UC Davis
UC Davis Electronic Theses and Dissertations

Title

Divergence, Selection, and Demographic History of Wolves in Eurasia using Genomic Data

Permalink

https://escholarship.org/uc/item/9022h7qv

Author

Hennelly, Lauren Mae

Publication Date

2022

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Divergence, selection, and demographic history of wolves in Eurasia using genomic data

By

LAUREN MAE HENNELLY
DISSERTATION

Submitted in partial satisfaction of the requirements of the degree of

DOCTOR OF PHILOSPHY

In

Ecology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Benjamin N. Sacks (Chair)

_____

Rachael Bay

_____

Graham Coop

Committee in Charge

2022

Table of Contents

Abstract

Species are often composed of evolutionarily distinct populations and lineages. Deciphering the evolutionary history of recently diverged species with multiple lineages remains challenging because genealogical discordance is common across the genome. Emerging genomic and statistical tools are allowing unprecedented insight into the evolutionary history of complicated species that have been notoriously difficult to resolve. The gray wolf exemplifies this challenge as it has multiple recognized lineages and has only recently diverged from other canids. In this dissertation, I use genome-wide data to study the evolutionary history of gray wolves in Eurasia, with a focus on investigating the phylogenomics, demographic history, and role of adaptive introgression.

For my first chapter, I investigate the genomic distinctiveness of wolves corresponding to two deeply divergent mitochondrial clades restricted to the Indian subcontinent and the Tibetan plateau, respectively. Using the first whole genome sequences of four Indian wolves along with those of two newly sequenced Tibetan wolves and 31 additional canids, I demonstrate Indian and Tibetan wolves are the two most deeply divergent wolf lineages and highly distinct from broadly distributed wolf populations corresponding to the mitochondrial Holarctic clade. Low-recombination regions of the genome provided evidence that the Indian wolf is the most basal wolf lineage, in contrast to the mitochondrial DNA, showing the Tibetan wolf as the most basal lineage. Our findings imply that southern regions of Asia have been important centers for gray wolf evolution and that the Indian wolf represents one of the world's most endangered and evolutionarily distinct wolf lineages.

My second chapter focuses on distinguishing secondary contact zones and investigating adaptive introgression among wolf lineages in Asia. I used 5 newly and 7 recently sequenced

wolf (*Canis lupus*) genomes from the lowland plains and high-altitude mountains of Pakistan, India, and Kyrgyzstan, along with 79 additional canid genomes, to explore the possibility that adaptive introgression from specialized basal wolf lineages into Holarctic wolves facilitated their range expansion. I detected three narrow secondary contact zones among the widespread Holarctic lineage and the divergent Indian and Tibetan lineages. Within West-Asian Holarctic wolves, I detected several gene regions that were highly differentiated compared to other Holarctic wolves and signals of higher-than expected levels of introgression from Indian or African wolves. Additionally, in the high-altitude Central Asian wolf, I found similar evidence for adaptive introgression from the Tibetan wolf including gene related to hypoxia adaptation in other mammals. Lastly, demographic analyses revealed Indian and Tibetan wolf lineages were likely isolated within separate glacial refugia in Asia while the Holarctic lineage underwent massive expansion events during the late Pleistocene.

In Chapter 3, I use genome-wide data from 5 newly sequenced gray wolves from Pakistan to more robustly investigate the locations of secondary contact zones and admixture among the three wolf lineages at these contact zones. Using a total of 8 wolves from Pakistan and an additional 48 canids, clustering and admixture analyses indicated high proportions of Indian ancestry was present in the lowlands of the Indus plains, Potwar plateau, and Western mountain ranges of Pakistan. Except for small amounts of Tibetan ancestry detected in two wolves from the Karakoram Mountains of northern Pakistan, the Tibetan lineage appeared to end to the east in the Ladakh region of India. These findings clarify the boundaries of the three divergent wolf lineages and highlight the conservation significance of Pakistan's wolf populations, especially the wolves in Sindh and Punjab that represent the Indian lineage. Overall,

this dissertation provides insight into the evolutionary and historical processes that shape genomic divergence and local adaptation in a wild and highly mobile species.

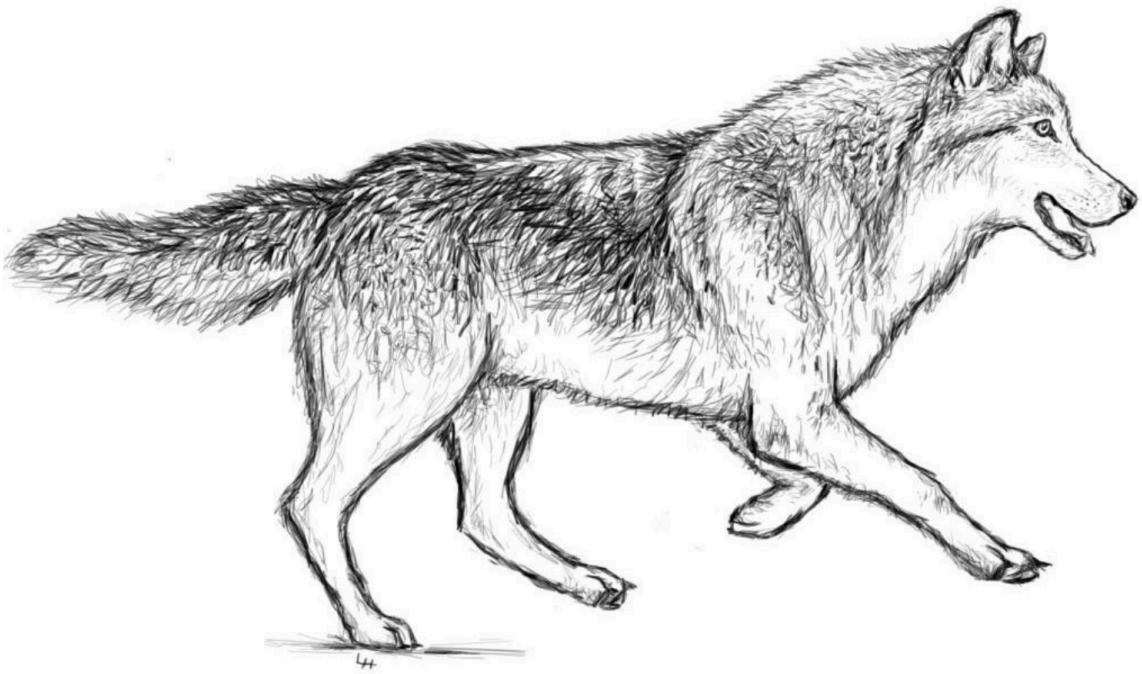
Acknowledgements

These past years working towards my PhD have been formative to my personal and professional life. This dissertation research has roots that were sowed long back before I started my PhD in 2016. I'm thankful for mentorship and guidance early on in my career that lead to my initial studies on gray wolves in India, which formed the basis and inspiration for my PhD work. I thank my friends and colleagues at the Wildlife Institute of India, particularly Bilal Habib, Shivam Shrotriya, Malem Ningombi, Indranil Mondal, as well as colleagues at The Grasslands Trust, the local families and forest department staff that helped during the initial field work, and the Fulbright Program. During my PhD, I started working with colleagues in Pakistan to study the endangered gray wolves that inhabit the country. I thank Hira Fatima and her family, Ghulam Sarwar, and many other colleagues who provided help, support, and whom I learned so much from about wildlife in Pakistan. I am deeply grateful for the opportunity and privilege to conduct fieldwork internationally and collaborate with so many wonderful colleagues in India and Pakistan.

Next, I thank my friends and colleagues at University of California, Davis. I thank my advisor, Benjamin N. Sacks, for providing guidance and support for my research, and very quick feedback on writing or always being available for chatting about research. I've learned so much from Ben over the years, and grateful for the time he spent in shaping me as a scientist. The Mammalian Ecology and Conservation Unit has been a great place to work and I've made many lifelong friends. I thank Stevi Vanderzwan, Zach Lounsberry, Cate Quinn, Julia Owen, Sophie Preckler-Quisquater, Tom Batter, Taylor Davis, Monica Serrano, Grace Rosburg-Francot, Carly White, Andrea Broad, and Cody Aylward for their company in the office and all being part of a

shared adventure of becoming scientists. I also thank my friends in the Graduate Group in Ecology for having such a supportive and fun environment for completing graduate school. Shotaro Nakamura provided so much support as well as fun adventures both in Davis and far away places. Thank you Sho for making graduate school full of fun and great memories.

Lastly, I thank my family for encouraging my love of science and gray wolves. My mom and dad brought me to wolf sanctuaries in Wisconsin, bought me books about wolves, and drove 6 hours to and from Battle Ground, Indiana so I could observe wolves for my 18th birthday. I'm forever grateful for their support and freedom to pursue a career I am passionate about. I thank my twin sister, Jessica, and two brothers, Kevin and Scott, for their support and for many days we spent exploring prairies, lakes, and forests in Illinois, which fostered my interest in conservation.



Chapter 1: Ancient divergence of Indian and Tibetan wolves revealed by recombination-aware phylogenomics

This chapter was submitted to and published in the *Molecular Ecology*.

Citation: Hennelly LM, Habib B, Modi S, Rueness EK, Gaubert P, Sacks BN. 2021. Ancient divergence of Indian and Tibetan wolves revealed by recombination-aware phylogenomics. Molecular Ecology 30:6687-6700.

Abstract

The gray wolf (*Canis lupus*) expanded its range across Holarctic regions during the late Pleistocene. Consequently, most gray wolves share recent (<100 kya) maternal origins corresponding to a widespread Holarctic clade. However, two deeply divergent (200-700 kya) mitochondrial clades are restricted, respectively, to the Indian subcontinent and the Tibetan Plateau, where remaining wolves are endangered. No genome-wide analysis had previously included wolves corresponding to the mitochondrial Indian clade or attempted to parse gene flow and phylogeny. We sequenced 4 Indian and 2 Tibetan wolves and included 31 additional canid genomes to resolve the phylogenomic history of gray wolves. Genomic analyses revealed Indian and Tibetan wolves to be distinct from each other and from broadly distributed wolf populations corresponding to the mitochondrial Holarctic clade. Despite gene flow, which was reflected disproportionately in high-recombination regions of the genome, analyses revealed Indian and Tibetan wolves to be basal to Holarctic gray wolves, in agreement with the mitochondrial phylogeny. In contrast to mitochondrial DNA, however, genomic findings suggest the possibility that the Indian wolf could be basal to the Tibetan wolf, a discordance potentially reflecting selection on the mitochondrial genome. Together, these findings imply that southern regions of Asia have been important centers for gray wolf evolution and that Indian and Tibetan

wolves represent evolutionary significant units (ESUs). Further study is needed to assess whether these ESUs warrant recognition as distinct species. This question is especially urgent regarding the Indian wolf, which represents one of the world's most endangered wolf populations.

1. Introduction

In North America and Eurasia, the most ancestral populations of species often occur in the southern portions of their ranges (Hewitt 2000, Petit et al. 2003). These southern regions tend to be the most climatically stable and therefore served as refugia during the heights of ice-age glaciations. In contrast, northern regions were associated with dramatic climate changes, fueling an alternation of range expansions and range contractions or extinctions in many species (Hewitt 2000, Hofreiter and Stewart 2009). Populations that today are widespread across northern latitudes often trace their origins to relatively recent, late Pleistocene expansion events from small founding populations and therefore reflect only a fraction of their ancestral diversity (Hundertmark et al. 2002, Statham et al. 2014, Palkopouou et al. 2016). Because southern Eurasia and North America also tend to be the most heavily populated by humans today, these evolutionarily significant populations also tend to be the most threatened.

The gray wolf (*Canis lupus*) exemplifies this problem. A mitochondrial perspective suggests that most gray wolves in the Northern Hemisphere today, hereafter the "Holarctic lineage," originate from one or more massive late-Pleistocene (<100 kya) population expansions and therefore carry only a fraction of their ancestral diversity (Ersmark et al. 2016; Koblmuller et al. 2016, Loog et al. 2020). Several southern populations of gray wolves in both North America and Eurasia survived the ice ages but are currently endangered (Boitani et al. 2018). In North America, Beringian wolf populations may have been replaced by a late-Pleistocene expansion originating from Asia (Leonard et al. 2007, but see Ersmark et al. 2016); wolf lineages basal to most

contemporary North American wolves survived south of the North American ice sheets, but were largely lost by the 20th century due to persecution (Leonard et al. 2005). The two most ancestral matrilines in extant gray wolves are restricted to southern regions of Asia: the Indian subcontinent and Tibetan plateau (Fig. S1) (Aggarwal et al. 2003, Sharma et al. 2004, Aggarwal et al. 2007). Indian and Tibetan wolf matrilines are estimated to have diverged from the Holarctic gray wolf clade up to 350,000 and 715,000 years ago, respectively (Sharma et al. 2004, Werhahn et al. 2018, Wang et al. 2020). Tibetan wolves face multiple threats in various regions of their distribution and Indian wolves are thought to number <3,000 individuals within India (Jhala 2003, Suryawanshi et al. 2013, Hennelly et al. 2015).

Because mitochondrial phylogenies only provide one genealogical view of wolf population history, comprehensive phylogeographic inferences must include analysis of nuclear DNA. The nuclear genome contains most of the genes that reflect a taxon's history and determine its evolutionary significance. Thus far, the nuclear genomic relationships among gray wolves remain unclear. First, no wolf sampled specifically from the lowland peninsular Indian subcontinent has been sequenced. The only putative Indian wolf sequenced was sampled from a zoo in Germany and lacked precise locality information (e.g. Fan et al. 2016); its mitochondrial haplotype clustered with Holarctic wolves rather than those of confirmed Indian wolves (Fig. S2), suggesting it was from Kashmir or further west rather than lowland peninsular India (Sharma et al. 2004). Second, the phylogenetic positioning of the Tibetan wolf has varied across studies. Whereas some studies have found the Tibetan wolf to be basal to Holarctic wolves, in agreement with mitochondrial patterns (Wang et al. 2020), others have found North American wolves to be basal to Tibetan and Eurasian Holarctic wolves (Fan et al. 2016, Wang et al. 2019). These discrepant results suggest that gene flow could be obscuring phylogenetic history. The

post-Pleistocene isolation of North American gray wolves versus ongoing gene flow among Eurasian wolf lineages, in particular, could account for the basal positioning of North American wolves in two of the previous studies (Fan et al. 2016, Wang et al. 2019). Therefore, a complete understanding of the nuclear relationships among gray wolves requires inclusion of a representative Indian wolf and explicit accounting for gene flow that could obscure the underlying phylogeny.

Because regions of the genome differ with respect to their susceptibility to introgression, phylogeny and gene flow leave distinct signatures on genomes that can be leveraged to investigate historical relationships among wolves (Li et al. 2019, Martin et al. 2019). One of the more systematic relationships relates to recombination (Nachman and Payseur 2012, Butlin et al. 2015, Martin and Jiggin 2017). Regions of the genome with high recombination can more rapidly decouple selectively neutral loci from deleterious loci, and therefore tend to harbor proportionally more introgressed ancestry (Nachman and Payseur 2012). Conversely, regions of the genome with low recombination tend to preserve the historical branching order of taxa (Pease and Hahn 2013). Low-recombination regions also have lower effective population sizes (N_e) and, therefore, more thoroughly sorted lineages than higher-recombination regions (Pease and Hahn 2013). Higher fidelity of low-recombination regions to the historical branching order has been documented in a range of species (Fontaine et al. 2015, Schumer et al. 2018, Martin et al. 2019, Li et al. 2019, Manuel et al. 2020).

Chromosomes also vary with respect to the relative frequencies of genomic regions that retain signals of introgression and phylogeny. In particular, sex chromosomes tend more than autosomes to reflect historical branching patterns. In mammals, the X chromosome tends to resist introgression because loci disproportionately contribute to reproductive isolation and

recombination is generally lower, causing stronger selection against introgressed loci that are less compatible with the local genetic background (Presgraves 2008, Bergero and Charlesworth 2009, Muirhead and Presgraves 2016). Lineage sorting also proceeds more rapidly for X-chromosome loci because males only carry one copy, which reduces its N_e (Schaffner 2004).

We sequenced 4 Indian and 2 Tibetan wolves, and an Indian golden jackal (*C. aureus*) and used these data along with 30 previously published whole genome shotgun sequences to investigate their genomic distinctiveness. We first reconstructed phylogenetic trees from mitogenomes, X chromosomes, and autosomes to evaluate broad-scale phylogenomic patterns. We then assessed regional patterns of gene flow across different Asian wolf populations with a focus on Tibetan and Indian wolves, including use of admixture graphs to parse phylogeny and admixture. Lastly, using the domestic dog recombination map (Auton et al. 2013), we investigated phylogenetic relationships explicitly in genomic regions of high, intermediate, and low recombination rates across the autosomes and X chromosome to elucidate positive and negative relationships between recombination rates and frequencies of particular topologies.

2. Methods

2.1 Laboratory and bioinformatic procedures

DNA was extracted from 4 Indian wolves, 2 Tibetan wolves, and 1 Asian golden jackal comprising 5 blood and 2 tissue samples using the DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA, USA) following the manufacturer's protocol. The Indian wolf samples were from Rajasthan (n = 1) and Maharashtra (n = 3), the 2 Tibetan wolf samples were from Ladakh of Jammu and Kashmir (n = 2), India, and the Asian golden jackal sample was from Uttarakhand, India. Libraries were prepared using the Illumina TruSeq DNA PCR-Free library preparation kit and sequenced using Illumina HiSeq 2500 paired end at 150-bp (Illumina, San Diego, CA,

USA). A dataset of canid species was assembled by including published samples from NCBI Sequence Read Archive (SRA), in addition to our newly sequenced samples (table S1.1). All raw reads were trimmed using Trim Galore version 0.6.5 using the following flag: --illumina (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)

To extract the mitogenome from the raw reads of our samples and 99 other data sets from NCBI SRA, we used a de novo assembler, NOVOPlasty version 3.8.3 (Dierckxsens et al. 2016), with a reference mitogenome of a Mongolian wolf (NCBI accession number KC896375). NOVOPlasty failed to assemble the mitogenome for a subset of our samples; to obtain the mitogenome of these samples, we aligned the raw reads using BWA MEM to the Mongolian wolf reference mitogenome. Specifically, we indexed the reference wolf mitogenome, used BWA MEM to align the raw reads to the indexed reference wolf mitogenome, removed PCR duplicates using MarkDuplicates version 2.18.25 in Picard tools (http://broadinstitute.github.io/picard/), and sorted the bam files using SAMtools version 1.9. A consensus fasta file was created using SAMtools mpileup and the vcf2fq command within vcfutils.pl of SAMtools.

For nuclear genomes, the reads were mapped to the domestic dog genome assembly (CanFam3.1) using the BWA MEM version 0.7.17.r1188 (Li and Durbin 2009). After alignment, PCR duplicates were identified and removed using the MarkDuplicates tool version 2.18.25 from the Picard suite. Our BAM files were then sorted and filtered to keep only properly paired reads (-F 1024) using SAMtools version 1.9 (Li et al. 2009). We called SNPs using freebayes version 9.9.2 (Garrison and Marth 2012) and we subsequently filtered by genotype quality (minGQ = 30) and removed indels using VCFtools 0.1.14 (Danecek et al. 2011). Sites with a mean depth of ≥850 for all individuals were removed to exclude paralogs from our dataset. Lastly, we estimated

the average depth of our BAM files using the command depth (--depth) in SAMtools version 1.9 (Li et al. 2009).

2.2 Mitochondrial phylogeny

After assembling mitogenomes from our samples and gray wolves, African wolves, golden jackals, Ethiopian wolf, coyotes, and dholes acquired through NCBI (Table S1.2), we aligned them using MUSCLE v3.8.31 (Edgar 2004), visually inspected the alignment using ClustalX 2.1, and removed the D-loop region, which resulted in a 15,437-bp length of the mitogenome. We partitioned the dataset into non-coding (tRNA and rRNA) and the 1st, 2nd, and 3rd codon position using the gene positions defined from the domestic dog mitogenome (NCBI accession NC_002008.4). The gene ND6 was included in the non-coding partition because it is transcribed in the reverse direction. We used BEAST v1.10.4 to construct two trees (Drummond and Rambaut 2007): one using only the 3rd codon positions to estimate divergence times while minimizing bias due to purifying selection on nonsynonymous sites (Subramanian and Lambert 2011), and another using all partitions to obtain node support. For both trees, we used the "Speciation: Birth-Death Process" tree prior and relaxed lognormal clock. We used the same partition scheme and substitution models previously determined for gray wolves: HKY+I for 1st codon position and non-coding, TN93+I for 2nd codon position, and TN93+G for the 3rd codon position (Loog et al. 2020). For the 3rd-codon position tree, we used a normal prior for the TMRCA of the African wild dog at 3.9 Ma (SD = 0.3 Ma) based on previous fossil and genetic analyses (Chavez et al. 2019). Each analysis was run for 50 million MCMC cycles, sampling every 5,000, and discarding the first 5 million states as burnin. We used Tracer v1.7.1 (Rambaut et al. 2018) to assess the outputs and all parameter estimates of the three analyses were above

200 effective samples (ESS). Finally, we used TreeAnnotator (v. 2.5.2.) to infer the maximum credibility trees and FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) to display the trees.

2.3 Nuclear analyses

To obtain a SNP data set consisting only of gray wolves (Holarctic, Indian, Tibetan), we filtered the variant calling format (VCF) file in Plink (v 1.90), retaining SNPs with a minimum allele count of 3, for which ≥90% of individuals had calls (--geno 0.1), and pruned for linkage disequilibrium >0.5 (--indep-pairwise 50 5 0.5), which resulted in ~3.5 million intraspecific SNPs (Purcell et al. 2007). We conducted a principal component analysis (PCA) in Plink using autosomal SNPs from gray wolves and domestic dogs. We estimated individual ancestry proportions with Admixture (v 1.3.0) using default setting (Alexander and Lange 2011). In datasets with discrete populations conforming to an island model, the cross-validation error can be used to determine the optimum (or correct) number of partitions. However, natural populations such as the system under study often manifest complex structure (e.g., hierarchical) that has no such optimum. In such cases, it is useful to examine multiple levels of K, each of which can provide unique insights. Thus, we conducted analyses including calculation of cross-validation errors for K ranging 2 to 12 to verify that configurations associated with higher K were nested within those of lower K, but highlighted K = 6, which was nested within lower levels of K and provided high resolution of populations and admixture among them.

To infer the phylogenetic relationships among gray wolves relative to other canid species, we extracted the autosomes and the X chromosome from our dataset and filtered to include only sites with calls in ≥90% individuals (--geno 0.1), resulting in approximately 30 million autosomal SNPs and ~1.2 million X chromosome SNPs. The VCF file was converted into Multisample Variant Call format (MVF) and subsequently fasta format with ambiguity codes

using MVFtools (http://www.github.com/jbpease/mvftools). We estimated the best model for the autosomal and X chromosome dataset using ModelFinder, as implemented in IQ-tree version 1.6.12 (Kalyaanamoorthy et al. 2017). According to Akaike Information Criterion, the top models estimated for autosomal and X chromosome trees were TVM+F+R2 and TVM+F+R4, respectively. Following model selection, we estimated the maximum likelihood tree for the autosomes and X chromosome using 1000 ultra-fast bootstraps (UFBoot) implemented in IQ-tree (Nguyen et al. 2014). The phylogenies were visualized using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

To disentangle incomplete lineage sorting and gene flow, we performed D-statistic analysis for the autosomes using ADMIXTOOLS (Patterson et al. 2012) and calculated f3 statistic using threepop in Treemix to test for introgression across the autosomes (Pickrell and Pritchard 2012). We converted the plink files into eigenstrat format using the convert script in the Eigensoft package (https://github.com/argriffing/eigensoft). We calculated the D-statistic and Z-scores for populations in the format: (North American (P1), Y, X; dhole (O)). Due to the presence of interspecific gene flow between gray wolves and other *Canis* species, we selected the dhole as the outgroup (Gopalakrishnan et al. 2018). Additionally, because there has been no gene flow between North American and Asian wolf populations in the past 12,000 years, our placement of North American wolves as P1 allowed us to evaluate the presence of recent gene flow among Asian wolf populations (Jakobsson et al. 2017). A negative D-statistic indicates gene flow occurring between X and Y, whereas a positive D-statistic indicates gene flow occurring between P1 and X. The statistical significance of the Z-score for each four-population calculation was assessed with weighted block jackknife tests using the default of 5-Mb block size for autosomes and 10-Mb block size for the X chromosome. If the Z-score is above 3 or below -3, it allows us

to reject the null hypothesis of no gene flow occurring between X and populations P1 and Y, respectively. We used the D-statistic analysis on the autosomes and X chromosome.

To explicitly model phylogeny and gene flow as distinct processes affecting genomic relationships, we modeled the admixture population history of gray wolves using qpGraph within the ADMIXTOOLS package (Patterson et al. 2012). Based on results of PCA, admixture analyses, f3 statistics, and D-statistics, we first assessed the best phylogenetic model describing the five least admixed gray wolf populations: North American, Central Asian, East Asian, Tibetan (excluding Ladakh and Qinghai), and Indian (as defined in Table S1.1). Specifically, we generated five admixture population models that included these primary gray wolf populations and other canids, but excluded wolves from Ladakh, Qinghai, and West Asia. We tested the five models both with and without incorporating admixture from coyotes into North American gray wolves (Sinding et al. 2018). We selected the best of these models as a phylogenetic scaffold on which to superimpose the admixed populations (Ladakh, Qinghai, and West Asian), and to characterize and quantify admixture. The admixture graph analyses were conducted using the autosomal SNPs with the default settings of qpGraph.

2.4 **Topology weighting and recombination**

To quantify the frequencies of different topologies across the genome, we used topology weighting by iterative sampling of subtrees, Twisst (Martin and Belleghem 2017). We excluded the two Tibetan wolves from Ladakh due to their admixed origins. Genotypes of the remaining gray wolf samples and the dhole were filtered using vcftools to include only biallelic sites and only sites with 100% of individuals present (--max missing 1). Our dataset was phased to infer haplotypes from SNP genotypes using Beagle (beagle.16May18.771.gar file) with default parameters. We constructed local neighbor-joining trees from SNPs extracted in 100-SNP

windows across the genome using PhyML 3.0 (Guindon et al. 2010). Exact weightings were computed for all inferred topologies using two sets of populations: 3 target populations (Tibetan wolf, Indian wolf, Holarctic wolf) and 4 target populations (Tibetan wolf, Indian wolf, Holarctic North American wolf, and Holarctic Asian wolf), both rooted to the dhole. Twisst estimates the relative frequency of occurrence (i.e. the weights) of each topology within each 100-SNP window, with a weighting of 1 indicating all 100 SNPs in the window reflect a single topology. Next, we downloaded and used a recombination map of the domestic dog to analyze the associations between topology and recombination rate (in cM/Mb) (Auton et al. 2013). To estimate the recombination rate within each 100-SNP window, we averaged the recombination rates found within the start and end of each window partitioned by Twisst across the genome. To infer the X chromosome and autosomal phylogenetic trees using regions of low recombination, we extracted the start and end positions of 100-SNP windows with an average combination rate ≤0.2 cM/Mb, and used these positions to create a bed file. We then extracted regions of the VCF file that overlapped the bedfile using bedtools (--intersect). We constructed phylogenetic trees using IQ-tree as described above.

3. Results

3.1 Broad phylogenomic patterns

We resequenced 4 Indian wolves, 2 Tibetan wolves, and an Asian golden jackal from India (Table S1.1). We first assembled mitogenomes of these samples, along with a gray wolf from Kyrgyzstan, 3 African wolves, and an Ethiopian wolf for which we only sequenced the mitogenome, and 93 other mitogenomes that we assembled from short reads in GenBank SRA or that were pre-assembled in the GenBank nucleotide database (Fig. 1a; Table S1.2). Similar to previous mitochondrial analyses (e.g., Sharma et al. 2004), Indian and Tibetan lineages were

basal to Holarctic gray wolves. Based on a 3rd-codon position tree calibrated to the African wild dog (*Lycaon pictus*), Indian and Tibetan wolves shared most recent common ancestors (MRCA) with Holarctic wolves 200 kya (95% HPD: 175-307 kya) and 496 kya (388-644 kya), respectively (Fig. S1.3). The Indian and Tibetan lineages were confined, respectively, to the lowland Indian subcontinent and the Tibetan plateau.

For nuclear genomic data, we aligned reads from our 6 wolf and golden jackal samples, along with 30 additional canids (Table S1.1), to the domestic dog genome (canFam3.1). After trimming adapters and removing paralogs and PCR duplicates, the average depth of each of the newly sequenced canids averaged 6.8×, resulting in 10.4× and 27.3× for our total sample of wolves from Ladakh and India, respectively (Table S1.1). After filtering out indels and SNPs with quality scores <30, we obtained 33 million autosomal and 1.2 million X-chromosome SNPs.

Similar to mitochondrial findings, principal component analysis (PCA) of gray wolves and admixture analysis revealed that Indian and Tibetan wolves were genomically distinct from each other and from Holarctic wolf populations (Fig.1.1b,c; Fig. S1.4). Also in line with mitochondrial patterns, Holarctic wolves spanning North America to the Middle East clustered closely together. Although West Asian and Indian wolves are currently classified as a single subspecies, *C. lupus pallipes*, West Asian wolves, including from Syria and Iran, clustered in the PCA and admixture analysis with Holarctic wolves rather than with wolves from India.

Despite distinctiveness of the Indian and Tibetan wolf clusters, admixture analyses showed evidence of gene flow. Although the minimum cross-validation error was associated with K = 2, successively higher partitions, up to K = 6, were nested in lower partitions, providing increasing resolution (Fig. S1.4). At K = 6, Tibetan wolves from the western edge of the Tibetan plateau – the Ladakh region – reflected admixture from Indian, West Asian, and Central Asian wolves (Fig. 1.1b; Fig. S1.4; Tables S1.3 & S1.4). Additionally, in agreement with previous findings from microsatellites (Werhahn et al. 2020), a Tibetan wolf from the northeastern edge of the Tibetan plateau – Qinghai in China – showed admixture from East and Central Asian gray wolves. Wolves from the center of the Tibetan plateau – the Tibetan Autonomous Region (TAR) – reflect the least amount of admixture from any adjacent wolf populations. Admixture analyses, assuming K = 3 and 4 clusters, also suggested some genetic connectivity between Indian and West Asian wolves (Fig. S1.4).

Regarding magnitudes, D-statistics and f3 statistics indicated that Indian wolves showed relatively greater, yet generally weak, gene flow with wolves of West and Central Asia than with the more geographically proximate Tibetan wolves (Fig. 1.1c; Tables S1.3 & S1.4). Otherwise, admixture among the four Holarctic populations (West Asian, Central Asian, East Asian, North American) was consistent with a continuous isolation-by-distance relationship (Fig. S1.5).

Consistent with the mitogenomes, maximum likelihood trees constructed from all autosomes and from the X chromosome showed Tibetan wolves to be the most basal of the gray wolves (Fig. 1.1b; Fig. S1.6). Otherwise, however, these trees were discordant from the mitogenome tree, in particular, clustering the Indian wolf with Holarctic wolves. The autosomal tree, in particular, revealed West Asian wolves to cluster with Indian wolves. Based on the Admixture, D-statistics,

and f3 results, however, this was attributable at least in part to gene flow from Indian wolves to West Asian wolves (Fig. 1.1c).

Based on the qpGraph analysis of the 5 primary gray wolf populations (i.e., excluding Ladakh, Qinghai, and West Asian wolves), the best-supported topology placed Indian wolf in the most basal position, followed by Tibetan wolf, and then the Holarctic populations (Fig. S1.7). Moreover, adding Ladakh, Qinghai, and West Asian wolves to this topology allowed us to quantify admixture (Fig. 1.2). The West Asian wolves (Iran, Saudi Arabia, and Syria) were primarily Holarctic, sharing 70% ancestry with Central Asian wolves, but also contained significant admixture (estimated at 30%) from Indian wolves. Wolves of Ladakh in the Western Himalayas contained significant three-way admixture among Central Asian, Tibetan, and Indian wolves, consistent with the geographic location at this three-way contact zone. Finally, Qinghai wolves were predominantly Tibetan, but with significant (estimated at 30%) admixture from East Asian wolves.

3.2 Topology weighting and recombination

To better distinguish signals of phylogeny and gene flow from ILS among Indian, Tibetan, and Holarctic wolves, we investigated the frequencies of the three possible topologies rooted to the dhole (*Cuon alpinus*; Fig. 1.3a). The expectation was that all topologies would occur in some frequency by chance due to ILS, but that topologies reflecting gene flow and phylogeny would occur more frequently. We quantified weights of each topology in terms of its estimated frequency in 100-SNP windows using Twisst (Martin and Belleghem 2017). On autosomes, the average weights were highest for the topology with the Indian wolf as basal (TC) but differences were slight (Fig. 1.3b). On the X chromosome, that same topology and the one with the Tibetan wolf as basal (TA) had similar average weights, both of which were higher than the one with

Holarctic wolves as basal (TB) (Fig. S1.8a,b,e). These findings accord with those of mitochondrial genomes in revealing the Holarctic wolf lineages on average to be most derived.

However, average frequencies of the topologies are informative primarily with respect to their occurrence beyond that expected by chance (ILS), providing little indication of their phylogenetic versus introgressive sources. To better parse phylogeny and gene flow, we took advantage of the relationship between recombination rate and introgressive versus phylogenetic signal. Because low-recombination regions of the chromosome tend to be more resistant than high-recombination regions to introgression (and to be less affected by ILS), they tend to harbor a comparably greater proportion of topologies reflecting the underlying phylogeny (Pease and Hahn 2013); in such cases, some of these low-recombination regions are expected to reflect high weights at or near 1. As a direct result of frequent recombination, which translates to shorter linkage blocks, it is comparably rare for high-recombination regions to reflect consistent topologies necessary to produce weights at or near 1. Frequencies of topologies with a weight of 1 were substantially higher for the topology with the Indian wolf in the basal position (TC) than the other two topologies, both on autosomes and the X chromosome (Fig. 1.3c). Continuous frequency distributions of topology weights more generally indicated the topology with the Indian wolf as basal (TC) to be associated with higher weights than that with the Tibetan wolf as basal (TA), both of which had higher weights than the topology with Holarctic wolves as basal (TB; Fig. S1.9a-c); the topology with the Holarctic wolf as basal (TB) also had the highest frequency of zero weights on both autosomes and the X chromosome. Kolmogorov-Smirnov tests indicated these differences were highly significant ($P << 0.001$; Fig. S9). Together, these findings suggest that the Indian wolf lineage is the most ancestral of the gray wolves, followed by that of the Tibetan wolf.

To more directly address the positioning of North American gray wolves relative to Eurasian populations, we repeated the analysis, but this time with Holarctic wolves divided into North American and Asian lineages, which have been isolated from each other since the flooding of the Beringian land bridge 12,000 years ago (Jakobsson et al. 2017). This resulted in 4 taxa (plus outgroup) with 15 possible topologies (Fig. 1.3d). Although differences among the average topology weights of autosomal loci were too slight to be informative, none of the three highest-weighted topologies (T12, T13, T9, respectively) on the X chromosome included North American wolves in the basal position despite their isolation from the other Eurasian wolves since the last ice age (Fig. 1.3e). Based on frequencies with high weights (>0.8), we found a similar but much stronger pattern favoring these three topologies on both autosomes and the X chromosome (Fig. 1.3f). These 3 topologies (T12, T13, T9) corresponded to both Indian and Tibetan wolf lineages as basal clades and North American and Asian Holarctic clades as sister taxa. Continuous frequency distributions of topology weights also indicated the topology with the Indian wolf as basal (T12) to be associated with higher weights than that with the Tibetan wolf as basal (T13), both of which had higher weights than the topology with North American wolves as basal (e.g., T7; Fig. S1.9d-f); the T7 topology (with North American wolves as basal) also had the second highest frequency of zero weights on the autosomes and the third highest frequency of zero weights on the X chromosome. As with the 3-group topologies, pairwise Kolmogorov-Smirnov tests among T7, T12, and T13 topology weight distributions were all highly significant ($P<<0.001$; Fig. S1.9). Thus, both Holarctic Asian and North American wolves formed the most derived lineages despite their 12 kya isolation on separate continents.

Lastly, we took advantage of the relationship between recombination rate and resistance to introgression to further distinguish topologies that reflect the original species topology. We

partitioned windows into regions of low (<0.2 cM/Mb), intermediate (0.2-2 cM/Mb), and high (>2 cM/Mb) recombination rates using the domestic dog recombination map (Auton et al. 2013). Windows with low, medium, and high recombination comprised ~18.8%, 62.3% and 18.2% of the total windows, respectively. Low-recombination regions were expected to most frequently retain the original species topology, whereas high-recombination regions were expected to more frequently facilitate introgression.

Consequently, we expected a negative relationship between recombination rate and topology weight for those windows reflecting the original species topology and positive relationships for those windows reflecting gene flow (Martin et al. 2019, Manual et al. 2020). Consistent with this prediction, the three topologies (T9, T12, T13) ranked highest above (Fig. 1.3f) also exhibited negative relationships between frequency of high-weighted topologies and recombination rates for autosomes and the X chromosome (Fig. 1.3g, h). For the X chromosome, low-recombination regions were dominated by T12 (Indian, then Tibetan as basal) and T13 (Tibetan, then Indian as basal), supporting that Indian and Tibetan wolves are ancestral lineages (Fig. S1.8c,d,e). In contrast, the three next-highest ranked topologies (T11, T10, T7, respectively) among autosomal regions exhibited positive relationships between topology weights and recombination rates, consistent with introgression (Fig. 1.3g,h). Finally, we inferred the autosomal and X chromosome phylogeny using only low recombination regions (<0.2 cM/Mb). While the low-recombination autosomal phylogeny was similar to the genome-wide autosomal phylogeny (Fig. S1.10), the low-recombination X chromosome strongly supported the Indian wolf as the most ancestrally diverging wolf lineage (Fig. 1.4, Fig. S1.8).

4. Discussion

We investigated the genomic distinctiveness of two southern Asian wolf populations that were previously found to exhibit divergent gray wolf matrilines. Previous phylogenomic studies of the gray wolf had not considered the Indian wolf and reached conflicting conclusions with respect to the topological positioning of the Tibetan wolf, possibly due to the confounding effects of gene flow (Fan et al. 2016, Li et al. 2019, Wang et al. 2019; Wang et al. 2020). Here, we used gray wolf nuclear sequence data representing wolves from all three major mitochondrial lineages and most of the Holarctic range in our efforts to disentangle ancestral relationships from recent gene flow. In agreement with mitochondrial phylogenies, we found that the two geographically restricted southern Asian populations were genomically distinct from the Holarctic lineage, the latter of which currently composes most populations of the species. Further, these southern Asian lineages were the most distantly divergent among extant gray wolves throughout their broad geographic range. In contrast to mitochondrial patterns, however, our findings from the qpGraph analyses and those of the low-recombination regions of the X chromosome suggest that the Indian wolf could represent an even more basal lineage than the Tibetan wolf. These results highlight the conservation significance of the remaining Indian and Tibetan wolf populations.

