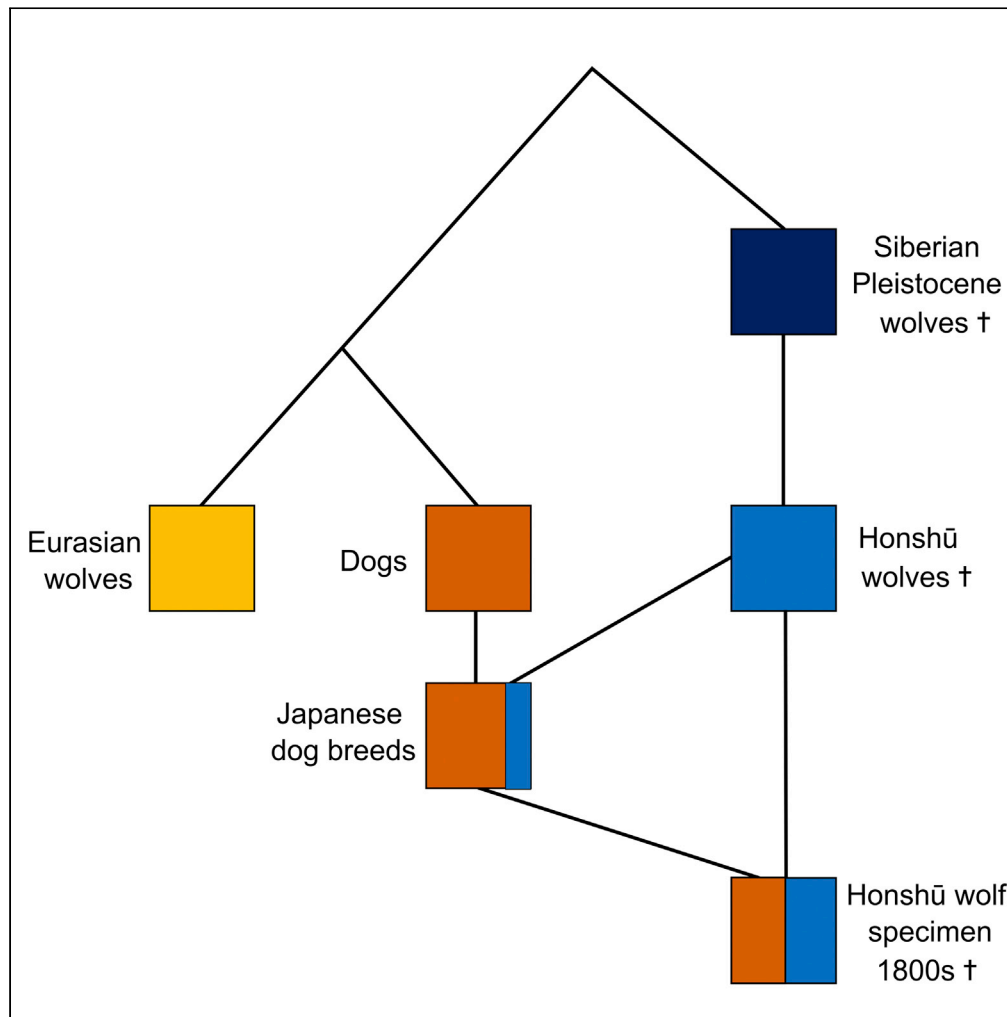**Article**

Extended survival of Pleistocene Siberian wolves into the early 20th century on the island of Honshū

Jonas Niemann, Shyam Gopalakrishnan, Nobuyuki Yamaguchi, Jazmín Ramos-Madrigal, Nathan Wales, M. Thomas P. Gilbert, Mikkel-Holger S. Sinding

jonas@palaeome.org (J.N.)
mhssinding@gmail.com (M.-H.S.S.)

HIGHLIGHTS

Generated 3.7✕ nuclear genome of the extinct Honshu wolf

The Honshū wolf belonged to the lineage of Siberian Pleistocene wolves

There was gene flow between Honshū wolves and Japanese dogs

Article

Extended survival of Pleistocene Siberian wolves into the early 20th century on the island of Honshū

Jonas Niemann,[1,2,8,*] Shyam Gopalakrishnan,[1,3] Nobuyuki Yamaguchi,[4] Jazmín Ramos-Madrigal,[1,3] Nathan Wales,[2] M. Thomas P. Gilbert,[1,3,5] and Mikkel-Holger S. Sinding[1,6,7,*]

SUMMARY

The Japanese or Honshū wolf was one the most distinct gray wolf subspecies due to its small stature and endemicity to the islands of Honshū, Shikoku, and Kyūshū. Long revered as a guardian of farmers and travellers, it was persecuted from the 17th century following a rabies epidemic, which led to its extinction in the early 20th century. To better understand its evolutionary history, we sequenced the nuclear genome of a 19th century Honshū wolf specimen to an average depth of coverage of 3.7×. We find Honshū wolves were closely related to a lineage of Siberian wolves that were previously believed to have gone extinct in the Late Pleistocene, thereby extending the survival of this ancient lineage until the early 20th century. We also detected significant gene flow between Japanese dogs and the Honshū wolf, corroborating previous reports on Honshū wolf dog interbreeding.

INTRODUCTION

Researchers have long debated the evolutionary origin of present-day dogs and Eurasian wolves, as well as their relationships to each other. While it is now clear from analyses based on genome-scale data sets that they are reciprocally monophyletic sister clades and thus dogs are not derived from any as yet identified extant Eurasian wolf lineage (Freedman et al., 2014), from what wolf population dogs were domesticated, and what the ancestral homeland was of their common ancestors, remains debated (Larson and Bradley, 2014). A further open question is the relationship of both lineages to other wolf groups. One key example is the enigmatic Pleistocene Beringian wolves that may have specialized on hunting megafauna (Leonard et al., 2007), several of whose genomes have recently been sequenced from subfossil materials, and found to represent lineages that fall outside of modern dogs and wolves (Ramos-Madrigal et al., 2020; Skoglund et al., 2015). Indeed, studies of both mitochondrial (Koblmüller et al., 2016; Loog et al., 2020) and nuclear genomes (Loog et al., 2020; Skoglund et al., 2015) recovered from Pleistocene and modern Eurasian wolves, suggest the Pleistocene lineages were lost and replaced by the modern lineages around the Pleistocene-Holocene transition in Siberia (Loog et al., 2020) - possibly associated with the extinction of much of the megafauna upon which the Pleistocene lineages may have depended (Leonard et al., 2007). One curious observation from such studies is the lack of evidence for admixture between these Pleistocene wolves and the ancestors of modern Eurasian wolves and most dog breeds (with the exception of some Asian dogs and Arctic sled dogs (Sinding et al., 2020; Skoglund et al., 2015). This is particularly striking given reports of widespread admixture between other coexisting species within the crown Canids (Gopalakrishnan et al., 2018), and suggests that the Pleistocene ancestors of present-day Eurasian wolves and dogs must have been physically isolated from the Siberian Pleistocene wolf lineages during the Last Glacial Maximum (LGM), prior to replacing them following their extinction after 14,000 years ago (Loog et al., 2020), likely at the transition at the Holocene (Leonard et al., 2007). This in turn raises the obvious, and currently unanswered, question, as to where this isolation could have occured?

The Japanese archipelago is one potential candidate for the LGM refugium of the ancestors of modern wolves and dogs, as land bridges between the Korean peninsula and Japan's largest island, Honshū, formed during the Pleistocene and the beginning of the Holocene (Ohshima, 1990). Hokkaidō, the second largest and northernmost island of Japan, was also connected to the Eurasian continent during periods of low sea level, which occurred for instance in the Late Pleistocene (Ohshima, 1990). Until their extinction at

[1]Section for Evolutionary Genomics, the GLOBE Institute, University of Copenhagen, Copenhagen, Denmark

[2]BioArch, Department of Archaeology, University of York, York, UK

[3]Center for Evolutionary Hologenomics, the GLOBE Institute, University of Copenhagen, Copenhagen, Denmark

[4]Institute of Tropical Biodiversity and Sustainable Development, University Malaysia Terengganu, Kuala Nerus, Terengganu 21030, Malaysia

[5]University Museum, NTNU, Trondheim, Norway

[6]Greenland Institute of Natural Resources, Nuuk, Greenland

[7]Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland

[8]Lead contact

*Correspondence: jonas@palaeome.org (J.N.), mhssinding@gmail.com (M.-H.S.S.)
https://doi.org/10.1016/j.isci.2020.101904

Figure 1. Geographical distribution of the extinct Honshū and Ezo wolf
The Tsugaru strait separates the former habitat of the Ezo wolf—Sakhalin and Hokkaidō—and the former habitat of the Honshū wolf—the islands of Honshū, Shikoku, and Kyūshū. Map created in ArcGIS, wolf shapes are in relative size (the wither height of the Honshu and Ezo wolf is 56-58cm and 70-80cm, respectively; Ishiguro, 2012)

the beginning of the 20th century, Japan was inhabited by two highly phenotypically distinct endemic wolf subspecies: the Japanese or Honshū wolf (*Canis lupus hodophilax*), and the Ezo wolf (*Canis lupus hattai*). While the Honshū wolf could be found on Honshū, Kyūshū, and Shikoku, the habitat of the Ezo wolf was restricted to Hokkaidō and Sakhalin (Figure 1) (Ishiguro et al., 2009). The Honshū wolf was among the smallest gray wolf subspecies in the world and appreciated in medieval Japan for killing crop-destroying wildlife (Fritts et al., 2003). A rabies epidemic in the 17th century caused an increase in wolf attacks, setting the human persecution of the Honshū wolf in motion, which culminated in their extinction by 1905 (Walker, 2009).

