Ancient Wolf Genome Reveals an Early Divergence of Domestic Dog Ancestors and Admixture into High-Latitude Breeds

Graphical Abstract

Highlights
- An ancient Siberian wolf yields a first draft genome sequence of a Pleistocene carnivore
- The 35,000-year-old wolf genome allowed recalibration of the lupine mutation rate
- Dog ancestors diverged from modern wolf ancestors at least 27,000 years ago
- Ancient Siberian wolves contributed to the ancestry of high-latitude dog breeds

Authors
Pontus Skoglund, Erik Ersmark, Eleftheria Palkopoulou, Love Dalén

Correspondence
skoglund@genetics.med.harvard.edu (P.S.), love.dalen@nrm.se (L.D.)

In Brief
Skoglund et al. recovered the genome sequence of a 35,000-year-old Siberian wolf. Calibration of the molecular clock suggests that the ancestors of modern dogs formed a distinct lineage prior to the peak of the last ice age. Siberian huskies and other northern dog breeds trace a part of their ancestry to the ancient Siberian wolf population.

Accession Numbers
PRJEB7788
Ancient Wolf Genome Reveals an Early Divergence of Domestic Dog Ancestors and Admixture into High-Latitude Breeds

Pontus Skoglund,1,2,*, Erik Ersmark,4,6 Eleftheria Palkopoulou,4,6 and Love Dalén4,*

1Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
2Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA
3Archaeological Research Laboratory, Department of Archaeology and Classical Studies, Stockholm University, 10691 Stockholm, Sweden
4Department of Bioinformatics and Genetics, Swedish Museum of Natural History, 10405 Stockholm, Sweden
5Department of Zoology, Stockholm University, 10691 Stockholm, Sweden
*Correspondence: skoglund@genetics.med.harvard.edu (P.S.), love.dalen@nrm.se (L.D.)
http://dx.doi.org/10.1016/j.cub.2015.04.019

SUMMARY

The origin of domestic dogs is poorly understood [1–15], with suggested evidence of dog-like features in fossils that predate the Last Glacial Maximum [6, 9, 10, 14, 16] conflicting with genetic estimates of a more recent divergence between dogs and worldwide wolf populations [13, 15, 17–19]. Here, we present a draft genome sequence from a 35,000-year-old wolf from the Taimyr Peninsula in northern Siberia. We find that this individual belonged to a population that diverged from the common ancestor of present-day wolves and dogs very close in time to the appearance of the domestic dog lineage. We use the directly dated ancient wolf genome to recalibrate the molecular timescale of wolves and dogs and find that the mutation rate is substantially slower than assumed by most previous studies, suggesting that the ancestors of dogs were separated from present-day wolves before the Last Glacial Maximum. We also find evidence of introgression from the archaic Taimyr wolf lineage into present-day wolf breeds from northeast Siberia and Greenland, contributing between 1.4% and 27.3% of their ancestry. This demonstrates that the ancestry of present-day dogs is derived from multiple regional wolf populations.

RESULTS AND DISCUSSION

The closest living relative of domestic dogs is the gray wolf, Canis lupus [1], but the number of domestication events, as well as their antiquity and geographical origin, is highly contentious [1–15]. While molecular estimates of the time of origin of the dog lineage are contingent on principally unknown mutation rates and generation times, the most recent genomic estimates of the divergence between wolves and dogs date to 11,000 to 16,000 years ago [13, 15, 17–19]. These estimates are in considerable discord with reported archaeological evidence of dog-like canids from before the Last Glacial Maximum, which date as far back as 36,000 years before present (BP) [6, 9, 10, 14, 16]. Furthermore, a recent study showed that gray wolves from as disparate locations as China, Israel, and Croatia were symmetrically related to modern-day dogs [15]. This observation suggests that dogs were domesticated prior to the diversification of present-day gray wolf populations or that the wild ancestors of dogs are now extinct. The latter scenario would be consistent with an earlier finding of a morphologically distinct wolf population adapted to megafaunal prey in Late Pleistocene Beringia [20], as well as mitochondrial DNA evidence for a Holocene replacement of European gray wolves [21]. One hypothesis could thus be that the wild ancestors of dogs were a genetically distinct wolf population that inhabited the Late Pleistocene steppe-tundra biome and that this population was subsequently replaced [19], possibly by a northward postglacial expansion of smaller-bodied wolves that gave rise to modern-day wolf diversity. To test this hypothesis, we sequenced a draft genome of a Late Pleistocene wolf from northern Siberia.

Shotgun sequencing was performed on a canid rib originally collected during an expedition to the Russian Taimyr Peninsula in 2010, from here on referred to as Taimyr 1. This rib was identified as coming from a wolf through sequencing of a short region of the mitochondrial 16S rRNA gene and was directly dated using accelerator mass spectrometry (AMS) radiocarbon dating to 30,920 ± 380 14C years BP, equivalent to ∼34,900 calendar years BP (35,000 years ago) (for details, see the Supplemental Experimental Procedures). Genomic libraries were constructed with and without treatment with the uracil-specific excision reagent (USER) enzyme mix [22]. This enzymatic treatment was employed in order to remove postmortem-derived uracil residues that introduce errors into ancient genome datasets. We sequenced these libraries on the Illumina HiSeq platform to a total average 1-fold sequencing depth (Figure S1), the majority of which is derived from the USER-treated libraries. Using the retained postmortem damage patterns at methylated CpG sites, we observed nucleotide misincorporation patterns in these sequences expected from DNA tens of thousands of years old (Figure S1) [23]. We assembled the mitochondrial genome to an average 182-fold sequencing depth and reconstructed a phylogeny relating the Taimyr 1 individual’s mitochondrial DNA lineage to those of modern-day dogs and wolves, as well as previously...
Published mitochondrial DNA sequences from ancient canids [14]. The mtDNA of Taimyr 1 forms a distinct lineage separated from the dog lineages (Figure 1), in agreement with the generally basal grouping of all ancient wolves that has been observed in previous studies [14, 20]. The sequencing depth of chromosome X is 50.4% of that expected of an autosome of similar size, demonstrating that the Taimyr individual was male.

A previous study used seven present-day canid genomes to show that all studied wolves share a common origin that excludes the ancestors of domestic dogs [15]. The divergence between the wolf and dog lineages was estimated to date back to 11,000–16,000 years ago, but this estimate relied on assumptions on the mutation rate, which has not been directly estimated for domestic dogs. A prediction of this model is that the ~35,000-year-old Taimyr individual would have lived long before the divergence of dogs and gray wolves, and thus would be symmetrically related to both populations. In a principal component analysis (Figure S2), the Taimyr individual is approximately intermediate to gray wolves and dogs, but this could be expected just from the extra genetic drift that has occurred in those lineages since the death of the Taimyr individual. To formally test the hypothesis that Taimyr 1 lived prior to the divergence between dogs and gray wolves, we used D statistics [24, 25], expecting symmetry (D = 0) under the null hypothesis corresponding to the population history (Andean fox, (Taimyr 1, (dog, wolf))). This test, as well as all other population genetic analyses reported here, used the USER-treated portion of the sequence data in order to minimize the effect of postmortem degradation. It should be noted that we also found consistent results for an even more conservative set of analyses restricted to transversion polymorphisms. Overall, we find that the data are consistent with the Taimyr wolf being symmetrically related to wolves and dogs for all pairs of present-day canids ([Z] < 3). Conversely, tests with the Taimyr individual being assigned to a clade with one of the present-day canids to the exclusion of another, corresponding to a history (Andean fox, (canid1, (Taimyr 1, canid2))), were all rejected ([Z] > 5).

