

<https://doi.org/10.1038/s42003-025-09379-9>

# Continent-wide view of genomic diversity and divergence in the wolves of Asia

Check for updates

Lauren M. Hennelly<sup>1,2,3,4</sup> ✉, Bárbara R. Parreira<sup>1</sup>, Ash Noble<sup>5</sup>, Camilla H. Scharff-Olsen<sup>1</sup>, M. Çisel Kemahlı Aytekin<sup>6</sup>, Çağan H. Şekercioğlu<sup>6,7,8</sup>, Pavel Kosintsev<sup>9,10</sup>, Ladislav Paule<sup>11</sup>, Pavel Hulva<sup>12,13</sup>, Hans K. Stenøien<sup>14</sup>, Bilal Habib<sup>15</sup>, Hira Fatima<sup>16</sup>, Ghulam Sarwar<sup>17</sup>, Samara P. El-Haddad<sup>18</sup>, Alexandra Youssef<sup>18</sup>, Frank Hailer<sup>19</sup>, Xin Sun<sup>1</sup>, Nuno Filipes Gomes Martins<sup>1</sup>, M. Thomas P. Gilbert<sup>1,14</sup>, Benjamin N. Sacks<sup>3</sup>, Mikkel-Holger S. Sinding<sup>20</sup> & Shyam Gopalakrishnan<sup>1</sup>

Grey wolves (*Canis lupus*) in Asia hold most of the species' genetic diversity and many endangered populations. However, a clear understanding of the evolutionary history of wolves in Asia is lacking, hindering their conservation. We used 98 whole genomes of wolves across Eurasia to better resolve their evolutionary history and conservation status. The strongest barriers to gene flow coincided with boundaries separating the three major wolf lineages - Indian, Tibetan, and Holarctic. Wolves in the central Asian mountain ranges belonged to the Holarctic lineage and share little ancestry with the nearby Tibetan lineage. In contrast, wolves from eastern Asia share population-wide ancestry with the Tibetan lineage, which may reflect an unsampled lineage similar to the Tibetan lineage. Wolves from southwestern Asia share population-wide ancestry with the Indian lineage, likely due to old (>6 kya) admixture events. Long-term declines and recent inbreeding have left Indian and Tibetan wolves with some of the lowest levels of genetic diversity and highest realized genetic loads. In contrast, adjacent populations have some of the highest genetic diversity, due in part to admixture along contact zones. Our study highlights southern regions of Asia as hotspots of wolf diversity and the need to conserve these remaining populations.

A comprehensive understanding of the evolutionary history of a species is essential to its conservation. Species often consist of distinct populations with genetic variation unevenly distributed across their range. Understanding the contribution of historical and human-induced changes in shaping present-day patterns of genetic variation is one of the key goals in evolutionary biology. This knowledge is especially important for prioritizing populations of concern and informing management of threatened species<sup>1</sup>.

Grey wolves (*Canis lupus*) are one of the most widely distributed mammals, making them an ideal system to study how historical and human-induced factors influence present-day genetic variation. Understanding these influences is particularly important because many wolf populations are endangered or threatened. For instance, some wolf populations in Europe and North America have low genetic diversity due to population declines and inbreeding, like the Mexican and Italian wolves<sup>2,3</sup>. Most studies on the population status and genetic diversity of grey wolves have focused on North American and European portions of their range<sup>3–8</sup>. Fewer studies have focused on wolves in Asia despite the region holding most of the species' phylogenetic diversity and many endangered

populations<sup>9</sup>. The lack of studies in Asia hinders conservation efforts, and a better understanding of how evolutionary distinct populations and genetic diversity vary across Asia can benefit conservation management and decisions.

Recent work has described three divergent wolf lineages—the Indian, Tibetan, and Holarctic lineages, all of which occur in Asia<sup>10–12</sup> (Fig. S1). These lineages diverged >100,000 years ago, after which the Indian and Tibetan (also called Himalayan) lineages became isolated within separate glacial refugia<sup>10,12</sup>. Wolves from the Holarctic lineage, which span from northern Europe to Siberia, as well as North America, were highly connected over the last 100,000 years, with notable population structure emerging in more recent times<sup>13</sup>. Analyses of ancient wolf genomes suggest wolves across northern Eurasia largely resembled a panmictic population over the last 100,000 years. This extensive genetic connectivity was mainly driven by the western expansion of Siberian ancestry throughout the late Pleistocene until 10,000 years ago<sup>13</sup>. While this recent work has provided a broad picture of wolf evolutionary history in Eurasia, few studies have examined how patterns of ancestry and genomic diversity are distributed across the continent,

A full list of affiliations appears at the end of the paper. ✉ e-mail: [lauren.hennelly@rice.edu](mailto:lauren.hennelly@rice.edu)

particularly in more southern regions of Asia, where the three main lineages occur in close proximity.

For example, based on a limited number of genomes, earlier work found wolves from southwestern Asia (defined as the region spanning Iran to the Arabian Peninsula and Türkiye) reflected Holarctic ancestry with minor ancestry from more divergent taxa, such as the Indian wolf and another canid species, the African wolf (*Canis lupaster*)<sup>11,14,15</sup>. Genomic analyses have confirmed a hybrid containing African wolf, grey wolf, and domestic dog ancestry from the Sinai Peninsula of Egypt<sup>14</sup>. Additionally, hybrids between the African wolves and dogs have been genetically confirmed from Senegal and Ethiopia, suggesting the two different species can interbreed<sup>16</sup>. However, a fuller understanding of wolves from southwestern Asia has been hampered by a lack of geographically representative samples from across this region. For wolves from eastern China and extinct wolves from Japan, previous work found they contain minor ancestry attributable to ancient Pleistocene wolves from ~35–50 thousand years ago (kya)<sup>17–19</sup>. In addition, both populations retain ancestry that is more divergent than ancient wolves from northern Eurasia that date back to 100 kya<sup>13</sup>. Together, these findings suggest various wolf populations in southern regions of Asia may retain ancestry distinct from northern Eurasian wolves.

We sought to address four gaps in our knowledge about wolves in Asia through building a genomic dataset that included 20 newly sequenced wolves sampled from key regions and 81 previously published wolf genomes from Eurasia and North America. Our objectives were to (1) characterize continental-level population structure and gene flow patterns of modern-day wolf populations across Eurasia, (2) clarify the evolutionary history of wolf populations in the central Asian mountain ranges and eastern regions of Asia (3) clarify the evolutionary history of wolves in southwestern Asia, and (4) characterize the genetic diversity, inbreeding levels, and genetic load of wolves throughout Eurasia to understand long-term and recent impacts on genetic diversity.

## Results

For this study, we resequenced new whole genomes of 20 wolves from various regions in Eurasia. These regions included southwestern Asia (Lebanon,  $n = 1$ ; Israel,  $n = 1$ , Türkiye=1), Indus plains of Pakistan ( $n = 2$ ), central Asian mountain ranges (Sulieman,  $n = 2$ ; Karakoram,  $n = 2$ ; Hindu Kush,  $n = 3$ ), the Eurasian steppe (Kazakhstan,  $n = 3$ ; Russia,  $n = 3$ ), and in Europe (Slovakia,  $n = 1$ ; Ukraine,  $n = 1$ ). We also conducted additional (i.e., deeper) sequencing of a previously sequenced Indian wolf from Central India. We compiled a dataset with these newly sequenced genomes and 95 previously sequenced wolves from North America ( $n = 3$ ) and Eurasia ( $n = 78$ ), as well as dogs ( $n = 4$ ), and 10 individuals from six other canid species: dhole, *Cuon alpinus* ( $n = 1$ ); Ethiopian wolf, *Canis simensis* ( $n = 1$ ); golden jackal, *Canis aureus* ( $n = 1$ ); African wolf, *Canis lupaster* ( $n = 4$ ); coyote, *Canis latrans* ( $n = 2$ ), Andean fox; *Lycalopex culpaeus* ( $n = 1$ ). This resulted in a total of 115 samples included in the study, which we aligned to the domestic dog genome assembly (canFam3.1) (Supplementary Data 1).

### Population structure and gene flow patterns of wolves in Eurasia

Principal component analysis (PCA) and inferred individual admixture proportions of 98 Eurasian wolves supported the presence of three major clusters corresponding to the three major extant wolf lineages in Eurasia: the Indian, Tibetan, and Holarctic wolves (Fig. 1a, b, Supplementary Figs. 2 and 3). Some wolves fell in-between these main clusters, forming clines between Tibetan and Holarctic as well as between Indian and Holarctic clusters. In both cases, these intermediate individuals came from areas near the corresponding contact zones between lineages. Intermediate individuals on the Tibetan-Holarctic cline included two wolves from Tajikistan (~24% Tibetan ancestry at  $K = 6$ ) and two wolves from Ladakh, which is a region on the southwestern edge of the Tibetan plateau (Supplementary Fig. 2). Intermediate individuals on the Indian-Holarctic cline included only wolves from Pakistan (Supplementary Fig. 4). In contrast, all 13 wolves from southwestern Asia clustered tightly together nearby to other

Holarctic wolves and exhibited ~15–20% Indian wolf ancestry at  $K = 6$ . A few wolves, most sampled near Africa, showed ancestry that did not cluster with grey wolves and dogs, possibly reflecting admixture from the African wolf.

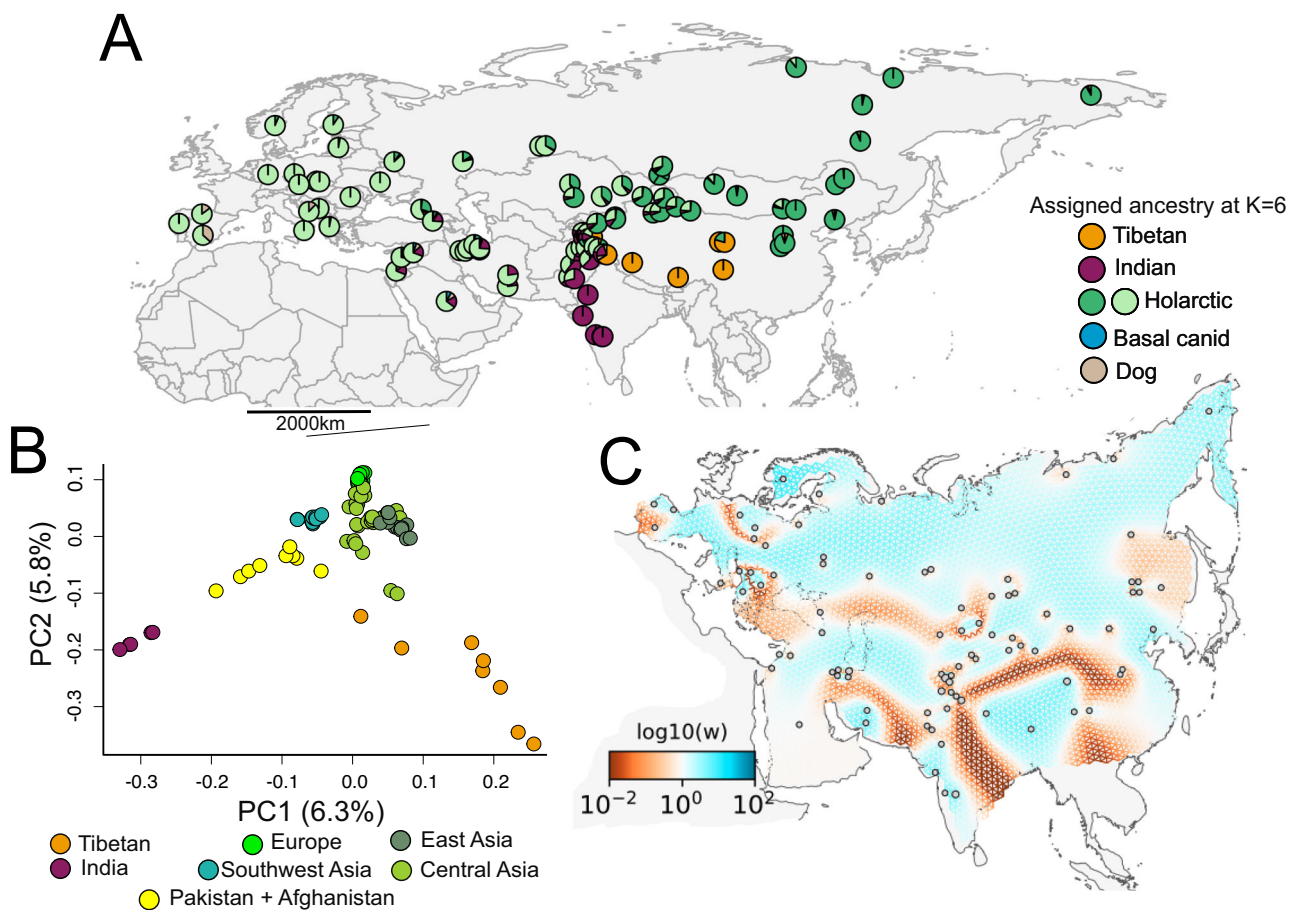
To better characterize geographic patterns of gene flow, we inferred the estimated effective migration surface (EEMS), which indicates regions of low or high gene flow relative to an isolation-by-distance model<sup>20</sup>. Mirroring the PCA and admixture results, the greatest EEMS barriers to gene flow in Eurasia corresponded to divisions among the three major lineages (Fig. 1c). The strongest barrier was along the Himalayan mountain range that separates the Indian and Tibetan lineages. A second strong barrier surrounds the Tibetan plateau, in agreement with our PCA and admixture results showing reduced gene flow between the Tibetan lineage and wolves in central and eastern Asia. We detected another barrier to gene flow positioned west of Pakistan; however, the exact position of the barrier is less clear due to a lack of sampling on the Eurasian steppe, such as in Turkmenistan and Uzbekistan. Otherwise, the population structure of Holarctic wolves across northern Eurasia can be modeled as isolation-by-distance (Fig. 1b, c)

### Evolutionary history of wolves in the central Asian mountains and eastern Asia

To explore the evolutionary history of wolves in the central Asian mountain ranges and eastern Asia, we first used  $D$  statistics to test for excess derived allele sharing with various wolf populations in Asia<sup>21</sup>. We observed wolves from both regions had a significant excess ( $Z$  score  $< -3$ ) of derived alleles shared with the Tibetan lineage, and not with any other taxa (Indian wolf, African wolf, golden jackal) (Supplementary Figs. 5–8). However, for wolves in the central Asian mountains, the extent of allele sharing with the Tibetan lineage varied widely among individuals (Fig. 2a). Only three out of 16 wolves showed significant allele sharing with the Tibetan lineage, in which all wolves from Kyrgyzstan and most wolves from Pakistan and Tajikistan had no detectable allele sharing with the Tibetan lineage (Fig. 2a; Supplementary Fig. 5). This high inter-individual variance in Tibetan derived alleles suggests gene flow with Tibetan wolves is limited, sporadic, and/or possibly recent. In contrast to wolves from central Asian mountains, we observe a consistent and significant excess of allele sharing with the Tibetan lineage for almost all wolves found in eastern China and Mongolia (Fig. 2a, Supplementary Figs. 5, 9).