The deep Tsugaru strait between Honshū and Hokkaidō is a major zoogeographical barrier between the two islands, also known as Blakiston's Line (Dobson, 1994). As a result, the fauna on Honshū, with its snow macaques (*Macaca fuscata*) and Asian black bears (*Ursus thibetanus*), has similarities to Southeastern Asia, while the fauna on Hokkaidō, which includes the Ussuri brown bear (*Ursus arctos lasiotus*), resembles the biological diversity in Northeastern Asia. As a consequence of this barrier, there is no evidence for an overlap between the habitats of the Japanese and the Ezo wolf that most likely colonized the Japanese archipelago from the Korean peninsula and Siberia, respectively.

The exact phylogenetic placement of both subspecies is speculative, as apart from osteological comparisons noting the striking morphological differences between Honshū wolves, Ezo wolves, and present-day wolves (Imaizumi, 1970; Ishiguro et al., 2010), only the mitochondrial genomes have been sequenced in previous studies (Matsumura et al., 2014; Koblmüller et al., 2016). These suggest a basal phylogenetic

placement of the Honshū wolf to all modern wolves, and a placement of the Ezo wolf in the North American wolf clade. The mitochondrial genome is however only one marker, and it does not allow the quantification of admixture, which is especially of interest given that both subspecies are potential candidate populations that link Pleistocene wolves and present-day Eurasian wolves.

We sequenced the nuclear genome of one of the two subspecies, the Honshū wolf (*Canis lupus hodophilax*), to reassess the relationship between Honshū wolves and other wolves and test the hypothesis that Japan was the LGM refugium for the ancestors of present-day wolves.

RESULTS

Gene flow between Honshū wolf, Pleistocene wolves, and present-day wolves and dogs

We resequenced the nuclear genome of a Honshū wolf sample provided by the Natural History Museum, London, to 3.7✕ coverage. The specimen had been shot in the wild in the 1800s in Chichibu District, Kotsuki, Northwest of Tokyo, Japan. We confirmed the authenticity of the ancient DNA based on the misincorporation and read length patterns (Figure S1) and determined the chromosomal sex of the specimen to be male (Figure S2). First, we investigated the evolutionary relationship between the historic Honshū wolf and other wolves and dogs with a whole-genome admixture analysis using NGSadmix (Figure 2A, see also Figure S3). The Honshu wolf specimen was found to have indistinguishable admixture profiles as Pleistocene wolves in the data panel (see Table S1), across all tested numbers of assumed ancestral populations. In contrast to all other modern wolf populations, we find that Pleistocene wolves contributed substantially to the Honshū wolf genome. Irrespective of the number of ancestry clusters inferred in the NGSadmix analysis, the Honshū wolf always derives the majority of its ancestry from the same cluster as the Pleistocene wolves (Figure S4).

To further explore the admixture landscape between the Honshū wolf and ancient and present-day wolf and dog populations, we used D-statistics to formally test for gene flow between these groups. The D-statistics provide support for excess allele sharing between the Honshū wolf and Greenland dogs, Asian dogs, Pleistocene wolves, and Chinese wolves (Figure 2B). We already observed shared genetic ancestry between Pleistocene wolves and the Honshū wolf in the NGSadmix analysis, so to further investigate wolf and dog populations that might be more genetically similar to the Honshū wolf than other Pleistocene wolves, we created a scatterplot with the D-statistics test with the Honshū wolf and Pleistocene wolf in H3, the Portuguese wolf in H1, and dogs and wolves in the panel in H2 (Figure 2B). The results suggest that the Honshū wolf and Pleistocene wolves are symmetrically related to modern Eurasian and North American wolves, with the exception of some Chinese wolves that share more alleles with the Honshū wolf than they do with any Pleistocene wolf. A potential explanation for this is the substantial admixture between East Asian wolves and dogs (Fan et al., 2016).

All dog individuals included in the D-statistics analysis share significantly more alleles with the Honshū wolf than with the Siberian Pleistocene wolf, with Japanese dogs and Greenland dogs having the closest genetic affinity with the Honshū wolf. Additionally, analysis using qpWave showed that while the Siberian Pleistocene wolves can be explained as a single migration stream equally related to dogs and wolves, the Honshu wolf needs ancestry from the dog lineage. We therefore hypothesize that our Honshū wolf individual was most likely admixed with Japanese dogs, as the excess of shared alleles with the Greenland dogs can be explained by the introgression from Pleistocene wolves to Arctic dogs (Skoglund et al., 2015), and the Chinese dogs could likewise be shown to have a significant wolf contribution (Fan et al., 2016).

Haplotype-aware clustering of Honshu wolf, Pleistocene wolves, and present-day dogs and wolves

In order to more robustly identify population structure among the wolf and dog samples, we used the haplotype-aware clustering tool fineSTRUCTURE (Lawson et al., 2012; see also Tables S2 and S3), which has been previously used on low-coverage ancient individuals (Martiniano et al., 2016, 2017). In the dendrogram based on a similarity matrix, the Honshū wolf was positioned in the same clade as three other Pleistocene wolves – Tumat, Yana, and Bunge-Toll, further corroborating our earlier findings (Figure 3A, see also Figure S5). To further verify our findings of genetic affinity of the Honshū wolf to the Pleistocene Siberian wolves, we performed unsupervised dimension reduction on the haplotype data using principal component analysis (PCA). The Honshū wolf formed an incline together with all Pleistocene wolves along the first
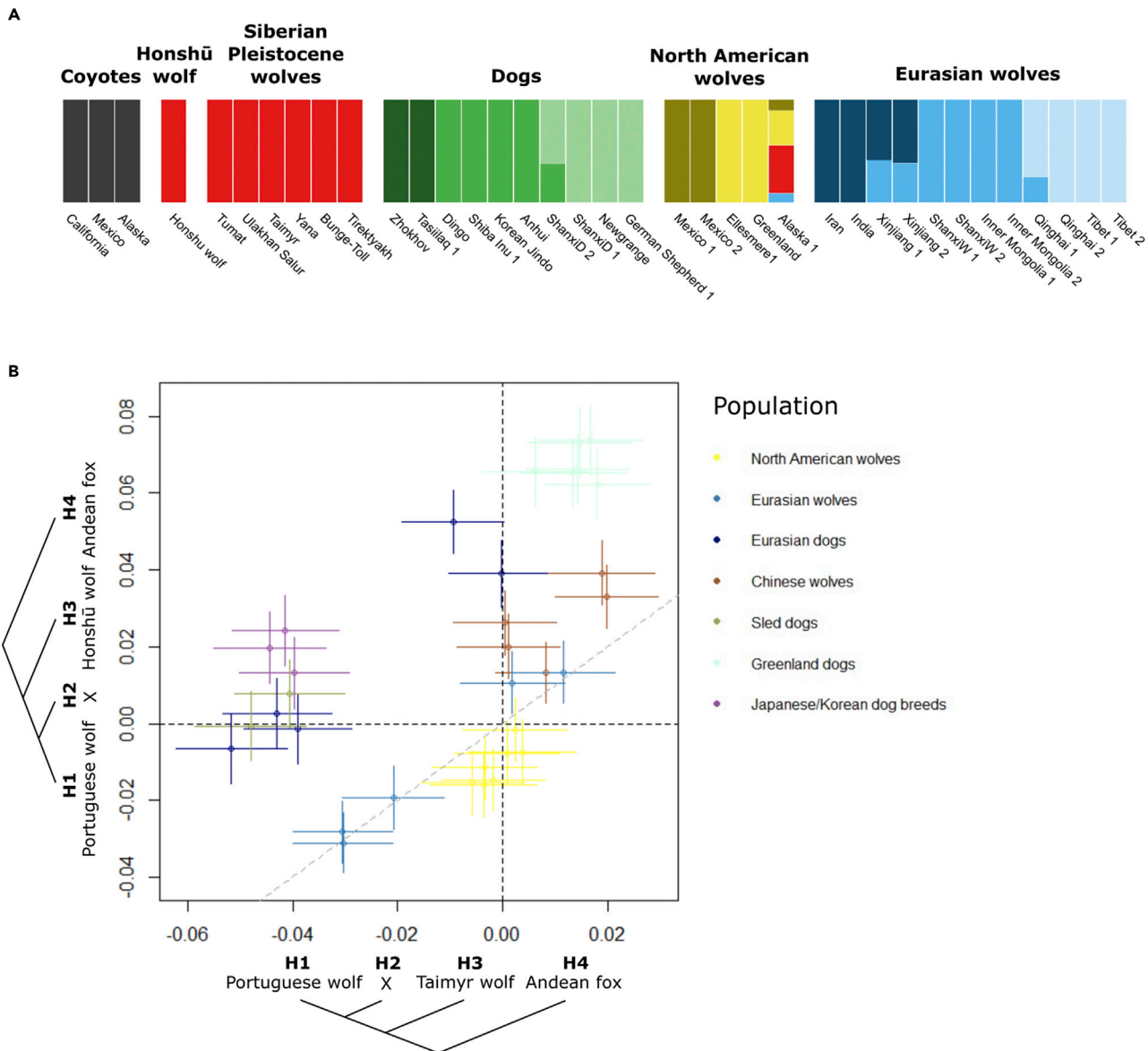
Figure 2. Gene flow between Honshū wolf, Pleistocene wolves, and present-day wolves and dogs

