Complex genomic ancestry in southern regions and drivers of continental-level genetic diversity in the wolves of Asia

Authors

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

[1]Section for Evolutionary Hologenomics, Globe Institute, University of Copenhagen, Denmark

[2]Center for Conservation Genomics, Smithsonian's National Zoo and Conservation Biology Institute, Washington, DC, USA

[3]Organisms and Environment, School of BiosSciences, Cardiff University, Cardiff, UK

[4]Department of Molecular Biology and Genetics, Koçc University, İlstanbul, Türkiye

[5]School of Biological Sciences, University of Utah, Salt Lake City, 84112, UT, USA

[6]Kuzey Doğa Society, Kars, Türkiye

[7]Ural Federal University, Yekaterinburg, Russia

[8] Institute of Plant and Animal Ecology, Ural Branch of Russian Academy of Science, Yekaterinburg, Russia

[9]Technical University in Zvolen, Zvolen, Slovakia

[10]Department of Zoology, Charles University, Prague, Czechia

[11]Biology and Ecology, University of Ostrava, Ostrava, Czechia

[12]NTNU University Museum, Norwegian University of Science and Technology, Trondheim, Norway

[13]Wildlife Institute of India, Dehradun, India

[14]University of Education – Attock Campus, Attock, Pakistan

[15]University of the Punjab, Lahore, Pakistan

[16]Lebanese Wildlife, El Metn, Lebanon

[17]Cardiff University-Institute of Zoology Joint Laboratory for Biocomplexity Research (CIBR), Beijing, China

[18]Department of Biology, University of Copenhagen, Copenhagen, Denmark

[19] Mammalian Ecology and Conservation Unit, Veterinary Genetics Laboratory, Department of Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, California, USA

Corresponding author: Lauren M. Hennelly laurenmaehennelly@gmail.com

Abstract

Gray wolves (*Canis lupus*) in Asia encompass most of the species' global genetic diversity and many endangered populations. However, a clear understanding of the evolutionary history of wolves from many parts of Asia, especially southern regions, is lacking. We used 98 whole genomes of wolves sampled across Eurasia to better resolve their evolutionary history by investigating phylogenetic and gene flow histories across the genome, and to characterize their demographic history and genetic diversity. The strongest barriers to gene flow coincided with boundaries separating the three major extant wolf lineages - Indian, Tibetan, and Holarctic. Wolves in the central Asian mountain ranges belonged to the Holarctic lineage, and despite their geographic adjacency only shared minimal ancestry with the Tibetan lineage. In contrast, wolves from eastern Asia share population-wide ancestry with the Tibetan lineage, which may reflect an unsampled lineage similar, but not exactly to, the modern-day Tibetan lineage. Wolves from southwestern Asia also share population-wide ancestry with the Indian lineage, likely due to old (>6kya) admixture events. Long-term historical declines over the last 100,000 years, geographic isolation, and recent inbreeding have resulted in the Indian and Tibetan wolves having some of the lowest levels of genetic diversity and highest realized genetic loads. In contrast, adjacent populations exhibit some of the highest genetic diversity, due in part to admixture along contact zones. Our study illustrates how using multiple approaches that consider heterogenous signals across the genome can more fully resolve the historical and contemporary processes that have led to present-day species' diversity.

Keywords: *Canis lupus,* demographic history, genetic diversity, gray wolf, phylogenomics

Introduction

Understanding the processes that shape genetic variation within a species is fundamental in evolutionary biology. Genetic variation is often unevenly distributed across a species' range, and assessing where and how it varies provides insight into underlying evolutionary processes. Because multiple evolutionary processes impact a species' genetic variation, it has long remained a methodological challenge to distinguish between these processes, such as intrapopulation and interspecies gene flow, genetic drift, and incomplete lineage sorting. The contribution of these processes also varies temporally over a species' evolutionary history, further challenging inferences. Recent methodological advances now consider genomes as mosaics of evolutionary history whereby different genomic regions have distinct evolutionary histories (Martin et al. 2017, Excoffier et al. 2021, Harris et al. 2022). Carefully interpreting these distinct signatures across the genome has recently improved our understanding of the evolutionary histories of notoriously complex groups (Figuiro et al. 2019, Lescroart et al. 2023, Jensen et al. 2023).

Widely distributed species offer a unique opportunity to study how genetic variation is shaped by historical and contemporary environmental changes. Because they are so widespread, their populations experience different histories of isolation and

interpopulation or interspecific gene flow, often in response to past and present environmental change. Applying new methodological approaches to widely distributed species can better resolve their evolutionary histories and offer additional insight into how periodic gene flow and isolation in different parts of their distribution has given rise to present-day patterns in genetic variation within species.

Gray wolves are one of the most widely distributed and best-studied large mammals, making them a promising system to explore how historical and contemporary processes shape present-day genetic variation. Due to their scientific and public appeal, they were one of the first large carnivores to be studied using genomic tools (vonHoldt et al. 2011, Stronen et al. 2013). Over the years, genomic work of wolves has grown into large-scale studies using numerous modern and ancient genomes. This has led to critical insights on recognizing the role of hybridization in speciation and impacts of the late Pleistocene on species' persistence (Gopalakrishnan et al. 2018, Bergstrom et al. 2022). However, most such studies have been on North American and European portions of the range. Fewer genomic studies have focused on wolves in Asia, which today holds most of their phylogenetic diversity and many endangered populations.

Gray wolves are composed of three extant divergent lineages – the Indian, Tibetan, and Holarctic – all of which occur in Asia (Wang et al. 2020, Hennelly et al. 2021, Wang et al. 2022). These lineages diverged > 100,000 years ago after which the Indian and Tibetan lineages became isolated within separate glacial refugia (Wang et al. 2020, Wang et al. 2022). In contrast, wolves of the Holarctic lineage, spanning northern latitudes from northern Europe to Siberia, as well as North America, were highly

connected by gene flow during the past 100,000 years (Bergstrom et al. 2022, **Figure S1**). This high gene flow resulted in genome-wide homogenization of wolves in northern Eurasia during the late Pleistocene (Bergstrom et al. 2022). While this has provided a broad picture of wolf evolutionary history in Eurasia, few studies have examined how ancestral composition and genetic diversity is distributed in wolves across Eurasia, especially in more southern regions of Asia where lineages occur in close proximity.

For example, earlier work based on from a handful of wolf genomes from southwestern Asia (defined as the region spanning Iran to the Arabian Peninsula to Türkiye) largely reflected Holarctic ancestry, with some minor ancestry from the Indian lineage, as well as from another canid species, the African wolf (*Canis lupaster*) (Gopalakrishnan et al. 2018, Hennelly et al. 2021). However, a fuller understanding of this population's history has been hampered by a lack of geographically spread samples from across southwestern Asia. For wolves in Japan and parts of China, previous work found they contain partial ancestry similar to ancient Pleistocene wolves from ~35-50kya, all of which fell within the Holarctic lineage based on mitochondrial DNA (Niemann et al. 2020, Madrigal et al. 2021, Segawa et al. 2022). In addition, some of their partial ancestry was found to be ancestral to late Pleistocene wolves from northern Eurasia, including ancient wolves dating back to 100 kya (Bergstrom et al. 2022). Together, this suggests various wolf populations in the southern regions of Asia may retain diversity that is ancestral to northern Eurasian wolves, and thus, were not completely homogenized by gene flow from northern wolves during the late Pleistocene (Bergstrom et al. 2022). It remains unclear what the geographical extent of ancestral diversity

153 (defined as ancestral to Holarctic wolves in the northern latitudes) within these wolf
154 populations is, and the spatiotemporal origin of this ancestry in their genomes. To
155 resolve this, additional samples and analyses are needed to clarify the ancestry
156 composition of wolves in southern regions of Asia, and what historical processes have
157 given rise to modern-day wolves in these southern regions.
158
159 We sought to address these four gaps in our knowledge about wolves in Asia through
160 building a genomic dataset that includes 20 newly sequenced wolves sampled from key
161 regions and 81 previously published wolf genomes from Eurasia and North America.
162 Our first objective was to characterize continental-level population structure and gene
163 flow patterns of wolves across Eurasia, with a focus on the southern regions of Asia.
164 Our second objective was to evaluate the presence and clarify the origin of ancestral
165 diversity (i.e. ancestral to Holarctic wolves in the northern latitudes) for wolf populations
166 in the central Asian mountain ranges and eastern regions of Asia by incorporating
167 recombination rate variation into phylogenetic analyses. Our third objective was similar
168 to objective two, but evaluating ancestral diversity within the southwestern Asian wolf
169 population using phylogenetic analyses and demographic modeling. For our fourth
170 objective, we use our genomic dataset to characterize the genetic diversity of wolves
171 throughout Eurasia, as well as inbreeding and genetic load.
172

173 Results

174
175 For this study, we resequenced new whole genomes of 20 wolves from various regions
176 in Eurasia. These regions included southwestern Asia (Lebanon, n=1; Israel, n=1,

177 Türkiye=1), Indus plains of Pakistan (n=2), central Asian mountain ranges (Sulieman,
178 n=2; Karakoram, n=2; Hindu Kush, n=3), the Eurasian steppe (Kazakhstan, n=3;
179 Russia, n=3), and in Europe (Slovakia, n=1; Ukraine, n=1). We also conducted
180 additional (i.e., deeper) sequencing of a previously sequenced Indian wolf from Central
181 India. We compiled these newly sequenced genomes with 81 previously sequenced
182 wolves from North America (n=3) and Eurasia (n=78), as well as dogs (n=5), and
183 related canid species (n=12) into a dataset, which we then aligned to the domestic dog
184 genome assembly (canFam3.1) (**Table S1**).
185

186 **Population structure and gene flow patterns of wolves in Eurasia**

187 Principal component analysis (PCA) and inferred individual admixture proportions of 98
188 Eurasian wolves primarily supported the presence of three major clusters corresponding
189 to the three major extant wolf lineages: the Indian, Tibetan, and Holarctic wolves
190 (**Figure 1A,1B, S2**). Some wolves fell in-between these main clusters, forming clines
191 between Tibetan and Holarctic as well as between Indian and Holarctic clusters. In both
192 cases, these intermediate individuals came from areas nearby the corresponding
193 contact zones between the lineages. Intermediate individuals on the Tibetan-Holarctic
194 cline included two wolves from Tajikistan (approximately 24% Tibetan ancestry at K=6)
195 and two wolves from Ladakh, which is a region on the southwestern edge of the Tibetan
196 plateau. Intermediate individuals on the Indian-Holarctic cline included only wolves from
197 Pakistan. In contrast, all 13 wolves from southwestern Asia clustered tightly nearby
198 other Holarctic wolves and exhibited ~15-20% Indian wolf ancestry at K=6. A few

199 wolves, most sampled near Africa, showed basal canid ancestry, likely reflecting
200 admixture with originating from the African wolves.
201
202 To better characterize geographic patterns of gene flow, we inferred the estimated
203 effective migration surface (EEMS), which indicates regions of low or high gene flow
204 relative to an isolation-by-distance model (Marcus et al. 2021). Mirroring the PCA and
205 admixture results, the most significant EEMS barriers to gene flow in Eurasia
206 corresponded to divisions among the three major lineages (**Figure 1C**). The strongest
207 barrier was along the Himalayan mountain range that separates the Indian and Tibetan
208 lineages. A second strong barrier surrounds the Tibetan plateau, in agreement with our
209 PCA and admixture results showing reduced gene flow between the Tibetan lineage
210 and wolves in central and eastern Asia. The positioning of another barrier west of
211 Pakistan was less clear, but may be artifactual due to our lack of sampling on the
212 Eurasian steppe, such as in Turkmenistan and Uzbekistan. Otherwise, the population
213 structure of Holarctic wolves across northern Eurasia can be modeled as isolation-by-
214 distance (**Figure 1B,C).**
215

216 Tibetan-like ancestry is rare in the central Asian mountains yet found 217 population-wide in eastern Asia

218 To explore the evolutionary history of wolves in the central Asian mountain ranges and
219 eastern Asia, we investigated the presence and potential source of ancestral diversity
220 (i.e. ancestral to wolves in northern Eurasia) within these two populations. We first used

221 D statistics to test for excess derived allele sharing with various wolf populations in Asia
222 (Green et al. 2010). For wolves in the central Asian mountains, we observed a wide
223 variance of allele sharing with the Tibetan lineage (**Figure 2A**). Apart from a few
224 individuals, most wolves in Pakistan, Tajikistan and Kyrgyzstan show no statistically
225 significant excess of derived alleles shared with the Tibetan lineage (**Figure 2A; Figure**
226 **S3**). A large inter-individual variance in Tibetan derived alleles may suggest that
227 admixture with Tibetan wolves occurs occasionally or is recent, rather than arising from
228 long-term admixture. In contrast, we observe a consistent and significant excess of
229 Tibetan wolf derived alleles across wolves from eastern China and Mongolia (**Figure**
230 **2A, Figure S3, S4**).
231
232 It can be difficult to identify the actual source using D statistics because any ancestry
233 more closely related to the Tibetan wolf will result in a negative D statistic. To better
234 understand the source and historical processes that may have resulted in Tibetan-like
235 ancestry, we inferred a set of phylogenetic trees from different regions of the genome.
236 We first inferred a multispecies coalescent tree in ASTRAL using 5,000 regions with a
237 length of 20kb across the autosomes using all 113 wild canid samples (Zhang et al.
238 2018). Tibetan wolves formed a well-supported branch basal to other Eurasian gray
239 wolves (**Figure 2B**). The rest of the wolves form weakly supported clades, including
240 Indian wolves which are clustered with wolves from Pakistan, and sister to those from
241 southwestern Asia (**Figure S5**).
242
243 Because the relationships inferred by genome-wide phylogenies can be impacted by
244 many processes, such as gene flow and incomplete lineage sorting (ILS), we leveraged

how genome architecture relates to phylogenetic signal. Briefly, the X chromosome and low recombination regions are more likely to retain the historical relationship due to higher selection against introgressed ancestry and higher lineage sorting (Schumer et al. 2018, Martin et al. 2019). In particular, we took advantage of the negative relationship between recombination and historical phylogenetic relationship that has been demonstrated across multiple taxa, including wolves (Li et al. 2019, Hennelly et al. 2021, vonHoldt et al. 2021, Feng et al. 2023, Jiang et al. 2024, Monthey and Spellman 2024).