To investigate the phylogenetic position of the Taimyr wolf further, we fitted a population model to the genomes of the Taimyr wolf and modern canids. We find that gene flow between present-day dogs and wolves after their initial divergence is required to explain the data (Supplemental Experimental Procedures). Once this gene flow is included in the model, the Taimyr wolf can be successfully fitted as belonging to the wolf lineage, the dog lineage, or the lineage ancestral to both wolves and dogs (Figure 2; Supplemental Experimental Procedures). In these three models, the Taimyr wolf is consistently placed very close to the split between the dog and wolf lineages, as measured in units of genetic drift (an $F_{ST}$ of 0 to 0.007). Thus, our data are consistent with a trifurcation of the dog, wolf, and Taimyr lineages, indicating that they all diverged at about the same time. This appears to be inconsistent with the hypothesis that dogs and wolves diverged only 11,000–16,000 years ago, since under this model it might be expected that the Taimyr wolf would be confidently placed on the lineage ancestral to the dog-wolf split, due to the substantial amount of genetic drift that most likely would have occurred since the death of the Taimyr wolf (35,000 years ago) until the split between dogs and wolves some 20,000 years later.

To estimate the divergence time between the Taimyr individual and present-day canids directly, and accounting for changes in effective population size, we estimated the probability $P(\text{derived} | \text{heterozygous})$ [25] of an individual A (such as the Taimyr individual) sharing a derived allele discovered as a heterozygote in a diploid present-day individual B. At sites where the Chinese wolf is heterozygous, the Taimyr individual carries the derived allele at 30.8% ± 0.1% of the sites, compared to ~32% for the Croatian wolf, dingo, and boxer (Figure 3; Table S2). We used this as a summary statistic to estimate the divergence time of the Taimyr lineage given a model of population history of the Chinese wolf inferred using the pairwise sequential Markovian coalescent method [26]. We find that calibration using the most
commonly assumed mutation rate of $1 \times 10^{-8}$ per generation and a 3-year gray wolf generation time [5, 15, 18, 19, 27] would imply that the Taimyr wolf diverged from the Chinese wolf 10,000-14,000 years ago (Figure 3), which is incompatible with its calibrated direct radiocarbon date of ~35,000 years BP. Instead, the mutation rate must be substantially slower in order to be compatible with the age of the Taimyr individual, and we find that the Taimyr divergence can be accommodated by a mutation rate of $0.4 \times 10^{-8}$ per generation (Figure 3). However, it should be noted that this assumes that the Taimyr wolf is directly ancestral to the Chinese gray wolf. If there was structure between the ancestors of the Chinese wolf and the Taimyr wolf, the mutation rate would have to be even slower, and as such a rate of $0.4 \times 10^{-8}$ per generation is conservative. We emphasize that this mutation rate is for non-CpG sites, since SNPs in CpG dinucleotide context were excluded from the variants called in the present-day genome. Alternatively, our results could indicate that the generation time is longer than 3 years, or some combination of slower mutation rate and a longer generation time. Regardless, this direct evidence suggests a longer timescale of wolf-dog population history and thus implies that the 11,000–16,000 years ago wolf-dog divergence inferred in a previous study [15] should be recalibrated to ~27,000–40,000 years ago.

To examine shared ancestry between the ancient Taimyr wolf and a larger set of modern-day dog populations, we used data from 48 dog breeds genotyped at ~170,000 SNPs [28] and computed $D$ statistics to assess whether each breed shared more alleles with the Taimyr wolf than a set of 15 modern-day gray wolves (Figure 4; Table S3). We found clear evidence of a closer relationship between the Taimyr wolf and the Siberian Husky ($Z = 4.3$, $p = 0.000009$), Greenland Sledge Dogs ($Z = 3.6$, $p = 0.00016$), and, to a lesser extent, Chinese Shar-Pei and Finnish Spitz ($p < 0.05$) compared to other dog breeds (Supplemental Experimental Procedures). To estimate the proportion of ancestry derived from the Taimyr wolf lineage in the Greenland Sledge Dogs, we fitted an admixture graph using the Andean fox, present-day gray wolves, and German Shepherds. We find that the best-fitting graph posits 3.5% Taimyr-derived ancestry in the Greenland Sledge Dogs but that an ancestry proportion ranging from 1.4% and 27.3% is consistent with the data (Supplemental Experimental Procedures). These results can be explained either by a very early presence of dogs in northern Eurasia or by the genetic legacy of the Taimyr individual being preserved in northern wolf populations until the arrival of dogs at high latitudes.

Extending our population history modeling to SNP array genotypes from a large set of gray wolf populations [11, 29], we further find that the majority of the ancestry of North American wolves also diverged from other wolves later than the Taimyr wolf lineage (Supplemental Experimental Procedures; Figure S4). This suggests that all extant gray wolf populations share a relatively recent origin, most likely sometime after the divergence of the Taimyr wolf lineage but prior to the inundation of the Bering Land Bridge and subsequent isolation of Eurasian and North American wolves.

In conclusion, our results provide direct evidence for a longer timescale for the divergence of the dog and wolf lineages than previously assumed, and thus suggest that dogs may have originated much earlier than commonly accepted. Such an early divergence is consistent with several paleontological reports of dog-like canids up to 36,000 years old [6, 9, 10, 14, 16], as well as the evidence that domesticated dogs most likely accompanied early colonizers into the Americas [30]. However, the initial divergence between the ancestors of dogs and gray wolves would not necessarily have had to coincide with domestication in the sense of selective breeding, since this human-mediated process could have occurred later or over an extended period.
introgression could have provided early dogs in high latitudes with phenotypic variation beneficial for adaptation to a new challenging environment.

**ACCESSION NUMBERS**

Sequence data for the Taimyr 1 genome has been submitted to the European Nucleotide Archive (http://www.ebi.ac.uk/ena) under accession number PRJEB7788.

**SUPPLEMENTAL INFORMATION**

Supplemental Information includes Supplemental Experimental Procedures, four figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2015.04.019.

**AUTHOR CONTRIBUTIONS**

P.S. and L.D. conceived and designed the study. L.D. conducted field work. E.P. and E.E. performed laboratory analyses. E.E. and P.S. performed the mtDNA analyses. P.S. analyzed the genomic data and interpreted the results with input from L.D. P.S. wrote the paper with input from all the other authors.