(A) Admixture plot for K = 6 ancestry components. Vertical bars represent single individuals. Different colors indicate the estimated ancestry components. The Honshū wolf forms a cluster with all other Pleistocene wolves (see also Figure S4).

(B) D-statistics scatterplot for the Portugese wolf in H1, samples from the reference data set in H2 (X), and Honshū wolf (y axis) or the Pleistocene wolf Bunge-Toll (x axis) in H3. Vertical and horizontal error bars correspond to three standard errors for the tests in the y- and x axis, respectively. The test involving samples with error bars that intersect the gray dotted line differ insignificantly between the Honshū wolf and Bunge-Toll in H3.

principal component (PC1). Among all the wolves included in the analysis, it placed closest to the dog cluster in the first two principal components (Figure 3B, see also Figure S6).

Japanese dog genome modeled as mixture of Honshū wolf and Chinese dog

To further examine the population history of the Honshū wolves, we tested eight putatively related populations: Japanese dogs, Chinese dogs, Greenland dogs, sled dogs, Honshū wolves, Pleistocene wolves, Eurasian wolves, and North American wolves. The chromosomes of a subset of each of these populations were then painted with the best fitting haplotypes of all remaining individuals. The resulting chromosome paintings could then be used as input for GLOBETROTTER (Hellenthal et al., 2014), which uses the
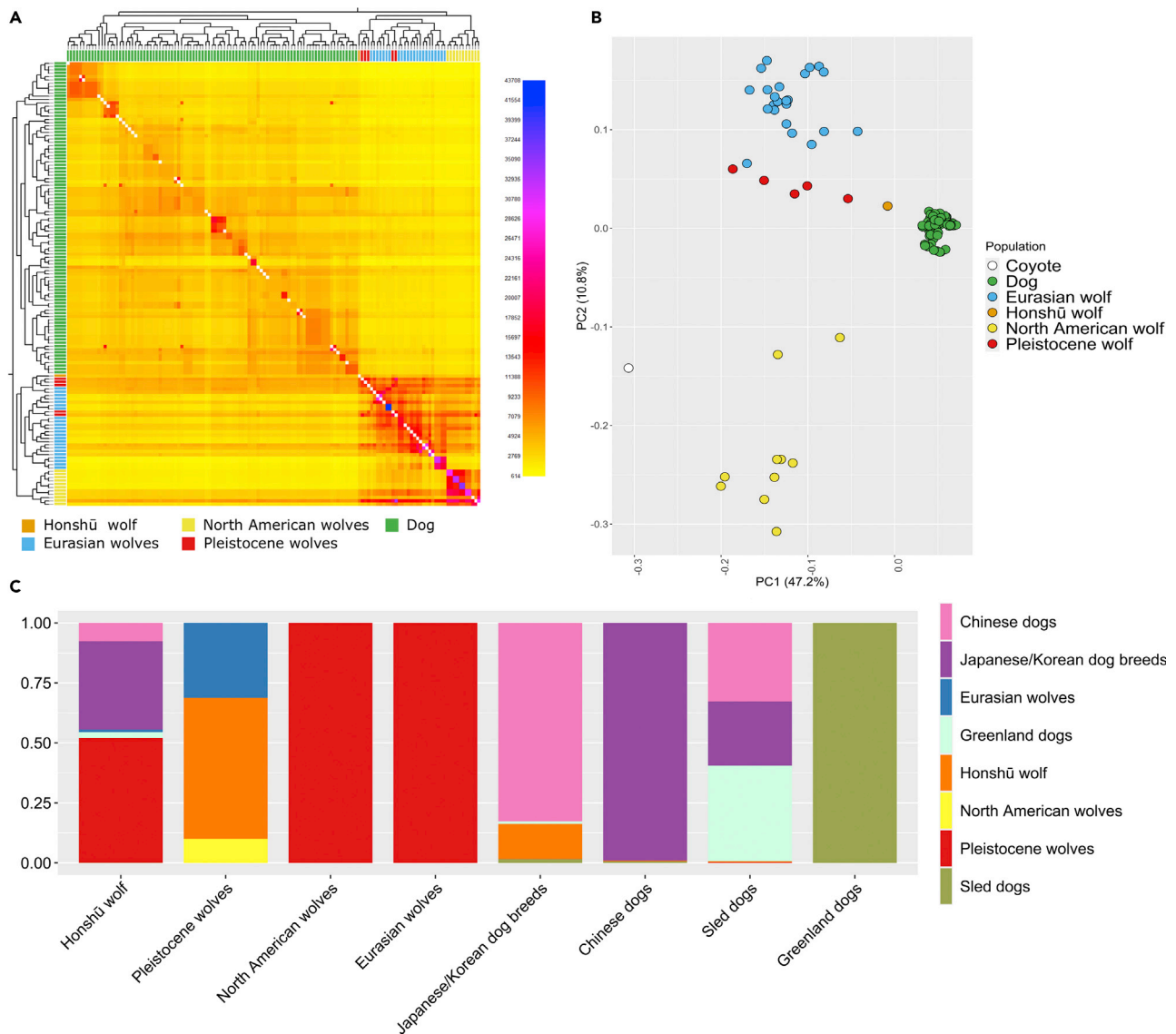
Figure 3. Haplotype-aware clustering methods modeling the ancestry of Honshū wolf and Japanese/Korean dogs

(A) Heatmap and dendrogram based on shared chromosome segments. Bottom left of the coancestry matrix is based on the unlinked model, i.e. not using genetic linkage information, while the top right shows values for the linked model (see also Figure S5). Higher values in the scale express a higher relatedness. Colored boxes at the axes indicate an individual's population (legend upper left).

(B) Principal component analysis based on the fineSTRUCTURE coancestry matrix. The Honshū wolf formed an incline with all other Pleistocene wolves in Principal component 2 (see also Figure S6).

(C) Estimated proportional contribution (y axis) from the surrogate populations (right) to the respective target population (x axis).

haplotype sharing information to describe and date admixture events involving pre-defined populations (surrogate populations) leading to the population of interest (target population).

As GLOBETROTTER requires the data of multiple individuals in the target population to infer admixture dates, we were unable to use the Honshū wolf as a target population. Instead, we chose to run GLOBE-TROTTER with Japanese dogs as the target population in order to potentially detect gene flow between the Honshū wolves and local dog populations. Using the Chinese dogs, Greenland dogs, sled dogs, Honshū wolves, Pleistocene wolves, Eurasian wolves, and North American wolves as surrogate populations, we estimated that the modern Japanese dog genome can be best described as a mixture of 93% Chinese dog and 7% Honshū wolf. The most likely scenario leading to this admixed Japanese dog population is a single

admixture event, occurring approximately 25 generations ago, between a population that is 9% Chinese dog and 91% Honshū wolf, and a population that is 100% Chinese dog. While these preliminary results indicate that Honshū wolves significantly contributed to modern Japanese dog genomes, it is most likely that the large contribution of Japanese dogs to the Honshū wolf genome confounds the results. Further studies with larger sample sizes of Honshū wolves is therefore needed to positively determine the introgression from Honshū wolves to Japanese dog breeds.

Japanese wolf genome modeled as mixture of Pleistocene wolf and Japanese/Korean dog

Finally, using the Markov chain Monte Carlo algorithm implemented in SOURCEFIND, we modeled each of the eight populations used in the GLOBETROTTER analysis—Japanese dogs, Chinese dogs, Greenland dogs, sled dogs, Honshū wolves, Pleistocene wolves, Eurasian wolves, and North American wolves—as a mixture of the remaining seven populations, i.e. all the populations except the one being modeled. The chromosome painting of the population of interest was split into 100 subsections, and each subsection was assigned to the best fitting counterpart from one of the other populations.

