Received: 29 January 2021 | Revised: 24 June 2021 | Accepted: 13 August 2021

DOI: 10.1111/mec.16127

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

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

Lauren M. Hennelly[1] | Bilal Habib[2] | Shrushti Modi[2] | Eli K. Rueness[3] | Philippe Gaubert[4] | Benjamin N. Sacks[1,5]

[1]Mammalian Ecology and Conservation Unit, Veterinary Genetics Laboratory, School of Veterinary Medicine, University of California, Davis, Davis, California, USA

[2]Department of Animal Ecology and Conservation, Wildlife Institute of India, Dehradun, Uttarakhand, India

[3]Department of Biosciences, University of Oslo, Oslo, Norway

[4]Laboratoire Evolution et Diversité Biologique (EDB), CNRS/UPS/IRD, Université Toulouse III Paul Sabatier – Bâtiment 4R1, Toulouse cedex 9, France

[5]Department of Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, Davis, California, USA

Correspondence
Lauren M. Hennelly, Mammalian Ecology and Conservation Unit, Veterinary Genetics Laboratory, School of Veterinary Medicine, University of California, Davis, CA, USA.
Email: lmhennelly@ucdavis.edu

Funding information
UK Wolf Conservation Trust; Wildlife Institute of India Grant-in Aid; National Science Foundation, Grant/Award Number: 1650042; Department of Science and Technology within the Ministry of Science and Technology of India; Explorers Club

Abstract

The grey wolf (*Canis lupus*) expanded its range across Holarctic regions during the late Pleistocene. Consequently, most grey wolves share recent (<100,000 years ago) maternal origins corresponding to a widespread Holarctic clade. However, two deeply divergent (200,000–700,000 years ago) 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 four Indian and two Tibetan wolves and included 31 additional canid genomes to resolve the phylogenomic history of grey 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 grey 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 grey 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.

KEYWORDS
Canidae, *Canis lupus*, gene flow, grey wolf, phylogenomics, recombination

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 & 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 grey wolf (*Canis lupus*) exemplifies this problem. A mitochondrial perspective suggests that most grey wolves in the Northern Hemisphere today, hereafter the "Holarctic lineage," originate from one or more massive late-Pleistocene (<100,000 years ago) 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 grey 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 grey wolves are restricted to southern regions of Asia: the Indian subcontinent and Tibetan plateau (Figure S1) (Aggarwal et al., 2003, 2007; Sharma et al., 2004). Indian and Tibetan wolf matrilines are estimated to have diverged from the Holarctic grey wolf clade up to 350,000 and 715,000 years ago, respectively (Sharma et al., 2004; Wang et al., 2020; Werhahn et al., 2018). Tibetan wolves face multiple threats in various regions of their distribution and Indian wolves are thought to number <3,000 individuals within India (Hennelly et al., 2015; Jhala, 2003; Suryawanshi et al., 2013).

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 grey 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 (Figure 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 grey 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 grey 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 & Payseur, 2012; Butlin, 2005; Marin & Jiggins, 2017). Regions of the genome with high recombination can more rapidly decouple selectively neutral loci from deleterious loci, and therefore tend to harbour proportionally more introgressed ancestry (Nachman & Payseur, 2012). Conversely, regions of the genome with low recombination tend to preserve the historical branching order of taxa (Pease & Hahn, 2013). Low recombination regions also have lower effective population sizes (N_e) and, therefore, more thoroughly sorted lineages than higher-recombination regions (Pease & 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; Li et al., 2019; Manuel et al., 2020; Martin et al., 2019; Schumer et al., 2018).