It can be difficult to identify the actual source of ancestry using  $D$  statistics, because any ancestry that is as divergent as the Tibetan wolf will result in a negative  $D$  statistic. To better understand the source of ancestry within wolves of eastern Asia, we inferred a set of phylogenetic trees from different regions of the genome. We first inferred a multispecies coalescent tree in ASTRAL using 1000 regions with a length of 20 kb across the autosomes of 107 individuals<sup>22</sup>. We found Tibetan wolves formed the earliest diverging lineage within grey wolves (Fig. 2b). The rest of the wolves in Eurasia and North America clustered together, including Indian wolves, which are clustered with wolves from Pakistan and sister to those from southwestern Asia (for a detailed tree with sample names, see Supplementary Fig. 10).

Because genome-wide phylogenies can be influenced by processes such as gene flow and incomplete lineage sorting (ILS), we leveraged how genome architecture relates to phylogenetic signal<sup>23</sup>. Briefly, the X chromosome and low recombination regions are more likely to retain the historical relationship due to differences in inheritance, effective population size ( $N_e$ ), and selection<sup>24–26</sup>. The X chromosome has a smaller  $N_e$  since males only carry one copy, resulting in faster lineage sorting and, in the case of male-mediated gene flow, may better retain the historical relationship<sup>27–30</sup>. Low recombination regions experience stronger selection against introgressed ancestry due to higher linkage and a lower  $N_e$ <sup>24,31</sup>. Therefore, we inferred phylogenetic trees using the X chromosome and low recombination regions to take advantage of their association with the historical phylogenetic relationship, as demonstrated in multiple taxa, including wolves<sup>11,32–36</sup>.



**Fig. 1 | Population genetic structure and genetic connectivity of wolves across Eurasia.** **A** Sample location and individual admixture proportions at  $K = 6$  for 98 wolves across Eurasia. The admixture proportions were estimated using 98 Eurasian grey wolves, 3 North American wolves, 4 dogs, and 10 individuals from six other canid species. **B** Principal component analysis using genotype likelihoods for 98

wolves across Eurasia. Colors indicate the geographic region to which each individual wolf belongs. **C** The estimated effective migration surface (EEMS) for 97 wolves across Eurasia. Colors of the EEMS correspond to lower-than-average (red) and higher-than-average (blue) effective migration as expressed as  $\log_{10}(w)$ . A lambda of 100 was used after performing cross-validation analyses.

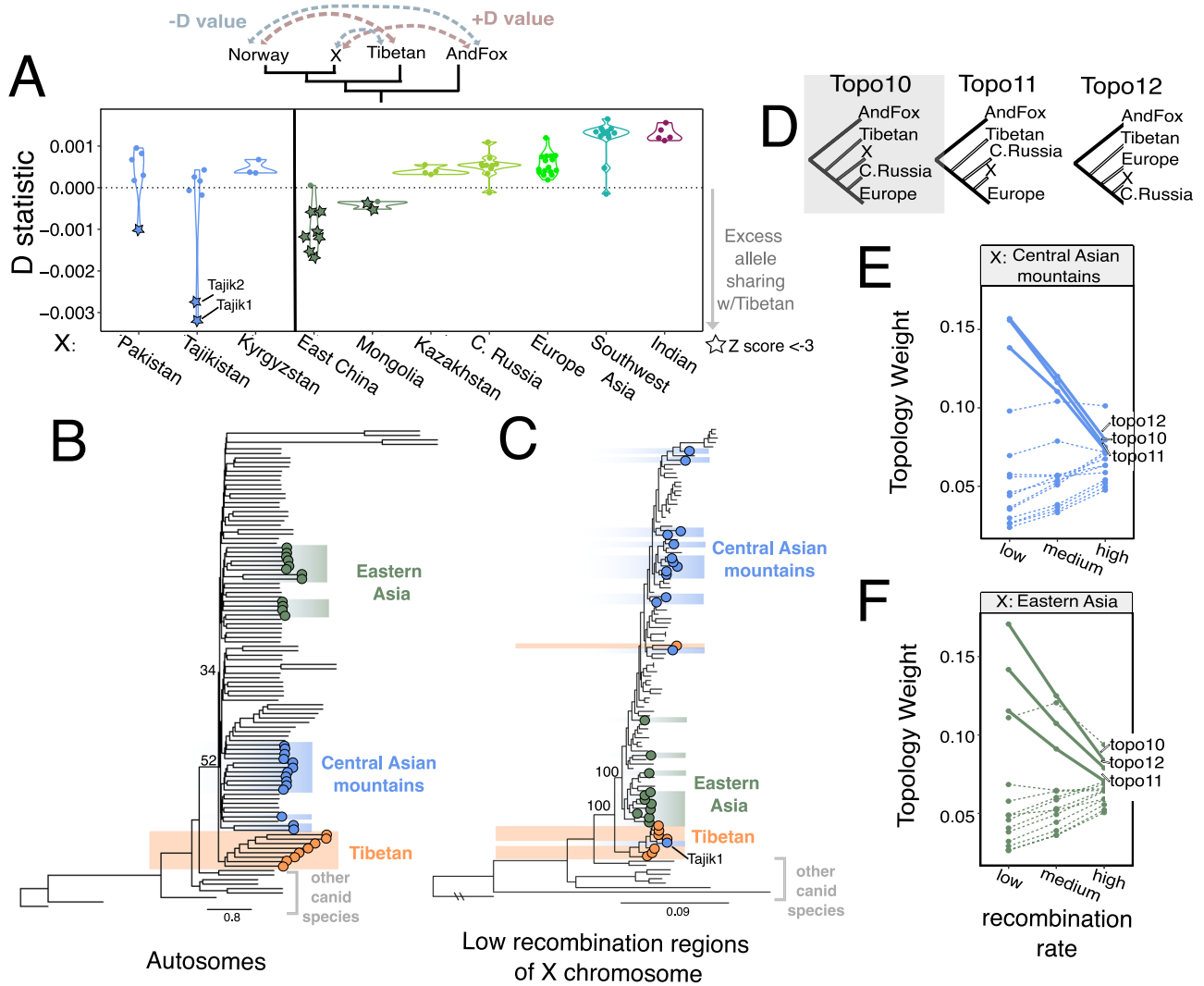
Our phylogenetic trees based on low-recombination regions (<0.2 cM/Mb) of the X chromosome and autosomes demonstrated that wolves from the Tibetan plateau form the earliest diverging branch within grey wolves. In contrast, Indian wolves clustered with wolves from Pakistan and southwestern Asia, a result that could be due to admixture between these populations (Fig. 2c, Supplementary Fig. 11). Interestingly, one of the two wolves from Tajikistan with the highest allele sharing with Tibetan wolves clustered within the Tibetan lineage, and all other wolves from Tajikistan and Kyrgyzstan clustered within the Holarctic lineage (Fig. 2c, for a detailed tree see Supplementary Fig. 12). Wolves from eastern Asia formed a well-supported clade that branched basally to the rest of the Holarctic lineage for the X chromosome phylogeny and phylogenies using low-recombination regions (Fig. 2c, Supplementary Figs. 11, 13). In addition, topology weighting analyses revealed topologies where eastern Asian wolves form the earliest diverging branch within the Holarctic lineage are more prevalent in low recombination regions of the X chromosome (Fig. 2d–f, Supplementary Fig. 14). Thus, while wolves in Tajikistan, Pakistan, and eastern Asia have similar  $D$ -values indicating significant derived allele sharing with Tibetan wolves, they have contrasting topological patterns. This is consistent with wolves from Tajikistan and Pakistan having experienced recent and/or periodic gene flow from Tibetan wolves near the secondary contact zone; however, this Tibetan ancestry has not spread widely through the population. In contrast, wolves in eastern China and Mongolia may have retained population-wide ancestry that is more divergent than wolves in northern Eurasia and similar to modern-day Tibetan wolves.

### Complex history of admixture within the wolves of southwestern Asia

Our estimated admixture proportions and  $D$  statistics suggest wolves in southwestern Asia shared ancestry with the Indian lineage and another canid species, the African wolf (Supplementary Figs. 2, 6, and 7). We detected nine wolves that showed significant allele sharing with the African wolf and these wolves were largely confined to areas most geographically close to the distributional extent of the African wolf range in the Sinai Peninsula of Egypt (Fig. 3b). We also detected one wolf from Syria having significant derived allele sharing with the golden jackal, as well as African wolf and the Indian wolf (Supplementary Fig. 8). Overall, our  $D$  statistic results suggest the divergent ancestry present in these nine individuals is most similar to the African wolf, as found in earlier studies<sup>14</sup>.

Interestingly, we found ancestry shared with the Indian lineage was nearly equal in all wolves in southwestern Asia (Fig. 3a, Supplementary Figs. 6, 9, 15). There appeared to be no decline in the admixture proportions of assigned Indian ancestry at  $K = 6$  from the contact zone in Pakistan until reaching the Caucasus mountain range and from Türkiye into Europe (Supplementary Fig. 2). There was also a lesser, yet significant ( $Z$  score < -3), signal of shared ancestry with the Indian lineage on parts of the Eurasian steppe in Central Asia, such as in Kazakhstan (Supplementary Figs. 6, 9). This geographically widespread signal contrasts with what we observed with Tibetan and African wolf ancestry, which declines quickly as distance increases from their respective distributions (Fig. 3b, c).

The balanced Indian wolf ancestry across southwestern Asia could be consistent with a historical admixture event between Indian and Holarctic