Our phylogenetic tree based on low-recombination regions (<0.2 cM/.Mb) of the X chromosome demonstrated that one of the two wolves from Tajikistan that showed the highest allele sharing with Tibetan wolves clustered within the Tibetan lineage, and all other wolves from Tajikistan and Kyrgyzstan clustered with other Holarctic wolves (**Figure 2C, Figure S6, Figure S7**). Wolves from eastern Asia formed a well-resolved clade that branched basally to the rest of the Holarctic lineage. Our phylogeny using low recombination regions (<0.2 cM/Mb) of the autosomes and the entire X chromosome recapitulates this pattern (**Figure S8, Figure S9**). In addition, topology weighting analyses show low recombination regions of the X chromosome are more resolved for topologies where wolves from eastern Asia form the basal branch within the Holarctic lineage (**Figure 2D, Figure S10**). Thus, while wolves in Tajikistan, Pakistan, and eastern Asia have similar D-values, they have contrasting topological patterns. This is consistent with wolves from Tajikistan and Pakistan having experienced recent and/or periodic gene flow from Tibetan wolves near the secondary contact zone, yet this

Tibetan ancestry has not yet spread widely through the population. In contrast, wolves in eastern China and Mongolia have retained population-wide ancestry that is more ancestral than wolves in northern Eurasia, yet appears distinct from modern-day Tibetan wolves.

Population-wide Indian ancestry within wolves of southwestern Asia

We detected complex ancestry patterns in wolves across southwestern Asia. Estimated admixture proportions and D statistics suggest that wolves in this region shared ancestry similar to the Indian lineage and another canid species, the African wolf (**Figure 1A,B, Figure 3A,B**). We detected nine wolves that showed significant derived allele sharing with the African wolf, which were largely confined to areas most geographically close to Egypt (Saudi Arabia, Syria, Lebanon) (**Figure S11, S12**).

Interestingly, we find ancestry shared with the Indian lineage was nearly equal in all wolves in southwestern Asia (**Figure 3A, S12, S13, S14**). There appeared to be no geographic decline in the admixture proportions of assigned Indian ancestry spanning the contact zone in Pakistan until reaching the Caucasus mountain range and from Türkiye into Europe, where no Indian wolf ancestry was detected (**Figure 3A, Figure S14**). There was also a lesser, yet significant, signal of ancestry similar to the Indian lineage on parts of the Eurasian steppe in Central Asia, such as in Kazakhstan (**Figure S14**). This geographically widespread signal contrasts what we observed with Tibetan and African wolf ancestry, which declines quickly as distance increases from their respective distributions (**Figure 3A**).

292
293 The balanced Indian wolf-like ancestry across southwestern Asia could be consistent
294 with an admixture event between Indian and Holarctic lineages. To investigate the origin
295 of this Indian-like ancestry, we used FASTSIMCOAL2 (Excoffier et al. 2021) to test three
296 demographic scenarios where (A) there is no introgression event between Indian
297 wolves and southwestern Asian wolves, (B) southwestern Asian wolves experience an
298 introgression event with Indian wolves, and (C) southwestern Asian wolves are formed
299 as a hybrid lineage between Indian and central Asian wolves (**Figure S15**). We
300 obtained similar likelihood scores for the three models, suggesting no one model is
301 strongly favored over the others (**Table S2**). However, topologies that included
302 bifurcation (Divergence and Introgression models) had the highest likelihoods. The
303 model that included introgression from Indian wolves to southwestern Asian wolves was
304 preferred, with a relative likelihood of 62% (**Table S2**). Estimates suggest a relatively
305 high introgression proportion (0.11) from India to southeast Asia, consistent with our D
306 statistic results (**Figure S14**). We estimate the introgression occurred around 6,000
307 years ago, relatively close to the divergence time between southwest and central Asian
308 populations about 10,300 years ago. Therefore, if introgression did occur, it likely
309 happened near the time of the split between southwest and central Asian populations,
310 explaining the similar likelihood between the divergence (A) and introgression (B)
311 models. Importantly, both models estimate similar dates for the split between
312 southeastern and central Asian lineages and all models date the divergence of Indian
313 wolves from Eurasian wolves to around 100,000 years before present, providing
314 robustness to time estimates. Additionally, both introgression and divergence models
315 produced similar estimates for population sizes, suggesting a small effective population

316 size (N_E) for India and southwest Asia, whereas central Asian populations and the
317 ancestral Asian population had a larger N_E (**Figure S15**). Our results indicate that the
318 Indian wolf ancestry within the genomes of southwestern Asian wolves reflects older
319 admixture (>6kya) and/or shared demographic history rather than appearing within the
320 past few last thousands of years.
321
322 Finally, to better assess how this Indian wolf-like ancestry relates to modern-day Indian
323 wolves, we inferred a set of phylogenetic trees to investigate the topological positions of
324 Indian and southwestern Asian wolves. In almost all inferred autosomal and X
325 chromosome trees, we observe the Indian lineage and wolves from Pakistan cluster
326 together, with almost all southwestern Asian wolves as a sister clade – all within the
327 Holarctic clade (**Figure S5, S6, S7, S8**). The clustering of the Indian lineage with wolves
328 in Pakistan implies they share more recent or prevalent ancestry, consistent with this
329 region being a prominent contact zone (Hennelly et al. 2023). When excluding wolves
330 from Pakistan and southwestern Asia, Indian wolves form the most basal branch of the
331 gray wolf phylogeny, in line with southwestern Asian wolves being the admixed
332 population (Wang et al. 2022, **Figure S17**). To investigate Indian-like segments within
333 southwestern Asian wolf genomes in more detail, we identified genomic regions across
334 the autosomes where Indian and southwestern Asian wolves were sister to each other
335 using topology weighting with Twisst (Martin and Belleghem 2017). For this, we
336 quantified the frequency of 105 possible topologies using 100-SNP windows across the
337 genome for five populations with the Andean fox (*Lycalopex culpaeus*) as an outgroup
338 (**Table S3**). The topology with the highest average weight was one where southwestern

339 Asian wolves were sister to Indian wolves, suggesting some of their genome shares a
340 more recent common origin rather than with other Holarctic wolves (**Figure 3C, Table**
341 **S3**).

342

343 We then inferred a maximum likelihood tree using only genomic regions that are most
344 reflective of Indian and southwestern Asian wolves being sister clades (>0.3 topology
345 weight for the highest supported topology; **Figure 3C**). In this maximum likelihood
346 phylogeny, we find the most basal wolf lineage consists of Indian and southwestern
347 Asian wolves, where southwestern Asian wolves fall outside of the diversity within the
348 modern-day Indian lineage (**Figure 3D, 3E, S18, S19**). One hypothesis for this finding is
349 that Indian-like ancestry within southwestern Asia may derive from a once more diverse
350 Indian lineage rather than arising directly from modern-day Indian wolves within the
351 Indian subcontinent. Alternatively, there could be some Holarctic ancestry within these
352 100-SNP windows, thus placing southwestern Asian wolves outside of modern-day
353 Indian wolf diversity. Our work emphasizes a complex evolutionary history of wolves in
354 southwestern Asia where additional studies are necessary to fully resolve it.

355
356
357
358 **Historical and present-day impacts on genetic diversity**
359
360 Our whole-genome dataset allowed us to assess patterns of genome-wide diversity of
361 wolves across Eurasia and North America. We calculated autosomal heterozygosity
362 using genotype likelihoods for 101 wolves from Eurasia (n=98) and North America
363 (n=3), and found wolves in Asia encompass a wide range of some of the lowest and

364 highest levels of genetic diversity among wolves in our dataset (**Figure 4A**). Along with
365 the well-known depleted heterozygosity of Mexican and Iberian wolves, we found
366 wolves in India and on the Tibetan plateau have some of the lowest documented
367 heterozygosity in our dataset. As observed in our Pairwise Sequentially Markovian
368 coalescent (PSMC) analysis, this remarkably low heterozygosity is in part due to their
369 long-term population declines in the last >100 kya rather than, or in addition to,
370 anthropogenic-related declines (Li and Durbin 2011) (**Figure 4B**). We also observe that
371 the wolf from Shanxi, China, had a lower population size throughout the Pleistocene
372 compared to other Holarctic wolves (**Figure 4B**). This may suggest the Shanxi wolf and
373 other wolves in southern China were less connected to northern wolf populations,
374 however, a greater number of higher coverage genomes are needed to rule out the
375 effects of low coverage. Otherwise, there is a shift to elevated heterozygosity coinciding
376 with secondary contact zones between the Holarctic lineage and the Indian and Tibetan
377 lineages, respectively. For example, admixed wolves in lowland Pakistan have higher
378 genome-wide heterozygosity than wolves of Central India, even though both
379 populations are similarly endangered and declining. We find the highest heterozygosity
380 among wolves in Asia is geographically clustered in Iran and Xinjiang of China (**Figure**
381 **4A**). These findings are concordant with our estimated inbreeding coefficients using
382 genotype likelihoods (Hanghoj et al. 2019) (**Figure S20).**
383
384 To further disentangle what historical and present-day processes may explain patterns
385 of genetic diversity in wolves, we inferred runs of homozygosity (ROH) for genomes with
386 moderate coverage (>15x) using BCFtools/RoH (Narasimhan et al. 2016). The lengths
387 of ROHs allowed additional insight into the timing of inbreeding. Combined with

388 knowledge of the generation time and recombination rate, we estimated the timing of
389 inbreeding, where, for example, a ROH length of 5Mb suggests inbreeding took place
390 ~7 generations ago (Thompson 2013). We find that Indian and Tibetan wolves show the
391 most extensive inbreeding between 800-102 years compared to all other wolves in our
392 study (**Figure S21**). Along with the highest number of short ROHs (0.1-1Mb), Indian and
393 Tibetan wolves show evidence of larger ROH as well (5-10Mb), suggesting their low
394 heterozygosity is a combination of long-term isolation over the last 200kya and
395 anthropogenic pressures in the last hundreds of years (**Figure 4, Figure 3B**). We find
396 the well documented and dramatic population declines of Mexican and Iberian wolves
397 occurred more recently over the course of the last century, in line with previous work
398 (**Figure S21**, Taron et al. 2021, Gomez-Sanchez et al. 2018). Likewise, we observe a
399 few individuals in Asia possess long ROHs (5-100Mb), implying the slightly reduced
400 genetic variation in these wolves is due to recent declines in the last 100 years (**Figure**
401 **S21**). Thus, in contrast to genetic diversity patterns in Europe, our results suggest much
402 of the variation in genetic diversity in Asia can be explained by historical processes,
403 coinciding with glacial refugia and secondary contact zones.
404
405 Because Indian and Tibetan wolves show low genetic diversity, we explored how
406 another component of genetic diversity – functional genetic diversity – compares to
407 other wolves across Asia. We were specifically interested in quantifying patterns of
408 deleterious variation that is expected to have a negative impact on fitness. Using SnpEff
409 to estimate functional impact of mutations, we quantified the total genetic load (total
410 number of derived alleles classified as high or moderate impact), realized load
411 (homozygous state of derived alleles in high or moderate impact), and masked load

