modern European dog, in that most European targets can be modelled with single-source C88 models. Most European targets receive 100% German Shepherd ancestry when German Shepherd is used as the source, but many receive slightly less (0-15% less) C88 ancestry when C88 is used instead, indicating that C88 is a very good but not perfect proxy for modern European dog ancestry (**fig. S13**). In the *qpAdm* analyses we present (**Fig 5A, Data S1**), we use results obtained with German Shepherd included as a source, as this seems to give slightly better model fits than when using C88, but, as outlined here, they are largely interchangeable.

In addition to these global *qpAdm* analyses, we performed an additional targeted analysis aiming to understand the sources of ancestry for Bronze Age and later dogs in Europe. For this we used as sources only ancient dogs predating the Corded Ware Culture and the Bronze Age, plus the Bronze Age Srubnaya dog as a proxy for steppe dogs that might possibly have entered Europe alongside migrating steppe pastoralist groups. We performed *qpAdm* analyses as above, with Bronze Age and a selection of modern European dogs as targets.

Testing for admixture from wolves and other canid species

We used f_4 -statistics to study how ancient and modern dog populations are related to present-day wolves. If statistics of the form $f_4(\text{Coyote}, \text{Wolf}; \text{Dog 1}, \text{Dog 2})$ deviate from 0, it suggests that some dog populations have a closer relationship than others to wolves, reflecting either ancestry deriving from a different wolf source population or post-domestication gene flow from wolves. Similarly, if statistics of the form $f_4(\text{Coyote}, \text{Dog}; \text{Wolf 1}, \text{Wolf 2})$ deviate from 0, it suggests that some wolf populations have a closer relationship to dogs. For these analyses, we used all ancient populations except the American dogs, as the affinity between these and the Coyote used as an outgroup confound these tests. For modern dogs we used all the 44 populations in the global representative set described above, except ChinaVillageShanxi which likely is an outlier with recent wolf admixture. We used 35 Eurasian gray wolves. To account for the multiple testing associated with computing all the possible combinations of f_4 -statistics of the above forms, we made quantile-quantile plots for each population to visually assess if the set of f_4 -values departed from what would be expected if they were normally distributed around 0.

We found that all dog populations display distributions of $f_4(\text{Coyote}, \text{Dog}; \text{Wolf 1}, \text{Wolf 2})$ values that greatly deviate from the expectation of zero (largest outlier $|Z| > 10$ for all dogs). We also found that most, but not all, wolf individuals display distributions of $f_4(\text{Coyote}, \text{Wolf}; \text{Dog 1}, \text{Dog 2})$ values that deviate from the expectation of zero (**fig. S7**). The observation of non-zero f_4 -statistics does not by itself allow us to assess if the direction of gene flow has been from dogs into wolves, from wolves into dogs, or in both directions, as these scenarios would all affect the distributions of both of the above forms of f_4 -statistics. However, we identify at least one wolf individual, Wolf35Xinjiang, with a distribution of $f_4(\text{Coyote}, \text{Wolf}; \text{Dog 1}, \text{Dog 2})$ values that is largely consistent with 0. The results for the Wolf08TibetanXinjiang individual is also largely consistent with 0 for all tests. Such an observation is not expected under a scenario of gene flow from wolves into dogs, as all wolf individuals would then display an elevated affinity to any dog population which received the gene flow. We also verified that Wolf35Xinjiang is indeed strongly on the same modern wolf lineage as the other genomes in the dataset: $f_4(\text{Coyote}, \text{Wolf35Xinjiang}; \text{GermanShepherdDog}, \text{Wolf21-M-02-15Scandinavia}) = 0.001392$, $Z = 14.3$. These results thus suggest that there has not been extensive gene flow from wolves into dogs, and that instead almost all analyzed wolf individuals harbour dog admixture.

We also incorporated data from the 35,000 year old Taimyr-1 wolf genome (4), which we processed from the published reads in the same way as described for the ancient dogs sequenced here. We could not repeat the full Q-Q plot analysis described above due to strong artifactual attraction between Taimyr-1 and ancient dogs relative to modern dogs, but instead included only pairs of ancient dogs in the analysis. While some technical noise still likely persists, the results are largely compatible with the Taimyr-1 individual being symmetrically related to the ancient dogs in the dataset (**fig S8**).

We also used f_3 - and f_4 -statistics to test for gene flow between dogs and other canid species (**table S5**). We find evidence for a previously reported affinity between ancient American dogs and Coyotes (21), although whether the signal is statistically significant or not depends on the choice of other dog used in the comparison, likely reflecting varying levels of technical noise related to e.g. sequencing coverage and reference bias.

We also used an outgroup f_3 -statistic to test for any differences between dogs in terms of basal ancestry, regardless of the source of that basal ancestry (**fig. S6**). The statistic $f_3(\text{Wolf35Xinjiang}, X; \text{CoyoteCalifornia})$ is expected to become smaller if some dog X carries ancestry from any source that is basal to the clade of dogs and wolves, as X would then share less genetic drift with members of this clade. Other than lower values for the ancient American dog, which has an affinity to Coyotes as discussed above, there is little variation across ancient and modern dogs in the values of this statistic. This is consistent with a lack of gene flow from other canid species into particular dog populations.

Amylase copy number estimation

The estimation of copy number of the AMY2B gene is complicated by the incomplete assembly of the locus in the canFam3.1 assembly. We queried the Ensembl database (148) for the locations in canFam3.1 of regions orthologous to the human AMY2B gene (ENSG00000240038), which included a region on chromosome 6 as well as three other, unplaced contigs (chrUn_AAEX03020568, chrUn_AAEX03022739, chrUn_AAEX03024353). We studied the read coverage along these regions and found that wolves and other canid species had reduced coverage along a previously identified (5) region of chromosome 6 (46,948,800-46,956,325) and the entirety of the three unplaced contigs, suggesting a lower copy number than the reference genome. To maximize power especially for low coverage ancient samples, we therefore used all of these regions, with a total size of about 75 kb, to obtain copy number estimates.

For modern wolves and other canids we used read alignments we had processed as described in the “Merging with modern genomes” section. We downloaded read alignments for a set of modern dogs from the DoGSD database (149) (<http://bigd.big.ac.cn/dogsdv2/>), with the following DoGSD sample IDs: 1735, 2972, 4669, Basenji, Dingo, Dog04, Dog06, Dog08, DQ1, DQ10, DQ2, DQ7, EG44, EG49, GS1, GS10, GS2, HR93, ID125, ID165, ID60, IN18, IN23, IN29, KM10, KM7, LB74, LB79, LB85, LJ1, LJ10, LJ3, LJ4, NA63, NA8, NA89, PG84, PT61, PT71, QA27, QA5, TM1, TM10, TW04, VN21, VN42, VN59, YJ1, YJ8, YJ9. For a previously published 4.8ky old Neolithic Irish dog (9), the read alignments we obtained from the publication did not include the three unplaced contigs, so we did not estimate copy number for this individual.

For each tested genome, we counted the number of reads mapping to the amylase regions identified above and the number of reads mapping to 75 randomly chosen windows of size 1kb from throughout the whole genome, and calculated the ratio of the former number of reads to the total number of reads. We obtained standard errors for these read count ratios using binomial confidence intervals, following a previous approach to infer biological sex from the ratio of reads mapping to the Y and X chromosomes (150). To convert these ratios to estimates of diploid copy

numbers, we made the assumption that the Eurasian gray wolves (excluding a few outlier individuals) that we tested had copy numbers of two, calculated the mean scaling factor needed to map the read count ratios of these wolves to copy numbers of two, and then multiplied the ratios of all individuals by this factor.