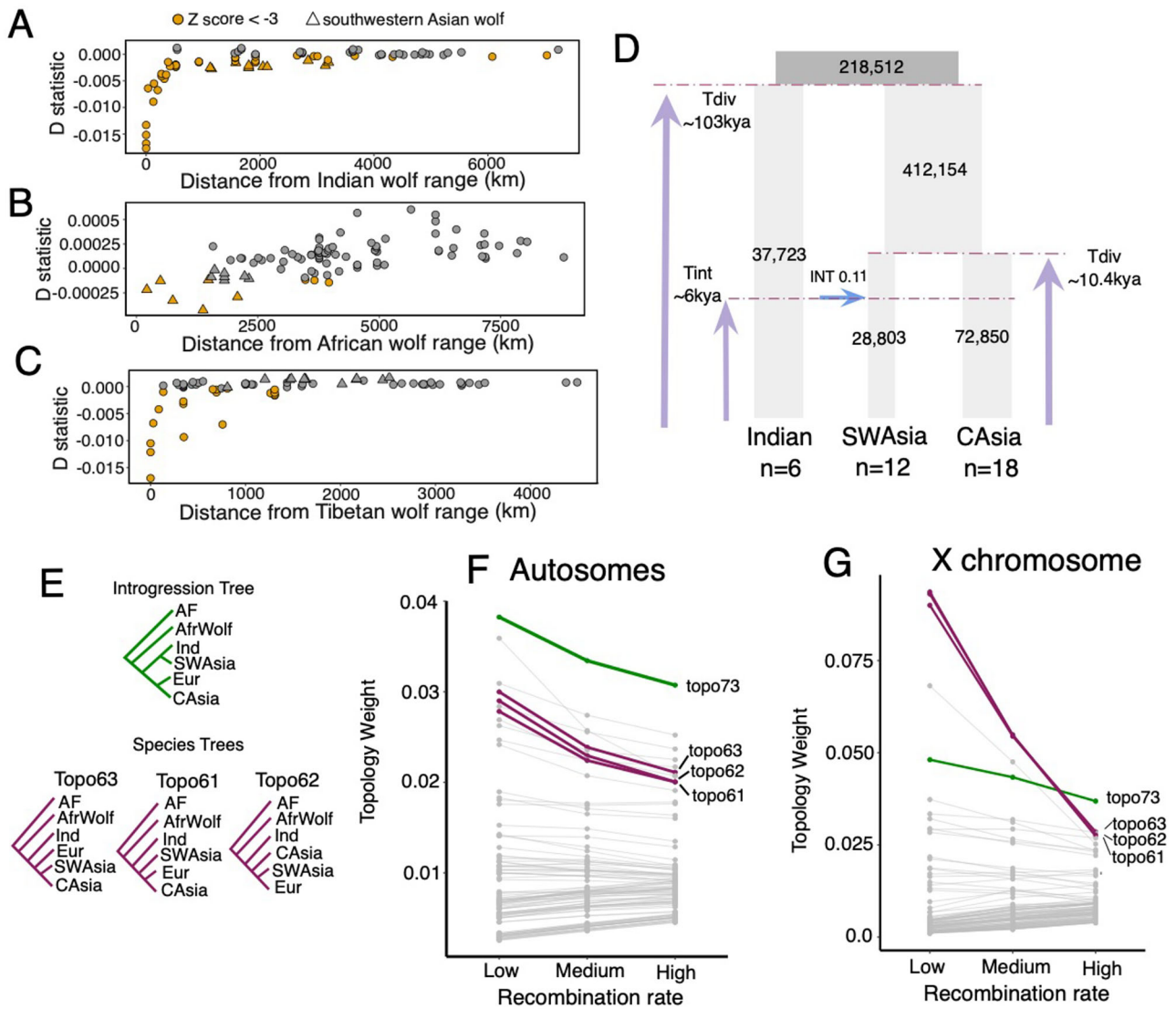


**Fig. 2 | Evolutionary history of wolves from central Asia and eastern Asia.** **A** *D* statistic results in estimating derived allele sharing between wolf individuals in Eurasia and the Tibetan lineage. The only populations that show significant allele sharing with Tibetan wolves are some wolves from Pakistan, Tajikistan, Mongolia, and eastern China. **B** Autosomal phylogeny using 97 wolves from Eurasia, 3 North American gray wolves, and 9 individuals from five other canid species. The coalescent-based tree was estimated using 1000 regions of 20 kb length that were randomly chosen across the dog reference genome. The normalized quartet score of selected nodes is shown. Colored circles and shaded color within the phylogeny indicate the individuals assigned to the Tibetan lineage (orange), central Asian mountains (high altitude Pakistan, Tajikistan, Kyrgyzstan; blue), and eastern Asia (eastern China, Mongolia, eastern Russia; green). **C** A maximum-likelihood tree

inferred with only low recombination (<0.2 cM/Mb) regions of the X chromosome using the same sample set as in **(B)**. Local posterior probability is labeled at selected nodes. **D** The three topologies with the highest average topology weightings across the X chromosome when considering wolves from the central Asian mountains or eastern Asia as the focal population X. The topology shaded in gray depicts a scenario where the focal wolf population is ancestral to wolves from Europe and Central Russia. Prevalence of 15 possible topologies in high, medium, and low recombination regions on the X chromosome where the focal population is from **E** central Asian mountains and **F** eastern Asia. For both central Asian and eastern Asian wolves, three topologies (solid lines) show a negative relationship between recombination rate and topology weighting, in contrast to the 12 other topologies (dotted lines).

lineages. To investigate the origin of Indian ancestry, we used *FASTSIMCOAL*<sup>27</sup> to test three demographic scenarios: (A) no introgression between Indian wolves and southwestern Asian wolves, (B) southwestern Asian wolves experienced an introgression event with Indian wolves, and (C) southwestern Asian wolves formed as a hybrid lineage between Indian and central Asian wolves (Supplementary Fig. 16). We obtained similar likelihood scores for the three models, suggesting no one model was strongly favored over the others (Supplementary Table 1). However, topologies that included bifurcation (models A and B) had the highest likelihoods. The model that included introgression from Indian wolves to southwestern Asian wolves (B) was preferred, with a relative likelihood of 62% (Supplementary Table 1). Estimates suggest a relatively high introgression proportion (0.11) from India to southwestern Asia, consistent with our *D* statistic results (Fig. 2d). We estimated that introgression occurred around 6000 years ago, relatively

close to the divergence time between southwestern and central Asian populations about 10,300 years ago. Therefore, if introgression did occur, it likely happened near the time of the split between southwestern and central Asian populations, explaining the similar likelihood between the divergence (A) and introgression (B) models. Importantly, both models estimate similar dates for the split between southwestern and central Asian populations, and all models date the divergence of Indian wolves from Eurasian wolves to around 100,000 years before present, providing robustness to time estimates. Additionally, both introgression and divergence models produced similar estimates for population sizes, suggesting a small effective population size (*N<sub>E</sub>*) for wolf populations in India and southwestern Asia, whereas central Asian populations and the ancestral Holarctic population in Asia had a larger *N<sub>E</sub>* (Fig. 16). Our results indicate that the Indian ancestry within southwestern Asian wolves reflects older admixture (>6 kya) and/or



**Fig. 3 | Complex history of admixture within the wolves of southwest Asia.** *D* statistic values plotted against geographic distance away from the **A** Indian wolf, **B** African wolf, and **C** Tibetan wolf distributional ranges, as defined in Fig. S1. A negative *D* statistic indicates increased derived allele sharing with the **A** Indian wolf, **B** African wolf, and **C** Tibetan wolf (A) as X based on the topology (((Norway wolf, Y), X), Andean fox). **D** The model with the highest likelihood inferred using fastsimcoal2 corresponds to an introgression event between the Indian lineage and wolves of southwestern Asia. Time is given in years and was obtained by assuming a

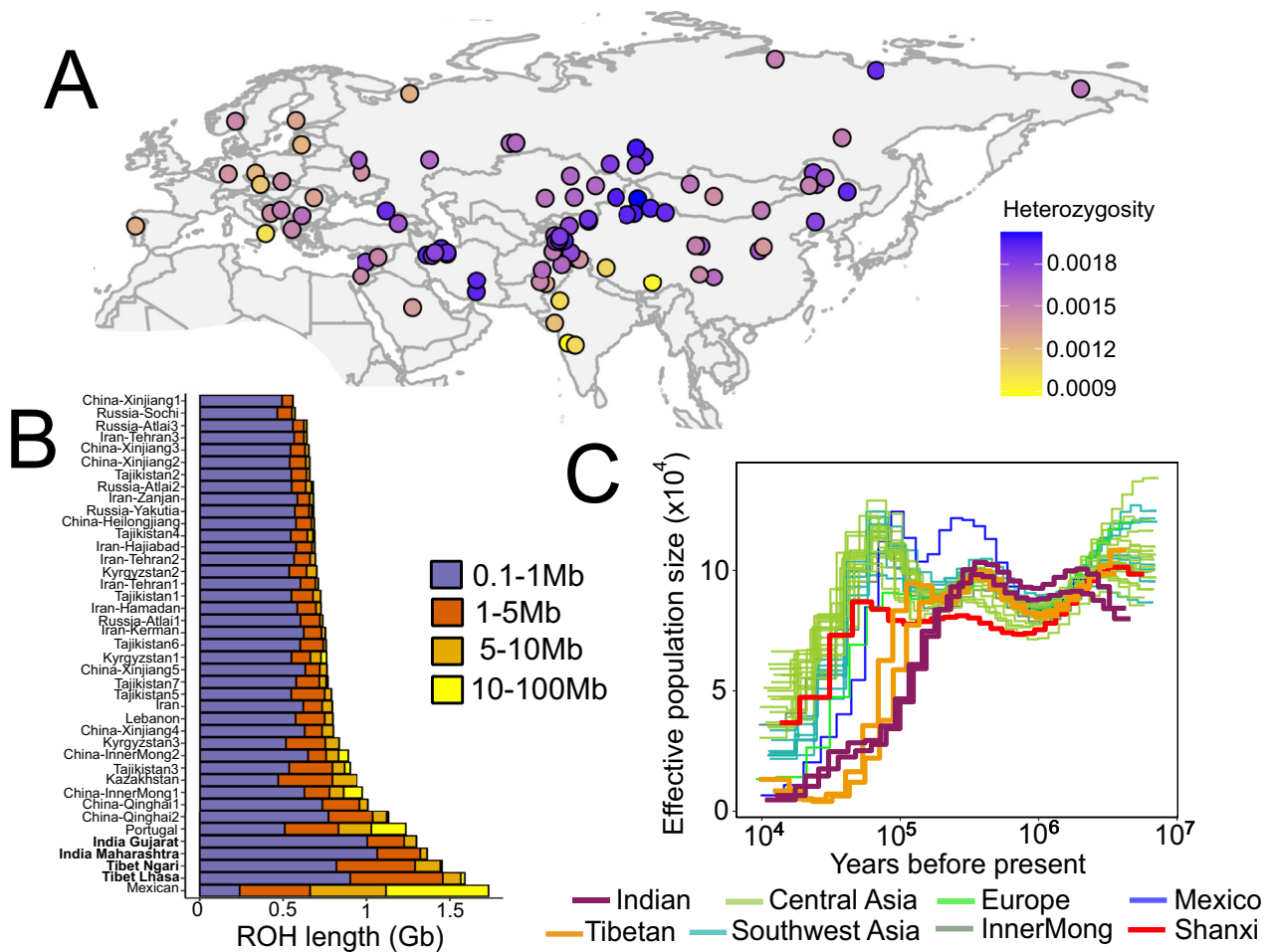
generation time of 4.4 years. **E** The topologies representing the introgression scenario (green), historical relationship (purple), and other processes such as ILS (gray) that were used to infer topology weights when using an outgroup (Andean fox, AF), the African wolf (AfrWolf), and gray wolf populations from India (Ind), southwestern Asia (SWAsia), central Asia (CAsia), and Europe (Eur). Average topology weights for each of 105 possible topologies found in low (<0.2 cM/Mb), medium (0.2–2 cM/Mb), and high (>2 cM/Mb) recombination rate regions for all 105 possible topologies across the **F** autosomes and **G** X chromosome.

shared demographic history rather than due to recent gene flow in the last thousands of years.

Finally, we inferred a set of phylogenetic trees to investigate the evolutionary history of southwestern Asian wolves in more detail. In both the autosomal tree and the X chromosome tree, we observe the Indian lineage and wolves from Pakistan and southwestern Asian wolves clustering as sister clades—nested all within the Holarctic clade (Supplementary Figs. 10 and 13). However, when excluding wolves from Pakistan and southwestern Asia, Indian wolves form the earliest diverging branch of the grey wolf phylogeny (Supplementary Fig. 16). This pattern is consistent with minor Indian wolf ancestry within southwestern Asian wolf genomes, causing phylogenetic affinity, resulting in an apparent sister relationship between the two taxa when forced into a bifurcating tree<sup>12</sup> (Supplementary Fig. 17).

To better investigate how phylogenetic signal is distributed across the genome, we quantified the frequency of 105 possible topologies using 100-

SNP windows consisting of wolves from India, southwestern Asia, Europe, and central Asia, the African wolf, and the Andean fox as an outgroup (Fig. 3e). The most frequent topology across the autosomes reflected introgression in which southwestern Asian wolves were sister to Indian wolves, including in low recombination regions (<0.2 cM/Mb) (Fig. 3g). The historical relationship, in which the Indian lineage is the earliest diverged, were among the highest weighted topologies and when considering all three topologies together, are the most prevalent phylogenetic signal in the autosomes. For the X chromosome, all topologies representing the historical relationship were most prevalent, including within low recombination regions (Fig. 3g). Unlike eastern Asian wolves, which consistently formed an earlier diverged clade within Holarctic wolves, our results indicate southwestern Asian wolves show two primary topological patterns: sister to the Indian lineage and clustered within the Holarctic clade. This may suggest southwestern Asian wolves share much of their evolutionary history with northern Eurasian wolves and derive part of



**Fig. 4 | Recent and long-term demographic history of wolves in Eurasia.**

**A** Estimates of genome-wide heterozygosity for 96 wolves from Eurasia. **B** Runs of homozygosity (ROH) for 41 wolves showing the sum of the length of ROH in four size ranges: 0.1–1 Mb (Short ROH), 1–5 Mb (Short-medium ROH), 5–10 Mb (medium-long ROH), and 10–100 Mb (long ROH). Individuals are arranged from

top to bottom by the sum length of ROH, with wolves representing the Indian and Tibetan lineages bolded. **C** Pairwise sequentially Markovian coalescent (PSMC) analysis of the demographic histories for 32 wolves from Eurasia and North America.