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 (Bergero & Charlesworth, 2009; Muirhead & Presgraves, 2016; Presgraves, 2008). 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 four Indian and two 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 broadscale 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 | MATERIALS AND METHODS

2.1 | Laboratory and bioinformatic procedures

DNA was extracted from four Indian wolves, two Tibetan wolves, and one Asian golden jackal comprising five blood and two tissue samples using the DNeasy Blood and Tissue Kit (Qiagen) 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). A data set of canid species was assembled by including published samples from NCBI Sequence Read Archive (SRA), in addition to our newly sequenced samples (Table S1). All raw reads were trimmed using Trim Galore v0.6.5 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 v3.8.3 (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 v2.18.25 in Picard tools (http://broadinstitute.github.io/picard/), and sorted the bam files using SAMtools v1.9. 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 v0.7.17.r1188 (Li & Durbin, 2009). After alignment, PCR duplicates were identified and removed using the MarkDuplicates tool v2.18.25 from the Picard suite. Our BAM files were then sorted and filtered to keep only properly paired reads (-F 1024) using SAMtools v1.9 (Li et al., 2009). We called SNPs using freebayes v9.9.2 (Garrison & 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 paralogues from our data set. Lastly, we estimated the average depth of our BAM files using the command depth (--depth) in SAMtools v1.9 (Li et al., 2009).

2.2 | Mitochondrial phylogeny

After assembling mitogenomes from our samples and grey wolves, African wolves, golden jackals, Ethiopian wolf, coyotes, and dholes acquired through NCBI (Table S2), 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 data set into non-coding (tRNA and rRNA) and the first, second and third 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 & Rambaut, 2007): one using only the third codon positions to estimate divergence times while minimizing bias due to purifying selection on nonsynonymous sites (Subramanian & 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 grey wolves: HKY+I for first codon position and non-coding, TN93+I for second codon position, and TN93+G for the third codon position (Loog et al., 2020). For the third-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 (v2.5.2.) 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 grey wolves (Holarctic, Indian, Tibetan), we filtered the variant calling format (VCF) file in Plink (v1.90), 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 grey wolves and domestic dogs. We estimated individual ancestry proportions with Admixture (v1.3.0) using default setting (Alexander & Lange, 2011). In data sets 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 grey wolves relative to other canid species, we extracted the autosomes and the X chromosome from our data set 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 data set using ModelFinder, as implemented in IQ-tree v1.6.12 (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 1,000 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/).