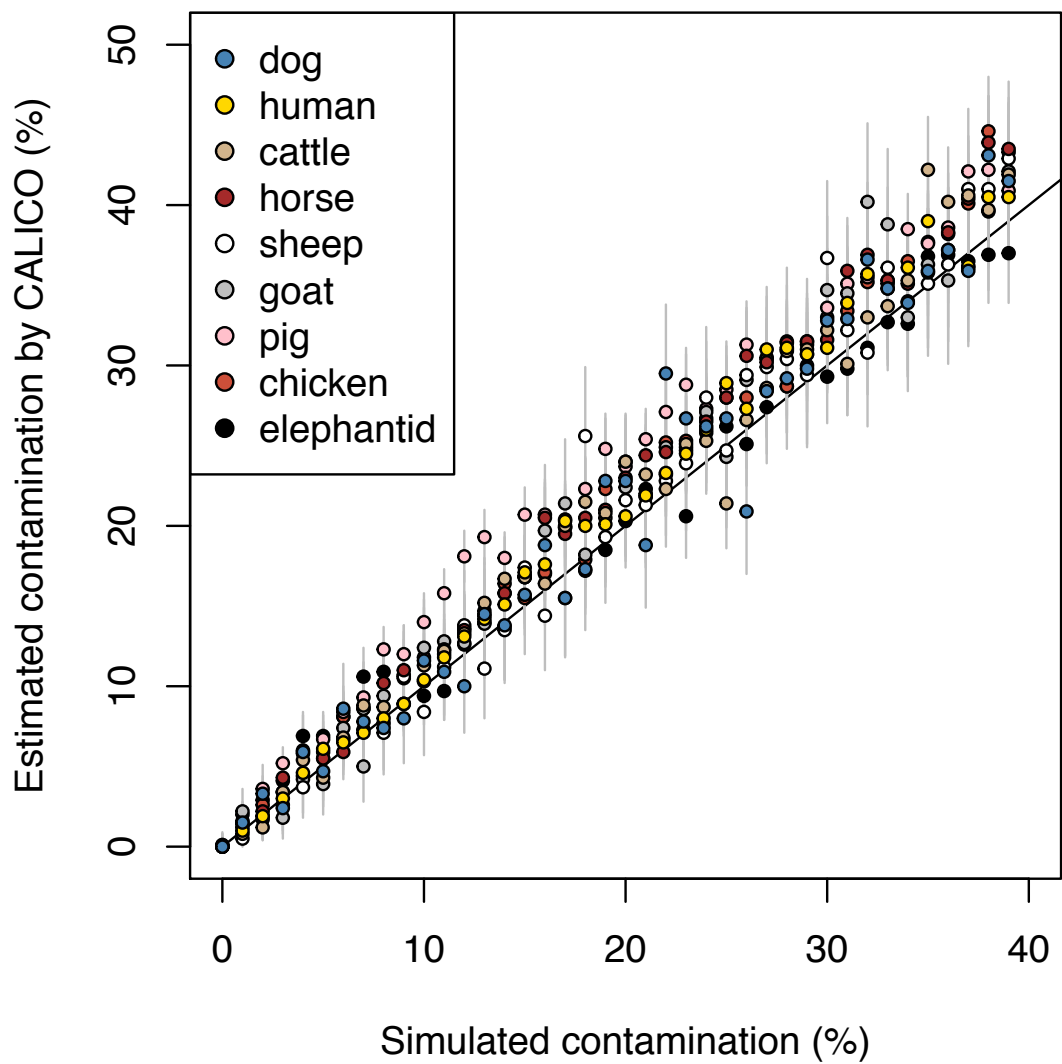


Fig. S1. Validation of mitochondrial DNA contamination in ancient animal genomes with CALICO.

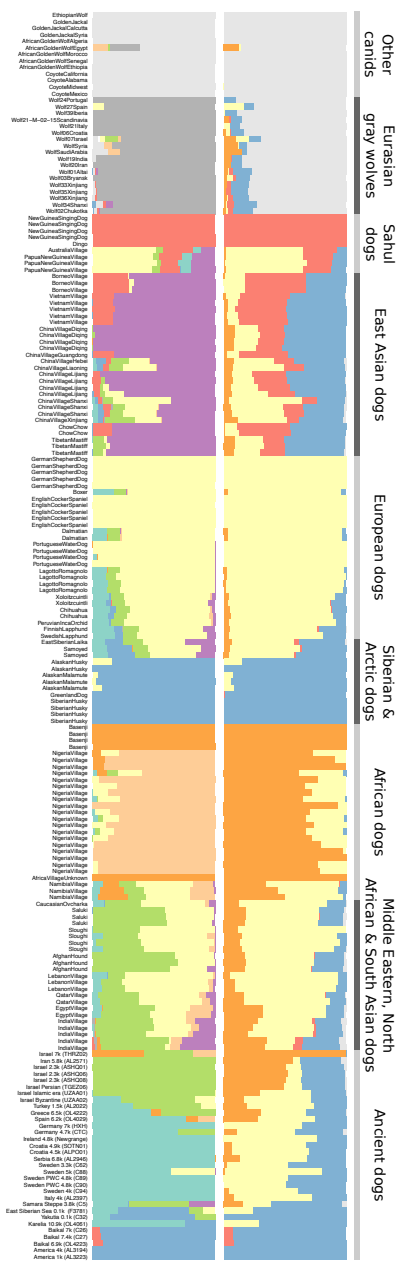


Fig. S2. Model based clustering of ancient and modern genomes, at $k=5$ and $k=9$.

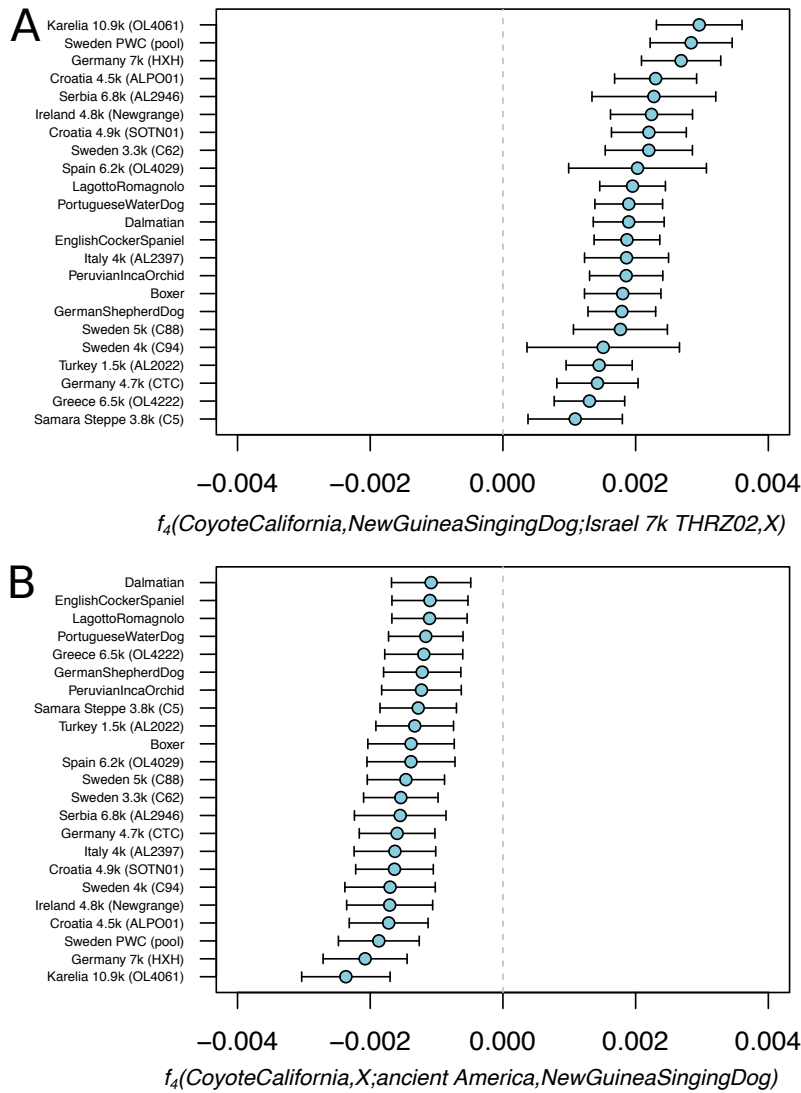


Fig. S3. East Eurasian and American affinities of European dogs. A) All analyzed ancient and modern European dogs have a stronger affinity to East Eurasian dogs, here represented by the New Guinea Singing Dog, than what early Levantine dogs have. Error bars represent ± 3 standard errors. B) All analyzed ancient and modern European dogs have a stronger affinity to ancient American dogs than to the New Guinea Singing Dog.