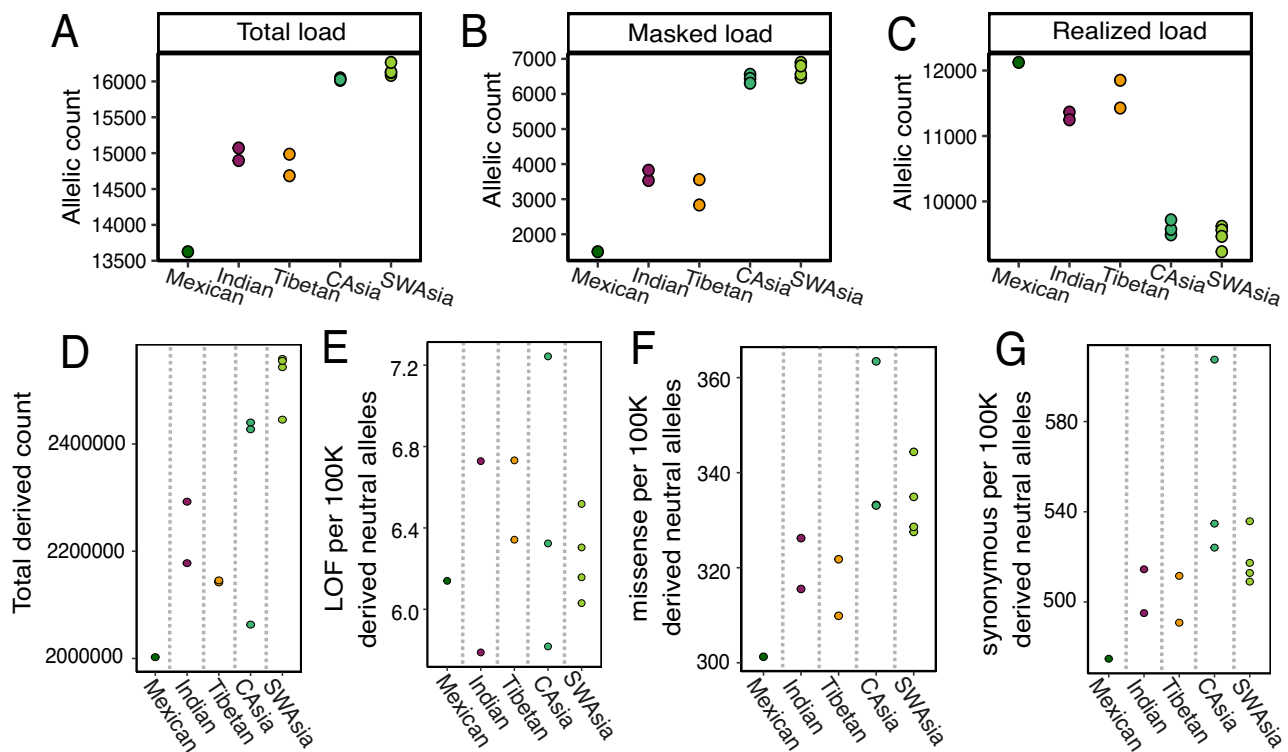
their ancestry from gene flow with an Indian wolf population similar to those found today.

### Historical and present-day impacts on genetic diversity

Our whole-genome dataset allowed us to assess patterns of genome-wide diversity of wolves across Eurasia and North America. We estimated autosomal heterozygosity using genotype-likelihoods for 97 wolves from Eurasia ( $n = 94$ ) and North America ( $n = 3$ ), and found wolves in Asia possess some of the lowest and highest levels of genetic diversity in our dataset (Fig. 4a). In addition to the well-known depleted heterozygosity in Mexican and Iberian wolves, we found exceptionally low heterozygosity in wolves from India and the Tibetan plateau. As observed in our pairwise sequentially Markovian coalescent (PSMC) analysis, this low heterozygosity is in part due to long-term population declines and/or a lack of genetic connectivity over the past >100 kya<sup>38,39</sup>. We also observe that the wolf from Shanxi, China, had a lower population size throughout the Pleistocene compared to other Holarctic wolves (Fig. 4c). This may suggest the Shanxi wolf and other wolves in southern China were less connected to northern wolf populations, however, a greater number of higher coverage genomes are needed to rule out the effects of low coverage and limited sampling. Otherwise, there is a shift to elevated heterozygosity coinciding with secondary contact zones between the Holarctic lineage and the Indian and Tibetan lineages, respectively. For example, admixed wolves in lowland Pakistan had higher genome-wide heterozygosity than wolves of Central

India, even though both populations are similarly endangered and declining. We found that the highest heterozygosity among wolves in Asia is geographically clustered in Iran and Xinjiang of China (Fig. 4a). These findings are concordant with our estimated inbreeding coefficients using genotype likelihoods<sup>40</sup> (Supplementary Fig. 18).

To further disentangle the historical and present-day processes shaping genetic diversity in wolves, we inferred runs of homozygosity (ROH) for genomes with coverage above 15× using BCFtools/RoH<sup>41</sup>. The lengths of ROHs allowed additional insight into the timing of inbreeding. Combined with knowledge of the generation time and assuming a constant recombination rate across the genome, we estimated the timing of inbreeding, where, for example, a ROH length of 5 Mb suggests inbreeding took place around 7 generations ago<sup>42</sup>. We find that Indian and Tibetan wolves show the most extensive inbreeding around 800–102 years ago compared to all other wolves in our dataset (Supplementary Fig. 19). Along with the highest number of short ROHs (0.1–1 Mb), Indian and Tibetan wolves show evidence of larger ROH as well (5–10 Mb), suggesting their low heterozygosity is a combination of long-term isolation over the last >100 kya and more recent declines in the last hundreds of years (Fig. 4b, Supplementary Fig. 19). We find the well documented and dramatic population declines of Mexican and Iberian wolves occurred more recently over the course of the last century, in line with previous work<sup>7,43</sup> (Supplementary Fig. 19). Overall, our results suggest much of the variation in genetic diversity in Asia can be explained by historical



**Fig. 5 | Genetic load in selected wolf individuals in Asia and North America.** **A** Total genetic load, **B** masked load, and **C** realized load for 13 wolf individuals within five populations. Wolves found in historically smaller populations (Mexican, Indian, and Tibetan) have a lower total load, lower masked load, and higher realized load. The total derived allele count of each wolf within the five populations, where generally Mexican, Indian, and Tibetan wolves show overall less derived alleles. The derived allele count of **D** includes all derived alleles, **E** Loss of function (LOF)

mutations classified as high impact, **F** missense mutation classified as moderate impact, and **G** synonymous mutations classified as low impact per 100,000 derived alleles found in non-coding regions, assumed to be neutral. When taking into account differences in the overall number of derived counts, Mexican, Indian, and Tibetan wolves show a reduced number of missense and synonymous mutations, but not for loss-of-function mutations, compared to central Asian and southwestern Asian wolves.

processes, including long-term isolation, shifts in connectivity coinciding with glacial refugia and secondary contact zones.

### Genetic load in wolf populations across Asia

Because Indian and Tibetan wolves show low genetic diversity, we explored how another component of genetic diversity—functional genetic diversity—compares to other wolves across Asia. Specifically, we quantified patterns of putative deleterious variation, which is expected to have a negative impact on fitness. Using SnpEff to estimate functional impact of mutations, we quantified the total genetic load (total number of derived alleles classified as high or moderate impact), realized load (homozygous state of derived alleles in high or moderate impact), and masked load (heterozygous state of derived alleles in high or moderate impact) for five selected wolf populations with assuming the derived alleles was the deleterious one<sup>44,45</sup>. Wolves from historically larger population sizes—central and southwestern Asia—had higher masked and total genetic load than wolves from historically smaller populations (Indian, Tibetan, Mexican wolves)<sup>46</sup> (Fig. 5a, b, Supplementary Figs. 20, 21). This is consistent with theoretical predictions that larger populations with higher genetic diversity should also possess more deleterious alleles in a heterozygous state<sup>47</sup>. In contrast, we found that Indian, Tibetan, and Mexican wolves have a higher realized load and a lower total genetic load (Fig. 5c) than the central Asian and southwestern Asian populations. A higher realized load in these small populations is expected because genetic drift and inbreeding will result in an increased proportion of homozygous genotypes, including deleterious ones.

Next, we conducted additional analyses to explore whether the lower total genetic load in the Indian and Tibetan wolves can be attributed to a loss of derived variants due to genetic drift in a historically small population and/or purging, whereby purifying selection removes recessive deleterious alleles

in a homozygous state<sup>48</sup>. To look for evidence of genetic drift versus purging, we asked whether deleterious alleles are lost at a higher frequency than the background neutral alleles (mutations in non-coding regions) for each individual wolf. In other words, does the fact that derived alleles are categorized as deleterious result in more frequent loss compared to neutral derived alleles within an individual, as expected for purging? Or are deleterious derived alleles lost at the same frequency as neutral derived alleles, suggesting the loss in deleterious derived alleles found in our total genetic load results could be due to genetic drift? We assessed the prevalence of sites associated with four impact categories: high impact (loss of function, LOF), moderate impact (missense), low impact (synonymous), and modifier (non-coding; here assumed to be selectively neutral). To account for differences in total derived alleles found in wolf individuals, we divided the total number of LOF, missense, and synonymous derived mutations by the total number of neutral derived alleles assumed to be under neutral processes (mutations in non-coding regions) for each individual.

We found some evidence that a lower total genetic load in Indian, Mexican, and Tibetan wolves may be due to genetic drift rather than purging via purifying selection; however, we suggest caution in interpreting our genetic load results. We do not observe Indian, Tibetan, and Mexican wolves having proportionally fewer high-impact mutations than wolves in central and southwestern Asia (Supplementary Fig. 5e). This observation suggests that genetic drift, rather than purging, may have removed putative strongly deleterious alleles at frequencies comparable to those of neutral derived alleles. Another reason for these results could be that LOF mutations were incorrectly annotated in the short-read assembled dog reference, as suggested by Smeds et al. (2023)<sup>49</sup>. In addition, among the LOF mutations that may be correct, there are so few LOF sites across the genome ( $n = 145$  on average in our dataset), and being such a small number, their frequency may

be more strongly influenced by genetic drift rather than purifying selection. For moderate and low impact mutations, we found some support that purging via purifying selection may have differentially removed moderate and low impact mutations. Specifically, we found Indian, Tibetan, and Mexican wolves have proportionally fewer missense and synonymous mutations per 100,000 derived neutral mutations compared to wolves in central and southwest Asia (Supplementary Fig. 5f, g). However, we emphasize that it is difficult to differentiate evidence of genetic drift and purging, and further work is needed to verify the functional impacts of our missense and synonymous mutations. Overall, we suggest caution when interpreting our genetic load results. Further work is needed to assess the accuracy of annotations from a short-read assembled genome, the functional impacts of mutations, and to increase the number of individuals included in each population.

## Discussion

We present a comprehensive whole-genome analysis of population structure, patterns of divergence, and genomic diversity of wolves in Asia. We find that the strongest barriers to gene flow in Eurasia coincide with contact zones among the Indian, Tibetan, and Holarctic lineage distributions. We show that wolf populations found in southern regions of Asia retain ancestries not found in Northern Eurasia, including components with some of these ancestries being similar to the African wolf, as well as Indian and Tibetan lineages. In contrast to the recurrent gene flow across northern Eurasia during the late Pleistocene, our work suggests this high connectivity did not completely replace existing wolf ancestry in southern regions of Asia<sup>13</sup>. Lastly, we find that broad-scale patterns of genetic diversity of wolves in Asia are largely explained by historical processes, such as long-term isolation within glacial refugia and contact zones.

Eurasia is the biogeographic cradle for grey wolf evolution, so it is not too surprising that southern regions of Asia have greater evolutionary diversity<sup>13,50,51</sup>. Previous work has also shown that the southern regions of Asia contain ancestry that is more divergent than modern and late Pleistocene wolves from northern Eurasia<sup>13,18,19,52</sup>. For instance, present-day wolves from southwestern Asia, eastern parts of China, and extinct wolves from Japan contain ancestry that is more divergent than a 100 kya ancient wolf genome from Siberia<sup>13</sup>. For eastern Asian wolves, we detected significant-derived allele sharing with the Tibetan lineage as P3, but not with the Indian lineage, African wolf, or golden jackal as P3 (Supplementary Figs. 5–8). While this suggests the origin of this ancestry within eastern Asian wolves is similar to the modern-day Tibetan lineage, the degree to which this ancestry is similar remains unclear. Our phylogenetic trees provide some clarity. Neither the X chromosome phylogenetic tree nor most frequent topologies using twist showed eastern Asian wolves clustering within or sister to the Tibetan lineage (Fig. 2f, Supplementary Fig. 14), as expected if gene flow between eastern Asian wolves and Tibetan wolves occurred recently<sup>53</sup>. Instead, our results showed eastern Asian wolves forming an independent lineage that diverged before the rest of the Holarctic wolves, yet after the Tibetan lineage. Their early diverging position is most evident with the X chromosome and low recombination phylogenies, a pattern that may reflect historical population structure during the late Pleistocene<sup>30</sup>. Thus, we hypothesize that wolves in eastern parts of Asia may have partial ancestry closely related but distinct to modern-day Tibetan wolves and were not fully replaced by extensive gene flow with northern Holarctic wolves during the late Pleistocene.

Consistent with this, previous work found historical genomes of wolves in southern China and Tibetan wolves occupy opposite ends of the PC2 axis, suggesting wolves from southern China may possess distinct ancestry from the Tibetan lineage<sup>54</sup>. In addition, a ~100,000 year old mitochondrial genome of a wolf from Eastern Asia represented a previously unidentified maternal lineage that diverged before other Holarctic wolves yet after the Tibetan lineage<sup>55</sup>. Fossils of wolves have been present across most regions of China over the last 100,000 years, including 12,000-year-old remains from the Southern province of Jiangxi<sup>56</sup>. During this time, much of southern China was drier, with mixed broadleaf forests occupying where monsoon

rain forests are now<sup>57</sup>. Thus, it is plausible that ancestry distinct from the modern-day Tibetan lineage could have existed within a glacial refugium in Southern China and has not been completely homogenized by expansion waves from northern Eurasian wolves. Incorporating ancient genomes from eastern China and the newly sequenced 35 kya Japanese wolf genome in future analyses would be needed to place this distinct ancestry into a spatiotemporal context during the Pleistocene<sup>19</sup>.

An unexpected result in our study was that ancestry from the Indian lineage is present in wolves across southwestern Asia. When and how the present-day population structure formed in southwestern Asia remains unclear, but our results offer some insights. Unlike eastern Asian wolves, which consistently form an early diverging lineage within Holarctic wolves, southwestern Asian wolves show differing phylogenetic relatedness across their genome—primarily clustering them within the Holarctic lineage or sister to the Indian wolf. However, the X chromosome was dominated by topologies reflecting the historical relationship with Indian wolves as the earliest diverging lineage, especially in regions of low (<0.2 cM/Mb) recombination. This historical relationship being most prevalent on the X chromosome is consistent with southwestern Asian wolves tracing their major genomic background back to the Holarctic lineage. Together, these results suggest ancient admixture with the Indian wolf contributed significantly to the genetic makeup of wolves in southwestern Asia. Ancient remains of wolves have been documented in southwestern Asia throughout the late Pleistocene, suggesting long-term continuity in wolf presence<sup>56,58–60</sup>. Genomes from these ancient wolf remains in southwestern Asia can directly test whether Indian ancestry was already present thousands of years ago, and whether there were shifts in Indian vs. Holarctic ancestry over the late Pleistocene.