We also found signals of gene flow between adjacent wolf populations in Asia. In particular, admixture analyses, the PCA, D and f3 statistics, and admixture graphs illuminated two important linkages. First, Tibetan wolves in Ladakh harbored significant admixture with nearby Indian and Holarctic wolves, consistent with the mitochondrial evidence pointing to this region as a 3-way contact zone (Sharma et al. 2004). Phenotypically, the wolves of Ladakh clearly reflect that of a Tibetan wolf in terms of its size, wooly coat, howl, and genetic adaptation to hypoxia (Lydekker 1900, Hennelly et al. 2017, Wang et al. 2020). To the north, previous studies

also have found admixture between Tibetan and lowland Asian wolves (Werhahn et al. 2020). In these cases, admixture occurs only at the margins of the Tibetan plateau, suggesting that strong selective pressures associated with high altitude habitats, natal habitat-biased dispersal, or genomic incompatibilities may hinder widespread introgression between Tibetan and Holarctic wolves.

Second, we detected significant signals of gene flow between West Asian and Indian wolves. Based on the qpGraph analysis as well as the observed relationships between topologies and recombination, this connectivity likely reflects admixture following secondary contact between Holarctic and Indian lineages. In particular, Indian and West Asian wolves exhibited contrasting patterns of ancestry across their genomes, particularly in the mitogenome and within low-recombination regions of the X chromosome. This pattern suggests West Asian wolves and Indian wolves reflect independent ancestry. However, West Asian and Indian populations also share ancestry across the autosomes that is distinct from Holarctic wolves and, therefore, consistent with gene flow from the Indian wolf westward. High connectivity spanning southwest Asia and the Indian subcontinent during the Pleistocene has been observed in other taxa, facilitated by relatively continuous and similar habitats that fostered migration during interglacial periods (Rohland et al. 2005, Blinkhorn et al. 2015, Jana and Karanth 2020). Thus, ancient gene flow likely connected Indian wolves and Holarctic wolves of West Asia, possibly during the contraction of the Thar desert during interglacials 80-130 kya or 27-60 kya (Blinkhorn et al. 2015). As reflected by their current subspecific classification (*C. lupus pallipes*), West Asian wolves and Indian wolves presently share many morphological characteristics consistent with their similar arid environments. Thus, it is conceivable that some shared ancestry linking these populations reflects selective introgression of Indian wolf genes into a Holarctic wolf genomic

background. Future investigation of this question would benefit from additional samples from Pakistan and other adjacent regions.

Our finding that in the lowest-recombination regions of the X chromosome and in the admixture graph analyses, the Indian wolf lineage was even more ancestral than that of the Tibetan wolf contrasted with the mitochondrial and averaged autosomal phylogenies. Mito-nuclear discordance is not uncommon in mammals (Toews and Brelsford 2012) and has been noted previously within the *Canis* clade (Gopalakrishnan et al. 2018). Such patterns can arise due to chance (i.e., incomplete lineage sorting), sex-based asymmetries, or differential patterns of selection on mitochondria and regions of the nuclear genome. Similar to other Tibetan species, certain mitochondrial genes of the Tibetan wolf confer adaptations to hypoxic conditions on the Tibetan plateau (Sun et al. 2014, Zhou et al. 2014, Liu et al. 2019). It is possible that the discordance between the mitogenome and X chromosome phylogeny could have arisen through adaptive introgression of an extinct Tibetan canid's mitogenome into the ancestral Tibetan wolf population (Wang et al. 2020). Despite discordant patterns, both nuclear and mitochondrial results were in agreement that Indian and Tibetan wolves diverged long before the expansion of Holarctic wolves, consistent with a potentially long period of isolation and independent evolution in southern Asia during the Pleistocene.

Several additional lines of evidence demonstrate that Indian and Tibetan wolves have a long history of endemism in southern Asia and support the hypothesis that southern regions of Asia have been important centers for gray wolf evolution (Aggarwal et al. 2003, Sharma et al. 2004). All gray wolf mitochondrial DNA samples, including ancient samples from throughout northern Eurasia and North America spanning the last >50,000 years, turned up no matrilines that were as ancestral as were modern Indian and Tibetan wolf matrilines (Leonard et al. 2007,

Ersmark et al. 2016, Loog et al. 2020, Meachen et al. 2020). In fact, all ancient mitogenomes thus far examined cluster within the Holarctic clade. The fossil record suggests gray wolves inhabited southern Asia continuously throughout the late Pleistocene (Kurten 1965, Dayan et al. 1992, Mashkour et al. 2008, Wang et al. 2016, Costa 2017). These observations strongly suggest that Indian and Tibetan wolf lineages experienced a long period of relative isolation in southern Asia.

Additional evidence indicates that the Tibetan wolf experienced a demographic history distinct from that of Holarctic wolves during the late Pleistocene. In particular, Tibetan wolves underwent a population decline coinciding with the population expansion of Holarctic wolves around 50 kya or more (Fan et al. 2016, Wang et al. 2020). This contrasting demographic pattern suggests that the Tibetan wolf was isolated within a separate refugium as Holarctic wolves expanded across Asia and North America during the late Pleistocene. Our calibrated estimates of branching points on the mitochondrial phylogeny and others (e.g., Sharma et al. 2004) suggest that the Holarctic lineages were derived from a wolf population no earlier than ~100,000 years ago and that the Indian and Tibetan wolves evolved somewhat independently up to an order of magnitude further back in time. This timeframe corresponds to evidence of periodic isolation within glacial refugia for many species on parts of the Tibetan Plateau and the Indian subcontinent (Yang et al. 2009, Roberts et al. 2014, Aradhya et al. 2017). Together, these observations suggest that southern Asia contained at least two long-term refugia for gray wolves. Such a pattern, whereby southern regions in Asia harbor ancestral diversity and northern latitudes reflect recent expansion events, has been found in Eurasian lynx (*Lynx lynx*), red foxes (*Vulpes vulpes*), and brown bears (*Ursus arctos*) (Rueness et al. 2014, Statham et al. 2014, Lan et al. 2017), and may be a common pattern (Hewitt 2000). In the future, demographic analyses such

as PSMC of an Indian wolf genome sequenced to a greater depth could provide additional insight into the timing of population contractions, expansions, and divergence among these three lineages.

4.1 Conservation implications

Our study underscores the importance of conserving remnant populations of Tibetan and Indian wolves and suggests the need to reassess taxonomic designations, which will significantly affect their conservation priority. For the Tibetan wolf, our work supports previous findings of its evolutionary distinctiveness and argues for its recognition on some level as an "evolutionary significant unit" (Werhahn et al. 2018, 2020; Wang et al. 2020). We observed mitochondrial and nuclear genomic distinctiveness between Tibetan and Holarctic wolves east and north of the Tibetan Plateau, as well as a potential contact zone with Indian and Holarctic wolves in Ladakh, India. Sampling more gray wolves from west of the Tibetan Plateau, such as Northern Pakistan, would provide insight into whether reproductive barriers could be hindering gene flow between Tibetan and Holarctic wolves, a consideration that would help inform whether species status is warranted for the Tibetan wolf.

An implication of our findings is that the Indian wolf could be far more endangered than previously recognized. In 2003, Indian wolves were thought to number around 2,000 to 3,000 individuals in India, with an unknown number of individuals in a declining population in Pakistan (Jhala 2003, Sheikh and Molur 2004). However, the current taxonomy does not distinguish Indian wolves from West Asian wolves, collectively considered *Canis lupus pallipes*, which spans much of southern and western Asia. In light of our nuclear genomic findings, which accord with previous mitochondrial studies, the ancestral Indian wolf distribution is much smaller and potentially restricted to the lowland peninsular Indian subcontinent (Sharma et al.

2004, Castello 2018, Hamid et al. 2019). The current broad geographic designation is largely due to similar phenotypes of Holarctic and Indian wolves occupying these arid regions. Based on our finding that West Asian wolves reflect admixture between Holarctic and Indian lineages, we hypothesize that phenotypic similarities resulted from selective introgression of Indian wolf genes conferring adaptation to their arid environment into the Holarctic genomic backgrounds of contemporary West Asian wolves. In the future, more intensive sampling of wolves spanning Pakistan to the Middle-East will help clarify locations of contact zones and, therefore, the range extent of the Indian wolf lineage. Additionally, closing this sampling gap for gray wolves will facilitate investigations for genomic heterogeneities potentially associated with selection on functional loci, such as those affecting body size, allometry, or reproductive phenology – all potentially associated with adaptations to arid environments. Thus, while our findings strongly support a taxonomic revision that differentiates West Asian and Indian wolves, a better understanding of these evolutionary relationships and of phylogenetic divergence times is essential to deciding whether to do so at the level of species (e.g., Aggarwal et al. 2007) or subspecies. In the meantime, we recommend the Indian wolf be considered an "evolutionary significant unit" in order to prioritize conservation efforts of this highly endangered lineage. Taken together, our work highlights that the Indian wolf constitutes one of the world's most endangered and evolutionarily distinct gray wolf populations.

5. References

Aggarwal R.K., Kivisild T., Ramadevi J., & Singh L. (2007). Mitochondrial DNA coding region sequences support the phylogenetic distinction of two Indian wolf species. *Journal of Zoolological Systematics and Evolutionary Research, 45*:163-172.

Aggarwal R.K., Ramadevi J., & Singh L. (2003). Ancient origin and evolution of the Indian wolf: evidence from mitochondrial DNA typing of wolves from Trans-Himalayan region and Peninsular India. *Genome Biology,* 4:P6.

Alexander D.H., & Lange K. (2011). Enhancements to the Admixture algorithm for individual ancestry estimation. *BMC Bioinformatics, 12*:246.

Aradhya M., Velasco D., Ibrahimov Z., Toktoraliev B., Maghradze D., Musayev M, … Preece J.E. (2017). Genetic and ecological insights into glacial refugia of walnut (*Juglans regia L.)*. *PLoS One, 12*:e0185974.

Auton A., Li Y.R., Kidd J., Oliveira K., Nadel J., Holloway J.K., … Boyko A.R. (2013). Genetic recombination is targeted towards gene promoter regions in dogs. *Plos Genetics, 9*:e1003984.

Bergero R., & Charlesworth D. (2009). The evolution of restricted recombination in sex chromosomes. *Trends in Ecology and Evolution, 24*(2):94-102.

Blinkhorn J., Achyuthan H., & Petraglia M.D. (2015). Ostrich expansion into India during the Late Pleistocene: Implications for continental dispersal corridors. *Palaeogeography, Palaeoclimatology, Palaeoecology, 417*:80-90.

Boitani L., Phillips M., & Jhala Y.V. (2018). *Canis lupus* (errata version published 2020). The IUCN Red List of Threatened Species.

Butlin R.K. (2005). Recombination and speciation. *Molecular Ecology, 14*:2621-2635.

Castello J.R. (2018). Wolf-like canids. In Canids of the world: Wolves, wild dogs, foxes, jackals, coyotes, and their relatives. (pp. 74-172). Princeton New Jersey: Princeton University press

Chavez D.E., Gronau I., Hains T., Kliver S., Koeplfi K.P., & Wayne R.K. (2019). Comparative genomics provides new insights into the remarkable adaptations of the African wild dog (*Lycaon pictus*). *Scientific Reports, 9*:8329.

Costa A.G. (2017). A new Late Pleistocene fauna from arid coastal India: Implications for inundated coastal refugia and human dispersals. *Quaternary International, 436*:253-269.

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., … Sherry S.T. (2011). The variant call format and VCFtools. *Bioinformatics, 27*:2156-2158.

Dayan T., Simberloff D., Tchernov E., & Yom-Tov Y. (1992). Canine carnassial: character displacement in the wolves, jackals, and foxes of Israel. *Biological Journal of the Linnean Society, 45*:313-331.

Dierckxsens N., Mardulyn P., & Smits G. (2016). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research, 45*:e18.

Drummond A.J., & Rambaut A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology, 7*: 214.

Edgar R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics, 5*:1-19.

Ersmark E., Klutsch C.F.C, Chan Y.L., Sinding M.H.S., Fain S.R., Illarionova N.A., … Savolainen P. (2016). From the past to the present: wolf phylogeography and demographic history based on the Mitochondrial control region. *Frontiers in Ecology and Evolution*, *4*:134. doi: 10.3389/fevo.2016.00134.

Fan Z., Silva P., Gronau I., Wang S., Armero A.S., Schweizer R.M., … Wayne R.K. (2016). Worldwide patterns of genomic variation and admixture in gray wolves. *Genome Research, 26*:163-173.

Fontaine M.C., Pease J.B., Steele A., Waterhouse R.M., Neafsey D.E., Sharakhov I.V., … Michell S.N. (2015). Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science, 347*(6217)1258524-1-1258524-6.

Garrison E., & Marth G. (2012). Haplotype-based variant detection from short-read sequencing *Arxiv*.

Gopalakrishnan S., Sinding M.H.S., Ramos-Madrigal J., Niemann J., Castuita J.A.S., Vieira F.G., … Gilbert M.T.P. (2018). Interspecific gene flow shaped the evolution of the genus *Canis*. *Current Biology, 28*:3441-3449.

Guindon S., Dufayard J.F, Lefort V., Anisimova M., Hordijk W., & Gascuel O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology, 59*:307-321.

Hamid A., Mahmood T., Fatima H., Hennelly L.M., Akrim F., Hussain A., & Waseem M. (2019). Origin, ecology and human conflict of gray wolf (*Canis lupus*) in Suleman Range, South Waziristan, Pakistan. *Mammalia, 83*:539-551.

Hennelly L.M., Habib B., & Lyngdoh S. (2015). Himalayan wolf and feral dog displaying mating behavior in Spiti Valley, India, and potential conservation threats from sympatric feral dogs. *Canid Biology and Conservation, 18*:27-30.

Hennelly L.M., Habib B., Root-Gutteridge H., Palacios V., & Passilongo D. (2017). Howl variation across Himalayan, North African, Indian, and Holarctic wolf clades: tracing divergence in the world's oldest wolf lineages using acoustics. *Current Zoology, 63*:341-348.

Hewitt G. (2000). The genetic legacy of the Quaternary ice ages. *Nature, 403*:907-913.

Hofreiter M., & Stewart J.R. (2009). Ecological change, range fluctuations, and population dynamics during the Pleistocene. *Current Biology, 19*:R584-R594.

Hundertmark K.J., Shields G.F., Udina I.G., Bowyer R.T., Danilkin A.A., & Schwartz C.C. (2002). Mitochondrial phylogeography of moose (*Alces alces*): late Pleistocene divergence and population expansion. *Molecular Phylogenetics and Evolution, 22*:375-387.

Jakobsson M., Pearce C., Cronin T.M., Backman J., Anderson L.G., Barrientos N., … O'Regan M. (2017). Post-glacial flooding of the Bering Land Bridge dated to 11 ca ka BP based on new geophysical and sediment records. *Climate of the Past*, *13*:991-1005.

Jana A., & Karanth P. (2020). Multilocus nuclear markers provide new insights into the origin and evolution of the blackbuck (*Antilope cervicapra*, Bovidae). *Molecular Phylogenetics and Evolution, 139*:106560.

Jhala YV. 2003. Status, ecology, and conservation of the Indian wolf *Canis lupus pallipes Skyes. J. Bombay Nat Hist Soc* 100:293-307.

Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., Haeseler A., & Jermiin L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods, 14*:587-589.

Koblmuller S., Vila C., Lorente-Galdos B., Dabad M., Ramirez O., Marques-Bonet T., … Leonard J.A. (2016). Whole mitochondrial genomes illustrate ancient intercontinental dispersals of grey wolves (*Canis lupus*). *Journal of Biogeography, 43*:1728-1738.

Kurten B. (1965). The carnivora of the Palestine caves. *Acta Zooogical Fennica, 107*:74.

Lan T., Gill S., Bellemain E., Bischof B., Nawaz M.A., & Lindqvist C. (2017). Evolutionary history of enigmatic bears in the Tibetan plateau-Himalayan region and the identity of the yeti. *Proceedings of the Royal Society B, 284*:20171804.

Leonard J.A., Vila C., & Wayne R.K. (2005). Legacy lost: genetic variability and population size of extirpated US grey wolves (*Canis lupus*). *Molecular Ecology, 14*:9-17.

Leonard J.A., Vilas C., Fox-Dobbs K., Koch P.I., Wayne R.K., & Valkenburgh B.V. (2007). Megafaunal Extinction and the Disappearance of a Specialized Wolf Ecomorph. *Current Biology, 17*:1146-1150.

Li G., Figueiro H.V., Eizirik E., & Murphy W.J. (2019). Recombination-aware phylogenomics reveals the structured genomic landscape of hybridizing cat species. *Molecular Biology and Evolution, 36*:2111-2126.

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

Li H., Handsaker B., Wyoker A., Fennell T., Ruan J., Homer N., Marth G., … Durbin R. (2009). 1000 Genomes Project Data Processing Subgroup, The sequence alignment/map format and SAMtools. *Bioinformatics, 25*:2078-2079.

Liu G., Zhao C., Yang X., Shang J., Gao X., Sun G., Dou H., & Zhang H. (2019). Comparative analysis of peripheral blood reveals transcriptomic adaptations to extreme environments on the Qinghai-Tibetan Plateau in the gray wolf (*Canis lupus chanco*). *Organisms Diversity and Evolution, 19*:543-556.

Loog L., Thalmann O., Sinding M.H.S., Schuenemann V.J., Perri A., Germonpre M., … Velasco M.S. (2020). Ancient DNA suggests modern wolves trace their origin to a Late Pleistocene expansion from Beringia. *Molecular Ecology, 29*:1596-1610.

Lydekker R. (1900). The Great and Small Game of India, Burma, and Tibet. Rowland Ward Limited, London.

Manuel M., Barnett R., Sandoval-Velasco M., Yamaguchi N., Vierira F.G., Mendoza M.L.Z., … Gilbert M.T.P. (2020). The evolutionary history of extinct and living lions. *Proceedings of the National Academy of Sciences of the United States of America, 117*:10927-10934.

Marin S.H., & Jiggins C.D. (2017). Interpreting the genomic landscape of introgression. *Current Opinion in Genetics and Development, 47*:69-74.

Martin S.H., Davey J.W., Salazar C., & Jiggins C.D. (2019). Recombination rate variation shapes barriers to introgression across butterfly genomes. *Plos Biology, 17*:e2006288.

Martin S.H., & Van Belleghem S.M. (2017). Exploring evolutionary relationships across the genome using topology weighting. *Genetics, 206*:429-438.

Mashkour M., Monochot H., Trinkaus E., Reyss J.L., Biglari F., Bailon S., … Abdi K. (2008). Carnivores and their prey in the Wezmeh Cave (Kermanshah, Iran): a Late Pleistocene refuge in the Zagros. *International Journal of Osteoarchaeology, 19*:678-694.

Meachen J., Wooller M.J., Barst B.D., Funck J., Crann C., Heath J., … Zazula G. (2020). A mummified Pleistocene gray wolf pup. *Current Biology, 30*:R1467-R1468.

Muirhead C.A., & Presgraves D.C. (2016). Hybrid incompatibilities, local adaptation, and the genomic distribution of natural introgression between species. *The American Naturalist,* 187:249-261.

Nachman M.W., & Payseur B.A. (2012). Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philosophicall Transactions of the Royal Society B, 367*:409-421.

Nguyen L.T., Schmidt H.A., Haeseler A., & Minh B.Q. (2014). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution, 32*:268-274.

Palkopoulou E., Baca M., Abramson N.I., Sablin M., Socha P., Nadachowski A., … Smirnov N.G.. (2016). Synchronous genetic turnovers across Western Eurasia in Late Pleistocene collared lemmings. *Global Change Biology, 22*:1710-1721.

Patterson N., Moorjani P., Luo Y., Mallick S., Rohland N., Zhan Y., … Reich D. (2012). Ancient admixture in human history. *Genetics 192*:1065-1093.

Pease J.B., & Hahn M.W. (2013). More accurate phylogenies inferred from low-recombination regions in the presence of incomplete lineage sorting. *Evolution, 67*:2376-2384.

Petit R.J., Aguinagalde I., Beaulieu J., Bittkau C., Brewer S., Cheddadi R., …. Vendramin G.G. (2003). Glacial Refugia: Hotspots but not melting pots of genetic diversity. *Science, 300*:1563-1565.

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

Presgraves D.C. (2008). Sex chromosomes and speciation in Drosophila. *Trends in Genetics 24*(7):336-343.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., … Sham P.C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics, 81*:559-575.

Rambaut A., Drummond A.J., Xie D., Baele G., & Suchard M.A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology 67*:901-904.

Roberts P., Delson E., Miracle P., Ditchfield P., Roberts R.G., Jacobs Z., … Petraglia M.D. (2014). Continuity of mammalian fauna over the last 200,000 years in the Indian subcontinent. *Proceedings of the National Academy of Science*, *111*: 5848-5853.

Rohland N., Pollack J.L., Nagel D., Beauval C., Airvaux J., Paabo S., & Hofreiter M. (2005).The population history of extant and extinct hyenas. *Molecular Biology and Evolution, 22*:2435-2443.

Rueness E.K., Naidenko S., Torsvik P., & Stenseth N.C. (2014). Large-scale genetic structuring of a widely distributed carnivore – the Eurasian lynx (*Lynx lynx*). *PLoS ONE, 9:*e93675.

Schaffner S.F. (2004). The X chromosome in population genetics. *Nature Reviews Genetics, 5*:43-51.

Schumer M., Xu C., Powell D.L., Durvasula A., Skov L., Holland C., … Przeworki M. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science,* 360:656-660.

Sharma D.K., Maldonado J.E., Jhala Y.V., & Fleischer R.C. (2004). Ancient wolf lineages in India. *Proceedings of the Royal Society B*. *271*:S1-4.

Sheikh K.M., & Molur S. (2004). Status and Red List of Pakistan's Mammals. *IUCN Pakistan*.

Statham M.J., Murdoch J., Janecka J., Aubry K.B., Edwards C.J., Soulsbury C.D., … Sacks B.N. (2014). Range-wide multilocus phylogeography of the red fox reveals ancient continental divergence, minimal genomic exchange, and distinct demographic histories. *Molecular Ecology, 23*:4813-4830.

Subramanian S., & Lambert D.M. (2011). Time dependency of molecular evolutionary rates? Yes and no. *Genome Biology and Evolution, 3*:1324-1328.

Sun J., Zhong H., Chen S.Y., Yao Y.G., & Liu Y.P. (2013). Association between MT-CO3 haplotypes and high-altitude adaptation in Tibetan chickens. *Gene, 529*:131-137.

Suryawanshi K.P., Bhatnagar Y.V., Redpath S., & Mishra C. (2013). People, predators and perceptions: patterns of livestock depredation by snow leopards and wolves. *Journal of Applied Ecology,* 50:550-560.

Toews D.P.L., & Brelsford A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology, 21*:3907-3930.

Wang G.D., Zhang M., Wang X., Yang M.A., Cao P., Liu F., … Zhang Y.P. (2019). Genomic approaches reveal an endemic subpopulation of gray wolves in Southern China. *iScience, 20*:110-118.

Wang L., Ma Y.P., Zhou Q.J., Zhang Y.P., Savolainen P., & Wang G.D. (2016). The geographical distribution of grey wolves (*Canis lupus*) in China: a systematic review. *Zoological Research 18*:315-326.

Wang M.S., Wang S., Li Y., Jhala Y.V., Thakur M.., Otecko N.O., … Wu D.D. (2020). Ancient hybridization with an unknown population facilitated high altitude adaptation of canids. *Molecular Biology and Evolution, 37*:2616-2629.

Werhahn G., Liu Y., Meng Y., Cheng C., Lu Z., Atzeni L., … Senn H. (2020). Himalayan wolf distribution and admixture based on multiple genetic markers. *Journal of Biogeography, 47*:1272-1285.

Werhahn G., Senn H., Ghazali M., Karmacharya D., Sherchan A.M., Joshi J., Kusi N., … Macdonald D.W. (2018). The unique genetic adaptation of the Himalayan wolf to high-altitude and consequences for conservation. *Global Ecology and Conservation, 16*:e00455.

Yang S., Dong H., Lei F. (2009). Phylogeography of regional fauna on the Tibetan plateau: a review. *Progress in Natural Science, 19*:789-799.

Zhou T., Shen X., Irwin D.M., Shen Y., Zhang Z. (2014). Mitogenomic analyses propose positive selection in mitochondrial genes for high-altitude adaptation in galliform birds. *Mitochondrion, 18*:70-75.

6. Main Figures

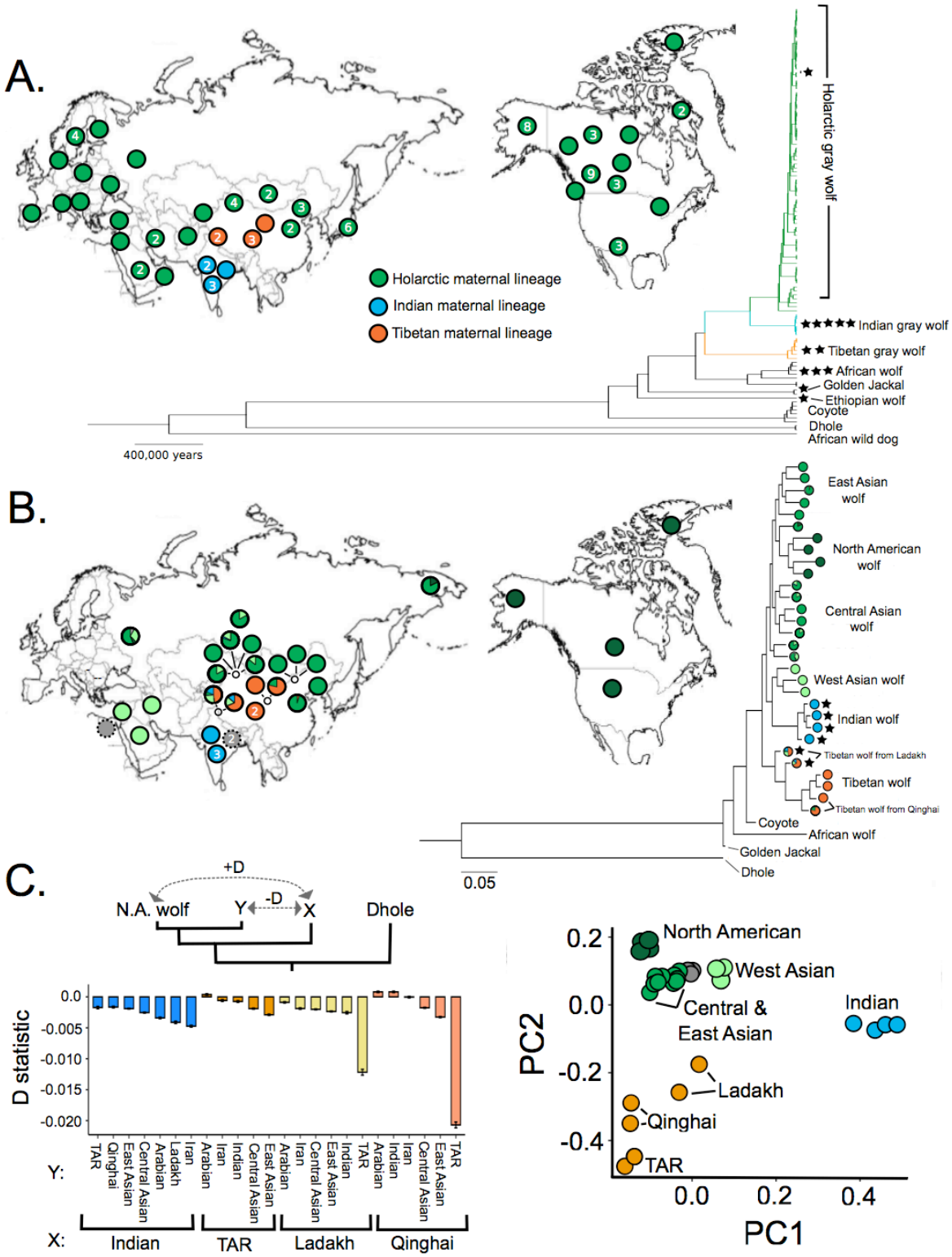


Figure 1. Geographic locations of samples with mitochondrial and nuclear genomic profiles**.** (A) Locations of 87 complete gray wolf mitogenomes and phylogenetic tree of these and 19 additional samples from 6 canid species (Table S2). The phylogeny was inferred using 15,437 bp of non-coding and coding regions of the mitogenome, excluding the D-loop region. We constructed the phylogenetic tree in BEAST v1.10.4 (Drummond and Rambaut 2007), rooted to the African wild dog (*Lycaon pictus*), and calibrated to a tree height of 3.9 Ma (SD = 0.3 Ma; Chavez et al. 2019). Numerals indicate numbers of samples within a location and stars indicate newly sequenced mitogenomes included in the study. (B) Locations of 33 gray wolf (colored circles) and dog (gray circles) whole genome sequences and autosomal phylogeny using 30 gray wolves and 4 canid species constructed using IQ-tree (version 1.6.12) with using 1000 ultra-fast bootstrap approximation (UFBoot) and a TVM+F+R2 substitution model inferred using ModelFinder within IQ-tree (Nyugen et al. 2014, Kalyaanamoorthy et al. 2017). All nodes received 100% ultrafast bootstrap support. Colors of the sample location circles correspond to population assignment results from the admixture analysis at K = 6 computed in Admixture (Alexander and Lange 2011) (Fig. S2). (C) Results of the D-statistic and principle component analyses (PCA), illustrating signals of gene flow between wolf populations. Negative D-statistics indicate gene flow between population Y and X, whereas positive D-statistics indicate gene flow between X and North American wolves. For the PCA, we used ~3.5 million SNPs consisting of 30 gray wolves and 3 dog samples (shown as gray circles).

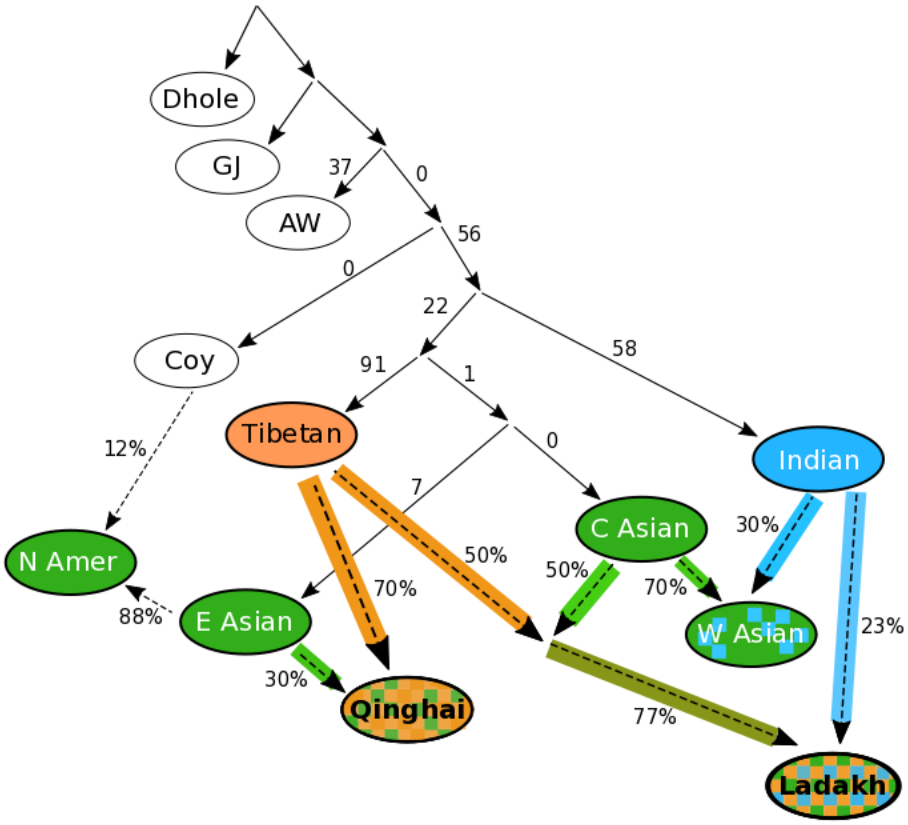


Figure 2. Graph illustrating estimates of admixture in wolves from Ladakh, Qinghai, and West Asia (W Asia) with respect to five primary wolf populations arranged according to the best-supported topology (Fig. S7). The admixture graph was inferred using qpGraph within the ADMIXTOOLS package (Patterson et al. 2012) and approximately 30 million autosomal SNPs. Percentages and colored arrows correspond to estimated ancestry proportions and admixture from primary wolf populations: Central (C) Asia, East (E) Asia, North (N) America, Tibetan, and Indian. Outgroups included the dhole, golden jackal (GJ), African wolf (AW), and coyote (coy). The displayed model had the highest support (lowest model score) among the five alternative full models corresponding to the foundational topologies shown in Fig. S7 but with Ladakh, Qinghai, and West Asian wolf populations and their admixture relationships added, as shown.

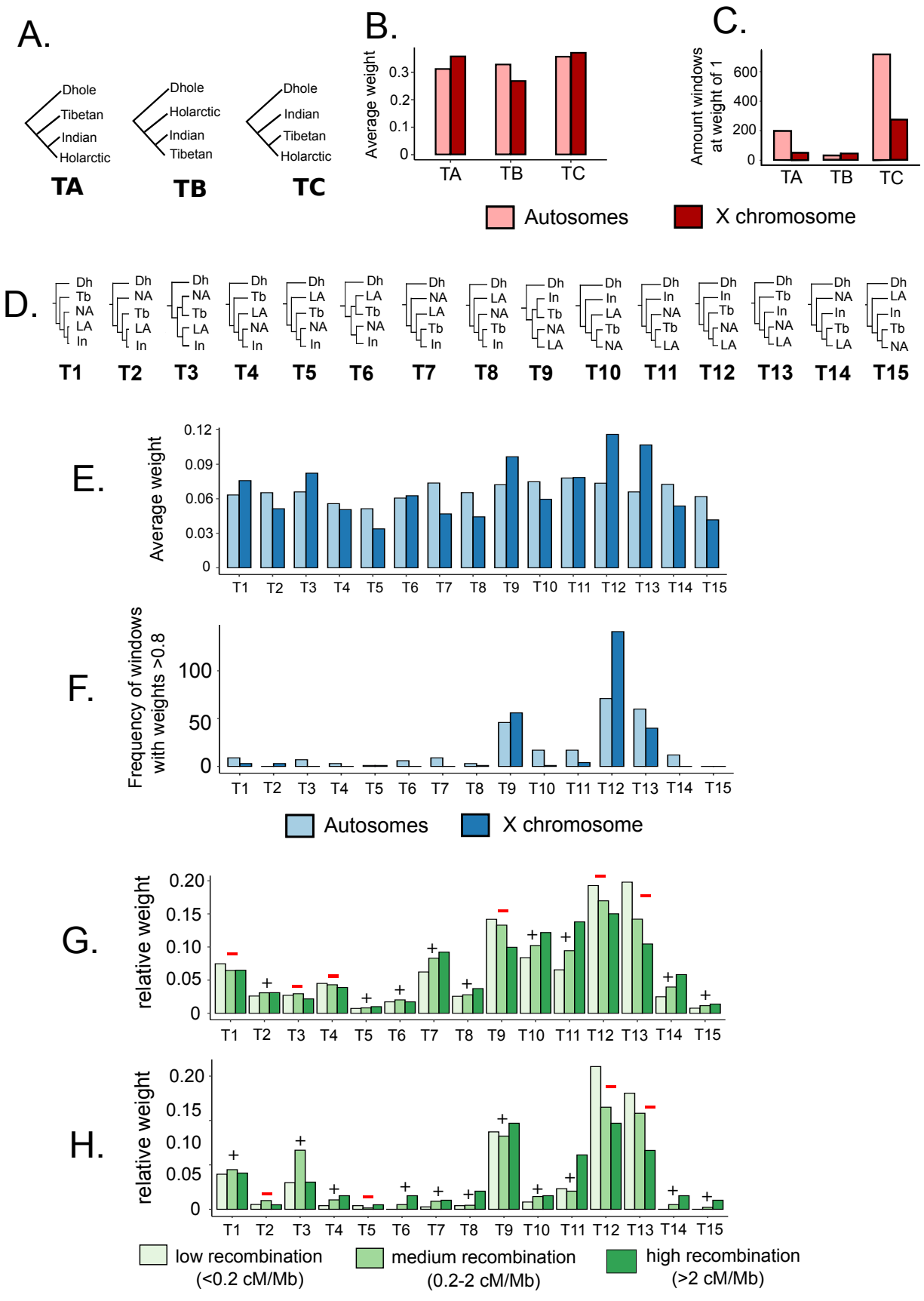

A.
Dhole
Tibetan
Indian
Holarctic
TA
Dhole
Holarctic
Indian
Tibetan
TB
Dhole
Indian
Tibetan
Holarctic
TC
B.
Average weight
0.3
0.2
0.1
0
TA
TB
TC
C.
Amount windows at weight of 1
600
400
200
0
TA
TB
TC
Autosomes
X chromosome
D.
T1
T2
T3
T4
T5
T6
T7
T8
T9
T10
T11
T12
T13
T14
T15
E.
Average weight
0.12
0.09
0.06
0.03
0
T1 T2 T3 T4 T5 T6 T7 T8 T9 T10 T11 T12 T13 T14 T15
F.
Frequency of windows with weights >0.8
100
50
0
T1 T2 T3 T4 T5 T6 T7 T8 T9 T10 T11 T12 T13 T14 T15
Autosomes
X chromosome
G.
relative weight
0.20
0.15
0.10
0.05
0
T1 T2 T3 T4 T5 T6 T7 T8 T9 T10 T11 T12 T13 T14 T15
H.
relative weight
0.20
0.15
0.10
0.05
0
T1 T2 T3 T4 T5 T6 T7 T8 T9 T10 T11 T12 T13 T14 T15
low recombination
(<0.2 cM/Mb)
medium recombination
(0.2-2 cM/Mb)
high recombination
(>2 cM/Mb)

Figure 3. Frequency of topologies within 100-SNP windows across the autosomes and X chromosome for Indian (In), Tibetan (Tb), and Holarctic gray wolves rooted to the dhole (Dh). (A) Three possible topologies composed of 3 gray wolf taxa used, (B) their average topology weights, and (C) frequencies of windows with weights of 1. (D) Fifteen possible topologies composed of 4 gray wolf taxa as above, but with Holarctic divided into lowland Asian (LA), and North American (NA) gray wolves, (E) their average topology weights, and (F) frequencies of windows with weights >0.8. (G, H) Proportions of 100-SNP windows with topology weights >0.5 in regions of low recombination (<0.2 cM/Mb), intermediate recombination (0.2-2cM/Mb), and high recombination (>2cM/Mb), for (G) autosomes and (H) the X chromosome. Relative weights within each recombination rate partition sum to 1 across the 15 topologies. Positive symbols above each topology on the x-axis of G and H represent a positive relationship between topology weights and recombination rate, whereas negative symbols represent a negative relationship between topology weights and recombination rates. Positive and negative relationships were calculated for each of the 15 topologies by subtracting the relative topology weight in the low recombination region by the relative topology weight in the high recombination region for a specific topology for the autosomes (G) and X chromosome (H).

