www.sciencemag.org/content/342/6159/871/suppl/DC1

Supplementary Materials for

Complete Mitochondrial Genomes of Ancient Canids Suggest a European Origin of Domestic Dogs

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. Giemsch, 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

*Corresponding author. E-mail: olatha@utu.fi

Published 8 November 2013, *Science* **342**, 871 (2013)
DOI: 10.1126/science.1243650

This PDF file includes

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/content/342/6159/page/suppl/DC1)

Materials and Methods

Samples

Ancient samples

A total of 28 ancient canids from sampling sites within Eurasia and the Americas were analyzed in this study (Table S1). The age of the samples ranged from 36,000 years ago to approximately 1,000 years ago. Whereas the taxonomic classification based on cranial morphometrics of some of the early putative dogs is currently a matter of debate (*2, 8, 9, 26, 27*), a consensus seems to exists with regard to younger dog specimens and ancient wolves (*2, 9, 12, 27, 28*). As it is not the focus of this paper to resolve this debate, we simply acknowledge this taxonomic uncertainty by assigning the specimens in question with a different color code in our figures (Figs. 1, S3, S9, S10) and providing the relevant citations. The ancient samples were prepared in two sets. The first set of samples (upper section of Tables 1, S1) were extracted in the ancient DNA facilities at the University of California Los Angeles (UCLA), USA and subsequently processed at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany. The second set of samples (lower panel of Tables 1, S1) were extracted and processed at the ancient DNA laboratory of the Institute for archaeological Sciences at the University of Tübingen, Germany. We have utilized two different DNA capture strategies as outlined below.

Modern samples

In order to establish a comprehensive data set we downloaded publicly available, complete mitochondrial genomes (mt-genomes) including 72 dogs, nine gray wolves (*Canis lupus*), and three coyotes (*Canis latrans*) (Table SI 2). We further added two complete mt-genomes recently generated by complete genome sequencing of the Basenji and Dingo and extracted the mitochondrial genomes of three Chinese dogs (*7*). In order to increase the geographic

distribution of wolves we generated additional 40 complete mt-genomes of wolves and also added a single mt-genome of a coyote (Table S2).

Sample preparation

Ancient Samples

Sample processing, UCLA

Fourteen ancient canids from various Pleistocene and Holocene excavation sites (upper panel in Table S1) were extracted according to ref. (*29*) in clean room facilities at the UCLA. These facilities are pressurized reducing air-influx, they are physically separated from laboratories in which modern DNA is treated and all used laboratory equipment is dedicated for this room. In brief, approximately 50 mg of bone material was powdered using mortar and pestle and subjected to an overnight lysis. DNA from the lysate was subsequently extracted utilizing a silica based method which yielded a 50 µl eluate. Each extraction series consisted of maximal six samples and was complemented by a mock extraction containing water instead of actual sample material.

DNA extracts were sent to the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany for library preparation, DNA capture and high-throughput sequencing. A pool of all extraction blanks was sent alongside the actual samples.

The capture arrays were customized using sequence information from the publically available dog genome (*21*). Mitochondrial genomes and nuclear segments including single nucleotide polymorphisms, diagnostic for dogs and wolves (*5*) were targeted in 60 bp fragments and 10x and 3x tiling density, respectively. We used Agilent's SureSelect (Agilent Technologies, USA) custom-made capture arrays on 14 ancient canids and 20 modern wolves. Although the mt-

genomes were captured and sequenced from the ancient specimens at mostly high coverage (Table 1), the nuclear regions were only retrieved at very low coverage, and genotypes could not be called. This difference likely reflects the higher abundance of mitochondrial than nuclear DNA in ancient specimen as well as the low fraction of endogenous DNA overall in the specimens. Consequently, a capture approach for nuclear segments across all 18 specimens would be logistically challenging, if even possible (but see (*30, 31*)).

Library preparation, hybridization enrichment and sequencing

Libraries were prepared from 15 µl of each ancient extract, a pool of extraction blanks and a H_2O control following the protocol detailed in ref. (*32*) using the modifications for ancient DNA described in ref. (*33*). No uracil-DNA-glycosylase / endonuclease VIII treatment was performed (*34*) to preserve the characteristic damage patterns of ancient DNA. Two sample-specific barcodes were introduced into each library using the double-indexing amplification scheme described in ref. (*33*).

Amplified libraries were pooled in equal mass ratios and subjected to two rounds of hybridization capture using 240k feature SureSelect capture arrays (Agilent) and following the protocol of Hodges et al. (*35*) with the following modifications: (i) different blocking oligos were used (BO4: 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-phosphate-3'; BO6: 5'-CAAGCAGAAGACGGCATACGAGAT-phosphate-3'; BO8: 5'-GTGTAGATCTCGGTGGTCGCCGTATCATT-phosphate-3'; BO10: 5'-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-phosphate-3') to adopt them to the adaptor sequences of double-indexed Illumina multiplex libraries, and (ii) libraries were reamplified after each round of capture using the primer pair IS5 and IS6 (*32*).

Sequencing was performed on a single lane of a flow cell on the Illumina's GA IIx platform. Molecules were sequenced for 75 cycles from both ends, and two 7-cycle index reads were carried out as described (*33*). Raw data processing was performed in the same manner as described for the sequences generated in Tübingen (see below). Table S3 denotes the number of filtered and overlap-merged sequences obtained from each sample. Notably, no sequences were recovered from the extraction or library negative controls and thus no contamination was detected.

Sample processing Tübingen, Germany

Fourteen ancient canid samples from Central Europe, Siberia and Northern America were analyzed in the ancient DNA laboratory in Tübingen (lower panel in Table S1). Seven samples were excavated at the archeological site Kesslerloch, Switzerland and dated to 14,500 ago (*11*). A putative dog like canid originated from the Razboinichy cave in Siberia (*13*) and three samples came from sites in Germany (Bonn-Oberkassel, Kartstein cave, Bedburg-Königshoven) and dated < 14,700 years ago (e.g. (*12, 14*)). The remaining three samples were excavated at archeological sites in Alaska and dated to 21,000-28,000 years ago (re-calibrated from ref. (*16*) using *Calpal*).

Approximately 50 mg bone powder was extracted for each sample in the clean room facilities at the University of Tübingen using the same silica-based method as used at UCLA (*29*). The three Alaskan samples were obtained as DNA extracts from the R. Wayne laboratory at UCLA, where they were extracted as described above. All pre-amplification steps of the library preparation were carried out in the clean room facilities. An Illumina multiplex protocol modified for ancient DNA was used to convert a 20 µl aliquot of each DNA extract into sequencing libraries (*32*). During all steps extraction and library blank controls were carried along and treated according to

the DNA extraction protocol. An amplification with two 'index' primers added sample-specific indexes to both library adapters, thus barcoding individual sample sequences so they could be unambiguously identified subsequent to multiplexed sequencing (*32*). The second individual amplification of each sample was performed in 100 µl reactions containing 5 µl library template, 2 units AccuPrime Pfx DNA polymerase (Invitrogen), 1 unit 10×PCR Mix and 0.3 µM of primer IS5 and IS6 spanning the sequences of the indexed libraries. The thermal profile is as follows: 2 min initial denaturation at 95°C, 3 to 9 cycles consisting of 15 sec denaturation at 95°C, 30 sec annealing at 60°C, 2 min elongation at 68°C and 5 min elongation at 68°C. After amplification, the products were purified using spin columns (Qiagen, Germany) and quantified on an Agilent 2100 Bioanalyzer DNA 1000 chip.

All amplified libraries and controls were enriched for the dog mitochondrial genome (*36*). To obtain bait-DNA for the molecular enrichment long-range products spanning the dog mitochondrial genome (NC_002008.4) were produced (*36*) using the following primer sets: dog2_for: GACAACACCTAATGACC-CACCAAA, dog2_rev: GTGCGTGCTTCATGGCCCTATTCAA, dog1_for: CCGCCAT-CTTCAGCAAACCCTCAAA, dog3_rev: GGATGCTCCTGCATGGGCCAGATT, dog3_for: GCATTCCCCCGAATAAATAACATGAGCTTC, dog1_rev: AGCGGTCATGGGCTTGGGT-TGA.

Purified modern dog DNA was used as DNA template and PCRs were performed using the Roche long-range PCR kit in 100 µl reactions (1 µl DNA template, 1 unit 10× PCR buffer 2, 0.4 mg/mL BSA, 3µl DMSO, 0.125 µM each dNTP, 7 U polymerase and 0.3 mM each primer). The thermal profile was as follows: initial denaturation at 92°C for 2 min followed by 10 cycles consisting of a 10 sec denaturation step at 92°C, annealing at 62°C for 30 sec and 6 min

elongation at 68°C, followed by 30 cycles using the same thermal profile with an additional increase of the elongation time by 20 sec each cycle and a 7 min final elongation at 68°C.

PCR products were purified using spin columns (Qiagen, Germany), quantified by NanoDrop and fragmented to 300 bp using a S220 Covaris machine (Duty cycle 10%, peak incident power (W) 140, cycles per burst 200, time (sec) 120). After ligation to biotinylated adapters the PCR products were immobilized on streptavidin-coated magnetic beads. The amplified libraries from Kesslerloch except K189 were pooled before enrichment as well as the Alaskan samples; all other libraries were enriched separately as previously described (*36*). The library molecules were eluted by NaOH melting after 48 h incubation at 65°C and quantified using qPCR (Lightcycler 480, Roche). After a final amplification for 13 to 16 cycles the libraries were quantified by an Agilent 2100 Bioanalyzer DNA 1000 chip.

The sequencing was carried out on the Illumina MiSeq platform by 2×150+8 +8 cycles using the MiSeq reagent kit v2 and the manufacturers' protocol for multiplex sequencing. The raw reads were called by the base caller *Ibis 1.1.1* after an alignment to the PhiX reference sequence to obtain a training data set for the base caller (*37*). Then they were filtered according to the individual indices and adapter and index sequences were removed. The paired end reads overlapping for at least 11 nucleotides were fused to one read considering only the base with the higher quality score at each position (*38*).

Modern samples

All modern samples were extracted at UCLA. The extraction facilities are physically separated by five floors from the ancient DNA laboratory. DNA was extracted following either standard Phenol/Chloroform procedures (*39*) or using Qiagen's DNeasy Blood and Tissue Kit (Qiagen, USA). DNAs of one set of modern wolves (2.18-18.10 µg in total) were shipped to the

laboratories of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany. There, DNA libraries were prepared following the protocol of Meyer and Kircher (*32*). As with the ancient samples, libraries were double-barcoded by amplification with unique pairs of indexing primers. A single round of multiplex hybridization enrichment was performed using the same experimental setup as described for the ancient samples. Sequencing was performed on the same run with the ancient samples as described above.

Another set of 21 wolves and 1 coyote was prepared for sequencing on a 454 GS FLX+ system (Roche, USA) at the UCLA Genotyping and Sequencing Core facility. Prior to sequencing, we amplified the complete mitochondrial genome in two overlapping segments by means of long range PCRs employing the Expand Long Range dNTPack (Roche, USA). 150 ng of genomic DNA was used in a 50 µl PCR consisting of 1x PCR buffer including Magnesium, 0.125 µM each dNTP, 3.5 U Expand Long Range Enzyme mix and 0.3 mM each primer (sequences available upon request from OT). Each individual mitochondrial segment was subjected to amplification under the following conditions: initial denaturation at 92 °C for 2 min; followed by 10 cycles each consisting of denaturation at 92 °C for 10 sec, primer annealing at 57 °C for 15 sec and elongation at 68 °C for 10 min; 27 cycles of denaturation at 92 °C for 10 sec, annealing at 57 °C for 15 sec and elongation at 68 °C for 10 min with an additional 20 sec per cycle. The amplification was finished with a final elongation at 68 °C for 7 min and a cooling step at 8 °C. PCR success was subsequently assessed under UV-light using an ethidium-bromide stained 1% agarose gel. In order to prepare individually barcoded sequencing libraries we followed the procedure suggested in ref. (*40*). PCR products were purified using Agencourt AMPure SPRI beads (Beckman Coulter, USA), quantified (NanoDrop, Thermo Scientific, USA) and the two long-range PCR segments pooled in equimolar ratios and the pool was sheared to the desired

fragment length by nebulization according to 454 GS FLX+ library preparation manual (Roche, USA). Since we prepared the samples for several libraries to be sequenced on a 454 GS FLX and a 454 GS FLX Titanium (Roche, USA) the protocols needed to be slightly adjusted to accommodate the increased fragment size of the latter platform. In general, we added an individual barcode to each canid sample by performing the following steps: blunt-end repair; ligation of adapters, which contain the barcode and the sequences necessary for successful sequencing on the 454 GS FLX+ machine; adapter fill in reaction; single library quantification using a Pico-green assay; pooling of the respective barcoded libraries; dephosphorylation and restriction digestion followed by a final small fragment removal. One µl of pooled libraries was quantified (*41*) on a LightCycler 480 (Roche, USA) using the following High Resolution Melting Master (Roche, USA) protocol for a 20 µl reaction: 1x MasterMix (dNTPs, polymerase, reaction buffer and HRM dye), 0.2 µM of each primer, 4.375 mM $MgCl_2$, and additional ddH_2O. The real time PCR was run as follows: pre-incubation at 95 °C for 10 min, 45 amplification cycles each with denaturation at 94 °C for 10 sec, annealing at 60 °C for 15 sec and elongation at 72 °C for 25 sec. The melting reaction had the following steps: 95 °C for 1 min, 40 °C for 1 min, 65-95 °C for 1 sec and cooling at 40 °C. The quantified libraries were subsequently processed according to 454 GS FLX+ manuals and sequenced on fractions ($1/16^{th}$) of full picotitre sequencing plates. Raw sequencing reads were adapter trimmed and filtered on the machine according to default 454 GS FLX+ parameters, de-tagged as described in ref. (*40*) and thus prepared for subsequent data-processing.

