PNAS
www.pnas.org

Supplementary Information for

Modern Arctic dog ancestry was shaped by several thousand years of Eurasian-wide trade and human dispersal

Tatiana R. Feuerborn[§1,2,3,4,5], Alberto Carmagnini[§6], Robert J. Losey[7,8], Tatiana Nomokonova[8], Arthur Askeyev[9], Igor Askeyev[9], Oleg Askeyev[9], Ekaterina E. Antipina[23], Martin Appelt[10], Olga P. Bachura[11], Fiona Beglane[12], Daniel G. Bradley[13], Kevin G. Daly[13], Shyam Gopalakrishnan[1], Kristian Murphy Gregersen[15], Chunxue Guo[16], Andrei V. Gusev[17], Carleton Jones[18], Pavel A. Kosintsev[11,19], Yaroslav Kuzmin[20], Valeria Mattiangeli[13], Angela Perri[21], Andrei V. Plekhanov[17], Jazmín Ramos-Madrigal[1], Anne Lisbeth Schmidt[10], Dilyara Shaymuratova[9], Oliver Smith[1,22], Lilia V. Yavorskaya[23], Guojie Zhang[16,24,25,26], Eske Willerslev[27,28,29,30], Morten Meldgaard[1,2,31], M Thomas P Gilbert[1,32], Greger Larson[33], Love Dalén[4,5], Anders J. Hansen[†1,2], Mikkel-Holger S. Sinding[†1,2,13,14], Laurent Frantz[†6,34]

Tatiana R. Feuerborn, Alberto Carmagnini, Laurent Frantz
Email: tatianafeuerborn@palaeome.org; carmagnini.alberto@gmail.com; laurent.frantz@qmul.ac.uk

This PDF file includes:

Supplementary Information Text

Materials

Sample information: Archaeological Sites

Veretye, Karelia, Russia

TRF-04-09:AL3253

The Veretye (ВерeТвё) [a.k.a. Veret'ye] Culture derives its name from the Veretye 1 site. It is located in Arkhangelsk Province, northern European Russia, on the bank of Kinema River ca. 1 km from its mouth; the river flows to Lake Lacha. Geographic coordinates for the Veretye 1 site are: 57,197296 ° N, 29,73683 ° E. The excavated area is ca. 1470 m2. The site has a single cultural component associated with the Mesolithic epoch (*1*, *2*). Planigraphically, there are remains of dwellings and other structures (of everyday life use), and scattered artefacts. Cultural material is located in oxygen-free peat layers (wetland site), and the preservation of bones, antlers and other perishable materials (wood, birch bark, and plant fibres) is generally very good. Material culture is represented by a large set of stone tools used for making items for hunting, fishing, and working the wood. No pottery is found. Numerous tools on bone, antler, and wood are also found.

Chronology of the Veretye 1 site is based on 14C dates obtained on different materials from the cultural layer: a) charcoal: 9600 ± 80 BP (Le-1469), 9050 ± 80 BP (GIN-4031), 8560 ± 120 BP (GIN-2452), 8520 ± 80 BP (GIN-4030), 8270 ± 100 BP (Le-1470), and 7960 ± 100 BP (Le-1471); b) antler: 9370 ± 80 BP (GIN-4833) and 8340 ± 120 BP (GIN-4832); and c) wood: 8750 ± 70 BP (Le-1472), 8550 ± 130 BP (GIN-2452), and 7700 ± 80 BP (Le-1773) (*2*). Oshibkina thinks that 14C dates on charcoal and worked wood of ca. 9600–8550 BP are the most closely associated with the cultural component (*1*).

Bones and skulls of 42 dogs were found at the Veretye 1 site, and it constitutes 12.6% of total animal bones from this site (*2*). One dog was sampled from the site for this study and underwent whole genome sequencing. A previously published genome from a second dog from the site was included in the dataset (*3*). The dog genome published by Bergstrom et al. was directly 14 C dated to 9575 ± 50 BP (OxA-36900), with a corresponding calendar age of 10,780–11,080 cal BP, a median age of approximatelty 10,930 cal BP (with ± 1 sigma; using IntCal13 dataset) (*3*). The dog remains sequenced for this study were recovered from the same context as the dog sequenced in Bergstrom et al. (*3*).

Ishkinino, Orenburg Province, Russia

TRF-04-04:AL2307

Iskinino is a Late Bronze Age site in the southern Urals of the Eurasian steppe belt, located within the Sukhaya Guberlya River valley (*4*). The site is situated in one of the richest mining areas of the Trans-Urals region where cobalt-copper-pyrite was exploited during the Bronze Age (*5*). Several small contemporaneous sites were located in close proximity to one another, collectively called Ishkinino cluster, that possibly operated as a large economic unit (*5*). The site consists of several dwellings. Study of animal bones shows that people were breeding both cattles and small stock (sheep and goats), and horses. The main activities were mining of copper ore and smelting of it. Animal bone 14C dates from this site are 3020 ± 150 BP (LE-8854) (or 1610–900 cal BC), 3190 ± 100 BP (LE-8855) (1730–1210 cal BC), and 2940 ± 200 BP (LE-9342) (1680–600 cal BC) (*4*). The overall time of human occupation is around 1500–1200 BC, considering relatively large standard deviations for the 14C dates. One dog from the site was sequenced for this study.

Ust'-Polui, Iamal-Nenets Region, Russia

TRF-05-04:736-6410, TRF-05-05:736-6430, TRF-05-07:736-6608, TRF-05-08:736-6648, TRF-05-10:736-4921, TRF-05-03:736-6581/3, TRF-05-06:736-7407, TRF-05-09:736-31679, TRF-05-11:736-5034

Genomic information was obtained from five dogs recovered from excavations at the site of Ust'-Polui in Iamal. The site was occupied during the Iron Age, with radiocarbon dates from the sites placing the occupation between 2,210 and 1,810 BP (*6*). Almost a quarter (21%) of the faunal remains present at the site belong to dogs, representing over 100 individuals (*6*, *7*). The site is thought to have been a ceremonial site rather than a settlement (*6*). The site shows evidence for young dogs having been eaten, possibly in the context of rituals while other dogs appear to have been intentionally buried (*7*). The site also contains materials typically associated with sledding, such as parts of harnesses and sleds (*7*). Nine dogs from the site were sequenced for this study, sufficient nuclear genome coverage was obtained for four of the nine samples, while sufficient mitochondrial genome coverage was obtained from five individuals.

Tiutei-Sale I, Iamal Peninsula, Russia

TRF-05-14:783-1020

Tiutei-Sale I is a medieval settlement located on the northwest shore of the Iamal Peninsula, overlooking a cove where several streams enter the Kara Sea (*8*). The site was occupied during multiple periods, between the 6th and 8th centuries CE and again between the twelfth and fourteenth centuries CE (*8*, *9*). The site was interpreted as a warm-season settlement where people focused on hunting reindeer and arctic fox, and occasionally procured marine resources such as walrus and seals. One dog was sampled for this study, deriving from the later phase of occupation, and it was directly dated to 1,111 +/- 30 BP, or 1,169 to 936 cal. BP.

Iarte 6, Iamal Peninsula, Russia

TRF-05-16:677-7393-7403

Iarte 6 is located on the open tundra of the central Iamal Peninsula. It is a medieval settlement with remains of seven house pits, and based on a large set of radiocarbon dates and dendrochronology, dates from 1016 to 1122 CE, or 934 - 828 cal. BP (*8*, *9*). Excavations at Iarte 6 produced one of the largest reindeer assemblages in the Arctic, numbering just over 22,000 specimens. It is unclear if the reindeer were wild, domesticated, or a combination of both. Remains of dogs/wolves at this location account for less than 1% of the total number of faunal remains found at this site (*8*, *9*). Two dogs from the site were sampled for this study.

Ust'-Voikar, Iamal-Nenets Region, Russia

TRF-05-17:1232-6837

Ust'-Voikar is a medieval town located southwest on the left bank of Gornyi Ob River in the southern portion of the Iamal-Nenets region of Northwest Siberia. This town was occupied between the thirteenth and twentieth centuries CE (*10*). The majority of the faunal remains found at the site belong to reindeer, but arctic fox, hare, dogs, and other species also were present (*10*). Dog remains were found throughout the excavated area and belong mostly to adult individuals, but a few juvenile skeletons were also found. One dog from the site was sequenced for this study and was directly dated to 857 +/- 30, calibrated to 900 to 693 cal. BP.

Endyrskoe 1, Khanty-Mansi Region, Russia

TRF-05-12:755-2100

This fortified habitation site is located about 70 km from the town of Niagan' on the bank of the Endyr' River in the Khanty-Mansi Region of Northwest Siberia (*11*, *12*). This is an Early Iron Age and medieval fortified site containing multiple dwellings, with occupation appearing to focus on the 6-7th and 11-16th centuries CE (*11*, *12*). The faunal remains from the site have not been reported, but the skulls of several dogs were present in the foundations of some of the dwellings. One specimen was sampled and sequenced for this study. This specimen was directly dated to 1460 +/- 30, calibrated to 1398 to 1302 cal. BP.

Peregrebnoe 1, Khanty-Mansi Region, Russia

TRF-05-13:203-508

This medieval period habitation site is located on the Ob River in the Khanty-Mansi Region of Northwest Siberia within the modern town of Peregrebnoe (*13*). This site is typologically dated to the 12th through mid-13thcenturies CE. The site was fortified and contained multiple dwellings. The faunal remains at the site primarily consist of fur-bearing animals but remains of domestic horses and cattle also are present. A total of 144 dog remains were found at the site. One

specimen was sampled and sequenced for this study, which was directly dated to 1093 +/- 31, calibrated to 1060 to 935 cal BP.

Bolgar, Volga River, Russia

TRF-04-10:AL2275

The site of Bolgar (Bolghar) is one of the capitals of the Volga Bulgaria (Volga–Kama Bulghar) state. It is located on the left bank of the Volga River, in the middle course of the basin. Cultural layer is dated from the end of the tenth century CE to the beginning of the fifteenth century CE. Before the Mongol invasion in the thirteenth century CE, it was one of the major cities of the Volga Bulgaria. After the defeat by the Mongol armies in 1236, this city became the center of the Bulgarian Ulus of the Golden Horde, one of the Mongol states. During the Golden Horde time, in the second half of the thirteenth century and in the fourteenth century, the Bolgar became significantly larger than in pre-Mongol time, and experienced its heyday. In its central part of site, 40 m to SW of the Cathedral Mosque, a craftsmen quarter appears with a large number of crafts: metallurgy, glass-making, processing of ornamental stones (in particular, amber), and bone-carving (*14*). In the fourteenth century CE, one of such crafts in this part of the city was the taming and breeding of specialized animals for hunting: bone remains of daytime birds of prey with an appropriate gender and age profile were recorded, as well as artifacts testifying about hunting with the help of prey birds. In the same quarter, at excavation pit No. 192, in 2013 in a large Depression 1 among typical “kitchen” remains, skeletons of 16 adult dogs that died at one time with traces of healing injuries were found; also, scattered bones of several puppies were recorded. According to our assumption, this was a collection of specially trained hunting dogs (*15*). A DNA sequenced sample comes from one of the dogs in this excavation pit.

Sample information: Ethnographic Samples

The 5th Thule Expedition 1921-1924 under Danish explorer Knud Rasmussen’s leadership travelled from Greenland to Arctic Canada, to Alaska and Siberia (*16*). The expedition only spent 48 hours in Siberia due to complications with visas. In order to fulfil the purpose of his expedition, Rasmussen in 1927, from the widow of the German antique dealer Eugen Alexander, purchased three Siberian collections including clothing with components made of dog fur. The collections were donated to the National Museum of Denmark. In the museum’s Siberian collections were also included items, exchanged with Russian institutions. Further information can be found on SkinBase (http://skinddragter.natmus.dk/Clothing). Several samples were attributed to a general location, such as Amur Delta (K.1-3/89375), Chukotka (TRF-02-01/K.613/19451), while others were attributed to the culture from which they were collected rather than a specific location, Chukchi (TRF-02-49/K.607/19448), Nivkh (TRF-02-21/K.1-1/89373_a), and Nenets/Khanty (TRF-02-25/K.3-7/90206_a). Dogs from the Kamchatka Krai, specifically dogs of the Koryaks were added to the National Museum of Denmark collections also from clothing made with dog fur in

1961, (TRF-02-53/K.1161b/67204, TRF-02-54/K.1160b/67202). Mitochondrial genomes from these individuals as well as other dogs from the National Museum of Denmark have been previously identified as dog fur(*17*). One ethnographic dog sample was obtained from the island of Sakhalin off the east coast of Russia, just north of Japan (TRF-01-51/ZMK 1054). Museum records at the Natural History Museum of Denmark show that this dog was collected in 1872.

Sample information: Comparative Dataset

Included in the analyses for comparison was a panel of publically available dogs (n=115), wolves (n=39), an Andean fox, and a black-backed jackal, see Dataset S2 for full list of accession numbers.

Methods

DNA was extracted from bone (n=17), tooth (n=3), and skin (n=9) samples to address the questions in this study (Dataset S1).

Sample Preparation

Samples obtained from bones and teeth were cleaned of surface contaminants prior to sampling to aid in the reduction of contaminant DNA. Surface cleaning was conducted through superficial drilling of the outer surface with a Dremel drill. In the case of samples obtained from teeth, both the outer surface and the enamel layer of the tooth were removed by drilling. Between 30 and 100 mg of bone powder was drilled for extraction from the bone sample. For bone samples which were derived from the petrous bone powder was drilled from the interior, for all other bone samples the densest part of the bone was drilled. For tooth samples 15 to 30 mg of powder was drilled from the cementum layer of the tooth.

Samples obtained from hides were based on macroscopic identification of dog hides used as materials for clothing identified garments which possessed dog fur stored in the ethnographic collections of the National Museum of Denmark. Further confirmation of the taxonomic assignment was performed with shallow shotgun sequencing in a previous study (*17*). Eight samples were taken from clothing which were believed to be dog hides. Samples were taken from the hide approximately the size of a grain of rice for DNA extraction. Sterile scalpels were used to remove hair from the hides, the hair was retained for other analyses (*17*).

Extraction

Bones and teeth processing involved a pre-digestion step, performed prior to the extraction in order to improve endogenous content of the samples. A 30 minute incubation of the bone/tooth powder and hide sample with 315 ul of 0.5M EDTA and 7.5 ul of 10 ng/ul proteinase K was

utilised for pre-digestion procedure. After the incubation the samples were centrifuged for 5 minutes at 13,000 rpm and the supernatant was removed. The undigested bone/tooth powder underwent an overnight digestion with 630 ul of 0.5M EDTA, 70 ul of 1M UREA, and 15 ul of 10 ng/ul proteinase K, following (*18*). The hide samples were incubated overnight at 37°C with a lysis buffer according to (*19*). All extracts were purified using Qiagen MinElute columns.

Library Build

Libraries were built for each sample for the shallow shotgun sequencing (screening) on the Illumina 2500 HiSeq platform. The single tube library build protocol, 'BEST', was used to prepare libraries with Illumina adapters and indexing primers (*20*). This protocol involves no purification steps after the extract has been purified until the library build is completed and ready for indexing PCRs. Several extractions underwent USER (Uracil-Specific Excision Reagent) treatment prior to undergoing the blunt end repair step, see Dataset S1, in order to remove uracil sites before library building. The USER treatment involved the addition of 4.8 µl of Thermolabile USER II (1U/ µl) to 32µl of the DNA extract and a 3 hour incubation at 37°C, followed by a purification with MinElute columns and elution in 32µl EB. As per the BEST library built protocol, after extraction the DNA undergoes blunt end repair with a denaturing step at 65°C, without purification the sample proceeds to the adapter ligation when adapters were ligated to the ends of the DNA fragments. Finally, after the adapter ligation the nicks in the strands are filled, then the library was purified using Qiagen MinElute columns. All indexing was performed with AmpliTaq Gold (2.5U uL-1) for non-USER treated libraries and USER treated libraries were indexed using Polymerase Pfu Turbo (2.5U uL-1).

Sequencing

The shallow shotgun screening was performed at Science for Life Laboratory in Stockholm, Sweden. After assessment of the clonality and endogenous DNA content further libraries were built for the BGISEQ-500 sequencing platform with platform specific adapters and indexing primers (*21*). Deeper shotgun sequencing was undertaken at the BGI sequencing facilities in Shenzhen, China on the BGI libraries. Additional Illumina sequencing was performed on the original Illumina shotgun sequencing libraries at the Science for Life Laboratory sequencing facilities in Stockholm, Sweden on the HiSeq X platform.

Data Preparation

Ancient genomes

All sequenced reads were mapped to the CanFam3.1 reference genome (*22*) using BWA aln (*23*, *24*) after adapters were removed and paired reads were merged with AdapterRemoval (*25*). The

mapped bam files were filtered with samtools and duplicates were removed with picard-tools mark duplicates (http://broadinstitute.github.io/picard/). Following duplicate removal, bam files corresponding to the same sample from different sequencing platforms were merged together using SAMtools (*26*). These reads were then re-mapped to VulVul2.2 (red fox assembly; accession: GCA_001887905.1) also using BWA. Quality control was performed using QualiMap v2.2.1 (*27*), 20 samples with at least 0.1x coverage of the nuclear genome were analysed in the downstream analyses.

Modern genomes

Raw reads from modern samples were downloaded from public repositories (Dataset S2) and aligned to canFam3.1 (dog reference genome) and VulVul2.2 (red fox assembly; accession: GCA_001887905.1) using BWA mem (*24*), with a realignment step as implemented in GATK (*28*).

Pseudo-haploidisation

Pseudo-haploid calling was performed on all samples generated in this study and the comparative reference panel using the -doHaploCall utility in ANGSD (*29*). We obtained SNPs coordinates, previously genotyped (67,850,544 SNPs) from a publically available VCF provided by the NHGRI Dog Genome Project (*30*). These SNP were extracted by ANGSD during the pseudo-haploid calling, all other SNPs were excluded from the dataset for the samples when aligned to the CanFam3.1 reference genome. During the pseudo-haploid calling, random bases were sampled with a minimum mapping quality of 20 and minimum base call of 20, transitions were discarded, genotypes were only recorded for sites which had at least 3x coverage, and the first and last five bases were trimmed to remove deaminated sites in ancient and historical samples.

angsd -doHaploCall 1 -doCounts 1 -minMapQ 20 -minQ 20 -minInd 1 -setMinDepth 1 -b bamlist.txt -minMinor -trim 5 -noTrans 1 -out SiberianDogs_Haploid -checkBamHeaders 0

The ANGSD output was converted to PLINK format with the haploToplink utility from ANGSD. In PLINK the dataset was filtered to remove CpG islands and filtered for linkage disequilibrium by removing SNPs with an R2 value greater than 0.1 with any other SNP within a sliding window of 50-SNPs and advanced by 10 SNPs each time (*31*). Sites were retained if covered by at least 75% of individuals for downstream analyses. This filtering resulted in the retention of 4,177,995 SNPs.

Data Analysis

PCA Analyses

PCAs (principal component analysis) were generated from the pseudo-haploid dataset using smartPCA (*32*, *33*). Ancient samples were projected onto two PCAs: Run-1 contained all canids in the reference panel including dogs and wolves while Run-2 contained only dogs.

Neighbour Joining Tree Construction

Neighbour joining trees were built from an identity-by-state matrix generated with PLINK (*34*, *35*) after the previously described filtering with samples with a minimum mean coverage of 0.5x. Only the sites genotype data for at least 75% of individuals were retained for the neighbour joining tree construction (4,177,995 SNPs). The neighbour joining trees were built using the black-backed jackal as an outgroup and 100 replicate trees were generated each using a random combination of 1,000,000 of the SNPs. The bootstrap support for each node of the tree was calculated by BOOSTER (*36*) using these 100 replicates.

Pairwise Distances

Identity-by-state (IBS) pairwise distances were calculated for each of the Siberian and Arctic dog using the filtered 4,177,995 sites in plink. The pairwise distances were plotted on maps according to the geographic origin of the comparative sample, intrasite comparisons were excluded from the maps, this was true for individuals from (Veretye and Ust'-Polui), see Fig. S4a-c. The IBS values were also plotted in ascending order for each ancient and historical Siberian dog (11,000 to 150 years ago) compared to the other ancient Siberian dogs and several ancient European and Near Eastern dogs with at least 0.5x coverage, see Fig. S4d-e. IBD for modern Arctic dogs (50 years ago to present) were plotted for ancient Siberian, European, and Near Eastern dogs, as well as historical Siberian dogs with at least 0.5x coverage, see Fig. S4f.

D-Statistics with qpDstat

D-statistics (D-stats) were calculated with qpDstat in AdmixTools (*37*) using the black-backed jackal as an outgroup. For investigation of West Eurasian ancestry in Arctic dogs D-stats were calculated of the form D(black-backed jackal, source; target, sister population). Source populations tested were represented by modern Portugese Village, ancient Near Eastern, ancient Europe, modern African dogs, and New Guinea Singing Dogs and the sister group used for comparison was the Zhokhov ancient Arctic dog or another Siberian/Steppe dog where stated. D-statistic calculations were based on 2,115,941 sites which had genotype information from at least 90% of individuals.

D-stats were also used to investigate wolf ancestry in all of the dogs sampled also using qpDstat in AdmixTools. The set of 2,115,941 filtered to only include sites with genotypes for at least 90% of individuals was used to calculate D-statistics. The black-backed jackal was set as an outgroup in the formula D(black-backed jackal, wolf; target, sister population). Individual wolves were tested as source populations for wolf ancestry representing Pleistocene wolves (n=2), modern Siberian wolves (n=2), modern European wolves (n=2), a modern Chinese wolf (n=1), and a Tibetian wolf (n=1).

Table S1, TreeMix sample list:

List of samples and their population names included in the analysis.