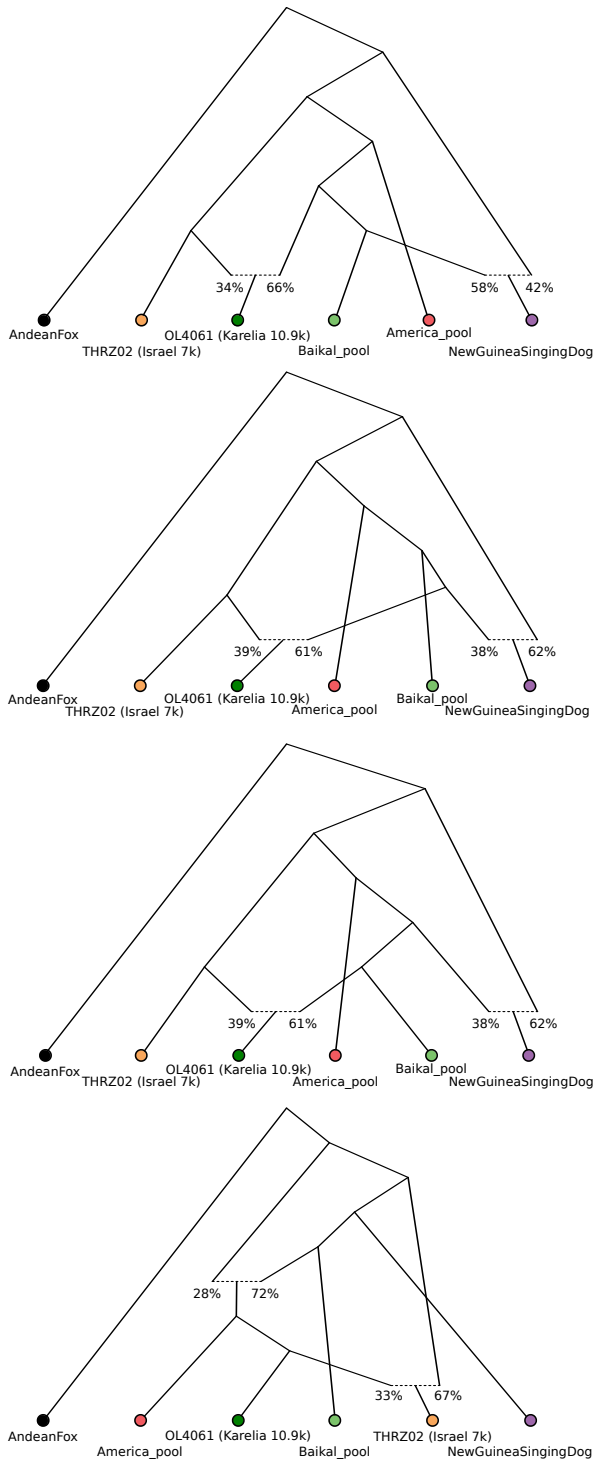


Fig. S4. Admixture graph modelling of dog population relationships. The top four graphs among all 135,285 possible graphs with six leaves and up to two admixture events. The fits between observed (black bars with ± 3 standard errors) and predicted (green or red points for fitting and deviating, respectively) f_4 -statistics are displayed to the right of each graph. The first graph is the only one with no outlier statistics. The second and third have very similar structures to the first one, and each have two minor outlier statistics. The fourth graph has a somewhat different structure, but has five outlier statistics.

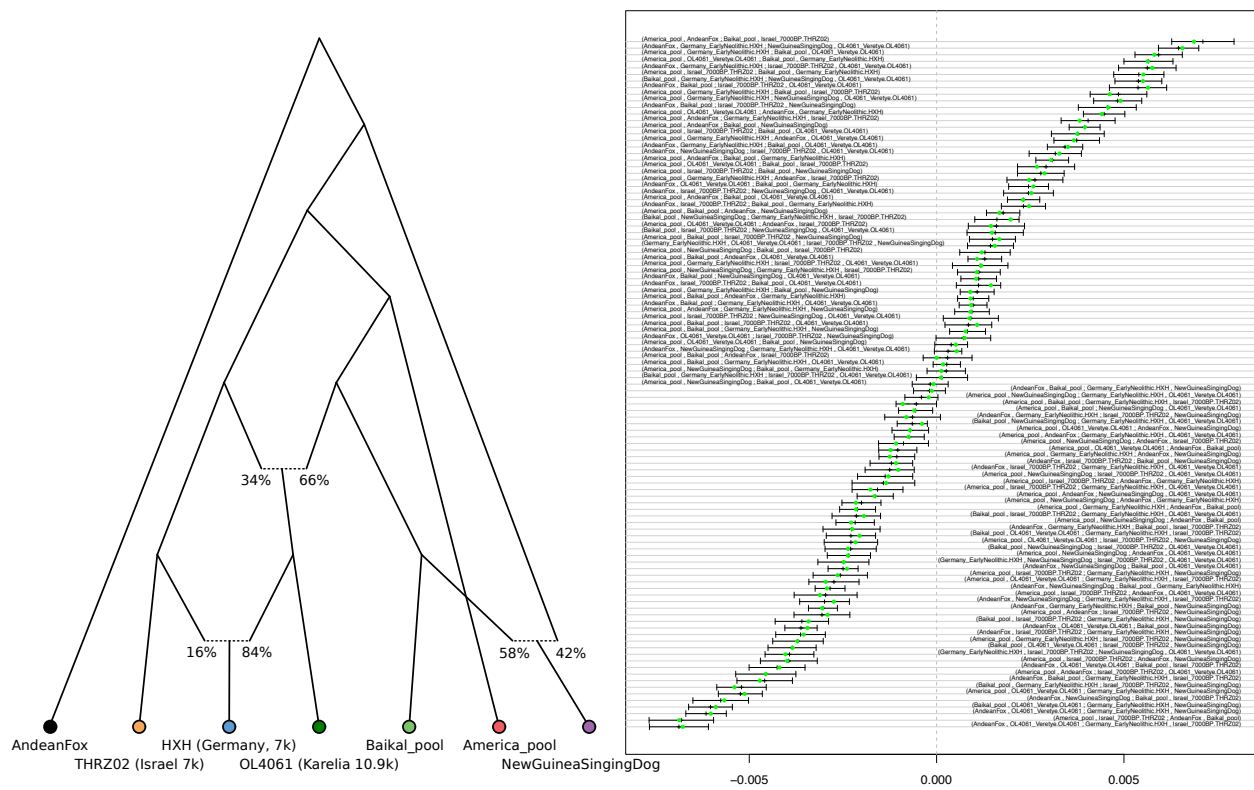


Fig S5. Extending the best fitting admixture graph to include an early European dog. The best fitting graph from the exhaustive testing with six populations (**fig. S4**) was extended by manually grafting on the earliest European dog genome (HXH, Herxheim, Germany, 7k BP) as a mixture of a lineage related to the early Levantine dog and a lineage related to the early Karelian dog. The fit between observed (black bars with ± 3 standard errors) and predicted (green or red points for fitting and deviating, respectively) f_d -statistics are displayed to the right of the graph. No outlier statistics were observed.

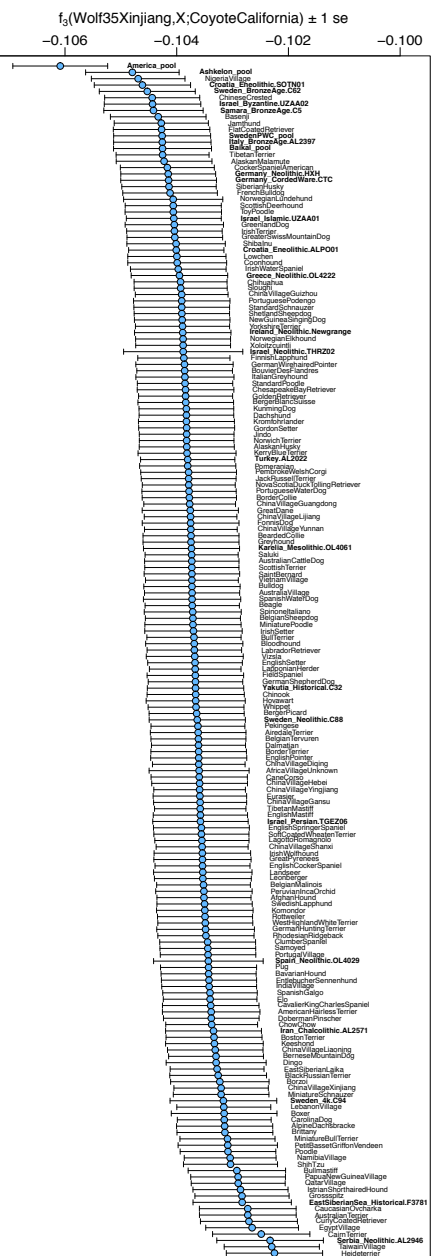


Fig. S6. Testing for basal ancestry across modern and ancient dogs. An outgroup f_3 -statistic of the form $f_3(Wolf35Xinjiang, X; CoyoteCalifornia)$ is used to quantify the amount of genetic drift shared with a wolf from Xinjiang, China. If a dog genome would contain ancestry from any lineage basal to the wolf and dog clade, this would result in a reduction in the value of this statistic. Conversely, if a dog genome would contain wolf ancestry, this would lead to an increase in the value of this statistic. Only sites ascertained as heterozygous in the CoyoteCalifornia individual were used for this analysis, and the fact that all values are negative is likely caused by this ascertainment and/or complex admixture in the Coyote. The ancient American dog population displays a clear reduction relative to other dogs, which is likely caused by an attraction between this population and the coyote (21), however this analysis does not inform on whether it is the American dog that has Coyote admixture or vice versa.