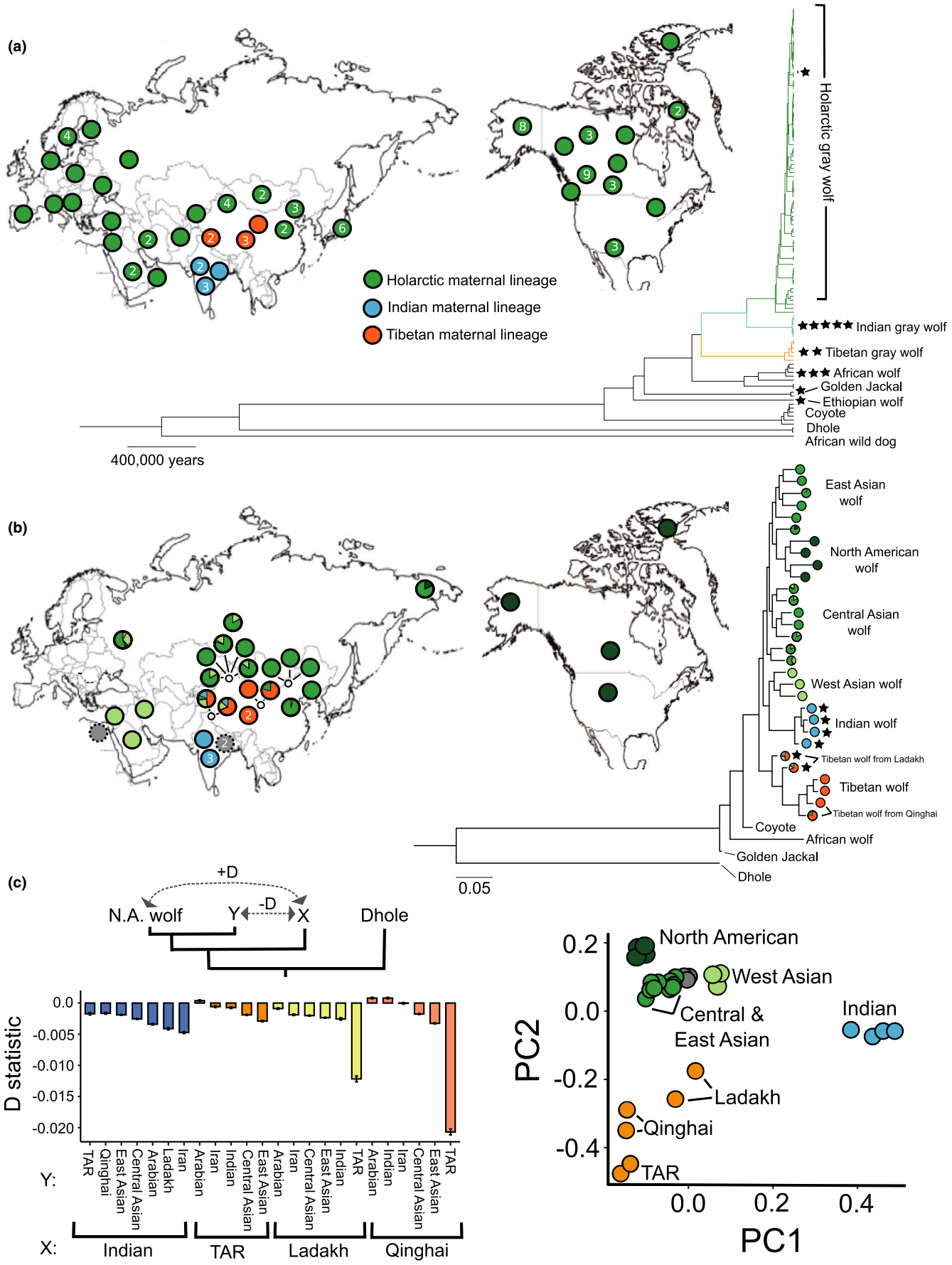
To disentangle incomplete lineage sorting and gene flow, we performed D-statistic analysis for the autosomes using ADMIXTOOLS (Patterson et al., 2012) and calculated f3 statistic using threepop in Treemix to test for introgression across the autosomes (Pickrell & Pritchard, 2012). We converted the plink 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 grey 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 grey 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 grey wolf populations: North American, Central Asian, East Asian, Tibetan (excluding Ladakh and Qinghai), and Indian (as defined in Table S1). Specifically, we generated five admixture population models that included these primary grey 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 grey 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 & Van Belleghem, 2017). We excluded the two Tibetan wolves from Ladakh due to their admixed origins. Genotypes of the remaining grey 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 data set was phased to infer haplotypes from SNP genotypes using Beagle (beagle.16May18.771.gar file) with default parameters. We constructed local neighbour-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: three target populations (Tibetan wolf, Indian wolf, Holarctic wolf) and four 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 recombination rate ≤0.2 cM/Mb, and used these

FIGURE 1 Geographic locations of samples with mitochondrial and nuclear genomic profiles. (a) Locations of 87 complete grey wolf mitogenomes and phylogenetic tree of these and 19 additional samples from six 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 & 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 grey wolf (coloured circles) and dog (grey circles) whole genome sequences and autosomal phylogeny using 30 grey wolves and four canid species constructed using IQ-tree (v1.6.12) with using 1000 ultrafast 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. Colours of the sample location circles correspond to population assignment results from the admixture analysis at $K = 6$ computed in Admixture (Alexander & Lange, 2011) (Figure 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 grey wolves and three dog samples (shown as grey circles) [Colour figure can be viewed at wileyonlinelibrary.com]