Sample:	**Population:**
HXH	Ancient Europe
Newgrange	Ancient Europe
Parknabinnia	Ancient Europe
TepeGhela	Levant 2.3kya
ASHQ01	Levant 2.3kya
ASHQ06	Levant 2.3kya
VillageDog_Borneo1	Modern Asia
VillageDog_Borneo2	Modern Asia
VillageDog_Borneo3	Modern Asia
Ishkinino	Steppe 3.2kya
Baikal1	Baikal 6.9kya
GSD_Aasiat1	Greenland Sled Dogs
GSD_Ilulissat1	Greenland Sled Dogs
GSD_Tasiilaq2	Greenland Sled Dogs
Yana1	Siberia Historical 0.1kya
Sakhalin1	Siberia Historical 0.1kya
Chuchki1	Siberia Historical 0.1kya
SiberianHusky1	Siberian Huskies
SiberianHusky2	Siberian Huskies
SiberianHusky3	Siberian Huskies
Veretye1	Veretye 10.9kya
Veretye2	Veretye 10.9kya
WolfSiberia1	Wolf Modern
WolfSiberia2	Wolf Modern
WolfChina11	Wolf Modern
CGG23	Wolf Pleistocene
Ust-Polui2	Iamal 2.0kya
Ust-Polui3	Iamal 2.0kya
Ust-Polui4	Iamal 2.0kya
Ust-Voikar	Iamal 1.0kya
Tiutei-Sale1	Iamal 1.0kya
Zhokhov	Zhokhov 9.5kya
PortauChoix	N. America 4.0kya
C.mesomelas	Jackal

TreeMix Analyses

TreeMix (*38*, *39*) was used to infer population splits and admixture events for the populations. Each population used for the TreeMix analyses contained one to three individuals depending on the available samples in the study and reference panel. Filtering was performed on the dataset before running TreeMix to remove all sites with missing data, resulting in the retention of 104,681 sites. TreeMix analyses were run with 100 bootstrap replicates over 500 blocks. Up to eight migration edges were modelled, the plotted replicate for each edge model was chosen based on the near mean residual score and the consistency of the modelled edges compared to other replicates. The individuals chosen as representatives of each population included in the analysis are listed in the following table (Table S1).

Table S2, Admixturegraph backboness:
List of samples forming the backbones used in the Admixturegraph analysis.

Backbone ID	Arctic Representative	Near East representative	European representative	Outgroup
backbone1	Zhokhov	TepeGhela	Newgrange	C.mesomelas
backbone2	Zhokhov	ASHQ01	Newgrange	C.mesomelas
backbone3	Zhokhov	TepeGhela	HXH	C.mesomelas
backbone4	Zhokhov	ASHQ01	HXH	C.mesomelas

Admixturegraph

We developed a wrapper for the Admixturegraph R package (*40*), using the snakemake workflow manager (*41*), to efficiently test similar models (backbone) across multiple samples (target). The workflow consists of three modules and relies on a series of conda environments which ensure the reproducibility of the results and the scalability of the analysis. The first module (subsetting.smk) requires a ped and map file as input. Here we used the file produced by ANGSD and filtered via PLINK as described in the pseudo-haploidisation section above with an extra filtering step based on minor allele frequencies (--maf 0.05) which resulted in a panel comprising 1,050,169 SNPs. For each combination of target/backbone, specific individuals are extracted from the ped using PLINK. These .map and .ped files serve as input for the second module (dstats.smk) which convert them in eigenstrat format using the Admixtools *convertf* utility and calculate the D-statistics for all possible quadruplets using qpDstats. The last module (modelling.smk) generates all input files for Admixgraph (R package), which is then employed to fit a series of predefined models. Our wrapper produces multiple outputs including the goodness of fit for each target/backbone pair as pdfs, and a heatmap representing the number of outliers under each backbone/target combination (e.g. Fig. S6b; Fig. S6c). A full description of the

workflow, the set of rules defining each module, as well as the python and R scripts used are available here https://github.com/a-karma/Arctic_Dogs.git.

We use this method to test multiple models using different backbones (see Table S2). For each backbone/target (all Siberian genomes) combination we tested a set of 6 models (Fig. S6a) and we determined the number of outliers. A D-statistic is considered an outlier if its expected value under a given model falls outside the 99% confidence interval centered around the observed statistic.

F4-ratio test

Admixture proportions were estimated using the qpF4ratio module of AdmixTools (*37*).

First we extracted the relevant individual from the pseudo-haplodized panel filtered panel (1,050,169 SNPs) described in the section above using plink. We then converted these output file in eigenstrat format by running the following commands:

plink1.9 --file arctic_dog_master --keep f4_ratio_keep_list.txt --dog --recode --out f4_ratio_run

convertf -p f4_ratio_run_convertf.par

Finally we perform the test by running:

qpF4ratio -p f4_ratio_run.par > f4_ratio_test.log

A consistent set of parameters was employed for all the analyses. An example of the *f4_ratio_run.par* is reported below.

genotypename: f4_ratio_run.eigenstratgeno
snpname: f4_ratio_run.snp
indivname: f4_ratio_run.ind
popfilename: f4_ratio_pop_list.txt
blgsize: 0.005

qpBrute

We performed an heuristic exploration of the admixture graph space using qpBrute (44), a python based utility that automates the building and fitting of admixture models using qpGraph (Admixtools)(*37*). The utility works in a stepwise fashion, by adding a new leaf at each iteration, until it exhausts the leaf nodes in the population list. When a new node can not be inserted on a subgraph without producing outliers (i.e. |Z| >=3 for the f-statistic) that sub-graph is discarded.

We conducted three separate analyses to assess:

A) continuity between Zhokhov and Baikal dogs

B) continuity between Iron Age and Medieval dogs from the Iamal peninsula

C) the ancestry of modern arctic breeds.

For each analysis we set up multiple runs which differ in the representative of the Ancient European population and/or the representative of the Ancient Arctic lineage. The list of samples and populations considered for each run are reported in Table S3.

In order to complete each run we first preprocessed the data by extracting the selected individuals from the pseudo-haplodized filtered panel of 1,050,169 SNPs, using the following PLINK command:

plink1.9 --file arctic_dogs_master --keep keep_list_qpbrute_run.txt --dog --recode --out qpbrute_run

We then converted the *qpbrute_run.map* and *qpbrute_run.ped* files using the Admixtools convertf utility and specifying the packedancestrymap as the desired output format in the parameter file for convertf:

convertf -p qpbrute_run_convertf.par

Finally, we modified the .ind output file manually by adding the desired population names and run the following commands:

conda activate qpbrute

python qpbrute.py --par qpbrute_run.par --prefix run_name_adm_graphs --pops list_of_population_names_involved_separated_by_a_space --out Out --threads 32

The file (*qpbrute_run.par*) contains all the parameters required by qpGraph. The following set of parameters was consistently employed through all the analyses:

indivname: run_name.ind
snpname: run_name.snp
genotypename: run_name.packedancestrymapgeno
outpop: NULL
useallsnps: YES
blgsize: 0.005
lsqmode: YES
diag: .0001
hires: YES

Table S3a, Samples involved in the qpbrute analysis A:

Table of genome and population names used in each qpbrute run.

Sample ID	Population Name	Run A1	Run A2	Run A3	Run A4	Run A5	Run A6
Zhokhov	Arctic	yes	yes	yes	yes	yes	yes
Baikal1	Baikal	yes	yes	no	no	yes	yes
Baikal2	Baikal	yes	yes	yes	yes	no	no
Baikal3	Baikal	yes	yes	yes	yes	no	no
Ishkinino	Steppe	yes	yes	yes	yes	yes	yes
Samara1	Steppe	yes	yes	yes	yes	yes	yes
ASHQ01	Levant	yes	yes	yes	yes	yes	yes
ASHQ08	Levant	yes	yes	yes	yes	yes	yes
ASHQ06	Levant	yes	yes	yes	yes	yes	yes
HXH	A_EU	yes	no	no	yes	yes	no
Parknabinnia	A_EU	no	yes	yes	no	no	yes
Newgrange	A_EU	yes	yes	yes	yes	yes	yes
C.mesomelas	Outgroup	yes	yes	yes	yes	yes	yes

Table S3b, Samples involved in the qpbrute analysis B:

Table of genome and population names used in each qpbrute run.

Sample ID	Population Name	Run B1	Run B2	Run B3	Run B4
Zhokhov	Arctic	yes	yes	yes	yes
Baikal2	Arctic	no	no	yes	yes
Baikal3	Arctic	no	no	yes	yes
Ishkinino	Steppe	yes	yes	yes	yes
Samara1	Steppe	yes	yes	yes	yes
ASHQ01	Levant	yes	yes	yes	yes
ASHQ08	Levant	yes	yes	yes	yes
ASHQ06	Levant	yes	yes	yes	yes
HXH	A_EU	no	yes	yes	no
Parknabinnia	A_EU	yes	no	no	yes
Newgrange	A_EU	yes	yes	yes	yes
Ust-Polui2	Iamal_Iron	yes	yes	yes	yes
Ust-Polui3	Iamal_Iron	yes	yes	yes	yes
Ust-Polui4	Iamal_Iron	yes	yes	yes	yes
Tiutei-Sale1	Iamal_Med	yes	yes	yes	yes
Ust-Voikar	Iamal_Med	yes	yes	yes	yes
C.mesomelas	Outgroup	yes	yes	yes	yes

Table S3c, Samples involved in the qpbrute analysis C:

Table of genome and population names used in each qpbrute run.

Sample ID	Population Name	Run C1	Run C2	Run C3	Run C4	Run C5
Zhokhov	Arctic	yes	yes	yes	yes	yes
Baikal2	Baikal	no	no	yes	no	no
Baikal3	Baikal	no	no	yes	no	no
Ishkinino	Steppe	yes	yes	yes	yes	yes
Samara1	Steppe	yes	yes	yes	yes	yes
ASHQ01	Levant	yes	yes	yes	yes	yes
ASHQ08	Levant	yes	yes	yes	yes	yes
ASHQ06	Levant	yes	yes	yes	yes	yes
HXH	A_EU	yes	no	no	no	no
Parknabinnia	A_EU	no	yes	yes	yes	yes
Newgrange	A_EU	yes	yes	yes	yes	yes
VillageDog_Borneo1	East_Asia	yes	yes	yes	no	no
VillageDog_Borneo2	East_Asia	yes	yes	yes	no	no
VillageDog_Borneo3	East_Asia	yes	yes	yes	no	no
NewGuineaSingingDog1	East_Asia	no	no	no	yes	yes
NewGuineaSingingDog2	East_Asia	no	no	no	yes	yes
NewGuineaSingingDog3	East_Asia	no	no	no	yes	yes
SiberianHusky1	Husky	yes	yes	yes	yes	no
SiberianHusky2	Husky	yes	yes	yes	yes	no
SiberianHusky3	Husky	yes	yes	yes	yes	no
GSD_Ilulissat1	GSD	no	no	no	no	yes
GSD_Qaanaaq1	GSD	no	no	no	no	yes
GSD_Tasiilaq2	GSD	no	no	no	no	yes
C.mesomelas	Outgroup	yes	yes	yes	yes	yes

Testing for Reference Bias

The aim here was to compare D-statistics computed from alignment to the dog (CanFam3.1) and fox (VulVul2.2) reference genomes. To do so, we first pseudo-haplodise the data aligned to the fox reference genome applying the following filtering thresholds using ANGSD: random bases were sampled with a minimum mapping quality of 20 and minimum base call of 20, transitions were discarded, initial and final three bases were trimmed to remove deaminated sites, and included only sites with coverage for at least 40 of the 47 (85.1%) individuals. These filters resulted in the retention of 978,502,332.

angsd -doHaploCall 1 -doCounts 1 -minMapQ 20 -minQ 20 -minInd 40 -setMinDepth 1 -b bamlist.txt -minMinor -trim 5 -noTrans 1 -out Sib_vulpes_40 -checkBamHeaders 0

Further filtering was performed before downstream analyses took place to retain only sites with at least a minor allele frequency of 5% and with data present for at least 75% of individuals. Following this filtering 6,912,616 SNPs were used in downstream analyses. D-statistics (D-stats)

were then calculated with qpDstat in AdmixTools (*37*) using the black-backed jackal as outgroup based.

The doAbbababa utility of ANGSD was also run to calculate D-statistics from the bam files mapped to the dog and fox reference genomes in blocks of 100,000 sites and quality filters.

angsd -doAbbababa 1 -doCounts 1 -rmTrans 1 -blockSize 100000 -bam bamlist.txt -useLast 1 -minQ 20 -minMapQ 20 -doCheck 0 -out <reference>

In order to compare the results from the admixture graph modelling analysis and f4-ratio analysis, we processed the angsd output (978,502,332 SNPs) in plink and applied the same filtering procedure by running:

plink1.9 --tfile ANGSD_output --make-bed --dog --missing-genotype N --output-missing-genotype 0 --out Sib_vulpes_40_haplo --allow-extra-chr --memory 100000

plink1.9 --bfile Sib_vulpes_40_haplo --allow-extra-chr --indep-pairwise 50 10 0.1 --threads 10 --memory 100000 --make-bed --out Sib_vulpes_40_to_prune

plink1.9 --bfile Sib_vulpes_40_to_prune --allow-extra-chr --extract Sib_vulpes_40_to_prune.prune.in --out Sib_vulpes_haplo_pruned --memory 100000 --make-bed

plink1.9 --bfile Sib_vulpes_haplo_pruned --allow-extra-chr --maf 0.05 --geno 0.25 --recode --out Sib_vulpes_haplo_pruned_maf5_geno25

We then renamed the contigs using a custom script (also available in the Utils folder of our wrapper) and extract the first 95 contigs by running:

plink1.9 --file Sib_vulpes_pruned_renamed --allow-extra-chr --chr-set 95 --chr 1-95 --recode --out Sib_vulpes_pruned_renamed_95_contigs

After these filtering steps, 4,430,009 SNPs were retained.

Supplementary Text

F4-ratio test

We calculated the f4-ratio to estimate the admixture proportions in the two Veretye dogs as:

f4(ASHQ01, Out; Veretye1, Zhokhov) : f4(ASHQ01, Out; TepeGhela, Zhokhov) = 0.347±0.034

f4(ASHQ01, Out; Veretye2, Zhokhov) : f4(ASHQ01, Out; TepeGhela, Zhokhov) = 0.290±0.031

f4(TepeGhela,Out; Veretye1, Zhokhov) : f4(TepeGhela, Out; ASHQ01, Zhokhov) = 0.290±0.031

f4(TepeGhela, Out; Veretye2, Zhokhov) : f4(TepeGhela, Out; ASHQ01, Zhokhov) 0.250±0.028

Values are all significant (as expected). The Near Eastern ancestry component of Veretye1 is slightly higher but this could also be the result of differences in coverage.

We estimated the proportion of Near Eastern ancestry in Steppe dogs using the f4-ratio test.

f4(Levant, Out; Steppe, Arctic) : f4(Levant, Out; Iran, Arctic) = 0.569±0.027

f4(Iran, Out: Steppe, Arctic) : f4(Iran, Out; Levant, Arctic) = 0.574±0.027

Where: Iran = TepeGhela; Levant = ASHQ08 and ASHQ01; Out = C.mesomelas; Steppe = Ishkinino and Samara1; Arctic = Zhokhov.

We used the f4-ratio to confirm previous findings and estimate the contribution of the Arctic lineage to the ancestry of Ancient European dogs (treated as a single population, including both HXH and Newgrange).

f4(Zhokhov, Out; Ancient EU, Levant) : f4(Zhokhov, Out; Baikal, Levant) = 0.336±0.013

f4(Baikal, Out; Ancient EU, Levant) : f4(Baikal, Out; Zhokhov, Levant) = 0.329±0.013

Where: Ancient EU = HXH, Newgrange, and Parknabinnia; Levant = ASHQ01, ASHQ06, and ASHQ08; Baikal = Baikal1, Baikal2, and Baikal3; Out = C.mesomelas.

We investigated this further by considering a single individual per population, and attempted to quantify whether the Arctic ancestry component of European dogs varies across individuals and whether it is closer to Baikal1 or Zhokhov.

f4(Zhokhov, Out; HXH, ASHQ01) : f4(Zhokhov, Out; Baikal1, ASHQ01) = 0.411±0.019

f4(Baikal1, Out; HXH, ASHQ01) : f4(Baikal1, Out; Zhokhov, ASHQ01) = 0.453±0.021

f4(Zhokhov, Out; Newgrange, ASHQ01) : f4(Zhokhov, Out; Baikal1, ASHQ01) = 0.288±0.019

f4(Baikal1, Out; Newgrange, ASHQ01) : f4(Baikal1, Out; Zhokhov, ASHQ01) = 0.312±0.021

f4(Zhokhov, Out; Parknabinnia, ASHQ01) : f4(Zhokhov, Out; Baikal1, ASHQ01) = 0.241±0.020

f4(Baikal1, Out; Parknabinnia, ASHQ01) : f4(Baikal1, Out; Zhokhov, ASHQ01) = 0.241±0.021

The Arctic ancestry component of European dogs indeed varies and seems to decrease over time. These results are generally consistent with the IBS pairwise distances analysis and the modelling with Admixturegraph (R package): amongst Ancient European dogs, HXH is the closest to Arctic dogs and the Zhokhov dog seems to be closer to the source of the Arctic component in European dogs.

Finally, we attempted to estimate the proportion of european ancestry in Baikal dogs (if any).

f4(HXH, Out; Baikal1, Zhokhov) : f4(HXH, Out; Newgrange, Zhokhov) = 0.104±0.026

f4(Newgrange, Out; Baikal1, Zhokhov) : f4(Newgrange, Out; HXH, Zhokhov) = 0.070±0.022

f4(HXH, Out; Baikal2, Zhokhov) : f4(HXH, Out; Newgrange, Zhokhov) = -0.069±0.048

f4(Newgrange, Out; Baikal2, Zhokhov) : f4(Newgrange, Out; HXH, Zhokhov) = -0.054±0.041

f4(HXH, Out; Baikal3, Zhokhov) : f4(HXH, Out; Newgrange, Zhokhov) = 0.006±0.045

f4(Newgrange, Out; Baikal3, Zhokhov) : f4(Newgrange, Out; HXH, Zhokhov) = 0.011±0.038

None of the values for the two oldest Baikal dogs is significantly different from zero while in the case of the youngest Baikal dog (Baikal1) we estimated a non-negligible admixture fraction. These results should be interpreted with caution though, given that the representatives of the Ancient European population are themselves admixed. Hence, we used two representatives of the Near eastern population as a proxy to estimate the non-Arctic component in Baikal dogs which confirmed the continuity of Baikal2 and Baikal3 as well as the attraction of Baikal1 towards Europe/Near East.

f4(ASHQ01, Out; Baikal1, Zhokhov) : f4(ASHQ01, Out; TepeGhela, Zhokhov) = 0.078±0.035

f4(TepeGhela, Out; Baikal1, Zhokhov) : f4(TepeGhela, Out; ASHQ01, Zhokhov) = 0.126±0.031

f4(ASHQ01, Out; Baikal2, Zhokhov) : f4(ASHQ01, Out; TepeGhela, Zhokhov) = -0.082±0.079

f4(TepeGhela, Out; Baikal2, Zhokhov) : f4(TepeGhela, Out; ASHQ01, Zhokhov) = -0.023±0.060

f4(ASHQ01, Out; Baikal3, Zhokhov) : f4(ASHQ01, Out; TepeGhela, Zhokhov) = -0.009±0.069

f4(TepeGhela, Out; Baikal3, Zhokhov) : f4(TepeGhela, Out; ASHQ01, Zhokhov) = 0.023±0.063

Qpbrute

We assessed whether including different combinations of Baikal genomes affected our conclusions that there was potentially an influx of Western Ancestry in these dogs between 6.9 and 7.5kya. To do so we ran qpbrute with different combinations of samples, 1) three Baikal dogs as a single population (Baikal_A1 and Baikal_A2), 2) using only the two oldest samples (Baikal_A3 and Baikal_A4), 3) using only the youngest / higher coverage sample (Baikal1) (Baikal_A5 and Baikal_A6). Visual inspections of the graphs leaving no outliers revealed that there are not substantially more models involving admixture from Western sources into the high coverage Baikal genome than in lower genomes.

In the case of Iamal dogs, although we observed a general increase in the number of fitting models, the addition of the two oldest Baikal dogs to the ancient Arctic population had little effect on their overall topologies. Most discrepancies were observed when using different

representatives of the European population (see Table S3b and Table S4). This effect of including different representatives for European dogs was even more pronounced in the Siberian dog analysis (see Siberian_C1 and Siberian_C2). The choice of the representative of the East Asan population (village dogs from Borneo or New Guinea singing dogs), however, had little effect (compare Siberian_C2 and Siberian_C4): fitting models are all variations of the first three graphs reported in Fig S7. These results are broadly consistent with the TreeMix analysis and represent three different solutions to resolve a polytomy at the root of dog phylogeny.

For each analysis, few chosen graphs were plotted in Fig. S7, Fig. S9, and Fig. S11. All other graphs can be found here: https://sid.erda.dk/share_redirect/DYIqytfNSR. A summary of each qpbrute run is reported in the following table (Table S4).

Table S4, qpbrute results summary:

The table reports the number of tested and fitting graphs for each run. Fitting models are categorised based on the number of admixture events (a) present in the graph.

Run ID	# Fitting, a=2	# Fitting, a=3	# Fitting, a=4	# Tested Graphs	% Fitting Graphs
Baikal_A1	4*	336	0	11,155	3,05%
Baikal_A2	4*	412	0	12,001	3,47%
Baikal_A3	0	428	0	14,676	2.92%
Baikal_A4	0	469	0	15,413	3,04%
Baikal_A5	4	394	0	13,448	2,96%
Baikal_A6	4	391	0	11,477	3,44%
Iamal_B1	0	3	330	61,744	0,54%
Iamal_B2	0	0	15	39,868	0,04%
Iamal_B3	0	0	20	40,718	0.05%
Iamal_B4	0	0	563	77,282	0.73%
Siberian_C1	0	0	1	40,527	0.0025%
Siberian_C2	0	0	94	55,603	0,17%
Siberian_C3	0	0	176	66,443	0,26%
Siberian_C4	0	0	124	60,141	0.21%
Siberian_C5	0	0	164	72,726	0.23%

* Two models were topologically identical.