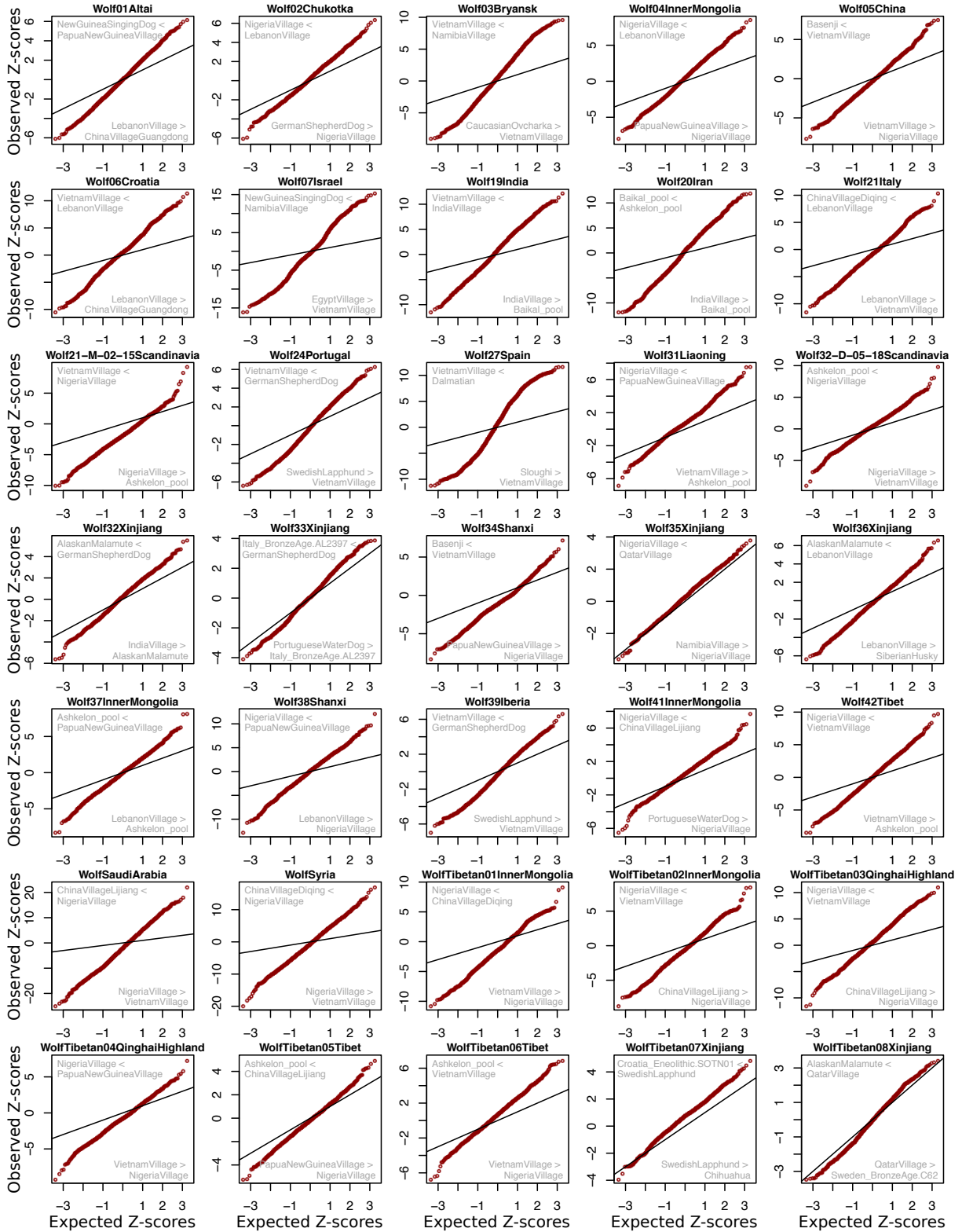


Fig S7. Wolves display widespread asymmetries to pairs of dogs. For each individual wolf, a quantile-quantile plot of the Z-scores from the f_4 -statistics $f_4(\text{CoyoteCalifornia}, \text{Wolf}; \text{Dog}, \text{Dog})$. Grey text displays the pairs of dogs producing the highest Z-scores, with inequality signs indicating which of the two dogs the wolf is closer to.

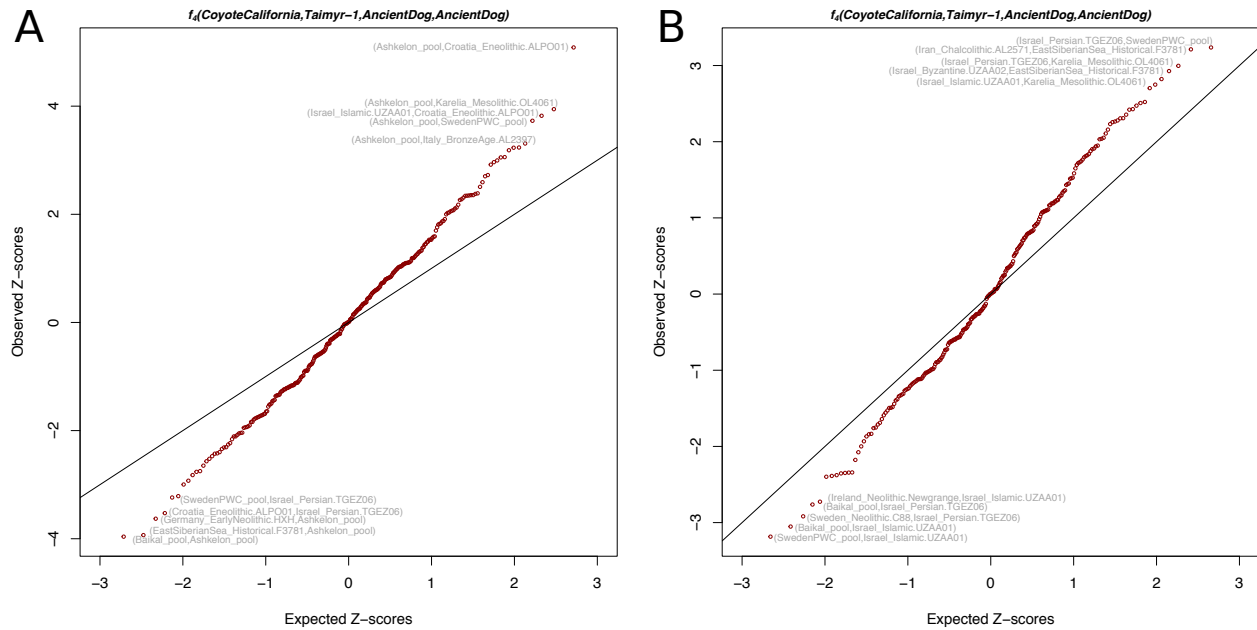


Fig S8. Relationship of the Pleistocene wolf Taimyr-1 to ancient dogs. A) A quantile-quantile plot of the Z-scores from the f_4 -statistics $f_4(\text{CoyoteCalifornia}, \text{Taimyr-1}; \text{Ancient Dog}, \text{Ancient Dog})$, where Taimyr-1 is a 35,000 year old wolf from Siberia (4). The pairs of dogs producing the strongest signals of asymmetry are indicated with grey text. Some deviation from the diagonal expectation of zero is observed, but it can not be ruled out that this is driven by technical noise arising from the direct comparison of ancient genomes. B) Same as in A, but excluding the two ancient dogs that drive most of the off-diagonal signal (Croatia_Eneolithic.ALPO01 and Ashkelon_pool). While there is no strong, statistically significant ($|Z| > 3$) signal, there might be a tendency for the strongest asymmetries to involve the Taimyr-1 wolf being slightly closer to Siberian dogs and northern European Mesolithic dogs than to Near Eastern dogs. This could be consistent with Pleistocene wolf gene flow into Siberian dogs, or with the presence of some basal admixture in Near Eastern dogs, e.g. similar to in some African dogs.