(a)
Holarctic maternal lineage
Indian maternal lineage
Tibetan maternal lineage
Holarctic gray wolf
Indian gray wolf
Tibetan gray wolf
African wolf
Golden Jackal
Ethiopian wolf
Coyote
Dhole
African wild dog
400,000 years
(b)
East Asian wolf
North American wolf
Central Asian wolf
West Asian wolf
Indian wolf
Tibetan wolf from Ladakh
Tibetan wolf
Tibetan wolf from Qinghai
Coyote
African wolf
Golden Jackal
Dhole
0.05
(c)
+D
-D
N.A. wolf
Y
X
Dhole
D statistic
0.0
-0.005
-0.010
-0.015
-0.020
Y:
TAR
Qinghai
East Asian
Central Asian
Arabian
Ladakh
Iran
Indian
X:
Indian
TAR
Ladakh
Qinghai
North American
West Asian
Central & East Asian
Indian
Ladakh
Qinghai
TAR
PC1
PC2
0.2
0.0
-0.2
-0.4
0.4

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 four Indian wolves, two Tibetan wolves, and an Asian golden jackal from India (Table S1). We first assembled mitogenomes of these samples, along with a grey wolf from Kyrgyzstan, three 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 preassembled in the GenBank nucleotide database (Figure 1a; Table S2). Similar to previous mitochondrial analyses (e.g., Sharma et al., 2004), Indian and Tibetan lineages were basal to Holarctic grey wolves. Based on a third-codon position tree calibrated to the African wild dog (*Lycaon pictus*), Indian and Tibetan wolves shared most recent common ancestors (MRCA) with Holarctic wolves 200,000 years ago (95% HPD: 175,000–307,000 years ago) and 496,000 years ago (388,000–644,000 years ago), respectively (Figure S3). 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 six wolf and golden jackal samples, along with 30 additional canids (Table S1), to the domestic dog genome (canFam3.1). After trimming adapters and removing paralogues 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). 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 grey wolves and admixture analysis revealed that Indian and Tibetan wolves were genomically distinct from each other and from Holarctic wolf populations (Figures1b,c; Figure S4). 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 (Figure S4). 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 (Figure 1b; Figure S4; Tables S3, S4). 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 grey wolves. Wolves from the centre 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 (Figure S4).

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 (Figure 1c; Tables S3, S4). Otherwise, admixture among the four Holarctic populations (West Asian, Central Asian, East Asian, North American) was consistent with a continuous isolation-by-distance relationship (Figure S5).

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 grey wolves (Figure 1b; Figure S6). 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 (Figure 1c).

Based on the qpGraph analysis of the five primary grey 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 (Figure S7). Moreover, adding Ladakh, Qinghai, and West Asian wolves to this topology allowed us to quantify admixture (Figure 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*; Figure 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 & Belleghem, 2017). On autosomes, the average weights were highest for the topology with the Indian wolf as basal (TC) but differences