Data processing

Sequence assembly

Ancient samples

The raw Illumina data were de-multiplexed using both indices. Only reads matching the bar-codes on both ends were used for further analysis. Forward and reverse reads were merged based on overlapping sequence to generate a single, intact read from each read pair. Read pairs that could not be merged in this way were not analyzed further.

For each sample, the merged reads were assembled using the following protocol. First, all reads were used to generate a preliminary mtDNA assembly using the reference dog mtDNA (NC_002008.4.fa) to seed a reference guided assembly using *mia* (*38*). Separately, a database of nuclear-mitochondrial insertion sequences (numts) was constructed by:

1. Constructing a blat alignment of the reference dog mtDNA sequence to the reference dog nuclear genome sequence.
2. Make a list of all hits >= 75% identity.
3. Combine regions that are within one mtDNA genome distance of one another
4. Using these coordinates, extract the corresponding dog nuclear genome sequence as a database of likely dog numt sequence.

In this way, 119 regions were extracted. The reads from each sample were then aligned, using the same *mia* command to each region. Then, for each read, all alignments against the preliminary mtDNA assembly and the putative numts were compared. Reads with a higher scoring alignment against a numt sequence were then excluded from a second round of assembly.

The command used for *mia* assembly was:

```
mia –r NC_002008.4.fa –f SAMPLE_READS.fa –c –I –s ANC_SUB_MAT –m OUTPUT.maln
–I NUMT_IDS.txt
```

Where SAMPLE_READS.fa are the merged reads, ANC_SUB_MAT is the default scoring matrix for aligning ancient DNA sequence data and NUMT_IDS.txt is the list of read IDs to be excluded because they are likely numt sequence as described above.

To mitigate the effect of deamination and sequencing error, the output assemblies were filtered to include only sites with at least 2-fold coverage. Default mia consensus calling was used.

Modern samples

Filtered and de-tagged reads were assembled into complete mt-genomes using two strategies. First, we employed the iterative mapping approach using a modified version of *mia* (*38*). This version was adjusted to handle long 454 reads of modern DNA origin and yielded a consensus sequence of all reads mapped against the reference mt-genome of a wolf (AM711902 (*42*)). The second strategy used 454's default programs *gsAssembler* for a *de-novo* assembly of the mt-genome and *gsMapper*, which was used to map reads against the reference mt-genome of the wolf. Both programs were used with default parameters. Only consensus sequences that had a minimum average coverage of 10-fold were further used and for consensus sites with ambiguities the respective assembly files were inspected by eye and a majority rule adopted. Sites with less than 2 fold coverage were masked out. The primer binding sites of the PCR-amplified mt-genomes were also evaluated manually and consensus nucleotides were retained from the respective reverse strands.

For two samples, Basenji and Dingo, mtDNA sequences were constructed from Illumina HiSeq paired end data. Sequence alignment and genotyping followed a previously implemented pipeline (see http://arxiv.org/abs/1305.7390.), modified to account for particular features of mitochondrial chromosome. Specifically, we did not impose a depth of coverage filter, as mtDNA depth of coverage is extremely high in next-generation sequencing experiments; for our

samples, coverage was several hundred fold per position or greater. Second, for the small number of sites at which genotypes were called as heterozygotes, we retained the higher frequency allele in the reconstructed mtDNA sequence. For two regions in which local alignment was ambiguous around called indels, we resolved ambiguities by performing multiple sequence alignments across a panel of 16 wolves and dog sequences constructed similarly from HiSeq data, as well as the reference genome (CanFam2, (*21*)) using the clustal W2 online server (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

Multispecies alignment

The mitochondrial genomes of 18 ancient canids were joined with complete mt-genomes of 77 modern dogs, 49 modern wolves and four coyotes. Newly generated sequences have been deposited at the publicly available database NCBI under the accession numbers: KF661036-KF661096 and a complete alignment is available as a Supplementary data file. The sequences were aligned using the software *mafft* applying the option that incorporates local pairwise alignment information (*linsi*) (*43*). Alignment ambiguities, which mostly affected the placement of insertion and deletions were inspected manually and corrected to minimize gaps by using *Bioedit v7* (*44*). A repeat unit variable in length (approximately 310 bp) in the control region of canids (*45*) was deleted from the complete multispecies alignment. In order to account for the different compartments of the mitochondrial genome and assign specific parameters to those in subsequent analyses, we extracted the following compartments and created a respective concatenated input file accordingly: tRNAs, rRNAs, protein coding genes and control region. With regard to the protein coding genes, we excluded overlapping segments and incomplete stop codons. Since *ND6* is coded on the L-strand we used the reverse complement of this gene in

order to keep the codons aligned correctly. Codon positions were separated into 1st , 2nd and 3rd codon bins.

Analyses

Phylogenetic reconstruction

In order to assess the phylogenetic informativeness of the multispecies alignment of all 148 canids we ran the likelihood mapping analysis in *TREE-PUZZLE v5.2 (46),* assuming the HKY substitution model (*47*) allowing for rate heterogeneity and with five gamma rate categories. Utilizing quartet topology weighting, this method provides a powerful tool to evaluate if sequence evolution occurred in a star-shape fashion or resulted in a completely resolved tree. We evaluated the complete dataset with 1,000,000 quartets. Whenever assigning the sequences into groups, this method can be used to investigate the support for the respective arrangement of the topology.

In order to infer the phylogenetic relationships of all canids we used a maximum likelihood approach implemented in *IQ-TREE* (*48*) and assessed the statistical support with 10,000 bootstrap steps. Alignment gaps and ambiguous positions coded as "N" were retained in the analysis but treated as missing data and did not contribute to the tree likelihood. In order to maximize the informativeness used in the analyses, we assigned the most likely substitution models, which we tested *apriori* using *jModelTest 2* (*49*) and inferred by the Akaike Information Criterion (AIC) (*50*) for each of the pre-defined compartments (see above).

Additional phylogenetic analyses were carried out in the program *MrBayes v3.2* (*51*) and executed through the University of Oslo Bioportal web portal (https://www.bioportal.uio.no). Similar to the maximum likelihood analysis missing data were coded as uninformative and did

not contribute to the analysis. The data were again partitioned into compartments, which were assigned independent models of molecular evolution. Various possible models of molecular evolution were sampled for each subset during the analysis by taking advantage of the model-jumping feature of *MrBayes v3.2* (*51*) through the command "*lset applyto = (all) nucmodel=4by4 nst=mixed rates=gamma*". Two independent MCMC analyses with four simultaneous chains (one cold and three heated) for each analysis were run for 10 million generations and the sampling of trees and parameters was set to every 1,000 generations. Convergence of the two runs was determined by the stationary distribution plot of the log likelihood values against number of generations and confirmed by the average standard deviation of split frequencies which in all the cases were lower than 0.05. We discarded the first 2,500,000 generations as burn-in and trees were summarized under the 50 percent majority rule method. We also tested whether constraining all dog haplotypes to be monophyletic resulted in a decisively worse likelihood based on Bayes Factors. We estimated the marginal likelihoods using the stepping-stone sampling method (*52*) implemented in *MrBayes v3.2* (*51*) with the number of generations increased to 20 million. The value of alpha was set to 0.4 and the number of steps was 50, which are default values.

Constraining all dog haplotypes to be monophyletic resulted in a decisively worse phylogenetic hypothesis than the unconstrained analyses gave: Stepping-stone Sampled Marginal Likelihood Estimate (unconstrained) = -38871.31, Stepping-stone Sampled Marginal Likelihood Estimate (Dogs monphyletic) = -39289.01, ln Bayes Factor = 835.4, log Bayes Factor = 362.8.

With regard to potentially deaminated sites; no special analytical treatment was applied to the phylogenetic analysis and when inspecting the branch lengths derived from sequences generated

from fossil or modern specimens, no obvious pattern relating to ambiguities in the sequences was detected (see Fig. S9).

Demographic inferences

To infer differences in the demographic histories of wolves and dogs summary statistics were calculated using *DnaSP v5* (*53*) (Table S4). We excluded two aberrant modern wolf sequences from this analysis (China1 and China3, see Table S2) since their phylogenetic positioning suggests only a distant relationship to all extant gray wolves and their taxonomic classification as a member of *Canis lupus* or a separate sub-species is a matter of debate (*54*) but not the focus of this study.

Estimating demographic change through time

Summary statistics calculated for the data analyzed here (Table S4) indicate that a simple model of constant population size is inappropriate to describe the coalescent process over the phylogenetic trees. For the purposes of inferring genealogies and exploring the demographic histories of dogs and wolves, we therefore assumed the flexible Bayesian skygrid model (*22*) as implemented in a development version of *BEAST v1.7.6* (*55*). This model has recently been shown to be the most accurate of the available "skyX" models, in particular when sequences from different temporal intervals are included in the analysis (*22*).

To estimate an evolutionary rate for the canid data set including all analyzed dogs and wolves, we performed a *BEAST* analysis with all complete mitochondrial genomes for which >50% of the sites in the mitochondrial genome were called as >2X (see section above). Mitochondrial genomes were partitioned into four concatenated categories based on their biological properties: tRNAs, rRNAs, coding regions (all gene regions), and the control region. The best fitting evolutionary model for tRNAs, rRNAs and the control region were selected

using the AIC (*50*) as implemented in *jModelTest 2* (*49*). Models were selected as follows: tRNAs: HKY+I; rRNAs: TN93+G; control region: HKY+G+I. For the coding region, we assumed the SDR06 model (*56*). Separate evolutionary rate parameters were estimated for each partition, however all four partitions informed the same tree (as the mitochondrion is a single, non-recombining locus). Mean, calibrated radiocarbon ages or stratigraphically assigned ages were used as priors for all ancient and historic samples (Table S1). We initially ran the model with both the uncorrelated log-normal relaxed molecular clock (UCLD (*57*)), and the strict molecular clock, however comparison of posterior estimates of the parameters of the UCLD model suggested this was an unnecessary over parameterization for these data, and the strict molecular clock was assumed in all subsequent analyses. For all *BEAST* runs, model parameters and trees were sampled every 6000 iterations over 60 million iterations in two independent MCMC chains. The first 10% of iterations were discarded as burn-in, and the remainders of the two chains were combined. We evaluated chain convergence to stationarity for all model parameters using *Tracer v1.5* (*58*). Maximum Clade Credibility (MCC) trees were summarized using *TreeAnnotator v1.7x* (available as part of the *BEAST* package) and visualized and manipulated in *FigTree v1.4* (http://tree.bio.ed.ac.uk/software/figtree/). For the demographic inference, we assumed 100 population size intervals and a cut-off of 80,000 years (other cutoff values and number of grid points were evaluated, with no observable effect on the resulting plot). For each analysis, we visualized the results of the skygrid analysis using an R script provided by Marc Suchard (UCLA).

To infer the recent demographic history for dogs, we performed subsequent analyses on the dogs comprising clade A and its associated ancient canids from the New World, as this is the most diverse and oldest dog clade. By restricting the analysis to clade A, we aimed to limit the

confounding influence of including both dogs and wolves in estimates of recent demographic history. For this analysis, we assumed evolutionary and coalescent parameters as described above, except the evolutionary rate was fixed to the mean rate estimated for each partition in the analysis of the full data set. MCMC chains and MCC trees were run and evaluated as described above.

Supplementary Results

Descriptive analyses of the sequences

A close inspection of the raw reads revealed a pattern typical for sequences generated from ancient materials (*59, 60*). Because we used only reads whose forward and reverse reads could be merged based on overlapping sequence, there is an effective upper limit on the length of DNA sequence fragments of 140 base pairs. Nevertheless, in the first panel of ancient sequences, there is a general correlation between the age of the sample and the average length of recovered mtDNA sequences (Fig. S1; top panel). The two exceptions to the trend are the USA *8,500* sample and the Russia *15,000* sample, which might indicate poor preservation. In contrast, the second panel of ancient canids showed little correlation between age and length (Fig. S1; bottom panel). This may be due to the narrower time range of this panel (12,500 to 33,500 years) compared to the first panel (1,000 to 36,000 years) and to the presence of permafrost samples from Alaska whose average fragment lengths were longer.

Another typical feature of ancient DNA is the rate of cytosine deamination (*61*), particularly at the edges of recovered DNA fragments (*34, 62*). We measured the rate of cytosine deamination at the first, most 5-prime position of each read by counting how often a read's first base aligned against a C in the consensus sequence and was observed in the read as a T, as deaminated cytosines are read as thymidines. As shown in Figure S2, we see a typical pattern of elevated cytosine deamination, consistent with previous analyses of ancient DNA datasets. We find that there is no obvious relationship in these data between the rate of cytosine deamination and age of the sample. However, we note that the permafrost samples have some of the lowest observed rates of cytosine deamination.