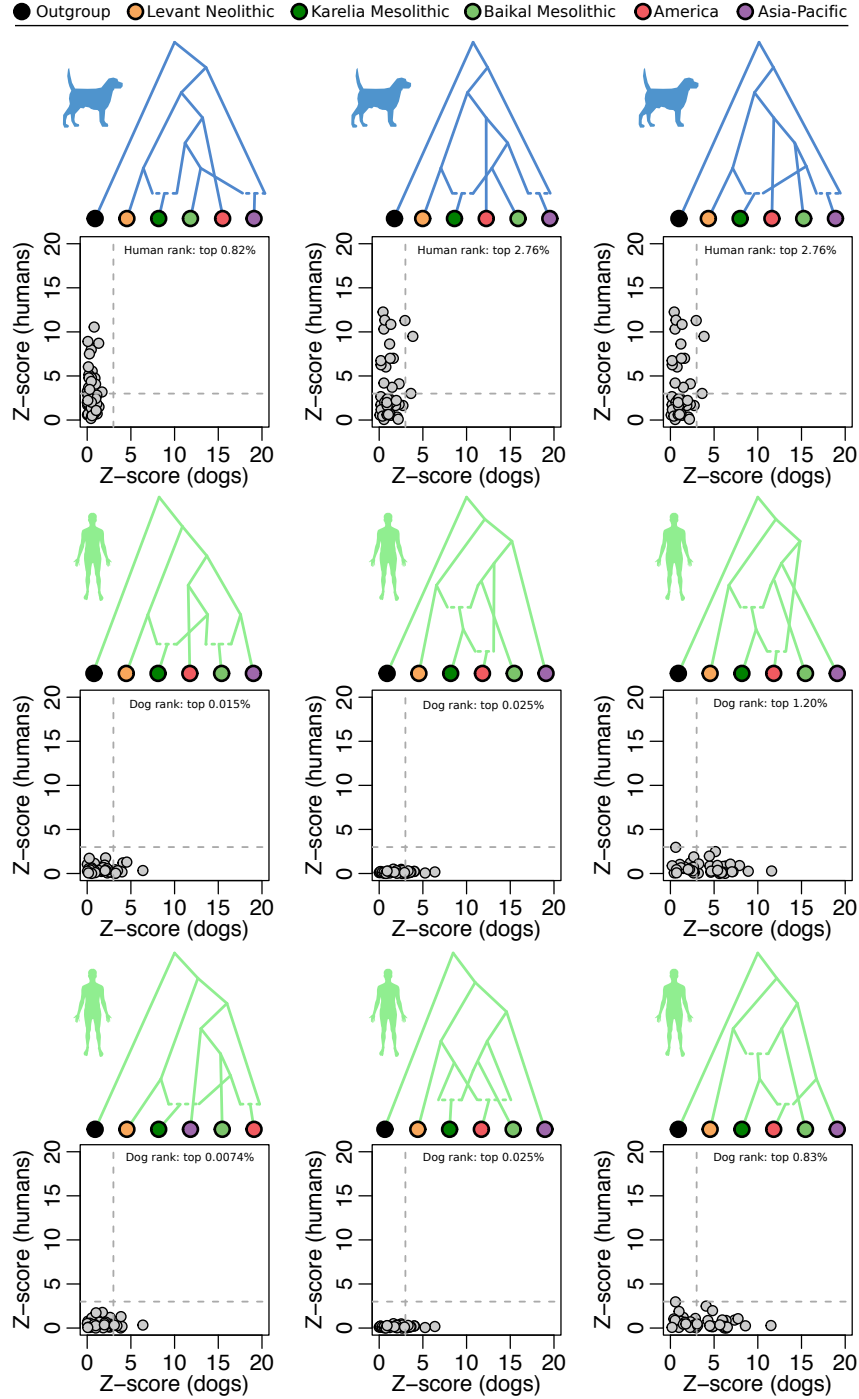


Fig. S9. Cross-species admixture graph fits. Exhaustive testing of all possible admixture graphs with six populations and up to two admixture events was performed independently on matched human and dog datasets. One graph that fit, and two graphs that nearly fit, the dog relationships were identified (**fig. S4**) and are displayed in blue. Six graphs that fit the human relationships were identified and are displayed in green. The fit of these graphs to data for the other species was then assessed. Under each graph the fit in dogs and in humans is displayed in a scatter plot, expressed as absolute Z-scores for the difference between the empirical f_4 -statistics and those predicted by the fitted model. Dashed lines indicate $|Z| = 3$. Text also indicates how well each graph ranks among all possible graphs in the other species.

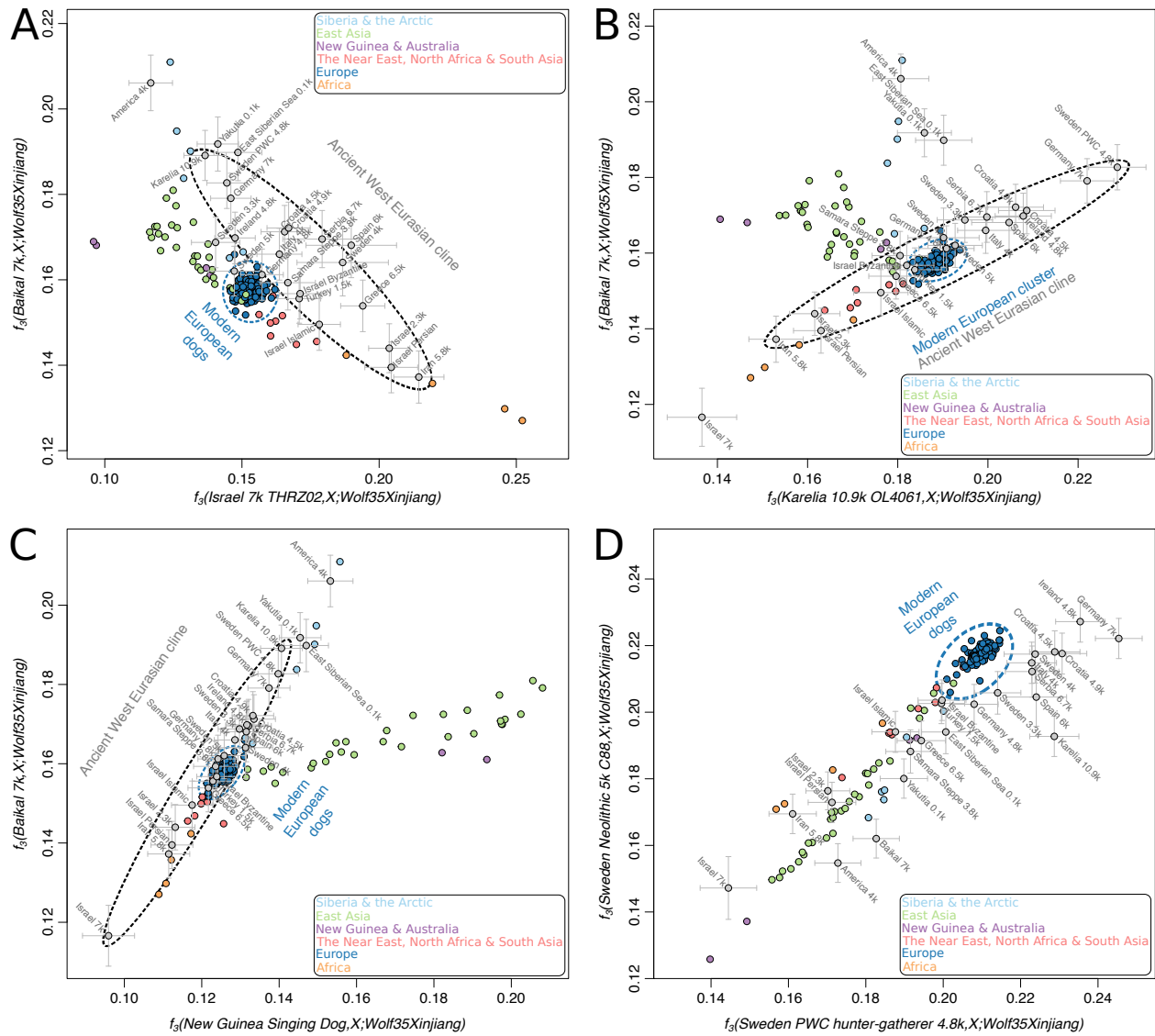


Fig S10. Patterns of shared genetic drift with ancient genomes. Values of outgroup f_3 statistics quantifying shared drift with selected genomes, using a gray wolf individual as the outgroup, are plotted against each other. Modern dog populations are displayed in symbols coloured by geographical region and ancient dogs are displayed in gray symbols. A) A cline of Levant versus Baikal related ancestry exists across ancient west Eurasian dog genomes, but not among modern European dogs. B) Ancestry related to the 10.9 ky old Karelian dog is present to the largest degree in ancient European dog genomes, but not to such high degrees in any present-day dogs. Present-day Siberian dogs instead share more drift with the ancient Lake Baikal dogs. C) The New Guinea Singing Dog and the ancient Lake Baikal dogs define two diverged branches of eastern Eurasian dog ancestry. D) Contrasting contemporaneous Swedish dogs from a Neolithic cultural context and a hunter-gatherer (Pitted Ware Culture) cultural context reveals an increased affinity for Near Eastern and African dogs in the former and Siberian dogs in the latter.