Together, our work uses a combination of methodological approaches to better resolve the evolutionary history of a wide-ranging species. For many species during the late Pleistocene, their populations underwent distinct evolutionary trajectories of isolation and expansion in response to shifting climates. The process by which populations do not or do come together, and how often, is a key force in speciation and evolution. Our study's findings of evolutionarily distinct wolf populations in southern regions may represent a common pattern in widespread species, as similarly documented in red foxes (*Vulpes vulpes*), Eurasian otters (*Lutra lutra*), and brown bears (*Ursus arctos*)<sup>29,61,62</sup>. While our study addresses gaps in our knowledge on wolves, many questions remain to be answered. For example, why some wolf populations seem to have been better connected with northern populations (southwestern Asia, eastern Asia) during the late Pleistocene, while others show little gene flow and prominent contact zones (Indian and Tibetan) is an open question. With many ancient and modern genomes now available, wolves are poised to serve as a valuable model to study how intra-population gene flow, interspecies hybridization, and isolation during the late Pleistocene have contributed to present-day species' diversity.

## Conservation implications

Our study has several implications for the conservation of various wolf populations in Asia. For taxonomy, our study supports previous calls that a taxonomic revision is necessary for wolves in Asia<sup>9,63,64</sup>. While wolves from India to Türkiye are collectively described as *Canis lupus pallipes*, our study supports past work indicating that the Indian lineage is confined to India and parts of Pakistan<sup>11,15,65</sup>. In particular, our work using whole-genome sequences adds support for the proposed distribution of Indian and Tibetan wolves based on mitochondrial or limited nuclear genetic data<sup>11,66,67</sup>. Formally describing the Indian wolf as a distinct taxon from wolves in southwestern Asia is supported by these findings. Such recognition also highlights that Indian wolves are much more endangered than is currently recognized. Recently, it was estimated that ~2800 individuals occur in India and likely <300 individuals in Pakistan<sup>68–70</sup>. This population estimate and their ongoing decline placed the Indian wolf as Vulnerable under the IUCN Red List, indicating they are at high risk of extinction in the wild<sup>70</sup>. For the wolves in southwestern Asia that are currently considered "*Canis lupus pallipes*", we

find this distinct population may be geographically limited by the Bosphorus Strait in Türkiye and the Caucasus mountain range (Figs. 1a,b and 2a). Further work is necessary to understand the population genetic structure in the arid areas of southern Central Asia, such as the Karakum Desert in Turkmenistan and the Kyzylkum Desert in Uzbekistan. We suggest the morphological and genomic distinctiveness of wolves in southwestern Asia warrants subspecies status, in line with recent recommendations<sup>9</sup>.

In addition, our study clarifies the distribution and informs the conservation status of the Tibetan lineage. We found that Tibetan ancestry is highly restricted to the Tibetan plateau, with populations just adjacent (Pakistan, Tajikistan, and lowland India) having little Tibetan ancestry. With additional inference using whole-genome data in this study, we propose that the most western distributional extent of the Tibetan lineage is in the eastern Ladakh region of India<sup>15,66,67</sup>. Clarifying their distribution using genetics has already significantly impacted the conservation status of the Tibetan wolf, where they were recently assessed as Vulnerable on the IUCN Red List<sup>71</sup>. Our genetic diversity estimates for Indian and Tibetan wolves also emphasize their precarious status, as they have among the lowest genetic diversity of wolves included in our study. While their long-term small population size may have buffered against acquiring deleterious variation as suggested by our analyses of genetic load, they still face lower adaptive potential that may hinder their ability to adapt to future environmental changes<sup>72</sup>.

Our study underscores the importance of robustly describing wolves in parts of China. Wolves have been documented as far South as Guangxi in China, a province that borders Vietnam<sup>26</sup> (Supplementary Fig. 1). In 1936, wolves were reported to exist in parts of Central and South China, but were already less common compared to Northern China<sup>73</sup>. A report in 2003 concluded wolves have largely been extirpated from regions of Guangxi where they once occurred in the 1990s, but they were possibly still present in Nanling National Nature Reserve<sup>74</sup>. In response to their decline in the country, the wolf was recently added in 2021 as a protected species under Class II of the List of Wildlife under Special State Protection of China. To the authors' knowledge, we are not aware of published estimates of the population size and status, or a morphological description, of wolves in the Southern parts of China. Given their evolutionary distinctiveness, our work highlights the need for more research on their distribution, genetics, morphology, and population status to conserve what populations remain.

Overall, our study emphasizes that southern regions of Asia hold most of the evolutionarily distinct populations within grey wolves, and are also the most endangered. The Indian lineage numbers at ~2800 individuals in India and the Tibetan lineage at ~3000 individuals<sup>68,71</sup>. There are fewer than ~750 wolves left on the Arabian Peninsula, an evolutionarily unique but understudied population where only one genome has been sequenced<sup>75</sup>. It is possible that wolves in Southern China have already been extirpated. Wolves in these southern regions of Asia are impacted by numerous human-induced threats, especially for the remaining populations in Southern China and India. Our study highlights the need to prioritize these populations for conservation to preserve the full spectrum of extant wolf diversity.

## Methods

### Specimen information and processing

This study includes 20 new whole genomes from wolves from Pakistan ( $n = 8$ ), Israel ( $n = 1$ ), Lebanon ( $n = 1$ ), Afghanistan ( $n = 1$ ), Türkiye ( $n = 1$ ), Ukraine ( $n = 1$ ), Slovakia ( $n = 1$ ), Kazakhstan ( $n = 3$ ), and Russia ( $n = 3$ ). Because wolves in Pakistan are listed as Appendix 1 under CITES, wolf samples from Pakistan were exported to the USA for sequencing under CITES permit number P-50/2020. Regional permits and permission to sequence samples for museum sample collection (Steinhardt Museum of Natural History, Natural History Museum of Denmark, Yekaterinburg Museum) were obtained for the rest of the wolf samples. For all wolf samples from Pakistan and Israel, we extracted DNA from these samples using the DNAeasy Blood and Tissue Kit (Qiagen) following the manufacturer's protocol. Indexed libraries were constructed using NEBNext Ultra DNA Library Prep Kit for Illumina according to the manufacturer's instructions, then cleaned and size selected using Agencourt AMPure XP beads

(Beckman Coulter). The libraries were then sequenced on an Illumina NovaSeq 6000 S4 flowcell and paired-end at 150-bp (Illumina) at the UC Davis Genome Center. For the rest of the new wolf samples, they were extracted using the KingFisher Duo Prime Purification System (ThermoFisher). Libraries were built using Beijing Genomics Institute (BGI) library protocols that were previously optimized<sup>76,77</sup> using 10uM adaptors. These samples were sent to BGI Copenhagen for library build and sequenced on 1/8 lane each on DNBSEQ at paired-end 150. The Indian wolf from Maharashtra originated from a previous study and was resequenced to a deeper depth following protocols in Hennelly et al. 2021<sup>11</sup>. Lastly, we sequenced a wolf from Afghanistan that derived from a museum specimen from the Natural History Museum of Denmark (Museum ID: 4499). This specimen was collected in 1949, and we sampled the cartilage of the specimen. Following Gilbert et al. 2007<sup>78</sup>, the skin cartilage tissue was digested in a DDT, proteinase K-based buffer. The digest was purified following Dabney et al. 2013<sup>79</sup> using a modified binding buffer<sup>80</sup>. The sample was built into a library using a modified single-tube protocol<sup>76,77</sup>. The sample was then sequenced on a BGI DNBSEQ. We constructed a dataset comprising an additional 95 previously published canid genomes, resulting in a total of 115 canid genomes in the study (Supplementary Data 1).

### Alignment, variant calling, and filtering

We used the Paleomix v.1.2.13.2 pipeline to align our raw reads to the dog reference genome Canfam3.1<sup>81</sup>. Specifically, we used the BWA mem algorithm for modern wolf samples and the BWA backtrack algorithm for the Afghan museum sample, and used the Genome Analysis Toolkit (GATK) indel realigner to generate an indel-realigned BAM file for each sample<sup>82,83</sup>. Variant calling was performed using the GATK version 4.2.5.0. First, we used GATK Haplotype Caller to perform variant calling on each sample, then used the resulting GVCF files to perform joint genotyping with all canid individuals combined using GenomicsDBImport and GenotypeGVCF in GATK. We obtained a set of high-quality single-nucleotide polymorphisms (SNPs) using the following filtering steps in GATK:  $QD < 2.0 \mid \mid SOR > 3.0 \mid \mid FS > 60.0 \mid \mid MQ < 40.0 \mid \mid MQRankSum < -12.5 \mid \mid ReadPosRankSum < -8.0$ . Sites with a mean depth of  $>1800$  for all individuals were removed to exclude paralogues from our dataset. Lastly, we removed indels from our dataset using SelectVariants in GATK (flag-select-type SNP).

Our resulting dataset included samples that varied in their average coverage (1.38x–43.46x). To account for lower coverage ( $<5\times$ ) samples and to fully use our dataset in the study, we used genotype likelihoods for the following analyses: principal component analysis with PCAngsd<sup>84</sup>, inferring individual admixture proportions with NGAadm<sup>85</sup>, inferring genome-wide heterozygosity with ANGSD<sup>86</sup>, and estimating the inbreeding coefficient with NgsRelate<sup>40</sup>. We also used a per-site depth cutoff at more than 1/2 the average depth for higher coverage samples ( $>15\times$ ) for the PSMC and ROH analyses, a minimum per-side depth of 10x for our genetic load and fastsimcoal2 analysis. Due to their low coverage, we excluded our wolf from the Afghanistan sample that was derived from a museum specimen and the Ethiopian wolf in our genotype calling with GATK, and only used this sample for analysis using genotype likelihoods.

### Principal component analysis and individual admixture proportions

To assess the genome-wide genetic structure of the autosomes, we conducted a principal component analysis (PCA) using PCAngsd<sup>85</sup> with our total set of Eurasian grey wolf genomes ( $n = 98$ ). We inferred the PCA using genotype likelihoods from ANGSD, in which we used the SAMtools model (-GL 1) and filtered the dataset by including only properly paired reads (-only\_proper\_paired 1) from autosomal bam files, excluded reads with excess of mismatches (-C 50) or mapping quality lower than 20 (-min-MapQ20), removed transitions (-noTrans 1), and keep bases with a sequence quality above 20 (-minQ 20)<sup>86</sup>. We also retained SNPs that were present in at least 90% of individuals present (-minInd 88), resulting in ~10 million autosomal SNPs. We obtained the covariance matrix from PCAngsd and used the R function *eigen* to infer the PCA.

To estimate individual ancestry proportions, we inferred genotype likelihoods using ANGSD and ran NGSadmix<sup>85</sup> that included (1) grey wolves, dogs, and other species for  $K = 2$  to  $K = 8$  clusters and (2) only grey wolves for  $K = 2$  to  $K = 6$  clusters. For the first run, we included grey wolves from Eurasia and North America ( $n = 101$ ), dogs ( $n = 4$ ), a golden jackal ( $n = 1$ ), coyotes ( $n = 2$ ), African wolves ( $n = 4$ ), an Ethiopian wolf ( $n = 1$ ), and a dhole ( $n = 1$ ) (Supplementary Data 1). For the second run, we used only grey wolves from Eurasia ( $n = 98$ ) and North America ( $n = 3$ ). We used the same criteria in ANGSD to infer the PCA, and retained SNPs that were present in at least 90% of individuals in both runs. This resulted in 10,626,992 SNPs and 10,907,278 SNPs after filtering for the first and second run, respectively. We calculated the cross-validation score (CV) using the command `-cv`, in which the highest score indicates the value of  $K$  for the model that has the best predictive accuracy. Because of the hierarchical structure of natural populations, we examined in detail multiple levels of  $K$ , where each can provide information on subpopulation structure. We present all levels of  $K$  and cross-validation scores in the supplemental materials, but we highlight  $K = 6$  because it provides higher resolution of the subpopulation structure and admixture proportions corresponding to interpopulation and interspecific admixture. Because of higher dog ancestry detected with NGSAdmix in the two Iberian wolf samples, we removed the two Iberian wolves from the heterozygosity analysis, ROH analysis, PSMC, and timing of inbreeding analysis.

### Assess barriers to gene flow for wolves across Eurasia

We inferred the estimated effective migration surface (EEMS) to assess spatial variation in rates of gene flow among grey wolf populations in Eurasia<sup>20</sup>. The FEEMS approach utilizes a stepping-stone model to calculate genetic dissimilarities between individuals based on spatial and genetic data. These genetic dissimilarities can be visualized to illustrate departures from strict isolation by distance, thereby detecting regions of high (i.e., corridors) and low (i.e., barriers to) gene flow. For the FEEMS, we only used our dataset of 97 Eurasian wolves, where we removed the low coverage wolf from Afghanistan, filtered the dataset using VCFtools version 1.16 in which we kept sites with a minimum allele count of 3 (`--mac 3`), all of individuals represented at a site (`--geno 0`), and pruned sites in high linkage disequilibrium by removing each SNP with a  $r^2$  value of greater than 0.5 (`--indep-pairwise 50 10 0.5`)<sup>87</sup>. This resulted in ~8.7 million autosomal SNPs across our Eurasian wolves. For gathering the geographic data for each wolf sample, we had some grey wolf samples that only had location information as the country or province within a country. In these cases, we took the center coordinate within that country or province.