Assessing Reference Bias in ancient Baikal genomes

Firstly, to test for reference bias we built a simple neighbour joining tree to compare the phylogenetic relationships seen in the dataset when aligned to the dog reference to the alignment of the data for the same samples to the red fox reference. The general topology and relationships

between samples in the neighbour joining tree largely stayed the same regardless of the reference (Fig. S12). We further investigate the reference bias effect by comparing D-statistics computed using alignment to the red fox and dog reference genome. Firstly, we tested whether the signal of gene flow from non-Arctic populations in early Holocene Baikal dogs could be driven by reference bias using ANGSD (Fig. S10). D-statistics, computed using the both Zhokhov and Baikal1 genomes aligned to the dog reference suggest that Baikal1 possess European like ancestry, which could be driven by the fact the reference genome is a European breed (boxer) (Fig. S10a). This signal disappears when computing the same combinations using the red fox reference genome suggesting that it is potentially driven by reference bias in this low coverage genome from near Lake Baikal (Fig. S10a). In fact while the signal from Steppe dogs and Near Eastern dogs becomes less significant, the test now reports more shared derived allele between Zhokhov and European/East Asian dogs. This suggests that, 1) there has been limited additional admixture from the Near East/Steppe into the Baikal lineage, 2) the Zhokhov dog belonged to a lineage that was closer to the source of Arctic ancestry in European/Asian dogs. It remains possible, however, that admixture from European/East Asian lineage in the Zhokhov lineage could also explain this - although given the age and location of this sample this scenario seems less likely. When D-statistics are generated on the pseudo-haploid dataset aligned to the VulVul2.2 reference using qpDstat a shift in statistically significant signals is also apparent (compare Dataset S3 and Dataset S4). Particularly in the case of Baikal1 there are several D-statistics that became significant when aligned to VulVul2.2 (ie. ancient European and Near Eastern dogs), while several lost their significance (ie. modern Europe and East Asia) (Dataset S4).

Assessing Reference Bias in admixture graph analyses

We performed the same admixture modelling analysis (as described in the Admixturegraph section above) using alignments to the red fox reference genome (VulVul2.2). The D-statistics were computed using all transversions from the largest 95 contigs (4,930,239 SNPs in total). Besides for a few minor differences (see below) the best models identified were the same between the two assemblies (see Fig. S6b; Fig. S6c). The exceptions were:

i) Nenets1: The red fox based analysis revealed a stronger affinity to European dogs.

ii) Bolgar1: The fit of models involving Near East admixture in this sample is reduced using the redfox assembly.

iii) Steppe dogs (Ishkinino and Samara1): we observed a general increase in the number of outliers in all models tested, with no models leaving no outliers when using alignment to the red fox assembly. In the case of Ishkinino the same models were preferred in both analyses while in the case of Samara1 the model involving admixture between Arctic and Near East was slightly preferred using the red fox assembly as reference.

Assessing Reference Bias in f4-ratio tests

We observed almost no differences when estimating the Near East ancestry component of Steppe dogs with f4-ratio.

f4(Levant, Out; Steppe, Arctic) : f4(Levant, Out; Iran, Arctic) = 0.594±0.030

f4(Iran, Out: Steppe, Arctic) : f4(Iran, Out; Levant, Arctic) = 0.70±0.023

Where: Iran = TepeGhela; Levant = ASHQ08 and ASHQ01; Out = C.mesomelas; Steppe = Ishkinino and Samara1; Arctic = Zhokhov.

The estimated Arctic ancestry component of Ancient European dogs is slightly higher and seems much closer to the Zhokhov lineage.

f4(Zhokhov, Out; Ancient EU, Levant) : f4(Zhokhov, Out; Baikal, Levant) = 0.358±0.015

f4(Baikal, Out; Ancient EU, Levant) : f4(Baikal, Out; Zhokhov, Levant) = 0.463±0.015

Where: Ancient EU = HXH, Newgrange, and Parknabinnia; Levant = ASHQ01, ASHQ06, and ASHQ08; Baikal = Baikal1, Baikal2, and Baikal3; Out = C.mesomelas.

The European/Near Eastern ancestry component in the two oldest Baikal dogs is still not significantly different from zero while in the case of Baikal1, the estimated admixture fraction increased substantially.

f4(Newgrange, Out; Baikal1,Zhokhov) : f4(Newgrange, Out; HXH, Zhokhov) = 0.33±0.016

f4(Newgrange, Out; Baikal2, Zhokhov) : f4(Newgrange, Out; HXH, Zhokhov) = -0.033±0.023

f4(Newgrange, Out; Baikal3, Zhokhov) : f4(Newgrange, Out; HXH, Zhokhov) = -0.026±0.021

f4(ASHQ01, Out; Baikal1, Zhokhov) : f4(ASHQ01, Out; TepeGhela, Zhokhov) = 0.198±0.034

f4(ASHQ01, Out; Baikal2, Zhokhov) : f4(ASHQ01, Out; TepeGhela, Zhokhov) = -0.021±0.050

f4(ASHQ01, Out; Baikal3, Zhokhov) : f4(ASHQ01, Out; TepeGhela, Zhokhov) = -0.077±0.054

Assessing Reference Bias from Pleistocene wolves

Reference bias also affected our power to detect gene flow from Pleistocene wolf into the Zhokhov dog as reported in a previous study (*42*). Here we show that using the red fox reference genome dramatically increases our power to detect gene flow from Pleistocene wolves into the Zhokhov lineage (Fig. S10b).

Limited Gene Flow from Wolves into other Siberian Dogs

Aside from the clear admixture signal from Pleistocene wolves into the Zhokhov lineage (see above), we found very limited evidence for introgression from both modern or Pleistocene wolves into Siberian. Several Siberian dogs gave statistically significant signals of Pleistocene wolf ancestry in their D-statistics when compared to village dogs from across India, East Asia, and Southeast Asia. The strongest signal for gene flow from Pleistocene wolves can be seen in the low coverage Veretye dog, while three historical dogs from the Amur Delta, Sakha Republic, and Chukchi show low but significant signals of Pleistocene wolf ancestry. Of the modern Arctic dogs, the Greenland Sled Dogs show the strongest signal that appears with both of the Pleistocene wolf sources tested (Fig. S13a), based on the Eurasian distribution and the seemingly universal signal for Pleistocene wolf ancestry in Greenland Sled dogs in the D-statistics this must have happened in Siberia before the ancestors of the Greenland Sled Dog departed for North America. These results corroborate previous analyses (*42*). When modern wolves from Chukotka and China are tested as sources the Z score is again significant (Fig. S13b).

Previous analyses of modern and ancient genomes, using f4 and D-statistics showed that most wolves are symmetrically related to all possible pairs of modern and ancient dogs, which indicates that gene flow mostly took place from dogs into wolf populations (*43*). Our results corroborate this finding, as with the exception of a few cases, we found that most wolves were symmetrically related to various dog combinations (Fig. S13b). By and large the historical samples from across Siberia do not show signs of modern wolf ancestry, an exception is a dog from the Amur Basin (AmurDelta1) that has comparatively more alleles shared with modern wolves than dogs from the same region, some African village dogs, Asian village dogs, and some breed dogs.

Gene Flow between Siberian and East Asian Dogs

A recent study of ancient Eurasian dogs revealed that East Asian dogs, as represented by the New Guinea Singing Dog, are the result of an ancient admixture event between an Ancient Arctic dog and an undefined ancient lineage (*3*). Overall, our results corroborate this finding. In fact, several qpBrute models computed for this study include admixture event(s) between ancient Arctic dog lineage, represented by the Zhokhov dog, and an unidentified ancient lineage that gave rise to modern East Asian dogs (Fig. S7c). Furthermore, several of our TreeMix analyses show migration edges between the ancient Zhokhov and Lake Baikal dog lineages and modern East Asian dogs (Fig. S3b). These signals for ancient admixture from the Arctic lineage into the ancestor of East Asian dogs likely resulted in issues resolving the root between Western Eurasian, Arctic and East Asian dog lineages (*44*) as seen in the neighbour joining tree (Fig. S2).

D-statistics of the form D(outgroup, East Asia; Zhokhov, Siberian dog), computed based on alignment to the dog reference (CanFam3.1) indicated that several ancient and historical Siberian dogs (Baikal1, Sakhalin1, AmurDelta1) had statistically significant signals ($Z<-3.3$) for gene flow from East Asian dogs when a village dog from Borneo (VillageDog_Borneo1) and/or a New

Guinea Singing dog (NewGuineaSingingDog1) were used as sources (Dataset S3). This signal, however, disappeared when D-statistics were computed from the dataset using the fox reference (VulVul2.2)(Dataset S4), indicating a potential reference bias issue. Most D-statistics, of the form D(outgroup, East Asia; Zhokhov, Siberian dog), in both alignments, resulted in significant values (Z>3.3) that instead suggest that the East Asian dogs are closer to the ancient Zhokhov dogs than to later Siberian dogs (Dataset S3; Dataset S4). This could be driven by either or both of these scenarios: gene flow from Western dogs into Siberian dogs since their TMRCA with Zhokhov, and/or early gene flow from a dog lineage, best represented by the Zhokhov genome, into the ancestor of East Asian dog. Together these results indicate that Siberian dogs did not receive extensive gene flow from East Asian dogs.

Mitochondrial DNA

Previous studies have found that most Arctic and Siberian mitochondrial genomes from dogs are related to the A-clade (*44–47*). The mitochondrial genomes from the Siberian dogs sequenced in this study generally fell into the A-clade in the maximum likelihood tree constructed with 1,000 bootstrap replicates. Only the two dogs from Veretye in Karelia (Veretye1, Veretye2) were found together with the ancient Arctic dogs from Zhokhov (Fig. S14). The other four of the five mitochondrial genomes sequenced from dogs of the Ust'-Polui site on the Iamal Peninsula carry A-clade mitochondrial haplotypes, specifically with A1a clade haplotypes, which are typical for Siberia and the Arctic as well as other regions of Eurasia (Fig. S14). However, one of the Iron Age dogs from Ust'-Polui (Ust'-Polui3) possesses a mitochondrial genome with a C-clade haplotype that is typically associated with Europe before the Neolithic. The C-clade haplotype was likely introduced to the region when gene flow occurred between the ancestors of the Iron Age dogs of the Iamal Peninsula and dogs from the further south. All three medieval Iamal dogs possessed A-clade mitochondrial genomes that cluster closely with several of the earlier Iamal dogs from Ust'-Polui. Mitochondrial genomes from historical dogs associated with Nenets and Khanty groups in western Russia from the early twentieth century also cluster closely with the Iron Age and Medieval Iamal dogs. Falling basal to the Iamal dogs was the Bronze Age Steppe dog from Ishkinino. A historical dog Sakhalin1, collected 1892 CE from Sakhalin Island, also possesses a C-clade mitochondrial genome.

Supplementary Information Figure Captions

Fig. S1 PCA:

Principal component analyses (PCAs) of modern dogs/wolves with ancient individuals projected onto the PCA. PCA were computed with smartPCA using 4,177,994 SNPs, up to 25% missingness was allowed per site. The geographical origin of the individual is indicated by the colour and the age of the individual is indicated by the shape. A) PCA plotting PC1-4 of dogs and wolves. B) PCA with PC1-4 plotting only dogs..

Fig. S2 Neighbour Joining Tree with >0.5x coverage samples:

Neighbour joining tree built from identity-by-state matrix of samples with >0.5x mean coverage of the nuclear genome, rooted with the black-backed jackal as the outgroup. The outgroup and wolves present during the construction of the tree are not shown in the figure. Branch colour reflects the region and/or time point at which the dog was living.

Fig. S3 TreeMix Models:

TreeMix analyses run with up to four migration edges. All sites with missing data have been removed from the analysis and low coverage individuals have been included, each population contains between one and three individuals. Models testing for each number of edges were run in 50 replicates. A) i-v: TreeMix tree with edges 0-4, respectively showing most frequent topology and edges. B) i-v: TreeMix tree with edges 5-8, respectively showing most frequent topology and edges.

Fig. S4 Pairwise Distances for Siberian dogs:

A-C) Identity-by-state (IBS) pairwise distances plotted on maps showing the approximate geographical origin of samples and their pairwise distance. The shape of the icon on the map denotes the age of the sample while the colour reflects IBS with the darkest colours showing the greatest affinity and lightest colours showing the smallest affinity. The black icon on the map shows where the individual being tested as the target in each test originates from. For sites with more than one individual intra-site IBS values have been excluded from the plot to highlight relationships outside of the site. A) Maps of IBS for samples from ancient Siberian dogs from 11,000 to 800 years ago. B) Maps of IBS for Siberian dogs from 150 to 50 years ago. C) Maps of IBS for modern Arctic dogs from 50 years ago to present. D-F) Identity-by-state (IBS) pairwise distances for Siberian dogs compared to other ancient dogs. The individual with the greatest affinity is plotted in red and all other comparisons are plotted in ascending order. D) IBS plotted in ascending order for each ancient Siberian dog compared to ancient Siberian, Near East, and European dogs with >0.5X coverage. E) IBS plotted in ascending order for historical Siberian dogs compared to ancient Siberian, Near East, and European dogs with >0.5X coverage. F) IBS plotted in ascending order for historical Siberian dogs compared to ancient Siberian, Near East, and European dogs as well as historical Siberian dogs with >0.5X coverage.

Fig. S5 D-Statistics for Gene Flow into Steppe Dogs:

D-statistics calculated with the qpDstat utility in AdmixTools where the black-backed jackal is used as the outgroup, D(Jackal, Source; Ishkinino, Samara). The z-score was plotted and statistical significance for admixture occurs above 3.3 or below -3.3. Positive scores reflect gene flow between the jackal and the target population or the Eurasian dog and the sister population. Negative scores reflect gene flow between the jackal and the sister population or the Eurasian dog and the target. Six populations were tested as a sources for gene flow into the Bronze Age and Medieval Steppe dog: ancient Iran (TepeGhela), ancient Israel (ASHQ01, ASHQ06, ASHQ08), New Guinea Singing Dogs (NewGuineaSingingDog1, NewGuineaSingingDog2, NewGuineaSingingDog3), ancient Europe (Newgrange, Dog1_PU, HXH), modern Europe (VillageDog_Portugal1, VillageDog_Portugal2), and ancient Siberia (Zhokhov, Baikal1).

Fig. S6a Admixture graphs tested

Schematic representation of the admixture graphs fitted to the data

Fig. S6b Outliers Full Heat Map (samples aligned to the dog reference genome, Canfam3.1)

Heat map based on the number of D-stats outliers under each model (see Fig. S6a). The value are averaged across 4 different backbones each including an C.mesomelas (Outgroup), the Zhokhov dog (Ancient Arctic), an ancient European lineage (either HXH or Newgrange) and an ancient Near Eastern lineage (TepeGhela or ASHQ01)

Fig. S6c Outliers Full Heat Map (samples aligned to the fox reference genome, VulVul2.2)

Heat map based on the number of D-stats outliers under each model (see (see Fig. S6a). The value are averaged across 4 different backbones each including an C.mesomelas (Outgroup), the Zhokhov dog (Ancient Arctic), an ancient European lineage (either HXH or Newgrange) and an ancient Near Eastern lineage (TepeGhela or ASHQ01)

Fig. S7 qpBrute Model with East Asian Dogs:

Four representative graph that describes the relationship between ancient European dogs (HXH or Prknabinna and Newgrange), Steppe dogs (Samara1 and Ishkinino), Levant dogs (ASHQ01, ASHQ06, ASHQ08), Ancient Arctic dogs (Zhokhov, or Zhokhov, Baikal2 and Baikal3), three modern Husky dogs and three East Asian dog (New Guinea Singing dogs or Village dog from Borneo).

Fig. S8 D−statistics (qpDstat) Gene Flow into Iamal dogs:

Boxplots of z score for qpDstat D-statistics for allele sharing between Iamal dogs and other dog populations. A.) Aggregated z scores for all Iron Age Iamal (Ust-Polui) dogs compared to the Early Holocene Zhokhov dog and Medieval Iamal dogs: Iarte6, Ust-Voikar, and Tiutei-Sale1. Comparisons show that the Ust-Polui dogs have greater allele sharing with non-Arctic dogs than the Zhokhov dog, the Ust-Polui dogs also have more allele sharing with Ancient Arctic dogs (Zhokhov & Baikal) than the three Medieval Iamal dogs. B.) Intersite comparisons of allele sharing to non-Iamal dogs for Medieval Iamal dogs. There is no statistically significant signal for additional allele sharing from outside of the Iamal region for any of the Medieval dogs.

Fig. S9 qpBrute Model for Baikal dogs:

Three representative graph that describes the relationship between the Arctic lineage (Zhokhov dog), the three samples from Baikal, ancient European dogs (A_EU), Steppe dogs (Samara and Ishkinino) and Levant dogs (ASHQ01, ASHQ06, ASHQ08) treated as separate populations. The models are topologically similar and show that models without admixture from non-Artic lineage into Baikal are sufficient to explain their ancestry.

Fig. S10 D−statistics (ANGSD) Gene Flow Reference Bias

D-statistics calculated by ANGSD comparing alignment of samples to the dog reference genome (CanFam3.1) and the red fox genome (VulVul2.2) A. D-stat results testing for additional gene flow into Baikal1 compared to Zhokhov. B.) D-stats calculated for gene flow from Pleistocene wolves into the Zhokhov dog.

Fig. S11 qpBrute Model for Iamal dogs:

Four representative graph (one for each qpbrute run) that describes the relationship between ancient European dogs (A_EU), Steppe dogs (Samara and Ishkinino), Levant dogs (ASHQ01,

ASHQ06, ASHQ08) and Ancient Arctic dogs, Iamal Iron Age dogs (Ust'-Polui2, Ust'-Polui3, Ust'-Polui4), and Iamal Medieval dogs (Tiutei-Sale1, Ust'-Voikar) are treated as separate populations. All models show different gene flow events experienced by the Iron Age and Medieval dogs of the Iamal region.

Fig. S12 Vulpes Neighbour Joining Tree

Neighbour joining tree built from identity-by-state matrix of samples with >0.75x mean coverage of the nuclear genome aligned to the red fox (VulVul2.2), rooted with the black-backed jackal as the outgroup. The outgroup and wolves present during the construction of the tree are not shown in the figure. Branch colour reflects the region and/or time point at which the dog was living.

Fig. S13 D−statistics (qpDstat) Gene Flow from Wolves into Siberian/Arctic Dogs

D-statistics as calculated by qpDstat to test for gene flow from A.) Pleistocene wolves (n=2) and B.) modern wolves (n=4) into Siberian and Arctic dogs. The plotted results testing four different wolves as source populations are coloured by the dog sample being tested. The red dashed line marks -3.3 and +3.3 corresponding to the statistically significant Z-score. The sister population, Y, was tested as modern village and breed dogs, ancient West Eurasian dogs, and all Siberian dogs in the dataset, in D(Jackal, wolf; X, Y).

Fig. S14 Phylogeny of mitochondrial genomes

A) Maximum likelihood tree of mitochondrial genomes built consensus tree from 500 bootstraps replicates from RaxML not showing the outgroup, Canis latrans. Identified dog clades are colour coded, mitochondrial genomes from samples generated in this study are labelled in red. B) Larger visualization of wolves and dog clades B-E. C) Larger visualization of dog clade A and related canids.

Supplementary Information Dataset Captions

Dataset S1, Siberian dog samples:

Dataset of dog genomes generated for this study with information about age, location, and coverage.

Dataset S2, Panel of dog and wolf genomes:

Dataset of publically available genomes used in this study with sample metadata, such as age and location of origin.

Dataset S3, D-statistics for non-Arctic ancestry in Siberian dogs:

Table of D-statistics as calculated by qpDstat to test for ancestry in Arctic Eurasian dogs from other Eurasian dog populations with data aligned to the CanFam3.1 reference genome. Ten Eurasian dogs were tested to represent potential admixture source populations. The D-stat results highlighted in green indicate where the Z-score shows statistically significance with a Z-score for the corresponding D above 3.3 or below -3.3. The table also lists the name of the site

the sample was recovered from, the region where the site is located, the age of the sample in years before present, and the mean nuclear coverage of the genome.

Dataset S4, D-statistics for non-Arctic ancestry in Siberian dogs:

A table of D-statistics, from dataset aligned to the VulVul2.2 reference genome, calculated by qpDstat to test for ancestry in Arctic Eurasian dogs from other Eurasian dog populations. Eight Eurasian dogs were tested to represent potential admixture source populations. The D-stat results highlighted in green indicate where the Z-score shows statistically significance with a Z-score for the corresponding D below -3.3 that were significant for both the data aligned to the CanFam3.1 and VulVul2.2 references. Results highlighted in yellow indicate instances where results were significant when aligned to CanFam3.1 but not for VulVul2.2 and results highlighted in blue indicate results that were not significant when aligned to CanFam3.1 but are significant when aligned to VulVul2.2. The table also lists the name of the site the sample was recovered from, the region where the site is located, the age of the sample in years before present, and the mean nuclear coverage of the genome.

SI References

1. S. V. Oshibkina, The Mesolithic Culture Veret'ye. Chronology and Periods. *Rossijskaâ arheologiâ*, 100–110 (2004).

2. S. V. Oshibkina, *Mezolit Vostochnogo Prionezhya. Kultura Veretye [The Mesolithic of the Eastern Shore of Lake Onega. The Veretye Culture]* (Institute of Archaeology, Russian Academy of Sciences, Moscow, 2006).

3. A. Bergström, L. Frantz, R. Schmidt, E. Ersmark, O. Lebrasseur, L. Girdland-Flink, A. T. Lin, J. Storå, K.-G. Sjögren, D. Anthony, E. Antipina, S. Amiri, G. Bar-Oz, V. I. Bazaliiskii, J. Bulatović, D. Brown, A. Carmagnini, T. Davy, S. Fedorov, I. Fiore, D. Fulton, M. Germonpré, J. Haile, E. K. Irving-Pease, A. Jamieson, L. Janssens, I. Kirillova, L. K. Horwitz, J. Kuzmanovic-Cvetković, Y. Kuzmin, R. J. Losey, D. L. Dizdar, M. Mashkour, M. Novak, V. Onar, D. Orton, M. Pasarić, M. Radivojević, D. Rajković, B. Roberts, H. Ryan, M. Sablin, F. Shidlovskiy, I. Stojanović, A. Tagliacozzo, K. Trantalidou, I. Ullén, A. Villaluenga, P. Wapnish, K. Dobney, A. Götherström, A. Linderholm, L. Dalén, R. Pinhasi, G. Larson, P. Skoglund, Origins and genetic legacy of prehistoric dogs. *Science*. **370**, 557–564 (2020).

4. V. V. Tkachev, The Ishkinino Bronze Age Mining Complex in the Southern Urals: Radiocarbon Dating. *Archaeology, Ethnology & Anthropology of Eurasia*. **47** (2019), pp. 38–47.