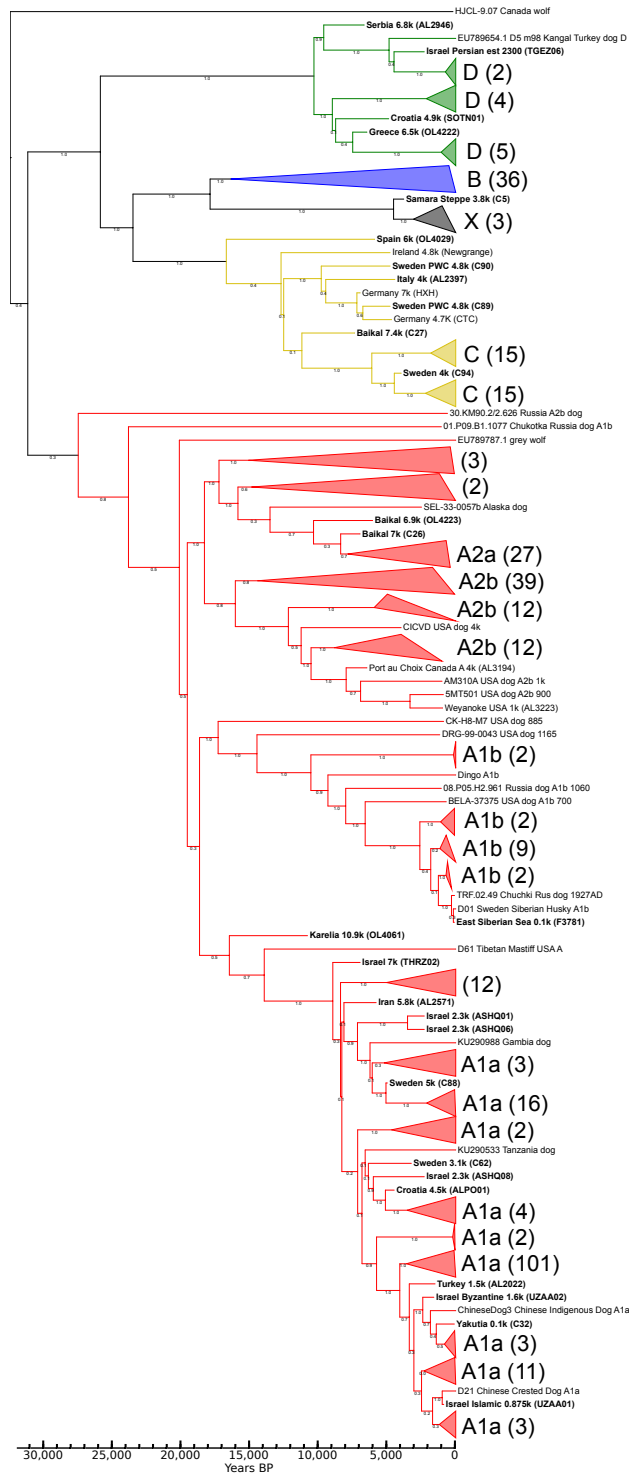


Fig S11. Time-scaled mitochondrial genome phylogeny. Bayesian phylogeny of both novel (bolded) and publically available dog and wolf mitochondrial genomes (n=381). Posterior support values are indicated on the nodes. All monophyletic clades extant to the novel sequences are collapsed, and all dog haplogroup and subclades of haplogroups are marked as such with sample numbers indicated in parentheses. The MRCA of haplogroup C dates to ~17.5 KYA, D to ~10KYA, B to ~2.5KYA, X to ~5KYA, and subclades A1a to at least ~8KYA, A1b to between ~17.5-10KYA, A2b to ~17KYA, and A2a to ~16KYA.

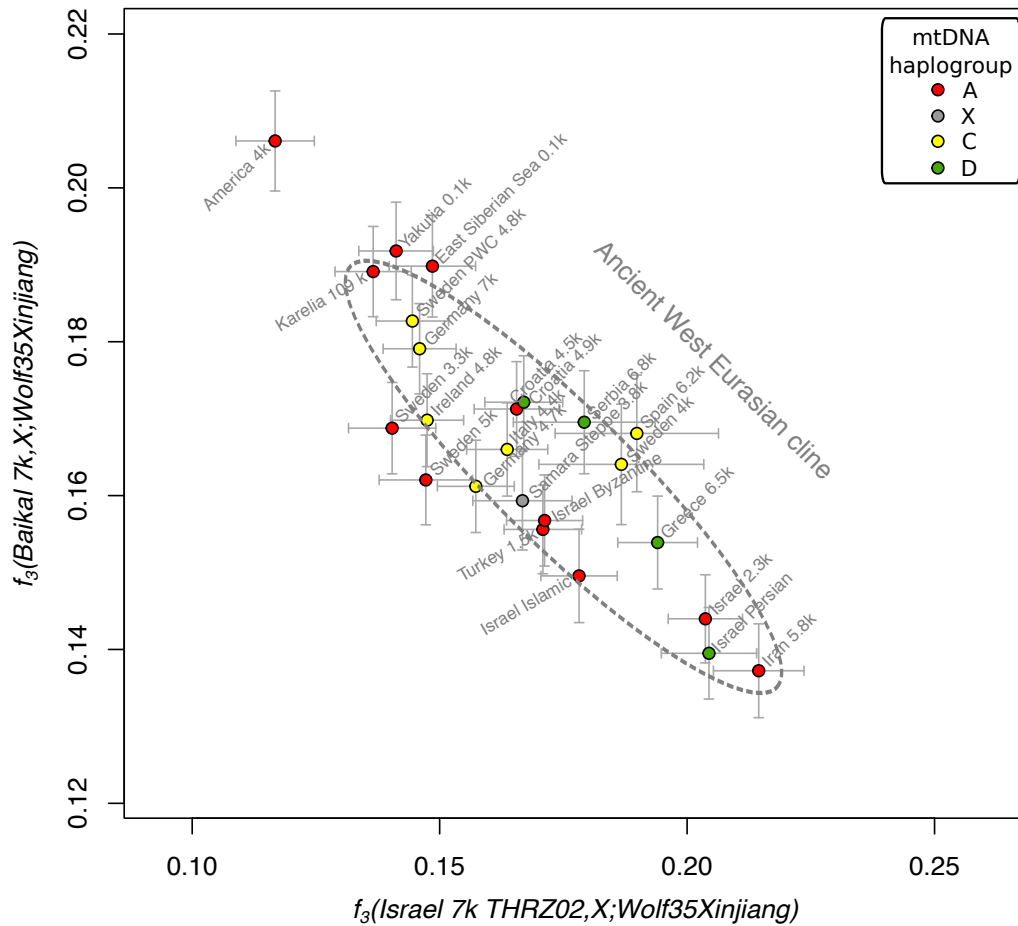


Fig S12. Mitochondrial haplogroups within the context of the ancient West Eurasian ancestry cline. The affinity of ancient dogs to a 7ky old Neolithic Levant individual is plotted against the affinity to 7ky old Lake Baikal dogs, as quantified using outgroup f_3 -statistics on the autosomal genomes. Each dog is then colored according to the mitochondrial haplogroup it carries.



Fig S13. A uniform ancestry of modern European dogs is largely approximated by the ancient dog C88 from Sweden. For the 108 analyzed modern European dogs, the horizontal axis displays the amount of their ancestry that is assigned to the German Shepherd, arbitrarily chosen to represent the modern European dog ancestry, in *qpAdm* analyses. The fact that most of them are assigned 100% German Shepherd ancestry, to the exclusion of all the ancient dogs in the dataset, suggests that most modern European dogs are a clade that does not retain any structure visible among the ancient dogs. The vertical axis shows how much of their ancestry is assigned to the ancient Swedish dog C88 when German Shepherd is kept out of the analysis. The fact that most dogs get 90-100% of their ancestry assigned to the C88 source shows that C88 has an ancestry that is very similar to modern European dog ancestry. A few breeds that are not assigned 100% ancestry with either source, and thus require some contribution from an additional source other than the modern European ancestry, are highlighted. These are Mexican breeds that likely have a small degree of pre-Columbian American ancestry (Chihuahua, Xoloitzcuintli, Chinese Crested), Scandinavian and Spitz dogs that likely have a small degree of Siberian-related ancestry (Swedish Lapphund, Norwegian Elkhound, Jamthund, Grossspitz) and the Rhodesian Ridgeback that likely has a small degree of African (related to the ancient Near Eastern dogs in our dataset) dog ancestry.

Table S1. Ancient dog samples included in this study.

Sample ID	Sample description	Location	Age years ago	Direct date uncal BP	Lat.	Lon.
AL2022	Turkey	Marmara region, Turkey	1565	1642 ± 27 (OxA-36894)	41.01	28.93
AL2397	Italy Early Bronze Age	Belverde di Cetona, Italy	4000	-	42.96	11.90
AL2571	Iran Chalcolithic	Tepe Ghela Gap, Iran	5826	5074 ± 33 (OxA-35324)	36.46	47.07
AL2946	Serbia Neolithic	Pločnik, Serbia	6839	5998 ± 32 (OxA-36896)	43.20	21.36
ALPO01	Croatia Eneolithic	Alymas-Podunavlje, Croatia	4500	-	45.53	18.95
ASHQ01	Israel Persian era	Ashkelon, Israel	2300	-	31.67	34.57
ASHQ06	Israel Persian era	Ashkelon, Israel	2300	-	31.67	34.57
ASHQ08	Israel Persian era	Ashkelon, Israel	2300	-	31.67	34.57
C26	Lake Baikal Mesolithic	Pad' Kalashnikova, Russia	7000	6122 ± 31 (Ox23910)	53.06	103.38
C27	Lake Baikal Mesolithic	Shamanka II, Russia	7400	6430 ± 35 (Ox20561)	51.70	103.71
C32	Yakutia historical	Bulgunnyakhtakh lake, Russia	100	-	70.70	138.21
C5	Steppe Bronze Age	Krasnosamarskoe, Russia	3800	-	52.80	51.10
C62	Sweden Bronze Age	Apalle, Sweden	3100	2945 ± 75 BP (Ua 8826)	59.70	17.60
C88	Sweden Neolithic	Frälsegården, Gökhem, Sweden	5000	-	58.10	13.24
C89	Sweden Pitted Ware	Ajvide, Gotland, Sweden	4800	-	57.17	18.12
C90	Sweden Pitted Ware	Ajvide, Gotland, Sweden	4800	-	57.17	18.12
C94	Sweden	Stora Förvar, Gotland, Sweden	4000	3680 ± 30 (Beta-440527)	57.29	17.97
F3781	Siberia historical	East Siberian Sea coast, Russia	100	50 ± 4 (OxA-34913)	69.62	164.33
OL4029	Spain Neolithic	Marizulo Cave, Gipuzkoa, Spain	6230	5390 ± 34 (OxA-36895)	43.27	-2.09
OL4061	Karelia Mesolithic	Veretye, Lake Lacha, Russia	10930	9575 ± 50 (OxA-36900)	61.26	38.90
OL4222	Greece Neolithic	Skoteini cave, Tharrounia, Euboea Island, Greece	6544	5743 ± 35 (OxA-36899)	38.51	23.98
OL4223	Lake Baikal Mesolithic	Pad' Kalashnikova, Russia	6900	6075 ± 32 (Ox23911)	52.69	103.68
SOTN01	Croatia Eneolithic	Sotin, Croatia	4900	4220 ± 40 (Beta-307577)	45.30	19.10
TGEZ06	Israel Persian era	Tel Gezer, Israel	NA	-	31.86	34.92
THRZ02	Israel Neolithic	Tel Hreiz, Israel	7000	-	32.74	35.05
UZAA01	Israel Islamic era	Uza, Israel	875	-	31.59	34.77
UZAA02	Israel Byzantine era	Uza, Israel	1600	-	31.59	34.77

Table S2. Properties of ancient dog genomes sequenced for this study. UDG refers to whether or not the DNA was treated with uracil-DNA glycosylase to repair deaminated cytosines. Depth refers to the average number of reads covering the autosomal genome. Sex was inferred from the read depth on chromosome X relative to the autosomes.