To run FEEMS, we first constructed a dense spatial grid using an edge width of 5 that covered the entire Eurasian continent. We initially performed a random initialization over the map to fix the estimate of the residual variance. We selected the tuning parameter, lambda, using leave-one-out cross-validation using 20 folds, where the lambda with the lowest cross-validation was selected. Our cross-validation test found a lambda value of 100 to be the optimal value of lambda (Supplemental Fig. 22).

### Phylogenomic relationships

We constructed the autosomal and X chromosome phylogeny that included all samples in our study except dogs and the Ethiopian wolf. We inferred the X chromosome phylogeny due to its differences in inheritance patterns and  $N_e$  compared to the autosome, providing additional insight into the evolutionary history of grey wolves. We filtered the autosomal and X chromosome VCF to exclude indels, keep sites with at least a minimum quality of 30 (`--minQ30`), keep only biallelic sites, and keep sites that had at least 90% individuals called (`--max-missing 0.9`). For the autosomal tree, we extracted 1000 randomly selected regions with a length of 20 kb that only included variant sites using bedtools random<sup>88</sup>. We converted each VCF file to a phylip file format using vcf2phylip.py (<https://github.com/edgardomortiz/vcf2phylip>). Next, we inferred a maximum-likelihood phylogenetic tree for each 20 kb region using IQ-Tree 1.6.12, where we estimated the best model using ModelFinder and used 1,000 ultra-fast bootstraps to infer each tree<sup>89,90</sup>.

To quantify what proportion of these randomly selected 1000 20 kb windows fell into different recombination rate regions, we downloaded the recombination map of the domestic dog from Auton et al. 2013<sup>91</sup> to use in our recombination rate analyses. We averaged the recombination rates within 100-SNP windows across the autosomes and X chromosome, as generated and detailed in a previous study<sup>11</sup>. We selected genomic regions that had an average recombination rate within each 100-SNP window of below 0.2 cM/Mb, considered as low recombination rate, 0.2 cM/Mb to 2 cM/Mb, considered as medium recombination rate, and above 2 cM/Mb, considered as high recombination rate. We used these positions of each 100-SNP window to create a bed file for low recombination, medium recombination, and high recombination regions. We then used these bed files to extract these genomic regions from our VCF using bedtools intersect. For our randomly selected 1000 20 kb regions, ~15.5% of the genomic regions fell within low recombination regions (<0.2 cM/Mb), ~61.3% fell within medium recombination regions (0.2 cM/Mb–2cM/Mb), and ~23.1% fell within high recombination regions (2 cM/Mb). This proportion is similar to the overall proportion across the autosomes that fall within low (<0.2 cM/Mb), medium (0.2–2 cM/Mb), and high (>2 cM/Mb) recombination categories, which were ~18.8%, 62.3% and 18.2% of the total 100-SNP windows across the autosomes, respectively, as reported in Hennelly et al. 2021. The inferred phylogenetic trees were used as input into ASTRAL 5.7.8, which inferred a species tree from a set of gene trees while taking into account gene tree discordance (Zhang et al. 2018, p. 90). For the X chromosome, we converted the filtered VCF file to a phylip file format using vcf2phylip.py and then inferred a maximum-likelihood phylogenetic tree using IQ-Tree 1.6.12, where we estimated the best model using ModelFinder and used 1000 ultra-fast bootstraps<sup>89,90</sup>.

In addition, we inferred phylogenetic trees of the X chromosome using only the low recombination regions across the X chromosome. This is because low recombination regions are more likely to retain the historical relationship due to higher selection against introgressed ancestry and higher lineage sorting<sup>24–26</sup>. We inferred the X chromosome phylogeny using only low recombination (<0.2 cM/Mb), using all individuals in the study except dogs and the Ethiopian wolf. We excluded the Ethiopian wolf from our analyses due to its low coverage. In addition, we inferred two more X chromosome phylogenies using only low recombination regions that excluded different wolf populations to investigate how phylogenetic relationships vary depending on what populations are included in the analysis. This is because the topology of genome-wide phylogenies can be influenced by many evolutionary processes, including admixture between different populations and/or species. Inferring phylogenetic trees with different sets of individuals can help reveal phylogenetic discordance among populations and understand the evolutionary processes underlying these discordances.

For these two X chromosome phylogenies using only low recombination regions, we inferred phylogenies where they included (1) all samples in our study excluding dogs, the Ethiopian wolf, wolves from Pakistan and Ladakh, and wolves from southwestern Asia and (2) all samples in our study excluding dogs, the Ethiopian wolf, wolves from Pakistan and Ladakh, and Indian wolves. We hypothesize that if Indian and southwestern Asian wolves are sister lineages as found in the autosomal phylogeny, then the removal of one population will not influence the topological relationship of the other due to their overall shared evolutionary histories. If the sister relationship between Indian and southwestern Asian wolves is due to minor ancestry via gene flow within southwestern Asian wolves, removing southwestern Asian wolves will result in the Indian lineage being the earliest diverging lineage. This is due to the Indian lineage no longer experiencing affinity towards the minor Indian ancestry present in southwestern Asian wolves.

For inferring the three sets of X chromosome phylogenies using only low recombination regions, we filtered all VCF files for sites with a minimum quality of 30 (`--minQ 30`), only biallelic sites, and at least 90% individuals present at each site (`--max-missing 0.9`). We then used our bed file containing coordinates to low recombination regions of the X chromosome to extract these genomic regions from our VCF using bedtools intersect. We

converted the VCF containing all low recombination regions to a phylip file format using `vcf2phylip.py` (<https://github.com/edgardomortiz/vcf2phylip>). We then ran IQ-Tree 1.6.12 with estimating the best model using ModelFinder and used 1000 ultra-fast bootstraps to infer each tree<sup>89,90</sup>.

### Assessing genome-wide gene admixture

To test for gene flow among various canid lineages and species, we performed *D*-statistic analyses using ADMIXTOOLS<sup>22</sup>. We filtered our dataset to exclude missing data (`--max-missing 1`), kept biallelic sites, and applied a minimum quality filter of 30 (`--minQ30`), along with the filters we applied using GATK. We converted plink files into eigenstrat format using the `convertf` script within ADMIXTOOLS. We calculated the *D*-statistic and *Z*-score for modern populations for four different topologies in the format (((P1, P2),P3),P4): ((Norwegian wolf MW005, X wolf); Indian wolf BH123); Andean fox), ((Norwegian wolf MW005, X wolf); Tibetan wolf TI32); Andean fox), ((Norwegian wolf MW005, X wolf); African wolf); Andean fox), ((Norwegian wolf MW005, X wolf); golden jackal); Andean fox). We selected the wolf from Norway as P1 because this population does not overlap with any closely related wild canid species, and low rates of wolf-dog hybridization in Scandinavia<sup>92</sup>. The *D*-statistic value will be negative for an excess of shared derived alleles between P2 and P3, and/or P1 and P4. Alternatively, the *D*-statistic will be positive for an excess of shared derived variants between P2 and P4, and/or P3 and P1<sup>93</sup>. For African wolves, we used multiple individuals as a population for P3, which consisted of all four of our African wolf samples from Kenya, Ethiopia, Morocco, and Algeria. For the golden jackal, we used the golden jackal from Syria, which does not show any signatures of recent admixture with any other canid species according to the individual admixture proportions (Supplemental Fig. 2).

We also quantified the frequency of different topologies in genomic windows across the genome using Twisst<sup>94</sup>. To prepare the dataset for the Twisst analysis, we first phased the filtered dataset that contained no indels, biallelic sites with a minimum quality of 30, and that were present in at least 90% of individuals using Shapeit v2.r904<sup>95</sup>. We used the dog genetic map from Auton et al. 2013<sup>91</sup> as input and a window size of 0.5 Mb. We converted the phased VCF to a geno file using the `parseVCF.py` script within [https://github.com/simonhmartin/genomics\\_general](https://github.com/simonhmartin/genomics_general). We then estimated phylogenies in 100-site windows using `phyml` with a general time reversible (GTR) substitution model with `phyml_sliding_windows.py` within [https://github.com/simonhmartin/genomics\\_general](https://github.com/simonhmartin/genomics_general). We summarized the relative prevalence of topologies using the `twisst.py` script. The set of topologies we summarized consisted of four population groups with Andean fox as an outgroup, focused on wolves in Central Asia and Eastern Asia, and five population groups with Andean fox as an outgroup, focused on southwestern Asian wolves (Supplementary Table 1). These result in 15 possible topologies for the four population groups, and 105 possible topologies using five population groups. We also quantified the frequency with which topologies occurred in windows of low (<0.2 cM/Mb), medium (0.2–2 cM/Mb), and high (>2 Mb/Mb) recombination rate regions. To do this, we averaged recombination rates within 100-SNP windows inferred in Hennelly et al. 2021—11 across the autosomes and X chromosome datasets. We then extracted 100-SNP windows inferred using `twisst` that overlapped within the low, medium, and high recombination regions for the autosomes and X chromosome using `bedtools` intersect. We averaged the topology weightings of each topology found within each recombination rate category for the autosomes and X chromosome.

### Demographic modeling with fastsimcoal2

In order to better understand the origin of southwestern Asian wolves, we compared three alternative demographic models and inferred demographic parameters under the FASTSIMCOAL2<sup>37</sup>, a composite-likelihood method that estimates historical introgression, divergence times, and population topologies from the site frequency spectrum using coalescent simulations.

We compared three demographic models (Supplemental Fig. 16) to better assess the demographic history of southwestern Asian wolves. The included three populations representing India, southwestern Asia and

central Asia samples, and were defined as (1) a divergent model with a strictly bifurcating topology separating Indian and Holarctic Asian lineages and latter southwestern Asia from central Asia populations, (2) an introgression model with the same strictly bifurcating topology followed by an introgression event from the Indian to the southwestern Asian population (3) a hybridization model, where southwestern Asian population was originated by hybridization between Indian and central Asian populations. Input files, including template and parameter estimation files for the three models, are available at Supplemental Fig. 25. A summary of all defined parameters and their search ranges is given in Supplementary Table 2.

We performed genotype calling using BCFtools, specifically keeping monomorphic sites<sup>96</sup>. We filtered our VCF containing all variants and monomorphic sites by the following parameters, keeping only those loci with >10× coverage, and >50% missing data, and removing sites SNPs violating Hardy–Weinberg equilibrium for heterozygous excess ( $p < 0.01$ ). We used the canFam 3.1 domestic dog assembly as the reference, resulting in the reference being more closely related to Holarctic wolves than to Indian wolves. Thus, we could not determine the ancestral state of each allele and therefore used the minor allele frequency spectrum (folded SFS). The mutation rate was set to  $4.5 \times 10^{-9}$  based on de novo mutations in a pedigree of wolves<sup>97</sup>.

To generate the 3D-SFS while ensuring no missing data and while considering local linkage disequilibrium (LD) patterns, we down-sampled individuals. Using the recombination map of the dog genome<sup>91</sup>, which estimates a recombination rate of 1.34 cM/Mb, we divided the genome into independent and unlinked blocks of 2 Mb. This approach ensures that sites within blocks are linked, while those between different blocks are unlinked. Within each block, we kept only SNPs without missing data by sampling three individuals from the Indian population, six from the southwestern Asia population, and nine from the central Asian population, ensuring no missing data for each sampled individual in that block. In total, we therefore retained 2,771,059 SNPs and 456,437 monomorphic sites, which we used to generate the joint 3D-SFS using the (folded) minor allele frequency spectrum.

FASTSIMCOAL2 estimates a composite likelihood, assuming independence between sites. Consequently, while the maximized likelihood converges to the correct parameter values, allowing for parameter estimation even from linked SNPs, the estimated likelihood is more accurate when derived from a set of unlinked loci<sup>37</sup>. Therefore, to compare the likelihoods obtained under different models, we need to recalculate the likelihood using unlinked SNPs. We used the parameter estimates from all retained SNPs to recalculate the likelihoods for each model. This recalculation was based on a 3D-SFS derived from a set of potentially independent SNPs, with one SNP selected per block, resulting in 1069 SNPs. We used this newly computed likelihood to apply the Akaike Information Criterion (AIC) to compare the three models. For this, we calculated delta AIC according to Excoffier et al. (2013)<sup>37</sup>, which represents the difference between the AIC value of each model and the minimum AIC value among all models. Further, we thoroughly examined the parameter values that optimized the likelihood for each model. Specifically, we compared divergence times between Holarctic and Indian lineages, the effective population sizes of each population, and the admixture proportions estimated under each model. This analysis aimed to determine whether the parameter estimates that optimized each model were consistent with the most likely model as indicated by the AIC. In addition, to evaluate whether the best-selected model could accurately reproduce the observed data, we conducted a visual examination comparing the expected and observed marginal one-dimensional SFS, as well as various two-dimensional SFS pairs.