Relationship to previously defined dog clades

Since the publication of the first large-scale, genetic analysis of modern dogs and wolves (*4*), attempts have been made to identify phylogenetic clusters prevalent in tree reconstructions of canids. Whereas the close phylogenetic arrangement of dogs and wolves undoubtedly supported a recent common ancestry of these two canids (*4*), a further distinction into monophyletic dog clades representing particular breeds or wolf clusters resembling their spatial distribution was hampered by weak statistical support. Consequently, these early attempts based on short fragments of the mitochondrial genome should be interpreted with caution and can at best be considered as poorly supported clusters. Not surprisingly, subsequent genetic studies on dogs and wolves that merely included more taxa (in particular more dogs) but did not significantly extend the sequence length, failed to provide additional statistical support for particular clusters and resulted in the dissolution of some and the addition of other clusters (*6, 63*). For example, the initial four defined dog clusters (I-IV in (*4*)) were later accompanied by another two (A-F in (*6*)) based on double the amount of sequence data (582 bp) and more than fourfold the number of sequenced dogs (654). Interestingly, Savolainen and colleagues acknowledged the weak statistical support of their dog clusters (Fig. 1 in (*6*)) but nevertheless used this arbitrary assignment as a proxy for phylogenetic interpretation, which surprisingly became commonly accepted.

The first study presenting strong statistical support for any dog clade was based on complete mitochondrial genomes sequenced on 14 dogs, six wolves and three coyotes (*64*). The four modern dog clades presented by Bjoernerfeld and colleagues were later recovered by Pang and colleagues (2009). The phylogenetic arrangement in the latter study was however only statistically supported by considering sequence information from the complete mitochondrial

genome (Fig. 2a in (*63*)). In agreement with the studies of Bjoernerfled and Pang, we provide further evidence for the existence of only four, statistically well-supported phylogenetic clades apparent in complete mitochondrial genome sequences of modern dogs. Aside for the use of a large number of dogs, we also include an unprecedented number of wolf genomes (49) covering the species' global distribution. We derive highly similar topologies with high statistical support values regardless of the method applied: maximum likelihood, coalescence or Bayesian based approaches. Intriguingly, repeating the phylogenetic analyses with exclusively 959 bp of the control region (Fig. S3) not only reduces the support values for nodes, but associates sequences such as the Belgian or the Altai specimens with modern domestic dogs, a result not supported in the complete mtDNA genome tree.

Consequently, although we can't exclude the possibility that other dog clades will emerge due to the sequencing of additional samples, our sampling has captured most of the phylogenetic diversity prevalent in modern dogs including young as well as divergent breeds. Confirmation of additional clades, such as clades E and F in Pang et al. (2009) may require further complete genome sequencing of additional dogs, but will not affect the conclusions presented in this paper.

Contamination of Belgium *36,000* with DNA from cow

Extraction of ancient specimens requires following strict rules in order to avoid contamination with contemporary DNA (*65*). This is especially worrisome when working with hominoids (*59, 66-69*) or ancient remains that are closely related to domesticated animals common in human environments such as dogs or cats. In order to minimize the contamination risk in our study we followed all relevant guidelines for working with ancient materials. The short read lengths in

conjunction with the phylogenetic analyses suggest that none of our samples is affected by modern DNA contamination.

However, we noticed an increased read coverage localized in small regions of the first few kilobases of the mt-genome of one of our ancient specimens (Belgium *36,000*). Alignment of these reads revealed that many were highly similar to bovine mtDNA. We subsequently re-aligned all reads from this sample and a similar, but unaffected sample (Belgium *30,000*) against the complete mitochondrial genomes of the dog and the cow. We compared the alignment scores of each read against the cow and dog mtDNA. Reads that failed to produce a positively scoring alignment had their scores arbitrarily set to 0. These results are shown in Figure S4.

Two sources of bovine DNA are imaginable: first, the use of BSA (Bovine Serum Albumin) in the subsequent processes of sequencing library preparation might have introduced such contaminants (*70*). Second and more plausible, this particular sample might have been contaminated by a treatment with glue that was based on cow collagen applied to the specimen during the 19th century. However, it remains unclear why only a particular sample and certain regions of the mitochondrial genome are affected and further investigations would be required to resolve this issue.

Likelihood mapping analysis

Likelihood mapping analyses (*71*) clearly suggest high phylogenetic informativeness of our data. A cumulative percentage of 92.3% of all possible quartets reveals a well-supported phylogeny, only 5.3% suggest a star-shape phylogeny and only 2.4% of the quartets remain unresolved (Fig. S5). It is worth noting that two (Belgium *36,000* and Alaska *21,000*) out of the 148 samples analyzed failed a chi-square test (p=3.56% and 2.07% respectively) comparing the nucleotide

composition of the respective sequences with a frequency distribution assumed in the maximum likelihood model. However, we did not attribute importance to this finding since we adopted a significance level of 1% and an in-depth inspection of the base composition did not reveal any apparent abnormalities of these two mt-genomes when compared to other ancient specimens or the proportions of nucleotides averaged over all modern mt-genomes (Fig. S6).

Altai dog mtDNA haplotype

We previously reported a phylogenetic analysis of the control region sequence from the putative dog from the Razboinichya cave (*13*), concluding that this canid's mtDNA is more closely related to modern dogs rather than to modern wolves (*72*). The more comprehensive analysis of the complete mtDNA presented here, however, strongly suggests a position at the root of a clade uniting two ancient wolf genomes, two modern wolves, as well as two dogs of Scandinavian origin (Fig. 1). An unambiguous delineation of the phylogenetic position of this Altai dog mtDNA as either dog or wolf is inconclusive, with 21.6% of the quartets in a maximum likelihood mapping analysis supporting a clustering with modern dogs and 21.1% a clustering with modern wolves (Fig. S7). Interestingly, the highest proportion of quartets supports a topology of the Altai dog clustering with other ancient canids to the exclusion of modern dogs and wolves, which is indicative of some shared mutations uniting all ancient canids. Furthermore, constraining the position of this specimen within Dog clade A (*72*) and re-running the phylogenetic analysis yielded a significantly less likely tree ($\log_{10}$ Bayes Factor 92.518). Notably, comparing the 413 nucleotides of the control region generated with conventional Sanger sequencing with the sequence obtained from the capture approach differed in two positions. Both positions are either an A-G or a C-T difference, and in combination with the high

number of molecules covering these sites in the newly generated sequence (Fig. S8), this pattern most likely indicates the amplification of a deaminated molecule in early PCR cycles in ref. (*72*).

Figures:

Fig S1: MtDNA fragment lengths.

The average, plus and minus one standard deviation of the fragment length of sequences in the assemblies is shown. The top panel shows the ancient canid samples in the first set, from youngest (left) to oldest (right). The bottom panel shows the ancient samples from the second set.

Fig S2: Rate of cytosine deamination in ancient canid data.

For each ancient canid sample in the first set (top panel) and second set (bottom panel) we counted the rate of observing a T in the first read position at which the mtDNA consensus base is a C.

Fig S3: Phylogenetic reconstruction generated employing a Bayesian approach and reducing the sequences to exclusively the mitochondrial control region.

The color code is adopted from Fig. 1 and values at the nodes indicate the respective statistical support. Highlighted are the four dog clades obtained from the analyses using the complete mt-genomes.

Fig S4: Alignment score of reads versus cow and dog mtDNA.

The Belgium *36,000* sample (top panel) was found to contain many reads with a better alignment score to cow mtDNA. On average, these reads were longer than those that align with better score against the dog mtDNA genome. In contrast, the Belgium *30,000* sample shows the typical pattern wherein reads align with high score against the dog mtDNA genome and poorly or not at all against the cow mtDNA genome.

Fig S5: Maximum Likelihood mapping of the complete 148 canid mt-genomes.

The fraction of quartets occupying the three corners support resolved, bi-furcating topologies, while the fraction in the center part of the triangle indicates star-shaped topology.

Fig S6: Nucleotide composition of each ancient specimen and the average of modern canids.

Fig S7: Maximum Likelihood mapping analysis of different phylogenetic arrangements of the Altai dog.

Each of the three corners show the fraction of quartets (1,000,000) supporting the respective phylogenetic arrangement, as depicted next the corners.

Fig S8: Coverage plot of the captured mt-genome of the Altai dog.

A) Base coverage generated from sequencing the captured mt-genome spanning the complete fragment previously generated by Sanger sequencing (*72*), plotted with number of reads per position. B) Number of reads spanning the two positions differing between the captured and the previously published sequences (*72*). In addition the red arrow highlights the coverage at the exact position as well as the number of reads sharing the previously sequenced nucleotide is indicated.

Fig S9: Consensus maximum likelihood tree representing the phylogenetic arrangement of all 148 canids.

For visibility reasons, the branches leading to the outgroup (coyotes) were truncated. The statistical support was evaluated with 10,000 bootstrap steps and is indicated at each node. The color pattern was adopted from Fig. 1 and the exact breed assignments or geographical localities of the specimens can be found in Table S1, S2.

Fig S10: Phylogenetic arrangement of modern and ancient dogs (blue) and wolf sequences (orange) as obtained from coalescent based, maximum likelihood and Bayesian methods.

The figure is identical to Fig. 1, except that 95% HPD's for the respective node ages are indicated as green bars.

Table S1: Ancient samples used in this study.
Upper section indicates samples of the set captured with Agilent technology and lower section shows samples enriched for mtDNA using custom designed biotinylated adapters (See Supplementary Information for more details). Gray and italic font indicates insufficient mitochondrial data.

ID	Country	Location	approximate age*	Morphological classification†	Accession Number	Museum ID, sample
Belgium *26,000*	Belgium	Trou des Nutons	26,000	Wolf-like (*2*)	KF661078	RBINS 2559, skull
Belgium *36,000*	Belgium	Goyet niveau 4	36,000	*Dog-like* (*2, 9, 27*)	KF661079	RBINS 2860, skull
Belgium *30,000*	Belgium	Goyet niveau 4	30,000	Wolf-like	KF661080	RBINS 2860-2, mandible
Russia *18,000*	Russia	Medvezya cave	18,000	Wolf-like	KF661081	RAS MS2
Russia *15,000*	Russia	Eliseevichi	15,000	Dog-like (*8*)	KF661082	RAS JAL4502, incisor
USA *8,500*	USA	Koster site, Illinois	8,500	Dog-like (*28*)	KF661083	
Argentina *1,000*	Argentina	Cerro Lutz (near Uruguay, Brasil)	1,000	Dog-like (*73*)	KF661084	
Russia *22,000*	Russia	Kostenki 4	22,000	Wolf-like	KF661085	tooth
USA *1,000*	USA	Florida	1,000	Dog-like (*74*)	KF661086	
ac4	*Belgium*	*Trou de Bailleux*	*postglacial*	*Wolf-like (2)*		*T.d.B.1, skull*
ac6	*Russia*	*Verholenskya Gora*	*13,000*	*Dog-like*		
ac9	*Israel*	*canids from Isreal*	*3,000*	*Dog-like*		
aca2	*Russia*	*Kostenki*	*25,500*	*Wolf-like*		
aca4	*Israel*	*canids from Isreal*	*7,000*	*Dog-like*		
Switzerland 1 *14,500*	Switzerland	Kesslerloch cave	14,500	Wolf-like (*11*)	KF661087	60
Alaska *28,000*	Alaska	Eastern Beringia	28,000	Wolf-like (*16*)	KF661088	AMNH FM 68008-6
Alaska *21,000*	Alaska	Eastern Beringia	21,000	Wolf-like (*16*)	KF661089	AMNH FM 67224
Alaska *20,800*	Alaska	Eastern Beringia	20,800	Wolf-like (*16*)	KF661090	AMNH FM 67216
Switzerland 2 *14,500*	Switzerland	Kesslerloch cave	14,500	Wolf-like (*11*)	KF661091	M001/62
Russia *33,500*	Russia	Razboinichya cave	33,500	*Dog-like* (*9, 13*)	KF661092	
Germany *14,700*	Germany	Bonn-Oberkassel	14,700	Dog-like (*12*)	KF661093	
Germany *12,500*	Germany	Kartstein cave	12,500	Dog-like (*20*)	KF661094	
Switzerland 3 *14,500*	Switzerland	Kesslerloch cave	14,500	Wolf-like (*11*)	KF661095	M001/58
IBK63	*Switzerland*	*Kesslerloch*	*14,500*	*Wolf-like (11)*		*M001/63*
JBK59	*Switzerland*	*Kesslerloch*	*14,500*	*Wolf-like (11)*		*59*
JK420	*Germany*	*Bedburg-Königshoven*		*Dog-like*		*420*
MRK54	*Switzerland*	*Kesslerloch*	*14,500*	*Wolf-like (11)*		*M001/54*
SBK189	*Switzerland*	*Kesslerloch*	*14,500*	*Dog-like (11)*		*189*

*in years before present
†Italic font indicates specimens with ambiguous classification

Table S2: Samples used with their respective accession numbers, breed and geographic origin.