5. V. V. Zaykov, A. M. Yuminov, V. V. Tkachev, Copper Mines, Chromite Copper Ores and Slags of the Ishkinino Bronze Age Archaeological Microregion, The Southern Urals. *Archaeology, Ethnology and Anthropology of Eurasia*. **40**, 37–46 (2012).

6. R. J. Losey, L. S. Fleming, T. Nomokonova, A. V. Gusev, N. V. Fedorova, S. Garvie-Lok, O. P. Bachura, P. A. Kosintsev, M. V. Sablin, Human and Dog Consumption of Fish on the Lower Ob

River of Siberia: Evidence for a Major Freshwater Reservoir Effect at the Ust'-Polui Site. *Radiocarbon*. **60**, 239–260 (2018).

7. R. J. Losey, T. Nomokonova, A. V. Gusev, O. P. Bachura, N. V. Fedorova, P. A. Kosintsev, M. V. Sablin, Dogs were domesticated in the Arctic: Culling practices and dog sledding at Ust'-Polui. *Journal of Anthropological Archaeology*. **51**, 113–126 (2018).

8. Fedorova, N.V., P.A. Kosintsev & W.W. Fitzhugh, "*Gone to the Hills": Culture of Iron Age Inhabitants of the Northwestern Shore of Iamal Peninsula* (Izd-vo Ekaterinburg, Ekaterinburg, 1998).

9. T. Nomokonova, R. J. Losey, A. V. Plekhanov, H. J. McIntyre, Iarte VI and Late Holocene Reindeer Remains from the Iamal Peninsula of Arctic Siberia. *Arctic Anthropol.* **55**, 56–75 (2018).

10. G. P. Vizgalov, A. V. Kardash, P. A. Kosintsev, T. V. Lobanova, Historical ecology of the population of the north of Western Siberia. *Nefteyugansk. Russian Federation. AMB Yekaterinburg* (2013).

11. G. F. Shafranov-Kutsev, *Iugoriia. Entsiklopediia Khanty-Mansiiskogo Avtonomnogo Okruga* (Tiumenskii Gosudarstvennyi Universitet, Khanty-Mansiisk, 2000), vol. 3 of *Iugry*.

12. Zykov A.P., Koksharov, S.F., in *Khanty-Mansiiskii Avtonomnyi Okrug v Zerkale Proshlogo*, I. A. Iakovlev, Ed. (Tomskii Gosudarstvennyi Universitet, Tomsk, 2003), vol. 1, pp. 110–136.

13. Morozov, V.M., Parkhimovich, S.G., in *Zapadnaia Sibir' v Drevnosti*, R. S. Vasilevskii, Ed. (Tiumen'skii Gisudarstvennyi Universitet, Tiumen, 1985), pp. 89–99.

14. D. Y. Badeev, V. Y. Koval, Studies of the Trade and Craft District of Medieval Bolgar. *Povolzhskaya Arkheologiya (The Volga River Region Archaeology)*. **1** (2018), pp. 270–289.

15. Y. L. V. Badeev D. Y., in *Kochevye Imperii Evrazii v Svete Arkheologicheskikh i Mezhdisciplinarnykh Issledovaniy (Nomad Empires of Eurasia in the Light of Archaeological and Multidisciplinary Studies)*, K. N. N. Bazarov B. V., Ed. (Ulan-Ude, Buryat Science Centre, Siberian Branch of the Russian Academy of Sciences, 2019), vol. 1, pp. 17–19.

16. T. Mathiassen, Knud Rasmussen's Sledge Expeditions and the Founding of the Thule Trading Station. *Geografisk Tidsskrift*. **37** (1934) (available at https://tidsskrift.dk/geografisktidsskrift/article/view/48004).

17. A. J. T. Harris, T. R. Feuerborn, M.-H. S. Sinding, J. Nottingham, R. Knudsen, A. Rey-Iglesia, A. L. Schmidt, M. Appelt, B. Grønnow, M. Alexander, G. Eriksson, L. Dalén, A. J. Hansen, K. Lidén, Archives of human-dog relationships: Genetic and stable isotope analysis of Arctic fur clothing. *Journal of Anthropological Archaeology*. **59**, 101200 (2020).

18. E. Ersmark, L. Orlando, E. Sandoval-Castellanos, I. Barnes, R. Barnett, A. Stuart, A. Lister, L. Dalén, Population Demography and Genetic Diversity in the Pleistocene Cave Lion (2015), , doi:10.5334/oq.aa.

19. M. T. P. Gilbert, L. P. Tomsho, S. Rendulic, M. Packard, D. I. Drautz, A. Sher, A. Tikhonov, L. Dalén, T. Kuznetsova, P. Kosintsev, P. F. Campos, T. Higham, M. J. Collins, A. S. Wilson, F. Shidlovskiy, B. Buigues, P. G. P. Ericson, M. Germonpré, A. Götherström, P. Iacumin, V. Nikolaev, M. Nowak-Kemp, E. Willerslev, J. R. Knight, G. P. Irzyk, C. S. Perbost, K. M. Fredrikson, T. T. Harkins, S. Sheridan, W. Miller, S. C. Schuster, Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science*. **317**, 1927–1930 (2007).

20. C. Carøe, S. Gopalakrishnan, L. Vinner, S. S. T. Mak, M. H. S. Sinding, J. A. Samaniego, N. Wales, T. Sicheritz-Pontén, M. T. P. Gilbert, Single-tube library preparation for degraded DNA. *Methods Ecol. Evol.* **9**, 410–419 (2018).

21. S. S. T. Mak, S. Gopalakrishnan, C. Carøe, C. Geng, S. Liu, M.-H. S. Sinding, L. F. K. Kuderna, W. Zhang, S. Fu, F. G. Vieira, M. Germonpré, H. Bocherens, S. Fedorov, B. Petersen, T. Sicheritz-Pontén, T. Marques-Bonet, G. Zhang, H. Jiang, M. T. P. Gilbert, Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing. *Gigascience*. **6**, 1–13 (2017).

22. 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. J. 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).

23. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013), (available at http://arxiv.org/abs/1303.3997).

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

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

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

27. K. Okonechnikov, A. Conesa, F. García-Alcalde, Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*. **32**, 292–294 (2016).

28. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

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

30. J. Plassais, J. Kim, B. W. Davis, D. M. Karyadi, A. N. Hogan, A. C. Harris, B. Decker, H. G. Parker, E. A. Ostrander, Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat. Commun.* **10**, 1489 (2019).

31. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

32. N. Patterson, A. L. Price, D. Reich, Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).

33. A. L. Price, N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, D. Reich, Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).

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

35. C. C. Shaun Purcell, *Plink1.9*, (available at www.cog-genomics.org/plink/1.9/).

36. F. Lemoine, J. -B. Domelevo Entfellner, E. Wilkinson, D. Correia, M. Dávila Felipe, T. De Oliveira, O. Gascuel, Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature*. **556** (2018), pp. 452–456.

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

38. J. K. Pickrell, N. Patterson, C. Barbieri, F. Berthold, L. Gerlach, T. Güldemann, B. Kure, S. W. Mpoloka, H. Nakagawa, C. Naumann, M. Lipson, P.-R. Loh, J. Lachance, J. Mountain, C. D. Bustamante, B. Berger, S. A. Tishkoff, B. M. Henn, M. Stoneking, D. Reich, B. Pakendorf, The genetic prehistory of southern Africa. *Nat. Commun.* **3**, 1143 (2012).

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

40. J. A. Cahill, I. Stirling, L. Kistler, R. Salamzade, E. Ersmark, T. L. Fulton, M. Stiller, R. E. Green, B. Shapiro, Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Mol. Ecol.* **24**, 1205–1217 (2015).

41. J. Köster, S. Rahmann, Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*. **28**, 2520–2522 (2012).

42. M.-H. S. Sinding, S. Gopalakrishnan, J. Ramos-Madrigal, M. de Manuel, V. V. Pitulko, L. Kuderna, T. R. Feuerborn, L. A. F. Frantz, F. G. Vieira, J. Niemann, J. A. Samaniego Castruita, C. Carøe, E. U. Andersen-Ranberg, P. D. Jordan, E. Y. Pavlova, P. A. Nikolskiy, A. K. Kasparov, V. V. Ivanova, E. Willerslev, P. Skoglund, M. Fredholm, S. E. Wennerberg, M. P. Heide-Jørgensen, R. Dietz, C. Sonne, M. Meldgaard, L. Dalén, G. Larson, B. Petersen, T. Sicheritz-Pontén, L. Bachmann, Ø. Wiig, T. Marques-Bonet, A. J. Hansen, M. T. P. Gilbert, Arctic-adapted dogs emerged at the Pleistocene–Holocene transition. *Science*. **368**, 1495–1499 (2020).

43. A. Bergström, Origins and Genetic Legacy of Prehistoric Dogs. *Science*.

44. M. Ní Leathlobhair, A. R. Perri, E. K. Irving-Pease, K. E. Witt, A. Linderholm, J. Haile, O. Lebrasseur, C. Ameen, J. Blick, A. R. Boyko, S. Brace, Y. N. Cortes, S. J. Crockford, A. Devault, E. A. Dimopoulos, M. Eldridge, J. Enk, S. Gopalakrishnan, K. Gori, V. Grimes, E. Guiry, A. J. Hansen, A. Hulme-Beaman, J. Johnson, A. Kitchen, A. K. Kasparov, Y.-M. Kwon, P. A. Nikolskiy, C. P. Lope, A. Manin, T. Martin, M. Meyer, K. N. Myers, M. Omura, J.-M. Rouillard, E. Y. Pavlova, P. Sciulli, M.-H. S. Sinding, A. Strakova, V. V. Ivanova, C. Widga, E. Willerslev, V. V. Pitulko, I. Barnes, M. T. P. Gilbert, K. M. Dobney, R. S. Malhi, E. P. Murchison, G. Larson, L. A. F. Frantz, The evolutionary history of dogs in the Americas. *Science*. **361**, 81–85 (2018).

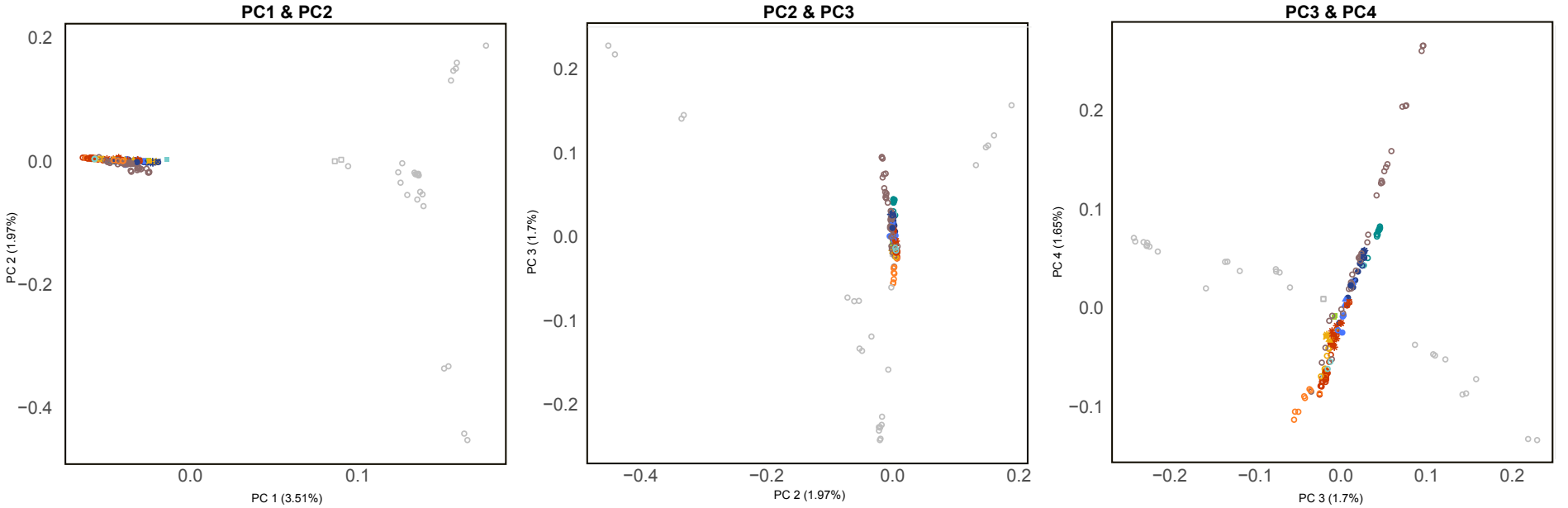
45. S. K. Brown, C. M. Darwent, B. N. Sacks, Ancient DNA evidence for genetic continuity in arctic dogs. *J. Archaeol. Sci.* **40**, 1279–1288 (2013).

46. S. K. Brown, C. M. Darwent, E. J. Wictum, B. N. Sacks, Using multiple markers to elucidate the ancient, historical and modern relationships among North American Arctic dog breeds. *Heredity* . **115**, 488–495 (2015).

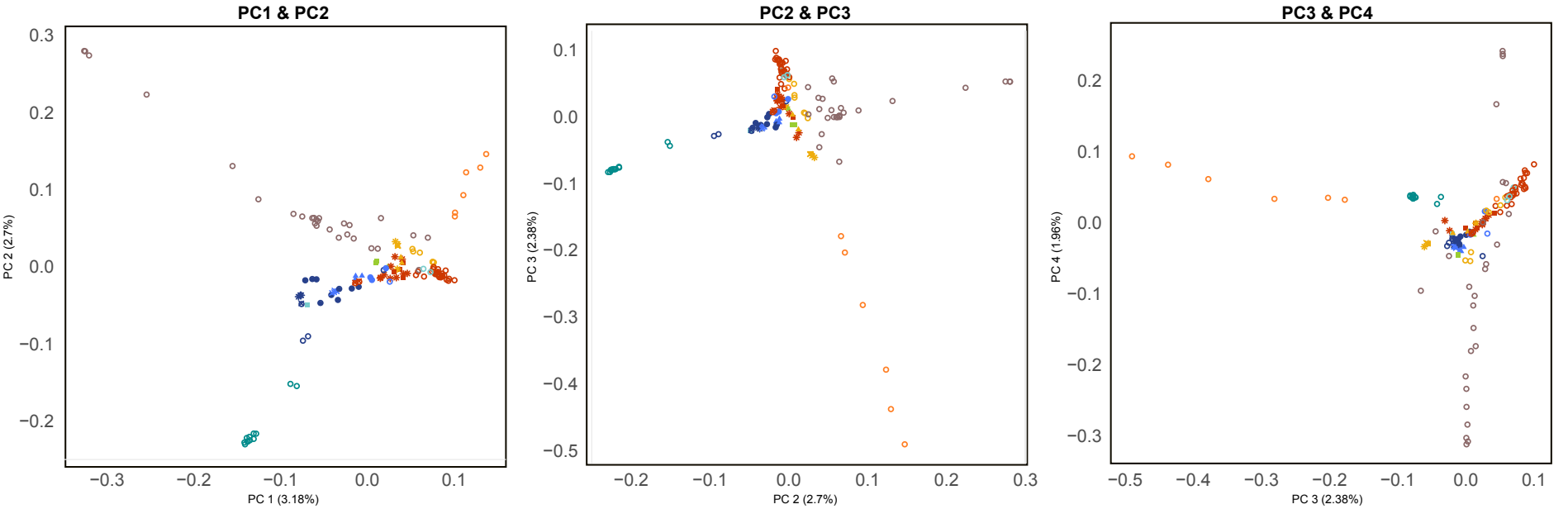
47. C. Ameen, T. R. Feuerborn, S. K. Brown, A. Linderholm, A. Hulme-Beaman, O. Lebrasseur, M.-H. S. Sinding, Z. T. Lounsberry, A. T. Lin, M. Appelt, L. Bachmann, M. Betts, K. Britton, J. Darwent, R. Dietz, M. Fredholm, S. Gopalakrishnan, O. I. Goriunova, B. Grønnow, J. Haile, J. H. Hallsson, R. Harrison, M. P. Heide-Jørgensen, R. Knecht, R. J. Losey, E. Masson-MacLean, T. H. McGovern, E. McManus-Fry, M. Meldgaard, Å. Midtdal, M. L. Moss, I. G. Nikitin, T. Nomokonova, A. H. Pálsdóttir, A. Perri, A. N. Popov, L. Rankin, J. D. Reuther, M. Sablin, A. L. Schmidt, S. Shirar, K. Smiarowski, C. Sonne, M. C. Stiner, M. Vasyukov, C. F. West, G. B. Ween, S. E. Wennerberg, Ø. Wiig, J. Woollett, L. Dalén, A. J. Hansen, M. T. P Gilbert, B. N. Sacks, L. Frantz, G. Larson, K. Dobney, C. M. Darwent, A. Evin, Specialized sledge dogs accompanied Inuit dispersal across the North American Arctic. *Proc. Biol. Sci.* **286**, 20191929 (2019).

Fig. S1: PCAs with PC1-PC4

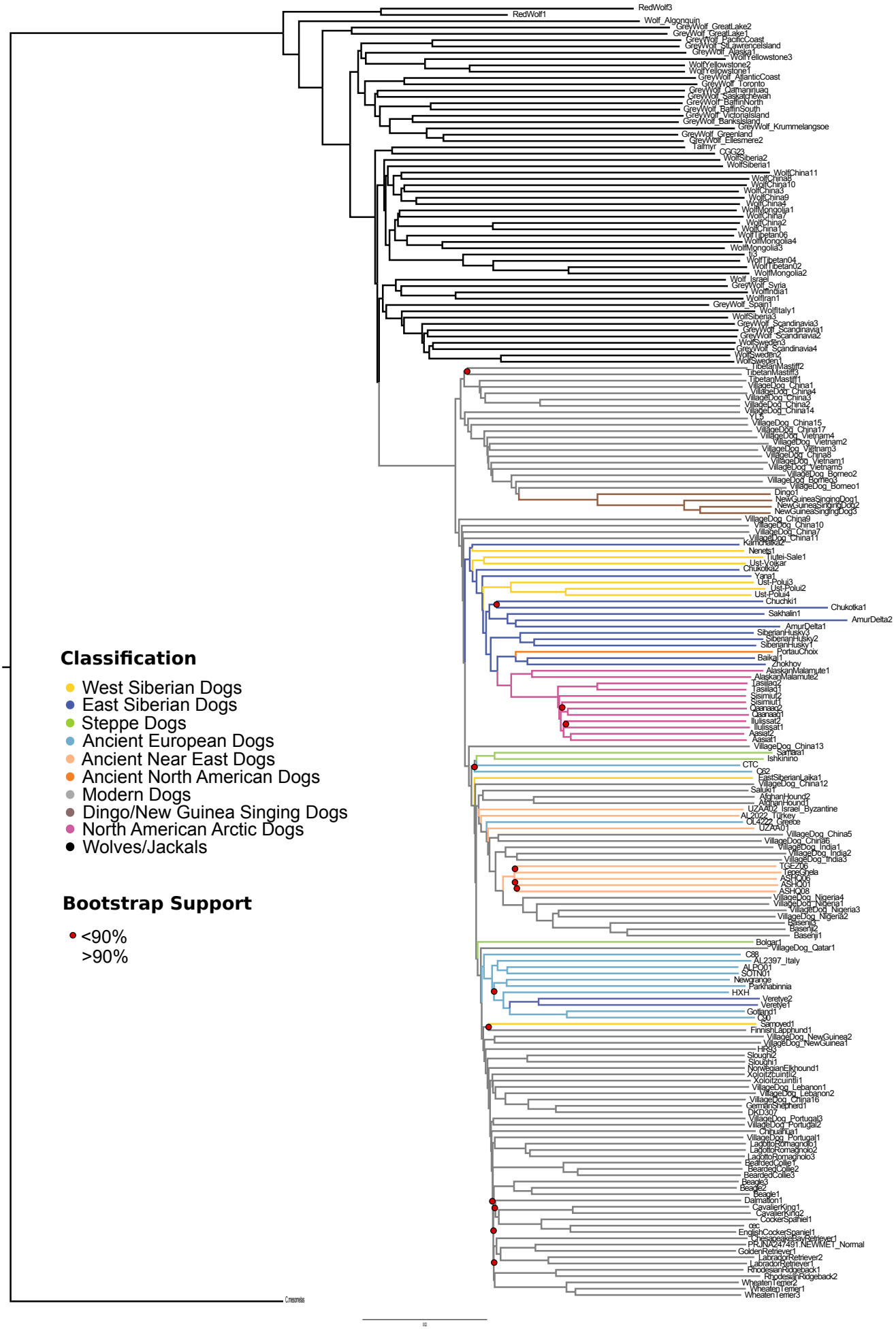
A. PCA of Dogs and Wolves



B. PCA of Dogs Only



SI Fig. 2 Neighbour Joining Treewith >0.5x



Classification

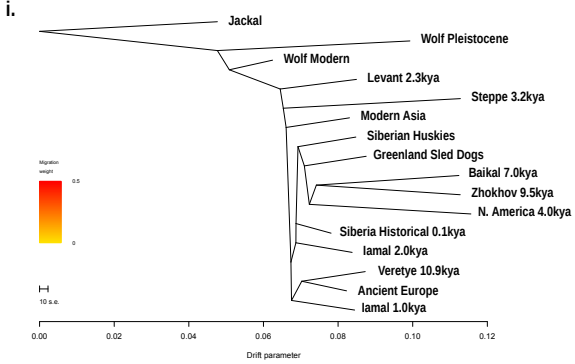
- West Siberian Dogs
- East Siberian Dogs
- Steppe Dogs
- Ancient European Dogs
- Ancient Near East Dogs
- Ancient North American Dogs
- Modern Dogs
- Dingo/New Guinea Singing Dogs
- North American Arctic Dogs
- Wolves/Jackals

Bootstrap Support

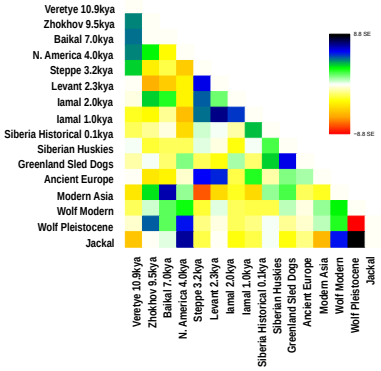
- <90%
  >90%

Fig. S3a: TreeMix Models Edges 0-4

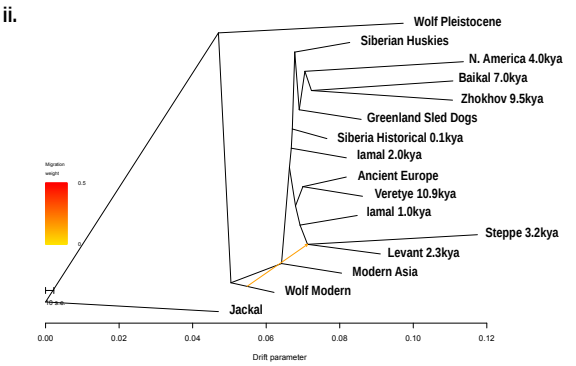
i.



