Ancient dog introgression into the Iberian wolf genome may have facilitated adaptation to human-dominated landscapes

Diana Lobo[1,2,3*], Hernán E. Morales[4,5], Cock Van Oosterhout[6], José Vicente López-Bao[7], Pedro Silva[1,3], Luis Llaneza[8], Carolina Pacheco[1,2,3], Diana Castro[1,3], Germán Hernández-Alonso[4], George Pacheco[4], John Archer[1,3], M. Thomas P. Gilbert[4,9], Nuno Ferrand[1,2,3,10], and Raquel Godinho[1,2,3,10*]

[1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal. [2]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, 4169-007 Porto, Portugal. [3]BIOPOLIS, Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal. [4]Center for Evolutionary Hologenomics, The GLOBE Institute, University of Copenhagen, 1353 Copenhagen, Denmark. [5]Section for Evolutionary Genomics, The GLOBE Institute, University of Copenhagen, 1353 Copenhagen, Denmark. [6]School of Environmental Sciences, Norwich Research Park, University of East Anglia, NR4 7TJ Norwich, UK. [7]Biodiversity Research Institute (CSIC - Oviedo University - Principality of Asturias) Oviedo University, E-33600 Mieres, Spain. [8]A.RE.NA, Asesores en Recursos Naturales, 27003 Lugo, Spain. [9]University Museum, Norwegian University of Science and Technology, 7012 Trondheim, Norway. [10]Centre for Ecological Genomics and Wildlife Conservation, Department of Zoology, University of Johannesburg, 2006 South Africa

*** Corresponding Authors**:

Diana Lobo: diana.lobo@cibio.up.pt

Raquel Godinho: rgodinho@cibio.up.pt

Running Title: Iberian wolf adaptation through dog introgression

Abstract

Understanding how large carnivores respond to increasingly human-dominated landscapes will determine their future adaptive potential. The Iberian wolf (*Canis lupus signatus*), a gray wolf subspecies endemic to the Iberian Peninsula (Portugal and Spain), has uniquely persisted in human-dominated landscapes, unlike many other wolf populations that faced widespread extinction across Europe during the 20th century. In this study, we conducted a comprehensive genome-wide analysis of 145 historical and contemporary Iberian wolf samples to investigate whether hybridization with domestic dogs resulted in genetic introgression. We identified a dog-derived block on Chromosome 2 in Iberian wolves, displaying signatures consistent with introgression and high nucleotide similarity among introgressed individuals. Additionally, our estimates place the average timing of introgression between 6,100 and 3,000 years ago, with low sequence divergence to dogs from the Iberian Peninsula suggesting a single local origin for the hybridization event. Using forward genetic simulations, we show that the introgressed haplotype is most likely being maintained in Iberian wolves by selection. The introgressed dog variants are located within the *MAST4* gene, which has been linked to neurological disorders, including cognitive and motor developmental delays, hinting at a potential role in cognitive behavior in Iberian wolves. This study uncovers a case of putative adaptive introgression from domestic dogs into wolves, offering new insights into wild canids' adaptation to human-dominated landscapes.

Introduction

Circa 10 thousand years ago (Kya), humans started to change the planet prompted by the advent of farming following the domestication of plants and animals (Frantz et al. 2020). Over time, large carnivores have been severely affected by these rapid environmental changes, mostly because of the increasing competition with humans (e.g., persecution mediated by livestock depredation), and many have gone (locally) extinct (Chapron et al. 2014). Notably, amidst these challenges, some species managed to change their behavior and ecology to adapt to human activities in human-dominated landscapes (Benazzo et al. 2017). A remarkable example of such resilience is illustrated by some populations of the Eurasian gray wolf (*Canis lupus*) living in highly human-dominated landscapes (Sazatornil et al. 2016; Rio-Maior et al. 2019). Such landscapes are found in the Iberian Peninsula (i.e., Portugal and Spain), where the human population density ranges from 20 to 400 inhabitants/km^2 in the wolf's distribution (INE 2017) (*vs.* European average: 97 inhabitants/km^2; Chapron et al. 2014).

The Iberian wolf subspecies (*Canis lupus signatus;* **Fig. 1C**) has diverged and remains isolated from other Eurasian wolves for approximately 10 Kya (Silva et al. 2020). Despite the widespread extinction of wolf populations in Europe during the 20th century (Mech and Boitani 2003), the Iberian wolf exhibits remarkable resilience despite intense human persecution (Sastre et al. 2011; Nores and López-Bao 2022). Notably, this subspecies not only persists in areas characterized by high human density but also exhibits a remarkable tolerance to high levels of human activity (Llaneza et al. 2012; Llaneza et al. 2016; Dennehy et al. 2021). Moreover, Iberian wolves show reduced levels of chronic stress compared to their Eastern European counterparts (Pereira et al. 2022). Genetic and spatial behavior analyses have unveiled peculiar lack of long-distance dispersal and cryptic population structure in Iberian wolves, with negligible

instances of dispersal beyond the range of the respective genetic cluster (Silva et al. 2018). This suite of features, particularly the lack of long-distance dispersal, hint at potential adaptations to increasingly human-dominated landscapes. In such environments, short-distance dispersal could confer advantage, as individuals that disperse long distances tend to be less risk-averse, traveling through unknown and more human densely populated areas, resulting in decreased survival rates (Murray et al. 2010; Morales-González et al. 2021). However, the evolutionary mechanisms driving this behavior remain unknown.