Dog A			Dog B			Dog C			Dog D			Eurasian wolves			North American wolves			Coyotes		
ID	**Accession number**	**Breed**	**ID**	**Accession number**	**Breed**	**ID**	**Accession number**	**Breed**	**ID**	**Accession number**	**Breed**	**ID**	**Accession number**	**Origin**	**ID**	**Accession number**	**Origin**	**ID**	**Accession number**	**Origin**
D01	DQ480499	Siberian Husky	**D26**	AY656743	Saint Bernard	**D09**	DQ480489	German Shepherd	**D03**	DQ480492	Jamthund	**China1**	NC011218	China	***Canada1***	*KF661056*	*Canada*	***coyote1***	*KF661096*	*USA*
D04	DQ480495	Cocker Spaniel	**D31**	AY656752	Standard Schnauzer	**D11**	DQ480501	Swedish Elkhound	**D34**	EU408288	NorwegianElkhound	**Mongolia**	GQ374438	Mongolia	***Canada2***	*KF661057*	*Canada*	**coyote2**	DQ480511	USA
D05	DQ480497	West Highland White Terrier	**D32**	AY656745	English Springer Spaniel	**D12**	DQ480493	Black Russian Terrier				**China2**	EU442884	China	**Canada3**	DQ480508	Canada	**coyote3**	DQ480510	USA
D07	DQ480491	Irish Setter	**D33**	AY656740	Kerry Blue Terrier	**D52**	EU408293	Pit Bull Terrier				**Sweden1**	AM711902	Sweden	***Alaska1***	*KF661058*	*Alaska*	**coyote4**	DQ480509	USA
D15	AY656755	Sapsaree	**D35**	EU408269	Doberman Pinscher	**D55**	EU408251	Blue Heeler				**Saudi Arabia1**	DQ480507	Saudi Arabia	***Canada4***	*KF661059*	*Canada*			
D16	AY656753	Irish Setter	**D40**	EU408252	Bolognese	**D60**	EU408267	Cocker Spaniel				**Saudi Arabia2**	DQ480506	Saudi Arabia	***Mexico1***	*KF661060*	*Mexico*			
D18	AY656741	Irish Setter	**D41**	EU408292	Poodle	**D63**	EU408279	Havanese				**Spain**	DQ480505	Spain	***Canada5***	*KF661061*	*Canada*			
D21	AY656754	Chinese Crested	**D42**	EU408260	Cardigan Corgi	**D84**	EU408291	Pomeranian				**Japan**	AB499824	Japan	***Canada6***	*KF661062*	*Canada*			
D22	AY656737	Basenji	**D44**	EU408254	Basset Hound	**ChineseDog1**	Wang2013	Chinese indigenous dog 1				***Finland***	*KF661038*	*Finland*	***Canada7***	*KF661063*	*Canada*			
D25	AY656744	English Springer Spaniel	**D46**	EU408307	Walker Hound							***Russia1***	*KF661039*	*Russia*	***USA1***	*KF661064*	*USA*			
D27	AY656749	Saint Bernard	**D47**	EU408303	unknown							***Sweden2***	*KF661040*	*Sweden*	***Mexico2***	*KF661065*	*Mexico*			
D28	AY656742	Old English Sheepdog	**D48**	EU408247	Australian Terrier							***China3***	*KF661041*	*China*	***Alaska2***	*KF661066*	*Alaska*			
D30	AY656747	Welsh Springer Spaniel	**D49**	EU408255	Basset Hound							***Israel1***	*KF661042*	*Israel*	***Alaska3***	*KF661067*	*Alaska*			
D56	EU408305	Viszla	**D50**	EU408268	Cocker Spaniel							***India****	*KF661043*	*India*	***USA2****	*KF661068*	*USA*			
D59	EU408282	Keeshond	**D51**	EU408258	Cockapoo							***Russia2***	*KF661044*	*Russia*	***USA3****	*KF661069*	*USA*			
D61	EU408300	Tibetan Mastiff	**D98**	FJ817364	Golden Retriever							***Poland1***	*KF661045*	*Poland*	***USA4****	*KF661070*	*USA*			
D65	EU408272	Dachshund	**D99**	FJ817363	Golden Retriever							***Russia3***	*KF661046*	*Russia*	***Alaska4****	*KF661071*	*Alaska*			
D66	EU408270	Dachshund										***Ukraine****	*KF661047*	*Ukraine*	***Alaska5****	*KF661072*	*Alaska*			
D67	EU408261	Chihuahua										***Italy****	*KF661048*	*Italy*	***Alaska6****	*KF661073*	*Alaska*			
D68	EU408246	American Cocker Spaniel										***Poland2****	*KF661049*	*Poland*	***Canada8****	*KF661074*	*Canada*			
D69	EU408304	unknown										***Oman****	*KF661050*	*Oman*	***Canada9****	*KF661075*	*Canada*			
D71	EU408295	Rottweiler										***Iran****	*KF661051*	*Iran*	***Canada10****	*KF661076*	*Canada*			
D72	EU408286	Miniature Dachshund										***Sweden3****	*KF661052*	*Sweden*	***Canada11****	*KF661077*	*Canada*			
D73	EU408249	Australian Shepherd										***China4****	*KF661053*	*China*						
D75	EU408274	English Mastiff										***Croatia****	*KF661054*	*Croatia*						
D78	EU408248	Australian Shepherd										***Israel2****	*KF661055*	*Isreal*						
D79	EU408264	Cairn Terrier																		
D80	EU408294	Pug																		
D81	EU408287	Newfoundland																		
D82	EU408302	Toy Poodle																		
D83	EU408290	Neapolitan Mastiff																		
D85	EU408289	Neapolitan Mastiff																		
D86	EU408263	Cavalier King Charles Spaniel																		
D87	EU408250	Bichon Frise																		
D88	EU408266	Cocker Spaniel																		
D89	EU408273	English Shepherd																		
D90	EU408275	French Bull Dog																		
D91	EU408265	Corgi																		
D93	EU408276	Great Dane																		
D94	EU408245	Akita																		
D95	EU408257	Brittany Spaniel																		
D96	FJ817358	Golden Retriever																		
D97	FJ817362	Golden Retriever																		
D102	FJ817359	Golden Retriever																		
D103	U96639	Sapsaree																		
Basenji	*KF661036*	*Basenji*																		
Dingo	*KF661037*	*Dingo*																		
ChineseDog2	Wang2013	Chinese indigenous dog 2																		
ChineseDog3	Wang2013	Chinese indigenous dog 3																		

Wolf samples that have been used on the capture array are highlighted with an asterisk and samples newly generated are italicized

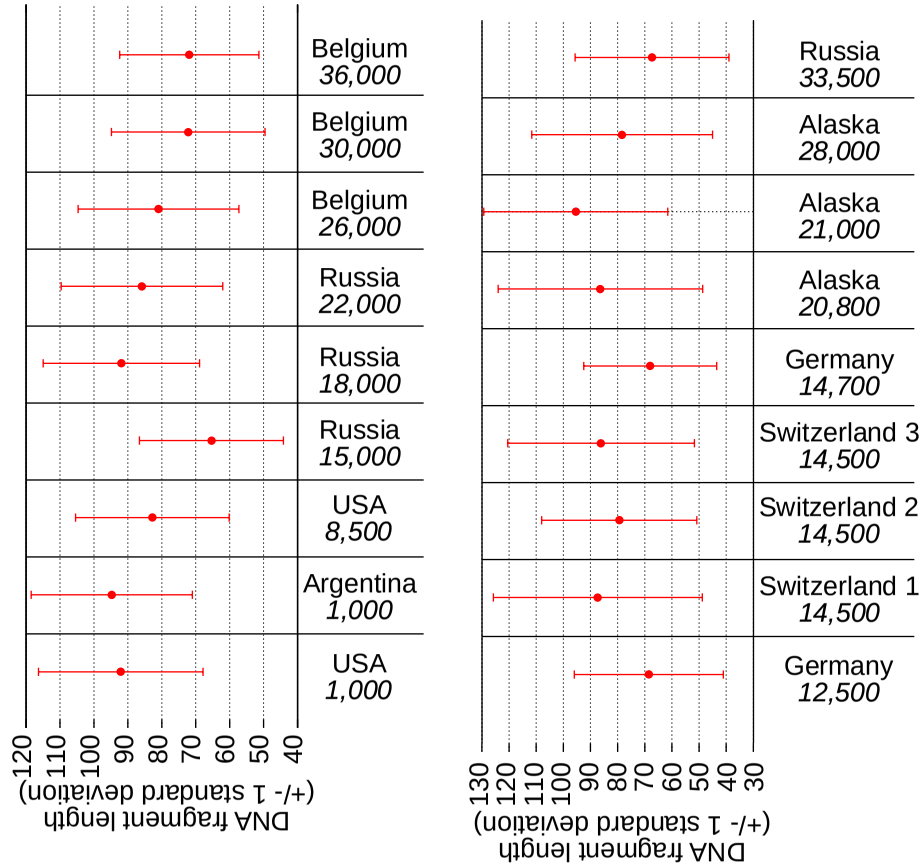
Table S3: Indexed samples and number of respective sequence reads. Samples in gray and italic font did not yield sufficient mitochondrial data.

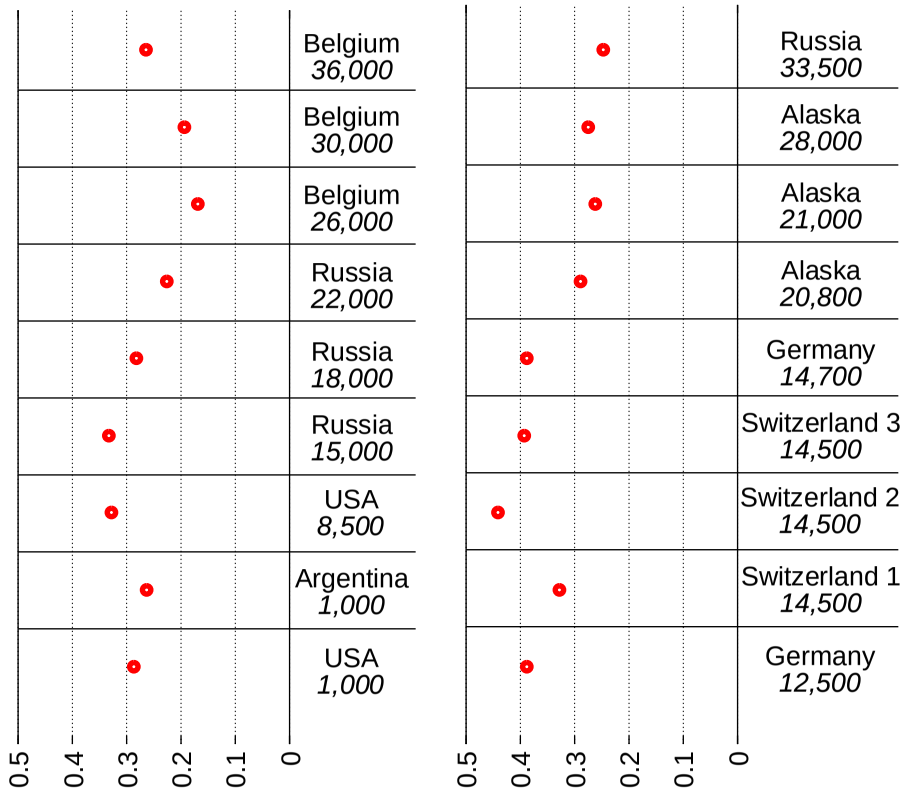
	P7 index	P5 index	#merged sequences
Belgium *26,000*	6	20	1,023,494
Belgium *36,000*	29	14	744,336
Belgium *30,000*	40	15	1,005,792
Russia *18,000*	45	21	5,954,383
Russia *15,000*	50	16	960,724
USA *8,500*	54	35	684,372
Argentina *1,000*	63	18	3,423,856
Russia *22,000*	71	22	1,252,086
USA *1,000*	80	23	2,540,176
ac4	42	33	936,222
ac6	49	34	503,088
ac9	62	17	1,096,127
aca2	72	19	993,509
aca4	91	24	563,538
Extraction blank	95	36	0
Library blank	96	37	0