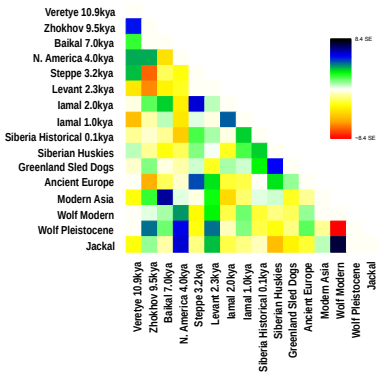
Edge 0



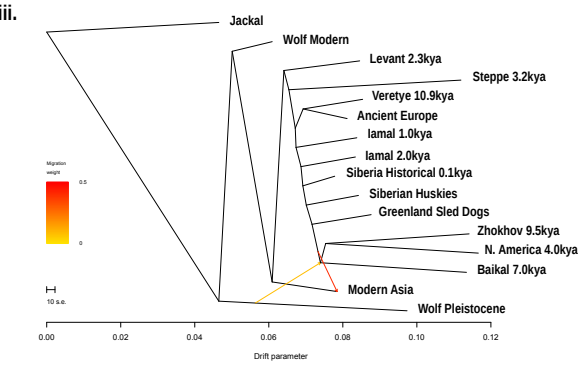
ii.



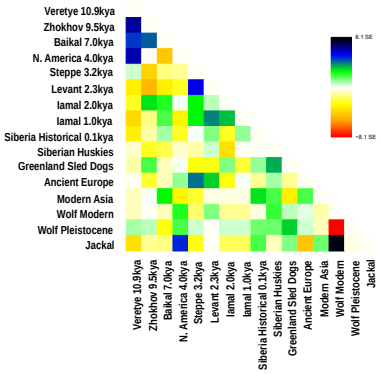
Edge 1



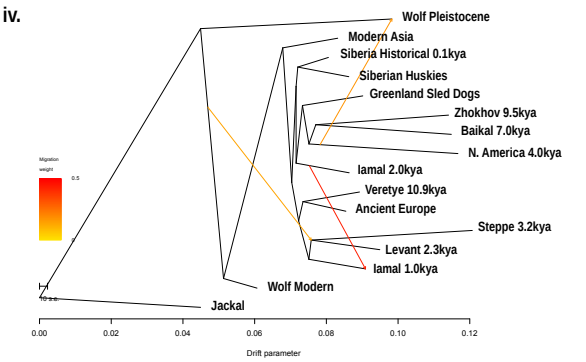
iii.



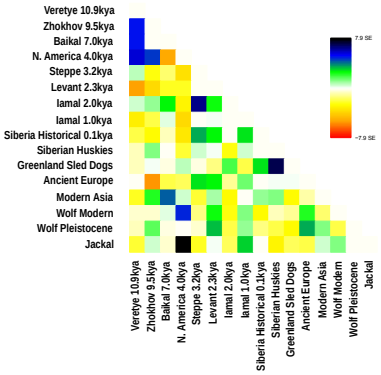
Edge 2



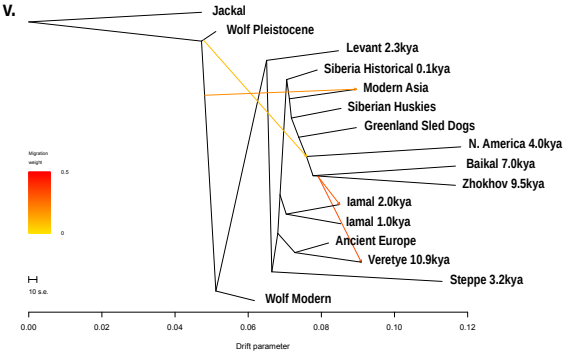
iv.



Edge 3



v.



Edge 4

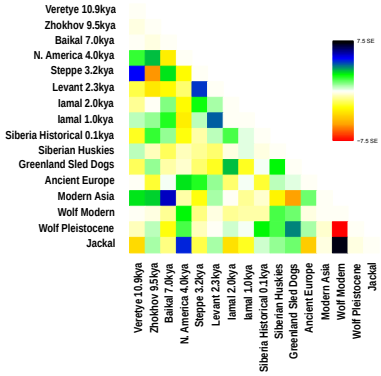
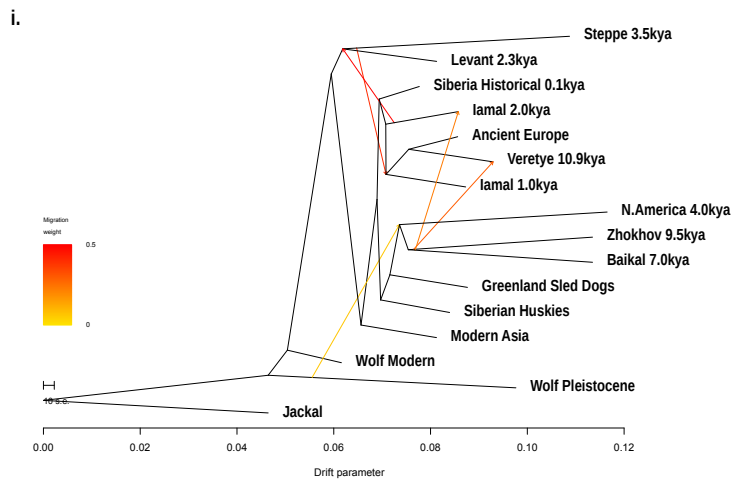
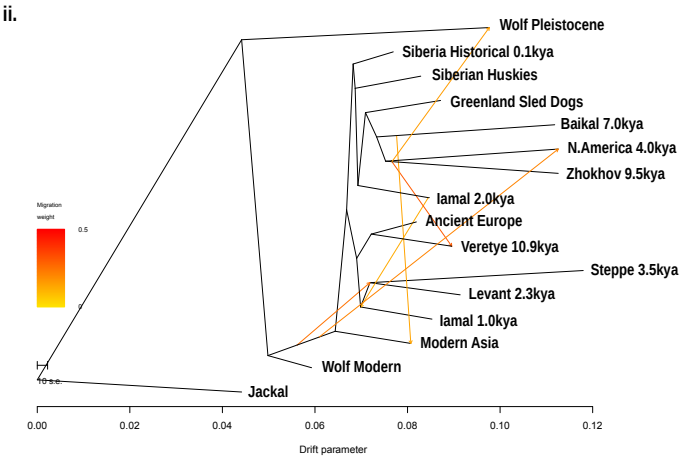
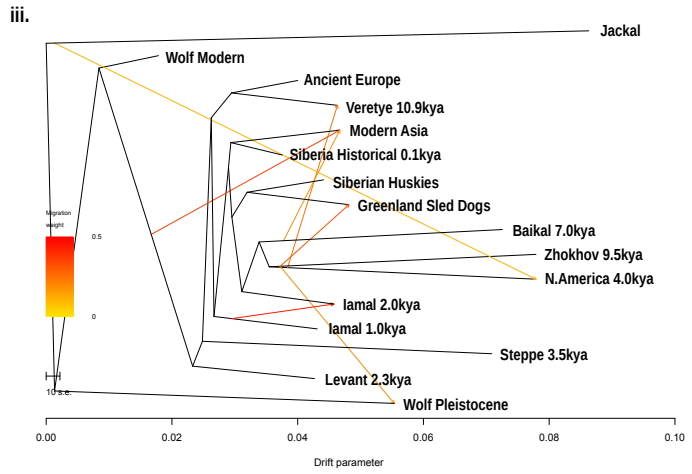
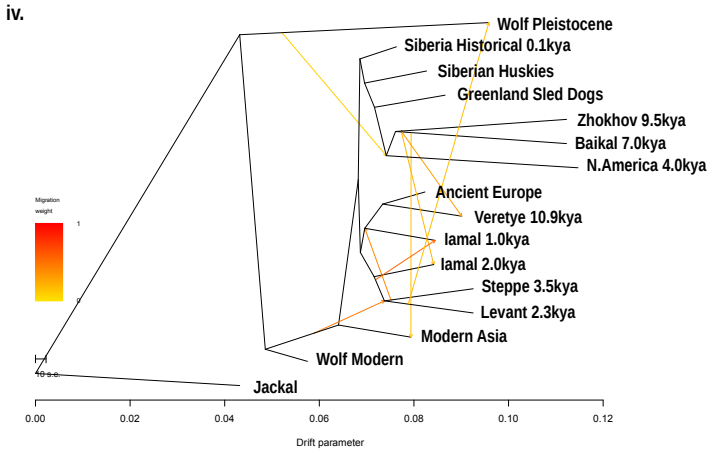


Fig. S3b: TreeMix Models Edges 5-8

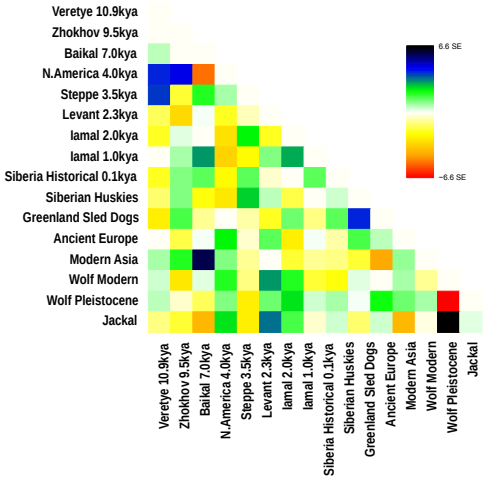




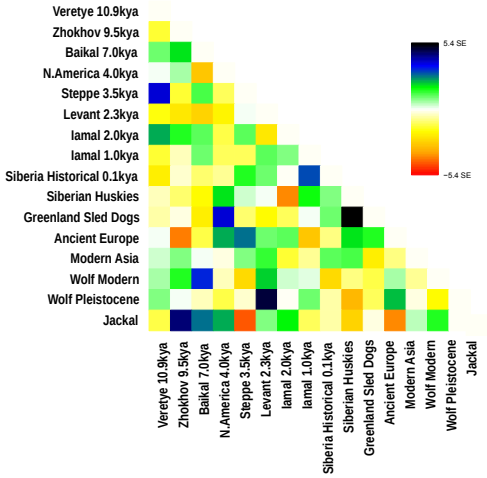




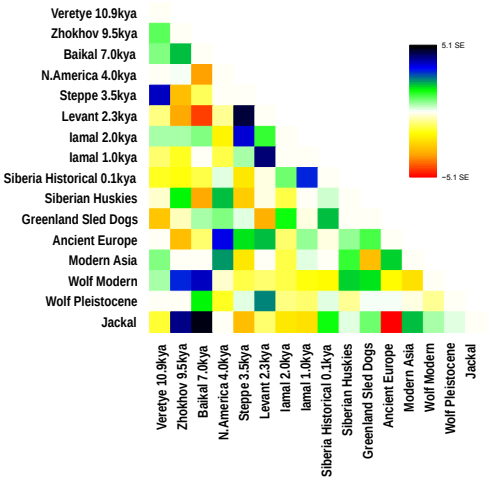
Edge 5



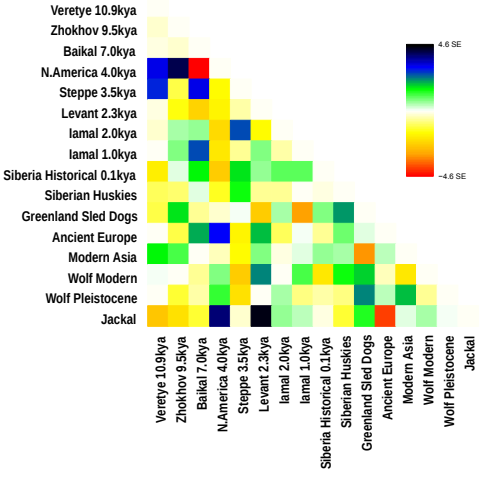
Edge 6



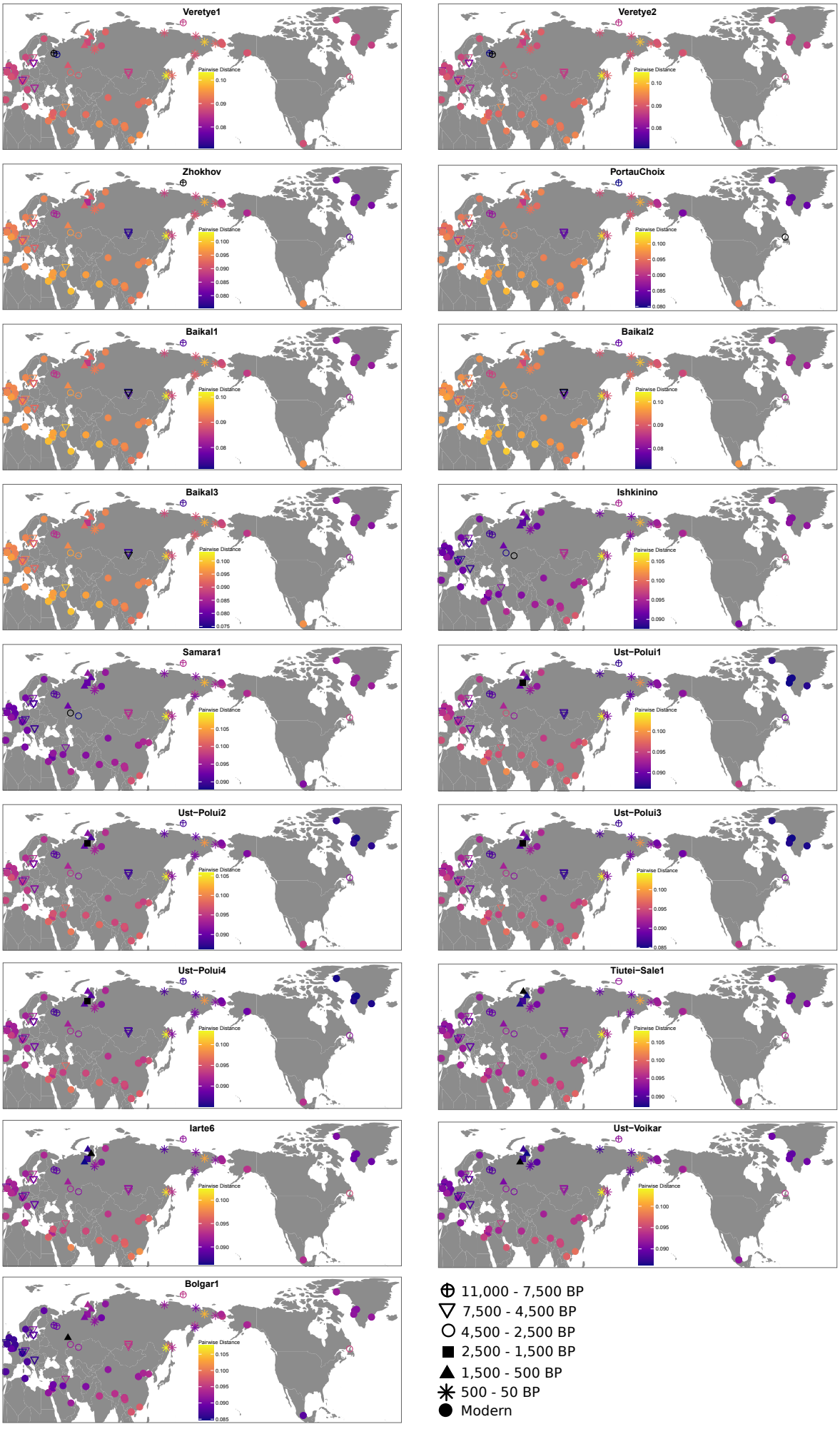
Edge 7



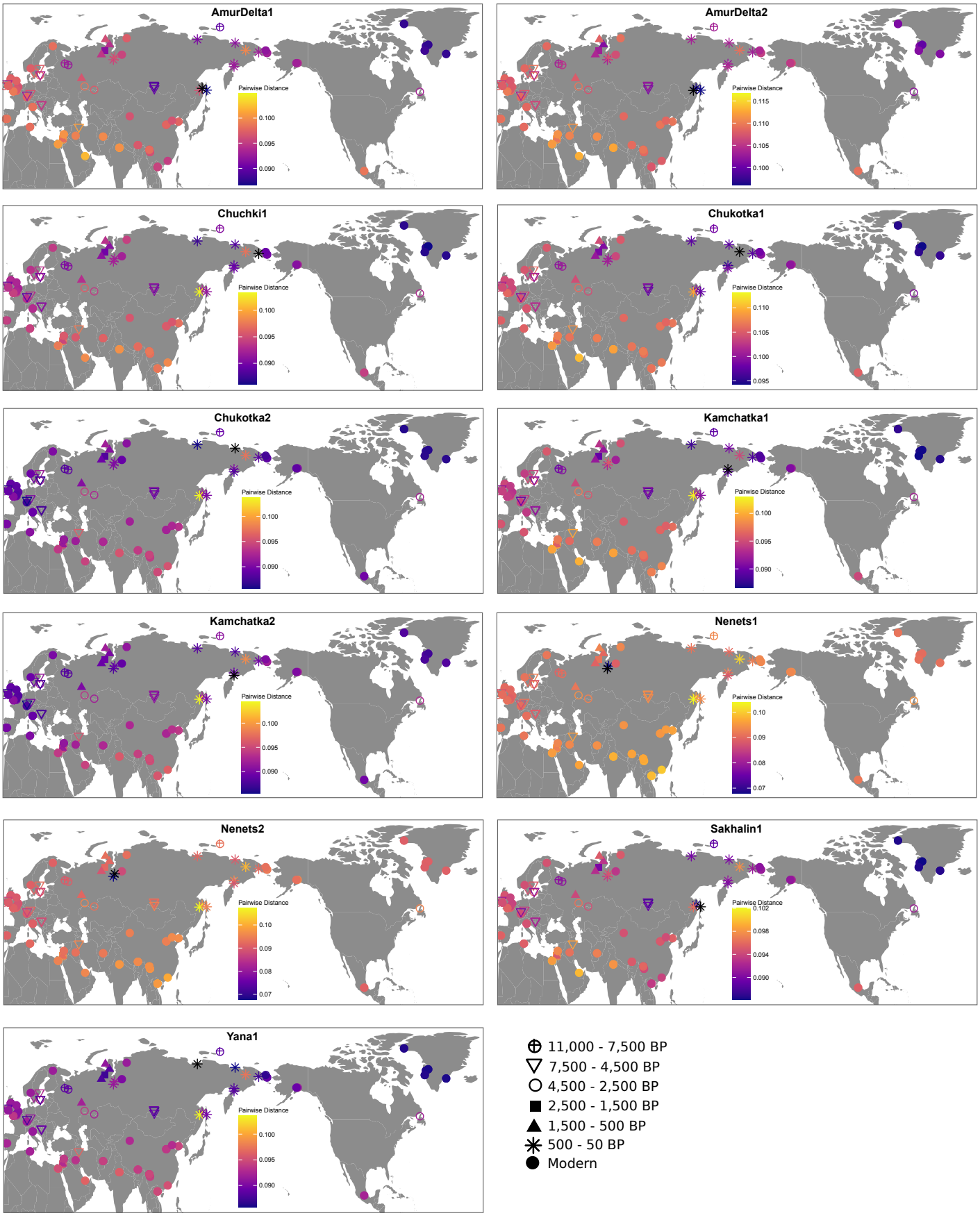
Edge 8



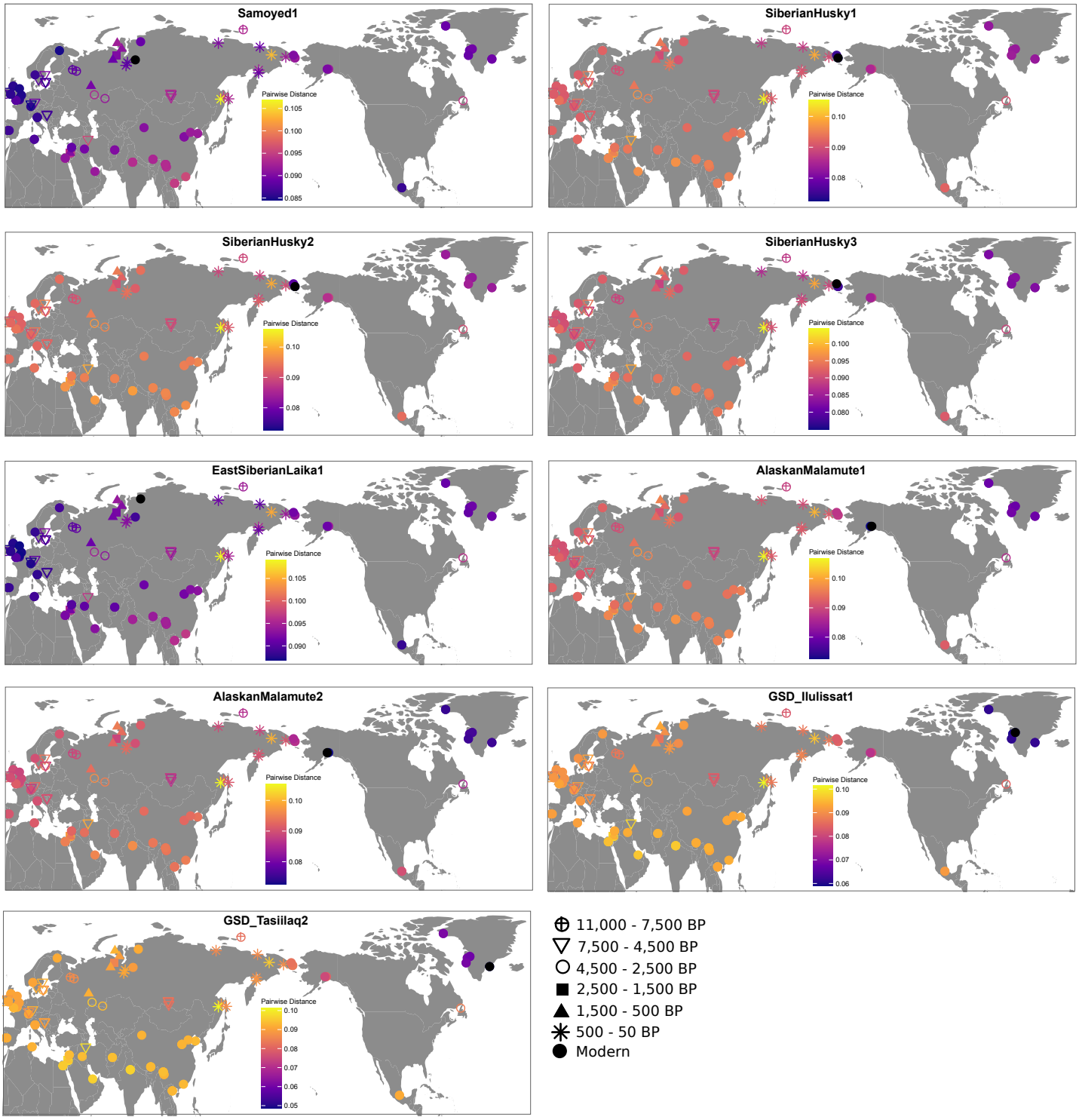
SIFig.4a: Identity−by−state Pairwise Distances (11,000 − 150 BP)