A consequence for wolves living in highly human-dominated landscapes is the occurrence of hybridization with domestic dogs (Vilà and Wayne 1999; Boitani 2003). Despite clear differences in behavior and morphology resulting from dog domestication, Eurasian gray wolves and dogs exhibit high genetic similarity (0.04 – 0.21% nucleotide differentiation; Lindblad-Toh et al. 2005) and their hybrids can backcross to both wild and domestic individuals (Godinho et al. 2011). In the Iberian Peninsula, historical and contemporary events of wolf-dog hybridization have been documented (Godinho et al. 2011; Fan et al. 2016; Torres et al. 2017; Pacheco et al. 2017; Gómez-Sánchez et al. 2018; Lobo et al. 2023). While hybridization has been mainly viewed as a conservation threat to wolves because the introgression of dog variants can disrupt their local adaptations (Rhymer and Simberloff 1996; Allendorf et al. 2001), more recent research has reported cases of adaptive introgression (Anderson et al. 2009; Schweizer et al. 2018; Pilot et al. 2021), suggesting it can also be a powerful force in the evolutionary response of wolves to new environmental conditions. Based on this, we hypothesize that ancient hybridization events and subsequent dog introgression could have influenced the evolutionary trajectory of the Iberian wolf, enabling its adaptation to human-dominated landscapes.

To test the "introgression fueled adaptation" hypothesis we conducted a comprehensive genomic analysis of 48 historical (1912 – 2005) and 97 contemporary samples of Iberian wolves, alongside 67 worldwide wolves and 131 dogs (**Figs. 1A** and **1B**; **Supplemental Table S1**). We used these data to: 1) assess the evidence for dog introgression in the Iberian wolf genome, 2) determine whether introgressed variants display signatures of selection, and 3) explore the functional role of introgressed variants to elucidate their potential association with adaptations to human-dominated landscapes.

Results

Presence of a dog block in Chromosome 2 of Iberian wolves

To unveil signatures of dog introgression in the Iberian wolf genome, we employed a population genomics approach, analyzing 85,000 genome-wide single nucleotide polymorphisms (SNPs) (hereafter 85K). These SNPs were genotyped using the Canine HD BeadChip microarray (Illumina, San Diego, CA, USA) in contemporary samples of non-closely related individuals of the Iberian wolf (N = 95), the Eurasian wolf (N = 55), and dogs (N = 120) (**Figs. 1A** and **1B**, see **Supplemental Table S1** and Methods for details). Using the criteria established by Lobo et al. (2023), we targeted wolf individuals measurably free of recent dog admixture to focus on older introgression events, setting a threshold of genome-wide dog ancestry below 10%. ADMIXTURE analysis clearly distinguished wolves from dogs at K=2, with an average genome-wide proportion attributed to dog below 1.6% in the wolf genome (**Supplemental Fig. S1A**). Similarly, all dogs had less then 1.5% of wolf ancestry in their genomes (**Supplemental Fig. S1A**), validating their use as a reference population. Principal component analysis (PCA) initially separated dogs and wolves and further distinguished Iberian wolves from other Eurasian wolf populations (**Supplemental Fig. S1B**), thus confirming the previously

documented genetic distinctiveness of Iberian wolves (Pilot et al. 2014; Silva et al. 2020).

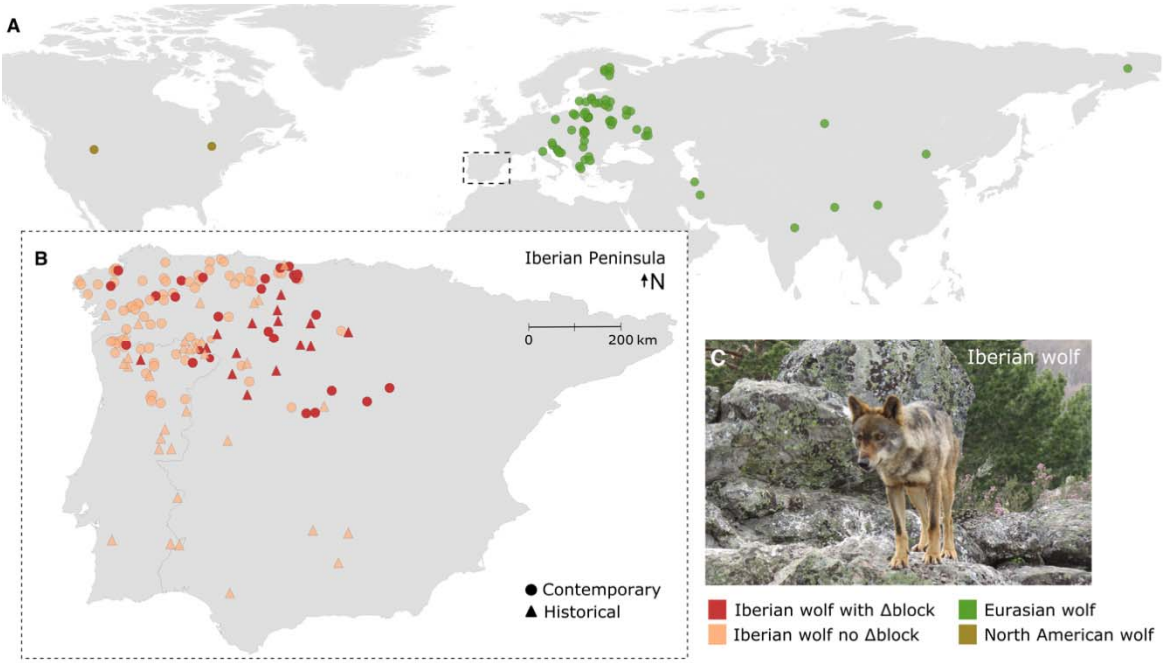


Figure 1. Wolf sample locations. (**A**) Geographical distribution of analyzed canid samples worldwide. Green and dark yellow dots denote samples from Eurasian and North American gray wolves, respectively. (**B**) Zoomed-in view of the sampling area within the Iberian Peninsula, outlined by the dashed line in (A). Iberian wolves with and without the dog Δblock are depicted in red and pink, respectively. Circles and triangles represent contemporary and historical samples, respectively. For additional sample details, refer to **Supplemental Tables S1, S3,** and **S4**. (**C**) Image of an Iberian wolf (*Canis lupus signatus*); photo credits to Raquel Godinho.