Using this method, we estimated that the Honshū wolf genome can be partitioned into a 52% contribution from Pleistocene wolves, 47% contribution from dogs, and a 1% contribution from present-day Eurasian wolves. Furthermore, we detected a 15% contribution from the Honshū wolf to the Japanese/Korean dog breed cluster but found no evidence for haplotype sharing between the Honshū wolf and Chinese dogs (Figure 3C). As explained above, the inference of shared ancestry in highly admixed and ill-defined populations such as wolves and dogs is computationally challenging, and the inclusion of more Honshū wolf genomes is necessary to obtain more statistically sound estimates of gene flow between dogs and the Honshū wolf. That being said, a previous mitochondrial study also documented the introgression from the Honshū wolf to some Japanese dogs (Ishiguro et al., 2009).

DISCUSSION

The results of our analyses show that the recently extinct Honshū wolf is not in the same phylogenetic clade as present-day Eurasian wolves and that only insubstantial gene flow occurred between present-day wolves and the Honshū wolf. We therefore deem it unlikely that the habitat of Honshū wolves was an LGM refugium for the common ancestor of modern wolves and dogs, as the colonization of Japan by the Honshū wolf is estimated to predate the LGM.

However, we made the unexpected discovery that the Honshū wolf specimen we sampled can be best described as a hybrid between Pleistocene wolves and Japanese dogs. Until now, Pleistocene wolves were thought to have gone extinct around the beginning of the Holocene, but the strong genetic affinity between Honshū wolves (*Canis lupus hodophilax*) and Pleistocene wolves suggests rather that the Japanese archipelago had been a refugium for Pleistocene wolves for thousands of years, where their descendants only went extinct about 100 years ago.

As the Honshū wolf specimen was one of the last of its kind after centuries of human persecution, which resulted in a drastic population decline in the 19th century, it is more than likely that the extent of dog introgression we detected was significantly lower in the Honshū wolf population before they were actively hunted. It is therefore necessary to sequence and analyze the genomes of additional Honshū wolf specimens, especially those that predate the population decline, to obtain a more accurate representation of the genetic makeup of the Honshū wolf. As of now, the high proportion of dog variants in the Honshū wolf specimen hinders our ability to quantify the extent of Honshū wolf introgression into Japanese dog breeds.

Finally, Hokkaidō and Sakhalin island remain potential candidates for LGM refugia, as our analyses only covered the more southern islands Honshū, Shikoku, and Kyūshū. Analyzing the yet understudied Ezo wolf genome might therefore be the key to resolve the mystery of the absent ancestors of present-day dogs and wolves.

Limitations of the study

This study is based on the genome of one Honshū wolf specimen, which we found to be admixed with Japanese dogs. The admixed nature of this individual limits our ability to quantify the genetic contribution from Honshū wolves to Japanese dog breeds.

CellPress
OPEN ACCESS

Resource availability

Lead contact

Further information, requests, and inquiries should be directed to and will be fulfilled by the Lead Contact, Jonas Niemann (jonas@palaeome.org).

Materials availability

This study did not generate new materials.

Data and code availability

The project accession number for the sequencing data project reported in this paper is ENA: PRJEB41490. The sample accession number reported in this paper is ENA: ERS5374233.

METHODS

All methods can be found in the accompanying Transparent methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101904.

AUTHOR CONTRIBUTIONS

M.T.P.G, M.-H.S.S. conceived the study. M.-H.S.S. did the ancient DNA lab work. J.N, S.G. performed the bioinformatic analysis. M.-H.S.S. contributed with sample collection. J.R.-M. provided computation expertise. J.N., M.-H.S.S., S.G., N.W., M.T.P.G. supervised the work. J.N, M.-H.S.S., S.G., J.R.-M., N.Y., M.T.P.G. interpreted the results. J.N, M.-H.S.S., S.G., M.T.P.G. wrote the manuscript with input from all authors. All authors read and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no conflict of interest.

REFERENCES

Dobson, M. (1994). Patterns of distribution in Japanese land mammals. Mammal Rev. *24*, 91–111.

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

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

Fritts, S.H., Stephenson, R.O., Hayes, R.D., and Boitani, L. (2003). Wolves and humans. In Wolves: Behavior, Ecology, and Conservation, L. David Mech and Luigi Boitani, eds. (University of Chicago Press), pp. 289–316.

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

Hellenthal, G., Busby, G.B.J., Band, G., Wilson, J.F., Capelli, C., Falush, D., and Myers, S. (2014). A genetic atlas of human admixture history. Science *343*, 747–751.

Imaizumi, Y. (1970). ニホンオオカミの系統的地位について. J. Mammalogical Soc. Jpn. *5*, 27–32.

Ishiguro, N. (2012). 絶滅した日本のオオカミの遺伝的系統. 日本獣医師会雑誌 *65*, 225–231.

Ishiguro, N., Inoshima, Y., and Shigehara, N. (2009). Mitochondrial DNA analysis of the Japanese wolf (Canis lupus hodophilax Temminck, 1839) and comparison with representative wolf and domestic dog haplotypes. Zoolog. Sci. *26*, 765–770.

Ishiguro, N., Inoshima, Y., Shigehara, N., Ichikawa, H., and Kato, M. (2010). Osteological and genetic analysis of the extinct Ezo wolf (Canis lupus hattai) from hokkaido island, Japan. Zoolog. Sci. *27*, 320–324.

Koblmüller, S., Vilà, C., Lorente-Galdos, B., Dabad, M., Ramirez, O., Marques-Bonet, T., Wayne, R.K., and Leonard, J.A. (2016). Whole mitochondrial genomes illuminate ancient intercontinental dispersals of grey wolves (Canis lupus). J. Biogeogr. 43, 1728–1738.

Larson, G., and Bradley, D.G. (2014). How much is that in dog years? The advent of canine population genomics. PLoS Genet. 10, e1004093.

Lawson, D.J., Hellenthal, G., Myers, S., and Falush, D. (2012). Inference of population structure using dense haplotype data. PLoS Genet. 8, e1002453.

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

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

Martiniano, R., Caffell, A., Holst, M., Hunter-Mann, K., Montgomery, J., Müldner, G., McLaughlin, R.L., Teasdale, M.D., van Rheenen, W., Veldink, J.H., et al. (2016). Genomic signals of migration and continuity in Britain before the Anglo-Saxons. Nat. Commun. 7, 10326.

Martiniano, R., Cassidy, L.M., Ó'Maoldúin, R., McLaughlin, R., Silva, N.M., Manco, L., Fidalgo, D., Pereira, T., Coelho, M.J., Serra, M., et al. (2017). The population genomics of archaeological transition in west Iberia: investigation of ancient substructure using imputation and haplotype-based methods. PLoS Genet. 13, e1006852.

Matsumura, S., Inoshima, Y., and Ishiguro, N. (2014). Reconstructing the colonization history of lost wolf lineages by the analysis of the mitochondrial genome. Mol. Phylogenet. Evol. 80, 105–112.

Ohshima, K. (1990). The history of straits around the Japanese Islands in the late-Quaternary. Quat. Res. (Daiyonki-kenkyu) 29, 193–208.

Ramos-Madrigal, J., Sinding, M.H.S., Carøe, C., Mak, S.S., Niemann, J., Castruita, J.A.S., Fedorov, S., Kandyba, A., Germonpré, M., Bocherens, H., et al. (2020). Genomes of Pleistocene Siberian Wolves Uncover Multiple Extinct Wolf Lineages. Curr. Biol. In press.

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

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

Walker, B.L. (2009). The Lost Wolves of Japan (University of Washington Press).