**ACKNOWLEDGMENTS**

We thank David Reich and Beth Shapiro for constructive comments on the population genetic analysis and a previous version of the manuscript. We also wish to thank John November, Rena Schweizer, Bridgett vonHoldt, Sofie Wagenius, and Matthew Webster for data access and technical assistance and Tom Gilbert, Maria Avila-Arcos, Ludovic Orlando, Mark Lipson, Iosif Laziridis, Iain Mathieson, Jan Storå, and Greger Larson for discussions. P.S. was supported by a scholarship from the Swedish Research Council (VR grant 2014-453). L.D. acknowledges funding from the Swedish Research Council (VR grant 2012-3869), as well as logistical support from the Swedish Polar Research Secretariat that organized the Taimyr 2010 expedition. The authors would like to acknowledge support from the Science for Life Laboratory, National Genomics Infrastructure, NGI, and UPPMAX for providing assistance in massive parallel sequencing and computational infrastructure (UPP-NEX project ID: b2014030). David Reich and Harvard Medical School provided additional computational resources.

Received: December 9, 2014
Revised: March 11, 2015
Accepted: April 10, 2015
Published: May 21, 2015

**Figure 3. Recalibrating the Lupine Mutation Rate using the Directly Dated Taimyr Genome**

We estimated the probability $F(A|B)$ for the Taimyr individual carrying the derived allele at positions where the Chinese wolf is heterozygous using empirical data. This statistic is informative on the coalescent time passed in the ancestry of the Chinese wolf since its divergence from the Taimyr wolf’s lineage. Since the Taimyr wolf must have separated at least 35,000 years ago, the age of the specimen obtained from a direct radiocarbon date, its divergence from the Chinese wolf can be used to calibrate the lupine mutation rate, in the sense that we can infer the maximum mutation rate that is consistent with the proximity of the Taimyr genome to the present-day Chinese wolf. To achieve this, we built calibration curves for $F(A|B)$ given a model of Chinese wolf effective population size history (inferred using pairwise sequentially Markovian coalescent (PSMC) analysis) and different mutation rates. Mutation rates per generation are shown in the legend. A generation time of 3 years is assumed. See also Table S2.

of time. On the other hand, a scenario of a much more recent timing of domestication would require that the majority of present-day dog ancestry originates from an extinct or presently unsampled wolf population. Regardless, we find that the ancestry of present-day dog breeds descends from more than a single domestication event, since high-latitude dog breeds such as the Siberian Husky and Greenland Sledge Dogs can trace part of their ancestry to the now-extinct Taimyr wolf lineage. This

**Figure 4. Introgression from a Population Related to the Ancient Taimyr Wolf into Northeast Siberian and New World Arctic Dog Breeds**

We computed the statistic $D(\text{ancient Taimyr wolf, present-day gray wolves; dog breed } X)$ with input from ~66,000 SNPs. The affinity of the Siberian Husky and Greenland Sledge Dogs to Taimyr is replicated also in statistics of the form $D(\text{Andean fox, Taimyr; present-day gray wolves, dog breed } X)$. See also Figure S4 and Tables S3 and S4.
REFERENCES


Ancient Wolf Genome Reveals an Early Divergence of Domestic Dog Ancestors and Admixture into High-Latitude Breeds

Pontus Skoglund, Erik Ersmark, Eleftheria Palkopoulou, and Love Dalén
Supplemental Data

Figure S1. Features of the Taimyr 1 genome (related to Figure 1). Evidence for postmortem damage for non-treated libraries and libraries treated with the USER enzyme mix, inside and outside of CpG context.
Figure S2. Analysis of 7 present-day canid genomes and the ancient Taimyr wolf (related to Figure 2). A) Principal component analysis. We used pseudo-haploid genotype calls of high-coverage genomes overlapped with the ancient Taimyr individual (red). B) PCA using dogs and wolves genotyped at ~170,000 SNPs.
Figure S3. Admixture graph population history models of the ancient Taimyr individual and present-day canids (related to Figure 2).
A) Rejected graph topology without gene flow. Note that even in this graph the internode distance between the Taimyr wolf and the dog-wolf ancestor node is only $F_{ST} = 0.006$
B) Admixture graph with regional dog-to-wolf gene flow and Taimyr basal to the dog-wolf ancestor.
C) Admixture graph with regional wolf-to-dog gene flow and Taimyr basal to the dog-wolf ancestor.
D) Dog-to-wolf gene flow and Taimyr on the wolf lineage.
E) Wolf-to-dog gene flow and Taimyr on the wolf lineage.
F) Dog-to-wolf gene flow and Taimyr on the dog lineage.
G) Wolf-to-dog gene flow and Taimyr on the dog lineage.

All graphs displayed in B-G provide a good fit to the data in the sense that the $f_4$ statistics that they predict are within 3 SEs of the empirically observed SEs. Nodes ANC, W0, D1 etc. are hypothesized ancestral populations in each graph. Numbers given for branch lengths are scaled as $F_{ST} \times 1000$, but note that drift lengths for admixture events (e.g. W2-W6 in B) are not mathematically determined.
Figure S4. Evidence of a recent origin of worldwide wolf populations (related to Figure 4).

A) Admixture graph fitted using TreeMix providing evidence that the Taimyr individual is substantially closer to present-day gray wolves than to coyotes, but that worldwide present-day gray wolves share substantial recent ancestry after the divergence of the Taimyr individual. Note that the long branch of the Taimyr individual is an artifact of haploid genotype calling and that this does not affect its relative covariance with other populations. B) $D$-statistics are consistent with Taimyr 1 being basal to all worldwide wolf populations, and tests where Taimyr 1 is more closely related to one gray wolf population than another are all rejected. Error bars are 3 standard errors of the $D$-statistic.
Table S1 (related to Figure 1). Sequencing statistics of the Taimyr nuclear and mitochondrial genome.

<table>
<thead>
<tr>
<th></th>
<th>All data</th>
<th>mitochondrial DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Seqs</td>
<td>MAPQ≥30</td>
</tr>
<tr>
<td>Non-treated</td>
<td>7,582,638</td>
<td>6,509,172</td>
</tr>
<tr>
<td>USER-treated</td>
<td>27,209,886</td>
<td>23,942,017</td>
</tr>
<tr>
<td>Total</td>
<td>34,792,524</td>
<td>30,451,189</td>
</tr>
</tbody>
</table>

Note: An asterisk denotes that mean sequencing depth was obtained by dividing the number of bp of sequences aligned to the genome (without mapping quality restrictions) with the number of bp (2,392,715,236) in the canFam3.1 assembly (http://useast.ensembl.org/Canis_familiaris/Info/Annotation, accessed 2 October 2014). Mean sequencing depth for the mtDNA was obtained by dividing the number of bp of sequences aligned to the mtDNA with 16,727, the number of bp in the canFam3.1 assembly of the dog mtDNA genome.
Table S2. Statistics used to estimate the coalescent time separating the divergence of the Taimyr individual from present-day gray wolves (related to Figure 3).