To identify signatures of dog ancestry across the Iberian wolf genome, we utilized local ancestry methods LAMP-ANC (Sankararaman et al. 2008; Pasaniuc et al. 2009) and ELAI (Guan 2014). In these analyses, Eurasian wolves and dogs were used as reference populations. Both local ancestry methods consistently assigned similar ancestries across the 38 autosomes, estimating that approximately 1% of the Iberian wolf genome carries dog ancestry (**Supplemental Figs. S2** and **S3**). This 1% of dog

ancestry results from small introgressed genomic regions mostly stochastically distributed across individuals (**Supplemental Figs. S2** and **S3**).

Unlike neutral ancient introgression, if introgression conveys a selective advantage, population genetic theory predicts that introgressed genomic regions should be present at high frequency within the population (Aguillon et al. 2022). To identify such pattern, we used the SNP-specific delta ancestry statistics (∆) (Tang et al. 2007). We detected a "dog block" on Chromosome 2 of Iberian wolves as the top genome-wide outlier, corresponding to 12 standard deviations (SD) from the mean ∆ancestry (**Fig. 2A**). This dog block was present in approximately 30% of the Iberian wolf population based on genome-wide SNP data, irrespective of the number of mixture generations considered (ranging from 10 to 1000; **Supplemental Fig. S4**). The block length is approximately 2.6 Mb (hereafter ∆block) between positions 50.02 – 52.62 Mb (**Fig. 2A; Supplemental Figs. S2** and **S3**), encompassing 78 SNPs and 22 protein-coding genes (CanFam 3.1 Ensembl annotation; **Supplemental Table S2**).

To further examine the haplotype structure and reconstruct phylogenetic relationships of ∆block within Iberian wolves and dogs, we used the 78 SNPs spanning this genomic region to construct both a neighbor-joining (NJ) tree and a median-joining network. The NJ tree revealed that all Iberian wolves with the ∆block (N = 21) were more closely related to dogs than to other wolves lacking the ∆block (**Supplemental Fig. S6A**). This relationship was further supported by a PCA (**Supplemental Fig. S6B).** Notably, Iberian wolves shared the same ∆block haplotype across the 78 SNPs (**Fig. 2B**), which is distinguished from a haplotype found in village dogs from Iberia by only eight nucleotide differences (**Fig. 2B**).

Prevalence of Δblock in historical Iberian wolf samples

For a recent-temporal perspective on the prevalence of the dog Δblock in the Iberian wolf population, we analyzed a set of 48 historical Iberian wolf samples spanning periods from 1912 to 2005 (Pacheco et al. 2022; **Fig. 1B**; **Supplemental Table S4**). These samples, genotyped for 18,000 genome-wide SNPs overlapping with the 85K SNP dataset, underwent local ancestry analysis using LAMP-ANC.

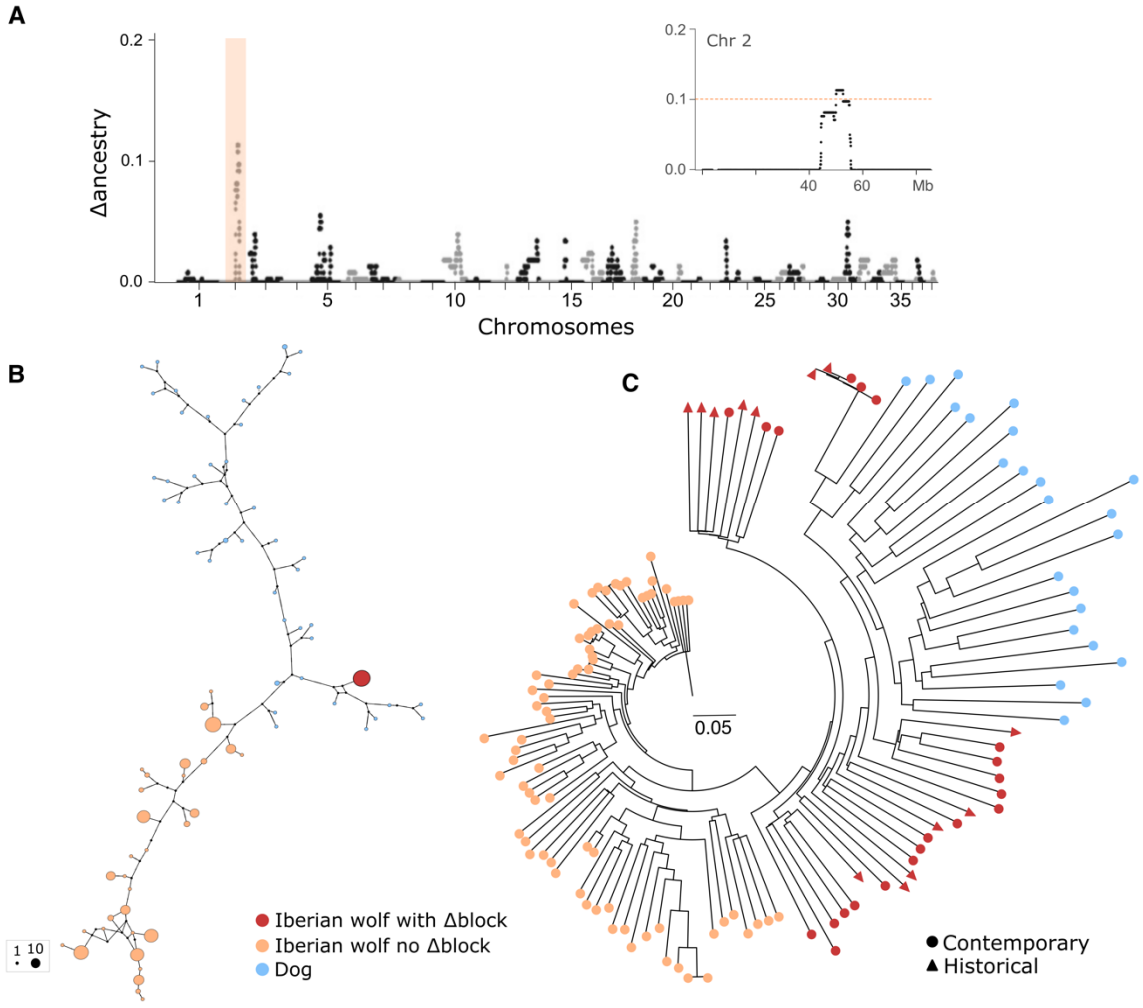


Figure 2. Evidence of a dog block in Iberian wolf Chromosome 2. (**A**) Δancestry scores, indicating the excess of dog ancestry across the Iberian wolf genome, based on ~85K SNPs from 95 contemporary samples. The Δblock position on Chromosome 2 is highlighted in orange. The inset offers a zoomed-in view of Chromosome 2, with the dashed line indicating the outlier detection cut-off (3 SD from the chromosome mean). (**B**) Median-joining network of Δblock