Supplemental Information

Extended survival of Pleistocene

Siberian wolves into the early 20th century

on the island of Honshū

Jonas Niemann, Shyam Gopalakrishnan, Nobuyuki Yamaguchi, Jazmín Ramos-Madrigal, Nathan Wales, M. Thomas P. Gilbert, and Mikkel-Holger S. Sinding
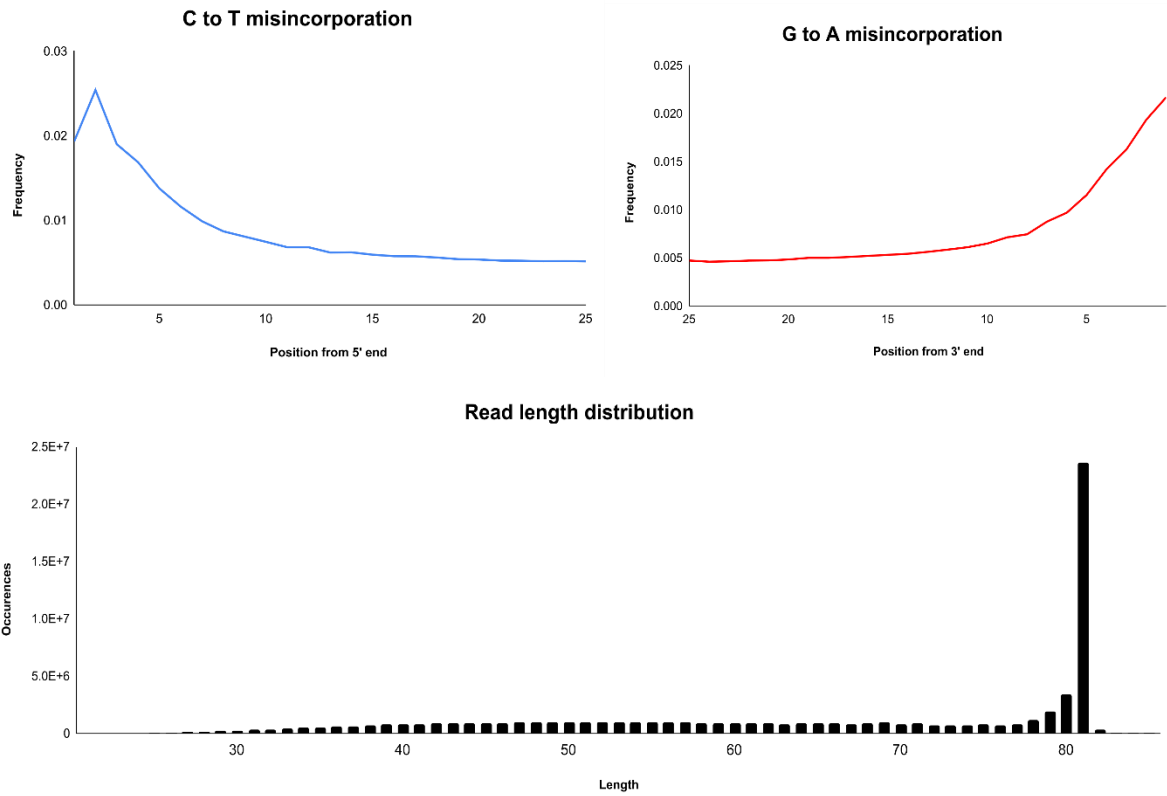
Supplemental Figures

Figure S1: Authentication of ancient DNA data from Honshū wolf, Related to Figure 2.
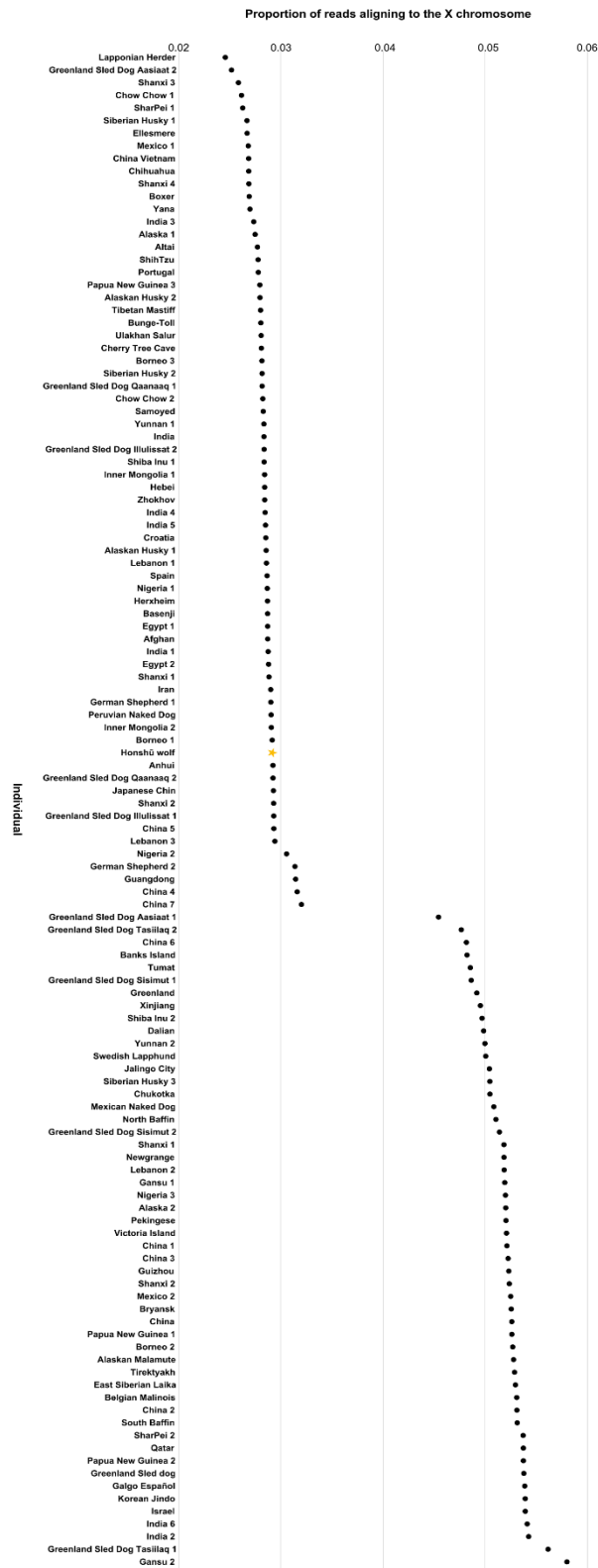
Figure S2: Sexing of the individuals in the reference panel and the Honshu wolf (orange star). The expected proportion of reads to the X chromosome is 0.053 for females and 0.029 for males, Related to Figure 2.