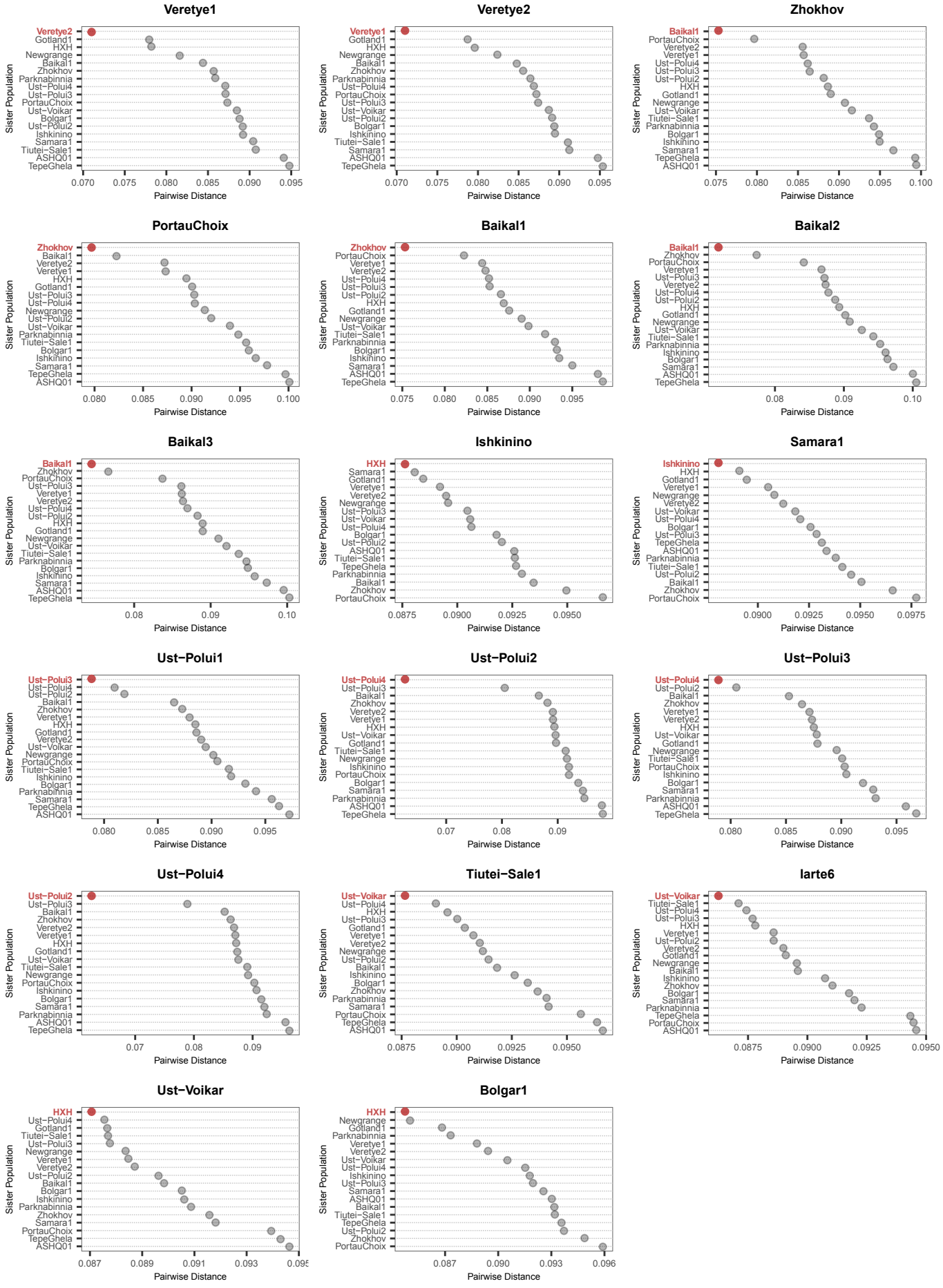
SIFig 4b: Identity−by−state Pairwise Distances (150 − 50 BP)



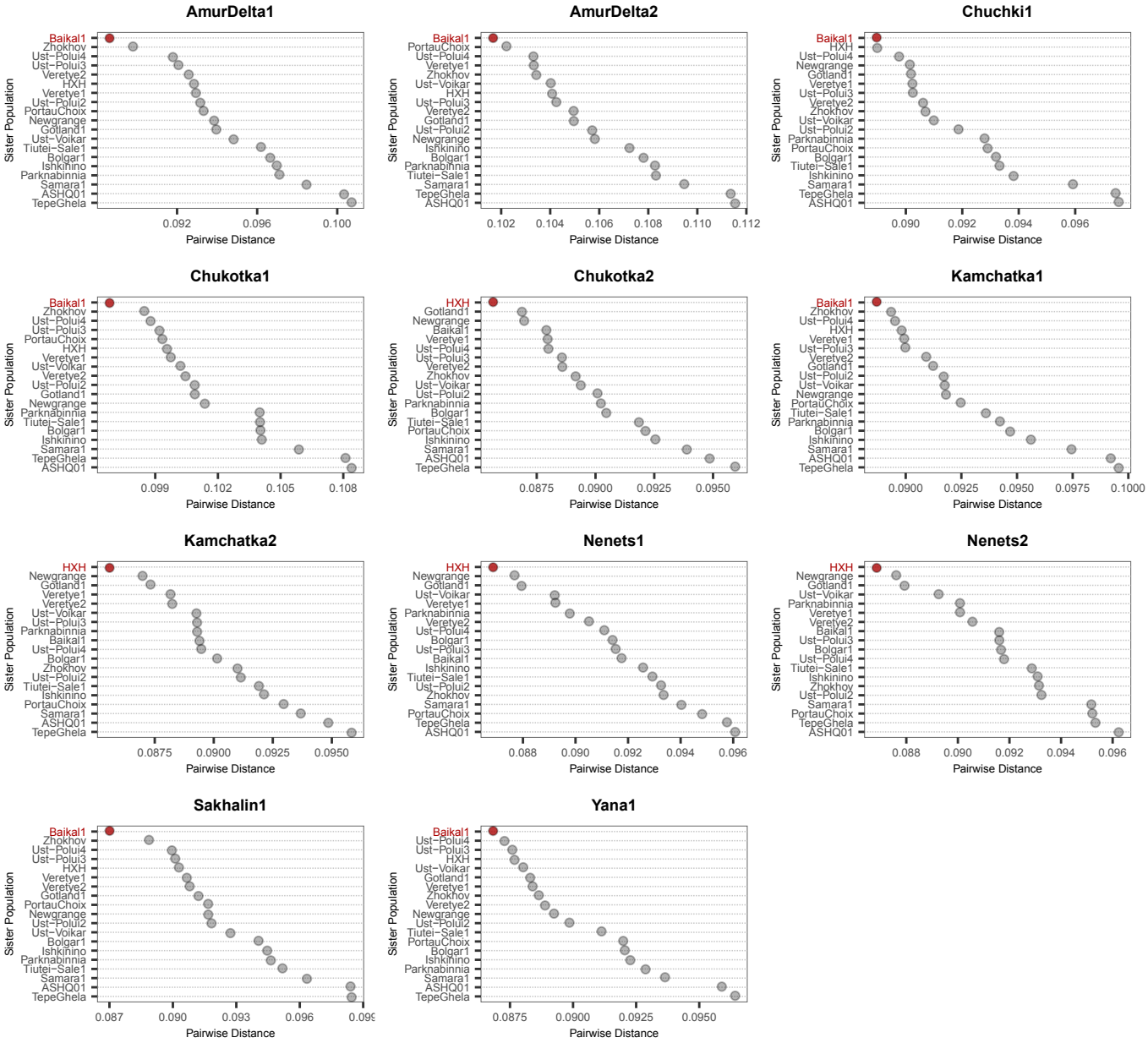
SIFig. 4c: Identity−by−state Pairwise Distances (50 − 0 BP)



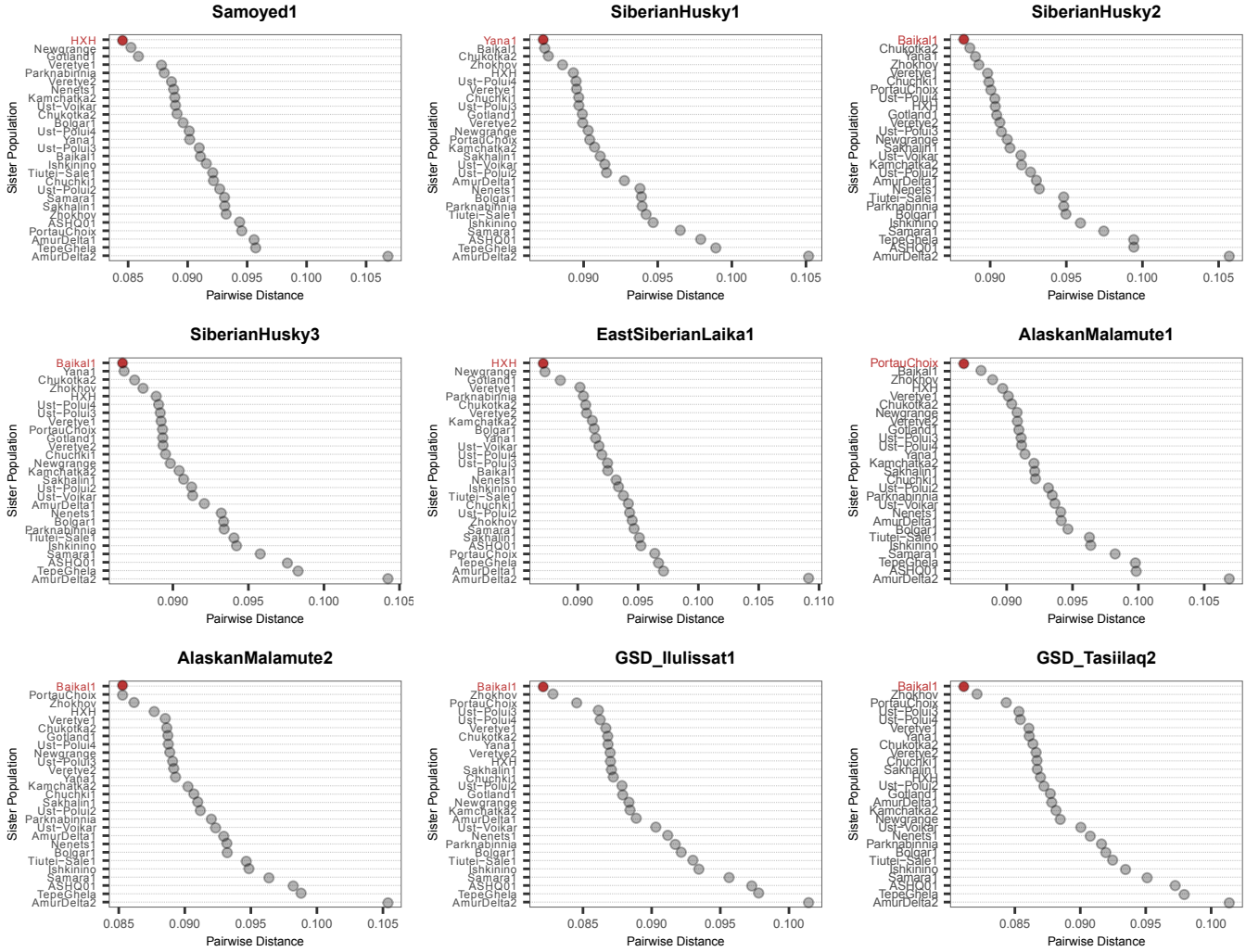
SIFig. 4d: Identity−by−state Pairwise Distances (11,000 − 150 BP)



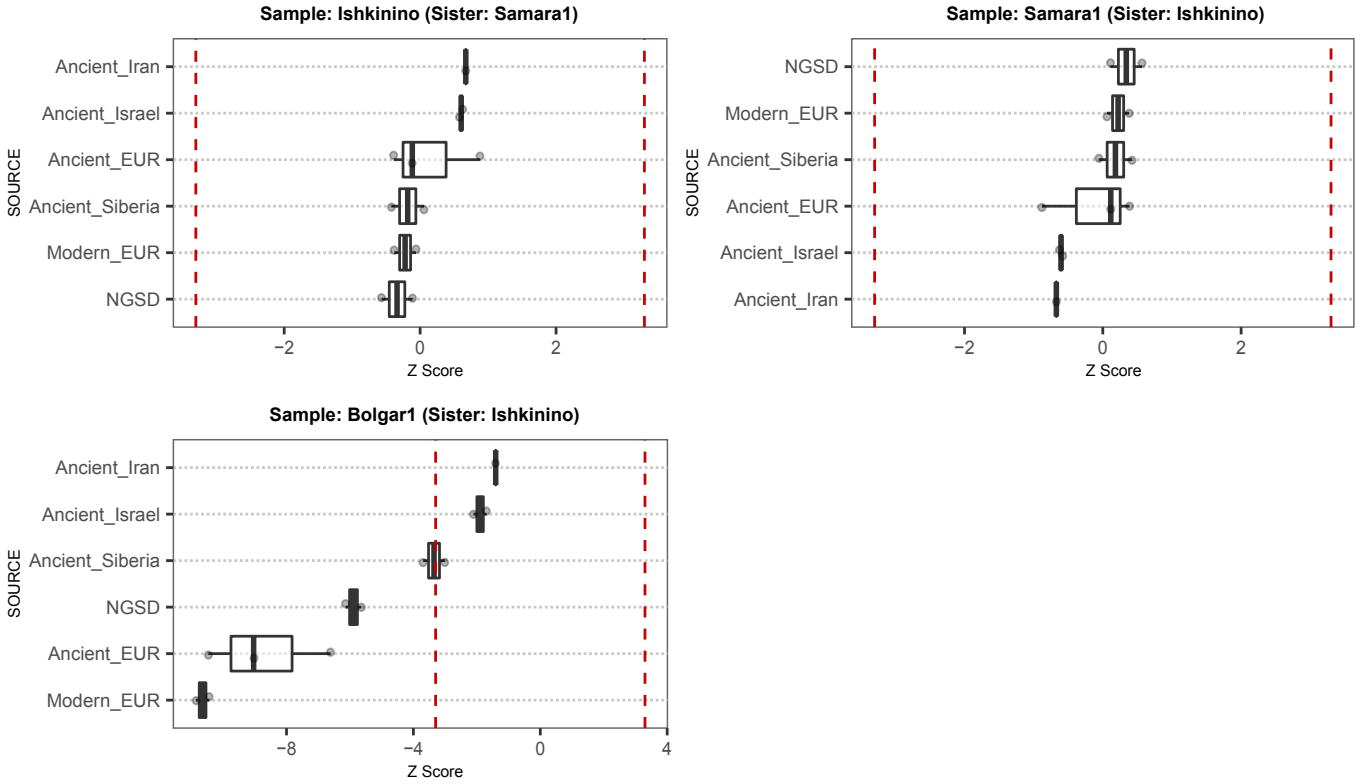
SIFig. 4e: Identity−by−state Pairwise Distances (150 − 50 BP)



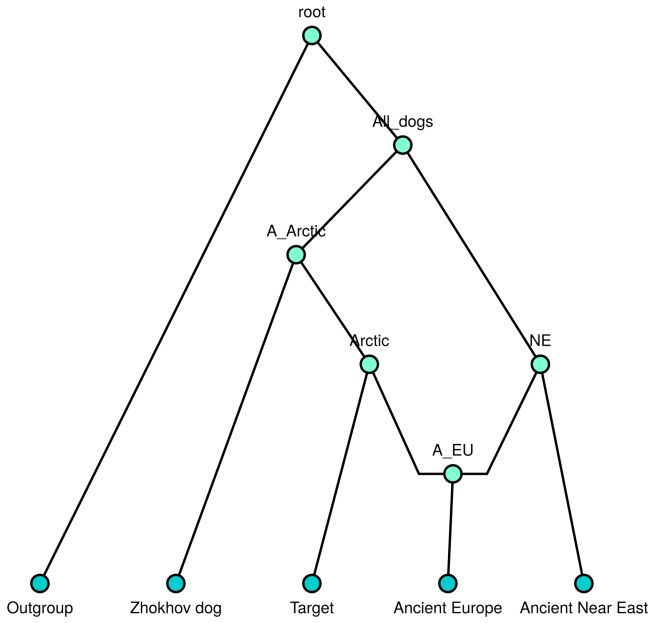
SIFig. 4f: Identity−by−state Pairwise Distances (50 − 0 BP)



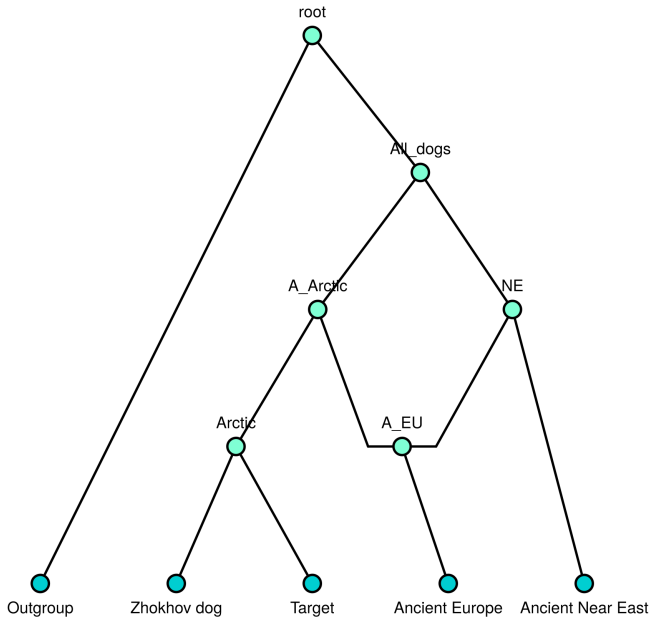
SIFig. 5: qpDstat Gene Flow into Steppe Dog



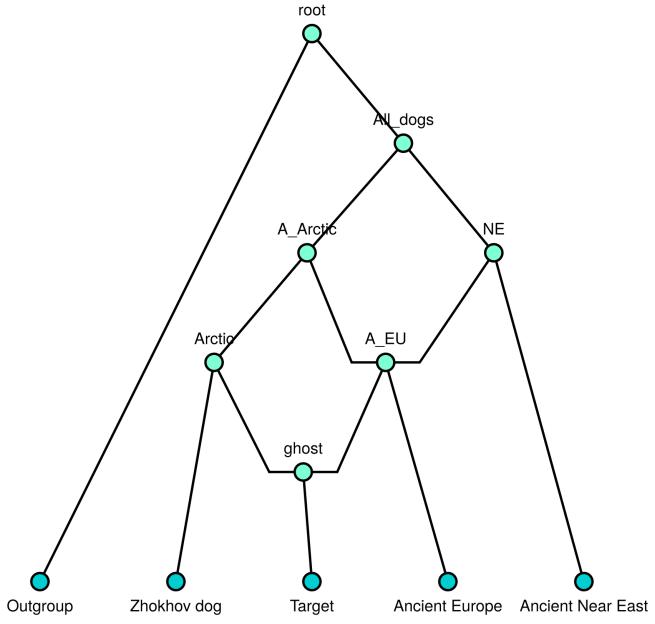
Divergence



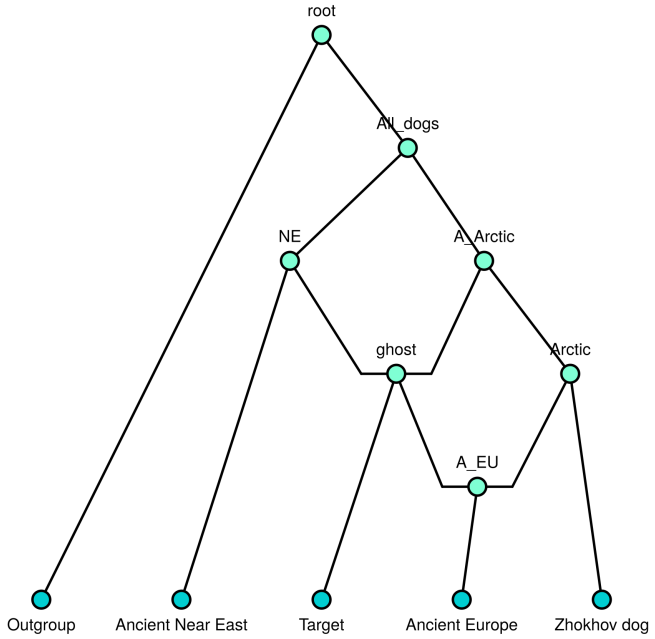
Div alternative



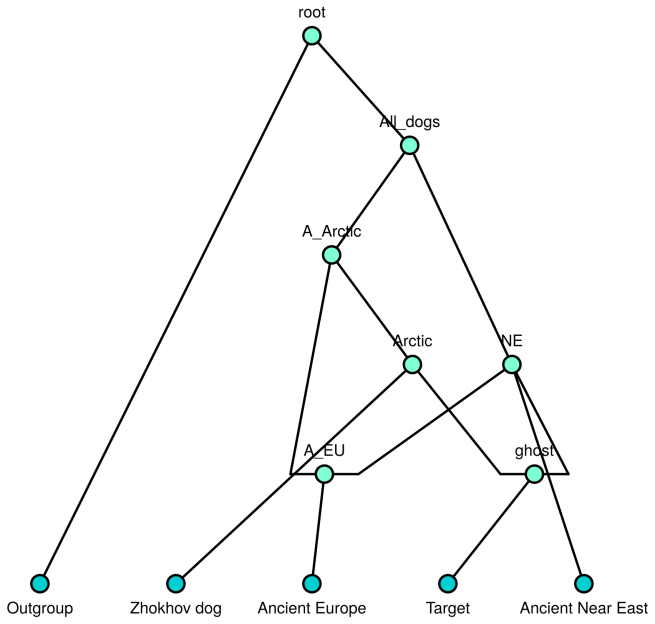
Admixture EU/AR



Old admixture AR/NE



Admixture AR/NE



Admixture EU/NE

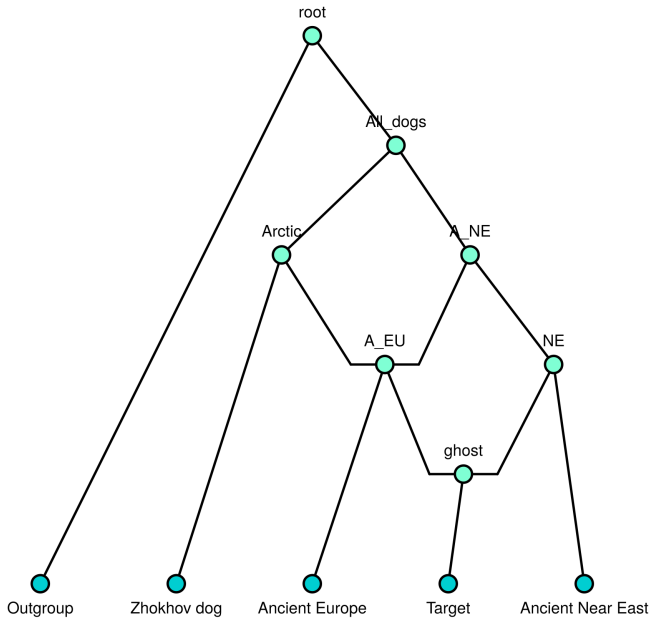


Fig. S6b Outliers Heat Map (samples aligned to the dog reference genome, Canfam3.1)