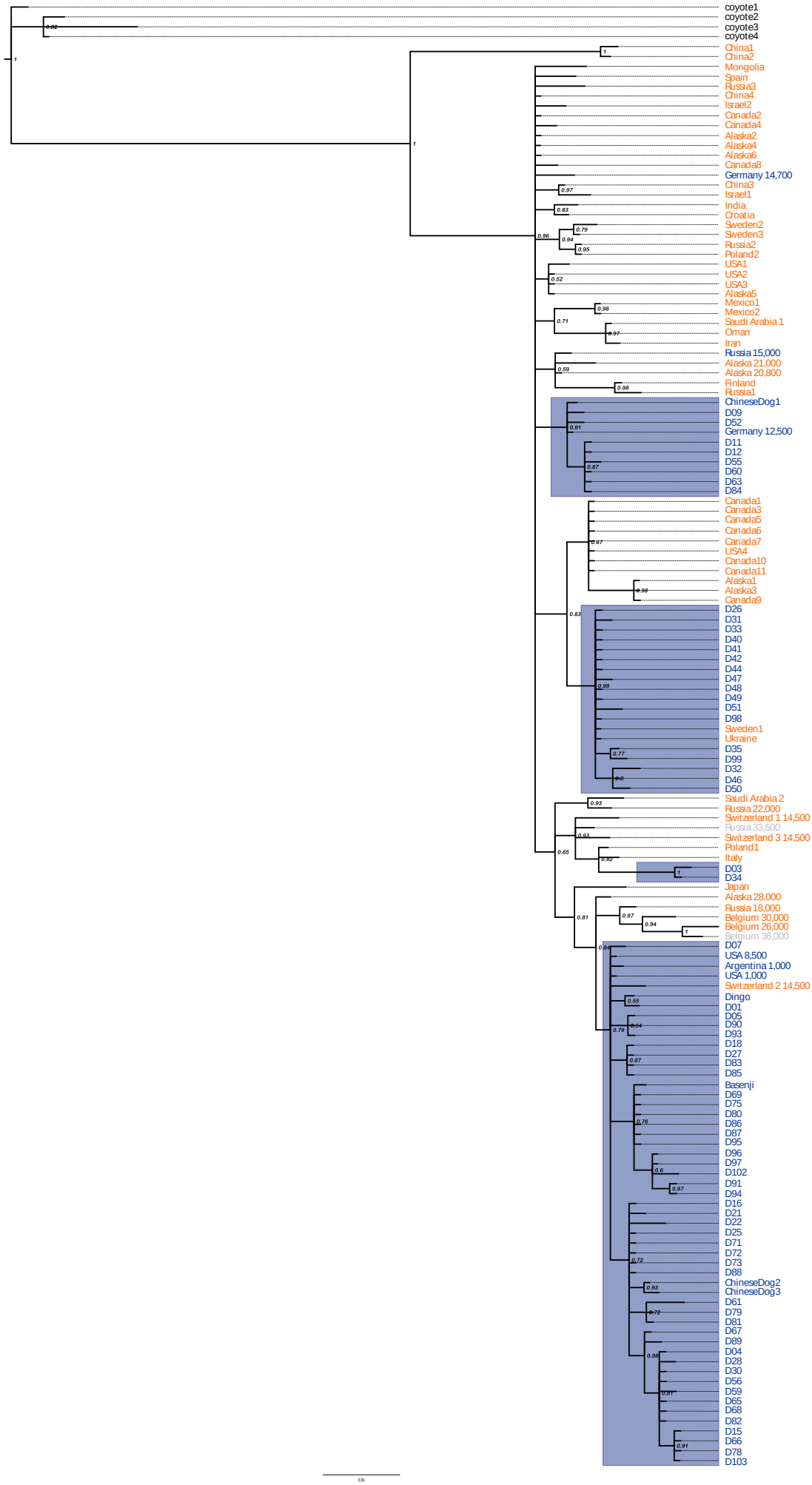
Table S4: Summary Statistics. Values highlighted in bold indicate significance ($p<0.05$).

Partition	# Mt-genomes	# Retained nucleotides*	# Haplotypes	Nucleotide diversity (π%) (*75*)	Nucleotide diversity (θ%) (*76*)	Tajima's D (*77*)	Fu & Li's D (*78*)	Fu & Li's F (*78*)
All canids	142	2,940	77	0.32	1.08	**-2.28**	**-4.78**	**-4.35**
Modern wolves	47	15,279	34	0.31	0.63	**-1.85**	-1.63	-2.05
Modern wolves and ancient wolves	57	7,060	39	0.34	0.9	**-2.27**	**-3.53**	**-3.64**
Eurasian wolves	24	15,836	22	0.397	0.63	-1.49	-1.22	-1.53
North American wolves	23	15,736	15	0.17	0.17	0.09	0.91	0.77
Modern dogs	77	15,759	75	0.33	0.51	-1.21	**-2.38**	-2.27
Modern dogs and ancient dogs†	83	7,317	72	0.3	0.64	**-1.85**	**-4.19**	**-3.85**

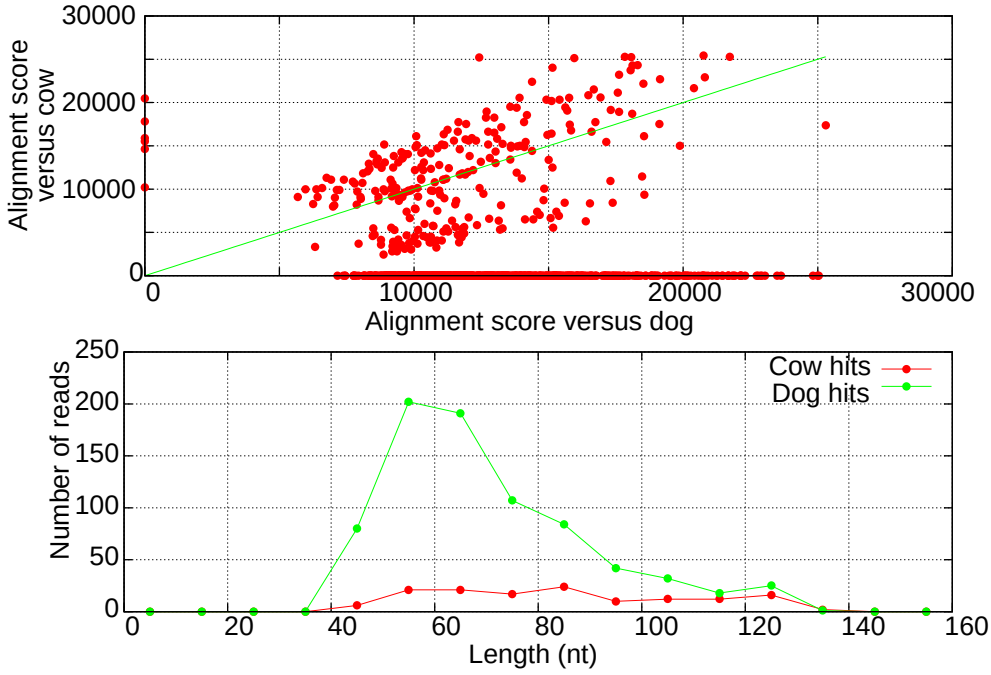
* Excluding sites with gaps and missing data
† Excluding 2 ancient specimens due to ambiguous classification (Belgium *36,000*; Russia *33,500*)


Belgium
36,000
Belgium
30,000
Belgium
26,000
Russia
22,000
Russia
18,000
Russia
15,000
USA
8,500
Argentina
1,000
USA
1,000
120
110
100
90
80
70
60
50
40
DNA fragment length
(+/- 1 standard deviation)
Russia
33,500
Alaska
28,000
Alaska
21,000
Alaska
20,800
Germany
14,700
Switzerland 3
14,500
Switzerland 2
14,500
Switzerland 1
14,500
Germany
12,500
130
120
110
100
90
80
70
60
50
40
30
DNA fragment length
(+/- 1 standard deviation)

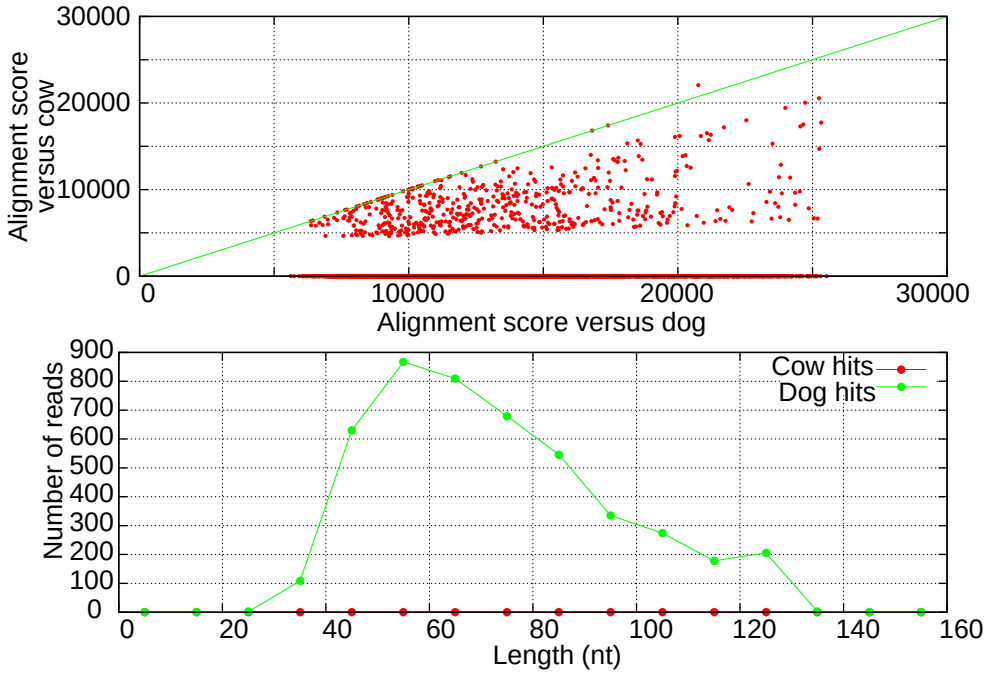

Belgium
36,000
Belgium
30,000
Belgium
26,000
Russia
22,000
Russia
18,000
Russia
15,000
USA
8,500
Argentina
1,000
USA
1,000
Russia
33,500
Alaska
28,000
Alaska
21,000
Alaska
20,800
Germany
14,700
Switzerland 3
14,500
Switzerland 2
14,500
Switzerland 1
14,500
Germany
12,500
0.5
0.4
0.3
0.2
0.1
0
0.5
0.4
0.3
0.2
0.1
0