### Long-term demographic history with PSMC

To infer the historical demography of wolves in Eurasia, we used the pairwise sequential Markovian coalescent (PSMC)<sup>38</sup>. PSMC uses a coalescent approach to estimate the history of change in effective population sizes over time. We only included the autosomal sequences of each grey wolf individual. We converted each bam file to a fasta-like consensus sequence by first

using the mpileup command with SAMtools and subsequently using BCFtools view -c to call variants and vcfutils.pl vcf2fq to convert the vcf file to fastq format (Danecek et al. 2021, p. 95). We excluded any reads that were <20 for minimum mapping quality and minimum base quality (-q 20 -Q 20) and excluded reads with excessive mismatches (-C 50). We also removed sites with more than double or less than a third of the average depth of coverage for each sample. We tested different combinations of parameters to infer the PSMC, which were “psmc -N25 -t15 -r5 -p 4 + 25\*2 + 4 + 6”, “psmc -N25 -t15 -r5 -p 2 + 2 + 25\*2 + 4 + 6”, and “psmc -N25 -t15 -r5 -p 1 + 1 + 1 + 1 + 25\*2 + 4 + 6” following previous studies on grey wolves and recent recommendations for inferring PSMC trajectories<sup>97</sup>. Some samples showed a false peak when using the parameter “psmc -N25 -t15 -r5 -p 4 + 25\*2 + 4 + 6”; therefore, we used the parameter “psmc -N25 -t15 -r5 -p 1 + 1 + 1 + 1 + 25\*2 + 4 + 6” for our final analyses (Supplemental Fig. 23).

To account for the limitations of our lower-coverage genomes (15–20×), which may miss heterozygous sites due to insufficient depth, we estimated the false negative rate (FNR) by downsampling a high-coverage grey wolf genome from the same geographic region to match the specific depth of the lower-coverage genome. To determine the best FNR, we visually compared PSMC plots from a high coverage (>20×) grey wolf from the same region with plots from its downsampled versions corrected using various FNR values (Supplemental Fig. 24). We chose the FNR that gave the closest match and applied it to the low coverage grey wolf genomes to infer their demographic history. For this, we used a mutation rate of  $4.5 \times 10^{-998}$  and a generation time of 4.4 years<sup>99</sup>.

### Genetic diversity and runs of homozygosity

We calculated the genome-wide heterozygosity for each wolf individual above 5x using ANGSD<sup>86</sup>. We first estimated the folded site allele frequency with the realSFS program in ANGSD (-doSaf) using genotype likelihoods with the SAMtools model (-GL 1) and using the Andean fox genome to determine ancestral state. We filtered each individual bam file to keep only autosomal chromosomes, only properly paired reads (-only\_proper\_paired 1), excluded reads with excess of mismatches (-C 50), excluded reads with a mapping quality (-minMapQ 20), excluded reads with a depth less than 4 (-setMinDepth 4), and kept bases with a sequence quality above 20 (-minQ 20). To calculate the genome-wide heterozygosity, we divided the number of heterozygous sites by the total number of sites for each sample.

We inferred runs of homozygosity (ROH) using BCFtools/RoH<sup>41</sup>. Along with the GATK filtering criteria, we included only biallelic SNPs that had all individuals present per site (--max-missing 1, --min-alleles 2, --max-alleles 2). We also included SNPs that had a depth of more than one-third and less than double the average depth of coverage for each sample, where we chose samples that were >15×. For the BCFtools/RoH analysis, we fixed the alternative allele frequency to 0.4 (-AFdflt 0.4; default setting) and used the dog recombination map from Auton et al. 2013<sup>91</sup> to account for recombination hotspots (--genetic-map). We kept ROHs that had a quality score of at least 80 and excluded ROHs that were below 100 kb in length in our analysis. We inferred the timing of inbreeding by the equation  $g = 100(2rL)$ , where  $g$  corresponds to generation time,  $L$  to the length of ROH in Mb, and  $r$  as the average genome-wide recombination rate (1.34 cM/Mb, averaged from the recombination rates in Auton et al. 2013<sup>91</sup>). We note this method does not account for recombination rate variation across the genome, and therefore, our inbreeding timing dates should be considered as estimates.

### Inbreeding coefficient

We inferred the per-individual inbreeding coefficient ( $F_{ind}$ ) using genotype likelihoods, where the inbreeding coefficient is the probability of identity by descent. We used ANGSD to infer genotype likelihoods and allele frequencies for samples above 4x coverage with the following parameters: -GL 2 -checkBamHeaders 0 -trim 0 -C 50 -baq 1 -skipTriallelic 1 -minMapQ 20

-minQ 20 -minInd 88 -uniqueOnly 1 -remove-bads 1 -noTrans 1 -doGlf 3 -doMajorMinor 1 -doMaf 1 -minMaf 0.0001 -SNP\_pval 1e-6. We then used NgsRelate v2 with option -F1 to infer the inbreeding coefficients for each Eurasian wolf individual<sup>40</sup>.

### Genetic load

We selected 14 individuals with >20× sequencing coverage from India ( $n = 2$ ), the Tibetan plateau ( $n = 2$ ), southwestern Asia ( $n = 4$ ), Central Asia ( $n = 4$ ), and the Mexican wolf ( $n = 1$ ). We selected these wolves to assess the genetic load of Indian and Tibetan wolves, in which we compared Indian and Tibetan wolf genetic load with two adjacent, genetically diverse populations, and a Mexican wolf individual with previously documented very low genetic diversity. We used SnpEff to annotate a VCF, including only biallelic non-indel sites with no missing data, a minimum depth of 10, a maximum depth of 60, a minimum allele count of 2, and a minimum sequence quality (--minQ) of 30.

To polarize our dataset, we followed the method by Smeds et al. (2022)<sup>49</sup> to map two outgroups to the dog reference genome at only sites with a minimum depth of 4 at both outgroup and included nonvariant sites. We used Dhole (*Cuon alpinus*) and Andean fox (*Lycalopex culpaeus*) as our outgroups. We then performed pseudo-haploidization of the outgroup genomes and kept only sites in which the allele agreed in both outgroups (Smeds et al. 2022). We used this to add the ancestral allele information to our annotated VCF, and assumed that the derived allele was the deleterious allele<sup>49</sup>.

We counted the number of derived alleles and whether they were in a homozygous or heterozygous state for variants predicted by SnpEff to have high, moderate, and low impact on protein function. High impact is predicted to cause loss of function or protein truncation (stop gain, frameshift mutation); moderate impact is predicted to change protein effectiveness (missense variant), and low impact is assumed harmless, such as a synonymous variant. We also calculated the following metrics: total load (the total number of derived alleles classified as high or moderate impact), realized load (homozygous state of derived alleles in high or moderate impact), and masked load (heterozygous state of derived alleles in high or moderate impact).

To investigate evidence of differential loss of derived variants for different types of mutations, we assessed the relative amounts of mutational categories given by SnpEff. These mutational categories are: Loss of Function (LOF) mutations that imply a high predicted impact on protein structure, missense mutations (moderate impact), synonymous mutations (low impact), and “modifier” mutations found in non-coding regions (no impact). We calculated the count of derived alleles and the number of heterozygous and homozygous derived allele calls in each of the four mutational categories for each wolf individual. To account for differences in overall derived allele counts among individuals, we calculated the average derived allele count per 100,000 derived neutral alleles for total mutational load, LOF, missense, and synonymous mutations. We considered the modifier category as neutral derived alleles. Specifically, we calculated mutational load as: (total derived alleles in LOF and missense/total derived alleles in modifier category) multiplied by 100,000.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

All raw reads are publicly available on the National Center for Biotechnology Information Sequence Read Archive under Project ID PRJNA1285574, with accession numbers SRR35174885- SRR34347608. Results and tables of the analyses can be found on GitHub: [https://github.com/hennelly/Asia\\_wide\\_wolf\\_genomics](https://github.com/hennelly/Asia_wide_wolf_genomics).

### Code availability

The bioinformatic scripts can be found on GitHub: [https://github.com/hennelly/Asia\\_wide\\_wolf\\_genomics](https://github.com/hennelly/Asia_wide_wolf_genomics).

Received: 12 February 2025; Accepted: 4 December 2025;  
Published online: 24 December 2025

## References

- Hogg, C. J. Translating genomic advances into biodiversity conservation. *Nat. Rev. Genet.* **25**, 362–373 (2024).
- Fredrickson, R. J., Siminski, P., Woof, M. & Hedrick, P. W. Genetic rescue and inbreeding depression in Mexican wolves. *Proc. R. Soc.* **274**, 2365–2371 (2007).
- Battilani, D. et al. Beyond population size: whole-genome data reveal bottleneck legacies in the peninsular Italian wolf. *J. Hered.* **116**, 10–23 (2025).
- Schweizer, R. M. et al. Genetic subdivision and candidate genes under selection in North American gray wolves. *Mol. Ecol.* **25**, 380–402 (2015).
- Kardos, M. et al. Genomic consequences of intensive inbreeding in an isolated wolf population. *Nat. Ecol. Evol.* **2**, 124–131 (2018).
- Robinson, J. A. et al. Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. *Sci. Adv.* **5**, eaau0757 (2018).
- Taron, U. H. et al. A sliver of the past: the decimation of the genetic diversity of the Mexican wolf. *Mol. Ecol.* **30**, 6340–6354 (2021).
- Lobo, D., Lopez-Bao, J. V. & Godinho, R. The population bottleneck of the Iberian wolf impacted genetic diversity but not admixture with domestic dogs: a temporal genomic approach. *Mol. Ecol.* **32**, 5986–5999 (2023).
- Werhahn, G., Senn, H., Macdonald, D. W. & Sillero-Zubiri, C. The diversity of genus *Canis* challenges conservation biology: a review of available data on Asian wolves. *Front. Ecol. Evol.* **10**, 10.3389 (2022).
- Wang, M. S. et al. Ancient hybridization with an unknown population facilitated high-altitude adaptation of canids. *Mol. Biol. Evol.* **37**, 2616–2629 (2020).
- Hennelly, L. M. et al. Ancient divergence of Indian and Tibetan wolves revealed by recombination-aware phylogenomics. *Mol. Ecol.* **30**, 6687–6700 (2021).
- Wang, M. S. et al. Genome sequencing of a gray wolf from Peninsular India provides new insights into the evolution and hybridization of gray wolves. *Genome Biol. Evol.* **2**, evac012 (2022).
- Bergstrom, A. et al. Grey wolf genomic history reveals a dual ancestry of dogs. *Nature* **607**, 313–320 (2022).
- Gopalakrishnan, S. et al. Interspecific gene flow shaped the evolution of the Genus *Canis*. *Curr. Biol.* **28**, P3441–P3449 (2018).
- Hennelly, L. M. et al. Genomic analysis of wolves from Pakistan clarifies boundaries among three divergent wolf lineages. *J. Heredit.* **115**, 339–348 (2023).
- Mallil, K. et al. Population genetics of the African wolf (*Canis lupaster*) across its range: first evidence of hybridization with domestic dogs in Africa. *Mamm. Biol.* **100**, 645–658 (2020).
- Niemann, J. et al. Extended survival of Pleistocene Siberian wolves into the early 20th century on the island of Honshu. *iScience* **24**, 101904 (2020).
- Ramos-Madrigal, J. et al. Genomes of Pleistocene Siberian wolves uncover multiple extinct wolf lineages. *Curr. Biol.* **31**, 198–206 (2021).
- Segawa, T. et al. Paleogenomics reveals independent and hybrid origins of two morphologically distinct wolf lineages endemic to Japan. *Curr. Biol.* **32**, 2494–2504 (2022).
- Marcus, J., Ha, W., Barber, R. F. & Novembre, J. Fast and flexible estimation of effective migration surfaces. *eLife* **10**, e61927 (2021).
- Green, R. E. et al. A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
- Zhang, C., Rabiee, M., Sayarri, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform.* **19**, 153 (2018).
- Burbrink, F. T., DeBaun, D., Foley, N. M. & Murphy, W. J. Recombination-aware phylogenomics. *Trends Ecol. Evol.* **40**, 900–912 (2025).
- Pease, J. B. & Hahn, M. W. More accurate phylogenies inferred from low-recombination regions in the presence of incomplete lineage sorting. *Evolution* **67**, 2376–2384 (2013).
- Schumer, M. et al. Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* **360**, 565–660 (2018).
- Martin, S., Davey, H., Salazar, C. & Jiggins, C. D. Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biol.* **17**, e2006288 (2019).
- Schaffner, S. T. The X chromosome in population genetics. *Nat. Rev. Genet.* **5**, 43–51 (2004).
- Webster, T. H. & Sayres, M. A. W. Genomic signatures of sex-biased demography: progress and prospects. *Curr. Opin. Genet. Dev.* **31**, 62–71 (2016).
- de Jong, M. J. et al. Range-wide whole-genome resequencing of the brown bear reveals drivers of intraspecies divergence. *Commun. Biol.* **6**, 153 (2023).
- de Jong, M. J. et al. Red deer resequencing reveals the importance of sex chromosomes for reconstructing Late Quaternary events. *Mol. Biol. Evol.* **42**, 1–17 (2025).
- Nachman, M. W. & Payseur, B. A. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philos. Trans. R. Soc. B* **367**, 409–421 (2012).
- Li, G., Figueiro, H. V., Eizirik, E. & Murphy, W. J. Recombination-aware phylogenomics reveals the structured genomic landscape of hybridizing cat species. *Mol. Biol. Evol.* **36**, 2111–2126 (2019).
- vonHoldt, B. M. et al. Persistence and expansion of cryptic endangered red wolf genomic ancestry along the American Gulf coast. *Mol. Ecol.* **31**, 5440–5454 (2021).
- Feng, C., Wang, J., Liston, A. & Kang, M. Recombination variation shapes phylogeny and introgression in wild diploid strawberries. *Mol. Biol. Evol.* **40**, msad049 (2023).
- Jiang, Z. et al. Gene flow and an anomaly zone complicate phylogenomic inference in a rapidly radiated avian family (Prunellidae). *BMC Biol.* **22**, 49 (2024).
- Monthey, J. D. & Spellman, G. M. Recombination rate variation shapes genomic variability of phylogeographic structure in a widespread North American songbird (Aves: *Certhia americana*). *Mol. Phylogenet. Evol.* **196**, 108088 (2024).
- Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. fastsimcoal2: Demographic inference under complex evolutionary scenarios. *Bioinformatics* **37**, 4842–4849 (2021).
- Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
- Mazet, O., Rodriguez, W., Grusea, S., Boitard, S. & Chikhi, L. On the importance of being structured: instantaneous coalescence rates and human evolution — lessons from ancestral population size inference?. *Heredity* **116**, 362–371 (2015).
- Hanghoj, K., Molke, I., Andersen, P. A., Manica, A. & Korneliusen, T. S. Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *GigaScience* **8**, 1–9 (2019).
- Narasimhan, V. et al. BCFtools/ROH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* **32**, 1749–2751 (2016).
- Thompson, E. A. Variation in meiosis, across genomes, and in populations. *Genetics* **194**, 301–326 (2013).
- Gomez-Sanchez, D. et al. On the path to extinction: inbreeding and admixture in a declining grey wolf population. *Mol. Ecol.* **27**, 3599–3612 (2018).
- Bertorelle, G. et al. Genetic load: genomic estimates and applications in non-model animals. *Nat. Rev. Genet.* **23**, 492–503 (2022).
- Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**, 80–92 (2012).