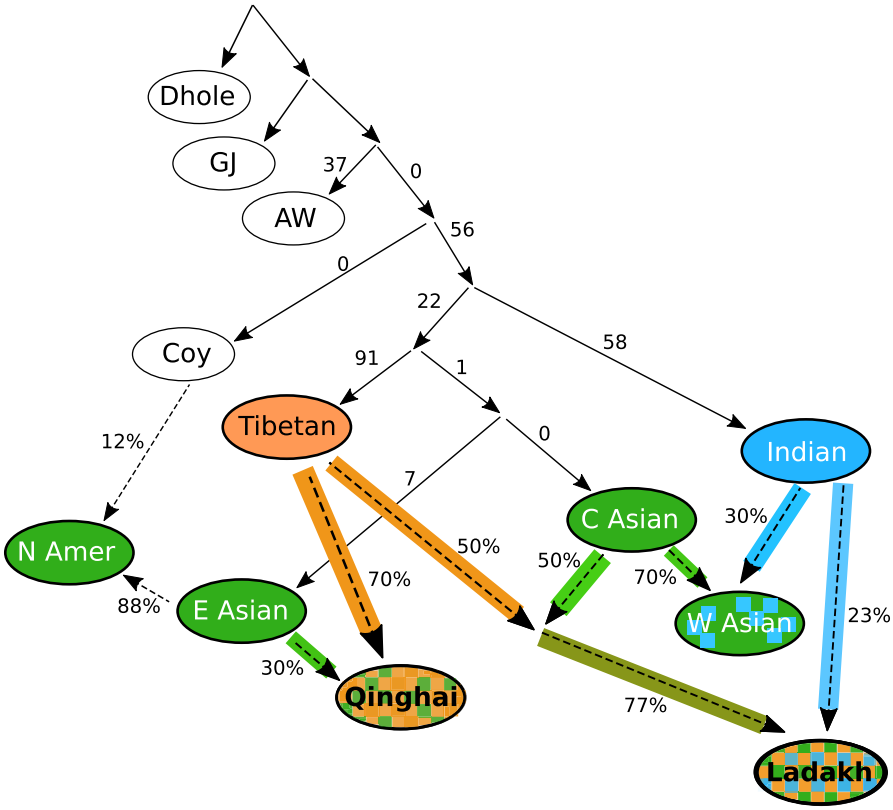


FIGURE 2 Graph illustrating estimates of admixture in wolves from Ladakh, Qinghai, and West Asia (W Asia) with respect to five primary wolf populations arranged according to the best supported topology (Figure S7). The admixture graph was inferred using qpGraph within the ADMIXTOOLS package (Patterson et al., 2012) and approximately 30 million autosomal SNPs. Percentages and coloured 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 Figure S7 but with Ladakh, Qinghai, and West Asian wolf populations and their admixture relationships added, as shown [Colour figure can be viewed at wileyonlinelibrary.com]

were slight (Figure 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) (Figures S8a,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 & 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 (Figure 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; Figures S9a–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 ($p << 0.001$; Figure S9). Together, these findings suggest that the Indian wolf lineage is the most ancestral of the grey wolves, followed by that of the Tibetan wolf.

To more directly address the positioning of North American grey 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 four taxa (plus outgroup) with 15 possible topologies (Figure 3d). Although differences among the average