412 (heterozygous state of derived alleles in high or moderate impact) for five selected wolf
413 populations with assuming the derived alleles was the deleterious one (Cingolani et al.
414 2012, Bertorelle et al. 2022). Wolves from historically large population sizes – central
415 and southwestern Asia – had higher masked and total genetic load than wolves from
416 historically smaller populations (Indian, Tibetan, Mexican wolves) (**Figure 5A,B, S22,**
417 **S23**). This is consistent with theoretical predictions that larger populations with higher
418 genetic diversity should also possess more deleterious alleles in a heterozygous state
419 (Mathur and DeWoody 2021). In contrast, we found that Indian, Tibetan, and Mexican
420 wolves have a higher realized load and a lower total genetic load (**Figure 5,C**) than the
421 other surveyed populations. A higher realized load in these small populations is
422 expected, because genetic drift and inbreeding will result in an increased proportion of
423 homozygous genotypes, including deleterious ones.
424
425 A lower total genetic load in the Indian and Tibetan wolves is intriguing, as this may
426 reflect a loss of derived variants due to stronger genetic drift in historically small
427 populations and/or purging (Dussex et al. 2023). Purging predominantly occurs when
428 recessive deleterious variation is exposed to purifying selection in a homozygous state.
429 To look for explore evidence of stronger genetic drift versus purging, we asked whether
430 deleterious alleles are lost at a higher frequency than the background neutral alleles
431 (mutations in non-coding regions) for each individual wolf. In other words, does the fact
432 that derived alleles are categorized as deleterious result in more frequent loss compare
433 to neutral derived alleles within an individual, as expected for purging via purifying
434 selection? Or are deleterious derived alleles lost at the same frequency as neutral

435 derived alleles, suggesting the loss in deleterious derived alleles found in our total
436 genetic load results could be due to genetic drift? We assessed the prevalence of sites
437 associated with four impact categories: high impact (loss of function, LOF), moderate
438 impact (missense), low impact (synonymous), and modifier (non-coding; here assumed
439 to be selectively neutral). To account for differences in total derived alleles found in wolf
440 individuals, we divided the total number of LOF, missense and synonymous derived
441 mutations by the total number of neutral derived alleles assumed to be under neutral
442 processes (mutations in non-coding regions) for each individual.
443
444 We found Indian, Tibetan, and Mexican wolves have proportionally fewer missense and
445 synonymous mutations per 100,000 derived neutral mutations compared to wolves in
446 central and southwest Asia. A reduced frequency of synonymous mutations may reflect
447 stronger genetic drift associated with a smaller long-term population size rather than
448 purifying selection via purging, as synonymous mutations are assumed to be nearly
449 neutral (**Figure 5F,G**). We emphasize that it is difficult to differentiate evidence of
450 genetic drift and purging, and further work is needed to verify the functional impacts of
451 our missense and synonymous mutations. For high impact mutations, we observe no
452 clear pattern among individuals (**Figure 5E**). One reason for this could be incorrectly
453 annotated LOF mutations in the short-read assembled dog reference, as suggested by
454 Smeds et al. (2023). In addition, among the ones that may be correct, there are so few
455 LOF sites across the genome (n=145 on average in our dataset) and being such a small
456 number, their frequency may be more strongly influenced by genetic drift, as seen with
457 killer whales (Kardos et al. 2023). Overall, we suggest caution when interpreting our
458 genetic load results. Further work is needed to assess the accuracy of annotations from

459 a short-read assembled genome, functional impacts of mutations, and to increase the
460 number of individuals included in each population.
461
462 # Discussion
463

464 In this study we conducted a comprehensive whole genome analysis on population
465 structure, ancestral composition, and genetic diversity of wolves in Asia. We find that
466 the strongest barriers to gene flow in Eurasia coincides with the Indian, Tibetan, and
467 Holarctic lineage distributions. We show that Holarctic wolf populations found in
468 southern regions of Asia possess genetic variation ancestral to wolves in Northern Asia
469 and that it is similar in the degree of its antiquity to that of the Tibetan and Indian
470 lineage. Despite evidence of recurrent gene flow across northern Eurasia during the late
471 Pleistocene, it appears this process of homogenization did not completely reach areas
472 in southern regions of Asia (Bergstrom et al. 2022). Lastly, for wolves in Asia we find
473 genetic diversity largely corresponds to historical processes, such as long-term isolation
474 within glacial refugia, contact zones, and long-standing environmental differences,
475 followed by anthropogenic processes.
476

477 Asia is the biogeographic cradle for gray wolf evolution, so it is not too surprising that
478 Asia harbors ancestral diversity in its southern regions. Previous work has also shown
479 that southern regions contain deeper local ancestry than both modern and some
480 Pleistocene wolf genomes from Northern Eurasia. Wolves from southwestern Asia,
481 eastern parts of China, and Japan contain ancestry that is more ancestral than Siberian
482 wolves living during the Pleistocene from 12 kya and even 100 kya (Bergstrom et al.

483 2022). In addition, a single wolf from Shanxi of eastern China had significant allele
484 sharing with ancient Siberian wolves from 50kya, 48kya, and 33kya (Ramos-Madrigal
485 2021). Recent work has emphasized the importance of carefully interpreting
486 Introgression analyses and to consider the possibility of “ghost lineages”, that is extinct
487 or unsampled lineages (Tricou et al. 2022). For eastern Asian wolves, we detected
488 significant introgression with the Tibetan lineage as P3, but not with the older diverged
489 Indian lineage or African wolf as P3. While this suggests the origin of this partial
490 ancestry within eastern Asian wolves is similar to modern-day Tibetan lineage, the
491 degree to which the partial ancestry is similar to the Tibetan lineage remains unclear.
492 Our phylogenetic trees provide some clarity. None of our trees nor most frequent
493 topologies using Twisst show eastern Asian wolves clustering within or sister to the
494 Tibetan lineage. Instead, low recombination regions on the X chromosome show they
495 form their own lineage that diverges before the rest of the Holarctic wolves, yet after the
496 Tibetan lineage. Thus, it is possible that wolves in southern China may contain ancestry
497 of an unsampled wolf lineage that was present in southern China and was distinct from
498 the Tibetan lineage.
499
500 Consistent with this, previous work shows historical genomes of wolves in Southern
501 China and Tibetan wolves occupy opposite ends of the PC2 axis, suggesting distinct
502 partial ancestries (Wang et al. 2019). Fossils of wolves have been present across most
503 regions of China over the last 100,000 years, including 12,000 year old remains from
504 the Southern province of Jiangxi (Wang 2016). During this time, much of southern
505 China was drier with mixed broadleaf forests occupying where monsoon rain forests are

506 now (Pinxian and Xiangjun 1994). Thus, it is plausible that ancestry that was distinct
507 from the Tibetan lineage could have existed within a glacial refugium in Southern China,
508 and has not been completely replaced by expansion waves from northern Eurasian
509 wolves. An interesting future avenue of research is whether partial ancestry of extinct
510 Honshu wolves may also be from this glacial refugia, and thus share a common origin
511 with wolves of southern China. Incorporating ancient genomes from eastern China and
512 the newly sequenced 35 kya Japanese wolf genome in future analyses would be
513 needed to place this ancestry into a spatiotemporal context during the Pleistocene.
514
515 An unexpected result in our study was uncovering population-wide ancestral diversity
516 present across southwestern Asia. Our analyses point towards the widespread
517 ancestral component is likely related to the Indian lineage, rather than from the African
518 wolf. This fits well with previous work suggesting part of the deep ancestry within
519 southwestern Asia is at least 100,000 years old, which is before the estimated
520 divergence time of the Indian lineage (Bergstrom et al. 2022, Wang et al. 2022). When
521 and how the present-day population structure formed in southwestern Asia remains
522 unclear, but our study offers some insights. Combined with the fact there have been no
523 Indian mitogenomes detected outside of the Indian subcontinent, our demographic
524 history results from fastsimcoal2 and PSMC finds no strong evidence that the Indian-like
525 ancestry within southwestern Asia is due to recent (pre-6kya) gene flow from the Indian
526 lineage. In addition, it appears that the X chromosome, including low recombination
527 (<0.2 cM/Mb) regions, of southwestern Asian wolves are more similar to the Holarctic
528 lineage. Along with our PSMC results, the major genomic background of southwestern
529 Asian wolves does seem to be that of the broader Holarctic lineage, possibly tracing

back to gene flow from northern latitudes. At the same time, we find the Indian-like genomic regions across the autosomes that do not fit into the modern-day diversity of the Indian lineage. These could be remnants of a once more diverse Indian lineage, or, despite our efforts, these genomic regions could contain some Holarctic ancestry that places southwestern Asian wolves nearby but not within Indian wolves. A more thorough analysis that uses multiple local ancestry approaches to classify Indian-like ancestry would be insightful, as the length distribution and its phylogenetic relationships can be more thoroughly assessed. More broadly, ancient remains of wolves have been documented in southwest Asia throughout the late Pleistocene (Kurten 1965, Dayan et al. 1992, Mashkour et al. 2008, Wang et al. 2016). Ancient genomes from these wolf remains in southwestern Asia can directly test hypotheses on whether there were shifts in ancestral composition over the late Pleistocene and if the Indian lineage was once more widespread and genetically diverse.

Together, our work uses a combination of methodological approaches to better resolve the historical and contemporary processes that shape genetic variation within a wide ranging species. For many species during the late Pleistocene, their populations underwent distinct evolutionary trajectories of isolation and expansion in response to shifting climates. The process in which populations don't or do come together, and how often, is a key force in speciation and evolution. Our study's findings of divergent ancestry persisting in southern regions may be common pattern in widespread species, where red foxes (*Vulpes vulpes*), Eurasian otters (*Lutra lutra*), and brown bears (*Ursus arctos*) all show evidence of cryptic divergent ancestry in southern regions of Asia

554 (Statham et al. 2014, Jong et al. 2023, du Plessis et al. 2023). While our study
555 addresses gaps in our knowledge on wolves, many questions remain to be answered.
556 For example, why some wolf populations seem to have been more susceptible to gene
557 flow from northern populations (southwestern Asia, eastern Asia) during the late
558 Pleistocene, while others show little gene flow and prominent contact zones (Indian and
559 Tibetan) is an open question. With so many ancient and modern genomes now
560 available, wolves are poised to serve as a valuable model to study how intra-population
561 gene flow, interspecies hybridization, and isolation during the late Pleistocene has
562 contributed to present-day species' diversity.
563
564 **Conservation Implications**
565
566 Our study has several implications for the conservation of various wolf populations in
567 Asia. For taxonomy, our study supports previous calls that a taxonomic revision is
568 necessary for wolves in Asia (Alvares et al. 2019, Werhahn et al. 2022, Krofel et al.
569 2022). While wolves from India to Türkiye are collectively described as *Canis lupus*
570 *pallipes*, our study emphasizes past work that the ancestral Indian lineage is confined to
571 India and parts of Pakistan (Hennelly et al. 2021, Hennelly et al. 2023). For the wolves
572 in southwestern Asia that are considered "*Canis lupus pallipes*", we find this distinct
573 population may be geographically limited by the Bosphorus strait in Türkiye and the
574 Caucasus mountain range (Figure 1AB, Figure 2A). Further work is necessary to
575 understand the population genetic structure in the arid areas of southern Central Asia,
576 such as the Karakum desert in Turkmenistan and Kyzylkum desert in Uzbekistan. We
577 suggest the morphological and genomic distinctiveness of wolves in southwest Asia

warrants subspecies status, in line with recent recommendations (Werhahn et al. 2022). For Indian and Tibetan wolves, the presence of gene flow barriers and likely reproductive barriers coinciding with lineage boundaries further supports the recognition of subspecies or species status (Werhahn et al. 2022).

Our study underscores the importance of robustly describing wolves in parts of China. Wolves have been documented as South as Guangxi in China, a province that borders Vietnam (Wang et al. 2016, **Figure S1**). In 1936, wolves were reported to exist in parts of Central and South China, but were already less common compared to Northern China (Sowerby 1936). A report in 2003 concluded wolves have largely been extirpated from regions of Guangxi that once occurred there in the 1990s, but they were possibly still present in Nanling National Nature Reserve (Fellowes et al. 2003). In response to their decline in the country, the wolf was recently added in 2021 as a protected species under Class II of the List of Wildlife under Special State Protection of China. To the authors' knowledge, we are not aware of published estimates of the population size and status, or a morphological description, of wolves in Southern parts of China. Given their evolutionary distinctiveness, our work highlights the need for more research on their distribution, genetics, morphology, and population status to conserve what populations remain.