Log-likelihood for each K
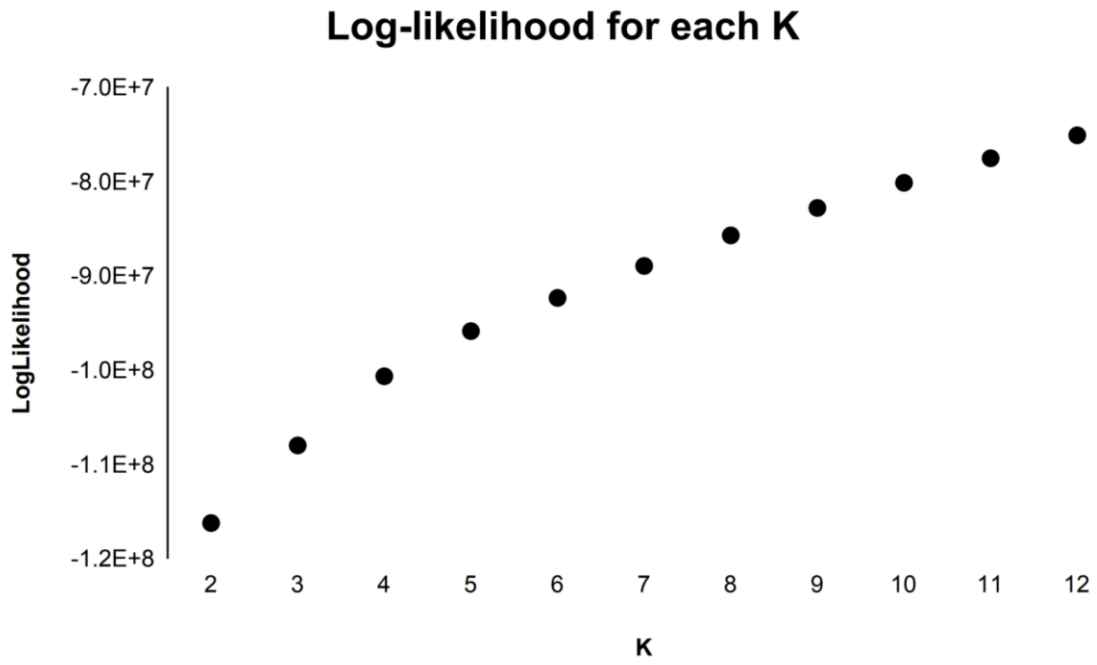
Figure S3: Log-likelihood for each K from the Admixture graphs (Figure S2), Related to Figure 2.
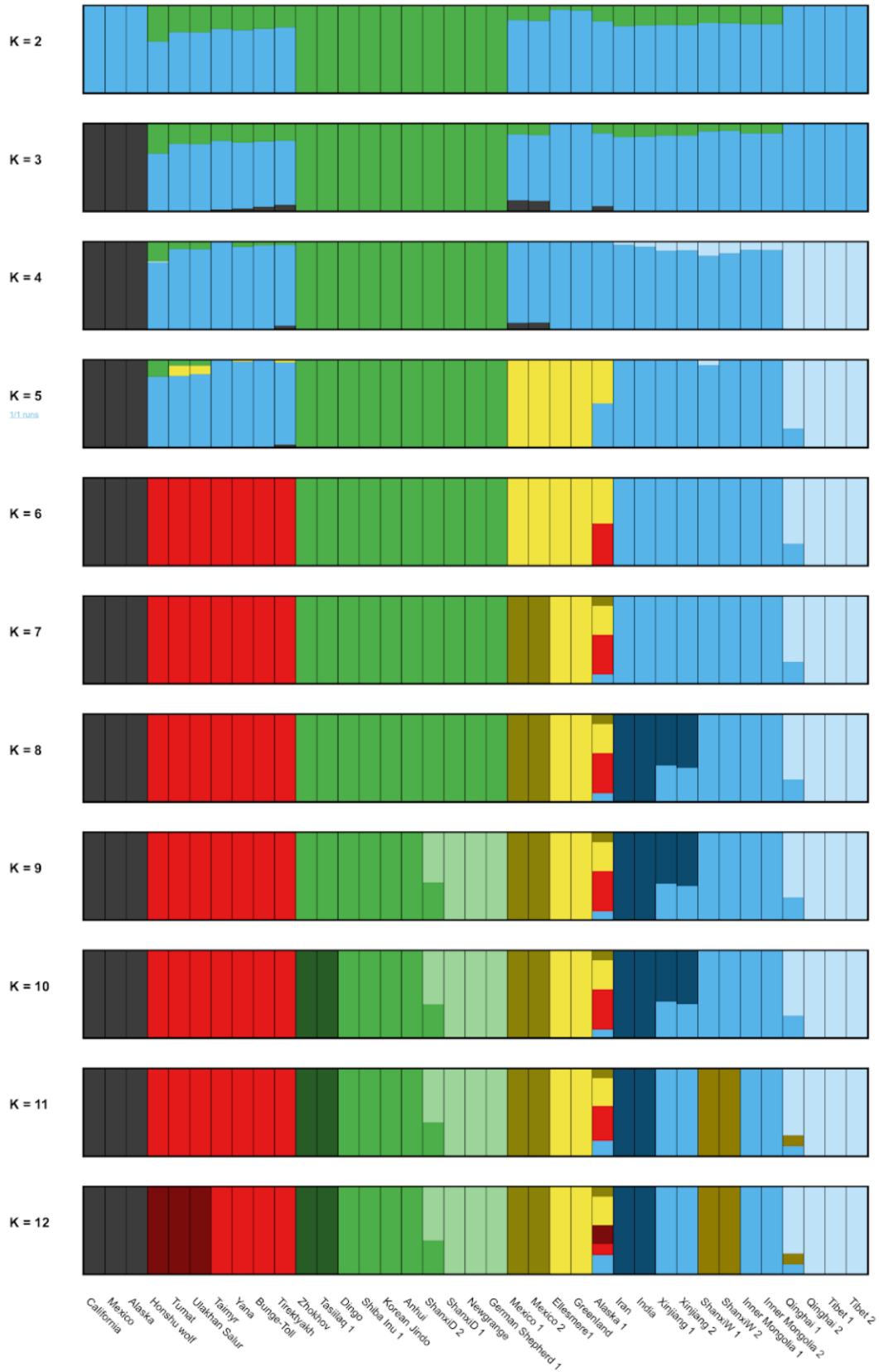
Figure S4: Extended Admixture graphs for K=2 to K=12 ancestral components, Related to Figure 2.
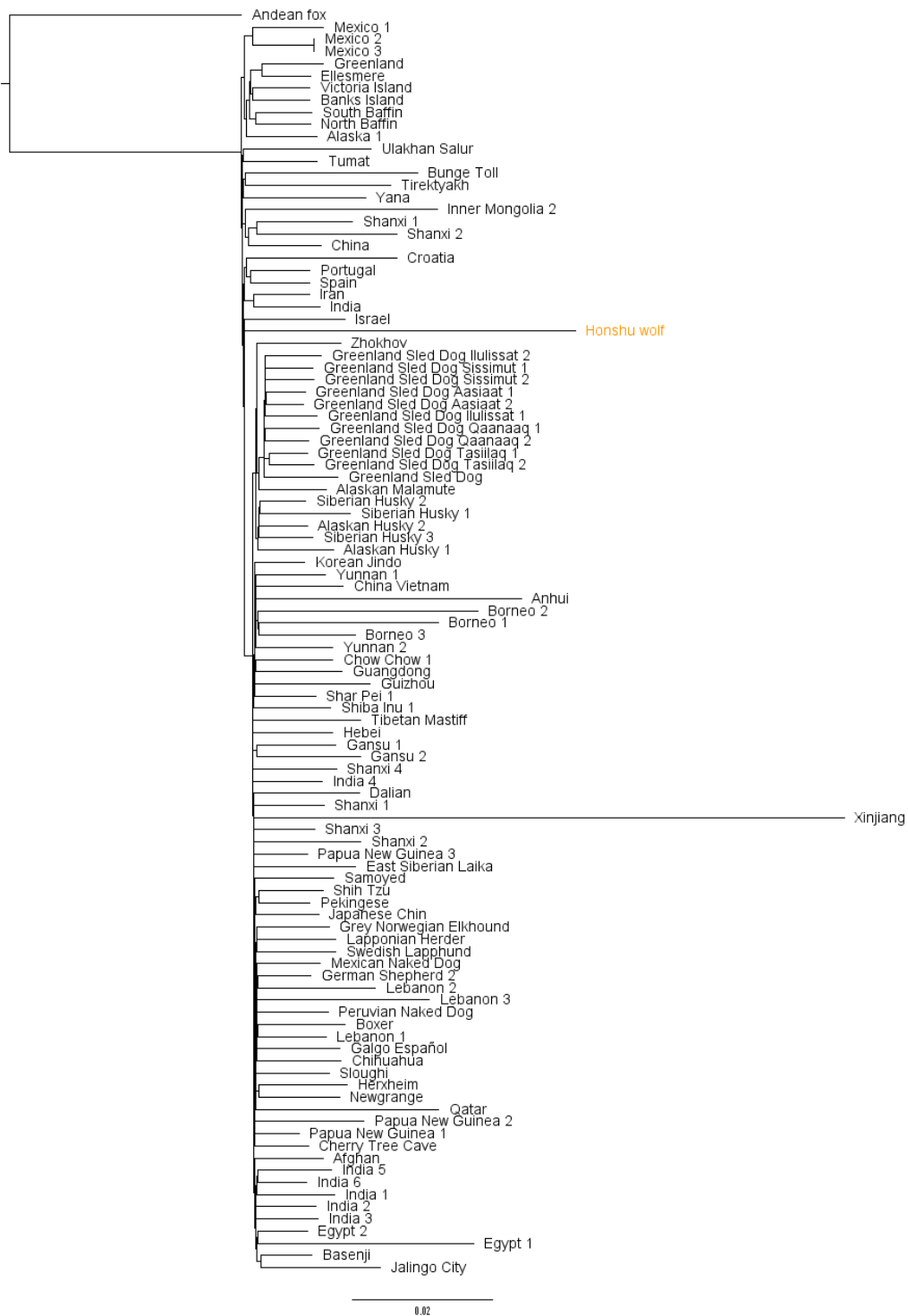
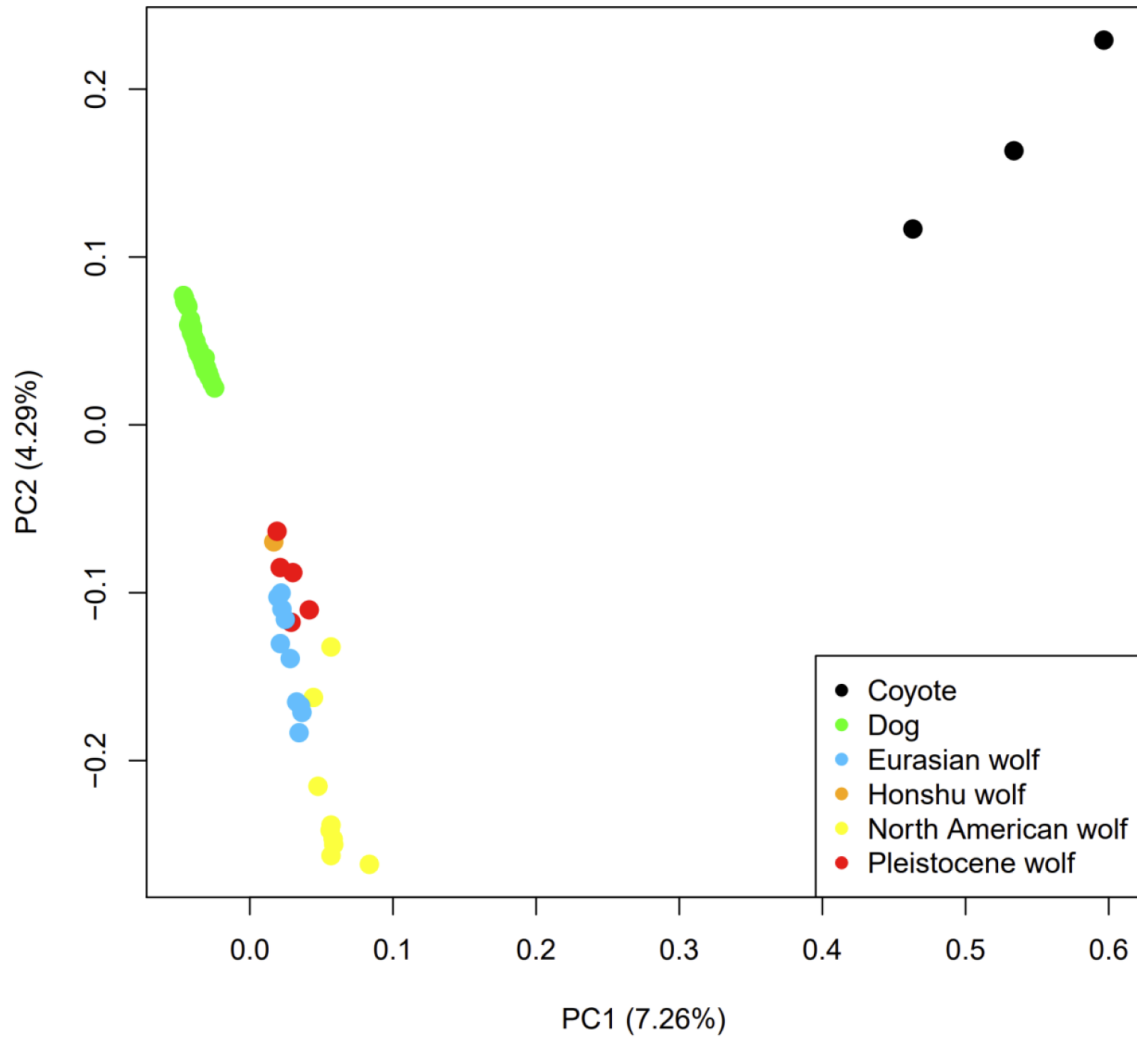**Figure S5: TreeMix output with 0 migration edges, Related to Figure 3.**