haplotypes (78 SNPs) for contemporary Iberian wolf samples and village dogs from Iberia. Circle size is proportional to the frequency of each haplotype. (**C**) Neighbor-joining (NJ) tree for the ∆block using contemporary and historical samples (subset > 60K SNPs; see Methods) of Iberian wolves and village dogs from Iberia. In panels B and C, Iberian wolves with and without the dog ∆block are colored in red and pink, respectively, while dogs are in blue. Circles and triangles denote contemporary and historical samples, respectively.

The ∆block was identified in 15 historical Iberian wolves (**Supplemental Figs. S7B** and **S7D**) and confirmed as the top genome-wide outlier region with dog ancestry based on ∆ancestry statistics (**Supplemental Figs. S7A** and **S7C**). The ∆block was traced back to the oldest record in a wolf sampled in 1945. Additionally, a NJ tree recreated for ∆block, encompassing contemporary and historical Iberian wolf samples and village dogs from Iberia, revealed that all wolves with the dog ∆block clustered together and closer to dogs (**Fig. 2C**). This clustering suggests a common origin for the ∆block through the same introgression event.

Evidence that the dog block is an introgressed variant using whole-genome data

To enhance the resolution provided by the 85K SNP dataset (averaging 1 SNP every 25 kb), we complemented our population-based dataset with whole-genome data. We sequenced the complete genome of 12 contemporary Iberian wolves (mean sequence coverage of 13.5×), each previously genotyped for the 85K SNPs, and added publicly available whole-genome data from contemporary samples of two additional Iberian wolves, 10 Eurasian gray wolves, two North American gray wolves, 11 dogs, one Golden Jackal, and an Andean Fox, used as outgroup species (**Figs. 1A** and **1B**; **Supplemental Table S3**). Employing the same local ancestry analysis in ELAI, we validated the presence of the dog ∆block in six Iberian wolf whole-genomes – L474, L588, L590, L844, Wolf24, and Wolf39 (**Fig. 3A**, **Supplemental Fig. S5**). However, only in two individuals, L588 and L590, was the ∆block previously identified using the SNP dataset,

illustrating its limited genomic resolution. Wolf39 and L590 carried the longest dog block (Chr 2: 43.5 – 55.1 Mb), whereas it appeared fragmented in the other four wolves.

Having validated the presence of the dog ∆block in Iberian wolves, we delved into a more detailed analysis of allele sharing patterns across Chromosome 2. Employing a sliding window analysis of the fraction of introgression using f_d statistic (Martin et al. 2015), we uncovered a significant excess of allele sharing between Iberian wolves with the ∆block and dogs within a 500 kb genomic region consistent with the ∆block (Chr 2: 52.00 – 52.50 Mb, f_d = 0.54; **Fig. 3B**). In a scenario of random sorting of ancestral variation, this significant excess of shared variation would not be expected in relation to Eurasian wolves, thus suggesting post-divergence gene flow between Iberian wolves and dogs.

To further dissect the evolutionary history of this genomic region, we compared the population phylogenetic trees for the entire Chromosome 2 and for the 500 kb region using TreeMix (Pickrell and Pritchard 2012). The Chromosome 2 tree (**Fig. 3C** and **Supplemental Fig. S8**) is consistent with the genome-wide tree (**Supplemental Fig. S8**), recapitulating the species tree topology. Upon introducing migration into TreeMix (*m* = 1) for the Chromosome 2 tree, gene flow from dogs into Iberian wolves with the ∆block was detected (**Fig. 3C**). A pronounced discordance in the 500 kb tree relative to the expected species tree grouped Iberian wolves carrying the ∆block together with dogs (**Fig. 3C**).

While the observed patterns of excess allele sharing and tree discordance align with the expectations of introgression, these can also emerge from alternative genomic processes, such as incomplete lineage sorting, population structure, or selection (Eriksson and Manica 2012; Smith and Kronforst 2013; Zheng and Janke 2018). However, only introgression should lead to exceptionally high levels of sequence identity

between the donor and recipient species. Such high sequence similarity indicates more recent coalescence of variation in this part of the genome in Iberian wolves and dogs.