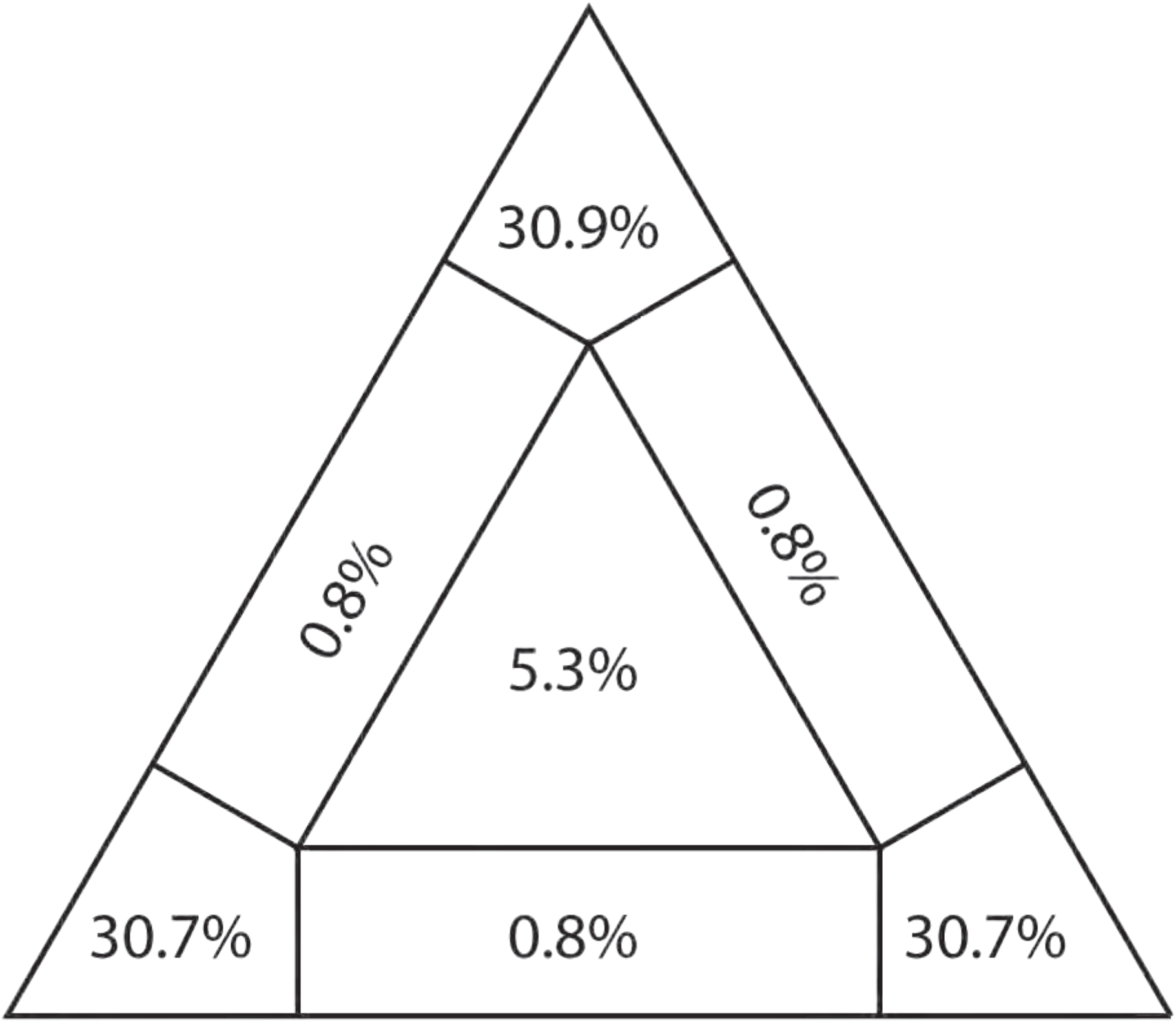

coyote1
coyote2
coyote3
coyote4
China1
China2
Mongolia
Spain
Russia3
China4
Israel2
Canada2
Canada4
Alaska2
Alaska4
Alaska6
Canada8
Germany 14,700
China3
Israel1
India
Croatia
Sweden2
Sweden3
Russia2
Poland2
USA1
USA2
USA3
Alaska5
Mexico1
Mexico2
Saudi Arabia 1
Oman
Iran
Russia 15,000
Alaska 21,000
Alaska 20,800
Finland
Russia1
ChineseDog1
D09
D52
Germany 12,500
D11
D12
D55
D60
D63
D84
Canada1
Canada3
Canada5
Canada6
Canada7
USA4
Canada10
Canada11
Alaska1
Alaska3
Canada9
D26
D31
D33
D40
D41
D42
D44
D47
D48
D49
D51
D98
Sweden1
Ukraine
D35
D99
D32
D46
D50
Saudi Arabia 2
Russia 22,000
Switzerland 1 14,500
Russia 33,500
Switzerland 3 14,500
Poland1
Italy
D03
D34
Japan
Alaska 28,000
Russia 18,000
Belgium 30,000
Belgium 26,000
Belgium 36,000
D07
USA 8,500
Argentina 1,000
USA 1,000
Switzerland 2 14,500
Dingo
D01
D05
D90
D93
D18
D27
D83
D85
Basenji
D69
D75
D80
D86
D87
D95
D96
D97
D102
D91
D94
D16
D21
D22
D25
D71
D72
D73
D88
ChineseDog2
ChineseDog3
D61
D79
D81
D67
D89
D04
D28
D30
D56
D59
D65
D68
D82
D15
D66
D78
D103

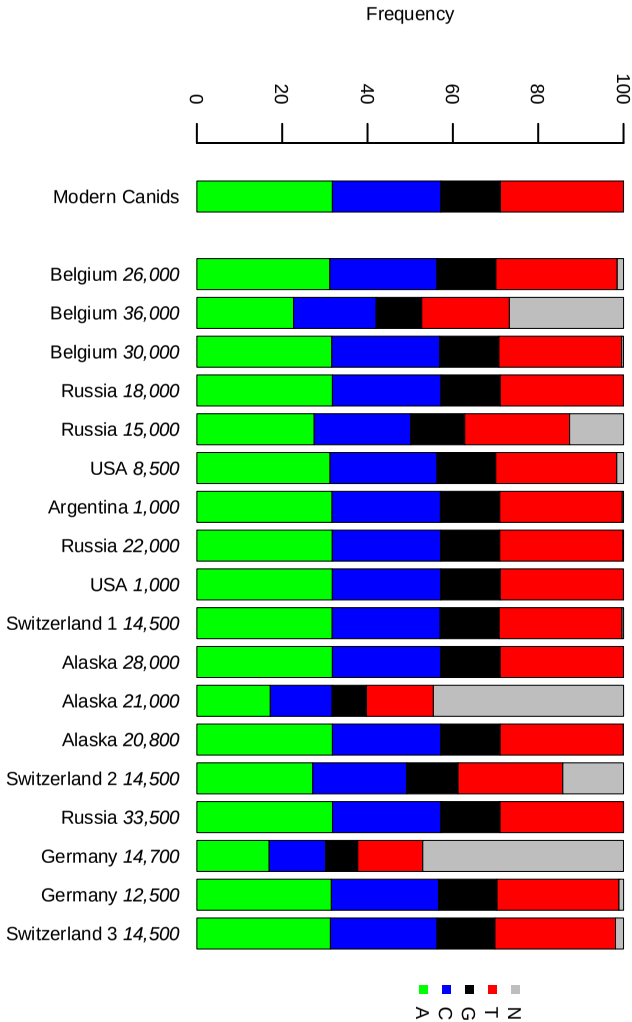
Belgium *36,000*

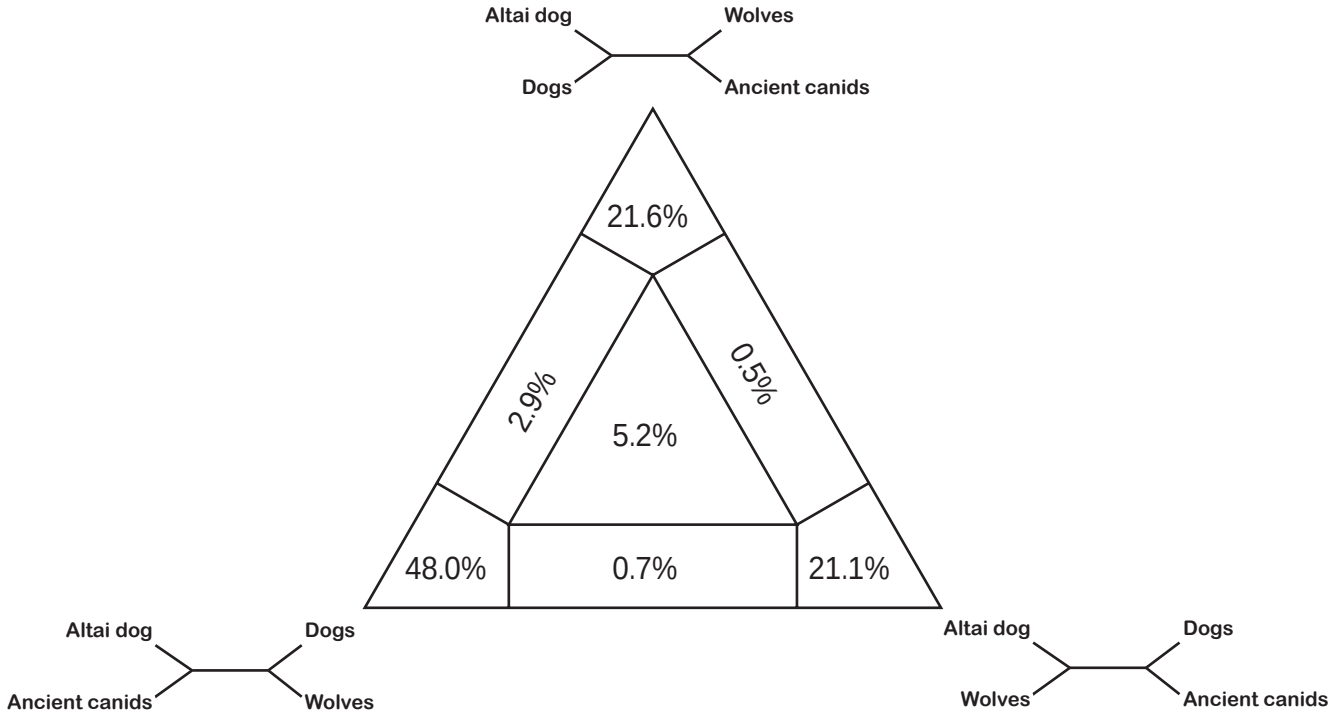


Belgium *30,000*

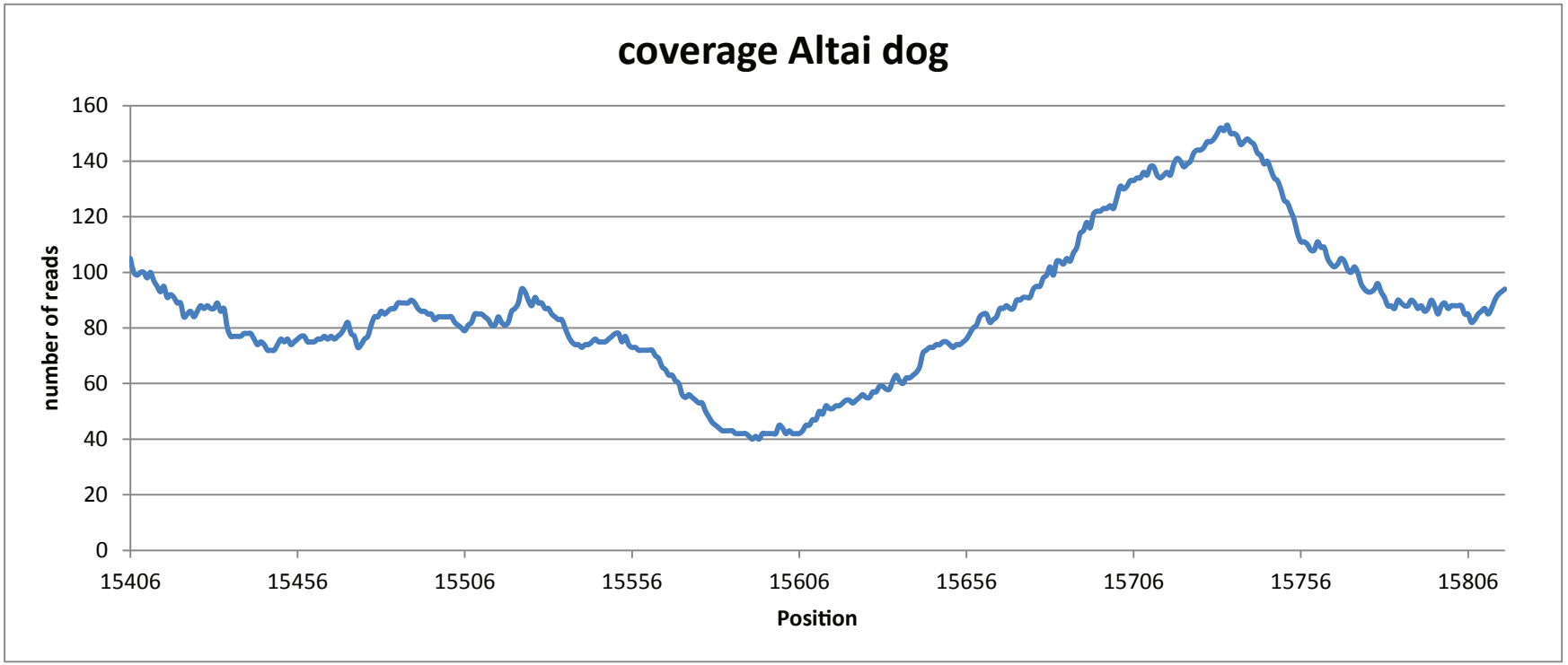



30.9%
0.8%
0.8%
5.3%
30.7%
0.8%
30.7%

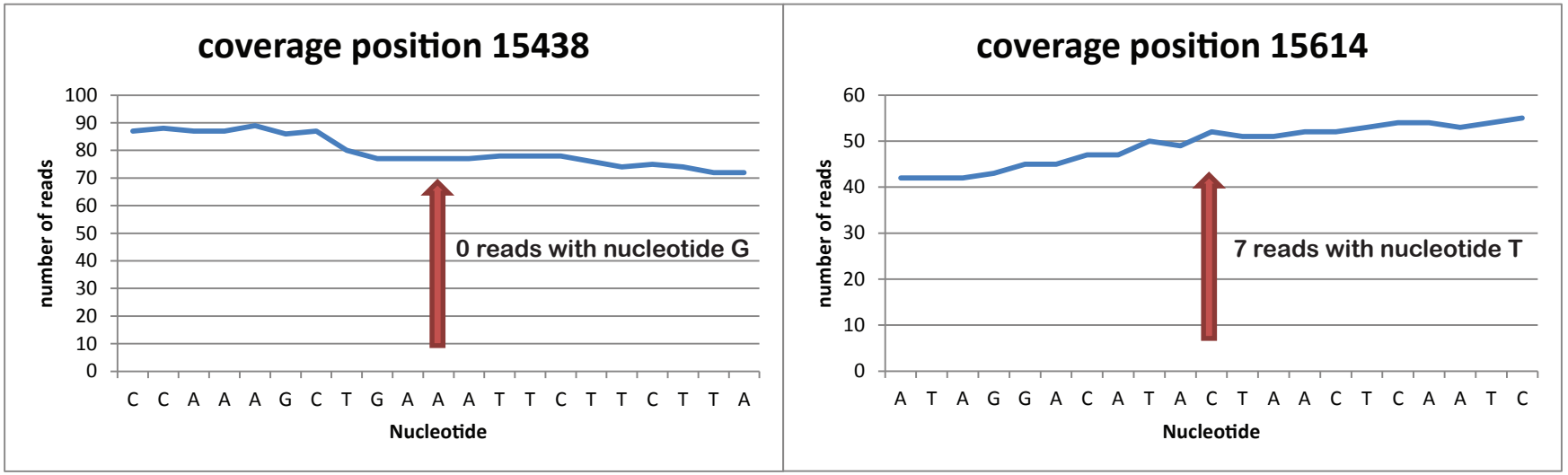

Frequency
0
20
40
60
80
100
Modern Canids
Belgium 26,000
Belgium 36,000
Belgium 30,000
Russia 18,000
Russia 15,000
USA 8,500
Argentina 1,000
Russia 22,000
USA 1,000
Switzerland 1 14,500
Alaska 28,000
Alaska 21,000
Alaska 20,800
Switzerland 2 14,500
Russia 33,500
Germany 14,700
Germany 12,500
Switzerland 3 14,500
A
C
G
T
N