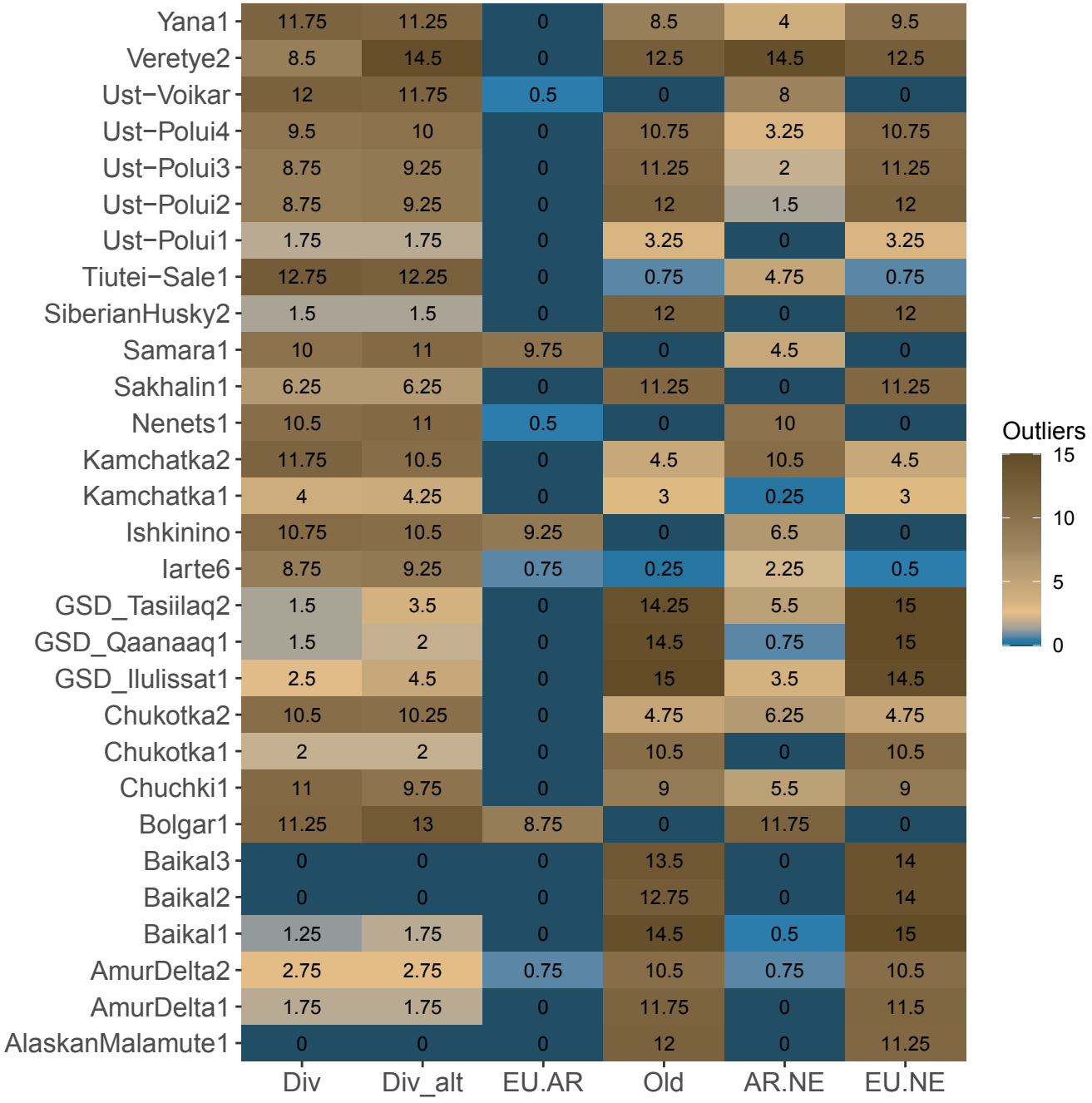
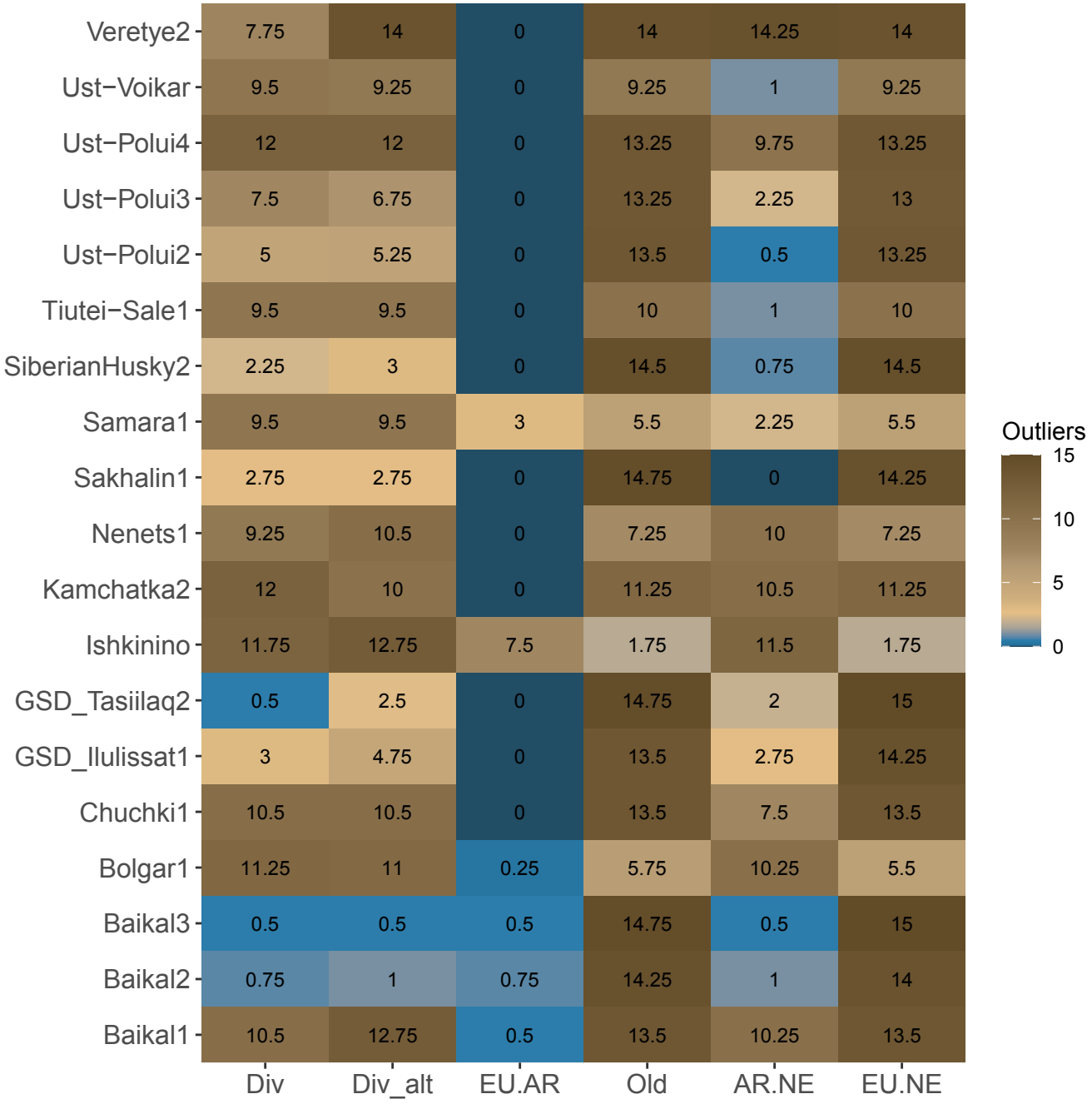
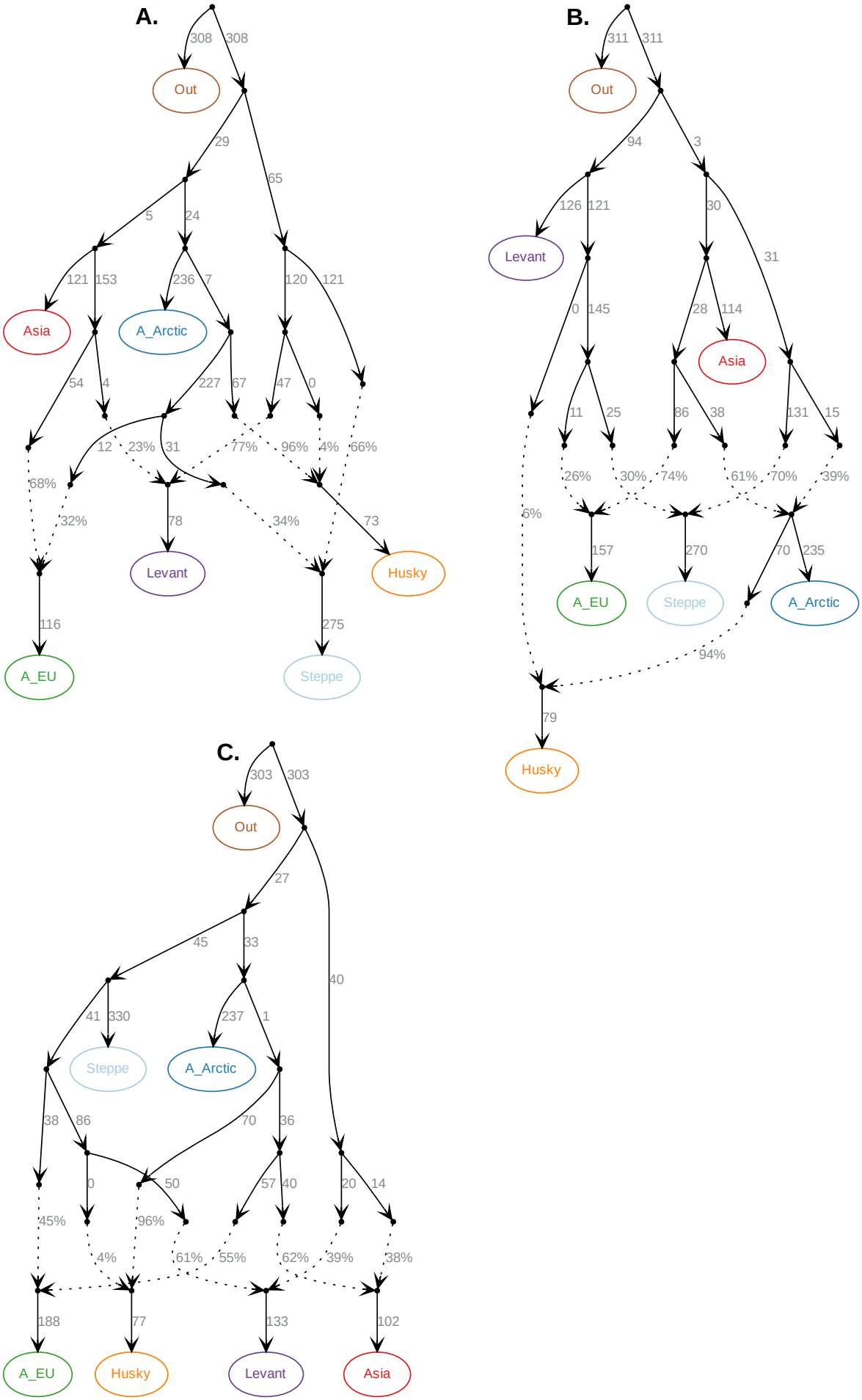


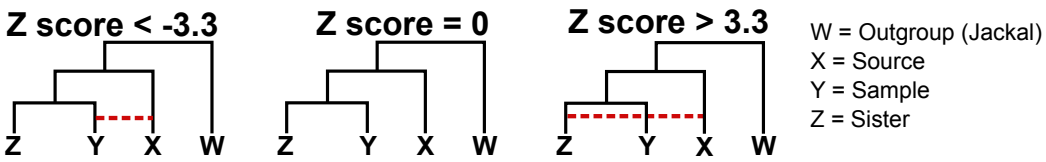
Fig. S6c Outliers Heat Map (samples aligned to the fox reference genome, VulVul2.2)



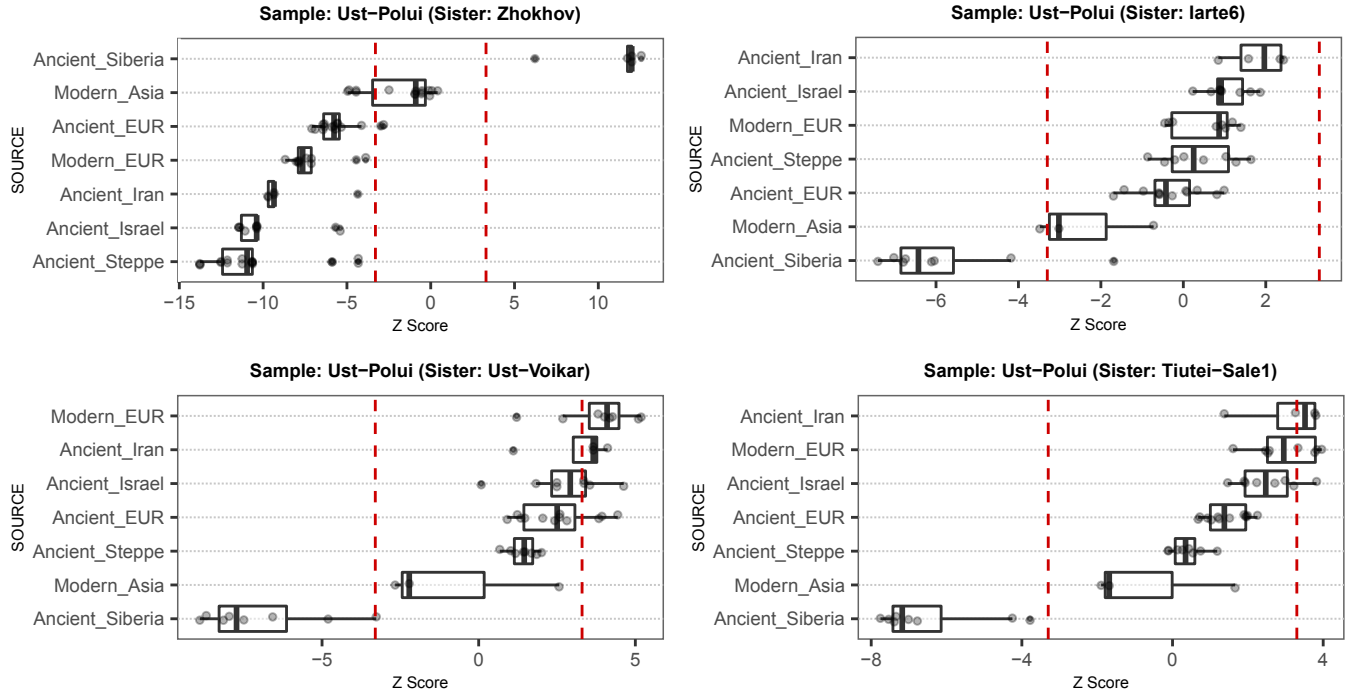
SIFig. 7: qpBrute Graphs for East Asian Dogs



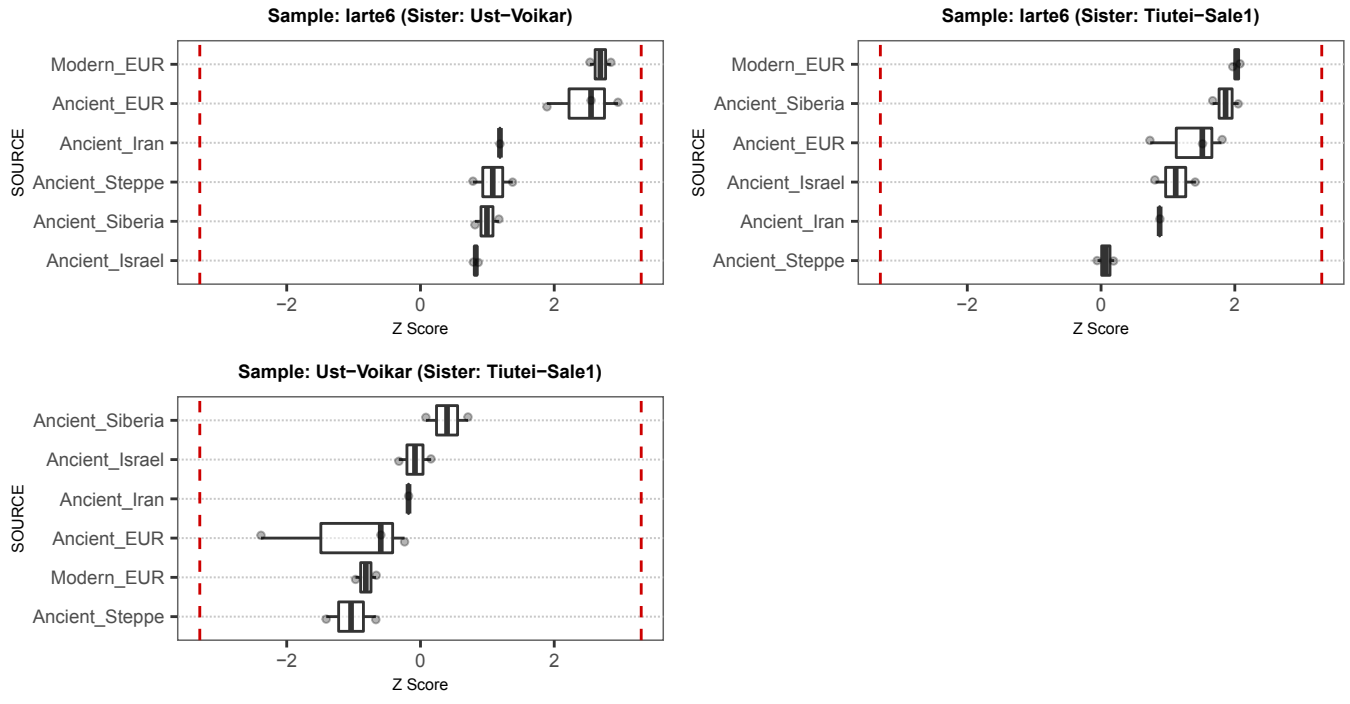
SIFig. 8: D-statistics (qpDstat) Gene Flow Into Iamal Dogs



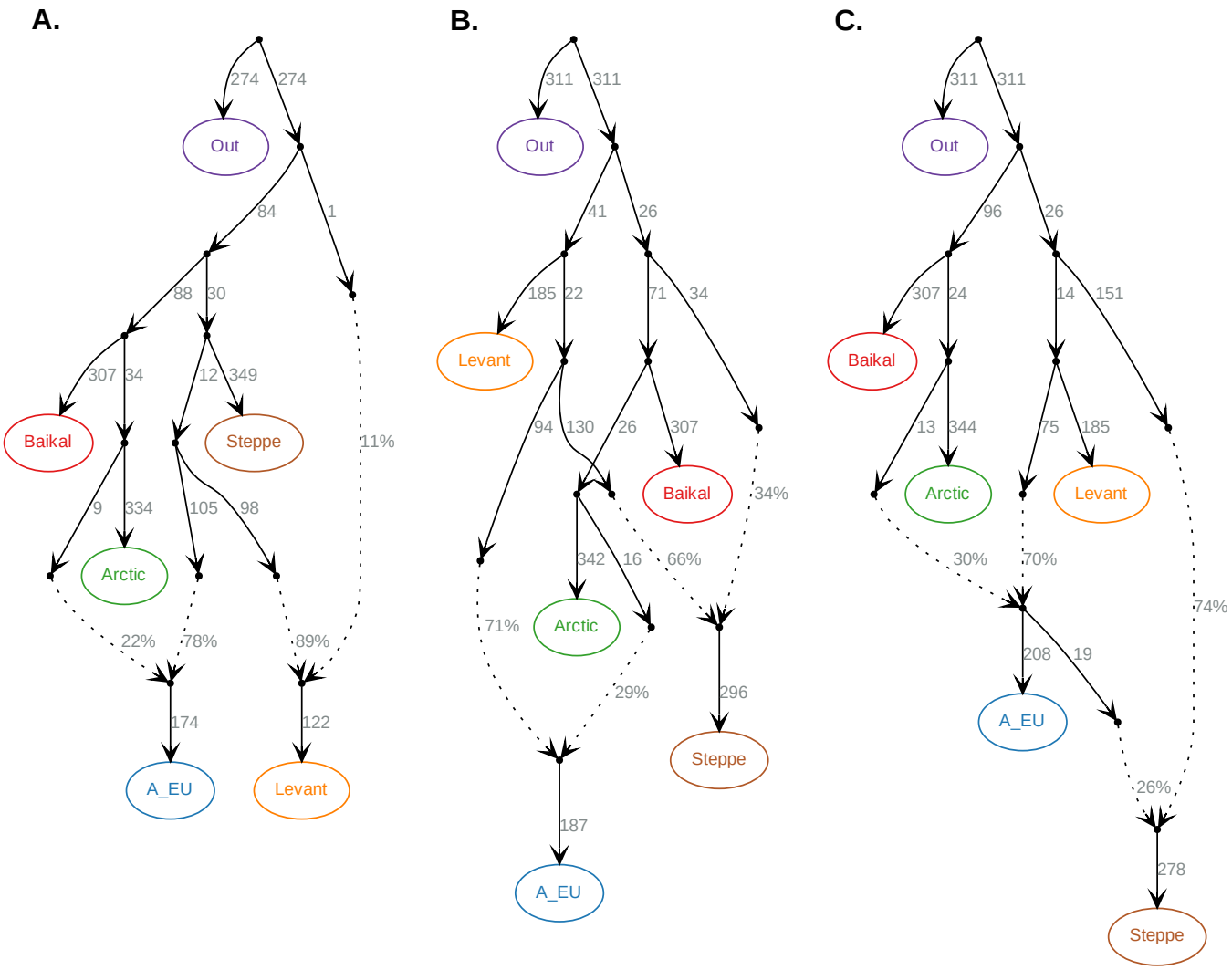
A.