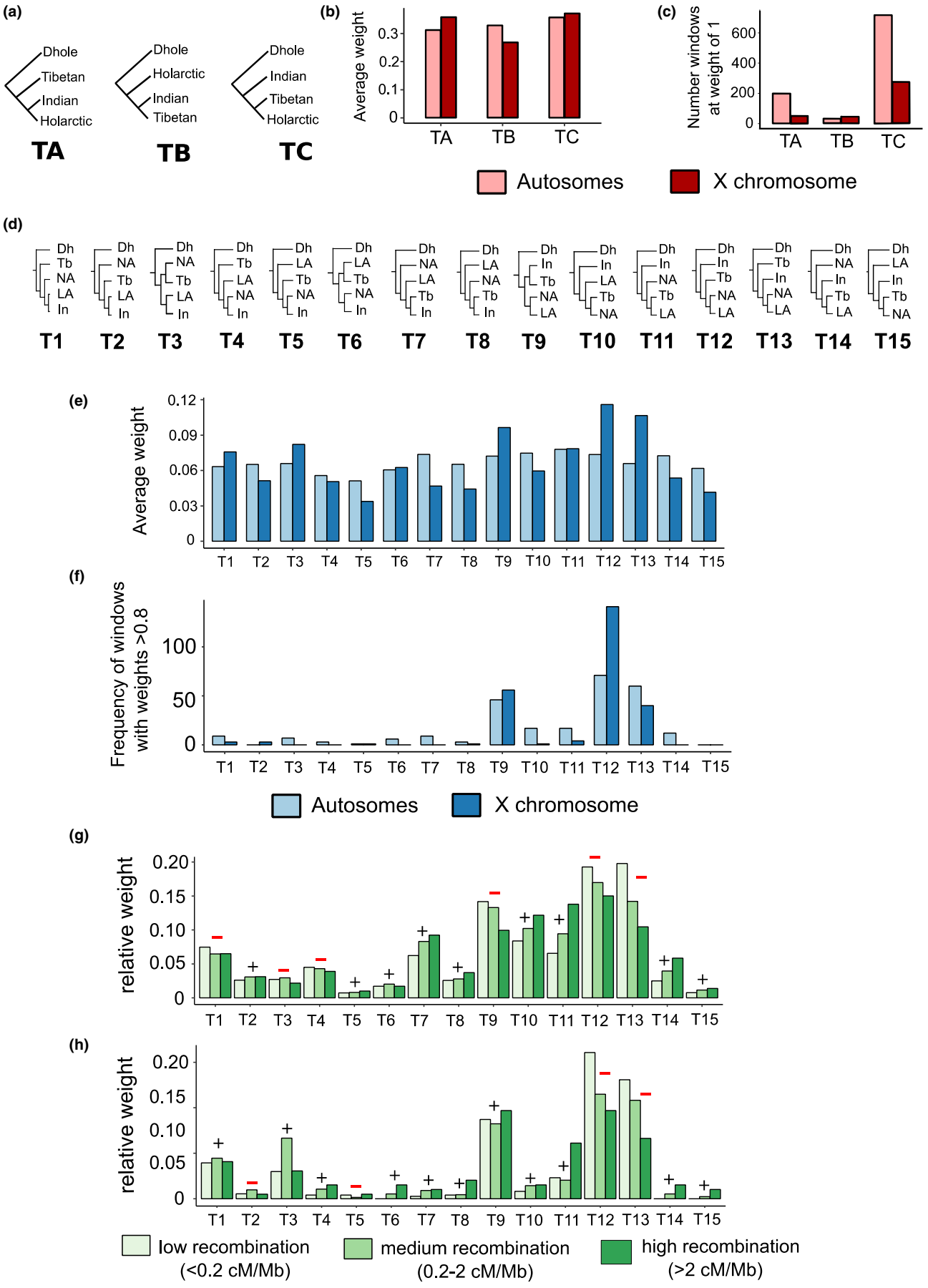

(a)
Dhole
Tibetan
Indian
Holarctic
TA
Dhole
Holarctic
Indian
Tibetan
TB
Dhole
Indian
Tibetan
Holarctic
TC
(b)
Average weight
0.3
0.2
0.1
0
TA TB TC
(c)
Number windows at weight of 1
600
400
200
0
TA TB TC
Autosomes
X chromosome
(d)
T1 T2 T3 T4 T5 T6 T7 T8 T9 T10 T11 T12 T13 T14 T15
(e)
Average weight
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 grey wolves rooted to the dhole (Dh). (a) Three possible topologies composed of three grey wolf taxa used, (b) their average topology weights, and (c) frequencies of windows with weights of 1. (d) Fifteen possible topologies composed of four grey wolf taxa as above, but with Holarctic divided into lowland Asian (LA), and North American (NA) grey 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 (0.2–2 cM/Mb), and high recombination (>2 cM/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) [Colour figure can be viewed at wileyonlinelibrary.com]

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 (Figure 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 (Figure 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; Figures S9d–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 three-group topologies, pairwise Kolmogorov-Smirnov tests among T7, T12, and T13 topology weight distributions were all highly significant ($p << 0.001$; Figure S9). Thus, both Holarctic Asian and North American wolves formed the most derived lineages despite their 12,000 years ago 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; Manuel et al., 2020). Consistent with this prediction, the three topologies (T9, T12, T13) ranked highest above (Figure 3f) also exhibited negative relationships between frequency of high-weighted topologies and recombination rates for autosomes and the X chromosome (Figure 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 (Figures S8c–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 (Figure 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 (Figure S10), the low-recombination X chromosome strongly supported the Indian wolf as the most ancestrally diverging wolf lineage (Figure 4, Figure S8).

4 | DISCUSSION

We investigated the genomic distinctiveness of two southern Asian wolf populations that were previously found to exhibit divergent grey wolf matrilines. Previous phylogenomic studies of the grey 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, 2020). Here, we used grey 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 grey 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.