Overall, our study emphasizes that southern regions of Asia hold most of the ancestral genetic diversity within gray wolves, and are also the most endangered. The Indian lineage numbers at ~2,700 individuals in India and the Tibetan lineage only ~3,000 individuals (Jhala et al. 2022, Werhahn et al. 2023). There are less than ~750 wolves

602 left on the Arabian peninsula, a genomically unique but understudied population where
603 only one genome has been sequenced (Bonsen et al. 2024). It is possible that wolves in
604 Southern China are already extinct or very close to it. Wolves in these regions are
605 impacted by numerous human-induced threats, especially for the remaining populations
606 in Southern China and India. Our study highlights the need to prioritize these
607 populations for conservation to preserve the full spectrum of extant wolf diversity.
608
609
610 # Methods
611
612 This study includes 20 new whole genomes from wolves from Pakistan (n=8), Israel
613 (n=1), Lebanon (n=1), Afghanistan (n=1), Türkiye (n=1), Ukraine (n=1), Slovakia (n=1),
614 Kazakhstan (n=3), and Russia (n=3). For all wolf samples from Pakistan and Israel, we
615 extracted DNA from these samples using DNAeasy Blood and Tissue Kit (Qiagen)
616 following the manufacturer's protocol. For the rest of the wolf samples, they were
617 extracted using Thermo Scientific KingFisher instrument following manufacturer's
618 protocol. The Indian wolf from Maharashtra originated from a previous study and was
619 resequenced to a deeper depth following protocols in Hennelly et al. 2021. For the gray
620 wolves from Pakistan and Israel, libraries were sequenced on an Illumina NovaSeq
621 6000 S4 flowcell and paired end at 150-bp (Illumina). For the rest of the newly
622 sequenced wolf samples, they were sent to BGI Copenhagen for library build and
623 sequenced on ⅛ lane each on DNBSEQ at paired end 150. We constructed a dataset
624 comprising an additional 105 previous published canid genomes, resulting in a total of
625 118 canid genomes in the study (**Table S1**).

Alignment, Variant Calling, and Filtering

We used the Paleomix v.1.2.13.2. pipeline to align our raw reads to the dog reference genome Canfam3.1 (Schubert et al. 2014). Specifically, we used the BWA mem algorithm and used the GATK indel realigner to generate an indel-realigned BAM file for each sample. Variant calling was performed using the Genome Analysis Toolkit (GATK) version 4.2.5.0 (McKenna et al. 2010). First, we used GATK Haplotype Caller to perform variant calling on each sample, then used the resulting GVCF files to perform joint genotyping with all canid individuals combined using GenomicsDBImport and GenotypeGVCF in GATK. We obtained a set of high quality single nucleotide polymorphisms (SNPs) using the following filtering steps in GATK: QD < 2.0 || 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). We excluded genomes that were derived from museum samples in the genotype calling, which was only the wolf from Afghanistan.

Principal Component Analysis and Individual Admixture Proportions

To assess genome-wide genetic structure of the autosomes, we conducted a principal component analysis (PCA) using PCangsd (Meisner and Albrecthtsen 2018) with our total set of Eurasian gray wolf genomes. We inferred the PCA using genotype likelihoods from ANGSD, in which we used the SAMtools model (-GL 1) and filtered the dataset by including only properly paired reads (-only_proper_paired 1) from autosomal

649 bam files, excluded reads with excess of mismatches (-C 50) or mapping quality lower
650 than 20 (-minMapQ20), removed transitions (-noTrans 1), and keep bases with a
651 sequence quality above 20 (-minQ 20) (Korneliussen et al. 2014). We also retained
652 SNPs that were present in at least 90% of individuals present (>90 out of xx total; -
653 minInd 90). We obtained the covariance matrix from PCAngsd and used the R function
654 *eigen* to infer the PCA.
655
656 To estimate individual ancestry proportions, we inferred genotype likelihoods using
657 ANGSD and ran NGSadmix (Skotte et al. 2013) using K=2 to K-9 genetic clusters. We
658 used the canids within the genus *Canis* in this analysis, which included an additional 3
659 North American gray wolves, 3 golden jackals, 2 coyotes, 4 African wolves, 1 Ethiopian
660 wolf, 5 dogs (**Table S1**). We used the same criteria in ANGSD as to infer the PCA,
661 however, we retained SNPs that were present in at least 90% of individuals. This
662 resulted in 578,343 SNPs after filtering.
663

664 Assess barriers to gene flow for wolves across Eurasia

665 We inferred the estimated effective migration surface (EEMS) to assess spatial variation
666 in rates of gene flow among gray wolf populations in Eurasia (FEEMS; Marcus et al.
667 2012). The FEEMS approach utilizes a stepping-stone model to calculate genetic
668 dissimilarities between individuals based on spatial and genetic data. These genetic
669 dissimilarities can be visualized to illustrate departures from strict isolation by distance,
670 thereby detecting regions of high (i.e. corridors) and low (i.e. barriers) to gene flow. For
671 the FEEMS, we only used our dataset of 95 Eurasian wolves, where we filtered the

672 dataset in which we kept sits with a minimum allele count of 3 (--mac 3), all of
673 individuals represented at a site (--geno 0), and pruned sites in high linkage
674 disequilibrium by removing each SNP with a r^2 value of greater than 0.5 (--indep-
675 pairwise 50 10 0.5). This resulted in ~8.7 million autosomal SNPs across our Eurasian
676 wolves. For gathering the geographic data for each wolf sample, we had some gray wolf
677 samples that only had location information as the country or province within a country.
678 In these cases, we took the center coordinate within that country or province.
679
680 To run FEEMs, we first constructed a dense spatial grid using an edge width of 5 that
681 covered the entire Eurasian continent. We initially performed a random initialization over
682 the map to fix the estimate of the residual variance. We selected the tuning parameter,
683 lambda, using leave-one-out cross-validation using 20 folds where the lambda with the
684 lowest cross-validation was selected. Our cross-validation test found a lambda value of
685 100 to be the optimal value of lambda (**Figure S24**).

686 Phylogenomic relationships

687 We constructed the autosomal and X chromosome phylogeny that included all wild
688 canid species except dogs and the Ethiopian wolf (**Figure 2B, S5, S6**). We filtered the
689 autosomal and X chromosome VCF to exclude indels, keep sites with at least minimum
690 quality of 30 (--minQ30), keep only biallelic sites, and kept sites that had at least 90%
691 individuals called (-max-missing 0.9). We then converted each vcf file into the phylip
692 format using the script vcf2phylip (Ortiz et al. 2019). We randomly selected 500 and
693 1,000 regions with a length of 20 kb across the autosomes and X chromosome,
694 respectively. We inferred a maximum-likelihood phylogenetic tree for each 20kb region

695 using IQ-Tree 1.6.12 where we estimated the best model using ModelFinder and used
696 1,000 ultra-fast bootstraps to infer each tree (Kalyaanamoorthy et al. 2017, Nguyen et
697 al. 2014). The inferred phylogenetic trees were used as input into ASTRAL 5.7.8, which
698 inferred a species tree from a set of gene trees while taking into account gene tree
699 discordance (Zhang et al. 2018).
700
701 In addition, we inferred phylogenetic trees of the X chromosome and autosomes using
702 only the low recombination regions across the genome. For the X chromosome, we
703 inferred five different runs containing different sets of canid individuals. These runs were
704 of: (1) including all wild canid species except dogs and the Ethiopian wolf, (2) all wild
705 canid species and wolf individuals excluding dogs, the Ethiopian wolf, wolves from
706 Pakistan, and wolves from Ladakh (3) all wild canid species, and wolf individuals
707 excluding dogs, the Ethiopian wolf, wolves from Pakistan and Ladakh, Indian wolves,
708 (4) all wild canid species and wolf individuals excluding dogs, the Ethiopian wolf, wolves
709 from Pakistan, and wolves from southwestern Asia, and (5) all wild canid species and
710 wolf individuals excluding dogs, the Ethiopian wolf, wolves from Pakistan, and two
711 wolves from that clustered with Indian wolves (wolf from Zanjan and Hamadan in Iran).
712 All VCF files were filtered for sites with a minimum quality of 30 (-minQ 30), only biallelic
713 sites, and at least 90% individuals present at each site (--max-missing 0.9). We
714 downloaded the recombination map of the domestic dog from Auton et al. 2013 and
715 used the recombination rate (cM/Mb) in our analyses. We averaged the recombination
716 rates within 100-SNP windows across the autosomes and X chromosome, as generated
717 in a previous study (Hennelly et al. 2021). We selected genomic regions that had an

718 average recombination rate of below 0.2 cM/Mb, and used these positions to create a
719 bed file, which we then extracted these genomic regions from our VCF.
720
721 We inferred low recombination phylogenies for each of the five runs with different
722 samples with IQ-Tree 1.6.12 (**Figure S6, Figure S16, Figure S17A,B,C**). We converted
723 the VCF containing all low recombination regions to a phylip file format using
724 vcf2phylip.py (https://github.com/edgardomortiz/vcf2phylip). We then ran IQ-Tree 1.6.12
725 with estimating the best model using ModelFinder and used 1,000 ultra-fast bootstraps
726 to infer each tree (Kalyaanamoorthy et al. 2017, Nguyen et al. 2014). Second, we
727 inferred low recombination autosomal and X chromosome phylogenies using ASTRAL
728 5.7.8 (**Figure 2C, Figure S7, Figure S8**). We split each SNP dataset into 10 kb regions,
729 and converted each of these regions to a phylip file (Zhang et al. 2018). We then
730 inferred a tree for each 10 kb region using IQ-Tree 1.6.12 with estimating the best
731 model using ModelFinder and used 1,000 ultra-fast bootstraps.
732

733 Assessing genome-wide gene admixture

734 To test for gene flow among various canid lineages and species, we performed D-
735 statistic analyses using AdmixTools (Patterson et al. 2012). We filtered our dataset to
736 exclude missing data (--max-missing 1), kept biallelic sites, and a minimum quality filter
737 of 30 (--minQ30), along with the filters we applied using GATK. We converted plink files
738 into eigenstrat format using the convertf script within ADMIXTOOLS. We calculated the
739 D-statistic and Z-score for modern populations for three different topologies in the
740 format (((P1, P2),P3),P4): (((Norwegian wolf MW005, X wolf); Indian wolf BH123);

741 Andean fox), (((Norwegian wolf MW005, X wolf); Tibetan wolf TI32); Andean fox),
742 (((Norwegian wolf MW005, X wolf); African wolf); Andean fox). The D-statistic value will
743 be negative for an excess of shared derived alleles between P2 and P3, and/or P1 and
744 P4. Alternatively, the D-statistic will be positive for an excess of shared derived variants
745 between P2 and P4, or P3 and P1. For African wolves, we used multiple individuals as
746 a population for P3, which consisted of all four of our African wolf samples from Kenya,
747 Ethiopia, Morocco, and Algeria.
748
749 We also quantified the frequency of different topologies in genomic windows across the
750 genome using Twisst (Martin and Van Belleghem 2017). To prepare the dataset for the
751 Twisst analysis, we first phased the filtered dataset that contained no indels, biallelic
752 sites with a minimum quality of 30 and that were present in at least 90% of individuals
753 using Shapeit v2.r904 (Delaneau et al. 2013). We used the dog genetic map from Auton
754 et al. 2013 as input and a window size of 0.5 Mb. We converted the phased VCF to a
755 geno file using the parseVCF.py script within
756 https://github.com/simonhmartin/genomics_general. We then estimated phylogenies in
757 100-site windows using phyml with a General Time Reversible (GTR) substitution model
758 with phyml_sliding_windows.py within
759 https://github.com/simonhmartin/genomics_general. We summarized the relative
760 prevalence of topologies using the twisst.py script. The set of topologies we
761 summarized consisted of four population groups with Andean fox as an outgroup
762 focused on wolves in Central Asia and Eastern Asia, and five population groups with
763 Andean fox as an outgroup focused on southwestern Asian wolves. These resulting in
764 15 possible topologies for the four population groups, and 105 possible topologies using

765 five population groups. We also quantified the frequency in which topologies occurred in
766 windows of low (<0.2cM/Mb) recombination rate regions. To do this, we partitioned our
767 our averaged recombination rates within 100-SNP windows inferred in Hennelly et al.
768 2021 across the autosomes and X chromosome datasets.
769
770 Finally, we inferred phylogenetic trees using genomic regions within wolves of
771 southwestern Asia that were designated to be Indian ancestry from our twisst results.
772 First, we normalized the Twisst results for all 105 topologies and calculated the mean
773 for each topology. The topology with the highest mean of 0.33 was topology 73 where
774 Indian and southwestern Asian wolves form sister clades as the most well-resolved
775 topology across the autosomes (**Table S3**). We extracted genomic regions that had a
776 topology weight of over 0.2 and 0.3 for topology 73 from a VCF containing all autosomal
777 SNPs using bedtools. This resulted in 299,560 SNPs and 40,887 SNPs found within
778 genomic regions that had over 0.2 and 0.3 topology weight for topology 73, respectively.
779 We converted two VCF datasets, one with genomic regions that had over 0.2 and 0.3
780 topology weight for topology 73, to a phylip file using vcf2phylip.py
781 (https://github.com/edgardomortiz/vcf2phylip). We then ran IQ-Tree 1.6.12 with
782 estimating the best model using ModelFinder and used 1,000 ultra-fast bootstraps
783 (Kalyaanamoorthy et al. 2017, Nguyen et al. 2014) (**Figure 3D,E, Figure S18, Figure**
784 **S19**).
785

786 **Demographic Modeling with fastsimcoal2**

787 In order to better understand the origin of southwestern Asian wolves, we compared
788 three alternative demographic models and inferred demographic parameters under the
789 FASTSIMCOAL2 (v2.7 Excoffier et al 2021), a composite-likelihood method that
790 estimates historical introgression, divergence times and population topologies from the
791 site frequency spectrum using coalescent simulations.