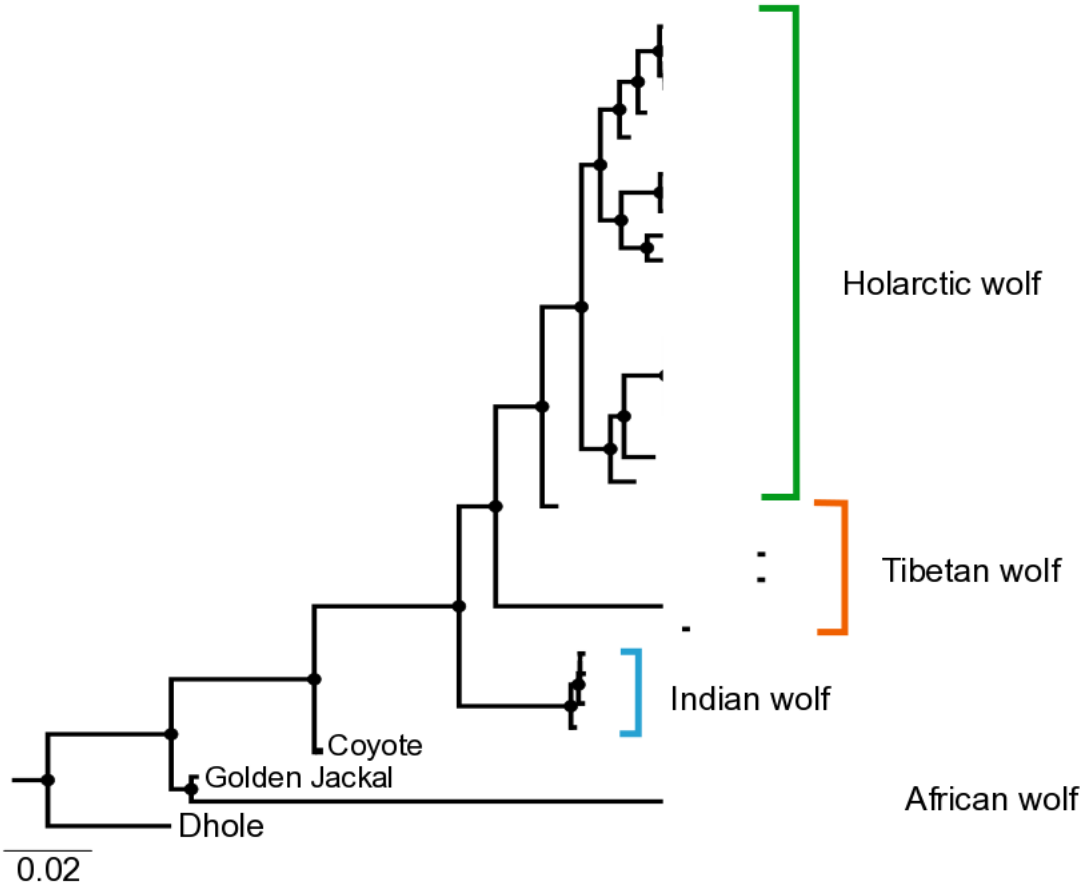


Figure 4. Phylogenetic tree of the X chromosome inferred using low recombination regions (<0.2 cM/Mb) across 30 gray wolves and 4 other canid species. The phylogeny was constructed using IQ-tree (version 1.6.12) with using 1000 ultra-fast bootstrap approximation (UFBoot) and a TVM+F+R3 substitution model inferred using ModelFinder within IQ-tree (Nyugen et al. 2014, Kalyaanamoorthy et al. 2017). We used ~210,000 SNPs that were located within low recombination regions of the X chromosome to infer the phylogenetic tree. Nodes with bootstrap support values of 100 are depicted with a black circle at corresponding nodes.

7. Supplemental Materials

Table S1.1. Sample information of whole genome-sequenced canids, including 7 sequenced in this study and 30 additional genomes assembled from short-read archives in GenBank. Wolf Population refers to the regional classification of 30 whole genome-sequenced gray wolves used in this study for the PCA, D statistic, f_3 statistic, and Twisst analyses.

GenBank ID	Other identifier	Species	Location	Depth	Wolf Population	Study
SRR14777848	BH1	Tibetan wolf	Domkhar, Sham, Ladakh, India	5.025	Ladakh	**This study**

SRR14777847	BH4	Tibetan wolf	Outside of Kargil, Ladakh, India	5.406	Ladakh	**This study**
SRR14777846	BH6	Indian wolf	Outskirts of Jodhpur, Rajasthan, India	5.93	Indian	**This study**
SRR14777845	BH123	Indian wolf	Gangewadi, Maharasthra, India	5.522	Indian	**This study**
SRR14777844	BH124	Indian wolf	Nannaj, Maharashtra, India	7.213	Indian	**This study**
SRR14777843	BH126	Indian wolf	Baramati, Maharashtra, India	8.665	Indian	**This study**
SRR14777842	BH127	Golden Jackal	Uttarakhand, India	6.724		**This study**
SRR1061818		Village Dog	Egypt	7.465		*1*
SRR1061963		Village Dog	India	5.539		*1*
SRR1061964		Village Dog	India	7.434		*1*
SRR2827601		Gray wolf	Xinjiang, China	10.105	Central Asian	*2*
SRR2827611		Gray wolf	Xinjiang, China	7.985	Central Asian	*2*
SRR3574870		Coyote	California, USA	4.948		*3*
SRR7107645		Gray wolf	Altai, Russia	9.896	Central Asian	*4*
SRR7107786		Gray wolf	Yellowstone, USA	24.879	North American	*5*
SRR7107906		Tibetan wolf	Luobulingka Zoo, Tibet (wild born), China	24.48	Tibetan	*7*
SRR7107907		Tibetan wolf	Xining Zoo, Qinghai (wild born), China	24.622	Qinghai	*6*
SRR7107908		Gray wolf	Haerbing Zoo, Inner Mongolia (wild born), China	23.617	East Asian	*6*
SRR7107909		Gray wolf	Kalamaili Nature Reserve (wild born), Xinjiang, China	22.152	Central Asian	*6*
SRR7107910		Tibetan wolf	Xining Zoo, Qinghai (wild born), China	24.098	Qinghai	*6*
SRR7107911		Gray wolf	Duzhishan Garden (wild born), Xinjiang, China	24.792	Central Asian	*6*
SRR7107912		Tibetan wolf	Kashi Garden, (wild born), Tibet, China	23.468	Tibetan	*6*
SRR7107913		Gray wolf	Haerbing Zoo, Inner Mongolia (wild born), China	20.578	East Asian	*6*
SRR7976426		African wolf	Kenya, Africa	23.601		*3*
SRR8049189		Dhole	Berlin Zoo, Germany	17.965		*7*
SRR8049193		Gray wolf	Breeding Centre for Endangered Arabian Wildlife, Saudi Arabia	13.146	West Asian	*7*
SRR8049194		Gray wolf	Breeding Centre for Endangered Arabian Wildlife, Syria	16.196	West Asian	*7*
SRR8049197		Gray wolf	10km south of Gunnar Island, Ellesmere Island, Canada	11.410	North American	*7*
SRR8066602		Gray wolf	Kobuk River, Alaska	9.510	North American	*8*
SRR8066605		Gray wolf	La Ronge, Saskatchewan, Canada	7.954	North American	*8*
SRR7107646		Gray wolf	Chukotka, Russia	7.439	East Asian	*4*

SRR7107647	Gray wolf	Bryansk, Russia	9.846	Central Asian	*4*
SRR1518518	Gray wolf	Iran	13.148	West Asian	*5*
SRR2827600	Gray wolf	Lioning, China	14.576	East Asian	*2*
SRR5168998	Gray wolf	Inner Mongolia, China	39.378	East Asian	*3*
SRR2827609	Gray wolf	Xinjiang, China	14.701	Central Asian	*2*
SRR2827603	Gray wolf	Shanxi, China	10.95	East Asian	*2*

Table S1.2. Sample information of 106 whole mitogenomes, including 13 sequenced in this study and 93 additional genomes either downloaded though Genbank or assembled using short-read archives of whole genome-sequenced canids in GenBank Short Read Archive (SRA).

Genbank Accession No	Other Identifier	Taxon	source	location	Reference
MZ433371	T2810	Ethiopian wolf	this study	Ethiopia	**this study**
MZ433379	BH127	Golden Jackal	this study	Uttarakhand, India	**this study**
MZ433369	T2243	African wolf	this study	Kheune, Senegal	**this study**
MZ433368	T1621	African wolf	this study	Mali	**this study**
MZ433370	T2257	African wolf	this study	Kheune, Senegal	**this study**
MZ433373	BH4	Gray wolf	this study	Domkhar, Sham, Ladakh, India	**this study**
MZ433372	BH1	Gray wolf	this study	Outside of Kargil, Ladakh, India	**this study**
MZ433374	BH6	Gray wolf	this study	Outskirts of Jodhpur, Rajasthan, India	**this study**
MZ433375	BH24	Gray wolf	this study	Daund, Maharashtra, India	**this study**
MZ433376	BH123	Gray wolf	this study	Gangewadi, Maharashtra, India	**this study**
MZ433377	BH124	Gray wolf	this study	Nannaj, Maharashtra, India	**this study**
MZ433378	BH126	Gray wolf	this study	Baramati, Maharashtra, India	**this study**
MZ433367	T151	Gray wolf	this study	Naryn, Kyrgyzstan	**this study**
MN071206.1	MS25	Gray wolf	Genbank	India	Loog et al. 2019
MNO71192.1	Ms9	Gray wolf	SRA	Alaska, USA	Sinding et al. 2018, assembly (Loog et al. 2019)
AM711902		Gray wolf	Genbank	Sweden	Arnason.et.al.2007
DQ480510		Coyote	Genbank	Colorado	Bjornfeldt et al. 2006

DQ480509	Coyote	Genbank	USA	Bjornfeldt et al. 2006
DQ480504	Gray wolf	Genbank	Sweden	Bjornfeldt et al. 2006
DQ480508	Gray wolf	Genbank	Canada	Bjornfeldt et al. 2006
DQ480507	Gray wolf	Genbank	Saudi Arabia	Bjornfeldt et al. 2006
GU063864.1	Dhole	Genbank	China	Chen L, Zhang HH. Direct submission
GQ374438	Gray wolf	Genbank	Inner Mongolia, China	Chen L, Zhang HH. Direct submission
MN071188	Gray wolf	Genbank	Saudi Arabia	Gopalakrishnan et al. 2018, assembly (Loog et al. 2019)
MN071204	Gray wolf	Genbank	Mexico	Gopalakrishnan et al. 2018, assembly (Loog et al. 2019)
NC_028427.1	African wild dog	Genbank	zoo: South Korea, Seoul Children's Grand Park	Hwang et al. 2015. Direct accession
KT448282.1	Dhole	Genbank		Koepfli et al. 2015
KT448275	Coyote	Genbank	Alabama	Koepfli et al. 2015
KT448276	Coyote	Genbank	California	Koepfli et al. 2015
KT448277	Coyote	Genbank	Illinois	Koepfli et al. 2015
KT448274	Golden Jackal	Genbank	israel	Koepfli et al. 2015
KT448272	African wolf	Genbank	Kenya	Koepfli et al. 2015
KT448273	African wolf	Genbank	Kenya	Koepfli et al. 2015
MN071187	Gray wolf	Genbank	Mongolia	Loog et al. 2019
MN071185	Gray wolf	Genbank	Afghanistan	Loog et al. 2019
MN071186	Gray wolf	Genbank	Denmark	Loog et al. 2019
MN071196	Gray wolf	Genbank	Banks Island, Canada	Loog et al. 2019
MN071198	Gray wolf	Genbank	Ellesmere Island, Canada	Loog et al. 2019
MN071205	Gray wolf	Genbank	Spain	Loog et al. 2019
MN071189	Gray wolf	Genbank	Turkey	Loog et al. 2019
AB499818	Gray wolf	Genbank	Japan, Kochi	Matsumura et al. 2014
AB499821	Gray wolf	Genbank	Japan, Kanagawa	Matsumura et al. 2014
AB499823	Gray wolf	Genbank	Japan, Hiroshima	Matsumura et al. 2014
AB499824	Gray wolf	Genbank	Japan, Hiroshima	Matsumura et al. 2014
AB499822	Gray wolf	Genbank	Japan, Nagano	Matsumura et al. 2014

AB499825	Gray wolf	Genbank	Japan	Matsumura et al. 2014
EU442884	Gray wolf	Genbank	Qinghai, China	Meng C and Zhang H. Direct Submission
FJ032363	Gray wolf	Genbank	Tibet, China	Meng et al. 2009
NC_027956.1	African wolf	Genbank	Atlas mountains, Morocco	Moliner et al. 2016
KT378606.1	African wolf	Genbank	Atlas mountains, Morocco	Moliner et al. 2016
EU789789	Coyote	Genbank		Pang et al. 2009
MN071199	Gray wolf	Genbank	Qamanirjuaq Canada	Sinding et al. 2018, assembly (Loog et al. 2019)
MN071201	Gray wolf	Genbank	Toronto Canada	Sinding et al. 2018, assembly (Loog et al. 2019)
MN071193	Gray wolf	Genbank	Altantic coast, Canada	Sinding et al. 2018, assembly (Loog et al. 2019)
MN071200	Gray wolf	Genbank	Saskatchewan, Canada	Sinding et al. 2018, assembly (Loog et al. 2019)
MN071195	Gray wolf	Genbank	Baffin Island, Canada	Sinding et al. 2018, assembly (Loog et al. 2019)
MN071194	Gray wolf	Genbank	Baffin Island, Canada	Sinding et al. 2018, assembly (Loog et al. 2019)
MN071197/SRR8049197	Gray wolf	Genbank	Ellesmere Island, Canada	Sinding et al. 2018, assembly (Loog et al. 2019)
MN071202	Gray wolf	Genbank	Victoria Island, Canada	Sinding et al. 2018, assembly (Loog et al. 2019)
MN071190	Gray wolf	Genbank	Alaska, USA	Sinding et al. 2018, assembly (Loog et al. 2019)
MN071191	Gray wolf	Genbank	St Lawrence Island, Alaska	Sinding et al. 2018, assembly (Loog et al. 2019)
KF661047	Gray wolf	Genbank	Ukraine , Kherson	Thalman.et.al.2013
KF661050	Gray wolf	Genbank	Oman	Thalman.et.al.2013
KF661051	Gray wolf	Genbank	Iran	Thalman.et.al.2013
KF661048	Gray wolf	Genbank	Italy	Thalman.et.al.2013
KF661049	Gray wolf	Genbank	Poland , Katovice	Thalman.et.al.2013, assembly (Loog et al. 2019)

KF661052	Gray wolf	Genbank	Sweden	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661040	Gray wolf	Genbank	Sweden	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661072	Gray wolf	Genbank	Alaska, USA	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661068	Gray wolf	Genbank	USA	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661069	Gray wolf	Genbank	USA	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661064	Gray wolf	Genbank	USA	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661066	Gray wolf	Genbank	Alaska USA	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661073	Gray wolf	Genbank	Alaska USA	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661057	Gray wolf	Genbank	Canada	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661071	Gray wolf	Genbank	Alaska USA	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661059	Gray wolf	Genbank	Canada	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661074	Gray wolf	Genbank	Canada	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661058	Gray wolf	Genbank	Alaska, USA	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661075	Gray wolf	Genbank	Canada	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661067	Gray wolf	Genbank	Alaska, USA	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661070	Gray wolf	Genbank	USA	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661077	Gray wolf	Genbank	Canada	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661063	Gray wolf	Genbank	Western Canada	Thalman.et.al.2013, assembly (Loog et al. 2019)

KF661062	Gray wolf	Genbank	Northern Canada	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661056	Gray wolf	Genbank	Canada	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661076	Gray wolf	Genbank	Canada	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661061	Gray wolf	Genbank	Northern Canada	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661060	Gray wolf	Genbank	Mexico	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661065	Gray wolf	Genbank	Mexico	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661038	Gray wolf	Genbank	Finland , Oulu	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661039	Gray wolf	Genbank	Western Russia , Kirov	Thalman.et.al.2013, assembly (Loog et al. 2019)
KF661054	Gray wolf	Genbank	Croatia	Thalmann et al. 2013
KF661042	Gray wolf	Genbank	Israel	Thalmann et al. 2013
SRR1518518	Gray wolf	Genbank	Iran	vonHoldt et al 2016
SRR5168998	Gray wolf	SRA	Inner Mongolia, China	vonHoldt et al. 2016
SRR2827603	Gray wolf	SRA	Shanxi, China	Wang et al. 2016
SRR2827615	Gray wolf	SRA	Shanxi, China	Wang et al. 2016
SRR2827611	Gray wolf	SRA	Xingjiang, China	Wang et al. 2016
SRR2827601	Gray wolf	SRA	Xingjiang, China	Wang et al. 2016
SRR2827609	Gray wolf	Genbank	Xingjiang, China	Wang et al. 2016
SRR2827612	Gray wolf	SRA	Inner Mongolia, China	Wang et al. 2016
KC896375	Gray wolf	Genbank	Mongolia	Zhang et al. 2013.
SRR7107906	Gray wolf	SRA	Luobulingka Zoo, Tibet (wild born), China	Zhang et al. 2014
KC461238	Gray wolf	Genbank	Altai in Xinjiang, China	Zhang et al. 2014
KF573616	Gray wolf	Genbank	captured in mountainous southwest of Lhasa in Tibet, China. Alt at 3,600m	Zhang et al. Direct submission

Table S1.3. Gene flow results using D statistic analyses in Admixtools (Patterson et al. 2012). We used a 5kb jackknife block size for the autosomal regions and 10kb jackknife block size for the X chromosome. Each test consisted of 18 individuals within the following populations: 4 North American wolves, 1 Iran wolf, 2 Qinghai wolves, 2 Ladakh wolves, 6 East Asian wolves, 2 Tibetan wolves, 2 Arabian wolves, 7 Central Asian wolves, 4 Indian wolves and 1 Dhole as an outgroup. P1 is North American wolf and O refers to the outgroup as Dhole, which the D statistic was calculated as (P1, P2;P3, O). A Z-score above 3 or below -3 indicates significance.

Genomic region	P2	P3	D statistic	SD	Z score	BABA	ABBA	nsnps
Autosome	Arabian	Indian	-0.0034	0.000125	-26.756	326053	397987	21547218
	East Asian	Indian	-0.0019	0.000071	-26.276	315298	355582	21547958
	Central Asian	Indian	-0.002546	0.000075	-33.942	315762	370616	21547958
	Ladakh	Indian	-0.004116	0.000191	-21.583	319344	408032	21544963
	Qinghai	Indian	-0.001647	0.000141	-11.706	335989	371474	21547117
	Tibetan	Indian	-0.001726	0.00019	-9.076	345464	382648	21541587
	Iran	Indian	-0.004736	0.000155	-30.589	310616	412521	21516367
	East Asian	Tibetan	-0.002895	0.000105	-27.651	325799	388166	21541587
	Central Asian	Tibetan	-0.001899	0.000091	-20.956	337848	378759	21541587
	Arabian	Tibetan	0.000386	0.000127	3.049	369362	361044	21540847
	Indian	Tibetan	-0.000742	0.000133	-5.598	366665	382648	21541587
	Iran	Tibetan	-0.000583	0.000146	-4.005	356467	369008	21510093
	East Asian	Ladakh	-0.002356	0.000079	-29.76	325718	376467	21544963
	Central Asian	Ladakh	-0.002016	0.000079	-25.468	333592	377037	21544963
	Arabian	Ladakh	-0.000887	0.000111	-7.992	357874	376978	21544223
	Indian	Ladakh	-0.002556	0.000191	-13.414	352958	408032	21544963
	Iran	Ladakh	-0.001885	0.00012	-15.682	345018	385578	21513379
	Tibetan	Ladakh	-0.012179	0.000471	-25.883	292102	554428	21538592
	East Asian	Qinghai	-0.003267	0.000098	-33.372	330566	400959	21547117
	Central Asian	Qinghai	-0.001764	0.000088	-19.954	346278	384290	21547117
	Arabian	Qinghai	0.000803	0.000119	6.772	379962	362669	21546377
	Indian	Qinghai	0.000803	0.000119	6.772	379962	362669	21546377
	Iran	Qinghai	-0.000045	0.00013	-0.342	367972	368932	21515553
	Tibetan	Qinghai	-0.020697	0.000472	-43.87	249657	695485	21540746
X chrom.	Arabian	Indian	-0.006996	0.002580	-2.712	11821	19963	1163773
	East Asian	Indian	-0.000826	0.000502	-1.646	11948	12909	1163799
	Central Asian	Indian	-0.005003	0.002009	-2.490	11067	16889	1163799
	Ladakh	Indian	-0.003342	0.001523	-2.195	12624	16490	1156627
	Qinghai	Indian	0.000415	0.001409	0.295	15478	14995	1163777
	Tibetan	Indian	0.001612	0.001576	1.023	16387	14511	1163597
	Iran	Indian	-0.008070	0.002645	-3.051	11026	20410	1162891
	East Asian	Tibetan	-0.003629	0.000767	-4.732	11436	15659	1163597
	Central Asian	Tibetan	0.000850	0.001605	0.530	14030	13041	1163597
	Arabian	Tibetan	0.004159	0.002274	1.829	16780	11941	1163571
	Indian	Tibetan	0.003131	0.002045	1.531	18155	14511	1163597
	Iran	Tibetan	0.003257	0.002364	1.378	16091	12304	1162690
	East Asian	Ladakh	-0.001190	0.000492	-2.422	12393	13770	1156627
	Central Asian	Ladakh	-0.001758	0.000400	-4.393	12847	14881	1156627
	Arabian	Ladakh	-0.000889	0.000709	-1.254	14761	15790	1156601
	Indian	Ladakh	0.001732	0.002047	0.846	18494	16490	1156627
	Iran	Ladakh	-0.001712	0.000734	-2.332	14085	16063	1155721
	Tibetan	Ladakh	-0.010615	0.002651	-4.004	14452	26728	1156425
	East Asian	Qinghai	-0.004000	0.000812	-4.924	11614	16269	1163777
	Central Asian	Qinghai	-0.000338	0.000720	-0.470	13813	14206	1163777

	Arabian	Qinghai	0.002583	0.001023	2.524	16426	13420	1163751
	Indian	Qinghai	0.003148	0.001360	2.315	18659	14995	1163777
	Iran	Qinghai	0.001801	0.001111	1.621	15765	13671	1162869
	Tibetan	Qinghai	-0.033522	0.002209	-15.178	6998	46004	1163575
Chr 1	Arabian	Indian	-0.003442	0.000363	-9.483	24736	30540	1685941
	East Asian	Indian	-0.001861	0.000190	-9.819	23724	26863	1686046
	Central Asian	Indian	-0.002665	0.000199	-13.394	23706	28198	1686046
	Ladakh	Indian	-0.005642	0.000622	-9.066	23268	32779	1685772
	Qinghai	Indian	-0.001471	0.000410	-3.589	25466	27946	1685983
	Tibetan	Indian	-0.002343	0.000527	-4.446	25887	29835	1685479
	Iran	Indian	-0.004547	0.000519	-8.758	23360	31014	1683508
	East Asian	Tibetan	-0.002470	0.000264	-9.359	24978	29142	1685479
	Central Asian	Tibetan	-0.001861	0.000199	-9.353	25708	28846	1685479
	Arabian	Tibetan	0.000513	0.000500	1.026	28581	27716	1685374
	Indian	Tibetan	-0.000718	0.000546	-1.315	28626	29835	1685479
	Iran	Tibetan	-0.000834	0.000423	-1.972	26941	28345	1682946
	East Asian	Ladakh	-0.002020	0.000222	-9.099	24976	28382	1685772
	Central Asian	Ladakh	-0.002087	0.000221	-9.442	25407	28926	1685772
	Arabian	Ladakh	-0.001400	0.000330	-4.235	27340	29699	1685667
	Indian	Ladakh	-0.003308	0.000719	-4.599	27203	32779	1685772
	Iran	Ladakh	-0.002312	0.000329	-7.022	26014	29905	1683234
	Tibetan	Ladakh	-0.008326	0.001417	-5.875	24392	38422	1685205
	East Asian	Qinghai	-0.003041	0.000245	-12.403	25301	30428	1685983
	Central Asian	Qinghai	-0.001684	0.000187	-9.001	26490	29329	1685983
	Arabian	Qinghai	0.001383	0.000395	3.502	29680	27347	1685878
	Indian	Qinghai	0.001611	0.000387	4.157	30661	27946	1685983
	Tibetan	Qinghai	-0.018740	0.001549	-12.098	19903	51488	1685416

Table S1.4. f3 statistic results using 30 gray wolf genomes. We calculated the f3 statistic using threepop in Treemix to test for introgression across the autosomes (Pickrell & Pritchard, 2012). The f3 statistic assesses whether the target population C is admixed between two source populations (A and B) using a product of the allele frequency difference between population C to population A and B. For the f3 statistic analyses using gray wolf populations, we selected a minor allele count of at least 3 to remove potential sequencing errors. We evaluated the f3 statistics using all combinations of 8 gray wolf populations: Arabian wolf, Iranian wolf, Indian wolf, Tibetan wolf, Qinghai wolf, Ladakh wolf, Central Asian wolf, East Asian wolf, and North American wolf. The Z-score for each population calculation was assessed using block jackknifing to accommodate the non-independence between loci, which we used a jackknife block size of 5 (-k 500) for the autosomes.

Topology combination (C;A,B)	f3statistic	sd	Zscore
Qinghaiwolf;EastAsianwolfTibetanwolf	-0.0199112	0.0001764	-112.878
Qinghaiwolf;TibetanwolfCentralAsianwolf	-0.0186617	0.00017602	-106.021
Qinghaiwolf;NorthAmericanwolfTibetanwolf	-0.0190279	0.00018352	-103.685
Qinghaiwolf;ArabianTibetanwolf	-0.0180488	0.00019297	-93.5309
Ladakhwolf;IndianwolfTibetanwolf	-0.0177171	0.00019784	-89.5536
Qinghaiwolf;IranTibetanwolf	-0.0177683	0.00020162	-88.1267

Ladakhwolf;ArabianTibetanwolf	-0.0159547	0.00019371	-82.3629
Qinghaiwolf;IndianwolfTibetanwolf	-0.0155508	0.00019307	-80.5434
Ladakhwolf;IranTibetanwolf	-0.0159219	0.00020673	-77.0168
Ladakhwolf;TibetanwolfCentralAsianwolf	-0.0126073	0.00017839	-70.6734
Ladakhwolf;NorthAmericanwolfTibetanwolf	-0.0121917	0.00018941	-64.3688
Ladakhwolf;EastAsianwolfTibetanwolf	-0.0107266	0.00018371	-58.3905
Ladakhwolf;QinghaiwolfIndianwolf	-0.0084333	0.00017117	-49.2686
Qinghaiwolf;LadakhwolfTibetanwolf	-0.0062671	0.00018527	-33.8273
Ladakhwolf;QinghaiwolfArabian	-0.004173	0.00017025	-24.5111
Ladakhwolf;IranQinghaiwolf	-0.0044207	0.00018043	-24.5007
Iran;ArabianIndianwolf	-0.0043241	0.00022612	-19.1231
Iran;ArabianCentralAsianwolf	-0.0029865	0.00018788	-15.8952
Iran;LadakhwolfArabian	-0.0032087	0.00020813	-15.4168
Iran;ArabianEastAsianwolf	-0.0027239	0.00019206	-14.1824
CentralAsianwolf;IranEastAsianwolf	-0.0011011	7.78E-05	-14.1536
Iran;ArabianTibetanwolf	-0.0032415	0.00023292	-13.9167
Iran;QinghaiwolfArabian	-0.002961	0.00021671	-13.6631
CentralAsianwolf;IndianwolfEastAsianwolf	-0.0009829	7.60E-05	-12.9353
CentralAsianwolf;ArabianEastAsianwolf	-0.0008386	6.94E-05	-12.0851
Iran;NorthAmericanwolfArabian	-0.0022073	0.00019937	-11.0714
Iran;NorthAmericanwolfIndianwolf	-0.0016853	0.00020953	-8.04358
Iran;IndianwolfCentralAsianwolf	-0.0007506	0.00020116	-3.73114
Iran;IndianwolfEastAsianwolf	-0.0006323	0.00020582	-3.07238
CentralAsianwolf;QinghaiwolfArabian	-0.0003318	0.00011513	-2.8818
CentralAsianwolf;IranQinghaiwolf	-0.0003573	0.000136	-2.627
Ladakhwolf;QinghaiwolfCentralAsianwolf	-0.0002127	0.00015367	-1.38395
CentralAsianwolf;NorthAmericanwolfIndianwolf	-7.27E-07	9.06E-05	-0.0080222
CentralAsianwolf;ArabianTibetanwolf	0.00028111	0.00013702	2.05158
CentralAsianwolf;IranTibetanwolf	0.00053615	0.00016585	3.23277
Ladakhwolf;QinghaiwolfNorthAmericanwolf	0.00056907	0.00016608	3.42651
CentralAsianwolf;IranNorthAmericanwolf	0.00093404	9.70E-05	9.632
Iran;QinghaiwolfIndianwolf	0.0030571	0.00024307	12.5768
CentralAsianwolf;NorthAmericanwolfTibetanwolf	0.00173877	0.00011482	15.1434
CentralAsianwolf;LadakhwolfNorthAmericanwolf	0.0013232	8.28E-05	15.9731
Ladakhwolf;QinghaiwolfEastAsianwolf	0.00291752	0.00016195	18.0146
Iran;IndianwolfTibetanwolf	0.00527456	0.00026945	19.575
CentralAsianwolf;NorthAmericanwolfArabian	0.00171322	8.53E-05	20.0754
CentralAsianwolf;QinghaiwolfNorthAmericanwolf	0.00210494	9.58E-05	21.9652
CentralAsianwolf;LadakhwolfEastAsianwolf	0.00163278	7.19E-05	22.7096
CentralAsianwolf;QinghaiwolfIndianwolf	0.00345039	0.00013101	26.3375

Iran;LadakhwolfIndianwolf	0.00706969	0.00024849	28.4504
Iran;LadakhwolfCentralAsianwolf	0.00565187	0.00019489	29.0001
Iran;LadakhwolfNorthAmericanwolf	0.00604103	0.00020537	29.4156
CentralAsianwolf;IranLadakhwolf	0.00385078	0.00012611	30.5353
CentralAsianwolf;LadakhwolfArabian	0.00362852	0.00010582	34.2897
CentralAsianwolf;EastAsianwolfTibetanwolf	0.00351345	0.00010169	34.5492
CentralAsianwolf;IndianwolfTibetanwolf	0.00656128	0.00016409	39.9859
Iran;TibetanwolfCentralAsianwolf	0.0089665	0.0002217	40.4446
Iran;NorthAmericanwolfTibetanwolf	0.00977123	0.00023993	40.7249
Iran;LadakhwolfEastAsianwolf	0.00838575	0.00020517	40.8717
Iran;NorthAmericanwolfCentralAsianwolf	0.00856861	0.00019011	45.0718
CentralAsianwolf;NorthAmericanwolfEastAsianwolf	0.00282882	6.14E-05	46.0508
Iran;QinghaiwolfCentralAsianwolf	0.00985994	0.000206	47.8647
Iran;QinghaiwolfNorthAmericanwolf	0.0110308	0.00022292	49.4828
CentralAsianwolf;QinghaiwolfEastAsianwolf	0.00476298	8.72E-05	54.5948
Arabian;IranTibetanwolf	0.013327	0.0002414	55.2075
Iran;EastAsianwolfCentralAsianwolf	0.0106038	0.00018998	55.8152
Iran;EastAsianwolfTibetanwolf	0.0135811	0.00023827	56.9982
Ladakhwolf;IndianwolfEastAsianwolf	0.0101564	0.00017306	58.6884
Arabian;IranQinghaiwolf	0.0130465	0.00022199	58.7706
CentralAsianwolf;IranIndianwolf	0.0102532	0.00016829	60.9266
Arabian;IranNorthAmericanwolf	0.0122928	0.00020037	61.351
Arabian;IranIndianwolf	0.0144096	0.00023354	61.7008
Iran;NorthAmericanwolfEastAsianwolf	0.0124985	0.00020156	62.0099
Ladakhwolf;NorthAmericanwolfIndianwolf	0.0114481	0.00018428	62.1229
Arabian;IranLadakhwolf	0.0132942	0.00021319	62.3579
Qinghaiwolf;LadakhwolfEastAsianwolf	0.0138979	0.000213	65.2497
CentralAsianwolf;ArabianIndianwolf	0.0089156	0.00013489	66.0948
Arabian;IranEastAsianwolf	0.0128094	0.00019152	66.8827
Arabian;IranCentralAsianwolf	0.013072	0.00018634	70.1528
Iran;QinghaiwolfEastAsianwolf	0.015724	0.00022316	70.46
Ladakhwolf;IndianwolfCentralAsianwolf	0.012772	0.00017638	72.4142
Qinghaiwolf;LadakhwolfNorthAmericanwolf	0.0162464	0.00022116	73.4596
Arabian;NorthAmericanwolfIndianwolf	0.0149315	0.00019208	77.7354
Qinghaiwolf;LadakhwolfCentralAsianwolf	0.0170281	0.00021635	78.7051
CentralAsianwolf;IranArabian	0.0124891	0.00015839	78.8493
Tibetanwolf;LadakhwolfQinghaiwolf	0.0171215	0.00021244	80.5952
Ladakhwolf;IranIndianwolf	0.0191745	0.00023684	80.9595
Ladakhwolf;QinghaiwolfTibetanwolf	0.0230825	0.00028129	82.0584
CentralAsianwolf;LadakhwolfIndianwolf	0.011671	0.00014147	82.4974

Ladakhwolf;ArabianIndianwolf	0.0180591	0.00021773	82.9446
Qinghaiwolf;IranLadakhwolf	0.0212362	0.00024806	85.6087
Qinghaiwolf;LadakhwolfArabian	0.0209884	0.00024224	86.6425
Arabian;IndianwolfEastAsianwolf	0.0165011	0.00018719	88.1528
Arabian;IndianwolfCentralAsianwolf	0.0166455	0.00018187	91.5266
Arabian;QinghaiwolfIndianwolf	0.0204276	0.0002195	93.063
Arabian;IndianwolfTibetanwolf	0.0229256	0.0002427	94.4611
Tibetanwolf;QinghaiwolfIndianwolf	0.0264052	0.00027151	97.2531
Ladakhwolf;IranEastAsianwolf	0.0178584	0.00018359	97.2711
NorthAmericanwolf;IranTibetanwolf	0.0204252	0.0002044	99.9269
Tibetanwolf;IranQinghaiwolf	0.0286227	0.00028593	100.104
Qinghaiwolf;LadakhwolfIndianwolf	0.0252488	0.00024929	101.283
Ladakhwolf;IranNorthAmericanwolf	0.0202032	0.00019849	101.782
Tibetanwolf;QinghaiwolfArabian	0.0289032	0.00028172	102.596
Iran;LadakhwolfQinghaiwolf	0.0306649	0.0002971	103.214
Tibetanwolf;QinghaiwolfNorthAmericanwolf	0.0298823	0.00027836	107.352
Ladakhwolf;ArabianEastAsianwolf	0.0183432	0.00017035	107.681
Tibetanwolf;QinghaiwolfCentralAsianwolf	0.0295161	0.00027303	108.104
NorthAmericanwolf;ArabianTibetanwolf	0.019391	0.00017916	108.234
NorthAmericanwolf;IranQinghaiwolf	0.0191656	0.00017475	109.675
Arabian;LadakhwolfIndianwolf	0.024688	0.00022467	109.887
Tibetanwolf;QinghaiwolfEastAsianwolf	0.0307656	0.00027673	111.178
Ladakhwolf;IranCentralAsianwolf	0.0205923	0.00018519	111.197
Arabian;LadakhwolfNorthAmericanwolf	0.0215425	0.0001905	113.085
Arabian;NorthAmericanwolfTibetanwolf	0.0253055	0.00021808	116.039
Ladakhwolf;NorthAmericanwolfArabian	0.0212046	0.00018152	116.819
NorthAmericanwolf;QinghaiwolfArabian	0.0184119	0.00015694	117.317
Ladakhwolf;ArabianCentralAsianwolf	0.0208146	0.00017243	120.712
Iran;LadakhwolfTibetanwolf	0.0421661	0.00034362	122.71
Arabian;LadakhwolfCentralAsianwolf	0.0219325	0.00017751	123.558
Qinghaiwolf;IndianwolfEastAsianwolf	0.0324876	0.00026148	124.244
Ladakhwolf;IranArabian	0.0294529	0.00023532	125.164
Arabian;TibetanwolfCentralAsianwolf	0.0252799	0.00020021	126.269
CentralAsianwolf;LadakhwolfQinghaiwolf	0.0246558	0.00019368	127.3
Qinghaiwolf;IranEastAsianwolf	0.0361771	0.00028267	127.985
Arabian;QinghaiwolfNorthAmericanwolf	0.0262846	0.00020391	128.905
Arabian;LadakhwolfEastAsianwolf	0.0244039	0.00018812	129.723
NorthAmericanwolf;IndianwolfTibetanwolf	0.0273851	0.00020883	131.137
Qinghaiwolf;ArabianEastAsianwolf	0.0364141	0.0002763	131.791
Qinghaiwolf;NorthAmericanwolfIndianwolf	0.0361278	0.0002725	132.578

NorthAmericanwolf;IranEastAsianwolf	0.0176979	0.00013161	134.477
Arabian;NorthAmericanwolfCentralAsianwolf	0.0238478	0.00017722	134.567
Arabian;EastAsianwolfTibetanwolf	0.029632	0.00021634	136.97
NorthAmericanwolf;QinghaiwolfIndianwolf	0.023908	0.00017425	137.203
Arabian;QinghaiwolfCentralAsianwolf	0.0258928	0.00018841	137.428
Qinghaiwolf;IranNorthAmericanwolf	0.0408703	0.00029675	137.724
NorthAmericanwolf;IranLadakhwolf	0.0241554	0.00017477	138.216
NorthAmericanwolf;ArabianEastAsianwolf	0.0171813	0.00012314	139.527
Qinghaiwolf;EastAsianwolfCentralAsianwolf	0.0369209	0.00026392	139.894
Qinghaiwolf;NorthAmericanwolfEastAsianwolf	0.0376448	0.00026902	139.935
NorthAmericanwolf;EastAsianwolfTibetanwolf	0.0215077	0.00015226	141.255
Qinghaiwolf;IndianwolfCentralAsianwolf	0.0382335	0.00026689	143.253
Qinghaiwolf;NorthAmericanwolfArabian	0.0416239	0.00028632	145.375
NorthAmericanwolf;TibetanwolfCentralAsianwolf	0.0208231	0.00014316	145.455
Arabian;NorthAmericanwolfEastAsianwolf	0.0275152	0.00018897	145.608
Ladakhwolf;NorthAmericanwolfCentralAsianwolf	0.0231199	0.00015838	145.977
Qinghaiwolf;IranCentralAsianwolf	0.0420412	0.0002875	146.229
NorthAmericanwolf;LadakhwolfArabian	0.023154	0.0001577	146.825
NorthAmericanwolf;IndianwolfEastAsianwolf	0.0187509	0.00012743	147.146
Qinghaiwolf;NorthAmericanwolfCentralAsianwolf	0.0395789	0.00026869	147.306
NorthAmericanwolf;IranIndianwolf	0.0318818	0.00021441	148.697
Arabian;EastAsianwolfCentralAsianwolf	0.0263996	0.00017587	150.113
Qinghaiwolf;ArabianCentralAsianwolf	0.0420157	0.00027971	150.212
Ladakhwolf;NorthAmericanwolfEastAsianwolf	0.0243159	0.00016138	150.671
CentralAsianwolf;LadakhwolfTibetanwolf	0.0370504	0.00024558	150.869
Qinghaiwolf;IranIndianwolf	0.048844	0.00032335	151.056
Ladakhwolf;EastAsianwolfCentralAsianwolf	0.0228103	0.00015032	151.741
Arabian;QinghaiwolfEastAsianwolf	0.0314944	0.00020442	154.068
Qinghaiwolf;ArabianIndianwolf	0.0474809	0.00030242	157.001
Indianwolf;IranTibetanwolf	0.039314	0.00024936	157.657
NorthAmericanwolf;IranArabian	0.0324037	0.00020398	158.859
NorthAmericanwolf;ArabianIndianwolf	0.029765	0.00018717	159.029
NorthAmericanwolf;LadakhwolfEastAsianwolf	0.0200426	0.00012575	159.38
NorthAmericanwolf;QinghaiwolfCentralAsianwolf	0.0204569	0.00012785	160.014
NorthAmericanwolf;IranCentralAsianwolf	0.0216278	0.00013413	161.241
Tibetanwolf;IranLadakhwolf	0.0561259	0.00034302	163.623
NorthAmericanwolf;QinghaiwolfEastAsianwolf	0.0223911	0.00013674	163.744
Tibetanwolf;LadakhwolfEastAsianwolf	0.0509306	0.00031027	164.149
Tibetanwolf;LadakhwolfNorthAmericanwolf	0.0523957	0.00031669	165.448
NorthAmericanwolf;ArabianCentralAsianwolf	0.0208486	0.00012538	166.284