46. Heffelfinger, J. R., Nowak, R. M. & Paetkau, D. Clarifying historical range to aid recovery of the Mexican wolf. *J. Wildl. Manag.* **81**, 766–777 (2017).
47. Mathur, S. & DeWoody, J. A. Genetic load has potential in large populations but is realized in small inbred populations. *Evol. Appl.* **14**, 1540–1557 (2021).
48. Dussex, N., Morales, H. E., Grossen, C., Dalen, L. & Oosterhout, C. Purging and accumulation of genetic load in conservation. *Trends Ecol. Evol.* **38**, P961–P969 (2023).
49. Smeds, L. & Ellegren, H. From high masked to high realized genetic load in inbred Scandinavian wolves. *Mol. Ecol.* **32**, 1567–1580 (2022).
50. Iurino, D. A. et al. A Middle Pleistocene wolf from central Italy provides insights on the first occurrence of *Canis lupus* in Europe. *Sci. Rep.* **12**, 2882 (2022).
51. Sotnikova, M. & Rook, L. Dispersal of the Canini (Mammalia, Canidae: Caninae) across Eurasia during the Late Miocene to Early Pleistocene. *Q. Int.* **212**, 86–97 (2010).
52. Gelabert, P. et al. Genome-scale sequencing and analysis of human, wolf, and bison DNA from 25,000-year-old sediment. *Curr. Biol.* **31**, 3564–3574 (2021).
53. Hibbins, M. S. & Hahn, M. W. Phylogenomic approaches to detecting and characterizing introgression. *Genetics* **220**, iyab173 (2021).
54. Wang, G. D. et al. Genomic approaches reveal an endemic subpopulation of gray wolves in Southern China. *iScience* **20**, 110–118 (2019).
55. Zhang M. et al. Ancient mitogenomes reveal the maternal genetic history of East Asian gray wolves (*Canis lupus*). *Integr. Zool.* 1–11 (2025).
56. Wang, L. et al. The geographical distribution of grey wolves (*Canis lupus*) in China: a systematic review. *Zool. Res.* **37**, 315–326 (2016).
57. Pinxian, W. & Xiangjun, S. Last glacial maximum in China: comparison between land and sea. *Catena* **23**, 341–535 (1994).
58. Dayan, T., Simberloff, D., Tchernov, E. & Yom-Tov, Y. Canine carnassial: character displacement in the wolves, jackals, and foxes of Israel. *Biol. J. Linn. Soc.* **45**, 313–331 (1992).
59. Kurten, B. The carnivora of the Palestine caves. *Acta Zool. Fenn.* **107**, 74 (1965).
60. Mashkour, M. et al. Carnivores and their prey in the Wezmeh Cave (Kermanshah, Iran): a Late Pleistocene refuge in the Zagros. *Int. J. Osteoarchaeol.* **19**, 678–694 (2008).
61. Plessis, S. J., Blaxter, M., Koepfli, K. P., Chadwick, E. A. & Hailer, F. Genomics reveals complex population history and unexpected diversity of Eurasian otters (*Lutra lutra*) in Britain relative to genetic methods. *Mol. Biol. Evol.* **40**, msad207 (2023).
62. Statham, M. J. et al. Range-wide multilocus phylogeography of the red fox reveals ancient continental divergence, minimal genomic exchange, and distinct demographic histories. *Mol. Ecol.* **23**, 4813–4830 (2014).
63. Alvares, F. et al. Old World *Canis* spp. with taxonomic ambiguity: workshop conclusions and recommendations. *Canid News* **21** <http://hdl.handle.net/10138/327703> (2019).
64. Krofel, M., Hatlauf, J., Bogdanowicz, W., Cambell, L. A. D. & Godinho, R. Towards resolving taxonomic lineages in wolf, dog, and jackal of Africa, Eurasia, and Australasia. *J. Zool.* **316**, 155–168 (2022).
65. Sillero-Zubiri, C. Family Canidae. In D.E. Wilson and R.A. Mittermeier (Eds.), *The Handbook of the Mammals of the World* (Lynx Edicions in association with Conservation International and IUCN, 2009).
66. Sharma, D. K., Maldonado, J. E., Jhala, Y. V. & Fleischer, R. C. Ancient wolf lineages in India. *Proc. R. Soc. B* **271**, S1–S4 (2004).
67. Werhahn, G. et al. Himalayan wolf distribution and admixture based on multiple genetic markers. *J. Biogeogr.* **47**, 1272–1285 (2020).
68. Jhala, Y. V., Saini, S., Kumar, S. & Qureshi, Q. Distribution, status, and conservation of the Indian Peninsula wolf. *Front. Ecol. Evol.* **10**, 814966 (2022).
69. Sheikh, K. M. & Molur, S. *Status and Red List of Pakistan's Mammals. Based on the Conservation Assessment and Management Plan* 312 (IUCN Pakistan, 2003).
70. Hennelly, L. M. et al. *Canis lupus* ssp. *pallipes* (The IUCN Red List of Threatened Species, 2025).
71. Werhahn, G. et al. *Himalayan Wolf: Canis lupus* spp. *chanco* (The IUCN Red List of Threatened Species, 2023).
72. Kardos, M. et al. The crucial role of genome-wide genetic variation in conservation. *Proc. Natl. Acad. Sci. USA* **118**, e2104642118 (2021).
73. Sowerby, A. C. *China's Natural History: A Guide to the Shanghai Museum* (Royal Asiatic Society of Great Britain and Ireland, North China Branch, 1936).
74. Fellowes, J. R., Chan, B. P. L., Lau, C. M. N., Sai-Chit, N. & Siu, G. L. P. 2003. *Report of Rapid Biodiversity Assessment at Shiwandashan National Nature Reserve and National Forest Park, Southwest Guanxi, China, 2000–2001* (Kadoorie Farm and Botanic Garden in collaboration with Guangxi Forestry Department, Guangxi Institute of Botany, South China Normal University, 2003).
75. Bensen, G. T. et al. Navigating complex geopolitical landscapes: challenges in conserving the endangered Arabian wolf. *Biol. Conserv.* **296**, 110655 (2024).
76. Caroe, C. et al. Single-tube library preparation for degraded DNA. *Methods Ecol. Evol.* **9**, 410–419 (2017).
77. Mak, S. S. T. et al. Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platform for paleogenomic sequencing. *GigaScience* **6**, 1013 (2017).
78. Gilbert, M. T. P. et al. Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science* **317**, 1927–1930 (2007).
79. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. USA* **110**, 15758–15763 (2013).
80. Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
81. Schubert, M. et al. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat. Protoc.* **9**, 1056–1082 (2014).
82. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
83. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
84. Meisner, J. & Albrechtsen, A. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* **210**, 719–731 (2018).
85. Skotte, L., Korneliussen, T. S. & Albrechtsen, A. Estimating individual admixture proportions from Next Generation sequencing data. *Genetics* **3**, 693–702 (2013).
86. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinform.* **15**, 356 (2014).
87. Danecek, P. et al. The variant call format and VCFtool. *Bioinformatics* **27**, 2156–2158 (2011).
88. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
89. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Haeseler, A. & Jermini, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
90. Nguyen, L. T., Schmidt, H. A., Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2014).
91. Auton, A. et al. Genetic recombination is targeted towards gene promoter regions in dogs. *PLoS Genet.* **9**, e1003984 (2013).
92. Smeds, L. et al. Whole-genome analysis provide no evidence for dog introgression in Fennoscandian wolf populations. *Evol. Appl.* **14**, 721–734 (2021).

93. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
  94. Martin, S. H. & Belleghem, S. M. V. Exploring evolutionary relationships across the genome using topology weighting. *Genetics* **1**, 429–438 (2017).
  95. Delaneau, O., Howie, B., Cox, A. J., Zagury, J. F. & Marchini, J. Haplotype estimation using sequencing reads. *AJGH* **93**, P687–P696 (2013).
  96. Danecek, P. et al. Twelve years of SAMtools and BCFTools. *GigaScience* **10**, giab008 (2021).
  97. Hilgers, L. et al. Avoidable false PSMC population size peaks occur across numerous studies. *Curr. Biol.* **35**, 927–930 (2025).
  98. Koch, E. M. et al. De novo mutation rate estimation in wolves of known pedigree. *Mol. Biol. Evol.* **36**, 2536–2547 (2019).
  99. Mech, D. L. & Barber-Meyer, S. Use of erroneous wolf generation time in assessments of domestic dog and human evolution. *Sci. Lett.* <http://science.sciencemag.org/content/352/6290/1228/tab-e-letters> (2017).
- sequencing efforts; L.M.H. led data analysis with assistance from B.R.P., A.N., X.S., N.F.G.M., and L.M.H. wrote the manuscript with input from all co-authors.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-025-09379-9>.

**Correspondence** and requests for materials should be addressed to Lauren M. Hennelly.

**Peer review information** *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Yi-Jyun Luo and Michele Repetto.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

### Acknowledgements

L.M.H. thanks the National Science Foundation Postdoctoral Research Fellowship (award number 2208950) for funding and support. The Norwegian Environment Agency (project 18088069) provided funding and support for sequencing efforts on newly sequenced wolf genomes. Ç.H.Ş. thanks to Fondation Segré and the Sigrid Rausing Trust for providing the majority of the funding for this project, H. Batubay Özkan and Barbara Watkins for their support of the Biodiversity and Conservation Ecology Lab at the University of Utah, and Bilge Bahar, Seha İşmen, Ömer Koç, Ömer Külahçioğlu, Burak Över, Emin Özgür, Suna Reyent, and Ceren Sağlam for supporting this project. Türkiye's Department of Nature Conservation and National Parks and the Ministry of Agriculture and Forestry granted the permit for Türkiye (No. 72784983-488.04-114100). P.H. thanks the Technology Agency of the Czech Republic (project SS07010447) for support. We thank the Museum of the Institute of Plant and Animal Ecology UB RAS for access to their collections. We thank Geraldine Werhahn for her input on the manuscript.

### Author contributions

L.M.H. conceived and designed the project with guidance from S.G., M.H.S.S., and B.N.S. C.S.O., M.H.S.S., and N.F.G.M. conducted laboratory work; H.F., G.S., B.H., F.H., S.P.E.H., C.K., Ç.H.Ş., P.K., H.S., M.H.S.S., L.P., P.H., A.Y., and M.T.P.G. provided logistics, field work, wolf samples, and

<sup>1</sup>Section for Evolutionary Hologenomics, Globe Institute, University of Copenhagen, Copenhagen, Denmark. <sup>2</sup>Center for Conservation Genomics, Smithsonian's National Zoo and Conservation Biology Institute, Washington, DC, USA. <sup>3</sup>Mammalian Ecology and Conservation Unit, Veterinary Genetics Laboratory, Department of Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, CA, USA. <sup>4</sup>Department of BioSciences, Rice University, Houston, TX, USA. <sup>5</sup>Organisms and Environment, School of Biosciences, Cardiff University, Cardiff, UK. <sup>6</sup>Department of Molecular Biology and Genetics, Koç University, Istanbul, Türkiye. <sup>7</sup>School of Biological Sciences, University of Utah, Salt Lake City, UT, USA. <sup>8</sup>Kuzey Doğu Society, Kars, Türkiye. <sup>9</sup>Ural Federal University, Yekaterinburg, Russia. <sup>10</sup>Institute of Plant and Animal Ecology, Ural Branch of Russian Academy of Science, Yekaterinburg, Russia. <sup>11</sup>Technical University in Zvolen, Zvolen, Slovakia. <sup>12</sup>Department of Zoology, Charles University, Prague, Czechia. <sup>13</sup>Department of Biology and Ecology, University of Ostrava, Ostrava, Czechia. <sup>14</sup>NTNU University Museum, Norwegian University of Science and Technology, Trondheim, Norway. <sup>15</sup>Wildlife Institute of India, Dehradun, India. <sup>16</sup>University of Education – Attock Campus, Attock, Pakistan. <sup>17</sup>University of the Punjab, Lahore, Pakistan. <sup>18</sup>Lebanese Wildlife, Jitta, El Metn, Lebanon. <sup>19</sup>Cardiff University-Institute of Zoology Joint Laboratory for Biocomplexity Research (CIBR), Beijing, China. <sup>20</sup>Department of Biology, University of Copenhagen, Copenhagen, Denmark.

✉ e-mail: [lauren.hennelly@rice.edu](mailto:lauren.hennelly@rice.edu)