792 We compared three demographic models (**Figure S15**) to better assess the
793 demographic history of southwestern Asian wolves. The included three populations
794 representing India, southwestern Asia and central Asia samples, and were defined as
795 (1) a divergent model with a strictly bifurcating topology separating Indian and Holarctic
796 Asian lineages and latter southwestern Asia from central Asia populations, (2) an
797 introgression model with the same strictly bifurcating topology followed by an
798 introgression event from the Indian to the southwestern Asian population (3) an
799 hybridization model, where southwestern Asian population was originated by
800 hybridization between Indian and central Asian populations. Input files including
801 template and parameter estimation files for the three models, are available at **Figure**
802 **S27.** A summary of all defined parameters and their search ranges are given in **Table**
803 **S4.**

804 We performed genotype calling using BCFtools specifically keeping monomorphic sites
805 (Danecek et al. 2021). We filtered our VCF containing all variant and monomorphic sites
806 by the following parameters, keeping only those loci with >10x coverage, and > 50%
807 missing data and removing sites SNPs violating Hardy-Weinberg equilibrium for
808 heterozygous excess ($p < 0.01$). After filtering, we retained 2,771,059 SNPs and
809 456,437 monomorphic sites, which we used to generate the joint 3D-SFS using the

810 (folded) minor allele frequency spectrum. We used the canFam 3.1 domestic dog
811 assembly as the reference, resulting in the reference being more closely related to
812 Holarctic wolves than to Indian wolves. Thus, we could not determine the ancestral
813 state of each allele and therefore used the minor allele frequency spectrum (folded
814 SFS). The mutation rate was set to 4.5×10^{-9} based on de novo mutations in a pedigree
815 of wolves (Koch et al. 2019). To generate the 3D-SFS while ensuring no missing data
816 and accounting for local linkage disequilibrium (LD) patterns, we down-sampled
817 individuals. Given an estimated recombination rate of 1.34cM/Mb using the
818 recombination map of the dog genome (Auton et al. 2013), we partitioned the genome
819 into independent and unlinked blocks of 2Mb, ensuring that sites within blocks were
820 linked, while those between blocks were unlinked. Within each block obtained, we
821 retained only SNPs without missing data by sampling three, six, and nine individuals
822 from the Indian, southwest, and central populations, respectively, ensuring that each
823 sampled individual had no missing data for that block. FASTSIMCOAL2 estimates a
824 composite likelihood, assuming independence between sites (i.e., it does not account
825 for linkage). Consequently, although the maximized likelihood converges to the correct
826 parameter values, the estimated likelihood approximates the true value when derived
827 from a set of unlinked loci (Excoffier et al., 2013). We utilized the parameter estimates
828 inferred from all retained SNPs to recalculate the likelihoods for each model using only
829 one SNP per block. This recalculation was based on a 3D-SFS derived from a set of
830 potentially independent SNPs, with one SNP selected per block. We used this newly
831 computed likelihood to apply the Akaike Information Criterion (AIC) to compare the
832 three models and computed the relative likelihood of models. For this we calculated

833 delta AIC according to Excoffier et al (2013), which represents the difference between
834 the AIC value of each model and the minimum AIC value among all models. Further we
835 thoroughly examined the parameter values that optimized the likelihood for each model.
836 Specifically, we compared divergence times between Holarctic and Indian lineages, the
837 effective population sizes of each population, and the admixture proportions estimated
838 under each model. This analysis aimed to determine whether the parameter estimates
839 that optimized each model were consistent with the most likely model as indicated by
840 the AIC. In addition, to evaluate whether the best-selected model could accurately
841 reproduce the observed data, we conducted a visual examination comparing the
842 expected and observed marginal one-dimensional SFS, as well as various two-
843 dimensional SFS pairs.
844

845 **Long-term demographic history with PSMC**

846 To infer the historical demography of wolves in Eurasia, we used the Pairwise
847 Sequential Markovian Coalescent (PSMC; Li and Durbin 2011). PSMC uses a
848 coalescent approach to estimate the history of change in effective population sizes over
849 time. We only included the autosomal sequences of each gray wolf individual. We
850 converted each bam file to a fasta-like consensus sequence by first using the mpileup
851 command with SAMtools and subsequently using BCFtools view –c to call variants and
852 vcfultils.pl vcf2fq to convert the vcf file to fastq format (Danecek et al. 2021). We
853 excluded any reads that were less than 20 for minimum mapping quality and minimum
854 base quality (-q 20 -Q 20) and excluded reads with excessive mismatches (-C 50). We
855 also removed sites with more than double or less than a third of the average depth of

856 coverage for each sample. We tested different combinations of parameters to infer the
857 PSMC, which were "`psmc -N25 -t15 -r5 -p 4+25*2+4+6`", "`psmc -N25 -t15`
858 `-r5 -p 2+2+25*2+4+6`" and "`psmc -N25 -t15 -r5 -p 1+1+1+1+25*2+4+6`"
859 following previous studies on gray wolves and recent recommendations for inferring
860 PSMC trajectories (Hilgers et al. 2024). Some samples showed a false peak when using
861 the parameter "`psmc -N25 -t15 -r5 -p 4+25*2+4+6`", therefore, we used the
862 parameter "`psmc -N25 -t15 -r5 -p 1+1+1+1+25*2+4+6`" for our final analyses
863 (**Figure S25**).
864
865 To account for our selected low coverage genomes (15-20x), we estimated the false
866 negative rate (FNR) by a downsampling high coverage gray wolf genome that were in
867 the same geographic region to the specific depth of the low coverage genome. To
868 determine the best FNR, we visually estimated the best correspondence between the
869 PSMC plots with the high coverage (>20x) regional wolf genome and downsampled
870 genomes with various FNR corrections (**Figure S26**). We then applied the best
871 estimated FNR to the low coverage gray wolf genomes to infer their demographic
872 history. We used a mutation rate of 4.5×10^{-9} (Koch et al. 2019) and a generation time
873 of 4.4 years (Mech and Barber-Meyer 2017).
874
875 **Genetic Diversity and Runs of Homozygosity**

876 We calculated the genome-wide heterozygosity for each wolf individual above 5x using
877 ANGSD (Korneliussen et al. 2014). We first estimated the folded site allele frequency
878 with the realSFS program in ANGSD (-doSaf) using genotype likelihoods with the
879 SAMtools model (-GL 1) and using the Andean fox genome to determine ancestral

880 state. We filtered each individual bam file to keep only autosomal chromosomes, only
881 properly paired reads (-only_proper_paired 1), excluded reads with excess of
882 mismatches (-C 50), excluded reads with a mapping quality (-minMapQ 20), excluded
883 reads with a depth less than 4 (-setMinDepth 4), and kept bases with a sequence
884 quality above 20 (-minQ 20). To calculate the genome-wide heterozygosity, we divided
885 the number of heterozygous sites by the total number of sites for each sample.
886
887 We inferred runs of homozygosity (ROH) using BCFtools/RoH (Narasimhan et al. 2016).
888 Along with the GATK filtering criteria, we included only biallelic SNPs that had all
889 individuals present per site (--max-missing 1, --min-alleles 2, --max-alleles 2). We also
890 included SNPs that had a depth of more than one third and less than double of the
891 average depth of coverage for each sample, where we chose samples that were >15x.
892 For the BCFtools/RoH analysis, we fixed the alternative allele frequency to 0.4 (-AFdflt
893 0.4) and used the dog recombination map from Auton et al. 201x to account for
894 recombination hotspots (--genetic-map). We kept ROHs that had a quality score of at
895 least 80 and excluded ROHs that were below 100 kb in length in our analysis.
896
897 To gain insight into the timing that inbreeding occurred, we calculated the expected
898 generations since inbreeding using the length of the ROHs, generation time of 4.4 years
899 (Mech and Barber-Meyer 2017), and an average recombination rate of 1.34 cM/Mb
900 calculated as the average recombination rate across the autosomes from the canfam3.1
901 genome assembly (Auton et al. 2013). Specifically, we used the formula: $g=100(2rL_{ROH})$
902 where g=generation interval, r=recombination rate (cM/Mb), and L=Length of ROH (Mb)
903 (Thompson et al. 2013). For example, a ROH length of 5Mb will have an expected time

904 of coalescence of 7.46 generations ago, corresponding to inbreeding occurring around
905 32 years ago using a wolf generation time of 4.4 years. Likewise, we estimate that a
906 ROH length of <0.16 Mb corresponds to inbreeding occurring >1,026 years ago.
907

908 ## Inbreeding coefficient

909 We inferred the per-individual inbreeding coefficient (F_{ind}) using genotype likelihoods,
910 where the inbreeding coefficient is the probability of identity by descent. We used
911 ANGSD to infer genotype likelihoods and allele frequencies for samples above 4x
912 coverage with the following parameters: -GL 2 -checkBamHeaders 0 -trim 0 -C 50 -baq
913 1 -skipTriallelic 1 -minMapQ 20 -minQ 20 -minInd 88 -uniqueOnly 1 -remove-bads 1 -
914 noTrans 1 -doGlf 3 -doMajorMinor 1 -doMaf 1 -minMaf 0.0001 -SNP_pval 1e-6. We then
915 used NgsRelate v2 with option -F1 to infer the inbreeding coefficients for each Eurasian
916 wolf individual (Hanghoj et al. 2019).
917

918 ## Genetic Load

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

926 depth of 60, a minimum allele count of 2 and a minimum sequence quality (--minQ) of
927 30.
928
929 To polarize our dataset, we followed the method by Smeds et al. (2022) to map two
930 outgroups to the dog reference genome at only sites with a minimum depth of 4 at both
931 outgroup and included nonvariant sites. We used Dhole (*Cuon alpinus*) and Andean fox
932 (*Lycalopex culpaeus*) as our outgroups. We then performed pseudo-haploidization of
933 the outgroup genomes and kept only sites in which the allele agreed in both outgroups
934 (Smeds et al. 2022). We used this to add the ancestral allele information to our
935 annotated VCF, and assumed that the derived allele was the deleterious allele (Smeds
936 et al. 2022).
937
938 We counted the number of derived alleles and whether they were in a homozygous or
939 heterozygous state for variants predicted by SnpEff to have high, moderate, and low
940 impact on protein function. High impact is predicted to cause loss of function or protein
941 truncation (stop gain, frameshift mutation); moderate impact is predicted to change
942 protein effectiveness (missense variant), and low impact is assumed harmless, such as
943 a synonymous variant. We also calculated the following metrics: total load (the total
944 number of derived alleles classified as high or moderate impact), realized load
945 (homozygous state of derived alleles in high or moderate impact), and masked load
946 (heterozygous state of derived alleles in high or moderate impact).
947
948

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

References

Alvares F, Bogdanowicz W, Campbell LAD, Godinho R, Hatlauf J, Jhala YV, Kitchener AC, Koepfli KP, Krofel M, Senn H, Sillero-Zubiri C, Viranta S, Werhahn G. 2019. Old World Canis spp. With taxonomic ambiguity: workshop conclusions and recommendations. Canid News 21

Auton A, Li YR, Kidd J, Oliveira K, Nadel J, Holloway JK, Boyko AR. 2013. Genetic recombination is targeted towards gene promoter regions in dogs. Plos Genetics, 9, e1003984.

Bergstrom A, Stanton DWG, Taron UH, Frantz L, Sinding MHS, Ersmark E, Pfrengle S, Cassatt-Johnstone M, Lebrasseur O, Girdland-Flink L, Fernandes DM, Ollivier M, et al. 2022. Grey wolf genomic history reveals a dual ancestry of dogs. Nature 607: 313-320.

Bertorelle G, Raffini G, Bosse M, Bortoluzzi C, Iannucci A, Trucchi E, Morales ME, Oosterhout C. 2022. Genetic load: genomic estimates and applications in non-model animals. Nature Review Genetics 23: 492-503.