Sample ID	Sample description	Sampled element	UDG	Depth	Endogenous DNA content	C>T at 1st pos	mtDNA contamination (95% CI)	mtDNA haplogroup	Sex
AL2022	Turkey	Skull	No	2.6	67.87%	38.0%	3.4% (1.4%-5.4%)	A	M
AL2397	Italy Early Bronze Age	Mandible	No	1.5	28.84%	18.4%	5.5% (3.6%-7.4%)	C	M
AL2571	Iran Chalcolithic	Cranium	No	1.1	50.87%	30.6%	4.7% (0.7%-8.7%)	A	M
AL2946	Serbia Neolithic	Mandible	No	0.2	20.50%	28.3%	10.0% (0.0%-23.1%)	D	F
ALPO01	Croatia Eneolithic	Petrous	No	1.6	18.98%	42.8%	2.7% (1.0%-4.4%)	A	M
ASHQ01	Israel Persian era	Cochlea	No	2.4	32.65%	40.3%	1.5% (0.4%-2.6%)	A	F
ASHQ06	Israel Persian era	Cochlea	No	1.8	24.10%	38.4%	1.3% (0.0%-2.6%)	A	F
ASHQ08	Israel Persian era	Petrous	No	2.2	28.45%	44.2%	0.9% (0.0%-1.9%)	A	M
C26	Lake Baikal Mesolithic	Rib	Yes	0.2	2.56%	18.1%	0.0% (0.0%-0.2%)	A	M
C27	Lake Baikal Mesolithic	Vertebra fragment	Yes	0.3	2.17%	18.6%	0.0% (0.0%-1.7%)	C	M
C32	Yakutia historical	Skull fragment	Yes	7.2	37.88%	1.3%	0.0% (0.0%-0.4%)	A	F
C5	Steppe Bronze Age	Tooth	Yes	0.6	7.00%	24.5%	0.3% (0.1%-0.5%)	X	M
C62	Sweden Bronze Age	Tooth	Yes	0.7	10.51%	24.5%	0.0% (0.0%-3.6%)	A	F
C88	Sweden Neolithic	Tooth	Yes	0.7	6.69%	16.0%	(0 informative sites)	A	M
C89	Sweden Pitted Ware	Tooth	Yes	2.2	22.51%	26.8%	0.1% (0.0%-0.2%)	C	M
C90	Sweden Pitted Ware	Tooth	Yes	0.6	7.16%	30.3%	0.1% (0.0%-0.2%)	C	M
C94	Sweden	Humerus	Yes	0.2	1.74%	19.4%	(0 informative sites)	C	M
F3781	Siberia historical	Tooth	Yes	0.8	6.62%	5.3%	0.0% (0.0%-0.1%)	A	M
OL4029	Spain Neolithic	Left femur	No	0.1	32.13%	22.6%	0.0% (0.0%-22.1%)	C	M
OL4061	Karelia Mesolithic	Tooth	No	1.8	59.79%	11.6%	2.8% (1.6%-4.0%)	A	M
OL4222	Greece Neolithic	Mandible	No	4.5	49.87%	38.6%	(0 informative sites)	D	M
OL4223	Lake Baikal Mesolithic	Rib	No	2.2	54.52%	21.9%	2.0% (1.5%-2.5%)	A	F
SOTN01	Croatia Eneolithic	Petrous	No	11.2	55.44%	36.8%	1.3% (0.8%-1.8%)	D	M
TGEZ06	Israel Persian era	Petrous	No	0.9	10.58%	44.9%	1.0% (0.0%-2.0%)	D	M
THRZ02	Israel Neolithic	Petrous	Yes	0.1	1.44%	51.9%	0.0% (0.0%-12.8%)	A	F
UZAA01	Israel Islamic era	Petrous	No	5.8	55.73%	33.6%	0.0% (0.0%-1.5%)	A	M
UZAA02	Israel Byzantine era	Petrous	No	5.5	55.02%	33.7%	1.4% (0.3%-2.5%)	A	F

Table S3. Previously published ancient dog genomes used in analyses. Columns marked with an asterisk (*) denote properties that were calculated or inferred in the present study, in the same way as in table S2, rather than obtained from the original studies.

Sample ID	Sample description	Location	Publication	Date BP	Depth*	mtDNA haplogroup*	Sex*
Newgrange	Ireland Neolithic	Newgrange, Ireland	Frantz et al.	4800	30.7	C	M
HXH	Germany Neolithic	Herxheim, Germany	Botigué et al.	7000	9.4	C	M
CTC	Germany Corded Ware	Cherry Tree Cave	Botigué et al.	4700	9.2	C	M
AL3194	American Port au Choix	Port au Choix, Newfoundland, Canada	Leathlobhair et al.	4000	1.9	A	M
AL3223	American Weyanoke	Weyanoke Old Town, Virginia, USA	Leathlobhair et al.	1000	0.4	A	F

Table S4. Human pairs. The pairings made between ancient dog genomes and published ancient human genomes matched as closely as possible in terms of space, time and cultural context. A few pairs involving modern populations were also used.

Joint label	Dog sample	Human sample
Baikal Mesolithic	Baikal 7000BP	Baikal Early Neolithic 6500BP (<i>120</i>)
Levant Neolithic	Levant Neolithic 7000BP	Levant Neolithic 8800BP (<i>41</i>)
Levant Bronze Age	Levant 2300BP	Caaninite Bronze Age 3700BP (<i>121</i>)
America	America 4000BP	Southwestern Ontario 4200BP (<i>122</i>)
Karelia Mesolithic	Karelia Mesolithic 10900BP	Karelia Mesolithic 7200BP (<i>123</i>)
Iran Chalcolithic	Iran Chalcolithic 5800BP	Iran Seh Gabi Chalcolithic 5500BP (<i>41</i>)
Srubnaya Bronze Age	Samara Bronze Age 3800BP	Srubnaya Bronze Age 3850-3200 BP (<i>123</i>)
Germany Neolithic	Germany Early Neolithic 7000BP	Stuttgart Neolithic 7000BP (<i>43</i>)
Germany Corded Ware	Germany Corded Ware 4700BP	Germany Corded Ware 4000 BP (<i>46</i>)
Sweden Neolithic	Sweden Neolithic 5000BP	Gökhem Neolithic 4900BP (<i>54</i>)
Sweden Pitted Ware Culture	Sweden Pitted Ware Culture 4800BP	Ajvide Pitted Ware Culture 4750BP (<i>54</i>)
Spain Neolithic	Spain Neolithic 6200BP	La Mina Neolithic 5700BP (<i>123</i>)
Croatia Eneolithic	Croatia Eneolithic 4900BP	Starcevo Eneolithic 5400 BP (<i>124</i>)
Greece Neolithic	Greece Neolithic 6500BP	Paliambela 6400 BP and Kleitos 6100BP (<i>125</i>)
Ireland Neolithic	Ireland Neolithic 4800BP	Ballynahatty Neolithic 5200BP (<i>44</i>)
Serbia Neolithic	Serbia Neolithic 6800BP	Serbia Neolithic 6400-7400BP (<i>124</i>)
Italy Bronze Age	Italy Bronze Age 4000BP	Northern Italy Bell Beaker 4000BP (<i>126</i>)
Asia-Pacific Modern	NewGuineaSingingDog	Ami (<i>127</i>)
Vietnam Modern	VietnamVillage	Kinh in Ho Chi Minh City, Vietnam (<i>127</i>)
China South Modern	ChinaVillage (Yunnan)	Chinese Dai in Xishuangbanna, China (<i>127</i>)
Europe West Modern	GermanShepherdDog	French (<i>127</i>)
Levant Modern	LebanonVillage	Palestinian (<i>127</i>)
India Modern	IndianVillage	Kapu (<i>127</i>)
Siberia East Modern	SiberianHusky	Chukchi (<i>127</i>)
Outgroup	CoyoteCalifornia	Mbuti (<i>127</i>)

Table S5. f -statistics testing for affinities between dog populations and other canids. Hand-picked f -statistics testing for affinities between particular dog populations and other canids.