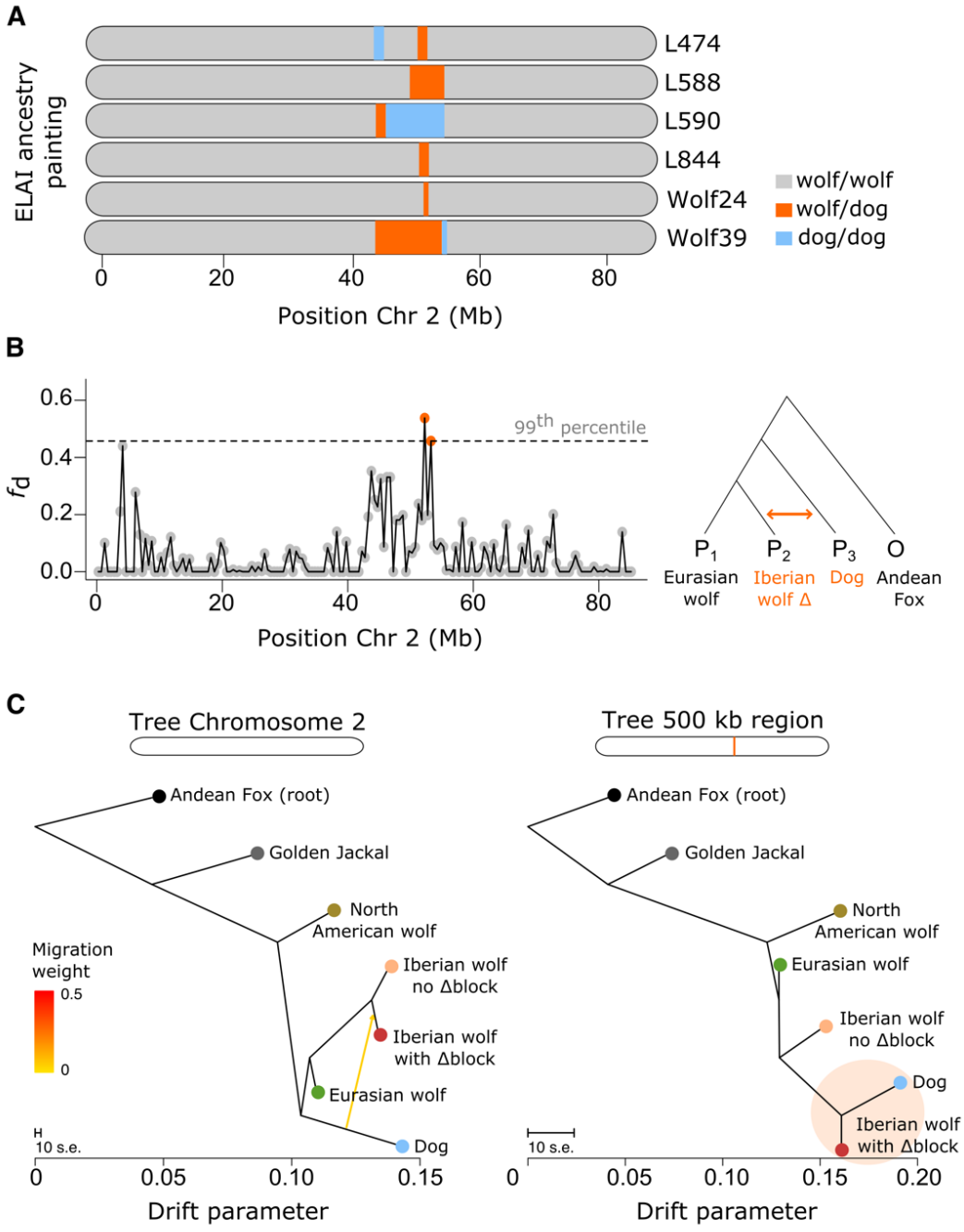


Figure 3. Signatures of excess allele sharing and phylogenetic discordance. (**A**) Representation of the Δblock on Chromosome 2 in six Iberian wolves using local ancestry analysis. Colors denote the attributed local ancestry: gray for homozygous wolf, orange for wolf/dog (heterozygous), and blue for homozygous dog. (**B**) Fraction of introgression (f_d) across

non-overlapping 500 kb windows on Chromosome 2. The test followed the phylogenetic arrangement depicted on the right: Eurasian wolves as P1, Iberian wolves with the ∆block as P2, and dogs as P3, with the Andean fox as the outgroup. The dashed line indicates the threshold for outlier regions, with orange dots representing 500 kb windows surpassing the 99th percentile. (**C**) Population trees estimated on Treemix for Chromosome 2 and the 500 kb region ranked as the top window in the f_d analysis, involving several canid species (Iberian, Eurasian, and North American wolves, dogs, Golden jackal, and Andean fox). The yellow arrow on the Chromosome 2 tree indicates migration (i.e., ancestry contribution) from dogs to Iberian wolves carrying the ∆block.

To further disentangle scenarios of incomplete lineage sorting and introgression, we compared levels of genetic differentiation (F_{ST}) and mean pairwise sequence divergence (d_{XY}) at the genomic background between Iberian wolves with the ∆block and dogs with those estimated for the shared 500 kb region. Low F_{ST} and low d_{XY} in the 500 kb region would support introgression over incomplete lineage sorting (Rosenzweig et al. 2016). Consistent with this scenario, F_{ST} levels across the 500 kb region were lower between Iberian wolves with the ∆block and dogs (0.05 *vs.* 0.27 at the genomic background level; Welch two-sample *t*-test: $p = 1.49\times10^{-9}$; **Supplemental Fig. S9**), but not between other Eurasian wolves and dogs (**Supplemental Fig. S9**). Similarly, d_{XY} levels across the 500 kb region were lower between Iberian wolves with the ∆block and dogs, particularly within a 100 kb window overlapping positions 52.4-52.5 Mb (0.0001 *vs.* 0.001 at the genomic background level; Welch two-sample *t*-test: $p = 1.28\times10^{-6}$; **Fig. 4C** and **Supplemental Fig. S10**), and higher between other Eurasian wolves and dogs (**Figs. 4B** and **4C**). Additionally, we also scanned Chromosome 2 haplotypes for regions of high sequence similarity between Iberian wolves, Eurasian wolves, and dogs, using HybridCheck (Ward and van Oosterhout 2016), and found the top outlier regions consistent with the position of the 500 kb region (**Supplemental Fig. S11**).