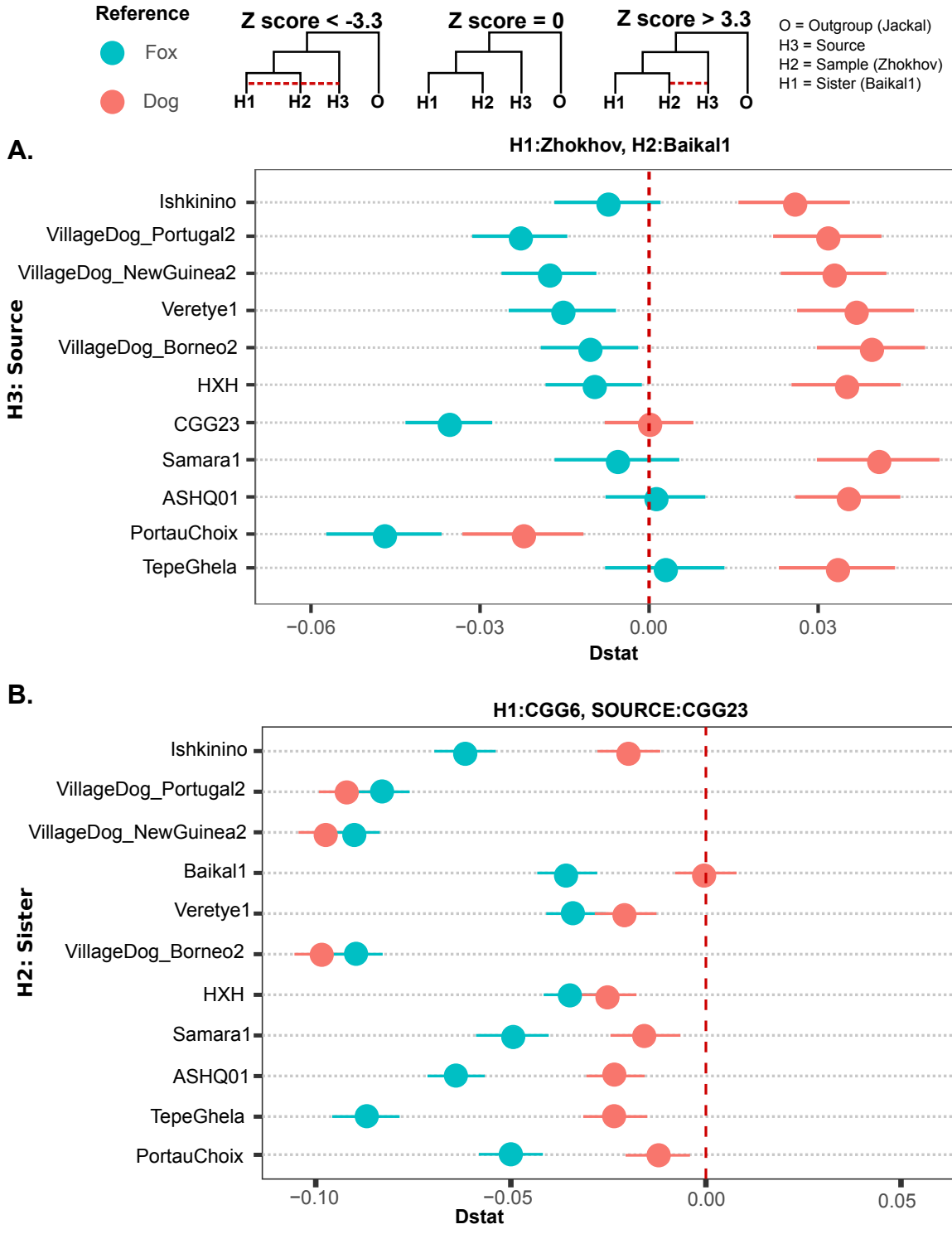
B.



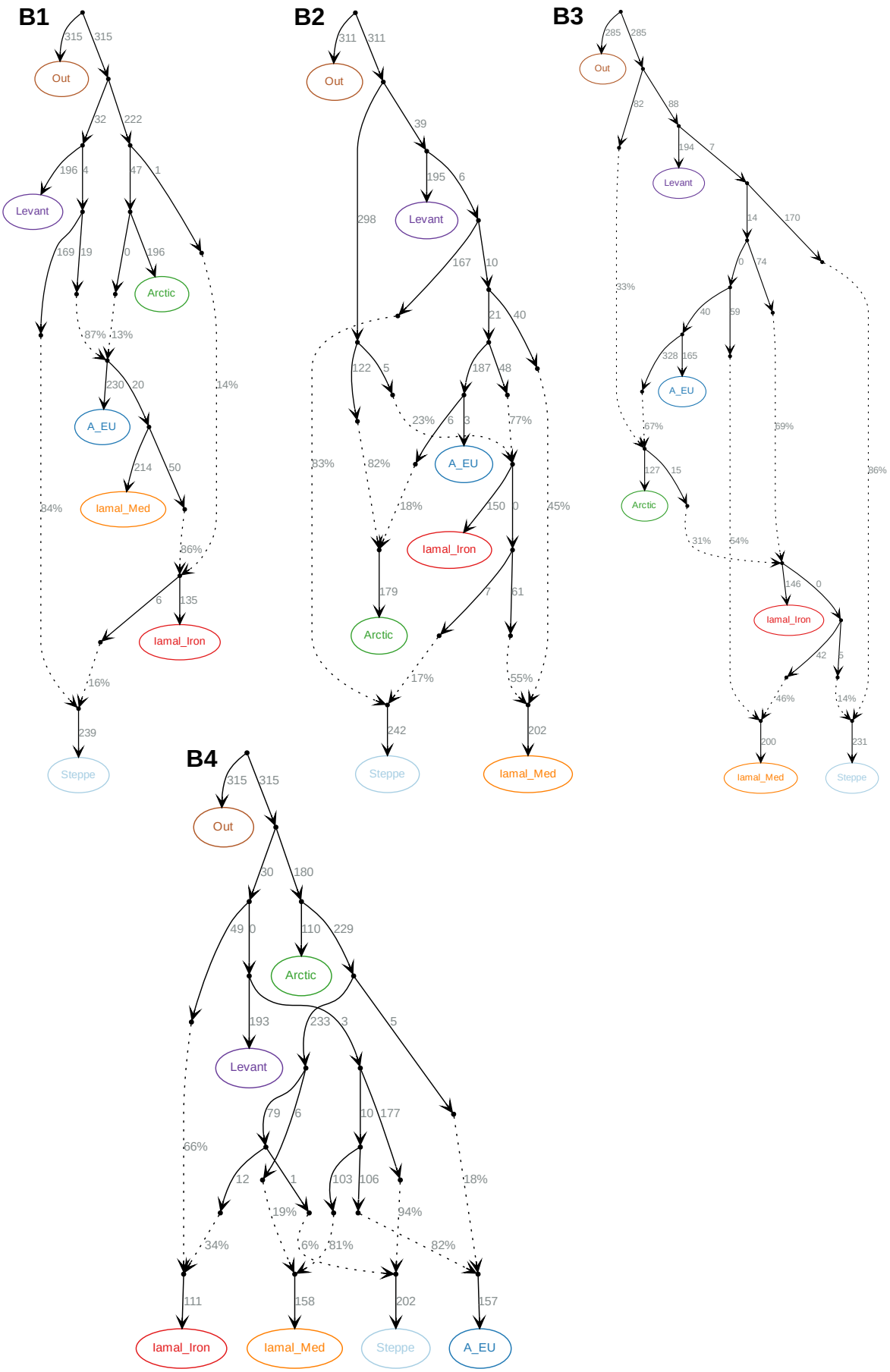
SIFig. 9: qpBrute Graphs for Baikal Dogs



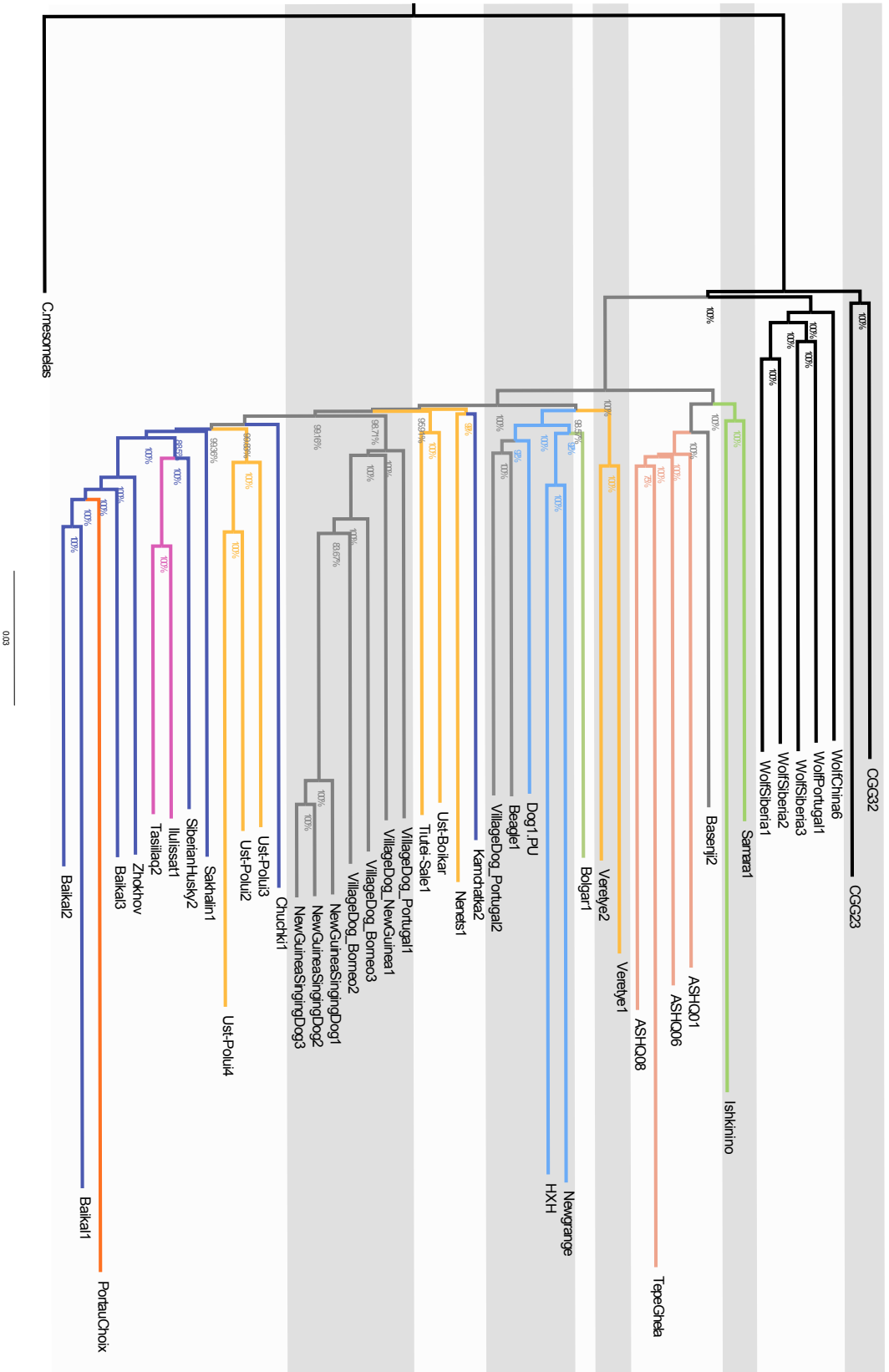
SIFig. 10 D-statistics (ANGSD) Gene Flow Reference Bias



SIFig. 11: qpBrute Graphs for Iamal Dogs



SIFig. 12 Vulpes Reference Neighbour Joining Tree



SIFig. 13 D-statistics (qpDstat) Gene Flow from Wolves

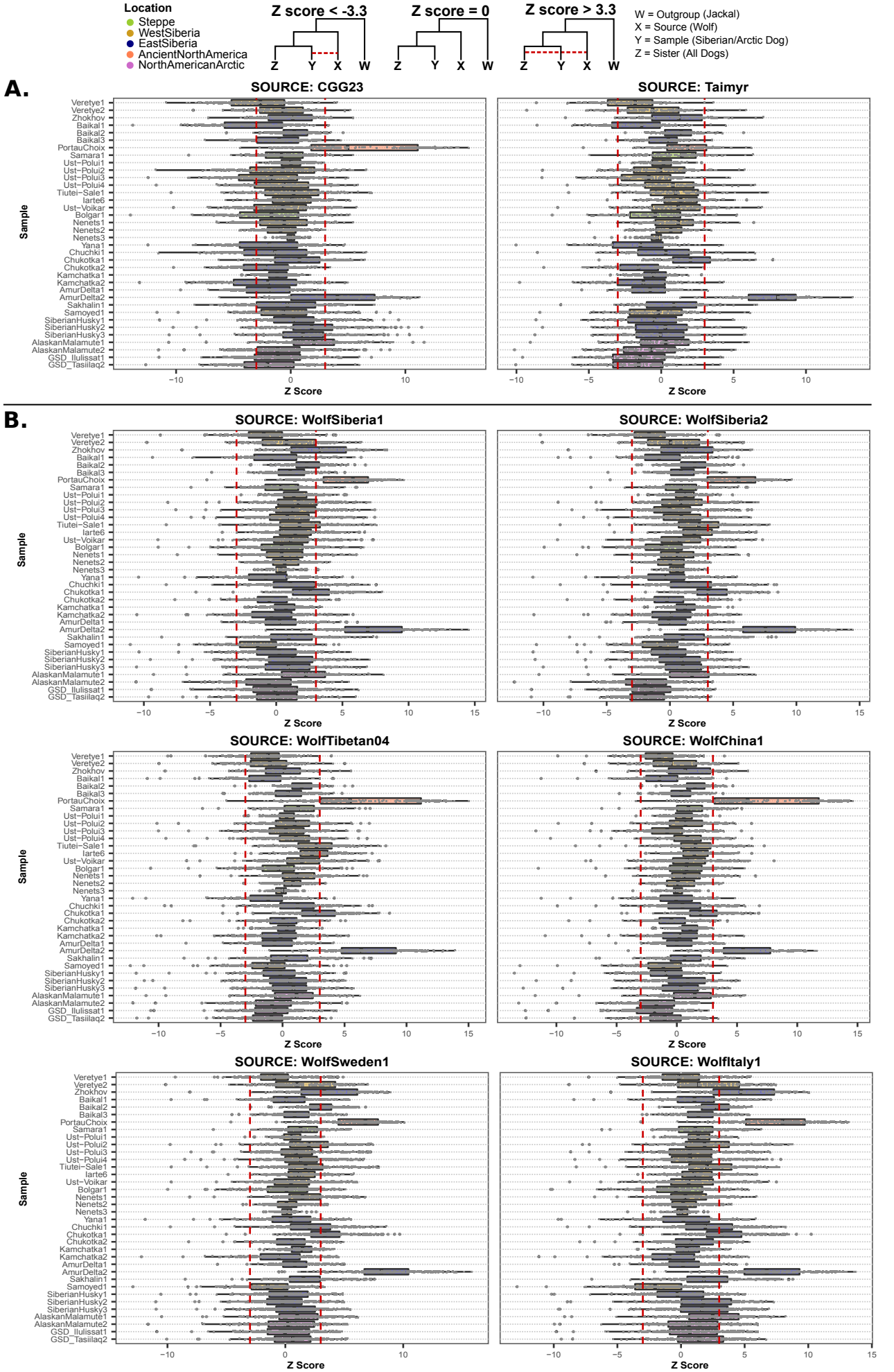


Fig. S14a Maximum Likelihood Phylogeny of Mitochondrial Genomes

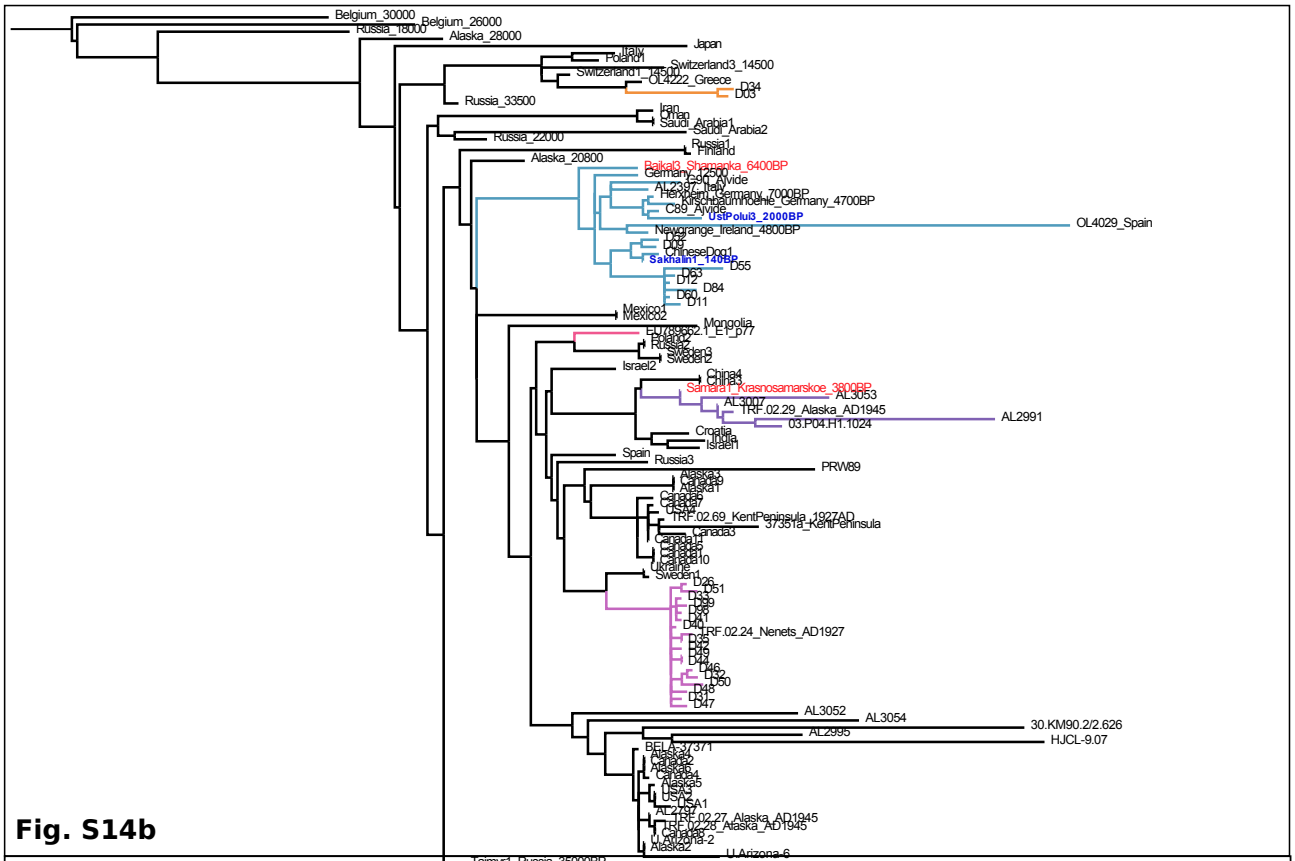


Fig. S14b

Fig. S14c

Dog Clade A
Dog Clade B
Dog Clade C
Dog Clade D
Dog Clade E
Dog Clade X

Nuclear Genomes from Siberia/Steppe

Genomes Generated in Study

Fig. S14b Maximum Likelihood Phylogeny of Mitochondrial Genomes

Dog Clade A
Dog Clade B
Dog Clade C
Dog Clade D
Dog Clade E
Dog Clade X

Nuclear Genomes from Siberia/Steppe

Genomes Generated in Study

Fig. S14c Maximum Likelihood Phylogeny of Mitochondrial Genomes