Qinghaiwolf;IranArabian	0.054862	0.00032953	166.487
Tibetanwolf;LadakhwolfArabian	0.0561587	0.00033522	167.526
Tibetanwolf;LadakhwolfCentralAsianwolf	0.0528113	0.0003125	168.998
Iran;QinghaiwolfTibetanwolf	0.0696694	0.00040467	172.164
Tibetanwolf;LadakhwolfIndianwolf	0.0579211	0.00033559	172.596
Indianwolf;IranLadakhwolf	0.0375189	0.00021712	172.8
NorthAmericanwolf;LadakhwolfCentralAsianwolf	0.0212387	0.00012278	172.978
NorthAmericanwolf;IndianwolfCentralAsianwolf	0.0225626	0.00013011	173.416
NorthAmericanwolf;LadakhwolfIndianwolf	0.0329104	0.00018941	173.757
Arabian;LadakhwolfQinghaiwolf	0.0469201	0.00026875	174.584
Indianwolf;IranQinghaiwolf	0.0415315	0.00023197	179.04
Indianwolf;ArabianTibetanwolf	0.0403966	0.00022142	182.447
NorthAmericanwolf;EastAsianwolfCentralAsianwolf	0.019733	0.00010734	183.836
Arabian;LadakhwolfTibetanwolf	0.0587018	0.0003149	186.413
Indianwolf;IranArabian	0.0489126	0.00025259	193.641
NorthAmericanwolf;LadakhwolfQinghaiwolf	0.0437895	0.00022584	193.898
Indianwolf;LadakhwolfArabian	0.0386342	0.00019824	194.889
CentralAsianwolf;QinghaiwolfTibetanwolf	0.0603456	0.00030868	195.498
Indianwolf;IranNorthAmericanwolf	0.0462739	0.00022879	202.258
NorthAmericanwolf;LadakhwolfTibetanwolf	0.0565503	0.00027613	204.794
Indianwolf;QinghaiwolfArabian	0.0428946	0.00020672	207.5
Indianwolf;IranEastAsianwolf	0.0452209	0.00021597	209.383
Tibetanwolf;IndianwolfEastAsianwolf	0.0788041	0.0003721	211.781
Tibetanwolf;IranEastAsianwolf	0.084711	0.00039899	212.312
Tibetanwolf;NorthAmericanwolfIndianwolf	0.0815609	0.00037934	215.006
Indianwolf;IranCentralAsianwolf	0.0453391	0.00021015	215.751
Tibetanwolf;IranNorthAmericanwolf	0.0885208	0.00040818	216.869
Tibetanwolf;IranIndianwolf	0.0930175	0.00042872	216.964
Tibetanwolf;ArabianEastAsianwolf	0.0852286	0.00039068	218.157
Tibetanwolf;IndianwolfCentralAsianwolf	0.0833004	0.00037641	221.304
Tibetanwolf;IranCentralAsianwolf	0.0893255	0.00040231	222.03
Tibetanwolf;ArabianIndianwolf	0.0919349	0.00040977	224.358
Tibetanwolf;NorthAmericanwolfArabian	0.089555	0.00039812	224.946
Indianwolf;NorthAmericanwolfTibetanwolf	0.0507706	0.00022356	227.101
Arabian;QinghaiwolfTibetanwolf	0.0859573	0.00037807	227.357
Tibetanwolf;ArabianCentralAsianwolf	0.0895806	0.00039319	227.831
Tibetanwolf;NorthAmericanwolfEastAsianwolf	0.0874383	0.00038196	228.923
Tibetanwolf;IranArabian	0.101534	0.0004426	229.404
Tibetanwolf;EastAsianwolfCentralAsianwolf	0.0863482	0.0003763	229.469
Indianwolf;LadakhwolfNorthAmericanwolf	0.0452453	0.00019606	230.769

Tibetanwolf;NorthAmericanwolfCentralAsianwolf	0.0881229	0.00037984	232.003
NorthAmericanwolf;QinghaiwolfTibetanwolf	0.0790637	0.00033544	235.703
Indianwolf;LadakhwolfTibetanwolf	0.0744105	0.00031519	236.085
Indianwolf;NorthAmericanwolfArabian	0.0483907	0.00020342	237.887
Indianwolf;LadakhwolfCentralAsianwolf	0.0439213	0.00018199	241.342
Indianwolf;ArabianEastAsianwolf	0.0468211	0.000194	241.346
Indianwolf;LadakhwolfQinghaiwolf	0.0651267	0.00026882	242.273
Indianwolf;TibetanwolfCentralAsianwolf	0.0490311	0.00020212	242.58
Indianwolf;LadakhwolfEastAsianwolf	0.046537	0.00019048	244.32
Indianwolf;ArabianCentralAsianwolf	0.0466768	0.00018887	247.133
Indianwolf;EastAsianwolfTibetanwolf	0.0535274	0.000216	247.816
Indianwolf;QinghaiwolfNorthAmericanwolf	0.0542477	0.00020865	259.99
Indianwolf;QinghaiwolfCentralAsianwolf	0.052142	0.00018883	276.13
Indianwolf;QinghaiwolfEastAsianwolf	0.0578878	0.00020324	284.828
Indianwolf;QinghaiwolfTibetanwolf	0.105926	0.00036403	290.983
Indianwolf;NorthAmericanwolfCentralAsianwolf	0.0555931	0.00018591	299.04
Indianwolf;NorthAmericanwolfEastAsianwolf	0.0594048	0.00019529	304.183
Indianwolf;EastAsianwolfCentralAsianwolf	0.0565752	0.00017751	318.722

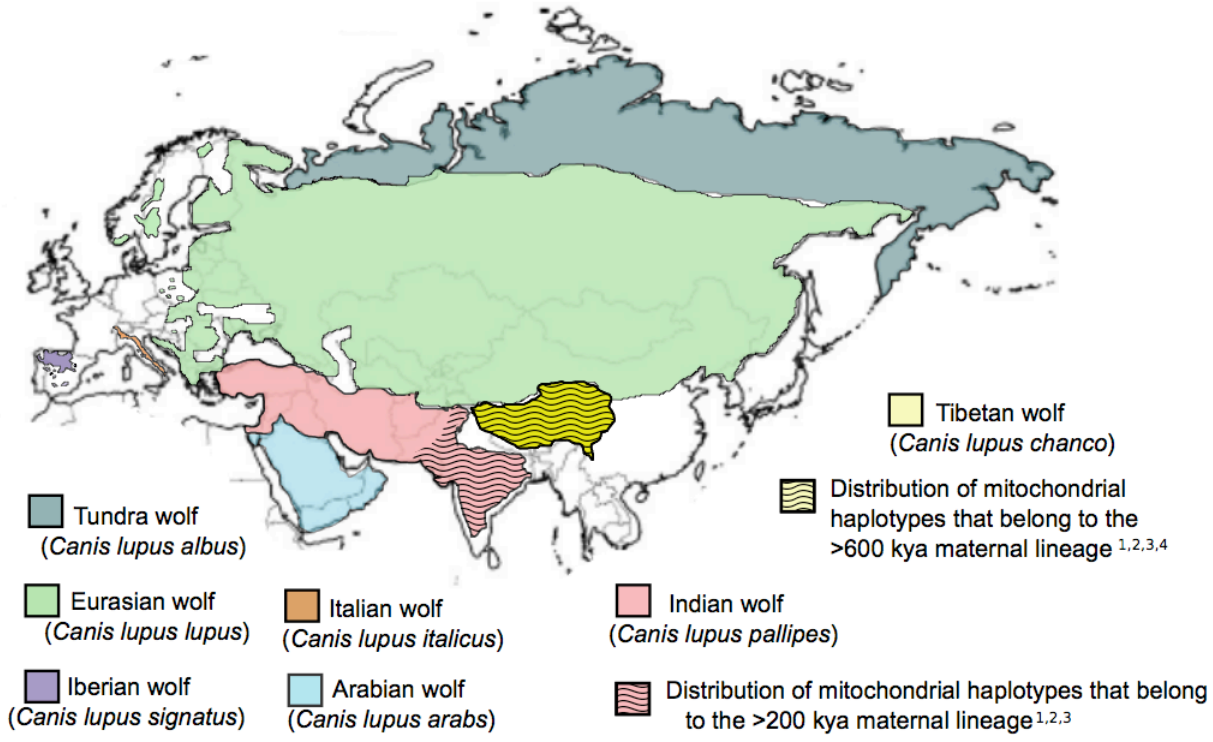


Figure S1.1. Distribution and taxonomy of gray wolves in Eurasia. The distribution of each currently recognized gray wolf subspecies in Eurasia (Castello 2018, Alvarez et al. 2019). Wavy lines indicate the distribution of evolutionarily distinct populations within the Indian wolf (*Canis lupus pallipes*) and Tibetan wolf (*Canis lupus chanco*) based on mitochondrial analyses (Sharma et al. 2004, Aggarwal et al. 2007, Esrmark et al. 2016, Werhahn et al. 2020).

Figure S1.2. Mitogenome of a previously whole-genome sequenced putative "Indian wolf" from Koln Zoo (Genbank SRA: SRR3574809; Fan et al. 2016) clusters within the Holarctic maternal wolf lineage. Due to the mitochondrial clustering of this wolf sample with Holarctic wolves and lack of specific locality data, we excluded this sample from our study. In contrast, the mitogenome of the Indian wolf that was sampled from the Natural History Museum of

Denmark (Genbank ID: MN071206; Loog et al. 2020) diverges before the Holarctic maternal wolf clade, which agrees with previous studies based from the mitochondrial D-loop region (Aggarwal et al. 2003, Sharma et al. 2004, Aggarwal et al. 2007). We constructed the phylogeny using 15,437 bp of the mitogenome in BEAST v1.10.4 (Drummond and Rambaut 2007) for 93 previously sequenced canid mitogenomes (supplementary table S2, Supplementary Materials) and rooted to the African wild dog (*Lycaon pictus*). We used the same methodology described in the main text except we used a strict clock and GTR substitution model inferred with MEGA v.10.0.5 (Kumar et al. 2018). The Bayesian posterior probability for relevant nodes of the phylogeny is provided.

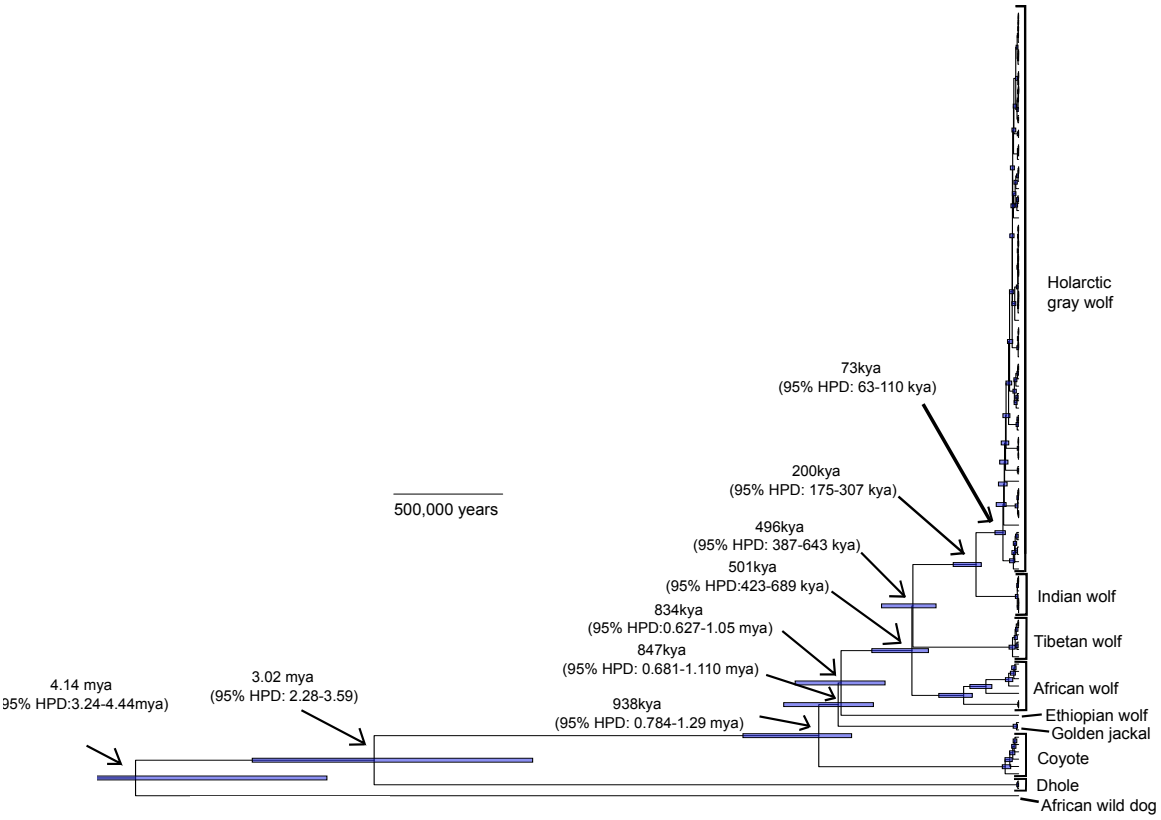


Figure S1.3. **Time-calibrated mitogenome phylogeny inferred using 3,627 bp of the 3rd codon position across the 106 mitogenomes.** We constructed the phylogenetic tree in BEAST v1.10.4 (Drummond and Rambaut 2007), rooted to the African wild dog (*Lycaon pictus*), and calibrated to a tree height of 3.9 Ma (SD = 0.3 Ma; Chavez et al. 2019). We used the substitution model TN93+G, a birth-death model, and relaxed normal clock (Loog et al. 2020). The mean divergence time with 95% HPD is shown for selected nodes.

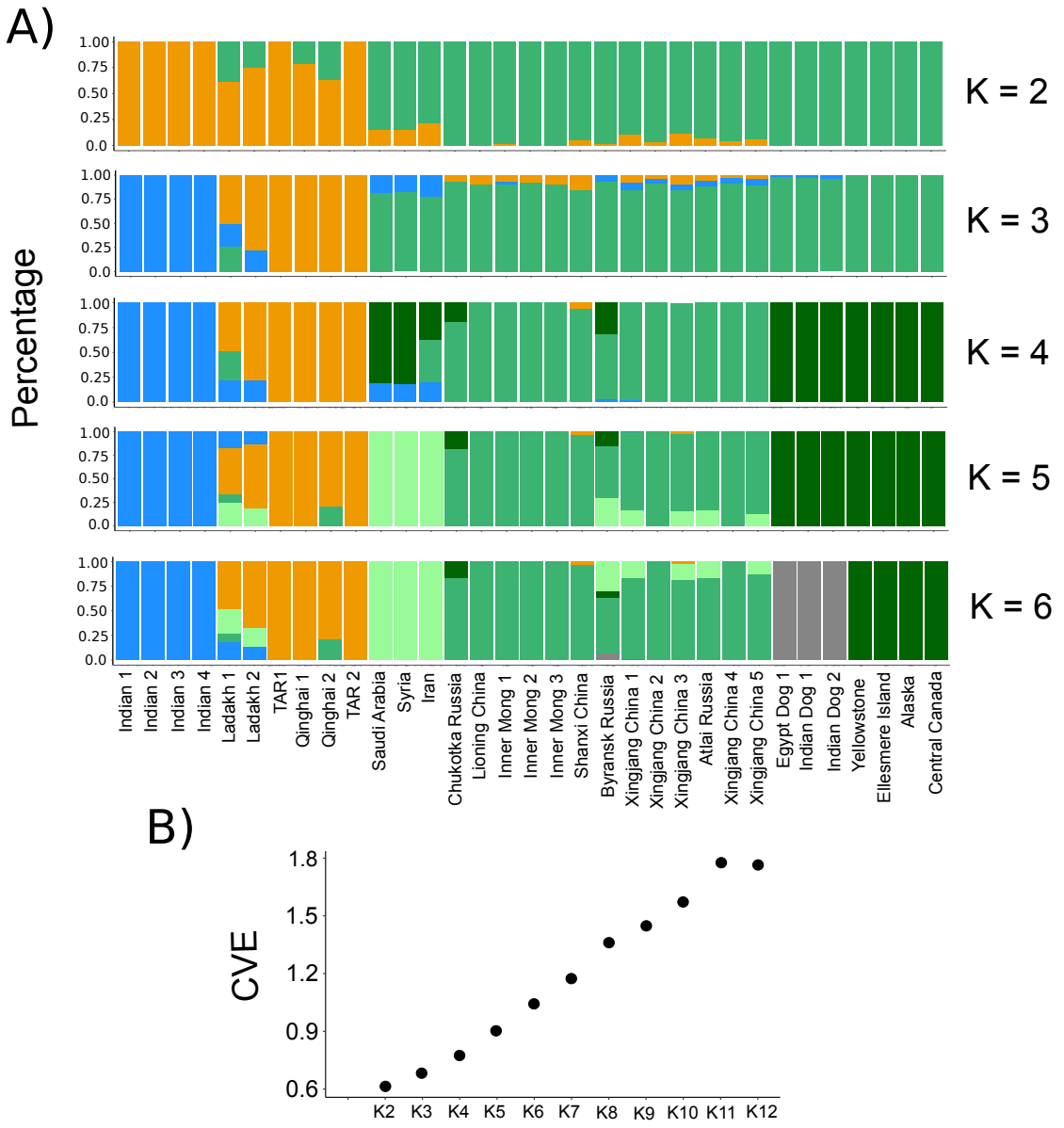


Figure S1.4. Individual ancestry proportions using K = 2 to K = 6 populations for gray wolves and dogs using Admixture version 1.3.0. (Alexander and Lange 2011) (A) We used ~3.5 million autosomal SNPs with a minimum allele count of 3, each site having at least 90% individuals present, and removed sites where linkage disequilibrium was greater than 0.5 using Plink version 1.90. Each bar represents an individual and colors within each bar represents the estimated ancestry belonging to a specific population. (B) The cross-validation error (CVE) for admixture runs with K = 2 to K = 12.

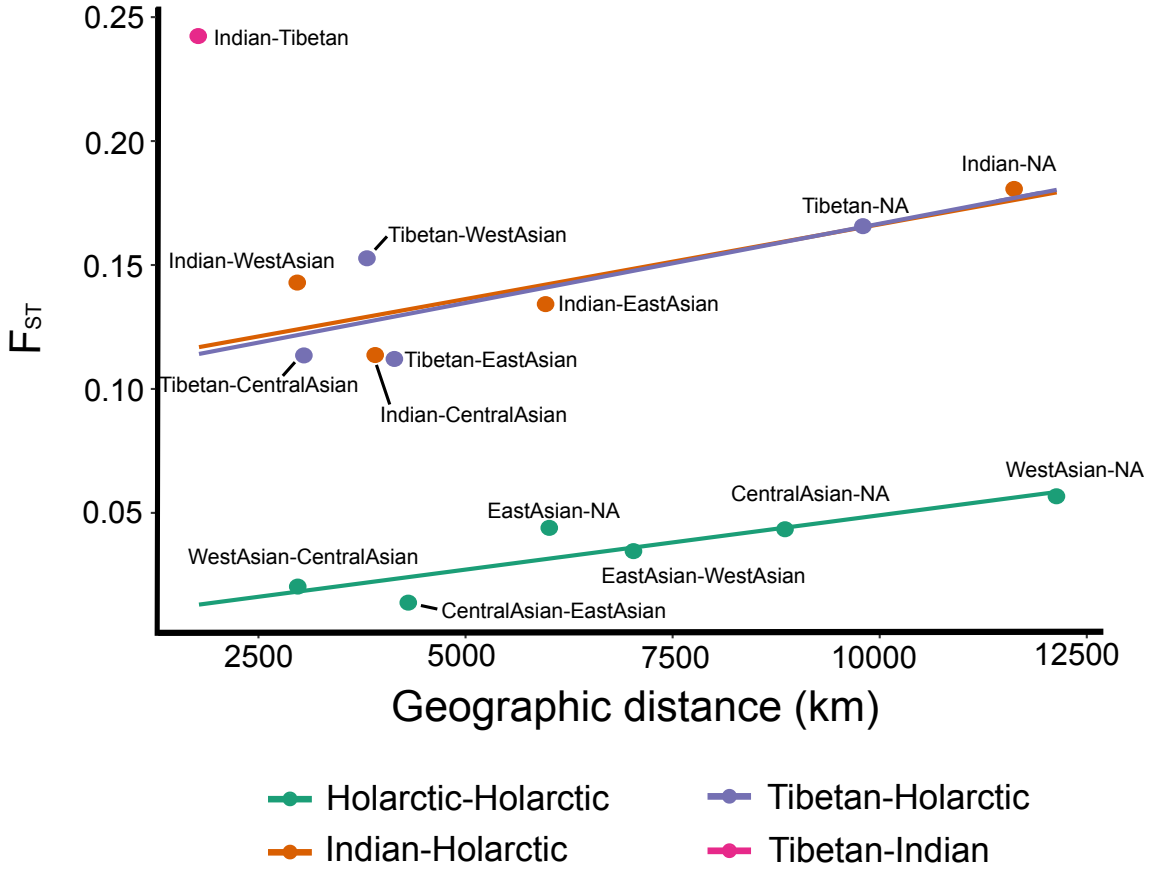


Figure S1.5. Relationship between genetic distance and geographical distance across 30 gray wolves. Genetic distance was measured as the Weir and Cockerham mean F_{ST} estimate using vcftools (Danecek et al. 2011). Geographic distance was measured between wolf populations at their center point using Google Earth Pro. If locations of wolf individuals were only given to the state or province level, we placed the location of the wolf individual at the center of its state or province. The Holarctic, Indian, and Tibetan population comparisons were fitted with a linear regression model. NA refers to North American wolf population. A Partial Mantel test using 999 permutations, which controlled for geographic distance, indicated that genetic distance ($F_{ST}/(1-F_{ST})$) was significantly greater (partial r = 0.84, p=0.001) between populations in different wolf clades (Indian-Holarctic, Tibetan-Holarctic, Tibetan-Indian) versus in the same wolf clade (Holarctic-Holarctic).

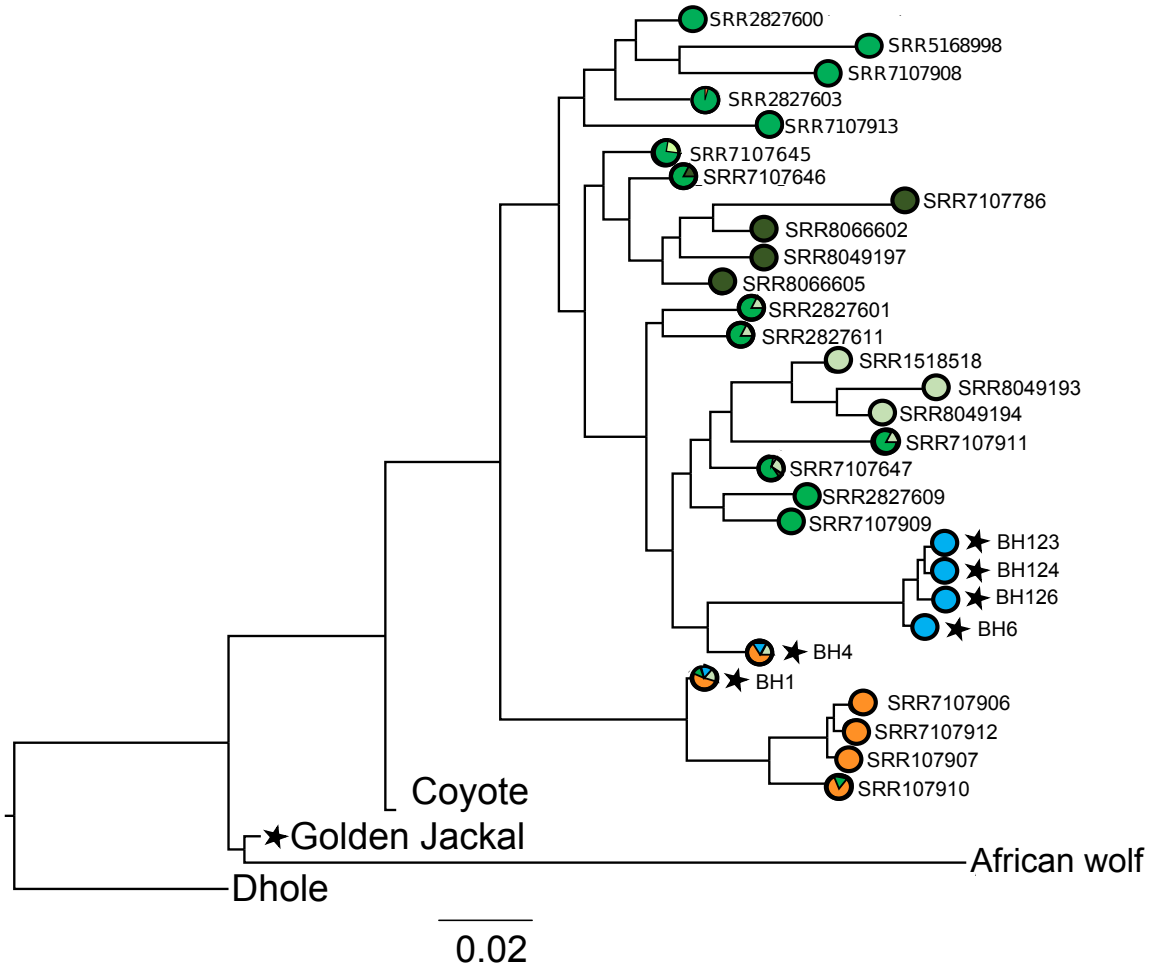


Figure S1.6. Phylogenetic tree of the X chromosome across 30 gray wolves and 4 other canid species. According to Akaike Information Criterion, we used the TVM+F+R4 substitution model, which we estimated using ModelFinder, as implemented in IQ-tree version 1.6.12 (Nguyen et al. 2014, Kalyaanamoorthy et al. 2017). We used 1,000 ultra-fast boostraps with UFBoot to infer the maximum likelihood phylogeny in IQ-tree, which resulted in ultrafast bootstrap support of >90 for all nodes. Starred samples represent newly sequenced canids included in this study. Colors associated to each sample correspond to the genetic clustering of K=6 of Fig S3.

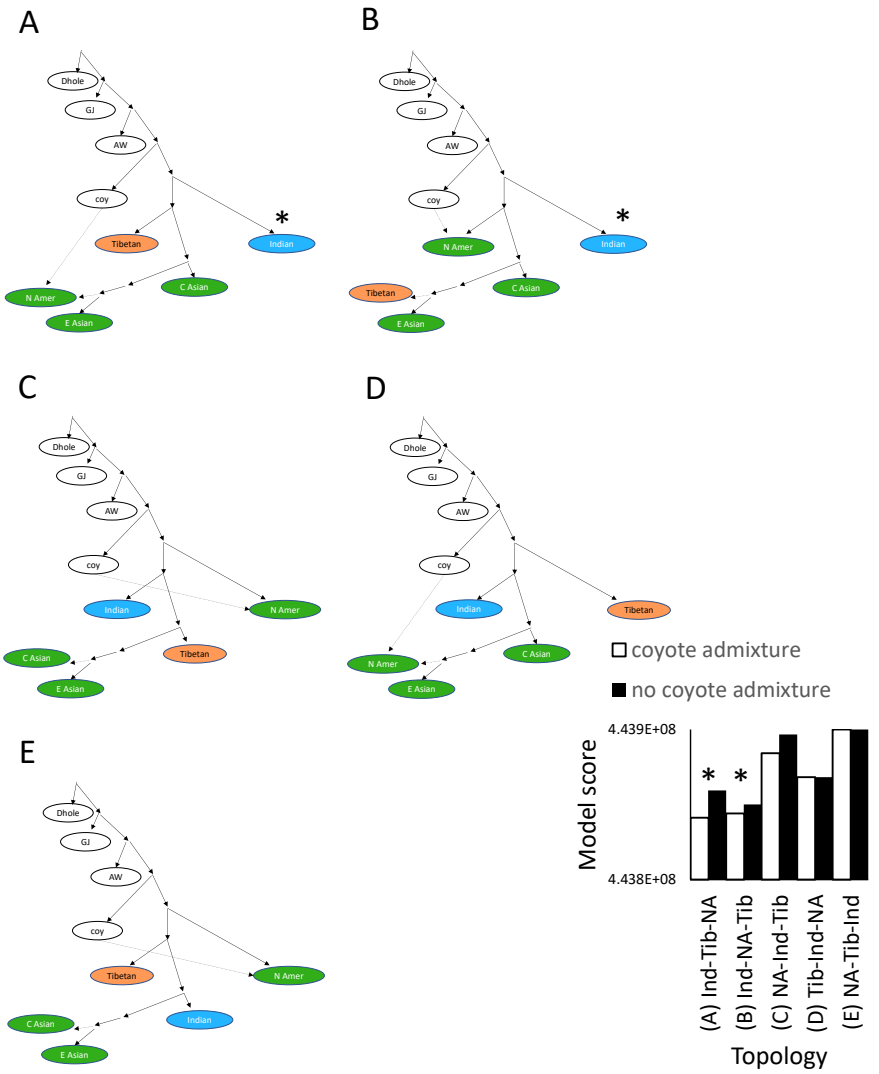


Figure S1.7. Five primary topologies tested using the least admixed (i.e., primary) gray wolf populations and including coyote introgression in North American (NA) wolves. Topologies include three populations of Holarctic gray wolves (green), Tibetan wolves (orange), and Indian wolves (blue). Models were run with and without admixture from coyotes into North American (NA) wolves. The bar graph shows model scores indicating highest support (lowest score) for the model with Indian wolf as basal, followed by the Tibetan wolf, showing the NA wolf most closely related to East Asian wolves (Ind, Tib, NA) but with introgression from coyotes. The top two models, both of which placed Indian wolf in the basal position (*), were better supported than any of the other seven models. Ladakh, Qinghai, and Arabian wolf populations were left out of this topological comparison as they were deemed to reflect admixture among various combinations of these primary populations, but were incorporated after selection of the best foundational topology (see Figure 2, main text).

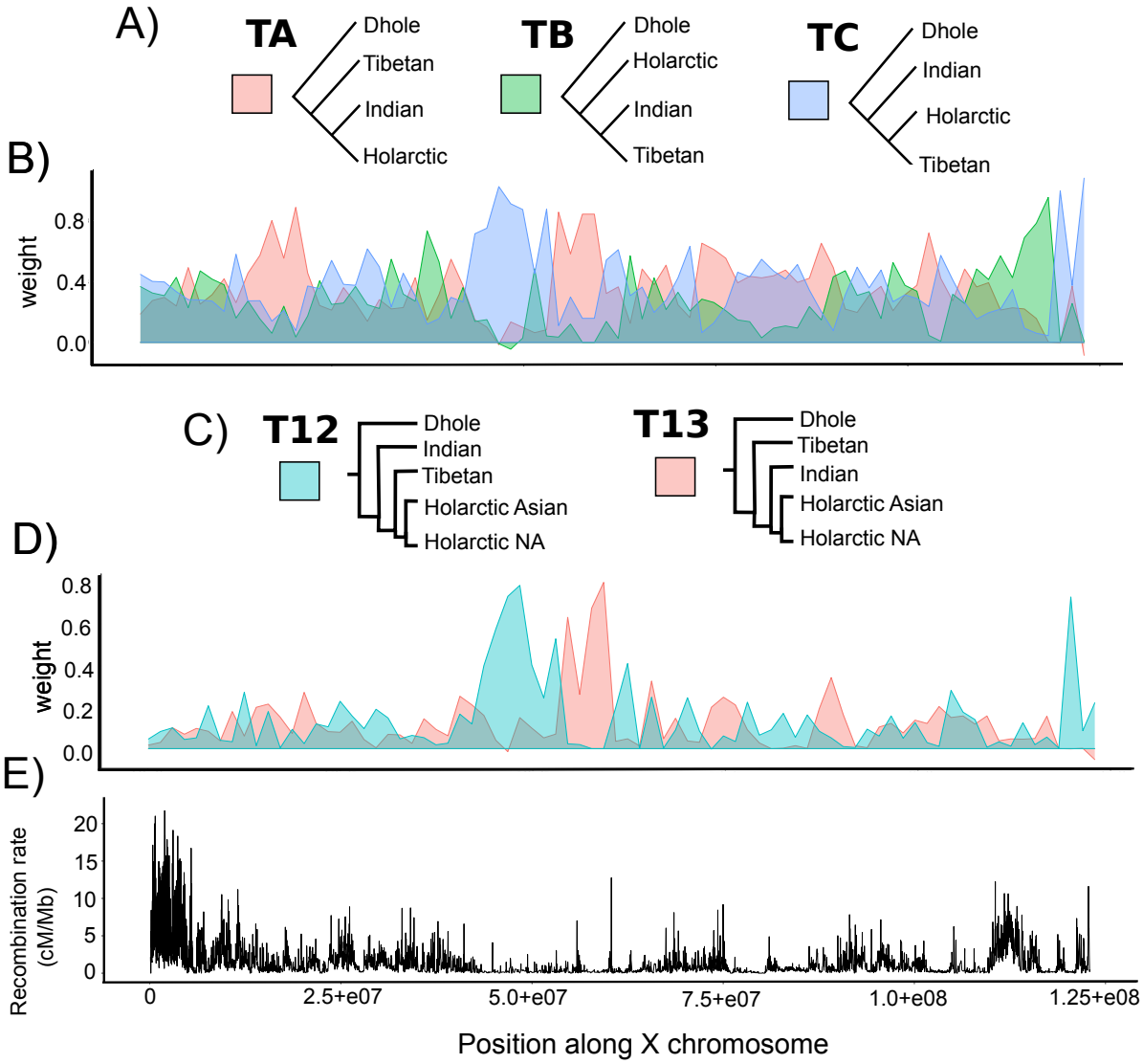


Figure S1.8. Variation of topology weights and recombination across the X chromosome. (A) Each of three possible topologies among 3 gray wolf populations (Indian, Tibetan, Holarctic) rooted to an outgroup (dhole) and (B) corresponding topology weights across the X chromosome. (C) Topology weights for the two best-supported of 15 possible topologies (T12, T13) among 4 gray wolf populations (Indian, Tibetan, Holarctic Asian, Holarctic North American) rooted to the dhole and (D) corresponding topology weights across the X chromosome. (E) Variation in recombination rate (cM/Mb) across the X chromosome. (B,D) Topology weights were estimated in Twisst (Martin and Belleghem 2017) and are represented by smoothed LOESS functions (span = 0.002). (E) The recombination rate for each point is the averaged recombination rate across 100-SNP windows that correspond to the Twisst analysis.

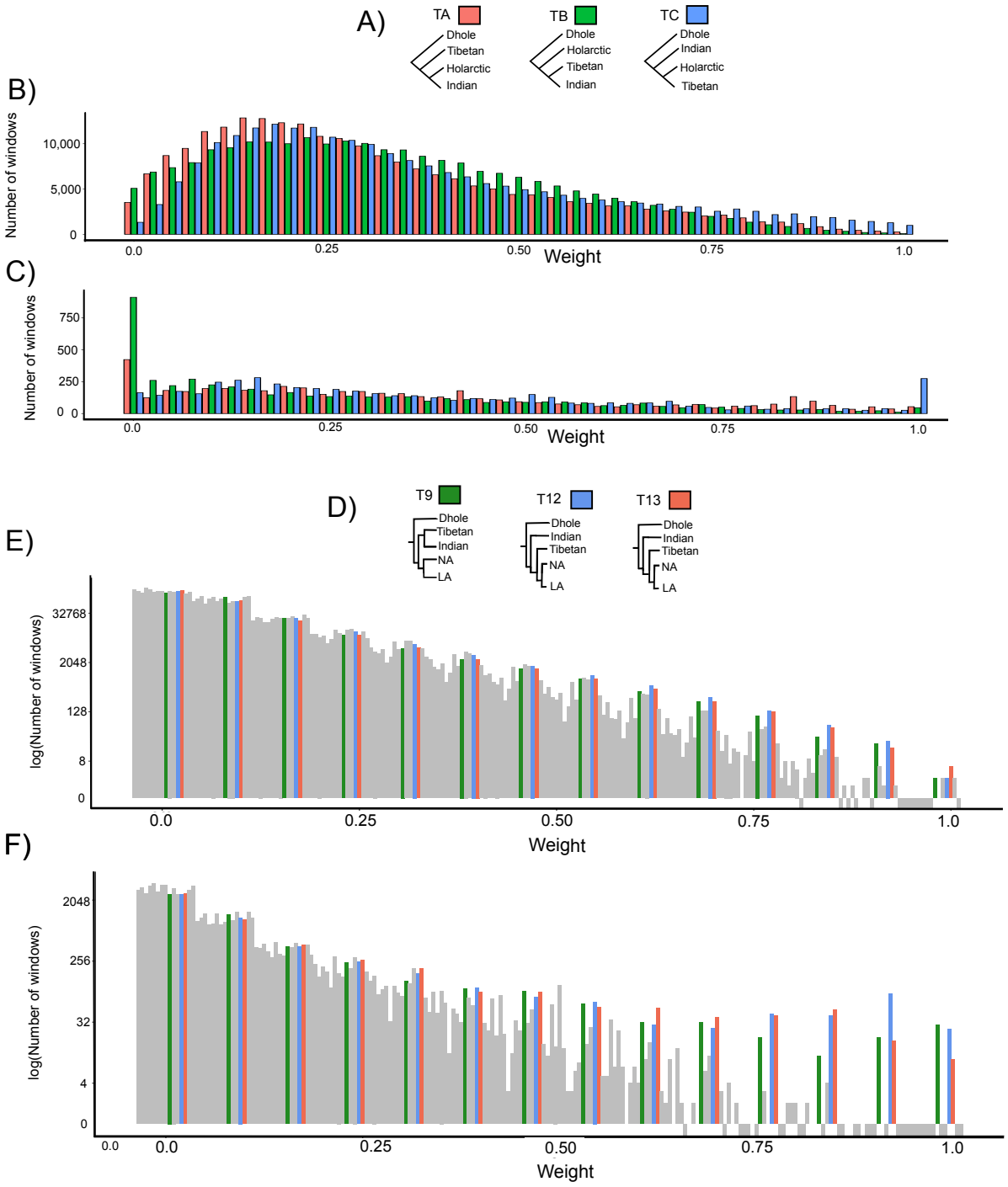


Figure S1.9. Continuous frequency distributions of specific topology weights using 3 (A-C) and 4 (D-F) wolf populations, in each case rooted to an outgroup (dhole). (A) The three possible topology weights using 3 gray wolf populations for (B) autosomes and (C) the X chromosome. (D) The top 3 topologies of 15 possible using 4 gray wolf populations (see Figure 3d in main text) and distribution of their weights across (E) autosomes and (F) the X chromosome. (B,C,E,F) Topology weights were estimated within 100-SNP windows using

Twisst (Martin and Bellegham 2017). All pairwise comparisons of the three color-coded distributions in B, C, E, and F were highly significant based on Kolmogorov-Smirnov tests ($P << 0.001$).