| B                | A      | $F(A_{\text{derived}}|B_{\text{heterozygote}})$ | Block jackknife SE | Loci       |
|------------------|--------|-----------------------------------------------|--------------------|------------|
| Croatian wolf    | Taimyr | 30.61%                                        | 0.11%              | 602,538    |
| Croatian wolf    | Chinese wolf | 31.35%                                    | 0.11%              | 1,353,073  |
| Croatian wolf    | boxer  | 31.41%                                        | 0.11%              | 1,391,231  |
| Croatian wolf    | dingo  | 31.35%                                        | 0.11%              | 1,308,264  |
| Chinese wolf     | Taimyr | 30.85%                                        | 0.12%              | 538,141    |
| Chinese wolf     | Croatian wolf | 31.90%                                   | 0.10%              | 1,133,764  |
| Chinese wolf     | boxer  | 31.78%                                        | 0.11%              | 1,244,847  |
| Chinese wolf     | dingo  | 32.14%                                        | 0.10%              | 1,154,013  |

Note: We estimate the probability $F(A_{\text{derived}}|B_{\text{heterozygote}})$ that a randomly chosen allele from individual A is derived at positions where another individual B displays a heterozygotic genotype. We estimate standard errors using a weighted block jackknife of 5 mb blocks, where each block is weighted by the number of informative loci.
Table S3 (related to Figure 4). Excess allele sharing between the Taimyr wolf and Asian dog breeds. $D(T,W;X,DP)$ abbreviates $D$(Taimyr, Wolf; Breed X, Boxer, Gordon Setter, Dobermann Pincher and Newfoundland). $D(AF,T;W,X)$ abbreviates $D$(Andean fox, Taimyr; Wolf, Breed X).

<table>
<thead>
<tr>
<th>Breed</th>
<th>$D(T,W;X,DP)$</th>
<th>$D(AF,T;W,X)$</th>
<th>n</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australian Shepherd</td>
<td>0.008</td>
<td>0.14</td>
<td>1</td>
<td>USA</td>
</tr>
<tr>
<td>Belgian Tervuren</td>
<td>-0.0048</td>
<td>-1.25</td>
<td>10</td>
<td>Belgium</td>
</tr>
<tr>
<td>Beagle</td>
<td>0.0040</td>
<td>0.99</td>
<td>12</td>
<td>Great Britain</td>
</tr>
<tr>
<td>Bernese Mountain Dog</td>
<td>-0.0010</td>
<td>-0.22</td>
<td>12</td>
<td>Swiss Alps</td>
</tr>
<tr>
<td>Border Collie</td>
<td>0.0043</td>
<td>1.21</td>
<td>16</td>
<td>English-Scottish Border</td>
</tr>
<tr>
<td>Border Terrier</td>
<td>0.0031</td>
<td>0.74</td>
<td>25</td>
<td>Northumberland, England</td>
</tr>
<tr>
<td>Boxer</td>
<td>0.0053</td>
<td>1.09</td>
<td>8</td>
<td>Germany</td>
</tr>
<tr>
<td>Brittany Spaniel</td>
<td>0.0022</td>
<td>0.65</td>
<td>12</td>
<td>Brittany, France</td>
</tr>
<tr>
<td>Cocker Spaniel</td>
<td>0.0030</td>
<td>0.72</td>
<td>5</td>
<td>Chihuahua, Mexico</td>
</tr>
<tr>
<td>Cavalier King Charles Spaniel</td>
<td>-0.0032</td>
<td>-0.58</td>
<td>2</td>
<td>Blenheim, England</td>
</tr>
<tr>
<td>Cockerspaniel</td>
<td>-0.0034</td>
<td>-1.00</td>
<td>14</td>
<td>York, England</td>
</tr>
<tr>
<td>Czechoslovakian Wolf Dog</td>
<td>-0.0139</td>
<td>-2.48</td>
<td>12</td>
<td>Czechoslovakia</td>
</tr>
<tr>
<td>Dachshund</td>
<td>0.0012</td>
<td>0.40</td>
<td>7</td>
<td>Germany</td>
</tr>
<tr>
<td>Dalmatian</td>
<td>0.0015</td>
<td>0.36</td>
<td>25</td>
<td>Dalmatia, Croatia</td>
</tr>
<tr>
<td>Doberman Pinscher</td>
<td>-0.0063</td>
<td>-1.71</td>
<td>12</td>
<td>Apolda, Germany</td>
</tr>
<tr>
<td>English Bulldog</td>
<td>-0.0008</td>
<td>-0.17</td>
<td>13</td>
<td>England</td>
</tr>
<tr>
<td>English Bull Terrier</td>
<td>-0.0071</td>
<td>-1.22</td>
<td>8</td>
<td>Birmingham, England</td>
</tr>
<tr>
<td>English Cocker Spaniel</td>
<td>-0.0061</td>
<td>-1.33</td>
<td>2</td>
<td>England</td>
</tr>
<tr>
<td>Elkhound</td>
<td>0.0035</td>
<td>1.01</td>
<td>12</td>
<td>Jämtland, Sweden</td>
</tr>
<tr>
<td>English Springer Cocker Spaniel</td>
<td>-0.0126</td>
<td>-2.93</td>
<td>3</td>
<td>England</td>
</tr>
<tr>
<td>English Setter</td>
<td>-0.0016</td>
<td>-0.41</td>
<td>12</td>
<td>England</td>
</tr>
<tr>
<td>Eurasian</td>
<td>0.0055</td>
<td>1.34</td>
<td>12</td>
<td>Germany</td>
</tr>
<tr>
<td>Flatcoated Retriever</td>
<td>0.0035</td>
<td>0.71</td>
<td>2</td>
<td>UK</td>
</tr>
<tr>
<td>Finnish Spitz</td>
<td>0.0101</td>
<td>2.29</td>
<td>14</td>
<td>Finland</td>
</tr>
<tr>
<td>Gordon Setter</td>
<td>0.0015</td>
<td>0.58</td>
<td>25</td>
<td>Scotland</td>
</tr>
<tr>
<td>Golden Retriever</td>
<td>0.0048</td>
<td>1.28</td>
<td>12</td>
<td>Scotland</td>
</tr>
<tr>
<td>Greyhound</td>
<td>0.0025</td>
<td>0.59</td>
<td>11</td>
<td>UK</td>
</tr>
<tr>
<td>German Shepherd</td>
<td>-0.0053</td>
<td>-1.15</td>
<td>12</td>
<td>Stuttgart, Germany</td>
</tr>
<tr>
<td>Greenland Sledge Dog</td>
<td>0.0173</td>
<td>3.64</td>
<td>14</td>
<td>Greenland</td>
</tr>
<tr>
<td>Siberian Husky</td>
<td>0.0208</td>
<td>4.27</td>
<td>1</td>
<td>Siberia, Russia</td>
</tr>
<tr>
<td>Irish Wolfhound</td>
<td>-0.0023</td>
<td>-0.48</td>
<td>2</td>
<td>Ireland</td>
</tr>
<tr>
<td>Jack Russell Terrier</td>
<td>0.0014</td>
<td>0.49</td>
<td>25</td>
<td>Oxford, England</td>
</tr>
<tr>
<td>Large Munsterlander</td>
<td>0.0014</td>
<td>0.24</td>
<td>23</td>
<td>Munster, Germany</td>
</tr>
<tr>
<td>Labrador Retriever</td>
<td>0.0013</td>
<td>0.38</td>
<td>17</td>
<td>Labrador, Canada</td>
</tr>
<tr>
<td>Mops</td>
<td>0.0092</td>
<td>1.65</td>
<td>12</td>
<td>China</td>
</tr>
<tr>
<td>Newfoundland</td>
<td>0.0026</td>
<td>0.89</td>
<td>2</td>
<td>Newfoundland, Canada</td>
</tr>
<tr>
<td>Nova Scotia Duck Tolling Retr.</td>
<td>0.0018</td>
<td>0.44</td>
<td>2</td>
<td>Nova Scotia, Canada</td>
</tr>
<tr>
<td>Rottweiler</td>
<td>-0.0004</td>
<td>-0.08</td>
<td>25</td>
<td>Rottweil, Germany</td>
</tr>
<tr>
<td>Samoyed</td>
<td>0.0067</td>
<td>1.36</td>
<td>3</td>
<td>Nenentsia, Russia</td>
</tr>
<tr>
<td>Sarloos</td>
<td>-0.0269</td>
<td>-5.15</td>
<td>11</td>
<td>Netherlands</td>
</tr>
<tr>
<td>Schipperke</td>
<td>0.0024</td>
<td>0.63</td>
<td>2</td>
<td>Belgium</td>
</tr>
<tr>
<td>Schnauzer</td>
<td>-0.0039</td>
<td>-1.00</td>
<td>12</td>
<td>Mecklenburg, Germany</td>
</tr>
<tr>
<td>Shar Pei</td>
<td>0.0127</td>
<td>2.98</td>
<td>6</td>
<td>Guangdong, China</td>
</tr>
<tr>
<td>Standard Poodle</td>
<td>0.0006</td>
<td>0.17</td>
<td>26</td>
<td>Germany</td>
</tr>
<tr>
<td>Yorkshire Terrier</td>
<td>0.0074</td>
<td>2.19</td>
<td>12</td>
<td>Yorkshire, England</td>
</tr>
<tr>
<td>Weimaraner</td>
<td>0.0015</td>
<td>0.34</td>
<td>0</td>
<td>Germany</td>
</tr>
</tbody>
</table>
Table S4 (related to Figure 4). Evidence for biased affinity to dogs for genome sequence data overlapped with SNP array data.