Figure S6: Principal Component analysis based on haploid data for samples in this study, Related to Figure 3.

Supplementary Tables

Table S1: qpWave results, Related to Figure 3.

Left population	Right population	Rank 0	Rank 1	Rank 2
Siberian Pleistocene wolves	Present-day wolves and dogs	0.0331104234	0.201958641	0.554427372
Siberian Pleistocene wolves and Honshū wolf	Present-day wolves and dogs	3.49E-26	0.192307254	0.428235247

Table S2: Mutation rate for the four chromosomes of ten individuals used to compute the weighted average, Related to Figure 3.

Chromosome	Mutation rate			
	1	2	3	4
Alaska 1 (W)	0.0086	0.0102	0.0091	0.0098
Banks Island (W)	0.0080	0.0087	0.0073	0.0061
Bunge-Toll (W)	0.0148	0.0146	0.0144	0.0158
Greenland (D)	0.0018	0.0014	0.0020	0.0020
Honshū (W)	0.0054	0.0048	0.0046	0.0049
India (W)	0.0096	0.0115	0.0097	0.0115
Japanese Chin (D)	0.0016	0.0020	0.0021	0.0021
Pekingese (D)	0.0037	0.0039	0.0034	0.0040
Portugal (W)	0.0085	0.0101	0.0090	0.0085
Shiba Inu 1 (D)	0.0032	0.0035	0.0030	0.0031

Table S3: Switch rate for the four chromosomes of ten individuals used to compute the weighted average, Related to Figure 3.

Chromosome	Switch rate			
	1	2	3	4
Alaska 1 (W)	27111.34	24623.14	27865.40	24252.75
Banks Island (W)	8810.56	7426.71	8509.89	6106.82
Bunge-Toll (W)	53379.92	48550.91	52917.27	50631.61
Greenland (D)	4208.91	3377.79	4565.10	3372.88
Honshū (W)	17299.89	15389.70	15069.95	14904.38
India (W)	10850.52	11956.99	10759.35	10946.93
Japanese Chin (D)	4175.42	4423.69	5070.07	4069.09
Pekingese (D)	8907.76	8262.15	6655.21	6382.16
Portugal (W)	11640.99	14283.94	11322.42	9887.85
Shiba Inu 1 (D)	7047.79	6006.83	6049.71	5935.66

Transparent Methods

Sampling collection

We sampled dry tissue from the inside of the paw of a tanned hide from a Honshū wolf, in the collections of the Natural History Museum - London, the specimen was shot in the 1800s in Chichibu District, Kotsuki, Japan and enter the museum records in 1886.

DNA extraction and shotgun sequence data generation

DNA was extracted, incorporated into libraries and sequenced as described for museum hides in Carøe et al.(Carøe et al. 2018) and sequenced on an Illumina 2500 (Illumina, San Diego, CA, USA) platform at the National High-throughput Sequencing Center, Copenhagen, Denmark.

Alignment

We used the PALEOMIX (v1.2.12)(Schubert et al. 2014) pipeline to process short reads obtained for all ancient and modern samples included in this study. As part of this pipeline, we trimmed the reads and removed adapters using AdapterRemoval2 (v2.2.0)(Schubert, Lindgreen, and Orlando 2016). Paired-end reads overlapping more than 10 base pairs - calculated using the sequences at the 3' end of the first read and the 5' end of the second read of the pair - were merged into a single long read (--collapse option). Adapter trimmed reads that were shorter than 25 bp were discarded. These processed reads were mapped against the wolf reference genome(Gopalakrishnan et al. 2017) and to the dog reference genome (CanFam3.1) using the alignment tool, bwa aln (v0.7.15; aln algorithm)(Li and Durbin 2009). Duplicate reads and reads that mapped to multiple locations in the reference genome were discarded using picard (v1.128, https://broadinstitute.github.io/picard). In order to improve the local mapping of reads that span indels, we used GATK (v3.8.0)(McKenna et al. 2010) to perform an

indel realignment step on the mapped reads for each of the samples, using no external indel databases. All analyses were performed using the alignments against the wolf reference genome, unless stated otherwise. The wolf reference genome was used in order to avoid potential reference biases when comparing a mix of ancient and modern samples.

Genotype likelihoods and admixture analysis

We computed the genotype likelihoods at variant sites of the wolf reference genome alignments using ANGSD v0.929-19(Korneliussen, Albrechtsen, and Nielsen 2014). Sites with base qualities lower than 20 and sequences with a mapping quality lower than 20 were discarded. Only biallelic transversions with data present in at least 30 out of the 37 samples were retained. All sites with minor allele frequencies below 0.01 were excluded. The final dataset consisted of 4,915,722 sites. The genotype likelihoods were then used to estimate admixture proportions between the different samples using NGSAdmix(Skotte, Korneliussen, and Albrechtsen 2013). The admixture proportions were estimated for 2 to 12 assumed ancestral populations. For each cluster, 100 replicates were computed and the admixture proportions of the replicate with the best likelihood were plotted using pong(Behr et al. 2016).

Chromosomal sex determination

The chromosomal sex of all samples was determined by calculating the proportion of mapped reads aligning to the X chromosome on the CanFam3.1 dog reference assembly. Due to the length of the X chromosome, a proportion of 0.053% of all aligned reads is expected in females that carry two copies of the X chromosome. For males, that carry a single copy of the X chromosome, the expected proportion of reads aligning to the X chromosome is 0.029, accounting for reads that originate from the Y chromosome aligning to the pseudoautosomal region of the X chromosome (about 6.6 Mb in dogs).

D-statistics

To further explore the gene flow between the Honshū wolf, Pleistocene wolves, and extant wolves and dogs, we computed D-statistics for the alignments against the wolf reference genome using ANGSD. Only biallelic transversions with a coverage higher than 3 and a base quality above 20 were considered. The Andean fox (*Lycalopex culpaeus*) was used as the outgroup for all D-statistics configurations. A weighted block jackknife procedure over 5Mb blocks was used to assess the significance of the tests. We visualized the D-statistics with the Portuguese wolf in H1, modern wolves and dogs in H2, and the Honshū wolf and Pleistocene wolf in H3 in a scatter plot.

TreeMix

We used TreeMix (Pickrell and Pritchard 2012) to investigate the phylogenetic relationship between the Honshū wolf, Siberian Pleistocene wolves, and present-day wolves and dogs.

Reads with a mapping quality lower than 30, bases with a quality lower than 20, and sites with a coverage lower than 3 were discarded from the analysis. For each individual we randomly sampled a read for every site. We considered individual samples and included 77 dogs, 26 wolves, and the Andean fox. The filtered dataset consisted of 1,374,592 segregating sites (after excluding transition sites). We ran TreeMix with 0 to 10 migration edges. We used the Andean fox as an outgroup. Since The TreeMix placement of the Honshū wolf is in a middle position between the dog and wolf clade, this is consistent with it being a mixture of Siberian Pleistocene wolf and dog ancestry. However, none of the migration edges involved the Honshū wolf, this is potentially due to the substantial gene-flow that exists among the other branches (Gopalakrishnan et al. 2018). We show the results for 0 migrations edges in Figure S6.