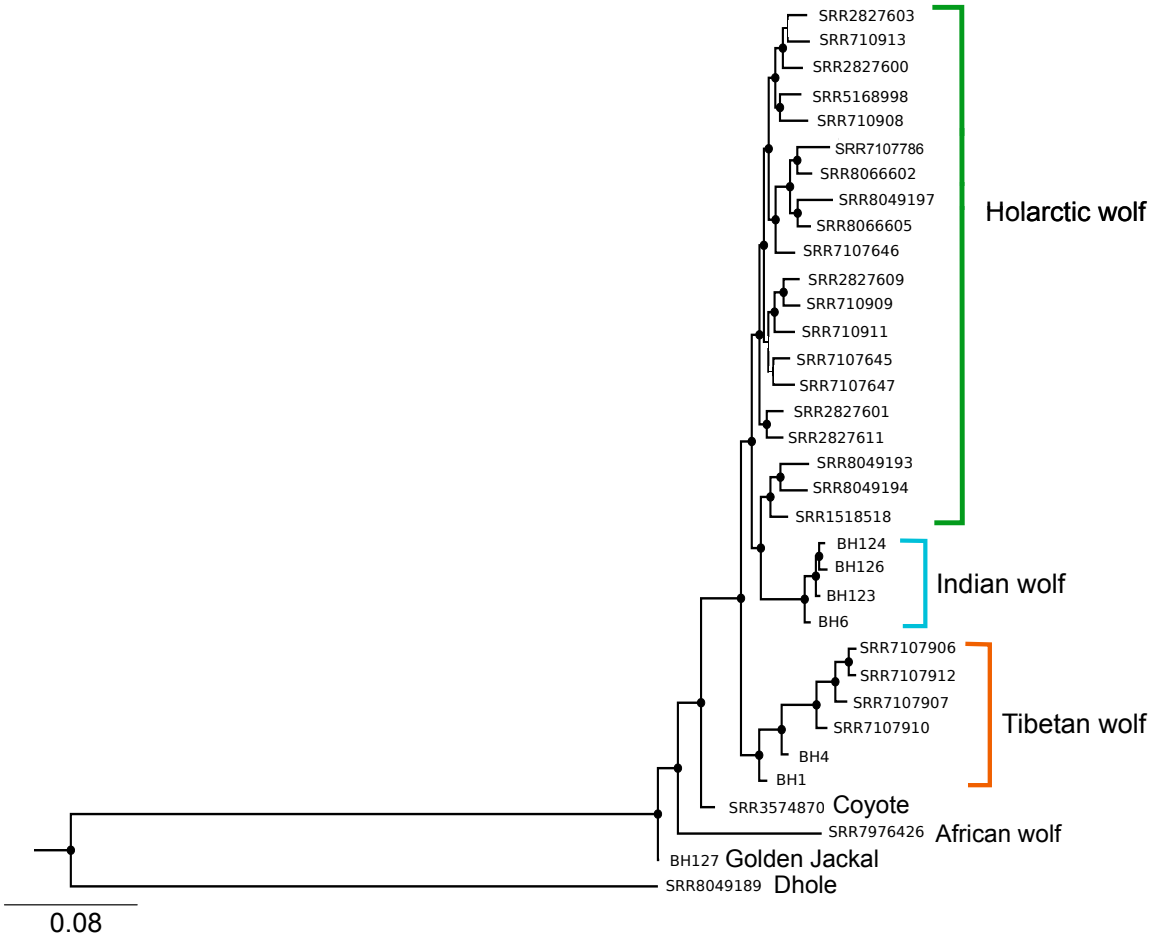


Figure S1.10. Phylogenetic tree of the autosomes using low recombination regions (<0.2 cM/Mb) across 30 gray wolves and 4 canid species. The phylogeny was constructed using IQ-tree (version 1.6.12) with using 1000 ultra-fast bootstrap approximation (UFBoot) and a TVM+F+R2 substitution model inferred using ModelFinder within IQ-tree (Nguyen et al. 2014, Kalyaanamoorthy et al. 2017). Nodes with bootstrap support values of 100 are depicted with a black circle at corresponding nodes. Sample names are provided at tip positions that correspond to table S1. We used ~5.8 million SNPs that were located within low-recombination regions of the autosomes to infer the phylogenetic tree.

7.1 References

Aggarwal RK, Kivisild T, Ramadevi J, Singh L. 2007. Mitochondrial DNA coding region sequences support the phylogenetic distinction of two Indian wolf species. *J Zool Syst Evol Res* 45:163-172.

Alvarez F, Bogdanowicz W, Campbell LAD, Godinho R, Hatlauf J, Jhala YV, Kitchener AC, Koepfli KP, Krofel M, Senn H, et al. 2019. Old world *Canis* spp. With taxonomic ambiguity: Workshop conclusions and recommendations. CIBIO, Vairao, Portugal, May 2019.

Auton A, Li YR, Kidd J, Oliveira K, Nadel J, Holloway JK, Hayward JJ, Cohen PE, Greally JM, Wang J, Bustamante CD, Boyko AR. 2013. Genetic recombination is targeted towards gene promoter regions in dogs. *Plos Genetics* 9:e1003984.

Ersmark E, Klutsch CFC, Chan YL, Sinding MHS, Fain SR, Illarionova NA, Oskarsson M, Uhlen M, Zhang Y, Dalen L, Savolainen P. 2016. From the past to the present: wolf phylogeography and demographic history based on the Mitochondrial control region. *Front. Ecol. Evol*. 4:134. doi: 10.3389/fevo.2016.00134.

Fan Z, Silva P, Gronau I, Wang S, Armero AS, Schweizer RM, Ramirez O, Pollinger Galaverni M, Del-Vecchyo DO, et al. 2016. Worldwide patterns of genomic variation and admixture in gray wolves. *Genome Research* 26:163-173.

Gopalakrishnan S, Sinding MHS, Ramos-Madrigal J, Niemann J, Castuita JAS, Vieira FG, Caroe C, Montero MM, Kuderno L, Serres A, et al. 2018. Interspecific gene flow shaped the evolution of the genus *Canis*. *Current Biology* 28:3441-3449.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018 MEGA X: Molecular Evolutionary Genetics Analysis across computing platorms. Molecular Biology and Evolution 25: 1547-1549.

Loog L, Thalmann O, Sinding MHS, Schuenemann VJ, Perri A, Germonpre M, Bocherens H, Witt KE, Castruita JAS, Velasco MS. 2020. Ancient DNA suggests modern wolves trace their origin to a Late Pleistocene expansion from Beringia. *Molecular Ecology* 29:1596-1610.

Sharma DK, Maldonado JE, Jhala YV, Fleischer RC. 2004. Ancient wolf lineages in India. *Proc Royal Soc B*. 271:S1-4.

Sinding MHS, Gopalakrishnan S, Vieira FG, Castruita JAS, Raundrup K, Jorgensen MPH, Meldgaard M, Petersen B, Sicheritz-Ponten T, Mikkelsen JB, et al. 2018. Population genomics of grey wolves and wolf-like canids in North America. *PLoS Genet* 14:e1007745.

vonHoldt BM, Cahill JA, Fan Z, Gronau I, Robinson J, Pollinger JP, Shapiro B, Wall J, Wayne RK. 2016. Whole-genome sequence analysis shows that two endemic species of Northern American wolf are admixtures of the coyote and gray wolf. *Science Advances* 2:e1501714.

Wang GD, Zhai W, Yang HC, Fan R, Cao X, Zhong L, Wang L, Liu F, Wu W, Cheng LG, et al. 2013. The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nature Communications* 4:1860.

Wang GD, Zhai W, Yang JC, Wang L, Zhong L, Liu YH, Fan RX, Yin TT, Zhu CL, Poyarkov AD, Irwin DM, Hytonen MK, Lohi H, Wu CI, Savolainen P, Zhang YP. 2016. Out of southern East Asia: the natural history of domestic dogs across the world. *Cell Research* 26:21-33.

Werhahn G, Liu Y, Meng Y, Cheng C, Lu Z, Atzeni L, Deng Z, Kun S, Shao X, Lu Q, et al. 2020. Himalayan wolf distribution and admixture based on multiple genetic markers. *Journal of Biogeography* 47:1272-1285.

Zhang W, Fan Z, Han E, Hou R, Zhang L, Galaverni M, Huang J, Liu H, Silva P, Li P, et al. 2014. Hypoxia adaptations in the grey wolf (Canis lupus chanco) from Qinghai-Tibet plateau. *PLoS Genet* 10:e1004466.

Chapter 2. Three-way contact zone reveals evidence for adaptive introgression among deeply divergent wolf lineages

Abstract

Gray wolves (*Canis lupus*) are comprised of three major lineages: the widespread Holarctic lineage and the geographically restricted Indian and Tibetan lineages. The Indian and Tibetan lineages are the most divergent lineages and are adapted to arid and high-altitude habitats, respectively. The Holarctic lineage expanded relatively recently during the late Pleistocene to a diverse range of habitats, including those similar to the two other specialized wolf lineages. We used 5 newly and 7 recently sequenced wolf (*Canis lupus*) genomes from the lowland plains and high-altitude mountains of Pakistan, India, and Kyrgyzstan, along with 79 additional canid genomes, to explore the possibility that adaptive introgression from specialized wolf lineages into Holarctic wolves facilitated their expansion. We detected three narrow secondary contact zones among the widespread Holarctic lineage and the divergent Indian and Tibetan lineages. Within West Asian Holarctic wolves, we detected several gene regions that were highly differentiated compared to other Holarctic wolves and signals of high levels of introgression from Indian or African wolves. Ancestry introgressed from the Indian wolf included the BRDT gene region, which may relate to changes in reproductive phenology as an adaptation to arid habitats. Ancestry introgressed from the African wolf was strongly indicated for the EXOC6 gene region, related to glucose metabolism and potentially conferring fitness in hot environments. In the high-altitude Central Asian wolf, we found similar evidence for adaptive introgression from the Tibetan wolf, including the COPS5 gene, which is known to confer hypoxia adaptation in other mammals. Our study highlights the potentially significant role of

adaptive introgression in facilitating expansion of the Holarctic gray wolf into novel environments.

1. Introduction

Climate change has been a powerful evolutionary driver of species diversity. Ice ages often caused range contractions, isolating populations in distinct environments, followed by range expansions during interglacial periods (Hewitt 2000). Depending on the time of isolation and evolution of reproductive barriers, secondary contact between expanding populations can allow novel alleles to pass between lineages. Such introgression can be adaptive when variants of genes underlying specialized functions evolving in one lineage enable another lineage to thrive in similar environments (Hedrick 2013). Termed adaptive introgression, this process has sometimes allowed species to expand into and rapidly adapt to new and extreme environments (Huerta-Sanchez et al. 2014, Jones et al. 2018, Mareues et al 2017, Racimo et al. 2017). The evolutionary role of adaptive introgression is particularly important to understand in the context of present-day climate change, which increasingly forces populations to adapt rapidly or perish.

We leveraged secondary contact zones among three divergent lineages of gray wolf (*Canis lupus*) to explore the role of adaptive introgression in their evolutionary histories. Gray wolves collectively are among the most widespread terrestrial mammals on earth, occurring over a wide range of climatic environments (Vila et al. 1999, Sharma et al. 2004, Wang et al. 2020, Hennelly et al. 2021, Wang et al. 2022). The Holarctic lineage is the most widespread of the three, and stems from an expansion across the Northern Hemisphere during the late Pleistocene (Loog et al. 2020, Ersmark et al. 2016, Koblmuller et al. 2016; Bergström et al. 2022). The Indian and Tibetan lineages diverged long before this expansion event and are currently restricted to the

Indian subcontinent and Tibetan plateau, respectively (Hennelly et al. 2021, Wang et al. 2022, Wang et al. 2020, Loog et al. 2020). Consistent with its widespread range across many habitats, the Holarctic lineage is the most phenotypically diverse. In arid regions of West Asia are small-bodied populations that appear morphologically similar to more distantly related wolves in other arid regions, those of the Indian lineage and the African wolf (*Canis lupaster*, a closely related species). Additionally, across the mountains of Central Asia, high-altitude populations of larger-bodied, thickly furred wolves are morphologically similar to the wolves of the Tibetan lineage. To be conservative, we have herein adopted the traditional nomenclature to describe all three lineages as conspecific, but note that their taxonomy remains in flux and could soon be decided otherwise (Krofel et al. 2022, Werhahn et al. 2022).

In this study, we used 91 newly and previously sequenced canid whole genomes to investigate the evolutionary history of the gray wolf in Eurasia. Our objectives were threefold: (1) characterize population structure, including more precisely mapping the contact zones among the Holarctic, Indian, and Tibetan lineages, (2) explore genomes for signatures of adaptive introgression, and (3) compare the historical demography of the three wolf lineages. We hypothesized that genes evolving in the Indian and Tibetan lineages that adapted them to their respective climatic extremes (arid and high-altitude habitats) were later shared with Holarctic wolf populations, enabling them to adapt and expand into similar environments. Because the African wolf exchanged gene flow with the Holarctic gray wolf in the past (Gopalakrishnan et al. 2018), we also investigated the role of the former species as a potential donor of adaptive genes to West Asian wolves.

2. Results

For this study, we resequenced the whole genomes of 5 wolves sampled near where the ranges of Holarctic, Indian, and Tibetan wolves meet: the lowland Indus plains in Pakistan (n = 1, DG Khan), the Karakoram mountains in Pakistan at approximately 2500 m altitude (n = 2), and the Tian Shan mountains of Kyrgyzstan at 3000 m altitude (n = 2). We also conducted additional (i.e., deeper) sequencing of a previously sequenced Indian wolf from Central India and included 3 additional Indian wolves and 2 Tibetan wolves from the southern boundary of their range in Ladakh, which were recently sequenced in a related study (Hennelly et al. 2021), along with 79 more previously sequenced wolves and related canids (Table S2.1). After aligning to the domestic dog genome assembly (canFam3.1) and filtering, we obtained mitogenomes for all newly sequenced samples and 54 million nuclear SNPs. As shown previously, both mitochondrial genomes and low-recombination regions of the nuclear genomes, particularly the X-chromosome, showed clear differentiation among the three wolf lineages (Hennelly et al. 2021). Despite their proximity to Indian and Tibetan wolf ranges, 4 of the newly sequenced wolves clustered in the Holarctic lineage based on mitochondrial and nuclear phylogenies, while one wolf from the lowland Indus plains (DG Khan) carried an Indian wolf mitogenome, but clustered in the Holarctic lineage of a nuclear DNA phylogeny (Figs. S2.1, S2.2).

2.1 Global phylogeography and major contact zones among wolves in Eurasia

Principal component analysis (PCA) and admixture analysis clustered most Holarctic wolves together as distinct from Indian and Tibetan wolves despite the distribution of the former across diverse habitats, including high-altitude alpine habitats in Central Asia, arid regions of West Asia, and boreal habitats throughout Eurasia (Figs. 2.1A, B). However, the newly sequenced

wolves from the lowland Indus plains and Karakoram mountains of Pakistan, as well as the two from Ladakh, exhibited ancestry deriving from two or all three wolf lineages, indicating the presence of inter-lineage gene flow near the contact zones (Figs. 2.1A, B).

To better characterize geographic patterns of gene flow and barriers near secondary contact zones, we inferred the estimated effective migration surface (EEMS), which indicates regions of low or high gene flow relative to an isolation-by-distance null model (Petkova et al. 2016). Results revealed significant barriers to gene flow centered in southern West and Central Asia corresponding to zones where lineages come into proximity (Figs. 2.1A, B). All three lineages meet in northern Pakistan around the Kashmir region, forming a three-way contact zone. The barrier in western Pakistan separated Indian and West Asian wolves. The wolf from the lowland Indus plains near DG Khan, situated on this contact zone, exhibited admixture between Indian and West Asian lineages (Figs. 2.1C, D), but wolves to either side of the contact zone clustered as distinct, consistent with a narrow hybrid zone where the two lineages met. Similarly, barriers indicated by EEMS surrounded the Tibetan plateau, corresponding in the north, west, and east to a transition between Tibetan wolves and Holarctic wolves of Central Asia, and to the south between Tibetan and Indian wolves. Wolves sampled on these contact zones (i.e., Ladakh, Karakoram) exhibit admixture among two or all three lineages corresponding to proximity to those ranges. However, wolves on either side of these inferred barriers, including from Kyrgyzstan, clustered distinctly with the corresponding lineage. Given the deep divergence among lineages and tens of thousands of years since Holarctic wolves expanded to meet Tibetan and Indian wolf ranges (e.g., Bergström et al. 2022), these patterns of gene flow are consistent

with stable hybrid zones representing semi-permeable barriers to gene flow (Barton and Hewitt 1985).

Despite the reduced gene flow at contact zones, wolves on either side occur in similar environments and share phenotypic traits, rendering the relatively abrupt population boundaries somewhat cryptic. Indian and West Asian wolves occur in hot and arid climatic zones, are small-bodied, and are characterized by short, thin hair (Fig 2.S3, Jhala 2004, Hefner and Geffen 1999). These two wolf populations are so phenotypically similar that they have been described as the same subspecies (*C. l. pallipes*; Castello 2018). Tibetan and high-altitude Central Asian wolves also share many traits likely to be adaptive in high-altitude landscapes, including thick fur, larger body size, and hypoxia tolerance, for example, as facilitated by a particular EPAS1 allele in the Tibetan wolf (Zhang et al. 2014, Werhahn et al. 2018, Werhahn et al. 2020).

2.2 Evidence of adaptive introgression in Holarctic wolves inhabiting arid and high-altitude environments

Next, we investigated evidence for adaptive introgression from (1) Indian wolves to West Asian wolves, and (2) Tibetan wolves to high-altitude Central Asian wolves. Because African wolves inhabit arid environments of northern Africa and experienced ancient gene flow with West Asian wolves (Gopalakrishnan et al. 2018), we also considered this species a possible source of adaptive introgression to West Asian wolves and included them in our analyses. For each focal Holarctic population (i.e., West Asian or high-altitude Central Asian), we first used a modified population branch statistic (mPBS; Yi et al. 2010) to identify genomic regions that were highly differentiated from those of other Holarctic populations in Eurasia, as would be expected for

regions introgressed from a distantly related donor population (Figs. S2.4, S2.5). We then used the f_d statistic (Martin et al. 2015) to identify subsets of these genomic regions indicating recent common ancestry (high frequency of shared alleles) with one of the divergent wolf lineages (i.e., Indian, African, or Tibetan) and therefore likely to reflect introgression from that source.

We identified 34 out of 21,671 total genomic regions with elevated (99^{th} percentile) mPBS and f_d values compared to a null expectation of 2 regions elevated by chance in both categories (i.e., $21{,}700 * [0.01]^2$). This suggests that most of the 34 genomic regions that were highly differentiated between West Asian and other Holarctic wolves reflected high levels of introgression from Indian into West Asian wolves (Fig. 2.2A). We also detected 29 genomic regions, including an exceptional outlier (two adjacent windows) at the EXOC6 gene showing evidence for elevated mPBS values and introgression from African wolves into the West Asian population (Fig. 2.2B).

One of the outlier gene regions exhibiting elevated introgression from Indian to West Asian wolves contained the BRDT and TGFBR3 genes on Chromosome 6 (Fig. 2.2C). The BRDT gene is essential for spermatogenesis and plays a key role in spermatid elongation (Berkovits and Wolgemuth 2013). Inhibiting BRDT directly reduces sperm number and motility, which is under study as a reversible and non-hormonal contraceptive for human males (Matzuk et al. 2012). For male gray wolves, spermatogenesis primarily occurs during breeding season and is reduced or entirely ceases outside of this season (Asa and Valdespino 1998). Both Indian and West Asian wolves have earlier breeding seasons than all other gray wolf populations, with Indian wolf breeding season being shifted approximately three to four months earlier than other gray wolves

(Mech 2002, Reichmann and Saltz 2013). Because BRDT expression has been shown to shift seasonally, it is possible that this gene plays a role in the timing of spermatogenesis for gray wolves (Aly et al. 2021).

As expected, relative to Indian wolf ancestry, signals of African wolf ancestry within West Asian wolves were much less frequent, both genome-wide and at outlier loci (Fig. 2.2B; Fig. S2.6). However, we detected a strong candidate region for adaptive introgression that contained the EXOC6 gene on chromosome 28 (Figs. 2.2B, D). The EXOC6 gene is involved in the translocation of an insulin-regulated glucose transporter (GLUT4), thus having an important role for glucose uptake and metabolism (Sano et al. 2015). In humans, this gene has a strong association with type-2 diabetes (Imamura et al. 2016, Sulaiman et al. 2022). For desert environments, the ability to modulate insulin response through insulin resistance plays an adaptive role to cope with starvation and reduce energy demands (Tsatsoulis et al. 2013, Rocha et al. 2021). Therefore, it is possible that that these genes within this approximately 200-kb region may facilitate survival of these wolves in arid and low-resource environments. Additionally, other candidate regions included NFAT5, a gene involved in osmotic stress response (Lopez-Rodriguez et al. 2004), and three additional genes previously found to be under selection in African wolves (Table S2.2) (Sarabia 2021).

For high altitude Central Asian wolves, we identified 32 genomic regions with elevated mPBS and f_d values (Fig. 2.3A), indicating these regions showed high differentiation with other Holarctic wolves and high introgression with Tibetan wolves. Among these candidate genomic regions, we found high differentation and elevated introgression from Tibetan to high altitude

Central Asian wolves of a 150-kb genomic region containing the COPS5 gene (Fig 2.3A, D). The COPS5 gene has a direct role in stabilizing the hypoxia-inducible factor that is required for hypoxia adaptation in humans (Bemis et al. 2004). It has been reported as a candidate gene for chronic mountain sickness and has been associated with high altitude adaptation for many species, including Tibetan yaks, and Tibetan and Andean human populations (Bigham et al. 2010, Chen et al. 2018, Stobdan et al. 2017). Thus, this genomic region also may confer adaptations to high altitude landscapes in wolves.

Despite its previous association with adaptation to high-altitude environments in Tibetan wolves and mastiffs (Zhang et al. 2014; Miao et al. 2017; vonHoldt et al. 2017), we did not detect a signal of elevated introgression of EPAS1 from Tibetan to high altitude Central Asian wolves in our outlier analysis (Fig. 2.3A; Fig S2.7). This gene is involved in the physiological response to hypoxia and is associated with high altitude adaptation for many species, including Tibetan wolves, dogs, and humans (Zhang et al 2014, Gou et al. Yi et al. 2010). To explore this region further, we phased and extracted the haplotypes in the EPAS1 region so that we could assess the frequency of the putatively adaptive haplotype (Wang et al. 2020, Miao et al. 2017). We detected a low frequency of divergent and putatively adaptive EPAS1 haplotypes in our high altitude Central Asian wolves and, as shown previously, wolves from Xinjiang (Fig. 2.3C; Zhang et al. 2014; Werhahn et al. 2018).

2.3 Contrasting demographic histories within Eurasian wolves

To contextualize our findings in evolutionary time, we examined the demographic history of Eurasian wolves representing the three major lineages using pairwise sequential Markovian

coalescent (PSMC; Li and Durbin 2011). To be comprehensive, we included reanalysis of wolf genomes that had been previously analyzed with PCMS (Wang et al. 2022, Fan et al. 2016). Altogether, we detected two contrasting patterns of demographic history among wolves in Eurasia: (1) the Indian and Tibetan lineages experienced continuous declines in their populations beginning 360 kya and 260 kya respectively, while (2) the Holarctic lineage underwent an expansion event beginning 70 kya (Fig. 2.4A, Table S2.1). The population declines of Indian and Tibetan lineages are consistent with long-term isolation in separate glacial refugia during the late Pleistocene, as has been previously suggested (Wang et al. 2020, Wang et al. 2022, Hennelly et al. 2021). In contrast, the population expansion for all other Eurasian wolves, including West and Central Asian wolves from Saudi Arabia, Kyrgyzstan, Syria, and Iran, suggests they derived together from late Pleistocene expansion events (Loog et al. 2020, Ersmark et al. 2016; Bergström et al. 2022).

The independence of demographic trends among the three lineages was further supported by runs of homozygosity (RoH) and heterozygosity analyses, which indicate that Indian and Tibetan wolves reflect smaller, more isolated populations than those of Holarctic wolves. Relative to other wolves, a much greater proportion of the genomes of Indian and Tibetan wolves from inner Tibet occurred within short RoH (0.1-1Mb), indicative of a history of a long-term bottleneck (Fig. 2.4B; Figs. S2.8, S2.9; Wang et al. 2022). In contrast, a lower overall proportion of the genomes of West and Central Asian Holarctic wolves occurred within RoH of any size and their cumulative RoH curves were similar to other Holarctic wolves in Eurasia (Fig. 2.4B; Fig. S2.8, S2.9). The cumulative RoH curves became differentiated among Holarctic wolves only at the longest RoH (10-100Mb), suggesting that, until recently, West and Central Asian wolves shared

the same demographic history as other Holarctic wolves. The difference among Holarctic wolf populations of proportions of the genome occurring in long RoH reflects recent history consistent with some populations (e.g., Mexican wolves) having become more inbred than others through anthropogenic causes (Fig. 2.3B). These results were further corroborated by our autosomal heterozygosity estimates, which showed wolves from West and Central Asia to have higher heterozygosity (mean 0.0018 +/- 0.0002 heterozygosity per kb) than wolves from India (mean 0.00125 +/-0.0001 heterozygosity per kb) and the inner Tibetan plateau (mean 0.0010 +/- 0.0001 heterozygosity per kb) (Table S2.1, Fig. S2.9**).**

3. Discussion

In the context of recent discoveries that Holarctic wolves underwent a massive late Pleistocene expansion and were highly connected throughout the past 100,000 years (Loog et al. 2020; Bergström et al. 2022), our study provides new insights into the role of hybridization with divergent lineages as a facilitator of that success. We also found abrupt secondary contact zones in southern West and Central Asia corresponding to where Holarctic wolves adjoin the ancient Indian and Tibetan lineages, in contrast to the high connectivity that defined Holarctic wolves throughout the late Pleistocene (Bergström et al. 2022). We then used these contact zones between morphologically and ecologically similar, yet evolutionarily distinct, lineages of gray wolves, as well as the African wolf, to explore the signatures of adaptive introgression. Although the approximate ranges of the three major wolf lineages were known prior to our study (Werhahn et al. 2020, Hennelly et al. 2021), we focused our sampling on gray wolves from ambiguous portions of their ranges to clarify where they came into contact, assess gene flow

across these zones, and to explore evidence for particular genes that were selectively introgressed from one lineage into another.

Regarding our first objective, our findings suggest that the three divergent lineages likely experience reproductive barriers that reduce gene flow across their contact zones. Rather than finding gradual clines in ancestry across contact zones, we found limited admixture giving way rapidly to distinct lineages on either side. For example, our EEMS analysis identified the strongest barriers to gene flow in Eurasia precisely at these contact zones. Additionally, the four wolf samples spanning the Ladakh-Karakoram region showed a steep genetic cline of Tibetan ancestry occurring in a relatively short distance (250 km), consistent with previous mitochondrial findings (Sharma et al. 2004). The narrowest zone flagged as a barrier in our EEMS analyses was that between Indian and Tibetan wolves. The contact zones involving Holarctic wolves in Asia likely formed during the late Pleistocene, coinciding with their expansion. If so, they must have remained relatively stable for tens of thousands of years, which contrasts starkly with the pervasive gene flow that continuously unified the genomes of gray wolves across the Holarctic north for the past hundred thousand years (Bergström et al. 2022). Based on our PCMS trajectories, along with a variety of evidence from other studies (Wang et al. 2020; Wang et al. 2021; Hennelly et al. 2021), the ranges of Indian and Tibetan wolves have likely been in close proximity for even longer than either with Holarctic wolves. The narrowness of contact zones suggested by our findings, along with evidence for their ancient formation, is therefore uncharacteristic of such a highly mobile species.

Together, these observations suggest the operation of some form or multiple forms of semi-permeable reproductive barrier among Holarctic, Indian, and Tibetan wolves. The current taxonomy, which treats the three lineages as conspecific, has been recently challenged to elevate Indian and Tibetan wolves to species level, consistent with recognition of the African wolf as a distinct species (Werhahn et al. 2022; Krofel et al. 2022). Our study suggested the existence of stable hybrid zones and supports the proposal that Indian and Tibetan wolves may warrant species status distinct from one another and the Holarctic gray wolf. Additional research to improve our understanding of the nature of reproductive barriers would further inform this decision.

Regardless of speciation among the three lineages, our findings support the hypothesis that adaptive introgression from the two locally adapted lineages and African wolf may have facilitated the expansion of Holarctic wolves. Northeastern Asia has been pinpointed as the likely source of Holarctic wolf migration throughout the late Pleistocene, suggesting the ancestors of these expansions were likely adapted to colder, low-altitude habitats (Loog et al. 2020; Bergström et al. 2022). In contrast, Indian and African wolves are adapted to hot, arid climates and Tibetan wolves to (or at least a portion of their ancestry; Wang et al. 2020) high-altitude environments. Despite the apparent presence of reproductive barriers among the three lineages, we found evidence that Holarctic wolves obtained alleles with functional significance though introgression from nearby locally adapted lineages or species, which may have allowed them to quickly adapt to novel environments.

For the Holarctic wolves of West Asia, we found evidence of Indian and African wolf ancestry at genes with functions that may facilitate adaptation to arid environments. These genomic regions included genes underlying glucose and fructose metabolism (EXOC6, GLYCTK), temperature regulation (NBEA), osmotic stress response (NFAT5), and Ultra-violet-damage response (ASCC3) (Howard et al. 2014, Kalim et al. 2017, Williamson et al. 2017). A notable difference between African and Indian wolves as donors of alleles to West Asian Holarctic wolves was the extreme signal for two outlier loci for the African wolf (EXOC6, GLYCTK) compared to a more continuous range of outliers for the Indian wolf. We suspect that the accumulation of drift associated with the long isolation time between African and West Asian wolves masked historical introgression of all but the most strongly selected loci. In contrast, Indian wolves continue to share intermittent gene flow with Holarctic wolves, which is expected to result in less power to distinguish neutrally diffusing versus selectively introgressed alleles, particularly those under low to moderate strengths of selection. Among the loci we detected with elevated introgression from Indian into West Asian wolves was the BRDT gene, which has been shown to affect spermatogenesis and its expression to shift seasonally in other species (Aly et al. 2021, Berkovits and Wolgemuth 2013). In contrast to most Holarctic wolves, West Asian and Indian wolves breed 3 to 4 months earlier, an adaptation to potentially enhancing neonatal survival in hotter climates (Mech 2002, Reichmann and Saltz 2013). Another possible benefit of an earlier breeding date is to facilitate gene flow with Indian wolves, potentially leading to greater introgression of a much broader range of coadapted or weakly locally adaptive alleles. The 3- to 4-month difference in reproductive phenology between Indian and most Holarctic wolves would otherwise be expected to pose a significant reproductive barrier; female canids are only in estrus for 7-10 days, and males only produce sperm for approximately 3 months, centered around estrus

(Asa and Valdespino 1998; Sacks 2005). Thus, in addition to potentially conferring higher fitness directly (e.g., through higher neonatal survival), a BRDT allele affecting the timing of sperm production also could have enabled the boreal-evolved wolf to obtain alleles from a suite of genes that confer higher fitness in arid environments.

We also found genes with Tibetan ancestry in high-altitude Central Asian Holarctic wolves that perform functions related to metabolism (ACBD6, SAR1B, SEC24A), bone mineralization (TCIRG1), vision (CRB1), and high-altitude adaptation (COPS5, EPAS1) (Deng et al. 2022, Georges et al. 2013, Ehrenberg et al. 2013, Bemis et al. 2004). One gene that seemed of particular interest was COPS5, which has a demonstrated role in other mammals for high-altitude adaptation (Bigham et al. 2010, Chen et al. 2018, Stodban et al. 2017).

Although we did not find a statistically meaningful signal of EPAS1, a gene previously shown to confer adaptive benefits to hypoxia in Tibetan wolves and other mammals (Witt and Huerta-Sanchez 2019), we did document some introgression of this allele into the high-altitude Holarctic wolf. In contrast, neither we nor any other study we are aware of has observed this allele in a low-elevation wolf. Thus, one possible explanation for not observing a higher frequency of this EPAS1 allele in our Central Asian high-altitude wolves relates to trade-offs in costs and benefits of this allele, along with codominant expression, as was hypothesized to explain an altitudinal gradient in EPAS1 variant frequency in snow leopards (Janecka et al. 2020). Our Holarctic wolf samples from the Central Asian mountains were collected at 2,500-3,000 m, which is lower than the average altitude of the Tibetan plateau (approximately 4,500 m). Moderate frequency of the adaptive hypoxia allele, mostly in heterozygous genotypes, were previously detected in wolves

from Kyrgyzstan and mid-altitudes of Qinghai of China, the latter of which is at the northeastern boundary of the Tibetan wolf range (Zhang et al. 2014, Werhahn et al. 2020). The frequency of the adaptive EPAS1 haplotype for humans is also closely correlated with altitude, with frequencies increasing from approximately 25% at 2500 m to 70% at 4500 m (Hackinger et al. 2016).

Our results add to a growing list of studies that suggest adaptive introgression from different canid species or lineages may have played a role in the gray wolf's ability to inhabit almost every type of terrestrial habitat (Wang et al. 2020, Bannasch et al. 2021, Anderson et al. 2009). Examination of larger numbers of gray wolf samples will also allow insight into the spread and extent of adaptive alleles across the gray wolf's geographic range and elucidate other evolutionary mechanisms of adaptation, such as via standing variation. Parallel studies of other canid groups, such as foxes of the *Vulpes* complex, could provide powerful comparative bases for further advancing our understanding of the combined roles of divergence, secondary contact, and selective introgression as an engine for persistence and adaptation to climate change.

4. Materials and Methods

4.1 Sample and sequencing procedures

We extracted DNA from one wolf from India, three wolves from Pakistan and two wolves from Kyrgyzstan comprising of three blood samples and three tissue samples using the DNAeasy Blood and Tissue Kit (Qiagen) following the manufacturer's protocol. The newly sequenced gray wolf samples were from Maharashtra in India (n=1), Punjab in Pakistan (n=1), Gilgit Baltistan in Pakistan (n=2), and Naryn region in Kyrgyzstan (n=2). The Indian wolf from

Maharashtra originated from a previous study and was resequenced to a deeper depth following the protocols in Hennelly et al. 2021. For the gray wolves from Pakistan, libraries were sequenced on an Illumina sequencer paired end at 150-bp (Illumina). For the Kyrgyzstan gray wolves, libraries sequenced on a BGISEQ-500 paired end at 100-bp. We assembled a dataset of an additional 87 canid genomes originating from published samples from NCBI Sequence Read Archive (SRA).

4.2 Alignment, variant calling, and filtering

Raw reads for all our samples were trimmed using Trim Galore v0.6.5 using the following settings: --paired --gzip (https://www.bioinformatics.babraham.ac.uk/proje cts/trim_galore/). We then mapped reads to the domestic dog genome assembly (CanFam3.1) with an attached gray wolf Y chromosome using BWA MEM v0.7.17. r1188 (Li and Durbin 2009). We attached the Y chromosome to CanFam3.1 by assembling Y-link scaffolds previously identified from the gray wolf assembly and added the assembled Y chromosome to CanFam3.1 (Smeds et al. 2019). After alignment, we identified and removed PCR duplicates using MarkDuplicates tool v2.18.25 from the Picard suite (http://broadinstitute.github.io/picard). The BAM files were then sorted and filtered to only keep properly paired reads (-F 1024) and with a minimum mapping quality score of 10 (-q 10) using SAMtools (Li et al. 2009). Variant calling was performed using the Genome Analysis Toolkit (GATK) version 4.2.3.0 (McKenna et al. 2010, DePristo et al. 2011). First, we used GATK HaplotypeCaller 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 GenotypeGVCFs in GATK. We obtained a set of high quality single nucleotide polymorphisms (SNPs) using the following filtering steps in GATK: QD < 2.0 || SOR

> 3.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || 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). This dataset resulted in ~54 million SNPs across the whole genome.

4.3 Population genetic structure

To explore genome-wide population genetic structure, we conducted a principle component analysis (PCA) and estimated individual ancestry proportions with the autosomal SNPs of gray wolves and dogs. Along with the filtering using GATK, we filtered the variant calling format (VCF) file for sites in which >90% of individuals had calls (-geno 0.1), removed overlapping deletions and symbolic alternate alleles, a minimum allele count of 3 (--mac 3), and a quality score of at least 30 (--minQ 30). In addition, we pruned sites in high linkage disequilibrium by removing each SNP with a r^2 value of greater than 0.5 with any other SNP within a 50 SNP sliding window, advanced by 10 SNPs at a time (-indep-pairwise 50 10 0.5). This resulted in 6.2 million autosomal SNPs in our gray wolf dataset and 7.2 million autosomal SNPs in our gray wolf and dog dataset. We inferred the PCA using plink v1.90 and the estimated individual ancestry proportions using Admixture v1.3.0 (Purcell et al. 2007, Alexander and Lange 2011).

4.4 Phylogenomic relationships

To assess phylogenomic relationships among gray wolves, we constructed a phylogenetic tree using the mitochondria and low recombination regions of the X chromosome. We extracted the mitogenome of our newly sequenced wolf genomes using a de novo assembler, NOVOPlasty v.3.8.3 (Dierckxsens et al. 2016), with a reference mitogenome of a Mongolia wolf (NCBI

accession number KC896375). We removed the Dloop region of the mitogenome, leaving 11,146bp with the Dloop region removed. Along with 118 additional gray wolf and other canid species mitogenome, we inferred the phylogenetic tree using BEAST v1.10.4 (Drummond and Rambaut 2007). For the tree, we used the "Speciation: Birth-Death Process" tree prior with a relaxed lognormal clock. We used the substitution model HKY and a normal prior on the root height of the tree for the TMRCA of the African wild dog (*Lycoan pictus*) at 3.9 Ma (SD=0.3 Ma). The analysis was run for 10 million MCMC cycles, sampling every 1,000, and discarding the first 1 million states as burnin. We used FigTree v1.4.4 to display the trees. For the low recombination regions of the X chromosome phylogeny, we extracted the X chromosome and filtered to include only sites with calls in >90% individuals (-max-missing 0.9), a minimum quality value of 30 (--minQ 30), and removed indels, resulting in ~2.3 million SNPs for the X chromosome. Next, we extracted regions of the VCF file that overlapped with low recombination regions (>0.2 cM/Mb) of the X chromosome based on the dog recombination map from Auton et al. 2013. This resulted in ~410,000 SNPs, which were used to infer the phylogenetic tree using IQ-Tree (Nyugen et al. 2014).

4.5 Broad genome-wide patterns across Asia

We inferred the estimated effective migration surface (EEMS) to assess spatial variation in rates of gene flow among gray wolf populations across Eurasia (Petkova et al. 2016). The EEMS 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 detected regions of high (i.e. corridors) and low (i.e. barriers) gene flow. For the EEMS, we only used Eurasian gray wolves, where we

filtered the dataset in which we kept sites with a minimum allele count of 3 (--mac 3), at least 90% of individuals represented at a site (--geno 0.1), 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). This resulted in ~5.8 million autosomal SNPs across our Eurasian gray wolves. For gathering the geographic data for each gray wolf sample, we had some gray 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. We used deme sizes of 500, 750, and 1000 with MCMC chains consisting of 15,000,000 iterations following 1,000,000 burnin and a thinning interval of 40,000. We ran each of the analysis with 500, 750, and 1000 deme sizes a total of three times and combined all runs for the final EEMS. We visually assessed our log posterior plots to verify convergence and visualized the results using the R package reemsplots2.