Altai dog
Wolves
Dogs
Ancient canids
21.6%
2.9%
0.5%
5.2%
48.0%
0.7%
21.1%
Altai dog
Dogs
Ancient canids
Wolves
Altai dog
Dogs
Wolves
Ancient canids

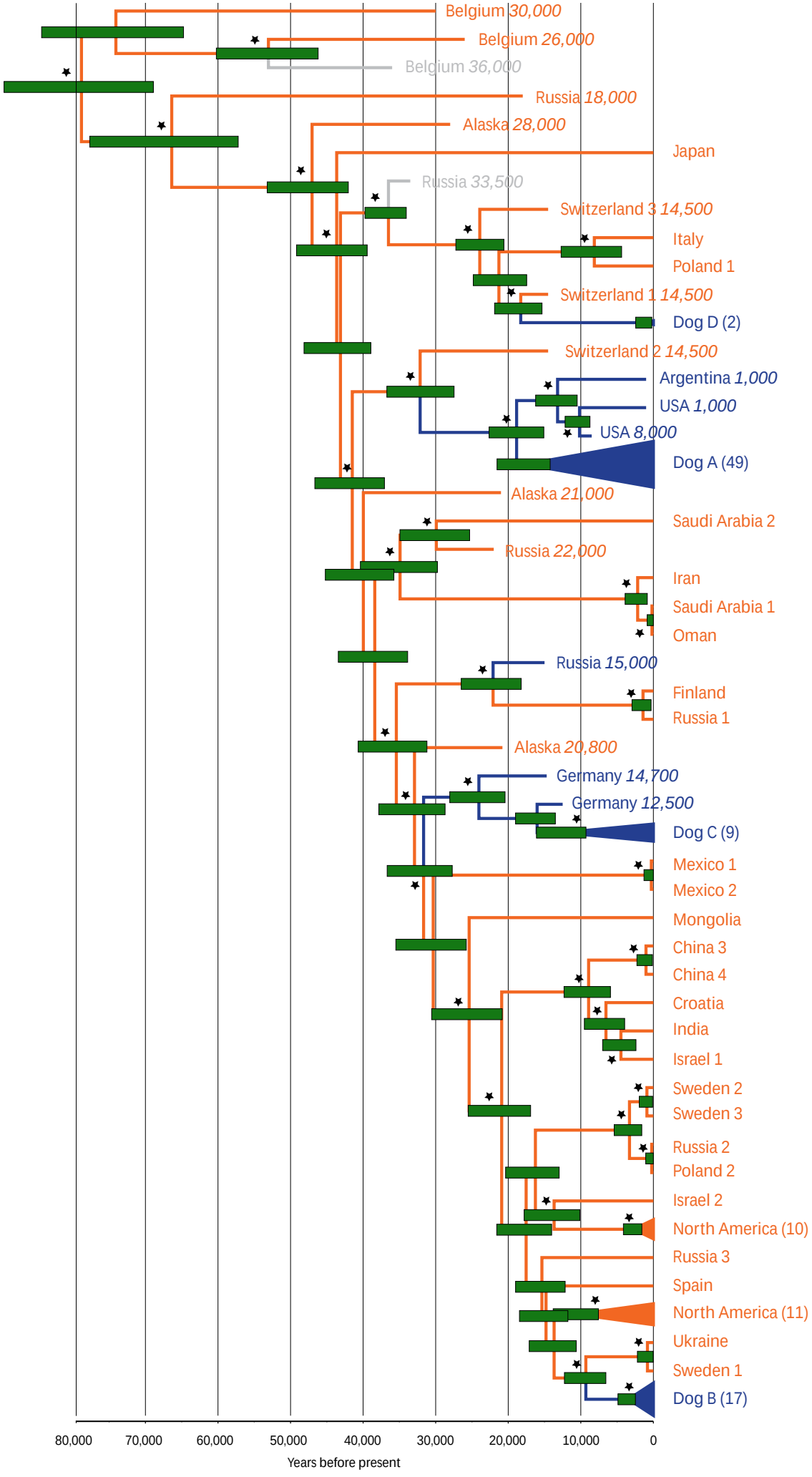
A)


coverage Altai dog
number of reads
160
140
120
100
80
60
40
20
0
15406
15456
15506
15556
15606
15656
15706
15756
15806
Position

B)


coverage position 15438
number of reads
100
90
80
70
60
50
40
30
20
10
0
C C A A A G C T G A A A T T C T T C T T A
Nucleotide
0 reads with nucleotide G
coverage position 15614
number of reads
60
50
40
30
20
10
0
A T A G G A C A T A C T A A C T C A A T C
Nucleotide
7 reads with nucleotide T


coyote3
coyote2
coyote1
coyote4
China1
China2
Belgium 30,000
Belgium 26,000
Belgium 36,000
Russia 18,000
Alaska 28,000
Japan
Russia 33,500
Poland1
Italy
Switzerland 3 14,500
Switzerland 1 14,500
D03
D34
Switzerland 2 14,500
D61
Argentina 1,000
USA 8,500
USA 1,000
Dingo
D01
D07
D18
D27
D83
D85
D05
D90
D93
Basenji
D22
D75
D69
D86
D80
D87
D95
D102
D91
D94
D96
D97
D25
ChineseDog2
ChineseDog3
D21
D79
D81
D71
D72
D88
D16
D73
D67
D89
D15
D103
D66
D78
D30
D28
D56
D59
D82
D68
D04
D65
Saudi Arabia 2
Russia 22,000
Iran
Saudi Arabia 1
Oman
Alaska 21,000
Russia 15,000
Finland
Russia1
Alaska 20,000
Germany 14,700
Germany 12,500
ChineseDog1
D09
D52
D55
D12
D60
D84
D11
D63
Mexico1
Mexico2
Mongolia
China3
China4
Croatia
Israel1
India
Spain
Russia3
Sweden2
Sweden3
Russia2
Poland2
Israel2
Alaska4
Alaska6
Canada2
Canada4
Alaska2
Canada8
Alaska5
USA2
USA1
USA3
Alaska1
Alaska3
Canada9
Canada1
Canada5
Canada10
Canada6
Canada3
USA4
Canada7
Canada11
Sweden1
Ukraine
D26
D51
D31
D40
D44
D49
D50
D32
D46
D42
D35
D48
D47
D33
D41
D98
D99
0.0070


Belgium 30,000
Belgium 26,000
Belgium 36,000
Russia 18,000
Alaska 28,000
Japan
Russia 33,500
Switzerland 3 14,500
Italy
Poland 1
Switzerland 1 14,500
Dog D (2)
Switzerland 2 14,500
Argentina 1,000
USA 1,000
USA 8,000
Dog A (49)
Alaska 21,000
Saudi Arabia 2
Russia 22,000
Iran
Saudi Arabia 1
Oman
Russia 15,000
Finland
Russia 1
Alaska 20,800
Germany 14,700
Germany 12,500
Dog C (9)
Mexico 1
Mexico 2
Mongolia
China 3
China 4
Croatia
India
Israel 1
Sweden 2
Sweden 3
Russia 2
Poland 2
Israel 2
North America (10)
Russia 3
Spain
North America (11)
Ukraine
Sweden 1
Dog B (17)
80,000
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
Years before present

References and Notes

1. G. Larson, J. Burger, A population genetics view of animal domestication. *Trends Genet.* **29**, 197–205 (2013).

2. 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).

3. S. J. Olsen, J. W. Olsen, The Chinese wolf, ancestor of new world dogs. *Science* **197**, 533–535 (1977).

4. 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).

5. 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).

6. P. Savolainen, Y. P. Zhang, J. Luo, J. Lundeberg, T. Leitner, Genetic evidence for an East Asian origin of domestic dogs. *Science* **298**, 1610–1613 (2002).

7. 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).

8. M. Sablin, G. Khlopachev, The earliest Ice Age dogs: Evidence from Eliseevichi 11. *Curr. Anthropol.* **43**, 795–799 (2002).

9. S. J. Crockford, Y. V. Kuzmin, Comments on Germonpré *et al*., *Journal of Archaeological Science* 36, 2009 "Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: Osteometry, ancient DNA and stable isotopes", and Germonpré, Lázničková-Galetová, and Sablin, *Journal of Archaeological Science* 39, 2012 "Palaeolithic dog skulls at the Gravettian Předmostí site, the Czech Republic." *J. Archaeol. Sci.* **39**, 2797–2801 (2012).

10. G. Larson, E. K. Karlsson, A. Perri, M. T. Webster, S. Y. Ho, J. Peters, P. W. Stahl, P. J. Piper, F. Lingaas, M. Fredholm, K. E. Comstock, J. F. Modiano, C. Schelling, A. I. Agoulnik, P. A. Leegwater, K. Dobney, J. D. Vigne, C. Vilà, L. Andersson, K. Lindblad-Toh, Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 8878–8883 (2012).

11. H. Napierala, H.-P. Uerpmann, A 'new' palaeolithic dog from central Europe. *Int. J. Osteoarchaeol.* **22**, 127–137 (2012).

12. G. Nobis, Der älteste Haushund lebte vor 14000 Jahren. *Umschau* **79**, 610 (1979).

13. N. D. Ovodov, S. J. Crockford, Y. V. Kuzmin, T. F. Higham, G. W. 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).

14. M. Baales, Umwelt und Jagdökonomie der Ahrensburger Rentierjäger im Mittelgebirge. *Mongraphien RGZM, Mainz* **38**, 106 (1996).

15. Supplementary materials are available on *Science* Online.

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

17. J. A. Leonard, R. K. Wayne, J. Wheeler, R. Valadez, S. Guillén, C. Vilà, Ancient DNA evidence for Old World origin of New World dogs. *Science* **298**, 1613–1616 (2002).

18. B. van Asch, A. B. Zhang, M. C. Oskarsson, C. F. Klütsch, A. Amorim, P. Savolainen, Pre-Columbian origins of Native American dog breeds, with only limited replacement by European dogs, confirmed by mtDNA analysis. *Proc. Biol. Soc.* **280**, 20131142 (2013).

19. H. Malmström, C. Vilà, M. T. 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).

20. 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).

21. K. Lindblad-Toh, C. M. Wade, T. S. Mikkelsen, E. K. Karlsson, D. B. Jaffe, M. Kamal, M. Clamp, J. L. Chang, E. J. Kulbokas 3rd, M. C. Zody, E. Mauceli, X. Xie, M. Breen, R. K. Wayne, E. A. Ostrander, C. P. Ponting, F. Galibert, D. R. Smith, P. J. DeJong, E. Kirkness, P. Alvarez, T. Biagi, W. Brockman, J. Butler, C. W. Chin, A. Cook, J. Cuff, M. J. Daly, D. DeCaprio, S. Gnerre, M. Grabherr, M. Kellis, M. Kleber, C. Bardeleben, L. Goodstadt, A. Heger, C. Hitte, L. Kim, K. P. Koepfli, H. G. Parker, J. P. Pollinger, S. M. Searle, N. B. Sutter, R. Thomas, C. Webber, J. Baldwin, A. Abebe, A. Abouelleil, L. Aftuck, M. Ait-Zahra, T. Aldredge, N. Allen, P. An, S. Anderson, C. Antoine, H. Arachchi, A. Aslam, L. Ayotte, P. Bachantsang, A. Barry, T. Bayul, M. Benamara, A. Berlin, D. Bessette, B. Blitshteyn, T. Bloom, J. Blye, L. Boguslavskiy, C. Bonnet, B. Boukhgalter, A. Brown, P. Cahill, N. Calixte, J. Camarata, Y. Cheshatsang, J. Chu, M. Citroen, A. Collymore, P. Cooke, T. Dawoe, R. Daza, K. Decktor, S. DeGray, N. Dhargay, K. Dooley, K. Dooley, P. Dorje, K. Dorjee, L. Dorris, N. Duffey, A. Dupes, O. Egbiremolen, R. Elong, J. Falk, A. Farina, S. Faro, D. Ferguson, P. Ferreira, S. Fisher, M. FitzGerald, K. Foley, C. Foley, A. Franke, D. Friedrich, D. Gage, M. Garber, G. Gearin, G. Giannoukos, T. Goode, A. Goyette, J. Graham, E. Grandbois, K. Gyaltsen, N. Hafez, D. Hagopian, B. Hagos, J. Hall, C. Healy, R. Hegarty, T. Honan, A. Horn, N. Houde, L. Hughes, L. Hunnicutt, M. Husby, B. Jester, C. Jones, A. Kamat, B. Kanga, C. Kells, D. Khazanovich, A. C. Kieu, P. Kisner, M. Kumar, K. Lance, T. Landers, M. Lara, W. Lee, J. P. Leger, N. Lennon, L. Leuper, S. LeVine, J. Liu, X. Liu, Y. Lokyitsang, T. Lokyitsang, A. Lui, J. Macdonald, J. Major, R. Marabella, K. Maru, C. Matthews, S. McDonough, T. Mehta, J. Meldrim, A. Melnikov, L. Meneus, A. Mihalev, T. Mihova, K.