<table>
<thead>
<tr>
<th>Test genome</th>
<th>full data</th>
<th>boxer asc.</th>
<th>boxer-poodle asc.</th>
<th>boxer-wolf asc.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$D$</td>
<td>$Z$</td>
<td>$D$</td>
<td>$Z$</td>
</tr>
<tr>
<td>Taimyr</td>
<td>-0.050</td>
<td>-4.006</td>
<td>-0.035</td>
<td>-1.937</td>
</tr>
<tr>
<td>CroatianWolf</td>
<td>0.028</td>
<td>3.609</td>
<td>0.052</td>
<td>4.479</td>
</tr>
<tr>
<td>ChineseWolf</td>
<td>-0.038</td>
<td>-4.767</td>
<td>-0.023</td>
<td>-1.872</td>
</tr>
<tr>
<td>GoldenJackal</td>
<td>-0.060</td>
<td>-6.276</td>
<td>-0.048</td>
<td>-3.638</td>
</tr>
</tbody>
</table>

Note: We show the statistic $D$(golden_jackal, Test genome; boxer, gray_wolf_Europe_Ukraine), computed on SNPs obtained from different ascertainment schemes. Positive values are coloured in blue and negative values in red.
Supplemental Experimental Procedures

Sample description
During the Taimyr Peninsula 2010 expedition organised by the Swedish Polar Research Secretariat, a partial rib (sample ID TX085) was collected at an ice complex site (N73°31'00", E104°32'20") along the Bolshaya Balakhnaya River. The rib was found ex situ and was only partial, making both its age and species identity uncertain. However, subsequent PCR amplification and sequencing of a part of the mitochondrial 16S rRNA gene [S1] established the rib as that from a gray wolf (Canis lupus) and AMS radiocarbon dating provided an age of 30,920±380 14C years BP (OxA 28059). This age is equivalent to ~34,900 calendar years BP when calibrated using the OxCal software v.4.2 [S2] and the IntCal 13 calibration curve [S3] (mean 34,888; median 34,865; range (95.4%) 35,659 - 34,142).

DNA extraction and sequence library preparation
Extraction of the sample and library preparation was performed at a specifically designated ancient DNA laboratory with high standards of sterility at the Swedish Museum of Natural History in Stockholm, where no work had previously been performed on canids. DNA was extracted using a silica-based method coupled with concentration on Vivaspin filters (Sartorius), following a modified version of protocol C in Yang et al. [S4], as described in Brace et al. [S5] Two sequencing libraries were prepared from 20μl of DNA extract using the multiplexed double-stranded library preparation protocol for Illumina sequencing by Meyer and Kircher [S6]. One of the libraries was treated with the USER enzyme (New England Biolabs) during the blunt-end repair step to remove uracils that derive from cytosine deamination [S7]. The non-USER-treated library was indexed and amplified with AmpliTaq Gold® (Life technologies) and the USER-treated library with the high-fidelity polymerase AccuPrime™ Pfx (Life technologies). Both libraries were purified with magnetic beads (Agencourt AMPure XP, Beckman Coulter) and their concentration was measured on the Bioanalyzer 2100 (Agilent) using high-sensitivity DNA chips. The non-USER-treated library was pooled together with a library from another sample at equimolar concentrations and the pool was sequenced on a single lane of an Illumina HiSeq 2500 flowcell with a paired-end 2x100bp setup in RapidRun mode. Sequencing of the USER-treated library was performed on 2 lanes of the Illumina HiSeq 2500 flowcell with a paired-end 2x100bp setup in HighOutput mode.

Ancient genomic data processing
For the sequencing data from libraries that were not treated with the USER enzyme mix, we merged read pairs where the expected index was observed using MergeReadsFastQ_cc.py [S8], requiring an overlap of at least 11 bp. We then mapped the sequence reads to the dog genome assembly (builds canFam2 and canFam3.1) using BWA version 0.5.9 [S9] with parameters -l 16500 -n 0.01 -o 2, and collapsed PCR duplicate reads with identical start and end coordinates into consensus sequences using FilterUniqueSAMCons.py [S8]. For the USER-treated sequencing data, we merged all read pairs using the software package SeqPrep 1.1 (URL: https://github.com/jstjohn/SeqPrep) using default parameters, and mapped to the dog genome assemblies as above, collapsing PCR duplicates using samtools rmdup [S10] with default parameters. For both data sets, we required mismatches to the reference genome to be less than 10% for each sequence, and discarded sequences of less than 35 bp.
Postmortem DNA damage

We computed postmortem DNA damage profiles from the libraries by modifying the PMDtools program [S11], allowing for separation of the data into two categories: 1) positions aligned to a nucleotide in CpG context and 2) positions aligned to a nucleotide in non-CpG context. This stratification is available in PMDtools v0.56 at https://code.google.com/p/pmdtools/. We used 23,942,017 sequences from the USER-treated data and 168,327 from the non-USER-treated data that mapped to chromosome 38 with a mapping quality of at least 30 to compute damage patterns for these two categories. We also computed damage patterns on 44,150 sequences with mapping quality of at least 30 from the USER-treated data that mapped to the mitochondrion, to test the notion that cytosines in mtDNA are unmethylated and thus all cytosine deamination events result in uracils. We required that each investigated base had a phred-scaled quality score of at least 30.