4.6 Detecting selection and introgression in West Asian and high altitude Central Asian gray wolves

We used a modified population branch statistic (mPBS) to identify genomic regions that are highly differentiated in West Asian and high altitude Central Asian wolves compared to other Holarctic wolf populations, suggesting these genomic regions are under selection (Yi et al. 2010). We first filtered our autosomal VCF file in which we kept sites with only gray wolves, a minimum allele count of 3 (--mac 3), at least 90% individuals represented at each site (--geno 0.1), only biallelic SNPs, and included only SNPs with a minimum depth of 3 (--minDP). Because the population branch statistics are derived from F_{ST} statistics, we first calculated pairwise F_{ST} using 100kb window size and 0kb sliding window size between Holarctic wolf populations in Eurasia as defined in **Table S1** using the script popgenWindows.py within

https://github.com/simonhmartin/genomics_general. We then extended the population branch statistic to include four populations to better represent the demographic history of Eurasian wolves (Loog et al. 2020). Our modified mPBS statistic assumes that each population diverged from a metapopulation independently, in which high values of mPBS indicate a long branch for focal population relative to the other three populations (Armstrong et al. 2021). We used the mPBS equation that was defined in Armstrong et al. 2021. For West Asian wolves as the focal population, we identified genomic regions that were highly differentiated as compared to three Holarctic wolf populations in Eurasia: European, Central Asian, and East Asian wolf populations. For high altitude Central Asian wolves as the focal population, we identified genomic regions that were highly differentiated as compared to three specific Holarctic wolf populations in Eurasia: West Asian, European, and East Asian wolf populations (**Table S1**). Due to the lack of information on the specific origin and altitude of the samples, we excluded gray wolves from Xinjiang in our selection analysis for high altitude Central Asian wolves.

To identify introgressed genomic regions, we used the f_d statistic (Martin et al. 2015). The f_d statistic quantifies shared ancestral variation and detects genomic regions that have an excess of shared ancestral variation between a set of focal populations. We applied the same filtering criteria as with calculating the mPBS statistic. We then calculated the f_d statistic for 100kb windows with 0 sliding window that allowed at least 100 biallelic SNPs per window using the script ABBABABAwindows.py within https://github.com/simonhmartin/genomics_general. We tested for introgression between P2 and P3 in the structure (((P1,P2),P3),P4) for the following topologies: (((Central Asia, West Asia), Indian),Dhole), (((Central Asia, West Asia), African wolf), Dhole), and (((East Asia, high altitude Central Asia), Tibetan), Dhole). We defined a

genomic regions as a candidate of adaptive introgression if the region was both in the top 1% of outliers in our mPBS and f_d statistic analysis for each of our three introgression analyses. To investigate selected genomic regions more in-depth, we extracted these regions and conducted additional analyses. For this, we inferred F_{ST}, d_{XY}, and pi statistics using 25kb windows with 5kb sliding window using the script popgenWindows.py within https://github.com/simonhmartin/genomics_general. We used the CanFam3.1 (GCA_000002285.2) genome assembly on Ensembl to document and annotate genes that were present within our list of candidate genomic regions that were both in the top 1% of outliers in our mPBS and f_d statistic analyses. Additionally, we phased the autosomal genome using Shapeit v2.r904 with providing the domestic dog genetic map to improve phasing (Auton et al. 2013). The phased vcf files were used to quantify the frequencies of different topologies within selected gene regions to visualize haplotype regions.

4.7 Demographic history

To infer the historical demography of gray wolves, we used the Pairwise Sequential Markovian Coalescent (PSMC; Li and Durbin 2011). PSMC uses a coalescent approach to estimate the history of change in population sizes over time. We only included the autosomal sequences of each gray wolf individual. We converted each bam file to a fasta-like consensus sequence by first using the mpileup command within SAMtools and subsequently using BCFtools view -c to call variants and vcfutils.pl vcf2fq to convert the vcf file to fastq format (Li et al. 2009). We excluded any reads that were less than 20 for minimum mapping quality and minimum base quality (-q 20 -Q 20) and excluded reads with excessive mismatches (-C50). We also removed sites with more than double and less than a third less than a third of the average depth of coverage for each

sample. The PSMC was inferred using the parameters "psmc -N25 -t15 -r5 -p4 + 25*2 + 4 + 6" following previous studies on gray wolves (Wang et al. 2020, Wang et al. 2022). To account for our selected low coverage genomes (<20x), we estimated the false negative rate (FNR) by downsampling two high coverage gray wolf genomes to 16x, 15x, and 10x. These three coverage thresholds correspond to the specific coverage of gray wolf genomes we selected to include in the PSMC. To determine the best FNR, we visually estimated for the best correspondence between the PSMC plots with the high coverage (>20x) wolf genomes and downsampled genomes with various FNR corrections (Fig. S2.10). We then applied the best estimated FNR to the low coverage gray wolf genomes to infer their demographic history (Li and Durbin 2011). We used a mutation rate of 4.5×10^{-9} (Koch et al. 2019) and a generation time of 4.4 years (Mech et al. 2017).

4.8 Genetic diversity and runs of homozygosity

We estimated the overall autosomal heterozygosity for gray wolf samples using mlRho, which estimates heterozygosity as the population mutation rate (theta) under an infinite sites model (Haubold et al. 2010). We excluded any reads that were less than 20 for minimum mapping quality and minimum base quality (-q 20 -Q 20), as well as excluded reads with excessive mismatches (-C50). Additionally, we excluded reads that were over twice the average depth or less than one third of the average depth. We only inferred the overall heterozygosity for samples that had an average depth of over 10x.

Next, we inferred runs of homozygosity (RoH) for gray wolves using two different methods: BCFtools/RoH and plink v1.90 (Narasimhan et al. 2016, Purcell et al. 2007). Along with the

GATK filtering criteria, we also included only biallelic SNPs that had >90% of individuals at each site (-max-missing 0.9), a minimum allele count of 3 (--mac 3), and a quality score of at least 30 (--minQ 30). We also included SNPs that had a depth of more than one third and less than double of the average depth of coverage for each sample. For the BCFtools/RoH analysis, we fixed the alternative allele frequency to 0.4 (-AF -dflt 0.4) and used the dog recombination map from Auton et al. 2013 to account for recombination hotspots (--genetic-map). We kept RoH that had a quality score of at least 80 and excluded RoH that were below100kb in length in our analyses. For our ROH analysis with plink, we used the following settings: --homozyg-window-het 2, --homozyg-window-missing 5, --homozyg-snp 100, --homozyg-kb 500, --homozyg-density 10, --homozyg-gap 100, --homozyg-window-threshold 0. Plink estimated a higher number of shorter RoH than Bcftools/RoH (Fig. S19), however, the two analyses gave qualitatively similar results.

5. References

Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. Nature 405-907-913.

Hedrick PW. 2013. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive introgression. Molecular Ecology 22: 4606-4618.

Huerta-Sanchez E, X Jin, Asan, Z Bianba, BM Peter, et al . 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 512: 194-197.

Marques JP, L Farelo, J Vilela, D Vanderpool, PC Alves, et al. . 2017. Range expansion underlies historical introgressive hybridization in the Iberian hare. Scientific Reports 40788

Jones MR, LS Mills, PC Alves, CM Callahan, et al. 2018. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. Science 360(6395): 1355-1358.

Racimo F, D Gokhman, M Fumagalli, A Ko, T Hansen, et al. 2017. Archaic adaptive introgression in TBX15/WARS2. Molecular Biology and Evolution.

Vilas C, IR Amorim, JA Leonard, D Posado, et al. 1999. Mitochondrial DNA phylogeography and population history of the gray wolf Canis lupus. Molecular Ecology 8(12): 2089-2103.

Sharma DK, JE Maldonado, YV Jhala, RC Fleischer. 2004. Ancient wolf lineages in India. Proceedings of the Royal Society B S1-S4.

Wang MS, M Thakur, YV Jhala, S Wang, S Yellapu, et al. 2022. Genome sequencing of a gray wolf from Peninsular India provides new insight into the evolution and hybridization of gray wolves. Genome Biology and Evolution 14(2): evac012

Wang MS, S Wang, Y Li, YV Jhala, M Thakur, NO Otecko, et al. 2020. Ancient hybridization with an unknown population facilitated high-altitude adaptation of canids. Molecular Biology and Evolution.

Hennelly LM, B Habib, S Modi, E Rueness, P Gaubert, BN Sacks. 2021. Ancient divergence of Indian and Tibetan wolves revealed by recombination-aware phylogenomics. Molecular Ecology 1-14.

Loog L, O Thalmann, MHS Sinding, VJ Schuenemann, A Perri, et al. 2020. Ancient DNA suggests modern wolves trace their origins to a late Pleistocene expansion from Beringia. Molecular Ecology 1596-1610.

Ersmark E, CFC Klutsch, YL Chan, MHS Sinding, SR Fain, et al. 2016. From the past to the present: wolf phylogeography and demographic history based on mitochondrial control region. Frontiers in Ecology and Evolution. 4: 134.

Koblmuller S, C Vila, B Lorente-Galdos, M Dabad, O Ramirez, et al. 2016. Whole mitochondrial genomes illuminate ancient intercontinental dispersal of grey wolves (Canis lupus). Journal of Biogeography 43(9): 1728-1738.

Bergstrom A, DWG Stanton, UH Taron, L Frantz, MHS Sinding, et al. 2022. Grey wolf genomic history reveals a dual ancestry of dogs. Nature https://doi.org/10.1038/s41586-022-04824-9

Krofel M, J Hatlauf, W Bogdanowicz, LAD Campbell, R Godinho, YV Jhala, AC Kitchener, et al. 2022. Towards resolving taxonomic uncertainties in wolf, dog, and jackal lineages of Africa, Eurasia, and Australasia. Journal of Zoology 316(3): 155-168.

Werhahn G, H Senn, DW Macdonald, C Sillero-Zubiri. 2022. The diversity of genus Canis challenges conservation biology: a review of available data on Asian wolves. Frontiers in Ecology and Evolution 315

Gopalakrishnan S, MHS Sinding, J Ramos-Madrigal, J Niemann, et al. 2018. Interspecific gene flow shaped the evolution of the Genus *Canis*. Current Biology 28(21): 3441-3449.

Petkova D, J Novembre, M Stephens. 2016. Visualizing spatial population structure with estimated effective migration surfaces. Nature Genetics 48(1): 94-100.

YV Jhala. 2003. Status, ecology, and conservation of the Indian wolf Canis lupus pallipes Skyes. The Journal of the Bombay Natural History Society. 100:293-307.

Hefner R, E Geffen. 1999. Group size and home range of the Arabian wolf (Canis lupus) in Southern Israel. Journal of Mammalogy 80(2): 611-620.

Zhang W, Z Fan, E Han, R Hou, L Zhang, et al. 2014. Hypoxia adaptations in the gray wolf (Canis lupus chanco) from Qinghai-Tibet Plateau. PloS Genetics 10(7): e1004466.

Werhahn G, H Senn, M Ghazali, D Karmacharya, et al. 2018. The unique genetic adaptation of the Himalayan wolf to high-altitudes and consequences for conservation. Global Ecology and Conservation e00455

Werhahn G, Y Liu, Y Meng, C Cheng, Z Lu, et al. 2020. Himalayan wolf distribution and admixture based on multiple genetic markers. Journal of Biogeography. 1272-1285.

Yi X, Y Liang, E Huerta-Sanchez, X Jin, et al. 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. Science 329:75-78.

Martin SH, JW Davey, CD Jiggins. 2015. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. Molecular Biology and Evolution 32(1): 244-257.

Berkovits BD, DJ Wolgemuth. 2013. The role of the double bromodomain-containing BET genes during mammalian spermatogenesis. Current Topics in Developmental Biology 102: 293-326.

Matzuk MM, MR McKeown, P Filppakopoulos, Q Li, L Ma, et al. 2012. Small-molecular inhibition of BRDT for male contraception. Cell 150(4): 673-684.

Berkovits BD, DJ Wolgemuth. 2013. The role of the double bromodomain-containing BET genes during mammalian spermatogenesis. Current Topics in Developmental Biology 102: 293-326.

Asa CS, C Valdespino. 1998. Canid reproductive biology: an integration of proximate mechanisms and ultimate causes. Amer Zool 38: 251-259.

Mech LD. 2002. Breeding season of wolves, Canis lupus, in relation to latitude. The Canadian Field Naturalist.

Reichmann A, D Saltz. 2003. The golan wolves: the dynamics, behavioral ecology, and management of an endangered pest. Israel Journal of Zoology 51:87-133.

Aly SM, SM Sharaf, AAI Hassanin, AS Griesh. 2021. Relation of githead seabream (*Sparus aurata*) seasonal reproductive activity to hematology, serum biochemistry, histopathology, and BRDT gene expression. Fish Physiology and Biochemistry 47: 961-977.

Sano H, GP Peck, S Blachon, GE Lienhard. 2015. A potential link between insulin signaling and GLUT4 translocation: association of Rab10-GTP with the exocyst subunit EXOC6/6b. Biochemical and biophysical research communications 465(3): 601-605.

Imamura M, A Takahashi, T Yamauchi, K Hara, K Yasuda, et al. 2016. Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes. Nature Communications 10531.

Sulaiman N, MY Hachim, A Khalique, AK Mohammad, SA Heialy, J Taneera. 2022. EXOC6 (Exocyst Complex Component 6) is associated with the risk of Type 2 Diabetes and Pancreatic B-cell dysfunction. Biology 11, 388.

Tsatsoulis A, MD Mantzaris, M Andrikoula. 2013. Insulin resistence: an adaptive mechanism becomes maladaptive in the current environment – an evolutionary perspective. Metabolism 62(5): 622-633.

Rocha JL, R Godinho, JC Brito, R Nielsen. 2021. Life in deserts: the genetic basis of mammalian desert adaptation. Trends in Ecology and Evolution 36(7): 637-650.

Sarabia C. 2021. The wolf and the desert: genomics and molecular ecology of the African golden wolf (*Canis lupaster*). PhD Thesis University of Sevilla.

Bemis L, DA Chan, CV Finkielstein, L Qi, PD Sutphin, X Chen, K Stenmark, AJ Giaccia, W Zundel. 2004. Distinct aerobic and hypoxic mechanisms of HIF-a regulation by CSN5. Genes and Development 18:739-744.

Bigham A, M Bauchet, D Pinto, X Mao, JM Akey, et al. 2010. Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. PloS Genetics 6(9): e1001116.

Chen N, Y Cai, Q Chen, R Li, K Wang, et al. 2018. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. Nature Communications 9(2337).

Stobdan T, Akbari A, Azad P, Zhou D, Poulsen O, Appenzeller O, Gonzales GF, Telenti A, Wong EHM, Saini S, Kirkness EF, Venter JC et al. 2017. New insights into the genetic basis of Monge's disease and adaptation to high-altitude. Mol Biol Evol 34:3154-3168.

Miao B, Z Wang, Y Li. 2017. Genomic analysis reveals hypoxia adaptation in the Tibetan Mastiff by introgression of the gray wolf from the Tibetan plateau. Molecular Biology and Evolution 34(3): 734-743.

vonHoldt B, Z Fan, DOD Vecchyo, RK Wayne. 2017. EPAS1 variants in high altitude Tibetan wolves were selectively introgressed into highland dogs. PeerJ 5:e3522

Gou X, Z Wang, N Li, F Qiu, Z Xu et al. 2014. Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. Genome Research 24:1308-1315.

Li H, R Durbin. 2011. Inference of human population history from individual whole-genome sequences. *Nature* (475) 493-496.

Fan Z, P Silva, I Gronau, S Wang, AS Armero, et al. 2016. Worldwide patterns of genomic variation and admixture in gray wolves. Genome Research 26:163-173.

Howard JT, SD Kachman, WM Snelling, EJ Pollack, et al. 2014. Beef cattle body temperature during climatic stress: a genome-wide association study. Int J Biometeorol. 58(7):1665-1672.

Kalim A, P Fitzsimons, C Till, M Fernando, P Mayne, JO Sass, E Crushell. 2017. Further evidence that D-glycerate kinase (GK) deficiency is a benign disorder. Brain and Development 39(6): 536-538.

Williamson L, et al. 2017. UV Irradiation induces a non-coding RNA that functionally opposes the protein encoded by the same gene. Cell 168(5): 843-855.

Sacks BN, BR Mitchell, CL Williams, HR Ernest. 2005. Coyote movements and social structure along a cryptic population genetic subdivision. Molecular Ecology 14(4): 1241-1249.

Deng L, Y Pan, Y Wang, H Chen, K Yuan et al. 2022. Genetic connections and convergent evolution of tropical indigenous peoples of Asia. Molecular Biology and Evolution 39(2):

Chen XW, H Wang, K Bajaj, P Zhang, ZX Meng, D Ma, et al. 2013. SEC24A deficiency lowers plasma cholesterol through reduced PCSK9 secretion. eLife e00444.

Georges A, J Bonneau, D Bonnefont-Rousselot, J Champiguelle, et al. 2011. Molecular analysis and intestinal expression of SAR1 genes and proteins in Anderson's disease (Chylomicron retention disease). Orphanet Journal of Rare Diseases 6

Ehrenberg M, EA Pierce, GF Cox, AB Fulton. 2013. CRB1: one gene, many phenotypes. Seminars in Ophthalmology 28:5-6.

Will KE, E Huerta-Sanchez.2019. Convergent evolution in human and domesticate adaptation to high-altitude environments. Philosophical transactions of the royal society B. 374(1777).

Janecka JE, C Hacker, J Broderick, et al. 2020. Noninvasive genetics and genomics shed light on the status, phylogeography, and evolution of the elusive snow leopard. In "Conservation Genetics of Mammals" by J. Ortega and JE Maldonado.

Hackinger S, T Kraaijenbrink, Y Xue, M Mezzavilla, Asan, et al. 2016. Wide distribution and altitude correlation of an archaic high-altitude adaptive EPAS1 haplotype in the Himalayas. Human Genetics 135: 393-402.

Bannasch DL, CB Kaelin, A Letko, R Loechel, et al. 2021. Dog colour patterns explained by modular promoters of ancient canid origin. Nature Ecology and Evolution. 5:1415-1423.

Anderson TM, BM vonHoldt, SI Candille, M Musiani, et al. 2009. Molecular and Evolutionary history of melanism in North American gray wolves. Science 323(5919): 1339-1343.

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

Smeds L, I Kojola, H Ellegren. 2019. The evolutionary history of grey wolf Y chromosomes. Molecular Ecology 28(9): 2173-2191.

Li H, B Handsaker, A Wysoker, T Fennell, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):207802079.

McKenna A, M Hanna, E Banks, A Sivachenko, et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 9:1297-12303.

DePristo MA, E Banks, Y Poplin, KV Garimella, JR Maguire, C Hartl. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43:491-498.

Alexander DH, Lange K. 2011. Enhancements to the Admixture algorithm for individual ancestry estimation. *BMC Bioinformatics* 12: 246.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, Bakker PIW, Daly MJ, Sham PC. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *AJHG 81*:559-575.

Dierckxsens N, Mardulyn P, Smits G. 2016. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research* 45:e18.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7: 214.

Auton A, Li YR, Kidd J, Oliveira K, Nadel J, Holloway JK, Hayward JJ, Cohen PE, Greally JM, Wang J, Bustamante CD, Boyko AR. 2013. Genetic recombination is targeted towards gene promoter regions in dogs. *Plos Genetics* 9:e1003984.

Nguyen LT, Schmidt HA, Haeseler A, BQ. 2014. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32:268-274.

Armstrong EE, A Khan, RW Taylor, A Gouy, G Greenbaum, A Thiery, JT Kang, SA Redondo, S Proft, G Barsh, C Kaelin, S Phalke, A Chugani, M Gilbert, D Miquelle, A Zachariah, U Borthakur, A Reddy, E Louis, OA Ryder, YV Jhala, D Petrov, L Excoffier, E Hadly, U Ramakrishnan. 2021. Recent evolutionary history of tigers highlights contrasting roles of genetic drift and selection. Molecular Biology and Evolution 38(6): 2366-2379.

Haubold B, P Pfaffelhuber, M Lynch.2010. mlRho – a program for estimating the population mutation and recombination rates from shotgun-sequenced diploid genomes. Molecular Ecology 277-284.

Narasimhan V, P Danecek, A Scally, Y Xue, et al. 016. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics 32(11)):1749-1751.

6. Main Figures

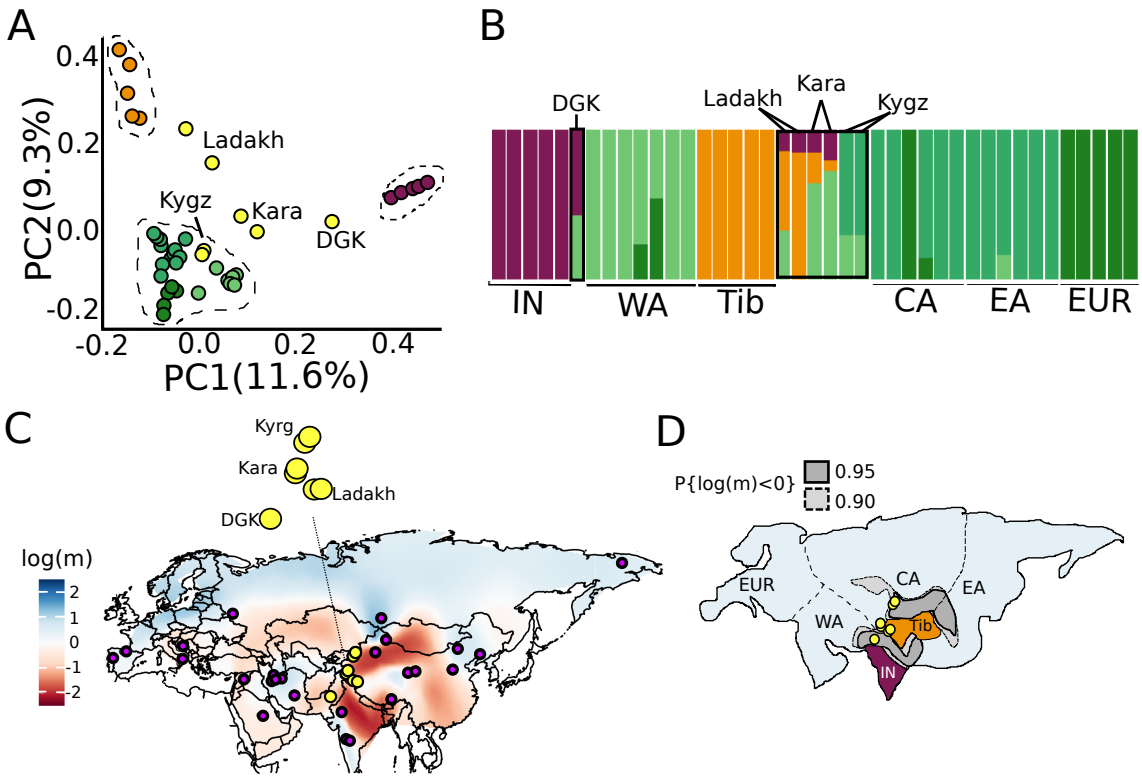


Figure 2.1 Population structure and barriers to effective migration for gray wolves across Eurasia. (A) Principal component analysis with Indian, Tibetan, and Holarctic lineages enclosed by dashed polygons. Newly sampled wolves that were situated on contact zones are shown in yellow, and include Kyrgyzstan (Kyrg), the Karakoram Mountains (Kara) and Dera Ghazi Khan (DGK) in Pakistan, and the Ladakh region in India. (B) Individual admixture proportions using K = 4 genetic clusters. Each bar represents an individual and colors represent the estimated ancestry attributed to each genetic cluster. The Indian (IN) and Tibetan (Tib) form their own genetic clusters separate from Eurasian Holarctic wolves, which include European (Eur), West Asian (WA), Central Asian (CA), and East Asian (EA) populations. (C) Estimated effective migration surface (EEMS) among gray wolves across Eurasia is shown with newly sequenced (yellow circles) and previously sequenced (purple circles) wolves superimposed. The colors of the EEMS surface correspond to lower-than-average (red) and higher-than-average (blue) effective m expressed as $\log_{10}$(m). (D) Ranges of Holarctic, Indian, and Tibetan wolves relative statistically significant barriers to gene flow detected in the EEMS analysis, where the posterior probability that migration (m) is < 1 (i.e., log[0]) migrant per generation was ≥90% (gray) and ≥95% (dark gray). All analyses were based on ~5.79 million SNPs from 41 Eurasian gray wolf samples.

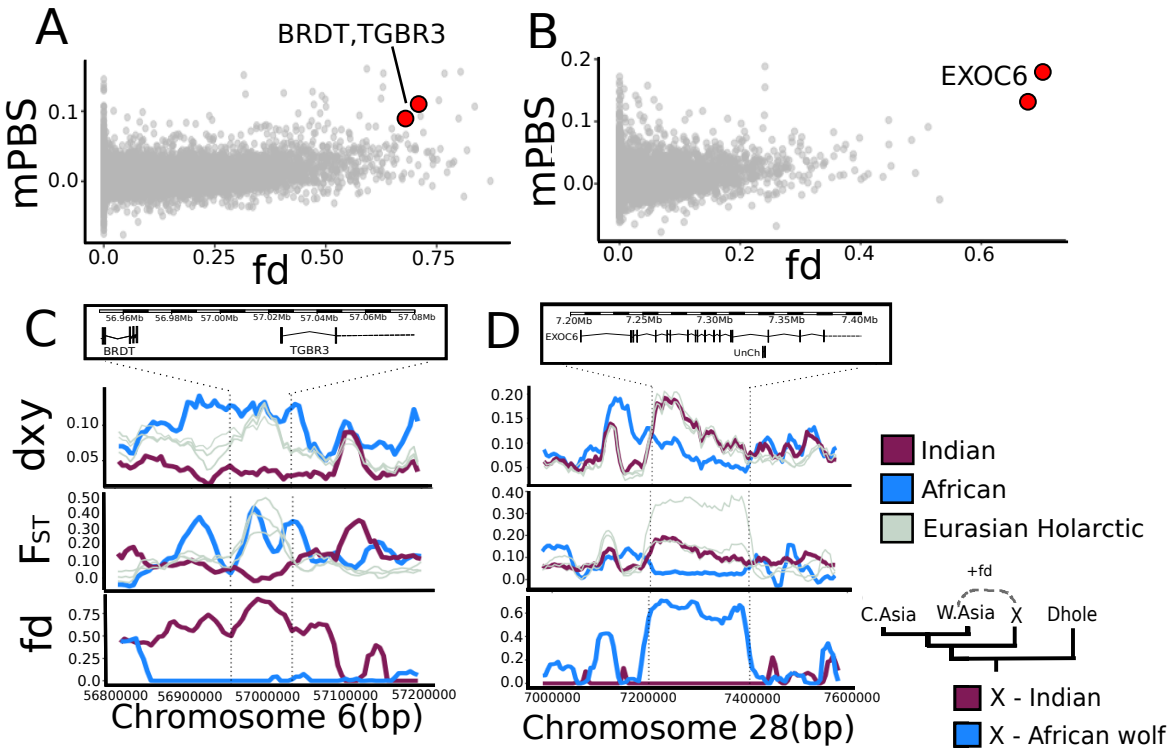


Figure 2.2 Two gene regions showing evidence of adaptive introgression from Indian or African wolves into West Asian wolves. (A,B) Plots showing relationships between modified population branch statistic (mPBS) and fd statistics inferred for 21,671 and 21,721 100-kb non-overlapping windows across the genome for analyses involving Indian and African wolves, respectively. Dashed lines represent 99th percentiles of both mPBS and fd values; red circles highlight selected examples with biological functions potentially under selection, including (A) an 80-kb region containing BRDT/TGBR3 suggestive of selective introgression from Indian into West Asian wolves and (B) a 200-kb region containing EXOC6 suggestive of selective introgression from African wolves into West Asian wolves. (C, D) Using 25-kb sliding (by increments of 5 km) windows, West Asian wolves showed elevated sequence distance (dxy) and allele-frequency differentiation (F_{ST}) relative to other Holarctic wolves at these loci, indicating likely origins from another source. At these same regions, West Asian wolves showed elevated signals of introgression (fd) at the (C) BRDT/TGBR3 region with Indian wolves and (D) EXOC6 region with African wolves, consistent with these regions originating in these divergent lineages and being selectively introgressed into West Asian wolves.

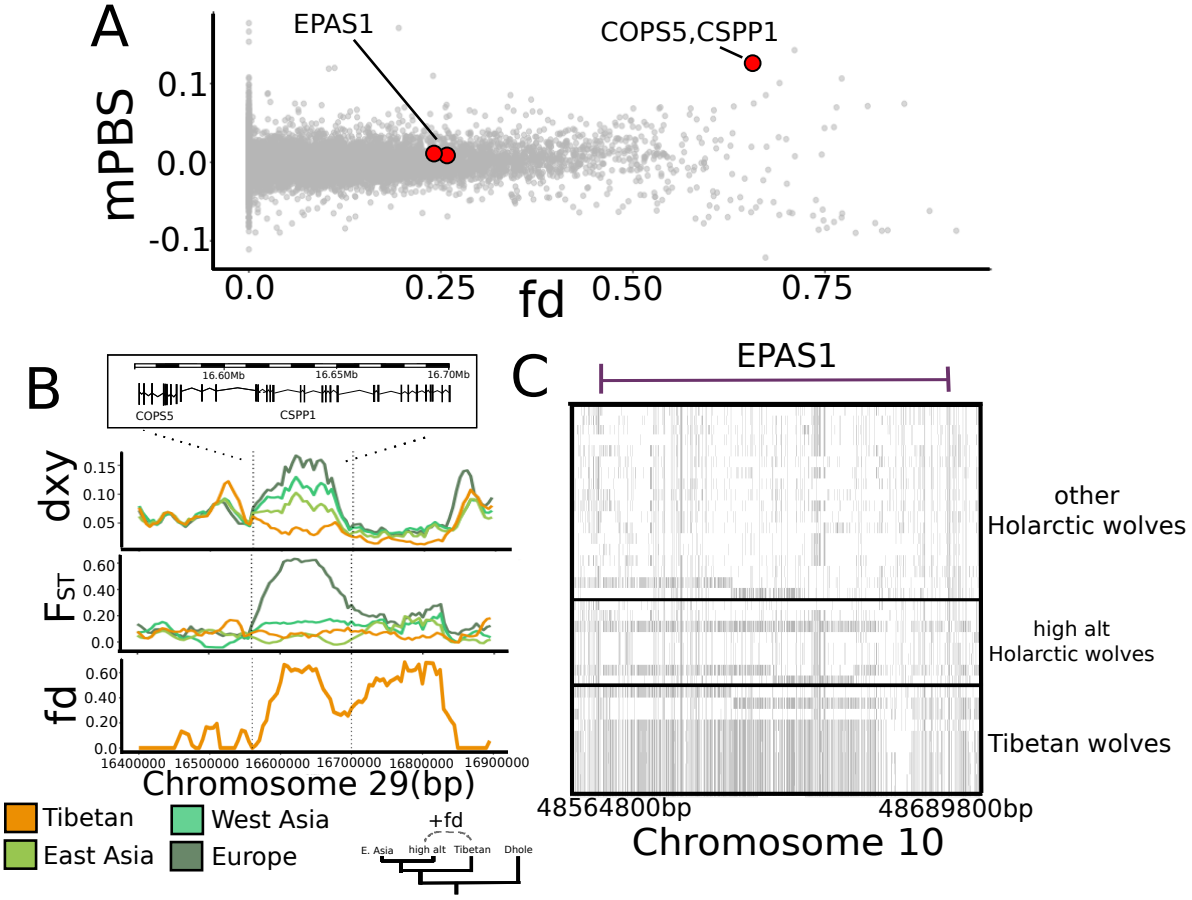
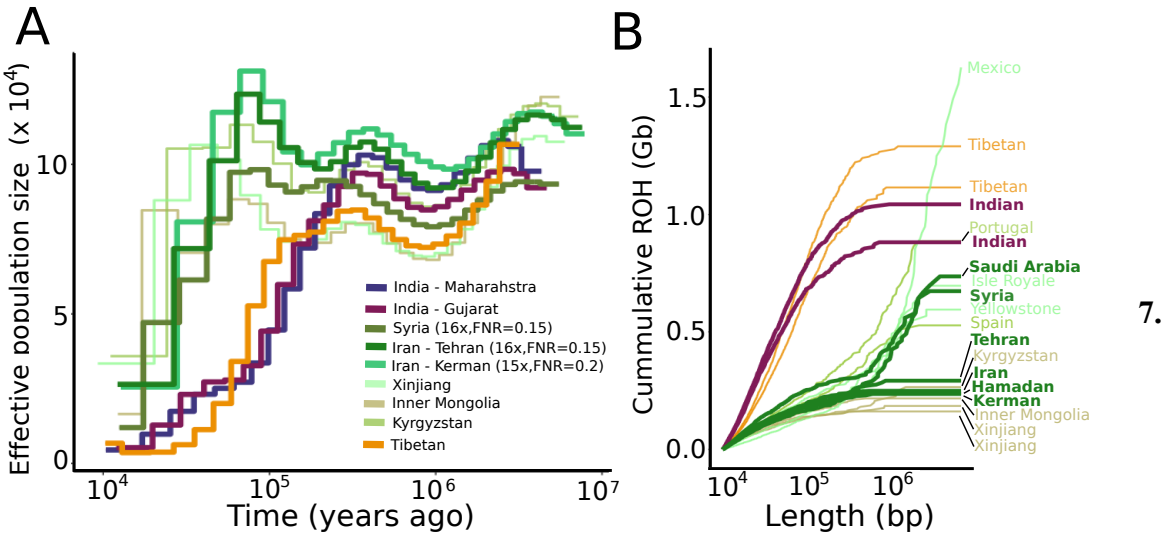


Figure 2.3 Evidence of adaptive introgression from Tibetan wolves into high altitude Central Asian wolves. (A) Plot showing relationships between modified population branch statistic (mPBS) and fd statistics for 21,669 100-kb non-overlapping windows across the genome. Dashed lines represent 99th percentiles of both mPBS and fd values; red circles highlight selected examples of gene regions that were either previously implicated as a locally adapted high-altitude gene region (EPAS1) or within the top 1% of both mPBS and fd values and have biological functions potentially under selection (COPS5/CSPP1). (B) Based on 25-kb sliding (by increments of 5 km) windows, the COPS5/CSPP1 gene region of high altitude Central Asian wolves showed elevated sequence distance (dxy) and allele-frequency differentiation (F_{ST}) from other Holarctic wolves, and an elevated signal of introgression (fd) from Tibetan wolves. (C) Phased haplotypes spanning 125 kb of the EPAS1 gene region in Holarctic and Tibetan wolves. Asterisks show haplotypes that have high homology with most Tibetan wolf haplotypes, which include high altitude Central Asian Holarctic wolves from the Karakoram and Tien Shan Mountains and one Holarctic wolf from Xinjiang from an unknown altitude. Each row represents one of the two haplotypes per individual. White color indicates reference allele and gray color indicates the derived allele. Although only 8 examples of low

altitude Holarctic wolves are shown, no other low altitude Holarctic wolves (n = 18) carried Tibetan EPAS1 haplotypes.



7.

Figure 2.4. Contrasting demographic histories of gray wolves in Eurasia. (A) Pairwise sequentially Markovian coalescent (PSMC) analysis of the demographic histories of 9 gray wolves representing multiple populations in Eurasia. Thicker lines indicate populations from West Asia and the Indian subcontinent. False negative rates (FNR) were estimated for a range of coverages by down-sampling two high-coverage (≥23x) wolf genomes and then used to correct plots for lower-coverage genomes (15–20x) from West Asia. The depths of the 9 wolf genomes used in the PSMC analysis ranged from 15x to 30x. We assumed a mutation rate of 4.5 x 10^{-9} (Koch et al. 2019) and a generation time of 4.4 years (Mech et al. 2017). **(B)** Cumulative lengths of runs of homozygosity (RoH) vs lengths of individual RoHs arranged from shortest (left) to longest (right) across genomes of 19 gray wolves, indicating higher frequencies of low- to medium-length RoHs in Indian and Tibetan wolves relative to all Holarctic wolves; by contrast, the most inbred Holarctic wolves, such as the Mexican gray wolf, showed most of cumulative length of RoHs in long (e.g., >1 Mb) sequences**.**

7. Supplemental Materials

Table S2.1. Sample information whole genome-sequenced canids, including 6 sequenced in this study and 85 additional genomes assembled from short-read archives in GenBank.