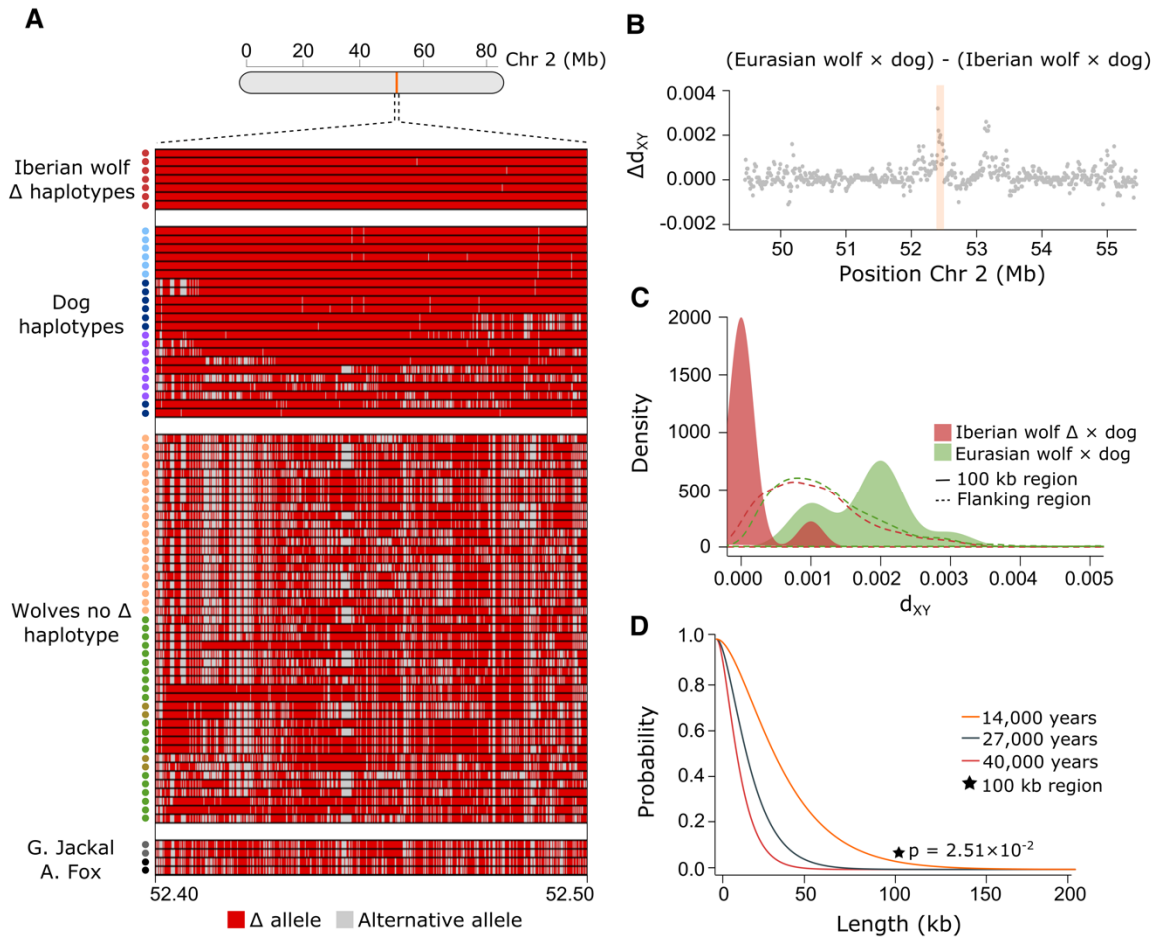


Figure 4. Haplotype sharing and sequence divergence between wolves and dogs. (**A**) Structure of the ~100 kb haplotype, represented by 593 SNPs, found within the 500 kb window with the lowest d_{XY} levels between Iberian wolves with the Δblock and dogs (Chr 2: 52.4-52.5 Mb). The top section shows the introgressed haplotype in Iberian wolves (red dots; L590 was homozygous across most positions, carrying two introgressed haplotypes); in the middle are the dog haplotypes (village dogs from Iberia in light blue dots, purebred dogs in dark blue, and village dogs from Asia and Africa in purple); followed by haplotypes found among wolves without the dog Δblock (Iberian in light orange dots, Eurasian in green, and North American in dark yellow). At the bottom are the Golden jackal and Andean fox haplotypes (gray and black dots, respectively). Each row represents a haplotype, and each position is colored according to whether it carries the introgressed allele (Δ in red) or the alternative allele (in gray). (**B**) Distribution of delta d_{XY} across all sites within the 6 Mb flaking region surrounding the introgressed block. This test identifies genomic regions with low sequence divergence specific to Iberian wolves and dogs and not in other Eurasian wolves (positive Δd_{XY} values). The orange bar indicates the position of the 100 kb region represented in (A). (**C**) Density distribution of d_{XY} between Iberian wolves with the Δblock

and dogs (red) and other Eurasian wolves and dogs (green) in the 100 kb region (full line) and in the genomic background (dashed line). (**D**) Probability distribution of maintaining a haplotype of 0-200 kb length due to ancestral shared variation, assuming a local recombination rate for the ∆block and three divergence time estimates between gray wolves and dogs (represented by distinct colors). The black star indicates the p-value associated with the introgressed haplotype (for 14 Kya).

We found that Iberian wolves with the ∆block and dogs shared a nearly identical 100 kb haplotype in the 100 kb window with low d_{XY} (52.4-52.5 Mb). This 100 kb region was otherwise highly polymorphic and possessed 593 segregating sites. Iberian wolves with the ∆block and dogs from the Iberian Peninsula differed by an average of only three nucleotide differences within the 100 kb haplotype (**Fig. 4A**). In contrast, this 100 kb haplotype was highly divergent from those found in other Eurasian wolves, with an average of 186 nucleotide differences across the 593 segregating sites (**Fig. 4A**). This supports that the dog ∆block originates from introgression due to a single hybridization event, and that the source was a dog from the Iberian Peninsula. The probability of finding only three nucleotide differences in a 100 kb haplotype with 593 segregating sites, where other Eurasian wolves possess on average 186 nucleotide differences, is exceptionally low (binomial test: $p = 3.90 \times 10^{-91}$). Moreover, within the 100 kb region, we identified 18 SNP positions where the ancestral allele was fixed in all gray wolves, while the derived allele was prevalent in dogs (75% frequency), nearly fixed in village dogs from Iberia (97%), and in Iberian wolves carrying the ∆block (60%) (**Supplemental Table S5**). These SNPs are positioned within a 5'UTR intron of the *MAST4* gene, which encodes a member of the microtubule-associated serine/threonine protein kinase (GeneCards, The Human Gene Database). A phylogenetic gene tree for *MAST4* unambiguously grouped Iberian wolves with the ∆block closer to dogs (**Supplemental Fig. S12**).