Biological sex

No Y-chromosome reference sequence is available from the canFam3.1 reference genome (the sequenced individual was a female boxer [S12]), but we found that only 832,657 sequences with mapping quality of at least 30 aligned to the 123,869,142 bp chromosome X sequence, compared to 1,635,494 sequences which aligned to the 122,678,785 bp chromosome 1. Normalizing by the length of the reference sequences, the number of X-chromosome alignments is (832,657 /123,869,142) / (1,635,494 /122,678,785) = 50.4% of what would be expected based on the chromosome 1 observation. For the Chinese wolf genome, the same fraction is 98.6%, whereas for the Dingo it is 53.2%. We thus conclude that the individual likely had only a single copy of chromosome X and thus was male.

Modern reference data curation and merging with the ancient wolf genome

We obtained the genomes of 6 canids sequenced using Illumina and Solid technologies by Freedman et al. [S13]. We obtained genotype calls directly from the authors and identified all polymorphic alleles between the 6 canids and the Boxer reference genome that passed both global and sample-specific quality filters implemented by the original study [S13]. Importantly, the filters applied by the original study excluded all positions that could possibly have been in CpG context. Since all discernible post-mortem derived damage in the USER-treated Taimyr wolf is due to methylated CpG sites (Figure S1), the remaining errors in the Taimyr sequence should all be on the magnitude expected from sequence errors, i.e. ~1/1000 given our phred-scaled base quality threshold of 30. We added haploid genotypes from the USER-treated Taimyr genome data by randomly sampling a single read at each locus to each identified reference locus [S14], requiring a mapping quality for each sequence of at least 30, and a minimum base quality of each base of 30. For the complete genomes, the number of SNPs where we called a base for the Taimyr wolf was 3,639,567 (41.9% of the total number of 8,686,809 SNPs) after restricting to the USER-treated data. To compare the Taimyr wolf to a broader diversity of modern-day dog breeds, we obtained a data set of 532 dogs from 48 breeds and 15 gray wolves genotyped at 169,066 SNP loci on the Illumina CanineHD array [S15]. For this data set, approximately 66,000 SNPs were covered by the USER-treated Taimyr data after applying quality filters as above. To compare the Taimyr wolf to a broader diversity of modern-day gray wolf populations, we used the CanMap data set [S16, S17] typed on the Affymetrix Canine version 2 genome-wide SNP mapping array, comprising a total of 1,235 canids genotyped at 47,934 SNPs, including 199 gray wolves and genotypes extracted from the 6 genome sequences first published by Freedman et al. [S13]. We obtained 21,687 autosomal SNPs for which we also had data from the Taimyr genome.
We processed the Andean fox data similarly to the Taimyr genome, mapping the first read in each pair to canFam2, canFam3 and canFam3.1 using BWA with default parameters.

**Mitochondrial DNA analysis**

We assembled a consensus sequence of the Taimyr mitochondrial genome restricting to data obtained from the USER-treated library. We obtained 46,113 sequences mapping to the mitochondrion of canFam2, 44,150 of which had a mapping quality of at least 30. The average sequencing depth for the mitochondrial genome without filtering was 182X, and no site was covered by less than three reads (two sites that were covered by three reads all agreed with the consensus base). We used the `vcfutils.pl` tool in the *samtools* suite [S10] to call a consensus sequence, with default parameters.

We performed a Bayesian phylogenetic analysis on the mitochondrial genome sequence of Taimyr 1 together with previously published mitochondrial genome sequences of both modern and ancient canids [S18]. A phylogeny was reconstructed using *BEAST* 1.8.0 [S19], applying the HKY + G model of nucleotide substitution and assuming constant population size as a coalescent tree prior. The posterior distribution of nodes, divergence times and substitution rates were estimated by Markov chain Monte Carlo, where samples were drawn every 1000 MCMC steps from a total of 10 million steps, following a discarded burn-in of 1 million steps. Convergence to the stationary distribution and sufficient sampling were checked by inspection of posterior samples and ESS values in *Tracer* 1.5.2 [S19]. Radiocarbon dates were used as internal calibration points applying the ‘Estimate’ option with no prior on the substitution rate. Two independent analyses were run and combined in *LogCombiner* 1.8.0, after which convergence between the two runs was checked using *Tracer* 1.5.2 [S19].

**Principal component analysis**

We performed principal component analysis using *EIGENSOFT* v4.0 [S20], excluding one locus from each pair in linkage disequilibrium, which was assessed with the \( r^2 \) statistic \((r^2 > 0 \) was excluded\). Since we did not make diploid genotype calls for the ancient Taimyr wolf genome, we pseudo-haploidized the modern-day data by randomly sampling a single allele at each site [S21]. We find that the two first principal components (PC) computed using Taimyr and the 7 present-day canid genomes see Taimyr 1 clustering intermediately to wolves and dogs (Figure S2A), as expected due to the 35,000 years of genetic drift that has occurred in dogs and wolves since the death of the Taimyr wolf. We note that the Taimyr wolf was included in the PC computations and not projected, as is often the practice for ancient humans [S22, S23]. This is motivated by the lack of postmortem damage outside CpG context in the Taimyr genome as well as the fact that only a single ancient individual is included, so there is no excess of missing data. This also serves as quality control, since if the Taimyr wolf appeared highly differentiated from all the present-day genomes in the PC analysis this would suggest that it is affected by sequencing- or processing errors. However, we see that the most extreme outliers are the golden jackal and the boxer, which might be expected from the evolutionary distance of the golden jackal to the other canids, and the extensive bottlenecks that have occurred in recent European dog breeds. Similar results are obtained for a PCA of Taimyr and present-day canids genotyped on the Illumina CanineHD array (Figure S2B).
**D-statistics**

*D*-statistics quantify excess correlations in allele frequencies that deviate from the expectation of a null model of a tree-like population history [S14, S24]. Given, for example, a history where dogs and wolves became isolated at a specific point with no subsequent gene flow, we expect that for a polymorphism discovered between two dogs, two wolf chromosomes that are polymorphic should be equally probable to carry the derived or ancestral allele. Extended to population-wide allele frequency data, we expect that the product \( (p_{\text{wolf1}} - p_{\text{wolf2}})(p_{\text{dog1}} - p_{\text{dog2}}) \) should be consistent with 0 when summed over many loci.