Sample Name	Wolf Population	Location	SRA ID	Study	estimated coverage
Gray wolf	Karakoram	near Jutal Village, Gilgit Baltistan, Pakistan 2019	**new**	**this study**	9.82
Gray wolf	Karakoram	Central Karakoram National Park, Gilgit Baltistan, Pakistan 2018	**new**	**this study**	10.79
Gray wolf	Indus River plains	Outskirts of Dera Ghazi Khan, Punjab, Pakistan	**new**	**this study**	9.79
Gray wolf	Central Asia	Pikertyk, Kyrgyzstan	**new**	**this study**	24.7
Gray wolf	Central Asia	Sarychat, Kyrgyzstan	**new**	**this study**	19.35
Indian wolf	Indian	Gangewadi, Maharasthra, India	**new**	**this study**	26.691
Indian wolf	Indian	Outskirts of Jodhpur, Rajasthan, India	SRR14777846	Hennelly et al. 2021	5.93
Indian wolf	Indian	Nannaj, Maharashtra, India	SRR14777844	Hennelly et al. 2021	7.213
Indian wolf	Indian	Baramati, Maharashtra, India	SRR14777843	Hennelly et al. 2021	8.665
Indian wolf	Indian	Velavadar Blackbuck National Park, Gujarat, India	SRR13985171	Wang et al. 2022	30.16
Tibetan wolf	Tibetan	Ganzi, Qinghai	SRR11085395	Wang et al. 2020	8.8
Tibetan wolf	Tibetan	Domkhar, Sham, Ladakh, India	SRR14777848	Hennelly et al. 2021	5.025
Tibetan wolf	Tibetan	Outside of Kargil, Ladakh, India	SRR14777847	Hennelly et al. 2021	5.406
Tibetan wolf	Tibetan	Luobulingka Zoo, Tibet (wild born), China	SRR7107906	Zhang et al. 2014	24.48
Tibetan wolf	Tibetan	Xining Zoo, Qinghai (wild born), China	SRR7107907	Zhang et al. 2014	24.622
Tibetan wolf	Tibetan	Xining Zoo, Qinghai (wild born), China	SRR7107910	Zhang et al. 2014	24.098
Tibetan wolf	Tibetan	Kashi Garden, (wild born), Tibet, China	SRR7107912	Zhang et al. 2014	23.468
Gray wolf	West Asia	Kerman, Iran	SRR12009566	Ghanatsaman et al. 2020.	14.99
Gray wolf	West Asia	Terhan, Iran	SRR12009567	Ghanatsaman et al. 2020.	16.08
Gray wolf	West Asia	Hamadan, Iran	SRR12009568	Ghanatsaman et al. 2020.	10.75
Gray wolf	West Asia	Iran	SRR1518518	Fan et al. 2016	13.148

Gray wolf	West Asia	Breeding Centre for Endangered Arabian Wildlife, Saudi Arabia	SRR8049193	Gopalakrishnan et al. 2018	13.146
Gray wolf	West Asia	Breeding Centre for Endangered Arabian Wildlife, Syria	SRR8049194	Gopalakrishnan et al. 2018	16.196
Gray wolf	West Asia	Neve Ativ, Golan Heights, Israel	SRR2149870	Freedman et al.	5.81
Gray wolf	Europe	Minho, Portugal	SRR7107782	Fan et al.	24.2
Gray wolf	Europe	Perkovic, Croatia, wild caught	SRR2149873	Freedman et al.	6.67
Gray wolf	Europe	Castilla y Leon, Spain	SRR7107785	Fan et al.	23.36
Gray wolf	Europe	Calabria, Italy	SRR7107779	Fan et al.	5.31
Gray wolf	Europe	Bryansk, Russia	SRR7107647	Wang et al. 2013	9.846
Gray wolf	Central Asia	Xinjiang, China	SRR2827601	Wang et al. 2016	10.105
Gray wolf	Central Asia	Xinjiang, China	SRR2827611	Wang et al. 2016	7.985
Gray wolf	Central Asia	Kalamaili Nature Reserve (wild born), Xinjiang, China	SRR7107909	Zhang et al. 2014	22.152
Gray wolf	Central Asia	Duzhishan Garden (wild born), Xinjiang, China	SRR7107911	Zhang et al. 2014	24.792
Gray wolf	Central Asia	Xinjiang, China	SRR2827609	Wang et al. 2016	14.701
Gray wolf	Central Asia	Altai, Russia	SRR7107645	Wang et al. 2013	9.896
Gray wolf	East Asia	Haerbing Zoo, Inner Mongolia (wild born), China	SRR7107908	Zhang et al. 2014	23.617
Gray wolf	East Asia	Haerbing Zoo, Inner Mongolia (wild born), China	SRR7107913	Zhang et al. 2014	20.578
Gray wolf	East Asia	Chukotka, Russia	SRR7107646	Wang et al. 2013	7.439
Gray wolf	East Asia	Lioning, China	SRR2827600	Wang et al. 2016	14.576
Gray wolf	East Asia	Inner Mongolia, China	SRR5168998	vonHoldt et al. 2016	39.378
Gray wolf	East Asia	Shanxi, China	SRR2827603	Wang et al. 2016	10.95
Gray wolf	North America	10km south of Gunnar Island, Ellesmere Island, Canada	SRR8049197	Gopalakrishnan et al. 2018	11.41
Gray wolf	North America	Kobuk River, Alaska	SRR8066602	Sinding et al. 2018	9.51
Gray wolf	North America	La Ronge, Saskatchewan, Canada	SRR8066605	Sinding et al. 2018	7.954
Gray wolf	North America	Isle Royale National Park, Michigan, USA	SRR8380700	Robinson et al.	24.68
Gray wolf	North America	Isle Royale National Park, Michigan, USA	SRR8380711	Robinson et al.	23.78
Gray wolf	North America	Mexican wolf	SRR7976431	vonHoldt et al. 2016	22.46
Gray wolf	North America	Yellowstone National Park, USA	SRR7107786	Fan et al. 2016	24.879

Dhole	Berlin Zoo, Germany	SRR8049189	Gopalakrishnan et al. 2018	17.965
Ethiopian wolf	Ethiopian wolf	SRR8049190	Gopalakrishnan et al. 2018	9.0546
Golden Jackal	Uttarakhand, India	SRR14777842	Hennelly et al. 2021	6.724
African wolf	Kenya	SRR7976426	vonHoldt et al. 2016	23.601
Coyote	California, USA	SRR3574870	vonHoldt et al. 2016	4.948
Coyote	Mexico	SRR8049187	Gopalakrishnan et al. 2018	
Red wolf	Red wolf	SRR7976417	vonHoldt et al. 2016	
Eastern wolf	Algonquin National Park, Canada	SRR8066610	vonHoldt et al. 2016	
Dog	Qahderijani dog from Iran	SRR12009565	Ghanatsaman et al. 2020	
Dog	Saluki dog from Iran	SRR12009563	Ghanatsaman et al. 2020	
Dog	Saluki dog from Iran	SRR12009564	Ghanatsaman et al. 2020	
Dog	Tibetan mastiff from Tibetan plateau china	SRR11085393	Wang et al. 2020	
Dog	Dog from Yushu, Qinghai, China	SRR11085401	Wang et al. 2020	
Dog	Egypt	SRR1061818	Auton et al. 2013	
Dog	India	SRR1061963	Auton et al. 2013	
Dog	India	SRR1061964	Auton et al. 2013	
Dog	Tibetan mastiff	SRR1138361	Guo et al. 2013	
Dog	Tibetan mastiff	SRR1138362	Guo et al. 2013	
Dog	Tibetan mastiff	SRR1138363	Guo et al. 2013	
Dog	Tibetan mastiff	SRR1138367	Guo et al. 2013	
Dog	Tibetan mastiff	SRR1138369	Guo et al. 2013	
Dog	Basenji	SRR2149861	Freedman et al. 2014	
Dog	Greenland Dog	SRR2827569	Plassais et al. 2019	
Dog	Bull Mastiff	SRR7107524	Plassais et al. 2019	
Dog	Mini Bull Terrier	SRR7107631	Plassais et al. 2019	
Dog	West Highland White Terrier	SRR7107636	Plassais et al. 2019	
Dog	Corgi	SRR7107802	Plassais et al. 2019	
Dog	Saint Bernard	SRR7107879	Plassais et al. 2019	
Dog	Shetland Sheepdog	SRR7107905	Plassais et al. 2019	

Dog	Jack Russell	SRR7107924	Plassais et al. 2019
Dog	Shiba Inu	SRR7107933	Plassais et al. 2019
Dog	Lowchen	SRR7107935	Plassais et al. 2019
Dog	Dauchound	SRR7107971	Plassais et al. 2019
Dog	Beagle	SRR7107976	Plassais et al. 2019
Dog	Alaskan Malamute	SRR7107992	Plassais et al. 2019
Dog	Bull Terrier	SRR7120150	Plassais et al. 2019
Dog	Scottish Terrier	SRR7120212	Plassais et al. 2019
Dog	Yorkshire Terrier	SRR7120292	Plassais et al. 2019
Dog	Dingo	SRR10596312	Zhang et al. 2020
Dog	German Shephard	SRR1134655	Guo et al. 2013
Dog	YuanJiang Dog	SRR1138314	Guo et al. 2013
Dog	YuanJiang Dog	SRR1138315	Guo et al. 2013
Dog	YuanJiang Dog	SRR1138333	Guo et al. 2013

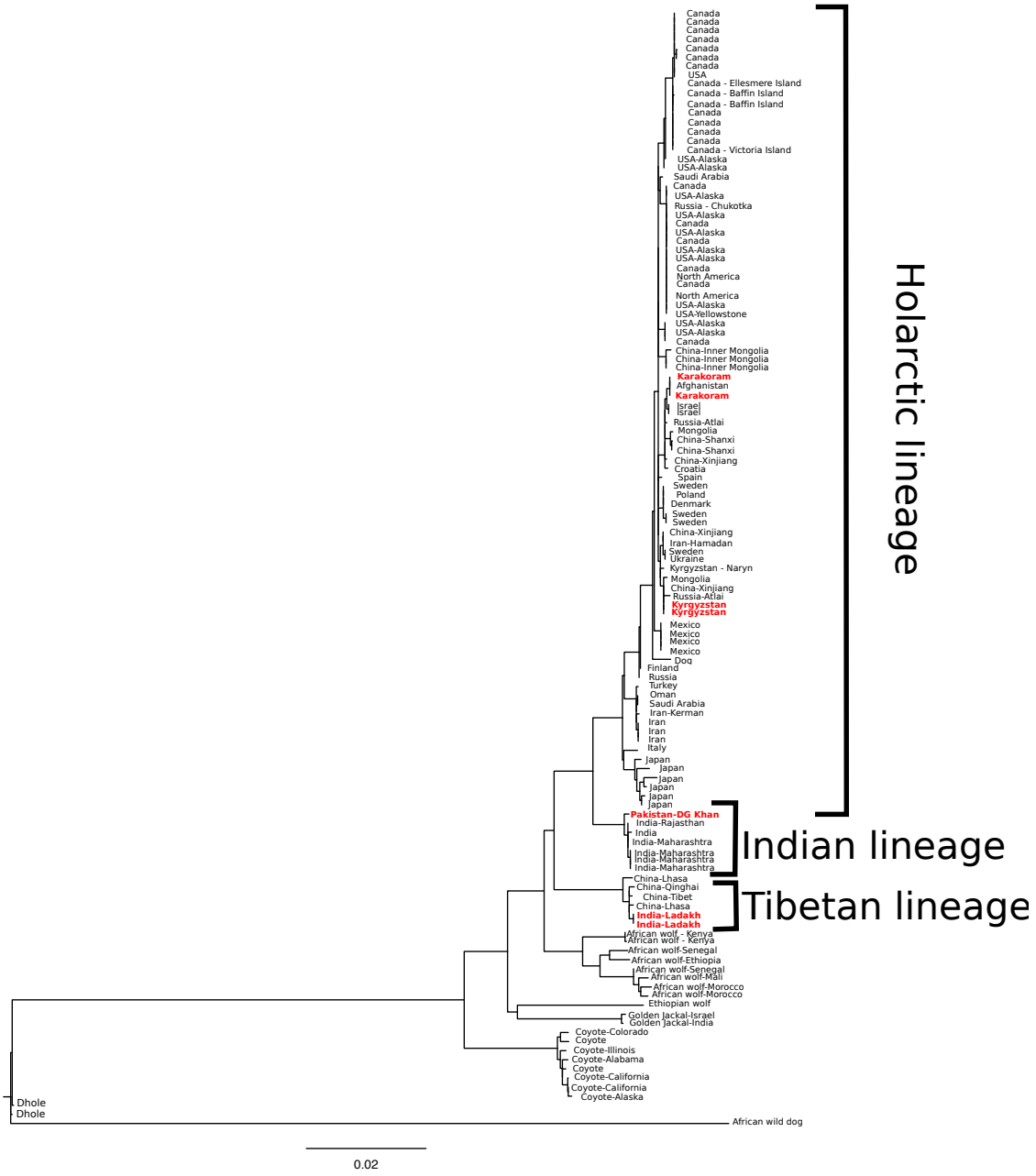


Figure S2.1 Mitogenome phylogeny inferred using 11,146 bp and consisting of 5 newly and 7 previously sequenced gray wolves at contact zones (red text) and an additional 116 individuals. We constructed the phylogenetic tree in BEAST v.1.10.4 (Drummond and Rambaut 2007) and used the substitution model HKY, a birth-death model, and a normal prior on the tree height at 3.9 Ma (SD=0.3) corresponding to the African wild dog (*Lycaon pictus*) divergence time (Chavez et al. 2019).

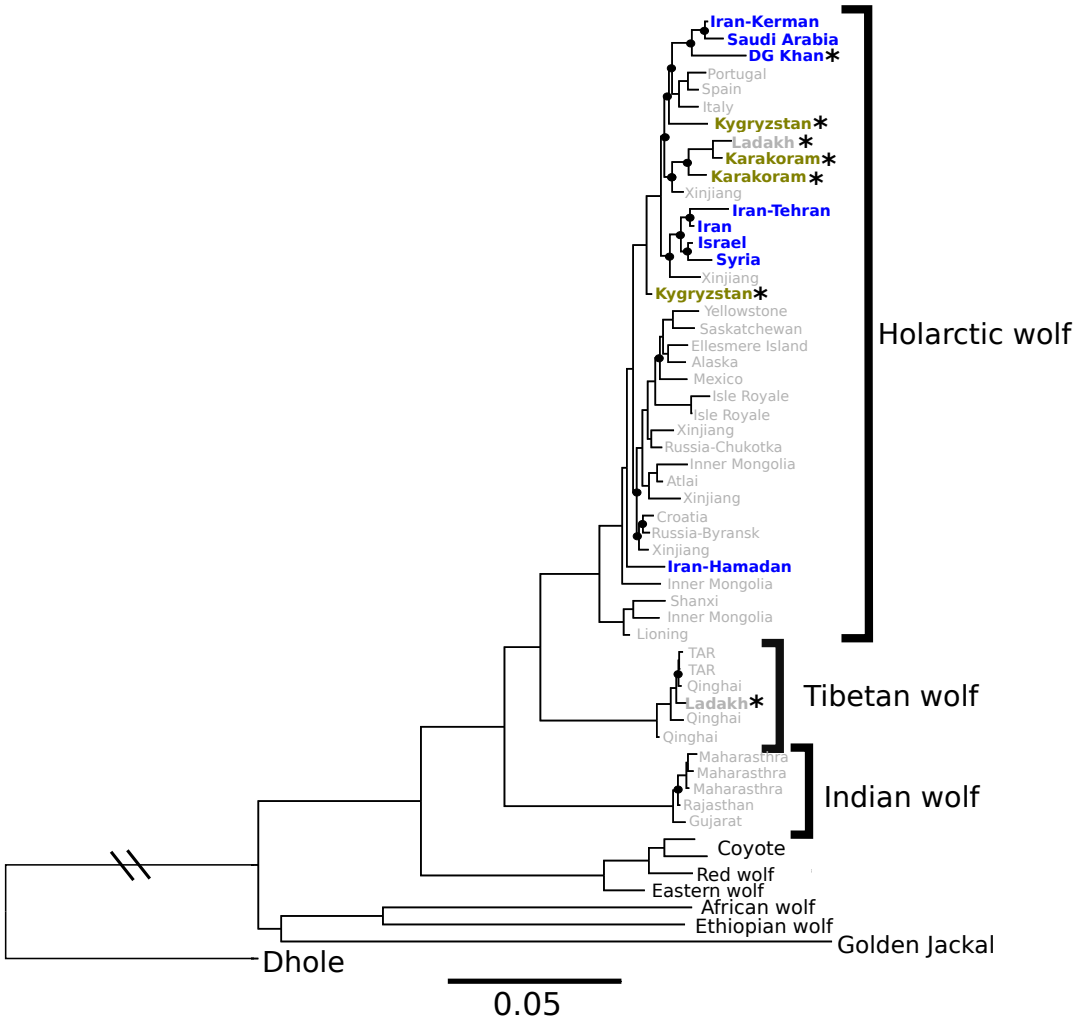


Figure S2.3 Phylogenetic tree of the X chromosome inferred using low recombination regions (<0.2 cM/Mb). We included 48 gray wolves along with 8 other canid species. The tree was constructed from ~410,580 SNPs located in low recombination regions of the X chromosome in using IQ-tree (v.1.6.12) using 1000 ultrafast bootstrap approximations (UFBoot) and a TVM+F+R2 substitution model inferred using ModelFinder within IQ-tree (Nyugen et al. 2014, Kalayaanamoorthy et al. 2017). Nodes with bootstrap support values <100 are depicted with a black circle at corresponding nodes. Gray wolf samples in blue text refer to Holarctic gray wolf samples from arid regions and samples in dark yellow text refer to high-altitude Holarctic gray wolves. Asterisks indicate the seven gray wolves that are located near the contact-zones.

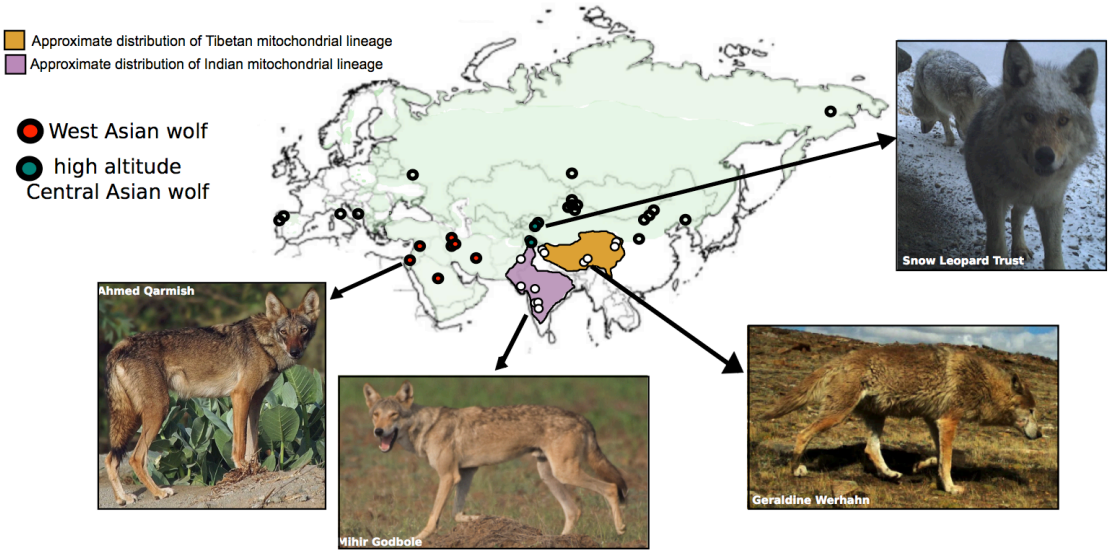


Figure S2.3. Map of Eurasian gray wolf samples included in the study and photographs of the Indian wolf, Tibetan wolf, and Holarctic wolves from West Asia and high-altitude mountains of Central Asia.

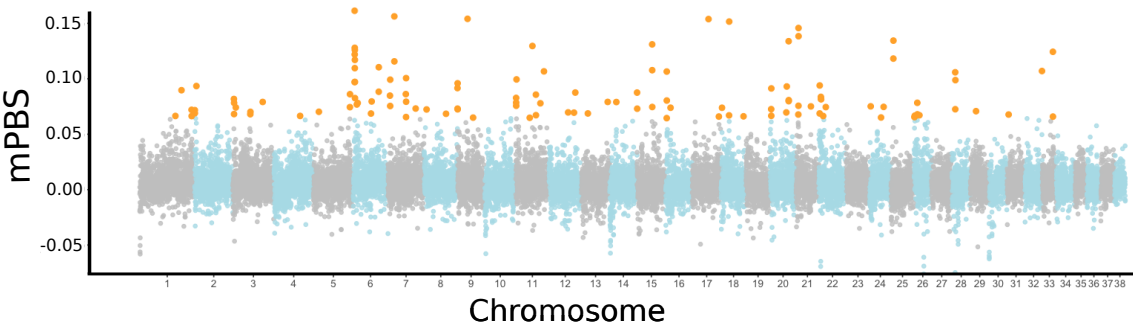


Figure S2.4. Genomic regions showing high differentiation between West Asian and other Holarctic gray wolf populations in Eurasia. Distribution of modified population branch statistic (mPBS) values inferred using 100-kb sliding windows with values in the 99th percentile shown in orange.

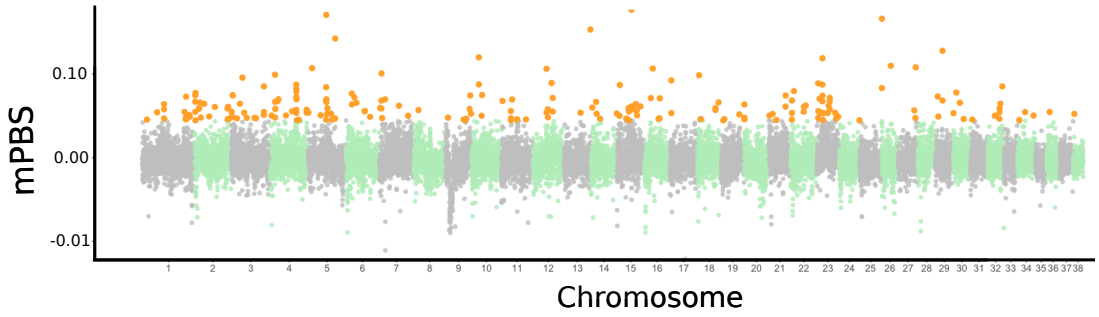


Figure S2.5. Genomic regions showing high differentiation between high altitude Central Asian and other Holarctic wolf populations in Eurasia. Distribution of modified population branch statistic (mPBS) values for 100-kb sliding windows with values in the 99th percentile shown in orange.

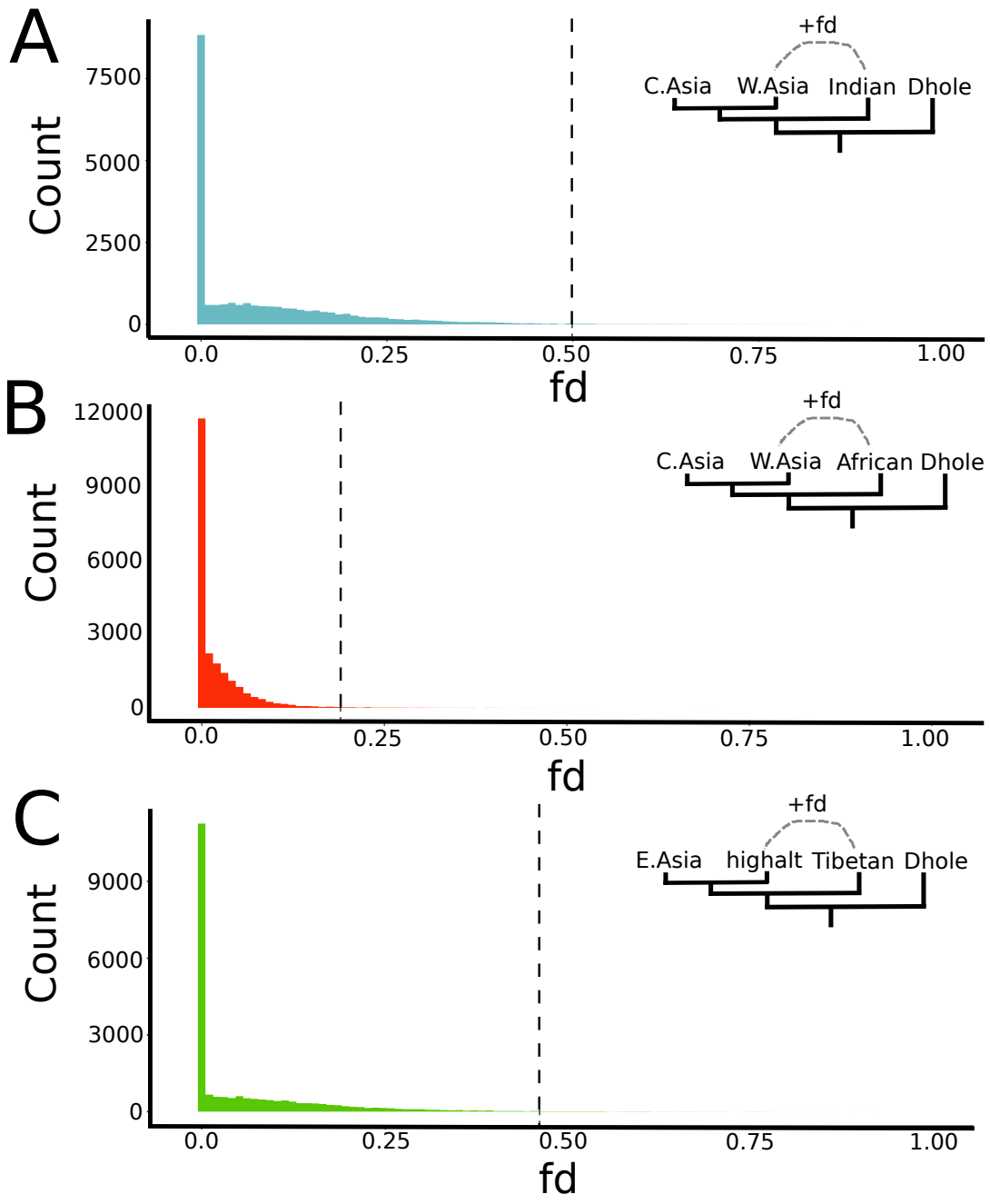


Figure S2.6. Histogram of the f_d statistic indicating levels of introgression inferred between canid taxa**.** The f_d statistic was estimated for 100-kb windows across the autosomes for each test for introgression between two taxa: (A) West Asian (W. Asia) and Indian wolf (mean f_d=0.087), (B) West Asian and African wolf (African, mean f_d= 0.022), and (C) high altitude Central Asian (C. Asia) to Tibetan wolf (mean f_d= 0.064), using either Central Asian or East Asian (E. Asia) wolf as the Holarctic sister taxon and dhole as the outgroup. Dashed lines on each graph indicate the top 1 percentile of the f_d distribution used to identify candidate genomic regions of selective introgression. E. Asia = East Asian wolf

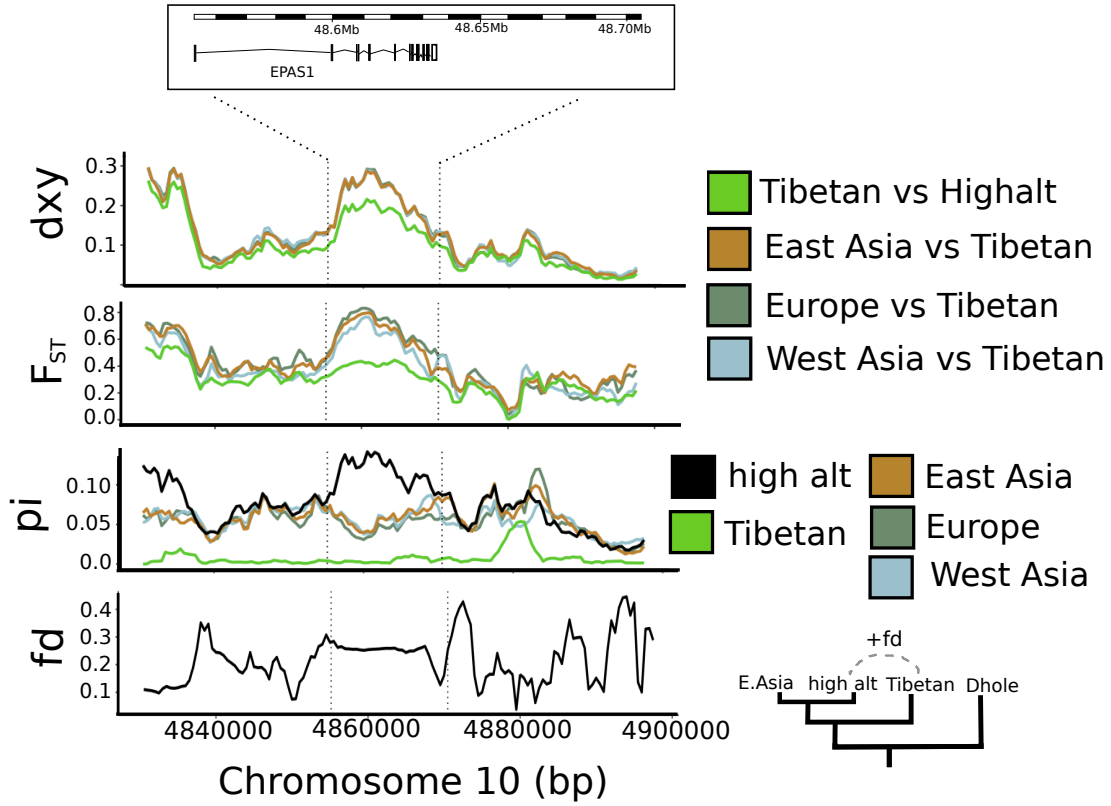


Figure S2.7. Signals of genetic similarity of a 153-kb region (between vertical dashed lines) containing EPAS1 gene introgressed from Tibetan into high altitude Central Asian wolves. Relative to other Holarctic wolves (from West Asia, East Asia, and Europe), high altitude Central Asian wolves were slightly less differentiated from Tibetan wolves based on nucleotide divergence (dxy) and genetic differentiation (F_{ST}); the nucleotide diversity was highest for high-altitude Central Asian wolves and lowest for Tibetan wolves, consistent with fixation in Tibetan wolves and admixture (high heterozygosity) in high-altitude Central Asian wolves in this 153-kb region. The fd statistic also was elevated throughout most of the 153-kb region, although did not differ significantly from flanking regions.

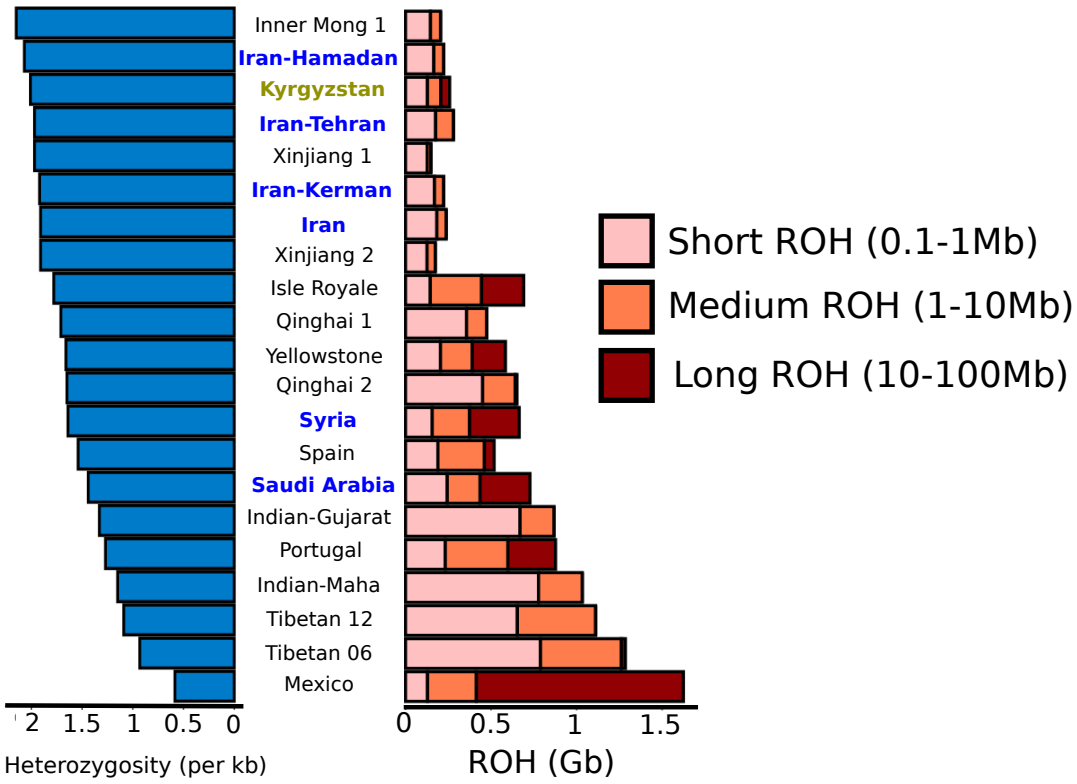


Figure S2.8. Heterozygosity and runs of homozygosity (RoH) in gray wolves. Left: Numbers of heterozygous sites per kb. Right: Sum of the length of RoH in three size ranges: 0.1-1 Mb (Short ROH), 1-10 Mb (Medium RoH), and 10-100 Mb (Long RoH). Individuals are arranged from top to bottom in decreasing order of estimated heterozygosity. Holarctic wolves inhabiting arid regions are labeled in blue letters and Holarctic wolves inhabiting high altitudes are labeled in dark yellow letters.

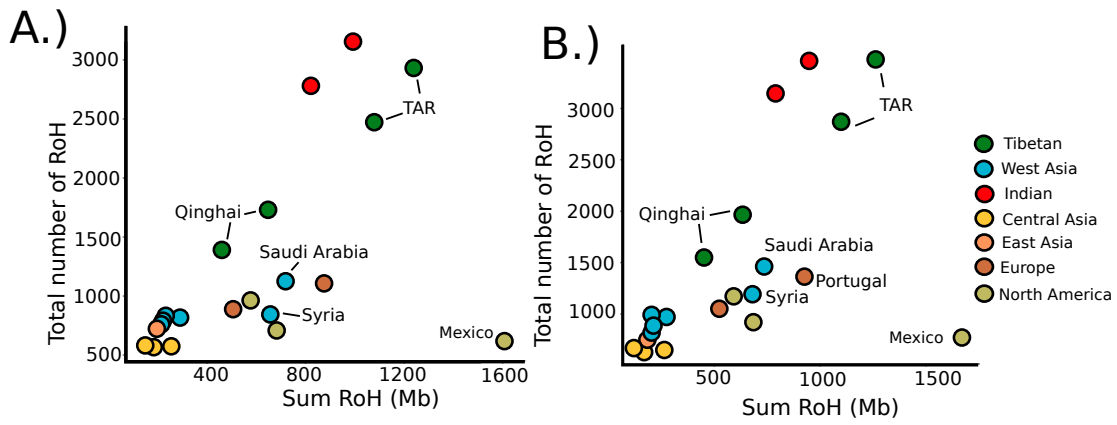


Figure S2.9. Plots of numbers of runs of homozygosity (RoH) verses cumulative length (Mb) of RoH for 21 gray wolves inferred using (A) Bcftools/ROH and (B) plink --homozyg. Indian wolves and Tibetan wolves from the Tibet Autonomous Region (TAR) had the highest number of RoH and more of their genomes in RoH than most other wolves, consistent with ancient bottlenecks. In contrast, the Mexican wolf had very few RoH, yet cumulatively more of its genome in RoH than any other wolf, reflecting recent inbreeding. While Plink estimated a higher number of shorter RoH than Bcftools/RoH), the two analyses gave qualitatively similar results.

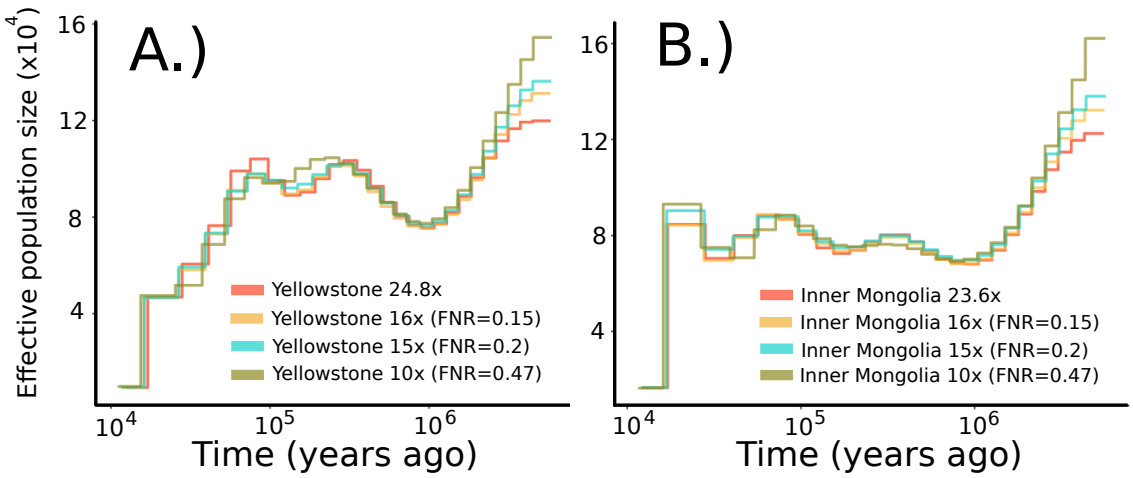


Figure S2.10. Down-sampling of genomes to select false negative rates (FNR) for low-depth corrections of pairwise sequentially Markovian coalescent (PSMC) demographic trajectories. Two gray wolf genomes were down-sampled to 16x, 15x, and 10x: (A) Yellowstone and (B) Inner Mongolia. The FNR corrections for each of three low coverage values (16x, 15, 10x) were selected as those corresponding to demographic trajectories that visually fit the demographic trajectories based on the full datasets (i.e., >20x). The demographic trajectories are shown with the selected FNR values.

Chapter 3. Genomic sequencing of wolves from Pakistan clarifies boundaries among three divergent wolf lineages

Abstract

Gray wolves are one of the most widely distributed terrestrial mammals and include three divergent lineages. The Holarctic lineage is the most widespread and best studied, particularly in North America and Europe. Less is known about wolves of Asia, where Holarctic, Tibetan, and Indian lineages form a three-way contact zone in the vicinity of Pakistan. Given the highly endangered status of the Indian wolf in neighboring India and relatively narrow distribution of Tibetan wolves, it is important to determine precise locations of these boundaries and the extent and location of admixture among. We sequenced 8 wolves from Pakistan to address these questions. Clustering and admixture analyses indicated a sharp cline along the western boundary or Pakistan of mostly Indian in the east to mostly Holarctic ancestry in the west. Wolves from the lowlands and Sulieman range of Pakistan contained high proportions of Indian wolf ancestry. Wolves found in the northern mountain ranges of Pakistan contained a mixture of Holarctic and Indian wolf ancestry, but Indian wolf ancestry was not found further north in Kyrgyzstan, placing the northern boundary of the Indian wolf range in northern Pakistan. Except for small amounts of Tibetan ancestry detected in two wolves from the Karakoram Mountains of northern Pakistan, this lineage appears to end to the east in the Ladakh region of India. Our results highlight the conservation significance of Pakistan's wolf populations, especially the remaining populations in Sindh and Southern Punjab which reflect the highly endangered and divergent Indian lineage.

1. Introduction

Understanding the evolutionary history and distribution of genetically distinct populations is important for conservation. Management decisions are often prioritized based on the taxonomic recognition of species and their populations. However, much of the world's biodiversity remains undescribed and data-deficient, which can lead to loss of taxa before they have been discovered (Mora et al. 2011). This problem is especially pronounced for many biodiverse countries near the equator, where climate change and increasing human-induced pressures threaten many data-deficient plant and animal species (Ecksein et al. 2018, Ali et al. 2022).

Although the gray wolf (*Canis lupus*) has a widespread distribution, its evolutionary diversity is disproportionately concentrated in southern Asia, where all of its three divergent lineages occur (Sharma et al. 2004; Wang et al. 2020; Wang et al. 2022; Hennelly et al. 2021, Ch 2). Two of these lineages, Indian and Tibetan, occur only in southern Asia and differ phenotypically from each other and most Holarctic gray wolves. The Holarctic lineage is the most widespread and bears a distinct evolutionary trajectory involving massive expansion events over the past 100,000 years (Bergström et al. 2022, Loog et al. 2020). The Indian wolf is small and is adapted to its hot, arid environment (Jhala 2003). The Tibetan wolf is large, carries a woolly coat and multiple genes thought to confer adaptation to high altitudes (Werhahn et al. 2018, Werhahn et al. 2020, Zhang et al. 2014). While most Holarctic wolves (which evolved in boreal environments) differ phenotypically from Indian and Tibetan wolves, Holarctic populations in West Asia and high-altitude portions of Central Asia bear phenotypic similarities likely gained through adaptive introgression with the Indian wolf or Tibetan wolf, respectively (Hennelly et al. Ch 2). Whether the three lineages constitute subspecies of gray wolf or distinct species is actively under

consideration, but in either case reflect evolutionarily significant units (Krofel et al. 2022, Werhahn et al. 2022; Hennelly et al. 2021).

The ranges of these three lineages converge in the vicinity of Pakistan, where the wolf represents one of the last remaining large carnivore species. Generally, wolves are classified as endangered and are thought to be in decline in Pakistan (Sheikh and Molur 2003). Phenotypically, wolves in the arid, lowland regions of Sindh and Punjab are similar to Indian and West Asian Holarctic wolves, whereas wolves in the northern Karakoram and Hindu Kush mountains of Gilgit-Baltistan and Khyber Pakhtunkhwa are larger and phenotypically similar to the Tibetan wolf. Wolves near the southwestern mountainous region of Baluchistan are morphologically intermediate between Indian or West Asian and Tibetan wolves (Roberts 1977).