Bonsen GT, Wallach AD, Ben-Ami D, Keynan O, Khalilieh A, Dahdal Y, Ramp D. 2024. Navigating complex geopolitical landscapes: challenges in conserving the endangered Arabian wolf. Biological Conservation 296: 110655.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, RM Davies, H Li. 2021. Twelve years of SAMtools and BCFTools. GigaScience 10(2): giab008

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

Delaneau O, Howie B, Cox AJ, Zagury JF, Marchini J. 2013. Haplotype estimation using sequencing reads. AJGH 93(4): P687-P696

Dussex N, Morales HE, Grossen C, Dalen L, Oosterhout C. 2023. Purging and accumulation of genetic load in conservation. Trends in Ecology and Evolution 38(10): P961-P969.

Exoffier L, N Marchi, DA Marques, R Matthey-Doret, A Gouy, VC Sousa. 2021. Fastsimcoal2: demographic inference under complex evolutionary scenarios. Bioinformatics 37(24): 4882-4885.

Fellowes JR, BPL Chan, CMN Lau, N Sai-Chit, GLP Siu. 2003. Report of rapid biodiversity assessment at Shiwandashan National Nature Reserve and National Forest Park, Southwest Guanxi, China, 2000-2001.

Feng C, Wang J, Liston A, Kang M. 2023. Recombination variation shapes phylogeny and introgression in wild diploid strawberries. Molecular Biology and Evolution 40:msad049.

Gomez-Sanchez D, Olalde I, Sastre N, Ensenat C, Carrasco R, Marques-Bonet T, Lalueza-Fox C, Leonard JA, Vila C, Ramirez O. 2018. On the path to extinction: inbreeding and admixture in a declining grey wolf population. Molecular Ecology 27:3599-3612.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly 6(2)

Gopalakrishnan S, Sinding MHS, Ramos-Madrigal J, Niemann J, Castruita JAS, Vieira FG, Caroe C, Montero MM, Kuderna L, et al. 2018. Interspecific gene flow shaped the evolution of the Genus *Canis*. Current Biology 28(21): P3441-3449.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, Hansen NF. A draft sequence of the Neandertal genome. Science 328:710-722.

Hanghoj K, Molke I, Andersen PA, Manica A, Korneliussen TS. 2019. Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. GagaScience 1-9.

Harris AJ, Foley NM, Williams TL, Murphy WJ. 2022. Tree House Explorer: a novel genome browser for phylogenomics. Molecular Biology and Evolution 39(6):msac130
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.

Hennelly LM, Sarwar G, Fatima H, Werhahn G, Abbas FI, Khan AM, Mahmood T, Kachel S, Kubanychbekov Z, Waseem MT, et al. 2023. Genomic analysis of wolves from Pakistan clarifies boundaries among three divergent wolf lineages. 115:339-348.

Hilgers L, Liu S, Jensen A, Brown T, Cousins T, Schweiger R, Guschanski K, Hiller M. 2024. Avoidable false PSMC population size peaks occur across numerous studies. BioRxiv.

Jensen A, Swift F, de Vries D, Beck RMD, Kuderna LFK, Knauf S, Chuma IS, Keyyu JD, Kitchener AC, Farh K, Rodgers J, Marques-Bonet T, Detwiler KM, Roos C, Guschanski K. 2023. Complex evolutionary history with extensive ancestral gene flow in an African primate radiation. Molecular Biology and Evolution 40(12): msad247
Jhala YV, Saini S, Kumar S, Qureshi Q. Distribution, status, and conservation of the Indian Peninsula wolf. Front Ecol Evol. 2022:10:814966.

Jiang Z, Zang W, Ericson PGP, Song G, Wu S, Feng S, Drovetski SV, Liu G, Zhang D, Saitoh T, Alstrom P, Edwards SV, Lei F, Qu Y. 2024. Gene flow and an anomaly zone complicate phylogenomic inference in a rapidly radiated avian family (Prunellidae). BMC Biology 22:49

Jong MJ, Niamir A, Wolf M, Kitchener AC, Lecomte N, Serydokin IV, Fain SR, Hagen SB, Saarma Y, Janke A. 2023. Range-wide whole-genome resequencing of the brown bear reveals drivers of intraspecies divergence. Communications Biology 6: 153.

Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods, 14, 587–589.
Kardos M, Zhang Y, Parsons KM, Yunga A, Kang H, Xi X, Liu X, Matkin CO, Zhang P, Ward EJ, et al. 2023. Inbreeding depression explains killer whale population dynamics. Nature Ecology and Evolution 7: 675-686.

Koch EM, RM Schweizer, TM Schweizer, DR Stahler, DW Smith, RK Wayne, J Novembre. 2019. De novo mutation rate estimation in wolves of known pedigree. Molecular Biology and Evolution 36(11): 2536-2547.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of next generation sequencing data. BMC Bioinformatics 15:356.

Krofel M, Hatlauf J, Bogdanowicz W, Cambell LAD, Godinho R. Towards resolving taxonomic lineages in wolf, dog, and jackal of Africa, Eurasia, and Australasia. J Zool. 2022:316:155–168.

Kurten B. 1965. The carnivora of the Palestine caves. Acta Zooogical Fennica 07, 74.

Lescroart J, Bonilla-Sanchez A, Napolitano C, Buitrago-Torres DL, Ramirez-Chaves HE, Pulido-Santacruz P, Murphy WJ, Svardal H, Eizirik E. 2023. Extensive phylogenomic discordance and the complex evolutionary history of the Neotropical cat Genus *Leopardus*. Molecular Biology and Evolution 40(12): msad255

Li G, Figueiro HV, Eizirik E, Murphy WJ. 2019. Recombination-aware phylogenomics reveals the structured genomic landscape of hybridizing cat species. Molecular Biology and Evolution 36(10): 2111-2126.

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

Marcus J, Ha W, Barber RF, Novembre J. 2021. Fast and flexible estimation of effective migration surfaces. eLife 10:e61927.

Martin S, Davey H, Salazar C, Jiggins CD. 2019. Recombination rate variation shapes barriers to introgression across butterfly genomes. PLoS Biology 17:e2006288.

Martin SH, Belleghem SMV. 2017. Exploring evolutionary relationships across the genome using topology weighting. Genetics 1:429-438.

Mashkour M, Monochot H, Trinkaus E, Reyss JL, 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

Mathur S, DeWoody JA. 2021. Genetic load has potential in large populations but is realized in small inbred populations. Evolutionary Applications 14(6): 1540-1557.

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

Mech DL, Barber-Meyer S. 2017. Use of erroneous wolf generation time in assessments of domestic dog and human evolution Sci ELett. Available from: http://science.sciencemag.org/content/352/ 6290/1228/tab-e-letters

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

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

Monthey JD, Spellman GM. 2024. Recombination rate variation shapes genomic variability of phylogeographic structure in a widespread North American songbird (Aves: Certhia americana). Molecular Phylogenetics and Evolution 196:108088.

Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. 2016. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics 32:1749-2751.

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

Niemann J, Gopalakrishnan S, Yamaguchi N, Ramos-Madrigal J, Wales N, Gilbert MTP, Sinding MHS. 2020. Extended survival of Pleistocene Siberian wolves into the early 20th century on the island of Honshu. iScience 24: 101904.

Ortiz EM. 2019. Vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. DOI: 10.5281

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Reich D. 2012. Ancient admixture in human history. Genetics, 192, 1065–1093.

Pinxian W, Xiangjun S. 1994. Last glacial maximum in China: comparison between land and sea. Catena 23: 341-535.

Plessis SJ, Blaxter M, Koepfli KP, Chadwick EA, Hailer F. 2023. Genomics reveals complex population history and unexpected diversity of Eurasian otters (*Lutra lutra*) in Britain relative to genetic methods. Molecular Biology and Evolution 40(11): msad207

Ramos-Madrigal J, Sinding MHS, Caroe C, Mak SST, Niemann J, Castruita JAS, Fedorov S, Kandyba A, Germonpre M, Bocherens H, et al. 2021. Genomes of Pleistocene Siberian wolves uncover multiple extinct wolf lienages. Current Biology 31:198-206.

Schubert M, Ermini L, Sarkissian CD, Jonsson H, Ginolhac A, Schaefer R, Martin MD, Fernandez R, Kircher M, et al. 2014. Characterization of ancient and modern genomes

by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nature Protocols 9: 1056-1082.

Schumer M, Xu C, Powell DL, Durvasula A, Skov L, Holland C, Przeworki M. 2018. Natural selection interacts with recombination to shape the evolution of hybrid genomes. Science 360:565-660.

Segawa T, Yonezawa T, Mori H, Kohno A, Kudo Y, Akiyoshi A, Wu J, Tokanai F, Sakamoto M, Kohno N, Nishihara H. 2022. Paleogenomics reveals independent and hybrid origins of two morphologically distinct wolf lineages endemic to Japan. Current Biology 32:2494-2504.

Skotte L, Korneliussen TS, Albrechtsen A. 2013. Estimating individual admixture proportions from Next Generation sequencing data. Genetics 3: 693-702.

Smeds L, Ellegren H. 2022. From high masked to high realized genetic load in inbred Scandinavian wolves. Molecular Ecology 32(7): 1567-1580.

Sowerby AC. 1936. China's natural history: a guide to the Shanghai Museum. Royal Asiatic Society of Great Britain and Ireland, North China Branch.

Statham MJ, Murdoch J, Janecka J, Aubry KB, Edwards CJ, Soulsbury CD, Berry O, Wang Z, Harrison D, Pearch M, et al. 2014. Range-wide multilocus phylogeography of the red fox reveals ancient continental divergence, minimal genomic exchange, and distinct demographic histories. Molecular Ecology 23(19): 4813-4830.

Stronen AV, Jedrzejewska B, Pertoldi C, Demontis D, Randi E, Niedzialkowska M, Pilot M, Sidorovich VE, et al. 2013. North-south differentiation and a region of high diversity in European wolves (*Canis lupus*). PLoS ONE 8(10): e76454

Taron UH, Salado I, Escobar-Rodriguez M, Westbury MV, Butschkau S, Paijmans JLA, vonHoldt BM, Hofreiter M, Leonard JA. 2021. A sliver of the past: the decimation of the genetic diversity of the Mexican wolf. Molecular Ecology 30:6340-6354.

Thompson EA. 2013. Variation in meiosis, across genomes, and in populations. Genetics 194:301-326.

Tricou T, Tannier E, Vienne DM. 2022. Ghost lineages highly influence the interpretation of introgression tests. Systematic Biology 71(5): 1147-1158.

vonHoldt B, Pollinger JP, Earl DA, Knowles JC, Boyko AR, Parker H, Geffen E, Pilot M, Jedrzejewski W, Jedrzejewska B, Sidorovich V, Greco C, et al. 2011. A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. Genome Research 21(8): 1294-1305.

vonHoldt BM, Brzeski KE, Aarema ML, Schell CJ, Rutledge LY, Fain SR, Shutt AC, Linderholm A, Murphy WJ. 2021. Persistence and expansion of cryptic endangered red wolf genomic ancestry along the American Gulf coast. Molecular Ecology 31:5440-5454.

Wang GD, Zhang M, Wang X, Yang MA, Cao P, Liu F, Lu H, Feng X, Skoglund P, Wang L, Fu Q, Zhang YP. 2019. Genomic approaches reveal an endemic subpopulation of gray wolves in Southern China. iScience 20:110-118.

Wang L, Ma YP, Zhou QJ, Zhang YP, Savolainen P, Wang GD. 2016. The geographical distribution of grey wolves (Canis lupus) in China: a systematic review. Zoological Research 37(6): 315-326.

Wang MS, Thakur M, Jhala Y, Wang S, Srinivas Y, Dai SS, Liu ZX, Chen HM, Green RE, Koepfli KP, Shapiro B. 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 2:evac012.

Wang MS, Wang S, Li Y, Jhala Y, Thakur M, Otecko NO, Si JF, Chen HM, Shapiro B, Nielsen R, YP Zhang, DD Wu. 2020. Ancient hybridization with an unknown population facilitated high-altitude adaptation of canids. Molecular Biology and Evolution 37:2616-2629.

Werhahn G, Hennelly LM, Lyngdoh S, Habib B, Viranta S, Shrotriya S. 2023. Himalayan wolf: Canis lupus spp chanco. The IUCN Red List of Threatened Species
Werhahn G, Senn H, Macdonald DW, Sillero-Zubiri C. The diversity of genus Canis challenges conservation biology: a review of available data on Asian wolves. Front Ecol Evol. 2022:10:10.3389.