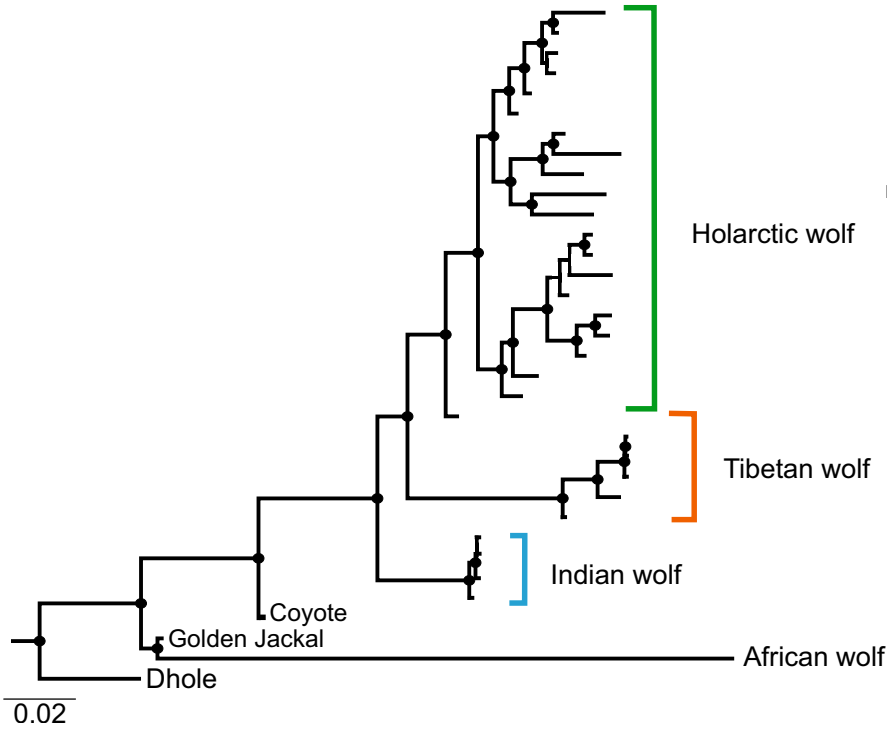


FIGURE 4 Phylogenetic tree of the X chromosome inferred using low recombination regions (<0.2 cM/Mb) across 30 grey wolves and 4 other canid species. The phylogeny was constructed using IQ-tree (v1.6.12) with using 1000 ultrafast bootstrap approximation (UFBoot) and a TVM+F+R3 substitution model inferred 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 [Colour figure can be viewed at wileyonlinelibrary.com]

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 harboured significant admixture with nearby Indian and Holarctic wolves, consistent with the mitochondrial evidence pointing to this region as a three-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 (Hennelly et al., 2017; Lydekker, 1900; 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 probably 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 (Blinkhorn et al., 2015; Jana & Karanth, 2020; Rohland et al., 2005). Thus, ancient gene flow probably connected Indian wolves and Holarctic wolves of West Asia, possibly during the contraction of the Thar desert during interglacials 80,000–130,000 or 27,000–60,000 years ago (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. Mitonuclear discordance is not uncommon in mammals (Toews & 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., 2013; 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 centres for grey wolf evolution (Aggarwal et al., 2003; Sharma et al., 2004). All grey 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 (Ersmark et al., 2016; Leonard et al., 2007; 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 grey wolves inhabited southern Asia continuously throughout the late Pleistocene (Costa, 2017; Dayan et al., 1992; Kurten, 1965; Mashkour et al., 2008; Wang et al., 2016). 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 50,000 years ago 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 (Aradhya et al., 2017; Roberts et al., 2014; Yang et al., 2009). Together, these observations suggest that southern Asia contained at least two long-term refugia for grey wolves. Such a pattern, whereby southern regions in Asia harbour 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*) (Lan et al., 2017; Rueness et al., 2014; Statham et al., 2014), 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.

5 | 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" (Wang et al., 2020; Werhahn et al., 2018, 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 grey 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 2000 to 3000 individuals in India, with an unknown number of individuals in a declining population in Pakistan (Jhala, 2003; Sheikh & 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 (Castello, 2018; Hamid et al., 2019; Sharma et al., 2004). 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 grey 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 grey wolf populations.