Lastly, we estimated the probability of maintaining an identical 100 kb haplotype without recombination in dogs and gray wolves since the time of divergence as extremely small (ranging from $p = 2.09\times10^{-6}$ to 2.51×10^{-2} for 40 - 14 Kya; Skoglund et al. 2015; Fan et al. 2016; Frantz et al. 2016; Perri et al. 2021; Bergström et al. 2022) (**Fig. 4D** and **Supplemental Fig. S13** for probability values using local or the chromosome-wide average recombination rates, respectively). Cumulatively, these findings provide compelling evidence that this haplotype is of dog origin rather than an ancestral shared polymorphism between wolves and dogs.

Age of ∆block introgression in Iberian wolves

After finding support for introgression in Iberian wolves, we reconstructed the evolutionary history of the introgressed dog block, aiming to determine its onset date. We employed the STARTMRCA method (Smith et al. 2018), which allows leveraging both the decay of linkage disequilibrium and the number of accumulated mutations in the introgressed 100 kb haplotype and flanking regions. Average TMRCA estimates ranged between ~1360 (861 – 1,857 $_{\text{min-max}}$) and ~670 (375 – 1,072 $_{\text{min-max}}$) generations ago, assuming local or the chromosome-wide average recombination rates, respectively, and a mutation rate of 4.5×10^{-9} (Koch et al. 2019). The choice of a smaller mutation rate (4×10^{-9}; Skoglund et al. 2015) resulted in similar time estimates (**Supplemental Fig. S14**). Assuming a mean generation time of 4.5 years for wolves (Mech et al. 2016), this places the time since introgression between ~6,100 and 3,000 years ago.

Signatures of selection support adaptive introgression

To critically assess whether the observed pattern of admixture within the 100 kb region in Iberian wolves could be explained by neutral evolution, we conducted individual-based

forward simulations using SLiM (Haller and Messer 2019). We explored multiple parameter combinations of number of migration events, number of migrants, and time since introgression (see Methods). In neutral simulations, the distinct combinations failed to reach the level of empirical admixture observed in the introgressed 100 kb region and in the genomic background (**Supplemental Figs. S15-S17**). Conversely, models that simulated adaptive introgression successfully recovered the observed empirical patterns of admixture in the introgressed 100 kb region (**Supplemental Fig. S18**). To investigate the genomic architecture and amount of selection required to retain an introgressed block we simulated a range (1-6) of SNPs inside the simulated introgressed haplotype with a range of selection coefficients (s). Among alternative adaptive introgression models, those simulating strong selection acting over more than three adaptive SNPs (with a combined selection of 0.1) had a better fit the observed level of admixture in this region (**Supplemental Figs. S18** and **S19**). Positive selection acting on multiple SNPs creates linkage that would have hindered the breakdown of the introgressed 100 kb haplotype. Moreover, we also evaluated distinct genetic dominance models – additive, dominant or overdominant – while considering different timings for the introgression event. We found that the overdominance model (i.e., heterozygote advantage) provided the best fit to the empirical admixture and allelic frequencies observed for the SNPs within the 100 kb region (**Supplemental Figs. S18** and **S19**). However, for certain timings for the introgression (between 400 and 900 generations) we also observed successful replicates in the dominance model under lower selection coefficients (**Supplemental Fig. S19**). Combined, our simulation results strongly favor selection, either positive or balancing selection, maintaining the 100 kb introgressed haplotype in the Iberian wolf.

Discussion

Gene flow has been demonstrated to be widely pervasive in canids (Gopalakrishnan et al. 2018). This is exemplified by events of adaptive introgression in genes associated with coat color, immune response (Anderson et al. 2009; Schweizer et al. 2018), and hypoxia (Miao et al. 2017; Wang et al. 2020) between wolves and their domestic counterparts. Dog introgression has also been suggested as a potential powerful force in the evolutionary response of wolves facing new anthropogenic pressures (Newsome et al. 2017; Pilot et al. 2021), despite no strong empirical evidence. Our study validates the introgression of dog genetic variants in a wolf population persisting for millennia in densely human populated areas (Llaneza et al. 2012; Llaneza et al. 2016; Dennehy et al. 2021). This finding suggests that introgression may be playing an important role in wolf adaptation to highly human-dominated landscapes.

Among Iberian wolves, the introgressed haplotype seems to coalesce into a single haplotype, strongly suggesting its origin from a single hybridization event. Such reduced nucleotide variability is unlikely to be explained by demographic events, such as the 1970s bottleneck experienced by the Iberian wolf (Sastre et al. 2011; Clavero et al. 2022; Nores and López-Bao 2022). Under this scenario, we would expect similar genome-wide signatures, which was not observed. Moreover, our study demonstrates that the introgressed haplotype was already present at the same frequency in the historical population, predating the bottleneck. SLiM simulations further demonstrate that, regardless of the amount or timing of gene flow simulated, neutral scenarios consistently failed to explain the differential levels of admixture observed in the introgressed haplotype. The presence of a single genomic region with dog ancestry consistently maintained at high frequency across the genome of Iberian wolves also suggests selection (Taylor and Larson 2019; Aguillon et al. 2022). Our simulations favored an overdominance model, suggesting heterozygote advantage, which raises the

possibility of balancing selection maintaining the introgressed haplotype. Although balancing selection can facilitate introgression by conferring advantage to novel alleles (Fijarczyk et al. 2018, Schweizer et al. 2018), our data do not allow us to exclude the possibility of positive selection. Variables such as historical population size, the timing of selection onset, genetic drift, and past genetic structure within the Iberian wolf population could influence the rate of allele fixation under positive selection, potentially explaining the observed intermediate allelic frequencies.

Our estimates suggest an ancient introgression event occurring between 6,100 and 3,000 years ago, with the high haplotype similarity to dogs from the Iberian Peninsula indicating a local origin for the hybridization event. This timeframe coincides with significant human-driven landscape changes, including widespread deforestation and increased agricultural activity in the Iberian Peninsula (Tereso et al. 2016). The growing presence of dogs in human settlements during this period may have facilitated hybridization between wolves and dogs (Albizuri et al. 2021). Notably, previous studies have suggested that hybridization was already evident during the Chalcolithic in Iberia (Catagnano 2016). Abrupt climatic events documented for this period in the Iberian Peninsula (Bernal-Wormull et al. 2023) may have also impacted wolf population dynamics, potentially facilitating hybridization with dogs (Lobo et al. 2023).