Few data exist on the genetic structure or ancestral origins of wolves in Pakistan, most of which is based on mitochondrial analysis of museum specimens. The Indian mitochondrial lineage has been documented from wolves collected in the lower-elevation regions of Sindh and Punjab, whereas Holarctic mitochondrial haplotypes have been documented from South Waziristan, a region within the Sulieman mountain range of western Pakistan, and in the Karakoram mountains of Gilgit Baltistan (Sharma et al. 2004, Ersmark et al. 2016, Hamid et al. 2019). The nearest documentation of Tibetan wolf haplotypes has been from the Ladakh region of India (Sharma et al. 2004). Consequently, it appears multiple wolf lineages potentially occur in Pakistan, where secondary contact zones may have formed upon the expansion of the Holarctic lineage. Although few genomic data exist for wolves in Pakistan, a recent study including 3 such

samples and many more from throughout Eurasia indicated the general region around Pakistan is likely where the three lineages came into contact (Hennelly et al., Ch 2).

In this study, we used a genotyping-by-sequencing (GBS) approach (Elshire et al. 2011) to sequence 5 new wolves from Pakistan and one from Israel, which we combined with previously sequenced wolves for a data set of 8 wolves from Pakistan and 49 other canid genomes to more precisely locate taxonomic range boundaries with respect to the three potential lineages from Pakistan, as well as to better characterize patterns of introgression across lineages.

2. Methods

2.1 Samples and sequencing procedures

The six new wolf samples used in this study included 5 from Pakistan: Chitral Gol National Park in the Hindu Kush mountain range (n = 2), the Sulieman mountain range near Kohlu in Baluchistan (n=1), South Waziristan (n=1), and the Potwar plateau near Kallar Syedan (n=1), along with one Arabian wolf (*Canis lupus arabs*) from Israel (Table S3.1). We also included three recently sequenced (Hennelly et al., Ch 2) wolf samples from Pakistan, one from the lowland Indus River plains near Dera Ghazi Khan (DG Khan) and two from Central Karakoram National Park in the Karakoram Mountain Range, as well as 33 additional previously sequenced samples from throughout Eurasia (Table S3.1, Fig. 3.1A).

We extracted DNA using the DNeasy Blood and Tissue Kit (Qiagen, CA, USA) following the manufacturer's protocol. For the six new samples, we digested 100 ng of DNA per sample at 37°C for 2 hours with a restriction enzyme (Nsil-HF, an equivalent to EcoT22I; New

England BioLabs Inc., Ipswich, MA). We then ligated both a common and uniquely barcoded adapter to each DNA sample (95 samples and 1 negative control per run). We pooled ligated samples into 8 separate sub-pools (12 samples each) before purifying via QiaQuick PCR columns (Qiagen Inc.) prior to the PCR reaction to minimize unevenness among individual libraries in the final sequencing pool. We then conducted 4 replicate PCR reactions for each of our 8 library sub-pools, for a total of 36 reactions to increase the complexity of our final library. Each 50 µL reaction contained 10 µL of purified adapter ligated DNA, 25 µL of NEB 2X Master Mix, 25 pmol of both forward and reverse PCR Primer, and dH_2O. The PCR conditions were as follows: 5 minutes at 72°C, 30 seconds at 98°C, followed by 18 cycles of 10 seconds at 98°C, 30 seconds at 65°C, and 30 seconds at 72°C, with a final extension at 72°C for five minutes. We purified PCR products with QiaQuick PCR columns, and quantified library concentrations with a Qubit fluorometer (Qiagen Inc.). All libraries were run on 1% agarose gels and a Bioanalyzer (Agilent Technologies, Santa Clara, CA) trace before pooling in equal masses for sequencing (96 libraries/lane) on an Illumina HiSeq4000 (SR100) at the University of California, Davis Genome Center DNA Technologies core.

2.2 Alignment and SNP dataset

Raw reads for all of our samples were trimmed using Trim Galore v0.6.5 (https://www.bioinformatics.babraham.ac.uk/proje cts/trim_galore/). We then mapped the reads to the domestic dog genome assembly (CanFam3.1) with an attached gray wolf Y chromosome using BWA MEM v0.7.17. r1188 (Li and Durbin 2009). We attached the Y chromosome to CanFam3.1 by assembling Y-link scaffolds previously identified from the gray wolf assembly listed in Table S4 in the supplemental materials of Smeds et al. (2019). We removed a 1.74-Mb

region of scaffold 242 that did not meet criteria for Y-linkage, and subsequently combined all scaffolds and added the assembled Y chromosome to CanFam3.1 (https://github.com/hennelly/Ancient-Divergence-Indian-Tibetan-Wolves-Paper/blob/main/Aligning_and_Filtering/Code_Attaching_Ychromosome_onto_DogRefGenome). For whole-genome resequenced (WGS) samples, we then identified and removed PCR duplicates using MarkDuplicates tool v2.18.25 from the Picard suite (http://broadinstitute.github.io/picard). All BAM files originating from WGS and GBS were then sorted and filtered to keep only properly paired reads (-F 1024) and with a minimum mapping quality score of 10 (-q 10) using SAMtools (Li et al. 2009). Because our dataset included both WGS and GBS data, we only retained reads of each WGS sample that overlapped with read positions in the GBS data. To achieve this, we first merged all GBS BAM files to obtain a list of all positions found in our GBS data and then subsetted each WGS bam files to keep reads that overlapped with those GBS positions (samtools view -L). Using these subsetted WGS bam files and GBS bam files, we computed genotype likelihoods using ANGSD 0.9.34 (Korneliussen et al. 2014). We excluded bases with base qualities <20 (-minQ 20), reads with mapping quality <20 (-minMapQ 20), reads with excessive mismatches (-C 50), and we retained sites with called in ≥90% of individuals.

2.3 Population structure and admixture patterns

To examine population differentiation among Eurasian wolves, we conducted a principal component analysis (PCA) using PCangsd (Meisner and Albrecthsen 2018). We then used NGSadmix to infer individual admixture proportions from K = 2 to 8 for our genomic dataset,

which included additional canid species and dogs (Skotte et al. 2013). We used genotype likelihoods deriving only from autosomal chromosomes for the PCA and admixture analyses.

2.4 Inferring gene flow and recent ancestry from D-statistics

To investigate patterns of allele sharing between different wolf populations, we calculated the D statistic (Green et al. 2010). Considering the topology (((P1,P2),P3),P4), a higher proportion of shared derived alleles between P2 and P3 compared to two sister populations P1 and P2 is indicative introgression between P2 and P3. To test for introgression between Eurasian wolves and Indian and Tibetan wolves, we inferred the D statistic to test for introgression between P2 and P3 in the topology structure (((P1, P2),P3),P4). With our wolf samples, we used the topologies (((Yellowstone wolf, X), Indian wolf), Dhole) and (((Yellowstone wolf, X), Tibetan wolf), Dhole), where X is an individual wolf from Eurasia. We also included an analysis to test for introgression between Eurasian wolves and a Central Asian wolf from Xinjiang, which we selected this individual to represent the Holarctic clade, for which we used the topology (((Yellowstone, X), Xinjiang), Dhole). We only used high coverage (>20x) Indian, Tibetan, and Xinjiang wolf individual genomes as P3 to test for introgression. To infer the D statistic, we used the (-doAbbababa) command in ANGSD using sites sequenced in ≥41 individuals with mapping and base-quality scores ≥20; we used only autosomal chromosomes and excluded reads with excessive mismatches (-C50) (Korneliussen et al. 2014). We examined the relationship between the magnitude of D-statistics and geographical distance separating each of the three P3 wolves from all other wolves in Eurasia.

3. Results

3.1 Population structure

We successfully sequenced the five new wolves from Pakistan and one from Israel using GBS and combined these with previously whole-genome resequenced wolves from Pakistan (n = 3) and elsewhere throughout Eurasia (n = 47). The average depth of our GBS samples was 41.5x, ranging 9.8x to 129.9x. After combining GBS data with subsetted WGS samples, we obtained 135,044 autosomal SNPs for the PCA and Admixture analyses, and 88,148 autosomal SNPs for the D statistic analysis.

3.2 Genome-wide patterns of ancestry for wolves in Pakistan

In the PCA, the 8 wolves from Pakistan ranged mostly between Indian and Holarctic wolf clusters, although one individual from the Karakoram Mountains was located slightly towards the Ladakh wolves, which were somewhat internal to all three clusters (Fig. 3.1B). Wolves from Pakistan that were geographically closer to lowland India fell closer on the PCA to those of the Indian lineage. Admixture analysis similarly showed high levels of ancestry shared between wolves from lower elevations of Pakistan and India (Fig 3.1C). The two wolves from Chitral assigned primarily to the Indian wolf cluster as well, but contained significant admixture from the Holarctic lineage. Wolves from the Karakoram Mountains of Pakistan and Ladakh region of India showed evidence of gene flow and three out of the four individuals showed admixture from all three wolf lineages. We did not detect any evidence of dog introgression in our wolf samples from Pakistan in NGSadmix at K = 2–8 (Fig. S3.1).

Similarly to the PCA and admixture analysis results, D statistics indicated that most wolves from Pakistan showed high portions of their genomes deriving from the Indian lineage (Table S3.2). Wolves from the lowland Indus plains of DG Khan, Potwar plateau near Kallar Syaden, and the Sulieman Mountains near Kolhu in Boluchistan showed the highest D-statistic values when using the Indian lineage as P3. In contrast, only one wolf from Pakistan, in the Karakoram Mountains, showed high D-statistic values when using the Tibetan lineage as P3.

To explore geographic patterns of gene flow, we plotted D statistics on separation distances between P2 and P3 wolves, using a North American Holarctic wolf in the P1 position. Gene flow between an Indian wolf (P3) and other Eurasian wolves (P2) showed a steep drop, but a continuous relationship, with geographic distance, consistent with a relatively narrow cline (Fig. 3.2A). Although West Asian wolves shared gene flow with Indian wolf, those more geographically distant, such as from Iran, especially in Kerman, had low D statistics with the Indian wolf (P3), consistent with a continuing shallower cline in Indian ancestry west of the Sulieman range of Pakistan. Gene flow between a Tibetan wolf and other Eurasian wolves was low at all distances except with other Tibetan (or Ladakh) wolves (Fig. 3.2B). The D statistics between a Holarctic wolf from Central Asia (Xinjiang, China, P3) and other Eurasian wolves (P2) were highest with other Central Asian wolves (P2) and declined with distance (Fig. 3.2C). Except for West Asian Holarctic wolves, however, even the when separated by >8,000 km (e.g., Central Asia vs Europe), the gene flow indicated between two Holarctic wolves was higher than with Indian or Tibetan wolves.

4. Discussion

Our genome-wide analyses using 8 wolves from Pakistan provide insight into the location of the boundaries of the three major wolf lineages, especially for the Indian and Tibetan lineage distribution. Wolves from the lowlands, Potwar plateau, Sulieman range, and the Hindu Kush range reflected genome-wide Indian ancestry based from our PCA, Admixture, and D statistic results. The two wolves from the Karakoram mountains are mostly Holarctic genome-wide ancestry with smaller amounts of Indian and Tibetan ancestry, consistent with proximity to the three-way contact zone among these lineages.

We detected most Indian genome-wide ancestry within Pakistan wolves of the lowlands, including the Indus river valley and the mountain ranges West of the Indus river. Among these wolves, individuals from the southern and low elevation regions of Pakistan showed the highest proportion of Indian ancestry compared to wolves. These results are consistent with museum-derived mitochondrial and morphological evidence (Sharma et al. 2004, Sarwar et al. 2021). We found wolves living in the northern mountain ranges of Pakistan, such as the Hindu Kush and Karakoram ranges, had the least amount of genome-wide Indian ancestry. These regions are characterized by mountainous terrain, high altitudes, and cold winters, which may have hindered significant introgression from the arid-adapted Indian lineage. Consistent with mitochondrial evidence from museum specimens, the wolf from the Potwar plateau showed a high proportion of ancestry from the Indian lineage despite being located at the Himalayan foothills (Sharma et al. 2004). Kallar Syedan lies at the junction between the Potwar plateau and the Himalayan mountain range, which has been previously hypothesized to represent a secondary contact zone for red foxes as well (Rafaqat et al. 2021). This junction region also represents the most northern

range of amphibian and reptile species found on the plains and the most southern extent of species found in the Himalaya (Masroor 2011). Thus, our data suggests that the Indian lineage may have been historically distributed across the Potwar Plateau, which likely represents the most northern distribution of the Indian lineage.

We observed a steep cline of Tibetan ancestry in Pakistan, whereby wolves in the Karakoram and Chitral mountains exhibited surprisingly little Tibetan ancestry despite living in similar high-altitude habitats. This observation is consistent with mitochondrial evidence showing a transition between Tibetan and Holarctic haplotypes occurring between Ladakh in India and Gilgit-Baltistan in Pakistan (Sharma et al. 2004). Using wolf genomes from across Eurasia has further emphasized the narrowness of contact zones between the Holarctic and Tibetan wolf that surrounds the Tibetan plateau (Ch 2). In fact, we observed little Tibetan genome-wide ancestry outside of Pakistan, such as in Kyrgyzstan and lowland China, suggesting Tibetan ancestry may decline relatively quickly outside of the core Tibetan distribution. As proposed in Chapter 2, our results strengthen the inference that some form of reproductive barrier may be hindering widespread introgression between Tibetan wolves of Ladakh and Holarctic wolves living within the northern mountains of Pakistan.

Our study elucidated multiple secondary contact zones among Holarctic, Indian, and Tibetan lineages. The Holarctic lineage underwent a massive late Pleistocene expansion while the Indian and Tibetan lineages were isolated within the Indian subcontinent and the Tibetan plateau, respectively (Loog et al. 2020, Wang et al. 2022, Wang et al. 2021). Pakistan is situated where the expanding Holarctic lineage would likely have abutted the Indian and Tibetan lineage

distribution, thus potentially forming contact zones. Based on PCA, admixture analyses, and D statistics, wolves of Iran, especially from Kerman, showed only slight Indian ancestry. Demographic analyses have also showed that wolves from West Asia, including Kerman, were part of the late Pleistocene expansion event of Holarctic wolves (Ch 2). Thus, it is conceivable that these boundaries between the Indian or Tibetan lineage and Holarctic lineage were formed during the late Pleistocene and have likely been stable for tens of thousands of years.

These results have implications for the taxonomy and conservation of wolves in Pakistan. Our genome-wide data indicate that the endangered Indian wolf is present throughout lowland Pakistan. Wolves most fully representing the Indian lineage are likely present in Sindh and Southern Punjab, where these populations are increasingly rare and declining (Rais et al. 2011, Sheikh and Molur 2003). Wolves along the Northern Sulieman range of South Waziristan, the Hindu Kush range near Chitral, and Potwar Plateau have a high proportion of Indian ancestry and a lower proportion of Holarctic ancestry. Conserving these admixed populations may contribute to preserving alleles and genetic variation deriving from the Indian lineage. Such admixed populations of other species have been recognized as potentially valuable reservoirs of genetic variation upon which natural selection can act (Supple and Shapiro 2018). Our study suggests that the western boundary of the Indian lineage lies between Baluchistan and Eastern Iran, whereas the northern boundary is likely at the Himalayan foothills. Our findings further suggest the western range extent of the Tibetan lineage corresponds to the Ladakh region of India, and that the wolves of northern Pakistan are primarily of the Holarctic lineage, albeit with some introgression from the Indian lineage. These findings highlight the urgency to conserve the remaining wolves that most fully represent the Indian lineage in Sindh and Punjab, where wolves

are extremely rare (Sharma et al. 2004, Saad et al. 2015). In fact, wolves may already be extinct in Southern Punjab and the Potwar plateau, highlighting the urgency to protect these remaining wolf populations (Sheikh and Molur 2003). A more comprehensive genomic assessment of the Indian wolf in Pakistan is needed to assess and identify potentially isolated, inbred, or otherwise at-risk populations in greatest need of conservation efforts. Together, our results confirm the presence of the endangered and divergent Indian lineage in Pakistan and informs delineation of wolf populations to aid in identifying populations of conservation concern.

5. References

Eckstein D, V Kunzel, L Shafer. 2018. Global Climate Risk Index. Bonn Germany.

Ali H, JU Din, L Bosso, S Hameed, M Kabir, M Younas, MA Nawaz. 2022. Expanding or shrinking? Range shifts in wild ungulates under climate change in Pamir-Karakoram mountains, Pakistan. Plos One 16(12): e0260031.

Sharma DK, JE Maldonado, YV Jhala, RC Fleicher. 2004. Ancient wolf lineages in India. Proceedings of the Royal Society B S1-S4.

Wang MS, S Wang, Y Li, Y Jhala, M Thakur, NO Otecko, et al. 2020. Ancient hybridization with an unknown population facilitated high-altitude adaptation of canids. Molecular Biology and Evolution 37(9): 2616-2629.

Wang MS, M Thakur, Y Jhala, S Wang, S Yellapu, SS Dai, et al. 2022. Genome sequencing of a gray wolf from Peninsular India provides new insights into the evolution and hybridization of gray wolves. Genome Biology and Evolution 14(2): evac012.

Hennelly LM, B Habib, S Modi, EK Rueness, P Gaubert, BN Sacks. 2021. Ancient divergence of Indian and Tibetan wolves revealed by recombination-aware phylogenomics. Molecular Ecology 30(24): 6687-6700.

Bergström A, DWG Stanton, UH Taron, L Frantz, MHS Sinding, et al. 2022. Grey wolf genomic history reveals a dual ancestry of dogs. Nature https://doi.org/10.1038/s41586-022-04824-9

Loog L, O Thalmann, MHS Sinding, VJ Schuenemann, A Perri, et al. 2020. Ancient DNA suggests modern wolves trace their origins to a late Pleistocene expansion from Beringia. Molecular Ecology 1596-1610.

Werhahn G, H Senn, M Ghazali, D Karmacharya, et al. 2018. The unique genetic adaptation of the Himalayan wolf to high-altitudes and consequences for conservation. Global Ecology and Conservation e00455

Werhahn G, Y Liu, Y Meng, C Cheng, Z Lu, et al. 2020. Himalayan wolf distribution and admixture based on multiple genetic markers. Journal of Biogeography. 1272-1285.

Zhang W, Z Fan, E Han, R Hou, L Zhang, et al. 2014. Hypoxia adaptations in the gray wolf (Canis lupus chanco) from Qinghai-Tibet Plateau. PloS Genetics 10(7): e1004466.

Krofel M, J Hatlauf, W Bogdanowicz, LAD Cambell, R Godinho, et al. 2022. Towards resolving taxonomic lineages in wolf, dog, and jackal of Africa, Eurasia, and Australasia. Journal of Zoology 316(3): 155-168.

Werhahn G, H Senn, DW Macdonald, C Sillero-Zubiri. 2022. The diversity of genus Canis challenges conservation biology: a review of available data on Asian wolves. Frontiers in Ecology and Evolution 315

Sheikh KM, S Molur. 2004. Status and Red List of Pakistan's Mammals. Based on the Conservation Assessment and Management Plan. 312pp. IUCN Pakistan.

Sharma et al. 2004, Ersmark et al. 2016, Hamid et al. 2019

Roberts TJ. 1977. The Mammals of Pakistan. Oxford University Press, England.

Ersmark E, CFC Klutsch, YL Chan, MHS Sinding, SR Fain, et al. 2016. From the past to the present: wolf phylogeography and demographic history based on mitochondrial control region. Frontiers in Ecology and Evolution. 4: 134.

Hamid A, T Mahmood, H Fatima, LM Hennelly, F Akrim, et al. 2019. Origin, ecology, and human conflict of gray wolf (Canis lupus) in Suleman range, South Waziristan, Pakistan. Mammalia 83(6): 539-551

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawaoto K, Buckler ES, et al. 2011. A robust, simple Genotype-by-Sequencing (GBS) Approach for high diversity species. Plos One 6(5): e19379.

Li H, R Durbin. 2011. Inference of human population history from individual whole-genome sequences. *Nature* (475) 493-496.

Smeds L, I Kojola, H Ellegren. 2019. The evolutionary history of grey wolf Y chromosomes. Molecular Ecology 28(9): 2173-2191.

Li H, B Handsaker, A Wysoker, T Fennell, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):207802079.

Korneliussen TS, A Albrechtsen, R Nielsen. 2014. ANGSD: Analysis of next generation sequencing data. BMC Bioinformatics 14(356)

Meisner J, A Albrechtsen. 2018. Inferring population structure and admixture proportions in low-depth NGS data. Genetics 210(2): 719-731.

Skotte L, TS Korneliussen, A Albrechtsen. 2013. Estimating individual admixture proportions from next generation sequencing data. Genetics 195: 693-702.

Sarwar G, AM Khan, FI Abbas, MT Waseem, LM Hennelly. 2021. First record of body morphometrics and chemical immobilization of wolves from Pakistan. Pakistan Veterinary Journal 41(4): 1-8.

Rafaqat I, T Mahmood, LM Hennelly, F Akrim, IU Zaman, H Fatima, et al. 2021. Phylogenetic statis and diet of red fox (Vulpes Vulpes griffithii) inhabiting Ayubia National Park, Pakistan. Brazilian Journal of Biology 84(2024).

Masroor R. 2011. An annotated checklist of amphibians and reptiles of Margalla Hills National Park, Pakistan. Pakistan Journal of Zoology 43(6):1041-1048.

Rais M, MZ Khan, D Abbass, G Akber, R Nawaz, M Islam. 2011. A qualitative study on wildlife of Chotiari Reservoir, Sanghar, Sindh, Pakistan. Pakistan Journal of Zoology 43(2): 237-247.

Supple MA, B Shapiro. 2018. Conservation of biodiversity in the genomics era. Genome Biology

Saad M, M Anwar, M Waseem, M Salim, Z Ali. 2015. Distribution range and population status of Indian gray wolf (Canis lupus pallipes) and Asiatic jackal (Canis aureus) in Lehri Nature Park, District Jhelum, Pakistan.

Rais M, MZ Khan, D Abbass, G Akber, R Nawaz, M Islam. 2011. A qualitative study on wildlife of Chotiari Reservoir, Sanghar, Sindh, Pakistan. Pakistan Journal of Zoology 43(2): 237-247.

Supple MA, B Shapiro. 2018. Conservation of biodiversity in the genomics era. Genome Biology 19(131)

Saad M, M Anwar, M Waseem, M Salim, Z Ali. 2015. Distribution range and population status of Indian gray wolf (*Canis lupus pallipes*) and Asiatic jackal (*Canis aureus*) in Lehri Nature Park, District Jhelum, Pakistan. The Journal of Animal and Plant Sciences. 25:433-440.

6. Main Figures

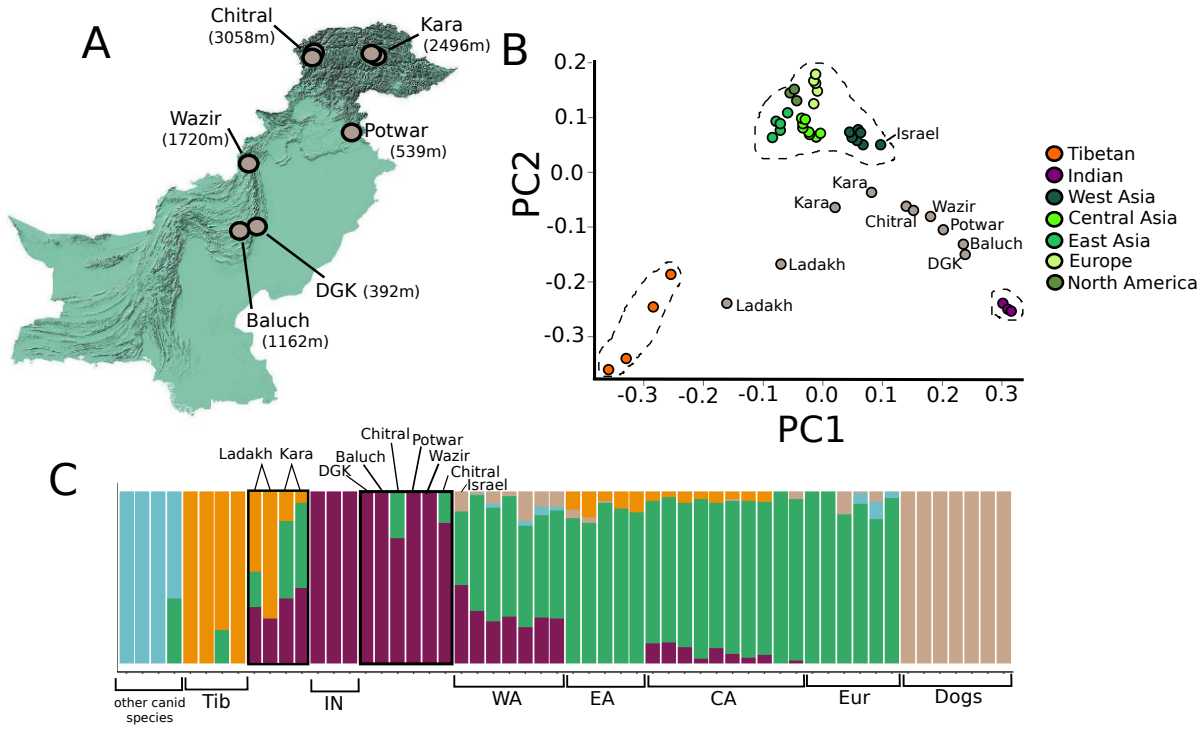


Figure 3.1. Gray wolf samples from Pakistan and population structure of Eurasian gray wolf populations. (A) Locations of gray wolf samples in Pakistan, which include wolves from Dera Ghazi Khan (DGK), Sulieman mountain range in Kohlu Baluchistan (Baluch) and South Waziristan (Wazir), Kallar Syedan on the Potwar plateau (Potwar), the Hindu Kush mountain range in Chitral-Gol National Park (Chitral), the Karakoram Mountains (Kara). (B) Principal component analysis consisting of 74,984 autosomal SNPs and 45 gray wolves showing the three lineages – Indian, Tibetan, and Holarctic –enclosed by a dashed polygon. Gray wolves that fall in among these lineages and that are situated on contact zones are shown in gray, which include our newly sequenced wolves from Pakistan and two wolves from Ladakh region in India. (C) Individual admixture proportions using 135,033 autosomal SNPs across 45 gray wolves, 7 dogs, and four other canid species. Each bar represents an individual and colors within each bar represents the estimated ancestry belonging to a specific population. Labels indicate the following populations: Tibetan (Tib), Indian (IN), West Asia (WA), East Asia (EA), Central Asia (CA), and Europe (Eur). The gray wolf labeled "Israel" indicates the newly sequenced gray wolf sample from Israel in the study.

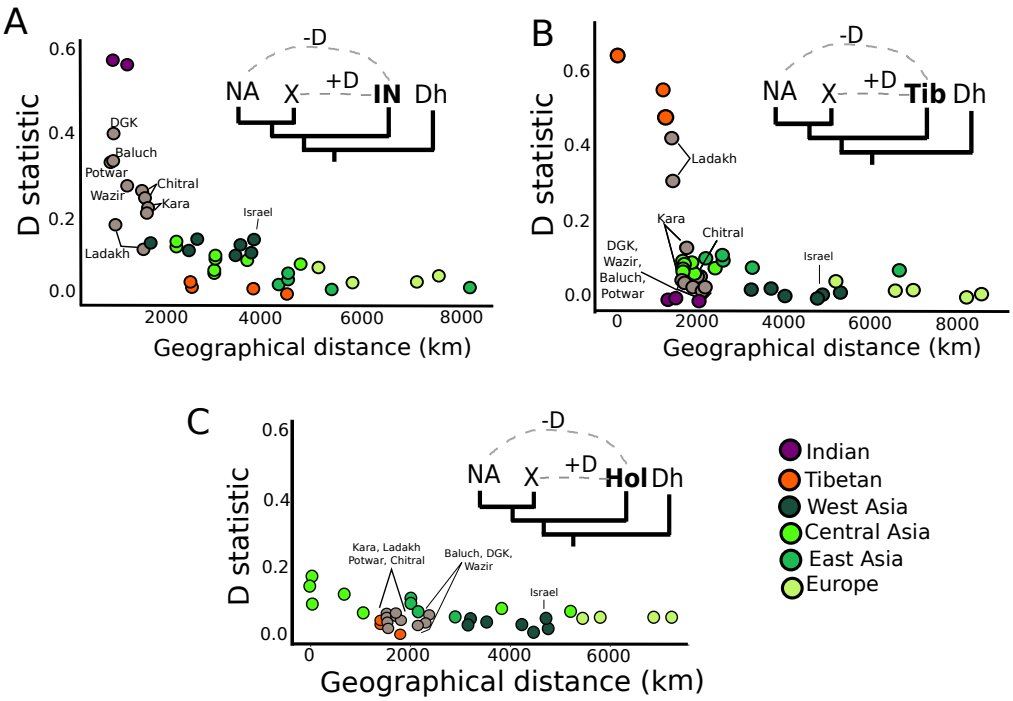


Figure 3.2. Relationship of gene flow (D-statistic) and pairwise geographic distance between each of 45 wolves (X) and (A) an Indian (IN), (B) a Tibetan (Tib), and (C) a Holarctic (Asia, Europe) wolf, in relationship to a North American gray wolf (NA) as sister taxon and dhole (Dh) as outgroup. (A) Gene flow between an Indian wolf and other Eurasian wolves shows a steep drop, but continuous relationship, with geographic distance. (B) Gene flow between a Tibetan wolf and other Eurasian wolves is low at all distances except with other Tibetan wolves. (C) Gene flow between a Holarctic wolf from Central Asia (Xinjiang, China) and other Eurasian wolves is highest with other Central Asian wolves and declines with distances. The D statistics were inferred using 88,148 SNPs. Most gray wolves from Pakistan appear to have high allele sharing with the Indian lineage but not with the Tibetan lineage; wolves from Pakistan were from Dera Ghazi Khan (DGK), Sulieman mountain range in Kohlu Baluchistan (Baluch) and South Waziristan (Wazir), Kallar Syedan on the Potwar plateau (Potwar), the Hindu Kush mountain range in Chitral-Gol National Park (Chitral), and the Karakoram Mountains (Kara). The gray wolf labeled "Israel" indicates the newly sequenced gray wolf sample from Israel in the study.

7. Supplemental Material

Table S3.1 Sample information of Genotype-by-Sequencing and whole genome-sequenced canids, including 7 newly sequenced using GBS canids in this study and 77 additional genomes assembled from short-read archives in GenBank.

Population	Species	WGS vs GBS	SRA number/GBS ID	Country of Origin	Location	Elevation (m)	Depth of GBS samples
Karakoram		WGS	S19-10678	Pakistan	Central Karakoram National Park, Pakistan, near Jutal Village	2496	
Karakoram		WGS	S19-10679	Pakistan	Central Karakoram National Park, Pakistan	2496	
Indus River plains		WGS	S19-10680	Pakistan	outskirts of Dera Ghazi Khan, Punjab, Pakistan	392	
Sulieman		GBS	S21-1037.merged.R1.fastq.gz	Pakistan	Kohlu, Baluchistan, Pakistan	1162	9.75
Hindu Kush		GBS	S21-1038.merged.R1.fastq.gz	Pakistan	Chitral Gol National Park, Khyber Pakhtunkhwa, Pakistan	3058	25.55
Potwar Plateau		GBS	S21-1039.merged.R1.fastq.gz	Pakistan	Kallar Syedan, Punjab, Pakistan	539	50.08
Sulieman		GBS	S21-1041.merged.R1.fastq.gz	Pakistan	South Waziristan, Khyber Paktunkhwa, Pakistan	1720	37.79
Hindu Kush		GBS	S21-1042.merged.R1.fastq.gz	Pakistan	Chitral Gol National Park, Khyber Pakhtunkhwa, Pakistan	32.39	19.19
West Asia		GBS	S21-2743.merged.R1.fastq.gz	Israel	Arabian wolf from Israel		129.86
Central Asia	Gray wolf	WGS	WLF179	Kyrgyzstan	Pikertyk, Kyrgyzstan		
Central Asia	Gray wolf	WGS	WLF180	Kyrgyzstan	Sarychat, Kyrgyzstan		
Indian	Indian wolf	WGS	SRR14777846	India	Outskirts of Jodhpur, Rajasthan, India		
Indian	Indian wolf	WGS	SRR14777843	India	Baramati, Maharashtra, India		
Indian	Indian wolf	WGS	SRR13985171	India	Velavadar Blackbuck National Park, Gujarat, India		
Tibetan	Tibetan wolf	WGS	SRR11085395	China	Ganzi, Qinghai		
Tibetan	Tibetan wolf	WGS	SRR14777848	India	Domkhar, Sham, Ladakh, India		
Tibetan	Tibetan wolf	WGS	SRR14777847	India	Outside of Kargil, Ladakh, India		
Tibetan	Tibetan wolf	WGS	SRR7107906	China	Luobulingka Zoo, Tibet (wild born), China		

Tibetan	Tibetan wolf	WGS	SRR7107907	China	Xining Zoo, Qinghai (wild born), China
Tibetan	Tibetan wolf	WGS	SRR7107910	China	Xining Zoo, Qinghai (wild born), China
Tibetan	Tibetan wolf	WGS	SRR7107912	China	Kashi Garden, (wild born), Tibet, China
West Asia	Gray wolf	WGS	SRR12009566	Iran	Kerman, Iran
West Asia	Gray wolf	WGS	SRR12009567	Iran	Terhan, Iran
West Asia	Gray wolf	WGS	SRR12009568	Iran	Hamadan, Iran
West Asia	Gray wolf	WGS	SRR8049193	Saudi Arabia	Breeding Centre for Endangered Arabian Wildlife, Saudi Arabia
West Asia	Gray wolf	WGS	SRR8049194	Syria	Breeding Centre for Endangered Arabian Wildlife, Syria
West Asia	Gray wolf	WGS	SRR2149870	Israel	Neve Ativ, Golan Heights, Israel
Europe	Gray wolf	WGS	SRR7107782	Portugal	Minho, Portugal
Europe	Gray wolf	WGS	SRR2149873	Croatia	Perkovic, Croatia, wild caught
Europe	Gray wolf	WGS	SRR7107785	Spain	Castilla y Leon, Spain
Europe	Gray wolf	WGS	SRR7107779	Italy	Calabria, Italy
Europe	Gray wolf	WGS	SRR7107647	Russia	Bryansk, Russia
Central Asia	Gray wolf	WGS	SRR2827611	China	Xinjiang, China
Central Asia	Gray wolf	WGS	SRR7107909	China	Kalamaili Nature Reserve (wild born), Xinjiang, China
Central Asia	Gray wolf	WGS	SRR7107911	China	Duzhishan Garden (wild born), Xinjiang, China
Central Asia	Gray wolf	WGS	SRR2827609	China	Xinjiang, China
Central Asia	Gray wolf	WGS	SRR7107645	Russia	Altai, Russia
East Asia	Gray wolf	WGS	SRR7107908	China	Haerbing Zoo, Inner Mongolia (wild born), China
East Asia	Gray wolf	WGS	SRR7107913	China	Haerbing Zoo, Inner Mongolia (wild born), China
East Asia	Gray wolf	WGS	SRR7107646	Russia	Chukotka, Russia
East Asia	Gray wolf	WGS	SRR2827600	China	Lioning, China
East Asia	Gray wolf	WGS	SRR2827603	China	Shanxi, China
North America	Gray wolf	WGS	SRR8049197	Canada	10km south of Gunnar Island, Ellesmere Island, Canada
North America	Gray wolf	WGS	SRR7107786	USA	Yellowstone National Park, USA
North America	Gray wolf	WGS	SRR7976431	USA or Mexico	Mexican wolf
	Dhole	WGS	SRR8049189		Berlin Zoo, Germany
	Ethiopian wolf	WGS	SRR8049190	Ethiopia	Ethiopia
	Golden Jackal	WGS	SRR14777842	India	Uttarakhand, India
	African wolf	WGS	SRR7976426	Kenya	Kenya

Coyote	WGS	SRR3574870	North America	North America	
Dog	GBS		Pakistan	Dog, Motor way Islamabad 5km before Thalian interchange, Punjab, Pakistan	18.145
Dog	WGS	SRR1061818	Egypt	Dog from Egypt	
Dog	WGS	SRR1061964	India	Village dog from India	
Dog	WGS	SRR1061963	India	Village dog from India	
Dog	WGS	SRR10596312	Australia	Dingo	
Dog	WGS	SRR12009563	Iran	Saluki dog from Iran	
Dog	WGS	SRR12009565	Iran	Qahderijani dog from Iran	

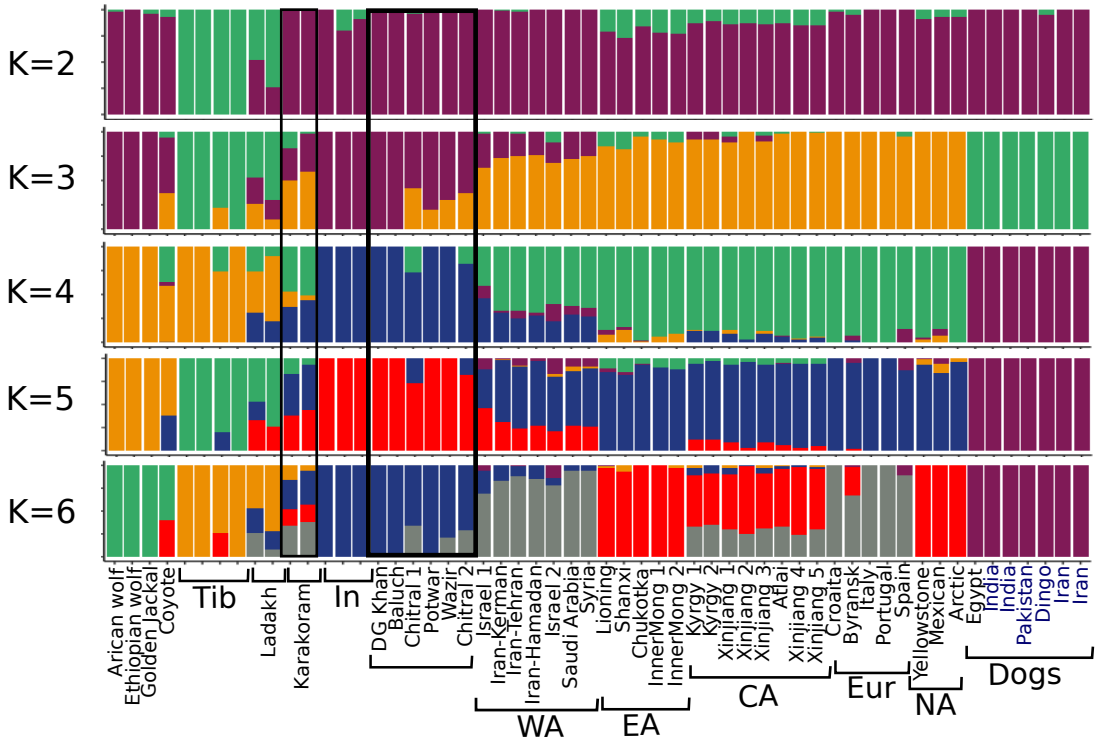


Figure S3.1. Individual admixture proportions using K=2 to K=6 for wolf-like canids with a focus on newly sequenced gray wolves from Pakistan. We used ~135,000 SNPs and 56 individuals to estimate individual admixture proportions through NGSadmix with genotype likelihoods inferred using ANGSD. Each bar represents an individual and colors within each bar represents the estimated ancestry belonging to a specific population. Gray wolf populations used in the analysis include Tibetan (Tib), Indian (In), West Asia (WA), East Asia (EA), Central Asia

(CA), Europe (EA), and North America (NA). Black rectangle around individuals indicate newly sequenced gray wolves from Pakistan.