ACKNOWLEDGEMENTS

We thank the Forest Department Government of Maharashtra and Jammu and Kashmir for permission to B. Habib for his projects on Ecology of Wolves in Maharashtra and Ladakh. We are

also thankful to Director, Dean and Research Coordinator of the Wildlife Institute of India. We appreciate Nucleome Informatics Private Limited for their assistance and support with sequencing Indian and Tibetan wolf genomes. We also thank The Grasslands Trust in Maharashtra for support and field work logistics. In addition, we thank N. Lescureux (Muséum National d'Histoire Naturelle de Paris, France), G. Dobigny (Centre de Biologie pour la Gestion des Populations, Montpellier, France), and S. Dufour and C. Bloch (Sylvatrop, France) for collecting canid samples. Laboratory work of P. Gaubert was conducted at Institut des Sciences de l'Evolution de Montpellier on the GenSeq platform of LabEx Centre Méditerranéen Environnement et Biodiversité (CeMEB). This work was supported by the Department of Science and Technology within the Ministry of Science and Technology of India for funding the Maharashtra Wolf Project and the Wildlife Institute of India Grant-in Aid for funding the research project on wolves in Ladakh. Additional funding was provided by UK Wolf Conservation Trust and the Explorers Club Student Grant and the Mammalian Ecology and Conservation Unit at UC Davis. L.M. Hennelly was supported by the National Science Foundation Graduate Research Fellowship (award no. 1650042) and a University of California, Davis Graduate Group of Ecology Fellowship. We also thank R. Bay and 3 anonymous reviewers for providing helpful comments on earlier drafts of the manuscript.

AUTHOR CONTRIBUTIONS

L.M.H. conceived and designed the study, compiled and analysed data, interpreted results, and drafted the manuscript. B.H. collected samples, and edited the manuscript. S.M. helped prepare samples for sequencing, and edited the manuscript. E.K.R. provided sequencing data, and edited the manuscript, P.G. provided genetic samples and sequencing effort, laboratory work, and edited the manuscript. B.N.S. conceived and designed the study, interpreted results, and drafted the manuscript.

DATA AVAILABILITY STATEMENT

Scripts used within this paper can be found on the associated GitHub public repository: https://github.com/hennelly/Ancient-Divergence-Indian-Tibetan-Wolves-Paper. The newly sequenced whole genome reads of one Asian golden jackal, four Indian wolves, and two Tibetan wolves used in this manuscript are available at the NCBI Sequence Read Archive (SRA) under the BioProject PRJNA736617 with Genbank Accession Nos SRR14777842- SRR14777848. The mitogenomes of 13 newly sequenced canids are available on NCBI under accession numbers MZ433367-MZ433379.

ORCID

Lauren M. Hennelly https://orcid.org/0000-0002-8805-2504
Benjamin N. Sacks https://orcid.org/0000-0003-0143-6589

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 Zoological 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 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., Klütsch, C. F. C., Chan, Y. L., Sinding, M.-H., Fain, S. R., Illarionova, N. A., Oskarsson, M., Uhlén, M., Zhang, Y.-P., Dalén, L., & 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. https://doi.org/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., Jiang, X., Hall, A. B., Catteruccia, F., Kakani, E., Mitchell, S. N., Wu, Y.-C., Smith, H. A., Love, R. R., Lawniczak, M. K., Slotman, M. A., Emrich, S. J., Hahn, M. W., & Besansky, N. J. (2015). Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, *347*(6217), 1258524. https://doi.org/10.1126/science.1258524

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, Y. V. (2003). Status, ecology, and conservation of the Indian wolf *Canis lupus pallipes Skyes*. *Journal of the Bombay Natural History Society*, *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.

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. *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. *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.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Hennelly, L. M., Habib, B., Modi, S., Rueness, E. K., Gaubert, P., & Sacks, B. N. (2021). Ancient divergence of Indian and Tibetan wolves revealed by recombination-aware phylogenomics. *Molecular Ecology*, 30, 6687–6700. https://doi.org/10.1111/mec.16127