Zhang C, Rabiee M, Sayarri E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19:153

Acknowledgements

L.M.H thanks the National Science Foundation Postdoctoral Research Fellowship (award number 2208950) for funding and support. The Norwegian Environment Agency (project 18088069) provided funding and support for sequencing efforts on newly sequenced wolf genomes. Ç.H.Ș. thanks Fondation Segré and the Sigrid Rausing Trust for providing the majority of the funding for this project, H. Batubay Özkan and Barbara

Watkins for their support of the Biodiversity and Conservation Ecology Lab at the University of Utah, and Bilge Bahar, Seha İşmen, Ömer Koç, Ömer Külahçıoğlu, Burak Över, Emin Özgür, Suna Reyent and Ceren Sağlamer for supporting this project. Türkiye's Department of Nature Conservation and National Parks and the Ministry of Agriculture and Forestry granted the permit for Türkiye (No. 72784983-488.04-114100). We thank the Museum of the Institute of Plant and Animal Ecology UB RAS access to their collections.

Author Contributions

L.M.H. conceived and designed the project with guidance from S.G., M.H.S.S., and B.N.S. C.S.O., M.H.S.S., N.F.G.M. conducted laboratory work, H.F., G.S., B.H., F.H., S.P.E.H., C.K., Ç.H.Ş., P.K., H.S., M.H.S.S., M.T.P.G. provided logistics, field work, wolf samples, and sequencing efforts, L.M.H. led data analysis with assistance from B.R.P., A.N., X.S., N.F.G.M, L.M.H. wrote the manuscript with input from all co-authors.

Data Availability

All raw reads have been submitted to the NCBI Project ID XXXX, with accession numbers XXXX-XXXX. The bioinformatic scripts and results from the phylogenetic analyses, D statistics, genetic load, and PSMC can be found on the github: https://github.com/hennelly/Asia_wide_wolf_genomics.

Figure Captions

Figure 1. Population genetic structure genetic connectivity of wolves across Eurasia. (A) Sample location and individual admixture proportions estimated using genotype likelihoods of ~300,000 autosomal SNPs at K=6 for 98 wolves across Eurasia. The individual admixture proportions were estimated using a set of 101 gray wolves, 5 dogs, and 12 genomes from various basal canid species. **(B)** Principal component analysis using genotype likelihoods of ~10 million autosomal SNPs for 98 wolves across Eurasia. Colors indicate geographic region of each the individual wolf belongs to. **(C)** The estimated effective migration surface (EEMS) using ~8.7 million autosomal SNPs for 97 wolves across Eurasia. Colors of the EEMS correspond to lower-than-average (red) and higher-than-average (blue) effective migration as expressed as $\log_{10}(w)$. A lambda of 100 was used after performing cross-validation analyses.

Figure 2. Tibetan-like ancestry in wolves from central Asia and eastern Asia. **(A)** D statistic results to estimate derived allele sharing between selected wolf individuals in Asia and the Tibetan lineage. The only populations that show significant allele sharing with Tibetan wolves are wolves from Pakistan, Kyrgyzstan, Mongolia, and eastern China. **(B)** Autosomal phylogeny using 98 wolves from Eurasia, 3 North American gray wolves and 12 genomes from other canid species. The tree was inferred using ASTRAL and estimated using 1,000 regions in 20kb length that were randomly chosen across the dog reference genome. Normalized quartet score of selected nodes are shown. Colored circles indicate the individuals assigned to the Tibetan lineage (orange), Central Asian mountains (high altitude Pakistan, Tajikistan, Kyrgyzstan), North American wolves (gray), and Eastern Asia (eastern China, Mongolia). **(C)** A tree inferred using ASTRAL with only low recombination (<0.2cM/Mb) regions of the X chromosome. A total of 513K SNPs were used, which we split into regions of 10kb length to assess gene tree discordance using ASTRAL. Normalized quartet score of selected nodes are shown. Eastern Asian wolves form a well-supported clade that is basal within gray wolves, except Tibetan wolves. **(D)** The three topologies with the highest average topology weightings across the X chromosome when considering wolves from the central Asian mountains or eastern Asia as the focal population (X). The topology shaded in gray depicts a scenario where the focal wolf population is ancestral to other Holarctic wolves from Europe and Central Russia. Prevalence of 15 possible topologies in high, medium and low recombination regions on the X chromosome to assess whether wolves from **(E)** central Asian mountains and **(F)** eastern Asia show evidence of ancestry that is ancestral to northern Holarctic wolves. For both central Asian and eastern Asian wolves, three topologies (solid lines) show a negative relationship between recombination rate and topology weighting, in contrast to the 12 other topology (dotted lines). **(F)** Eastern Asian wolves have more thoroughly sorted windows when they form a lineage ancestral to other northern Holarctic populations (topology 10) in low recombination regions compared to wolves from the mountains of central Asian mountains.

Figure 3. Persistence of Indian-like ancestry in wolves of southwest Asia. **(A)** Declines in autosomal ancestry proportion at K=6 assigned to the Indian lineage (purple), the Tibetan lineage (orange) and African wolf (*Canis lupaster*) (blue) for selected wolves as geographic distance increases from the distributional extent of these three taxa. Assigned Indian ancestry does not decline with geographic distance for wolves in southwest Asia (purple triangles). **(B)** The model with the highest likelihood inferred using fastsimcoal2 corresponding to an introgression event between the Indian lineage and wolves of southwest Asia. **(C)** The topology with the highest average topology weight when testing the relationship between 4 wolf populations and the African wolf, resulting in 105 possible topologies inferred using 100-SNP windows with Twisst. **(D)** A maximum likelihood phylogeny using autosomal genomic regions in which the topology weight of topology where Indian and southwest Asian wolves are sister (topology 73) was above 0.3 (~99.99th percentile; ~40kb length total). The tree was inferred using IQ-tree where the best model was TVM+F+R5 through ModelFinder and 1000 bootstraps were used. Local posterior probability is labeled at selected nodes. **(E)** A maximum likelihood phylogeny that excludes wolves from Pakistan and Ladakh, which was inferred using autosomal genomic regions where the topology weight of topology 73 was above 0.3 (~99.99th percentile; ~40kb length total).

Figure 4. Recent and long-term demographic history of wolves in Eurasia. (A) Estimates of genome-wide heterozygosity inferred using ANGSD for 99 wolves from Eurasia**. (B)** Runs of homozygosity (ROH) for 43 wolves showing the sum of the length of ROH in four size ranges: 0.1-1Mb (Short ROH), 1-5Mb (Short-medium ROH), and 5-10Mb (medium-long ROH), and 10-100Mb (long ROH). Individuals are arranged from top to bottom by the sum length of ROH with wolves most representing the Indian and Tibetan lineages bolded. **(C)** Pairwise sequentially Markovian coalescent (PSMC) analysis of the demographic histories of 33 wolves representing multiple populations in Eurasia and North America. Thicker lines indicate population belonging to the Indian and Tibetan lineages, and a single individual from Shanxi, China. We assumed a mutation rate of 4.5 x 10^{-9} and a generation time of 4.4 years.

Figure 5. Genetic load in selected wolf individuals in Asia and North America. **(A)** Total genetic load, **(B)** masked load, and **(C)** realized load for 13 wolf individuals wihtin five populations. Wolves found in historically smaller populations (Mexican, Indian, and Tibetan) have a lower total load, lower masked load, and higher realized load. The total derived allele count of each wolf within the five populations, where generally Mexican, Indian, and Tibetan wolves show overall less derived alleles. The derived allele count of **(E)** Loss of function (LOF) mutations classified as high impact, **(F)** missense mutation classified as moderate impact, and **(G)** synonymous mutations classifed as low impact per 100,000 derived alleles found in non-coding regions, assumed to be neutral. When taking into account differences in overall number of derived count, Mexican, Indian, and Tibetan wolves show reduced number of missense and synonymous mutations, but not for loss of function mutation, compared to central Asian and southwest Asian wolves.

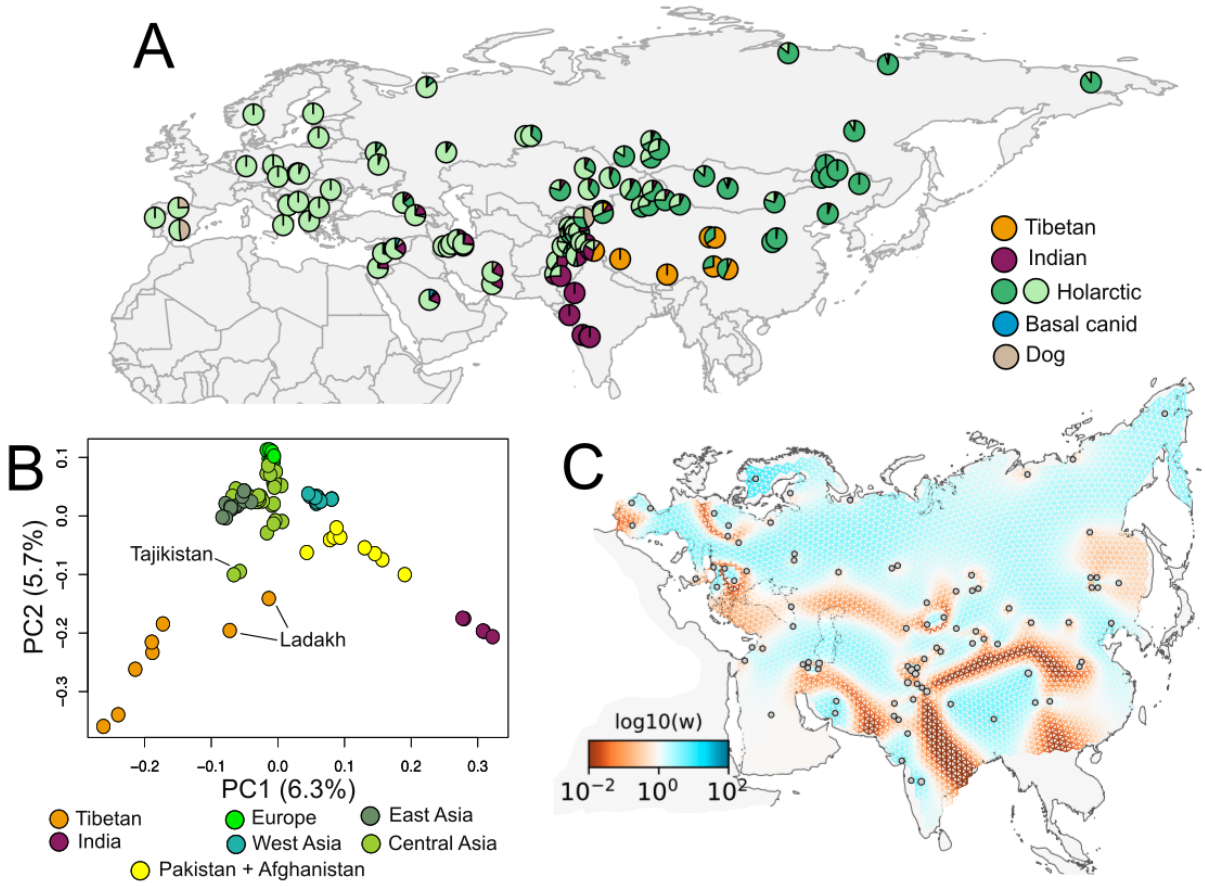

Figure 1. Population genetic structure genetic connectivity of wolves across Eurasia.