To compute *D*-statistics on empirical data, we used the estimation framework described in [S25]

\[
\text{Numerator} = (p_A - p_B) - (p_X - p_Y) \\
\text{Denominator} = (p_A + p_B - 2p_Ap_B)(p_X + p_Y - 2p_Xp_Y)
\]

where \( p_A \) is the frequency of one arbitrarily chosen allele in population \( A \) at marker \( i \). To obtain genome-wide estimates, the *Numerator* and *Denominator* is summed for all \( n \) markers [S25, S26] and \( D = \text{Numerator} / \text{Denominator} \). We obtained standard errors by performing a weighted block jackknife over 5 Mb blocks in the genome.

The use of an Andean fox (*Lycalopex culpaeus*) genome [S27] as the outgroup for most of these analyses was motivated by the evidence for gene flow between the golden jackal and other canids [S27], and we also excluded the Israeli wolf and Basenji genomes for this reason [S13], but were still left with one dog and one wolf from both Western- (Boxer and Croatian wolf) and Eastern Eurasia (Dingo and Chinese wolf).

**Admixture graphs of population history using complete genomes**

We used the program *ADMIXTUREGRAPH* [S24, S25] which uses *f*-statistics of allele frequency correlations between samples to assess whether a fitted admixture graph of population history is consistent with the data. *ADMIXTUREGRAPH* optimizes the fit of a proposed admixture graph in which each node can be descended either from a mixture of two other nodes, or from a single ancestral node from which it may be separated by genetic drift. *ADMIXTUREGRAPH* optimizes the fit between predicted and empirical *f*\( _2 \) statistics of the form \( f_2(A, B) = (p_A - p_B) \), where \( p_A \) and \( p_B \) are the allele frequencies of populations \( A \) and \( B \), respectively, and the statistic is summed over all \( n \) loci [S24, S25]. To assess the fit of a given model to the data, all possible *f*\( _4 \) statistics \( f_4(A, B; X, Y) = (p_A - p_B) - (p_X - p_Y) \) for the empirical data and the fitted model are compared. Following previous studies, we consider statistics which deviate by normalized \( Z \)-scores > 3 between predicted and empirical statistics as evidence against the tested graph hypothesis.

We first tested a simple tree-like model where the Taimyr individual is basal to all present-day wolves and dogs, with the two dogs and the two wolves forming separate clades (Figure S3A). We found that this model was inconsistent with the data, with 8 *f*\( _4 \) statistics predicted by the model to be zero deviating by between 3 < \( |Z| < 7.3 \) standard errors from zero. We also note that this model was in fact fitted as a trifurcation between the dog lineage, and the Chinese and Croatian wolf lineages, respectively, since the drift inferred from the divergence of wolves and dogs to the divergence of the two wolves was 0. Many statistics implied an excess affinity between the dingo and the Chinese wolf, e.g. \( f_4(\text{Andean fox, dingo}; \text{Chinese wolf, Croatian wolf}) \).
= -0.005328 (Z = -4.0), as well as an excess affinity between the Croatian wolf and the boxer, e.g. \( f_4(\text{Andean fox, boxer; Chinese wolf, Croatian wolf}) = 0.004249 \) (Z = -3.2). Thus we next tested two modified models where either the dingo or the boxer had partial Chinese wolf-related ancestry, and Croatian wolf-related ancestry, respectively, or, the Chinese wolf and the Croatian wolf had partial dingo-related and boxer-related ancestry, respectively. Both these models were good fits to the data with no \( f_4 \) statistics deviating from predicted values by more than 3 standard errors, respectively (Figure S3B and S3C). In addition, models of dog-to-wolf, or wolf-to-dog gene flow, where Taimyr was on the lineage leading to either dogs or wolves also fit the data with no significant deviations (Figure S3D, S3E, S3F, and S3G). However, all these posit drift corresponding to \( F_{ST} < 0.01 \) between the Taimyr branch and the original divergence of wolves and dogs. We therefore conclude that our graph modeling suggests that the wolf, dog, and Taimyr lineages all diverged at about the same time. In the main text, we report the model that posits dog-to-wolf gene flow and Taimyr as basal to the two other lineages (in effect a trifurcation due to zero-length drift), since it requires admixture at a lower rate. However, some bidirectional gene flow is probably more biologically realistic [S13], in which case some combination of the two models, each with lower admixture proportions in any given direction, is more likely.

**Divergence time estimation and calibration of dog-wolf divergence**

We used an approach for estimating the population divergence of the Taimyr individual’s lineage from modern canids (scaled in coalescent time) that has previously been used to estimate the divergence time between Neandertals and modern humans [S28]. This approach is based on detection of heterozygous positions in the genome of a single present-day individual \( B \), and then estimating the probability \( F(A|B) \) of a second (ancient) genome \( A \) carrying the derived allele at a randomly chosen chromosome. We estimate standard errors using a weighted block jackknife of 5 mb blocks, where each block is weighted by the number of informative loci. Since this approach only samples a single chromosome from the ancient individual, population size changes in the lineage specific to the population the ancient individual belonged to do not enter into the probability of it carrying the derived allele. However, the proportion of derived alleles in \( B \) that are not present in \( A \) is also affected by the genetic drift that has occurred in the ancestry of \( B \).

To account for genetic drift, we built a calibration curve for \( F(A|B) \) taking demographic history into account by estimating historical changes in effective population size in the Chinese wolf using the \( PSMC \) method [S29]. We chose the Chinese wolf for this analyses since a previous study showed that the resolution of the inference for the other canid genomes was poorer due to insufficient recombination events in the most recent time period [S13]. We used genotypes inferred by the original authors, requiring a genotype quality of at least 20. We ran the \( PSMC \) inference using the parameters `-N20 -t10 -r5 -p "64*1"`, limiting the last coalescence to 10 times the effective population size and inferring \( N_e \) over all 64 atomic time intervals. We then simulated 900 mb under the inferred PSMC population size history using \( MaCS \) [S30], over a grid of divergence times for a single haploid lineage \( A \) and computed \( F(A|B) \) for each simulation replicate. Initial calibration of coalescent time units to chronological years assumed a generation time of 3 years and a mutation rate of \( 1 \times 10^{-8} \) per bp per generation, corresponding to \( 3.33 \times 10^{-9} \) mutations per bp per year, which yields a starting \( N_e \) of the gray wolf of \( \sim 7000 \). We investigated a range of calibrations of the per-generation mutation rate in order to identify a range that would
be compatible with the age of the Taimyr wolf. We note however that the slower mutation rate that we infer could also be explained by a longer generation time interval than 3 years. Regardless, this longer generation time would have the same effect of recalibrating models of dog-wolf divergence to a longer time scale. We also note that SNPs in CpG context were excluded from the present-day SNP panel, and so the upper bound on the mutation rate inferred here excludes such sites.