Statistic	Value	Z-score
<i>Affinity between African Golden Wolves and African Dogs:</i>		
$f_4(\text{CoyoteCalifornia, AfricanGoldenWolfAlgeria; Basenji, NewGuineaSingingDog})$	-0.0006	-8.175
$f_4(\text{CoyoteCalifornia, AfricanGoldenWolfEgypt; Basenji, NewGuineaSingingDog})$	-0.0022	-20.321
$f_4(\text{CoyoteCalifornia, AfricanGoldenWolfEthiopia; Basenji, NewGuineaSingingDog})$	-0.0008	-7.764
$f_4(\text{CoyoteCalifornia, AfricanGoldenWolfMorocco; Basenji, NewGuineaSingingDog})$	-0.0006	-8.126
$f_4(\text{CoyoteCalifornia, AfricanGoldenWolfSenegal; Basenji, NewGuineaSingingDog})$	-0.0005	-7.407
$f_3(\text{Basenji, } X; \text{NigeriaVillage}), X = [\text{Golden Jackal, Dhole, North American Wolves, Coyotes, Ethiopian Wolf, African Golden Wolf}]$	< -0.015	< -9
$f_3(\text{X, AfricanGoldenWolfAlgeria; AfricanGoldenWolfMorocco}), X = \text{Worldwide dogs}$	< -0.0075	< -3
<i>No genome-wide evidence for Tibetan wolf admixture in Tibetan dogs</i>		
$f_4(\text{CoyoteCalifornia, Wolf42Tibet; TibetanMastiff, ChinaVillageDiqing})$	-0.000015	-0.573
$f_4(\text{CoyoteCalifornia, Wolf42Tibet; TibetanTerrier, ChinaVillageDiqing})$	0.000168	3.144
$f_4(\text{CoyoteCalifornia, Wolf42Tibet; ShihTzu, ChinaVillageDiqing})$	0.000023	0.324
<i>Affinity between Coyotes and American dogs</i>		
$f_4(\text{AndeanFox, CoyoteCalifornia; America_pool, Baikal_pool})$	-0.000281	-3.890
$f_4(\text{AndeanFox, CoyoteCalifornia; America_pool, Karelia_Mesolithic.O14061})$	-0.000461	-5.673
$f_4(\text{AndeanFox, CoyoteCalifornia; America_pool, NewGuineaSingingDog})$	-0.000114	-1.465
$f_4(\text{AndeanFox, CoyoteCalifornia; America_pool, Germany_Neolithic.HXH})$	-0.000452	-6.085

Table S6. Modern dog populations displaying evidence of admixture in the form of negative f_3 -statistics. Modern dogs were tested as targets with all possible combinations of all ancient dogs and a globally representative selection of 15 modern dog populations as sources. All targets with at least one significantly ($Z < -3$) negative statistic are included, and for each of these, the three combinations of sources producing the smallest f_3 values are displayed.

Target	Source1	Source2	f_3	Z
AustraliaVillage	Dingo	GermanShepherdDog	-0.08529	-11.212
AustraliaVillage	NewGuineaSingingDog	GermanShepherdDog	-0.082696	-11.636
AustraliaVillage	Dingo	PortugueseWaterDog	-0.077833	-10.414
CarolinaDog	NewGuineaSingingDog	GermanShepherdDog	-0.063129	-21.139
CarolinaDog	NewGuineaSingingDog	EnglishCockerSpaniel	-0.061362	-20.941
CarolinaDog	NewGuineaSingingDog	PortugueseWaterDog	-0.060886	-21.204
ChinaVillageHebei	Sweden_Neolithic.C88	NewGuineaSingingDog	-0.027449	-8.456
ChinaVillageHebei	NewGuineaSingingDog	GermanShepherdDog	-0.027304	-7.975
ChinaVillageHebei	Turkey.AL2022	NewGuineaSingingDog	-0.026413	-7.946
ChinaVillageLiaoning	NewGuineaSingingDog	GermanShepherdDog	-0.037463	-14.516
ChinaVillageLiaoning	Sweden_Neolithic.C88	NewGuineaSingingDog	-0.032687	-11.619
ChinaVillageLiaoning	NewGuineaSingingDog	EnglishCockerSpaniel	-0.030601	-12.471
ChinaVillageLijiang	NewGuineaSingingDog	GermanShepherdDog	-0.040672	-36.016
ChinaVillageLijiang	Sweden_Neolithic.C88	NewGuineaSingingDog	-0.035942	-30.126
ChinaVillageLijiang	NewGuineaSingingDog	Dalmatian	-0.034538	-28.137
ChinaVillageShanxi	NewGuineaSingingDog	GermanShepherdDog	-0.057129	-41.713
ChinaVillageShanxi	Dingo	GermanShepherdDog	-0.050047	-29.072
ChinaVillageShanxi	GermanShepherdDog	ChinaVillageGuizhou	-0.045035	-34.279
ChinaVillage	NewGuineaSingingDog	GermanShepherdDog	-0.041189	-33.264
ChinaVillage	Sweden_Neolithic.C88	NewGuineaSingingDog	-0.039581	-30.338
ChinaVillage	Turkey.AL2022	NewGuineaSingingDog	-0.03693	-27.906
ChinaVillageYingjiang	NewGuineaSingingDog	GermanShepherdDog	-0.049556	-39.52
ChinaVillageYingjiang	Sweden_Neolithic.C88	NewGuineaSingingDog	-0.045691	-38.101
ChinaVillageYingjiang	Turkey.AL2022	NewGuineaSingingDog	-0.045213	-36.516
ChinaVillageYunnan	Turkey.AL2022	NewGuineaSingingDog	-0.033005	-17.607
ChinaVillageYunnan	Sweden_Neolithic.C88	NewGuineaSingingDog	-0.032295	-17.018
ChinaVillageYunnan	Samara_BronzeAge.C5	NewGuineaSingingDog	-0.032075	-16.092
Jindo	NewGuineaSingingDog	GermanShepherdDog	-0.0219	-4.378
Jindo	NewGuineaSingingDog	EnglishCockerSpaniel	-0.020675	-4.131
Jindo	Sweden_Neolithic.C88	NewGuineaSingingDog	-0.02014	-3.982
LebanonVillage	Israel_Neolithic.THRZ02	GermanShepherdDog	-0.021238	-8.824
LebanonVillage	Iran_Chalcolithic.AL2571	GermanShepherdDog	-0.018645	-9.252
LebanonVillage	Ashkelon_pool	GermanShepherdDog	-0.013212	-6.99
NamibiaVillage	Germany_Neolithic.HXH	Basenji	-0.033793	-14.771
NamibiaVillage	Sweden_Neolithic.C88	Basenji	-0.031593	-14.265
NamibiaVillage	Germany_Neolithic.HXH	Basenji	-0.031512	-12.841
NigeriaVillage	Sweden_Neolithic.C88	Basenji	-0.03069	-15.84
NigeriaVillage	SwedenPWC_pool	Basenji	-0.029342	-14.672
NigeriaVillage	Ireland_Neolithic.Newgrange	Basenji	-0.028738	-14.029
PapuaNewGuineaVillage	NewGuineaSingingDog	GermanShepherdDog	-0.061108	-23.566
PapuaNewGuineaVillage	NewGuineaSingingDog	EnglishCockerSpaniel	-0.057408	-22.523

PapuaNewGuineaVillage	NewGuineaSingingDog	Dalmatian	-0.05604	-21.105
TaiwainVillage	NewGuineaSingingDog	GermanShepherdDog	-0.066158	-28.074
TaiwainVillage	NewGuineaSingingDog	Dalmatian	-0.060695	-26.85
TaiwainVillage	Sweden_Neolithic.C88	NewGuineaSingingDog	-0.060664	-27.926
TibetanMastiff	Turkey.AL2022	NewGuineaSingingDog	-0.004472	-3.212
TibetanMastiff	Iran_Chalcolithic.AL2571	NewGuineaSingingDog	-0.004222	-2.613
TibetanMastiff	Sweden_Neolithic.C88	NewGuineaSingingDog	-0.003051	-2.079
Xoloitzcuintli	Dingo	GermanShepherdDog	-0.013908	-6.326
Xoloitzcuintli	GermanShepherdDog	ChinaVillageGuizhou	-0.009384	-4.852
Xoloitzcuintli	America_pool	GermanShepherdDog	-0.00843	-3.976

Data S1. qpAm results (additional Excel file). Sheet 1: For each target, the top five models obtained with the “comprehensive” set of possible sources and outgroups, and the top five models obtained with the “core” set of possible sources and outgroups, are displayed. Models are ranked by their p-values, but prioritizing models with fewer sources. Sheet 2: Results for the targeted analysis aiming to understand the sources of ancestry for Bronze Age and later dogs in Europe.