Principal Component analysis based on haploid data

We performed a Principal Component analysis using smartPCA (Patterson, Price, and Reich 2006) with 73 dogs, 3 coyotes, and 25 wolves. We filtered reads with a mapping quality below 30 and a base quality below 20. We then sampled a random read for each sample and each site and excluded transitions, resulting in a dataset of 4,880,683 segregating sites.

qpWave

We use qpWave (Reich et al. 2012) to further test whether the Honshu wolf carried ancestry from the dog lineage. For each sample and each site we sampled a random read after discarding reads with mapping quality lower than 30 and base quality lower than 20. All transitions were excluded. The final dataset consisted of 2,541,450 segregating sites. qpWave was run with the *allsnps* parameter set to YES. For a given group of populations (left side of the test), qpWave tests if they derive from at least N migration streams from a second group of populations (right side of the test). We expect the test to be consistent with a single migration stream if all samples in the left and the right side of the test have not exchanged gene flow after these two groups diverged from each other. For the left side of the test we used the five Siberian Pleistocene wolves (Yana RHS, Bunge-Toll-1885, Tirekhtyakh, Ulakhan Sular and Tumat 2) with and without the Honshu wolf. For the right side of the test we included four Eurasian grey wolves (Iberian, Portuguese, Croatian and Bryansk) and three dogs (Boxer, German Shepherd and Belgium Malinois). When considering the five Siberian Pleistocene wolves as the populations in the left side of the test, we could not reject the model with a single migration wave (p-value = 0.0331104234). However, when considering the five Siberian wolves together with the Honshu wolf as the populations in the left side of the test, we were able to reject a scenario with a single migration stream (p-value = 3.49164347e-26) but not a scenario with two migrations streams (p-value = 0.192307254). Furthermore, the Honshu wolf showed the highest residuals together with the three dogs, consistent with the D-statistic results (Fig. 2 b). The latter suggests that while the five

Siberian Pleistocene wolves are symmetrically related to the wolves and dogs tested, the Honshu wolf needs additional ancestry from the dog lineage.

Genotype calling

For each sample, we use the alignments to the dog reference genome to generate a VCF file using GATK's HaplotypeCaller (v3.8.0,(Poplin et al. 2018)). We ran HaplotypeCaller for each sample separately using a minimum base quality score of 20 and a minimum mapping quality score of 30. Further, we ran haplotype caller with the options "--output_mode EMIT_ALL_SITES --ERC BP_RESOLUTION" to obtain genotype calls at all sites including sites that were not variable in the sample. Subsequently, we generated a GVCF file using the GenerateGVCFs function in GATK, while still outputting genotypes at all sites. As a final step in the variant calling, we combined the GVCFs from the different samples to get joint variant calls using the SelectVariants function, and at this step, we retained only bi-allelic SNP variants while discarding indels, multi-allelic SNPs and structural variants.

The same filters that were used for the Outgroup F3 statistics analysis were applied. The final dataset consisted of 30,466,729 sites. In order to more sensitively detect gene flow and admixture, we simultaneously phased the filtered variant sites of the 136 individuals using ShapeIt2(Delaneau et al. 2013). The recombination maps for each chromosome of the dog genome were downloaded from https://github.com/clcampbell/dog_recombination(Campbell et al. 2016).

fineSTRUCTURE

To obtain an estimate for the global mutation and switch rate we ran ChromopainterV2 on four chromosomes of ten individuals and calculated a weighted average. These estimates were then used in a second ChromopainterV2 analysis to identify shared haplotypes among the samples, whereas each individual can be a donor and recipient of haplotypes. A Markov chain Monte Carlo

(MCMC) clustering algorithm was then used with 1 million burn in iterations followed by another 1 million iterations to cluster individuals based on their haplotype sharing. Every 1000th iteration was sampled. The MCMC iteration with the highest observed posterior likelihood was used to infer a phylogeny using a hill-climbing algorithm with 10,000 iterations. A Principal Component Analysis (PCA) was then performed on the resulting linked coancestry matrix.

GLOBETROTTER

Based on the fineSTRUCTURE clustering, we chose seven populations to infer potential admixture events occuring in the ancestral history of Japanese and Korean dog breeds. The previously estimated global mutation and switch rates were used to run Chromopainter v2 with the target and surrogate populations as recipients and the remaining populations as donors.

SOURCEFIND

The Chromopainter v2 results for the previously defined eight wolf and dog populations were furthermore used to predict the admixture proportions for each population, where the remaining six populations act as potential source populations. Unlike GLOBETROTTER, SOURCEFIND(Chacón-Duque et al. 2018) uses a Markov chain Monte Carlo algorithm to estimate the most likely mixture model resulting in the target population. A total of 200,000 MCMC iterations were used, with 50,000 burn-in iterations. Every 5,000th iteration was then sampled, and the mean of each source population contribution was calculated for the resulting 30 sampled iterations.

Supplemental References

Behr, A. A., Liu, K. Z., Liu-Fang, G., Nakka, P., Ramachandran, S. (2016). pong: fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics* , *32*(18), 2817–2823.

Campbell, C. L., Bhérer, C., Morrow, B. E., Boyko, A. R., & Auton, A. (2016). A Pedigree-Based Map of Recombination in the Domestic Dog Genome. *G3* , *6*(11), 3517–3524.

Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S. S. T., Sinding, M. H. S., Samaniego, J. A., Wales, N., Sicheritz-Pontén, T., Gilbert, M. T. P. (2018). Single-tube library preparation for degraded DNA. In *Methods in Ecology and Evolution* (Vol. 9, Issue 2, pp. 410–419).

Chacón-Duque, J.-C., Adhikari, K., Fuentes-Guajardo, M., Mendoza-Revilla, J., Acuña-Alonzo, V., Barquera, R., Quinto-Sánchez, M., Gómez-Valdés, J., Martínez, P. E., Villamil-Ramírez, H., Hünemeier, T., Ramallo, V., de Cerqueira, C. C. S., Hurtado, M., Villegas, V., Granja, V., Villena, M., Vásquez, R., Llop, E., … Ruiz-Linares, A. (2018). Latin Americans show wide-spread Converso ancestry and imprint of local Native ancestry on physical appearance. In *Nature Communications* (Vol. 9, Issue 1).

Delaneau, O., Howie, B., Cox, A. J., Zagury, J.-F., Marchini, J. (2013). Haplotype Estimation Using Sequencing Reads. In *The American Journal of Human Genetics* (Vol. 93, Issue 4, pp. 687–696).

Gopalakrishnan, S., Castruita, J. A. S., Sinding, M.-H. S., Kuderna, L. F. K., Raikkonen, J., Petersen, B., Sicheritz-Ponten, T., Larson, G., Orlando, L., Marques-Bonet, T., Hansen, A. J., Dalen, L., & Gilbert, M. T. P. (2017). The wolf reference genome sequence (Canis lupus lupus) and its implications for Canis spp. population genomics. *BMC Genomics*, *18*.

Gopalakrishnan, S., Sinding, M.-H. S., Ramos-Madrigal, J., Niemann, J., Samaniego Castruita, J. A., Vieira, F. G., Carøe, C., Montero, M. de M., Kuderna, L., Serres, A., González-Basallote, V. M., Liu, Y.-H., Wang, G.-D., Marques-Bonet, T., Mirarab, S., Fernandes, C., Gaubert, P., Koepfli, K.-P., Budd, J., … Gilbert, M. T. P. (2018). Interspecific Gene Flow Shaped the Evolution of the Genus Canis. *Current Biology*, *28*(21), 3441-3449

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

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* , *25*(14), 1754–1760.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303.

Patterson, N., Price, A. L., Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, *2*(12), e190.

Pickrell, J. K., Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, *8*(11), e1002967.

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M. J., Neale, B., MacArthur, D. G., Banks, E. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. In *bioRxiv* (p. 201178).

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M. V., Rojas, W., Duque, C., Mesa, N., García, L. F., Triana, O., Blair, S., Maestre, A., Dib, J. C., Bravi, C. M., Bailliet, G., Corach, D., Hünemeier, T., … Ruiz-Linares, A. (2012). Reconstructing Native American population history. In *Nature* (Vol. 488, Issue 7411, pp. 370–374).

Schubert, M., Ermini, L., Der Sarkissian, C., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M. D., Fernández, R., Kircher, M., McCue, M., Willerslev, E., Orlando, L. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature Protocols*, *9*(5), 1056–1082.

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

Skotte, L., Korneliussen, T. S., Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, *195*(3), 693–702.