**Recent shared ancestry of worldwide gray wolves**

To investigate the relationship between the Taimyr wolf and a larger set of gray wolf populations, including New World gray wolves and coyotes typed on the Affymetrix Canine version 2 genome-wide SNP mapping array [S16], we first fitted admixture graphs using the heuristic approach employed by TreeMix. We excluded dogs from this analysis due to the observation of biases towards dogs when overlapping genome sequence data to this data set (see below). We fitted between 0 and 3 admixture edges using a selected set of wolf and coyote populations from across the covered distribution that all had appreciable sample size, estimating standard errors of the covariance matrix using blocks of 30 contiguous SNPs. We found that assuming 0 or 1 migration edge resulted in fitted covariances that deviated from empirical values of more than 3 SEs, suggesting poor fit, but that the maximum deviation for 2 migration edges was 2.4, suggesting good fit. These two migration edges featured North American gray wolf ancestry in the coyote, in agreement with previous results [S16], as well as gene flow from a basal canid into the Israeli wolf, also in agreement with previous results [S13] and our ADMIXTUREGRAPH modeling using the complete genome data (Figure S2).

This best-fitting model (Figure S4) places the ancient Taimyr wolf as being basal to all gray wolves, but sharing a substantial amount of history with the present-day gray wolves after their divergence from the coyote. To test this further, we computed $D(\text{golden jackal, Taimyr; wolf1, wolf2})$ for the selected wolf populations (excluding the Israeli wolf due to its basal ancestry) and found that these statistics were all consistent with 0 (Figure S4). In contrast, there is evidence for all present-day wolf populations sharing genetic drift with each other, which is not shared with the Taimyr wolf, since statistics of the form $D(\text{golden jackal; wolf1, wolf2, Taimyr})$ are significantly negative. Since the Taimyr wolf genome is consistent with being basal to gray wolves from the Middle East, China, Europe and North America, one possible historical scenario is that the majority of gray wolf ancestry today stems from an ancestral population that lived less than 35,000 years ago, but we note that we cannot exclude that this ancestral population diverged from the population of the Taimyr wolf earlier than its lifetime.

**Admixture into high-latitude dog breeds**

To test for admixture between the Taimyr wolf and a large set of modern dog breeds, we first tested the null hypothesis $D(\text{Andean fox, Taimyr; present-day gray wolves, dog breed } X)$ using 15 present day gray wolves and 48 dog breeds genotyped on the Illumina CanineHD array (Table S3). This null hypothesis assumes the topology inferred for the complete genome sequence data: that the Taimyr individual is basal to present-day gray wolves and dogs. A significantly positive statistic can be interpreted as excess derived allele sharing between the Taimyr individual and the dog breed $X$. We found that the two most significant statistics were indeed positive, and involved the Greenland sledge dog ($Z = 2.96$) and the Siberian Husky ($Z = 2.2$), two dog breeds that both originate from arctic human populations.
To increase power to detect a genetic affinity to the Taimyr wolf, we computed $D$ (present-day gray wolves, Taimyr; dog pool, dog breed $X$) which tests if allele frequency differences between a dog breed $X$ and a pool of present-day breeds (Newfoundland, Boxer, Gordon Setter and Doberman Pincher) are correlated to the allele frequencies of either present-day gray wolves or the Taimyr individual. Positive statistics can be interpreted as an excess affinity of dog breed $X$ to the Taimyr individual, whereas negative statistics can be interpreted as an excess affinity to the present-day gray wolves. We again find strong evidence of Taimyr affinity in the Siberian Husky ($Z = 4.27$) and Greenland sled dog ($Z = 3.64$), but also in the Shar Pei ($Z = 2.98$) and the Finnish Spitz ($Z = 2.29$), who are other 'ancient' breeds associated with high latitudes. In contrast, the Dutch Saarloos wolfdog was closer to present-day gray wolves than Taimyr 1 ($Z = -5.2$), in agreement with documented historical crossbreeding with wolves in this breed.

To estimate the proportion of Taimyr-derived ancestry in the Greenland sled dog (the Siberian Husky is represented by a single individual in our data so is limited in power for such an inference), we used ADMIXTUREGRAPH to fit a graph consisting of the Andean fox outgroup, the Taimyr individual, present-day gray wolves, the Greenland sled dog and the German Shepherd. We used exactly the same topology as inferred for our genome-wide data, that the Taimyr wolf is basal to present-day gray wolves and dogs, but fitted the Greenland sled dog as being comprised both of dog and Taimyr-related ancestry. We then incrementally tested different proportions of ancestry, and found that a proportion between 1.4% and 27.3% resulted in no deviations of more than $|Z| = 2$ between empirical $f_4$ statistics and those predicted by the model.

Comparison between complete genomes and the Affymetrix Canine SNP mapping array

We also wanted to corroborate the affinity between the Taimyr individual and gray wolves using data from the Affymetrix Canine version 2 genome-wide SNP mapping array [S16, S17]. However, we found evidence for biases towards dogs in genome sequences merged into this data set. This can be seen for example in the statistic $D$ (golden jackal, Chinese wolf genome; boxer, gray wolf Ukraine) = -0.038 ($Z = 4.8$), which suggests that even the high-quality Chinese wolf genome shares more alleles with the boxers genotyped on the SNP array than with Ukrainian gray wolves genotyped on the SNP array (the golden jackal data in these tests was also SNP array genotypes).

Replacing the Chinese wolf genome with Taimyr, Croatian wolf, and golden jackal genomes, we find that both the Taimyr individual and the golden jackal show a similar attraction to the boxer (Table S4). In contrast, the Croatian wolf genome is correctly identified as being more closely related to the Ukrainian wolf SNP array data. This suggests that the extra drift shared between the two European wolves is enough to overcome the bias that is causing the attraction of e.g. the Chinese wolf to dogs in the SNP array data. To study this further, we stratified the data by SNPs that were ascertained by Lindblad-Toh et al. [S12] as either $i$) being polymorphic in the boxer reference genome (13,302 SNPs), $ii$) polymorphic between the boxer reference genome and a standard poodle individual (24,930 SNPs), and $iii$) polymorphic between the boxer reference genome and shotgun sequences from wolves (634 SNPs). While the uncertainty is large, the bias seems exacerbated for the SNPs ascertained as boxer-wolf differences, for which even the Croatian wolf genome appears closer to the boxer than to Ukrainian wolves ($D = -0.14$, $Z = 2.2$) (Table S1).
We conclude that using the SNP array data for this purpose is unreliable. One explanation could be that this is due to reference alignment bias in the Chinese wolf genome, since the reference genome is a boxer individual. Importantly however, this could not explain our results based on genome sequences that the Taimyr individual is closer to wolves, since reference alignment bias would be expected to cause it to be closer to dogs. Given the careful filtering of the genome sequence data and lack of ascertainment bias, we suggest that the genome sequence data is more trustworthy for these purposes, and the observation that e.g. D(goldenJackal, Chinese wolf; basenji, boxer) using genome sequence data is consistent with 0 suggests that reference alignment bias is not a major issue. Other explanations could include ascertainment bias, or allelic biases in the array genotyping.

References