Miller, R. Mittelman, V. Mlenga, L. Mulrain, G. Munson, A. Navidi, J. Naylor, T. Nguyen, N. Nguyen, C. Nguyen, T. Nguyen, R. Nicol, N. Norbu, C. Norbu, N. Novod, T. Nyima, P. Olandt, B. O'Neill, K. O'Neill, S. Osman, L. Oyono, C. Patti, D. Perrin, P. Phunkhang, F. Pierre, M. Priest, A. Rachupka, S. Raghuraman, R. Rameau, V. Ray, C. Raymond, F. Rege, C. Rise, J. Rogers, P. Rogov, J. Sahalie, S. Settipalli, T. Sharpe, T. Shea, M. Sheehan, N. Sherpa, J. Shi, D. Shih, J. Sloan, C. Smith, T. Sparrow, J. Stalker, N. Stange-Thomann, S. Stavropoulos, C. Stone, S. Stone, S. Sykes, P. Tchuinga, P. Tenzing, S. Tesfaye, D. Thoulutsang, Y. Thoulutsang, K. Topham, I. Topping, T. Tsamla, H. Vassiliev, V. Venkataraman, A. Vo, T. Wangchuk, T. Wangdi, M. Weiand, J. Wilkinson, A. Wilson, S. Yadav, S. Yang, X. Yang, G. Young, Q. Yu, J. Zainoun, L. Zembek, A. Zimmer, E. S. Lander, Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–819 (2005).

22. M. S. Gill, P. Lemey, N. R. Faria, A. Rambaut, B. Shapiro, M. A. Suchard, Improving Bayesian population dynamics inference: A coalescent-based model for multiple loci. *Mol. Biol. Evol.* **30**, 713–724 (2013).

23. J. A. Tennessen, A. W. Bigham, T. D. O'Connor, W. Fu, E. E. Kenny, S. Gravel, S. McGee, R. Do, X. Liu, G. Jun, H. M. Kang, D. Jordan, S. M. Leal, S. Gabriel, M. J. Rieder, G. Abecasis, D. Altshuler, D. A. Nickerson, E. Boerwinkle, S. Sunyaev, C. D. Bustamante, M. J. Bamshad, J. M. Akey, Broad GO, Seattle GO, on behalf of the NHLBI Exome Sequencing Project, Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012).

24. Z. L. Ding, M. Oskarsson, A. Ardalan, H. Angleby, L. G. Dahlgren, C. Tepeli, E. Kirkness, P. Savolainen, Y. P. Zhang, Origins of domestic dog in southern East Asia is supported by analysis of Y-chromosome DNA. *Heredity* **108**, 507–514 (2012).

25. J. Alroy, A multispecies overkill simulation of the end-Pleistocene megafaunal mass extinction. *Science* **292**, 1893–1896 (2001).

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

27. M. Germonpré, M. V. Sablin, V. Després, M. Hofreiter, M. Lázničková-Galetová, R. E. Stevens, M. Stiller, Palaeolithic dogs and the early domestication of the wolf: A reply to the comments of Crockford and Kuzmin (2012). *J. Archaeol. Sci.* **40**, 786–792 (2013).

28. D. F. Morey, M. D. Wiant, Early holocene domestic dog burials from the North American Midwest. *Curr. Anthropol.* **33**, 224 (1992).

29. N. Rohland, M. Hofreiter, Ancient DNA extraction from bones and teeth. *Nat. Protoc.* **2**, 1756–1762 (2007).

30. H. A. Burbano, E. Hodges, R. E. Green, A. W. Briggs, J. Krause, M. Meyer, J. M. Good, T. Maricic, P. L. Johnson, Z. Xuan, M. Rooks, A. Bhattacharjee, L. Brizuela, F. W. Albert, M. de la Rasilla, J. Fortea, A. Rosas, M. Lachmann, G. J. Hannon, S. Pääbo, Targeted investigation of the Neandertal genome by array-based sequence capture. *Science* **328**, 723–725 (2010).

31. L. Orlando, A. Ginolhac, G. Zhang, D. Froese, A. Albrechtsen, M. Stiller, M. Schubert, E. Cappellini, B. Petersen, I. Moltke, P. L. Johnson, M. Fumagalli, J. T. Vilstrup, M.

Raghavan, T. Korneliussen, A. S. Malaspinas, J. Vogt, D. Szklarczyk, C. D. Kelstrup, J. Vinther, A. Dolocan, J. Stenderup, A. M. Velazquez, J. Cahill, M. Rasmussen, X. Wang, J. Min, G. D. Zazula, A. Seguin-Orlando, C. Mortensen, K. Magnussen, J. F. Thompson, J. Weinstock, K. Gregersen, K. H. Røed, V. Eisenmann, C. J. Rubin, D. C. Miller, D. F. Antczak, M. F. Bertelsen, S. Brunak, K. A. Al-Rasheid, O. Ryder, L. Andersson, J. Mundy, A. Krogh, M. T. Gilbert, K. Kjær, T. Sicheritz-Ponten, L. J. Jensen, J. V. Olsen, M. Hofreiter, R. Nielsen, B. Shapiro, J. Wang, E. Willerslev, Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**, 74–78 (2013).

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

33. M. Kircher, S. Sawyer, M. Meyer, Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).

34. 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).

35. E. Hodges, Z. Xuan, V. Balija, M. Kramer, M. N. Molla, S. W. Smith, C. M. Middle, M. J. Rodesch, T. J. Albert, G. J. Hannon, W. R. McCombie, Genome-wide in situ exon capture for selective resequencing. *Nat. Genet.* **39**, 1522–1527 (2007).

36. T. Maricic, M. Whitten, S. Pääbo, Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* **5**, e14004 (2010).

37. M. Kircher, U. Stenzel, J. Kelso, Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biol.* **10**, R83 (2009).

38. R. E. Green, A. S. Malaspinas, J. Krause, A. W. Briggs, P. L. Johnson, C. Uhler, M. Meyer, J. M. Good, T. Maricic, U. Stenzel, K. Prüfer, M. Siebauer, H. A. Burbano, M. Ronan, J. M. Rothberg, M. Egholm, P. Rudan, D. Brajković, Z. Kućan, I. Gusić, M. Wikström, L. Laakkonen, J. Kelso, M. Slatkin, S. Pääbo, A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416–426 (2008).

39. J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, ed. 2, 1989).

40. M. Meyer, U. Stenzel, M. Hofreiter, Parallel tagged sequencing on the 454 platform. *Nat. Protoc.* **3**, 267–278 (2008).

41. M. Meyer, A. W. Briggs, T. Maricic, B. Höber, B. Höffner, J. Krause, A. Weihmann, S. Pääbo, M. Hofreiter, From micrograms to picograms: Quantitative PCR reduces the material demands of high-throughput sequencing. *Nucleic Acids Res.* **36**, e5 (2008).

42. U. Arnason, A. Gullberg, A. Janke, M. Kullberg, Mitogenomic analyses of caniform relationships. *Mol. Phylogenet. Evol.* **45**, 863–874 (2007).

43. K. Katoh, K. Kuma, H. Toh, T. Miyata, MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518 (2005).

44. T. A. Hall, BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95 (1999).

45. P. Savolainen, L. Arvestad, J. Lundeberg, mtDNA tandem repeats in domestic dogs and wolves: Mutation mechanism studied by analysis of the sequence of imperfect repeats. *Mol. Biol. Evol.* **17**, 474–488 (2000).

46. H. A. Schmidt, K. Strimmer, M. Vingron, A. von Haeseler, TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502–504 (2002).

47. M. Hasegawa, H. Kishino, T. Yano, Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174 (1985).

48. B. Q. Minh, M. A. T. Nguyen, A. von Haeseler, Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).

49. D. Darriba, G. L. Taboada, R. Doallo, D. Posada, jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **9**, 772 (2012).

50. H. Akaike, in *Proceedings of the Second International Symposium on Information Theory*, Tsahkadsor, Armenia, N. Petrov and S. Caski, Eds. (Akadeniai Kiado, Budapest, 1973), pp. 267–281.

51. F. Ronquist, M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, J. P. Huelsenbeck, MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).

52. W. Xie, P. O. Lewis, Y. Fan, L. Kuo, M.-H. Chen, Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst. Biol.* **60**, 150–160 (2011).

53. P. Librado, J. Rozas, DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).

54. C. Meng, H. Zhang, Q. Meng, Mitochondrial genome of the Tibetan wolf. *Mitochondrial DNA* **20**, 61–63 (2009).

55. A. J. Drummond, M. A. Suchard, D. Xie, A. Rambaut, Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).

56. B. Shapiro, A. Rambaut, A. J. Drummond, Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol. Biol. Evol.* **23**, 7–9 (2006).

57. A. J. Drummond, S. Y. Ho, M. J. Phillips, A. Rambaut, Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).

58. A. Rambaut, A. Drummond, Tracer, version 1.4, Computer program and documentation distributed by the author; http://beast.bio.ed.ac.uk/Tracer (2007).

59. R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H. Fritz, N. F. Hansen, E. Y. Durand, A. S. Malaspinas, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Prüfer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Höber, B. Höffner, M. Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Z. Kucan, I. Gusic, V. B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. de la Rasilla, J. Fortea, A. Rosas, R. W. Schmitz, P. L. Johnson, E. E. Eichler, D. Falush, E. Birney, J. C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D.

Reich, S. Pääbo, A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).

60. S. Sawyer, J. Krause, K. Guschanski, V. Savolainen, S. Pääbo, Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS ONE* **7**, e34131 (2012).

61. M. Hofreiter, V. Jaenicke, D. Serre, A. von Haeseler, S. Pääbo, DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res.* **29**, 4793–4799 (2001).

62. M. Stiller, R. E. Green, M. Ronan, J. F. Simons, L. Du, W. He, M. Egholm, J. M. Rothberg, S. G. Keates, N. D. Ovodov, E. E. Antipina, G. F. Baryshnikov, Y. V. Kuzmin, A. A. Vasilevski, G. E. Wuenschell, J. Termini, M. Hofreiter, V. Jaenicke-Després, S. Pääbo, Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 13578–13584 (2006).

63. J.-F. Pang, C. Kluetsch, X. J. Zou, A. B. Zhang, L. Y. Luo, H. Angleby, A. Ardalan, C. Ekström, A. Sköllermo, J. Lundeberg, S. Matsumura, 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).

64. S. Björnerfeldt, M. T. Webster, C. Vilà, Relaxation of selective constraint on dog mitochondrial DNA following domestication. *Genome Res.* **16**, 990–994 (2006).

65. M. Hofreiter, D. Serre, H. N. Poinar, M. Kuch, S. Pääbo, Ancient DNA. *Nat. Rev. Genet.* **2**, 353–359 (2001).

66. R. E. Green, A. W. Briggs, J. Krause, K. Prüfer, H. A. Burbano, M. Siebauer, M. Lachmann, S. Pääbo, The Neandertal genome and ancient DNA authenticity. *EMBO J.* **28**, 2494–2502 (2009).

67. R. E. Green, J. Krause, S. E. Ptak, A. W. Briggs, M. T. Ronan, J. F. Simons, L. Du, M. Egholm, J. M. Rothberg, M. Paunovic, S. Pääbo, Analysis of one million base pairs of Neanderthal DNA. *Nature* **444**, 330–336 (2006).

68. J. D. Wall, S. K. Kim, Inconsistencies in Neanderthal genomic DNA sequences. *PLoS Genet.* **3**, 1862–1866 (2007).

69. A. W. Briggs, J. M. Good, R. E. Green, J. Krause, T. Maricic, U. Stenzel, C. Lalueza-Fox, P. Rudan, D. Brajkovic, Z. Kucan, I. Gusic, R. Schmitz, V. B. Doronichev, L. V. Golovanova, M. de la Rasilla, J. Fortea, A. Rosas, S. Pääbo, Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325**, 318–321 (2009).

70. J. A. Leonard, O. Shanks, M. Hofreiter, E. Kreuz, L. Hodges, W. Ream, R. K. Wayne, R. C. Fleischer, Animal DNA in PCR reagents plagues ancient DNA research. *J. Archaeol. Sci.* **34**, 1361–1366 (2007).

71. K. Strimmer, A. von Haeseler, Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6815–6819 (1997).

72. A. S. Druzhkova, O. Thalmann, V. A. Trifonov, J. A. Leonard, N. V. Vorobieva, N. D. Ovodov, A. S. Graphodatsky, R. K. Wayne, Ancient DNA analysis affirms the canid from Altai as a primitive dog. *PLoS ONE* **8**, e57754 (2013).

73. A. Acosta, D. Loponte, C. G. Esponda, Primer registro de perro doméstico prehispánico (*Canis familiaris*) entre los grupos cazadores recolectores del humedal de Paraná inferior (Argentina). *Antipod. Rev. Antropol. Arqueol.* **13**, 175–199 (2011).

74. B. E. Worthington, thesis, Florida State University (2008).

75. M. Nei, *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York, 1987).

76. G. A. Watterson, On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**, 256–276 (1975).

77. F. Tajima, Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).

78. Y. X. Fu, W. H. Li, Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709 (1993).