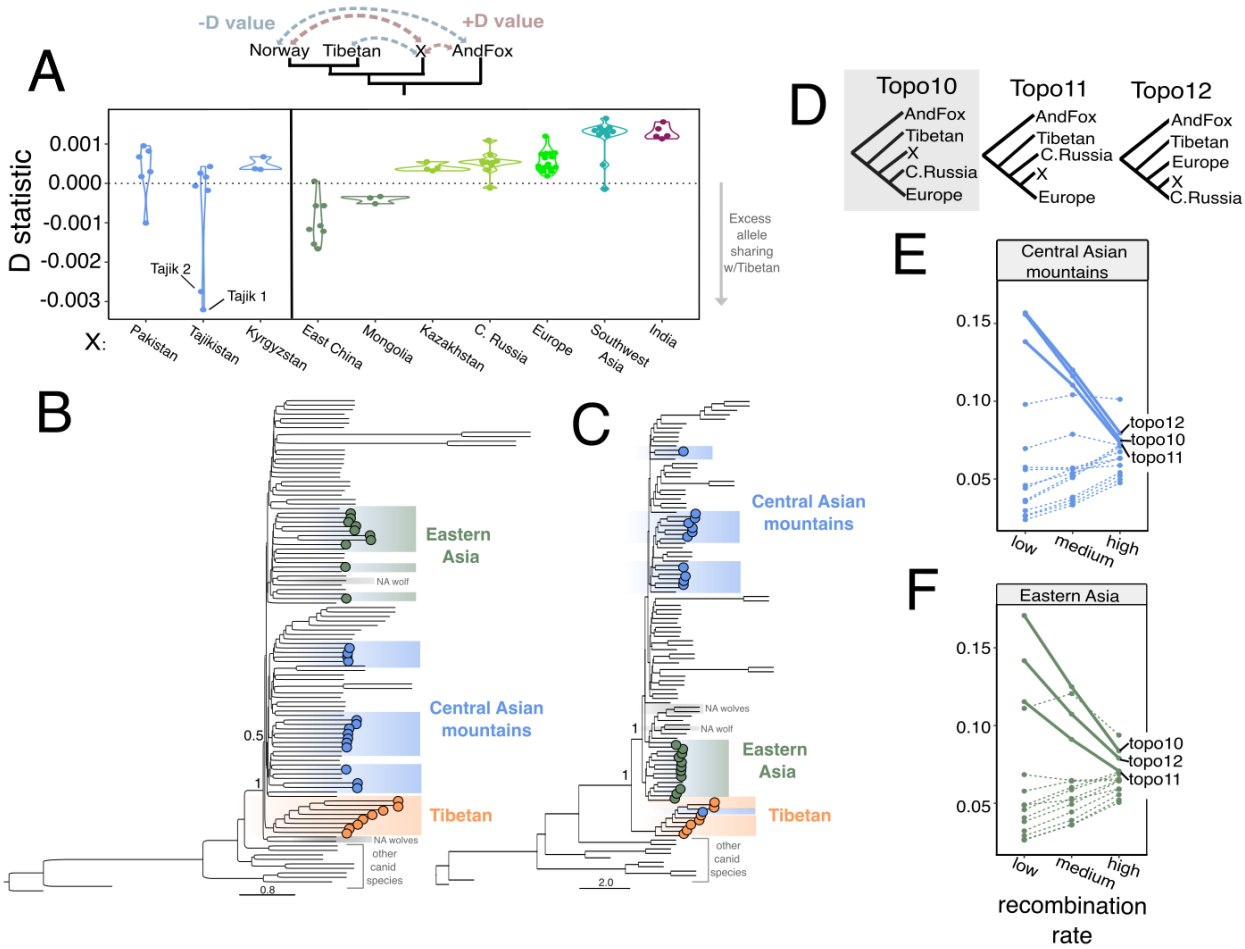

Figure 2. Tibetan-like ancestry in wolves from central Asia and eastern Asia.

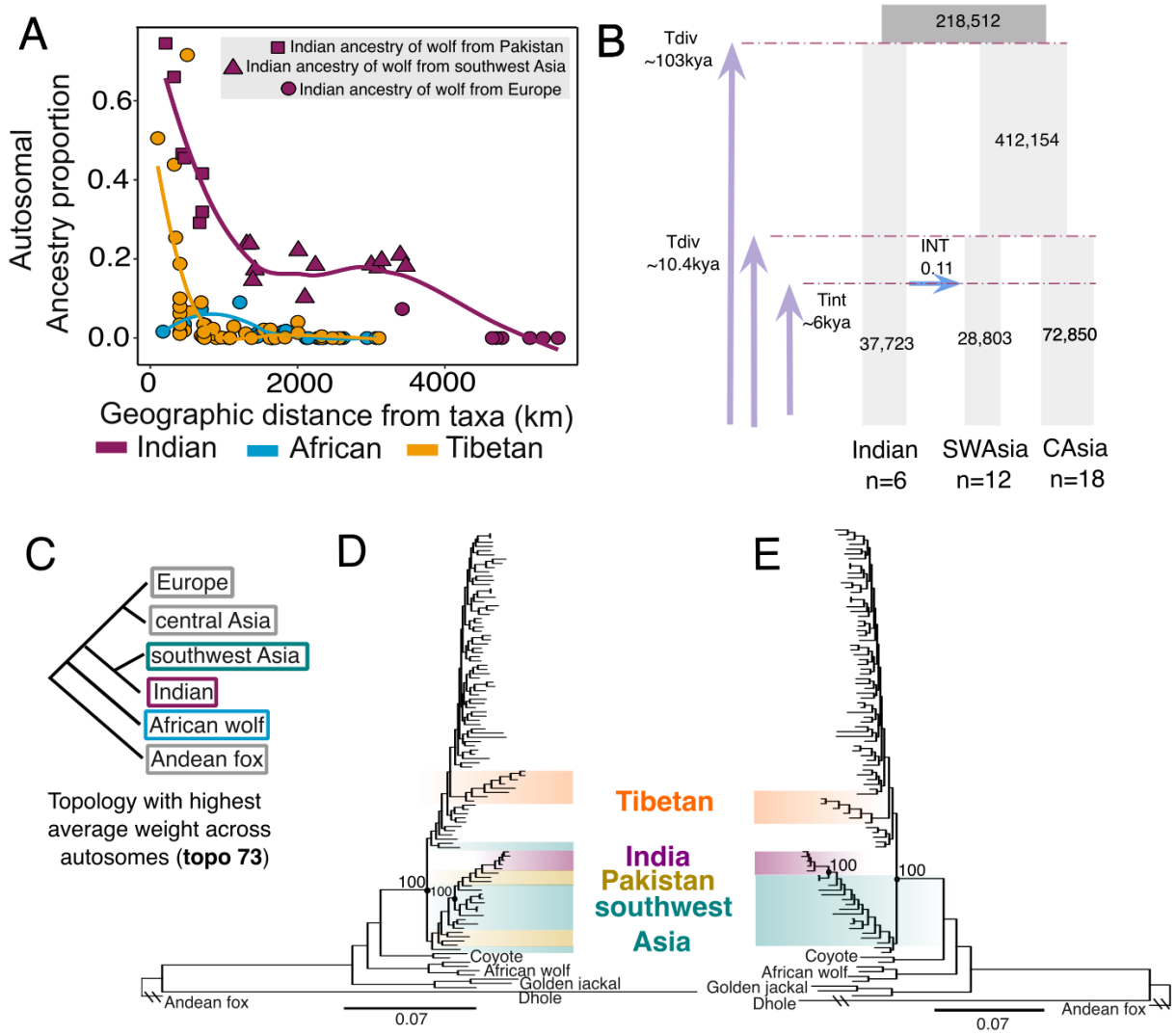

Figure 3. Persistence of Indian-like ancestry in wolves of southwest Asia.

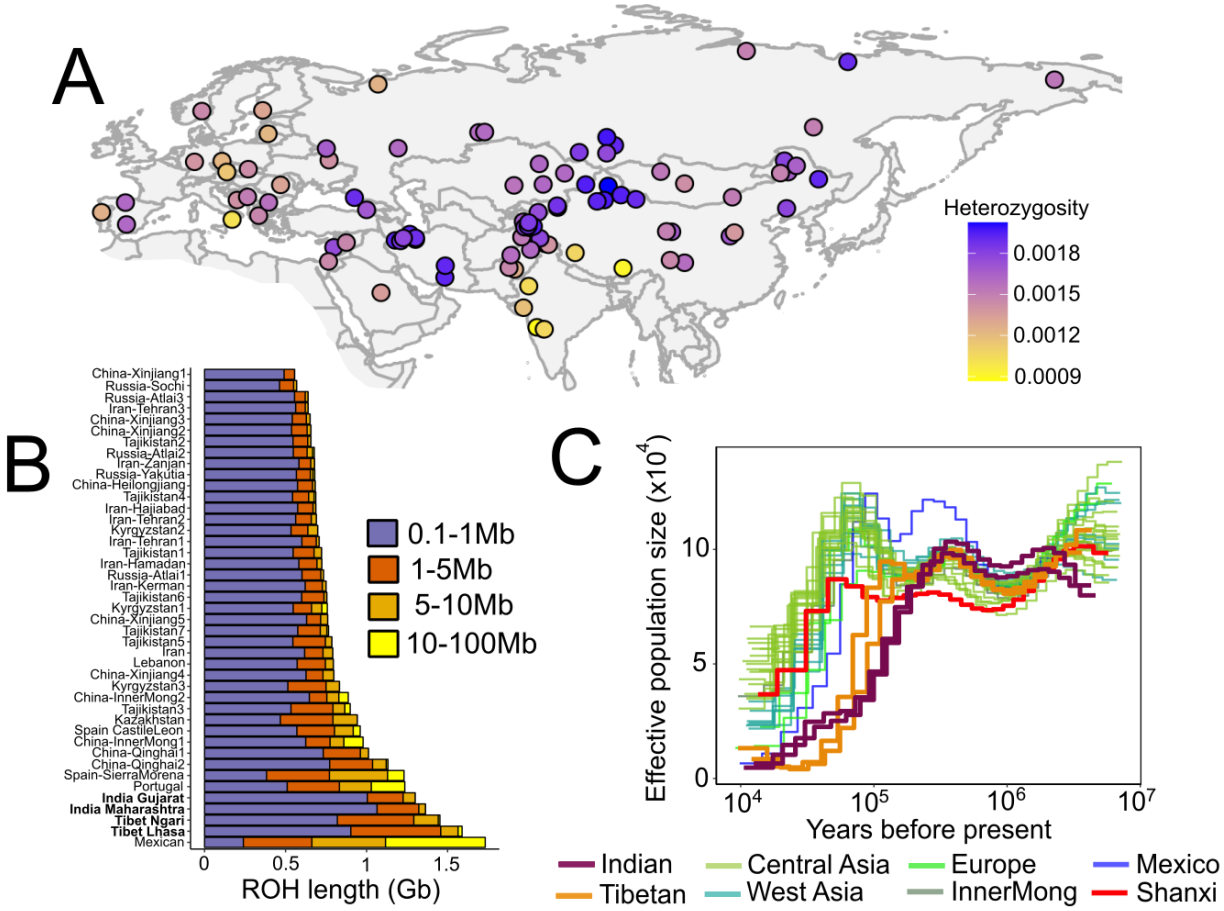

Figure 4. Recent and long-term demographic history of wolves in Eurasia.

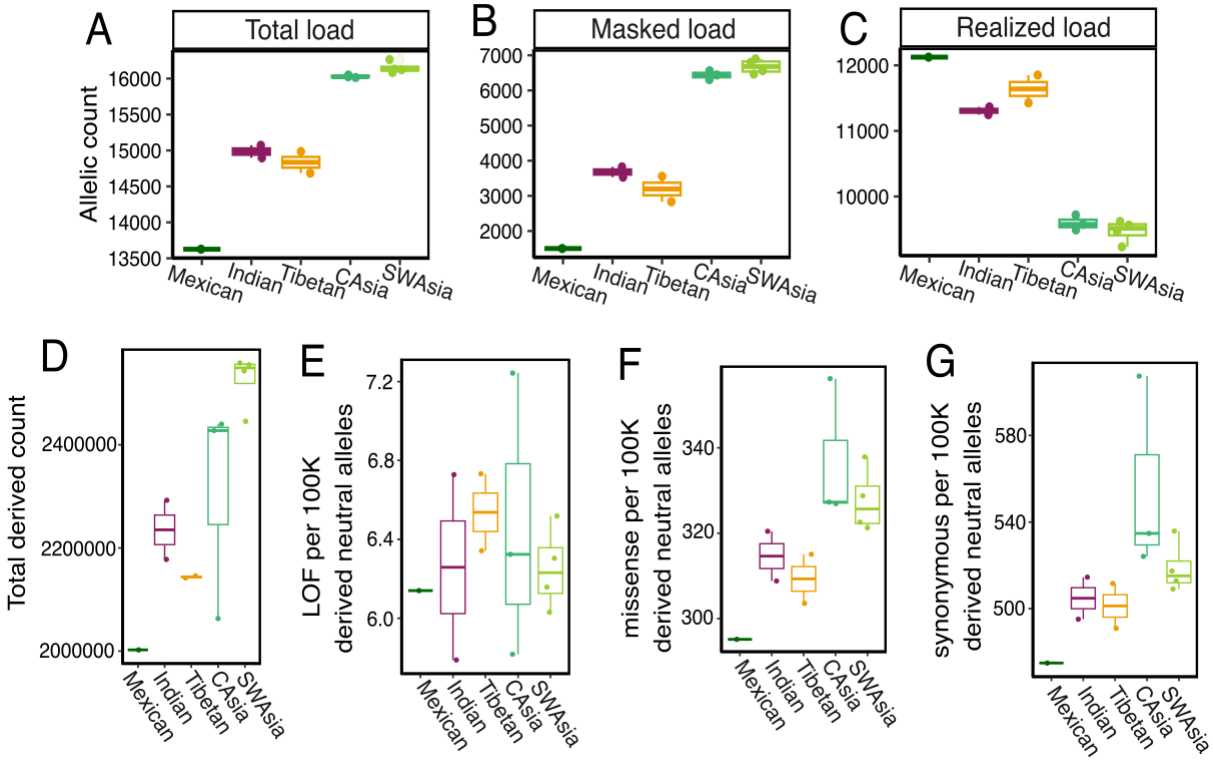

Figure 5. Genetic load in selected wolf individuals in Asia and North America.