One of the key challenges in investigating adaptive introgression is establishing a clear link between introgressed variants and specific functional or phenotypic traits in the recipient species (Suarez-Gonzalez et al. 2018; Jones et al 2018; Grant and Grant 2019; Taylor and Larson 2019; Ferreira et al. 2023). The dog variants found on Chromosome 2 in Iberian wolves are located within a 5'UTR intron of the *MAST4* gene. Increasing empirical evidence reveals that *MAST4* is associated with several neurologic disorders, including developmental cognitive and motor delays, as well as infantile spasms (Mala

Cards, Human Disease Database; Strupp et al. 2020; Zhang et al. 2023). Additionally, *MAST4* appears to be differentially expressed in the prefrontal cortex of atypical cases of frontotemporal lobar degeneration (Martins-de-Souza et al. 2012). Although this gene has been primarily studied in the context of neurobiology, emerging research has also linked it to bone development (Cui et al. 2022; Kim et al. 2022) and spermatogenesis (Lee et al. 2021), with knockout mice exhibiting reduced body size and increased infertility.

An experimental demonstration of the adaptive function of *MAST4* in a protected wild large carnivore like the Iberian wolf is not feasible. However, although we did not directly assess the functional significance of the introgressed variants in the *MAST4* gene, we hypothesize that these variants may be linked to immature cognitive development in Iberian wolves, mirroring the juvenile cognitive phenotype typically retained by domestic animals into adulthood (Wilkins et al. 2014). While speculative, under this hypothesis, such immature cognitive development might partially contribute to the lack of long-distance dispersal observed in Iberian wolves, a behavior also common in domestic dogs and wolf pups (Jimenez et al. 2017; Morales-González et al. 2021). The frequency of introgression – approximately 30% using the SNP data, or 43% using the whole-genome dataset – suggests that the dog-derived genetic contribution alone cannot fully account for the dispersal behavior of Iberian wolves, and other factors are likely also at play (Silva et al. 2018). We note, however, that the frequency may be underestimated due to the limited genomic resolution of the SNP panel, as demonstrated by comparisons with whole-genome data. Future research, including functional genomic experiments in model vertebrate species (e.g., Bono et al. 2015; Li et al. 2024), will be essential to test our hypothesis and further explore the potential role of *MAST4* in the cognitive behavior of Iberian wolves.

Methods

Population genomics approach: contemporary samples

Canine HD SNP BeadChip data

We generated genome-wide SNP data for a comprehensive sample set comprising 95 Iberian wolves, five Eurasian wolves, and 62 dogs (**Supplemental Table S1**). Iberian wolf samples, collected between 1996 and 2017, were obtained from muscle and blood samples (mainly from road kills), spanning the entire wolf distribution range in the Iberian Peninsula (**Fig. 1B**). The sampling strategy was designed to be representative of the entire Iberian wolf population. Dog samples, sourced from local shelters and collaborators, included muscle, blood and buccal swabs. Eurasian wolf muscle samples were donated by collaborators, and no animals were sacrificed for this study. Total genomic DNA was extracted using the QIAGEN DNeasy Blood & Tissue Kit (Qiagen) and quantified on the Qubit DNA quantification system (Thermo Fisher Scientific) using Qubit broad range assay reagents, following the manufacturer's protocol. DNA concentration across all samples was normalized to 50 ng/µl to be genotyped for approximately 170,000 genome-wide SNPs, using the Canine HD BeadChip microarray (Illumina, Inc., San Diego, CA, USA). GenomeStudio software (Illumina) was employed for genotype calling following Illumina's guidelines. Our dataset was expanded with two additional datasets genotyped using the same technique: 58 dogs, including 30 breeds, from the LUPA project (Vaysse et al. 2011; Lequarré et al. 2011) and 50 European wolves (Stronen et al. 2015) (**Supplemental Table S1**). The three datasets were merged using PLINK v.1.9 (Purcell et al. 2007) after converting the SNP coordinates of the two additional datasets to the CanFam3.1 dog genome assembly using the *liftOver* tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver). SNP distribution and density along the

genome were verified using the R/Bioconductor package *karyoploteR* (Gel and Serra 2017) in R (R Development Core Team 2017). The final dataset encompassing 95 Iberian wolves, 55 Eurasian wolves, and 120 dogs, excluded closely related individuals (identity-by-descent > 0.5). Only autosomal SNPs without multiple positions were retained, and filters for high call rates per sample (> 0.95) and per SNP (> 0.98) were applied. Loci with minor allele frequency (MAF) below 0.01 were removed, resulting in a final dataset of ~85,000 SNPs. For global ancestry inference analyses (PCA and ADMIXTURE), we applied a light pruning on linkage disequilibrium (LD) using $r^2 \geq 0.8$, with a sliding-window size of 50 SNPs, shifted and recalculated every 10 SNPs (LD-pruned dataset resulted in ~79,000 SNPs). All filtering processes were implemented in PLINK.

Global ancestry inference

Global ancestry proportions were initially explored through a Principal Component Analysis (PCA) using PLINK. Subsequently, we used ADMIXTURE v.1.3 (Alexander et al. 2009) to estimate ancestry proportions (q) with a maximum likelihood model, focusing on K = 2 to distinguish between wolves and dogs. Separate runs were conducted for Iberian and Eurasian wolves. ADMIXTURE was run with the entire LD-pruned dataset in 2000 iterations, implementing a 10-fold cross-validation procedure (Alexander and Lange 2011).

Local ancestry analysis

We employed local ancestry methods to infer the identity of individual chromosomal blocks within the Iberian wolf genome (Liu et al. 2013; Geza et al. 2019), testing two distinct methods: LAMP-ANC v.2.5 (Sankararaman et al. 2008; Pasaniuc et al. 2009)

and ELAI (Guan 2014). LAMP-ANC, a non-LD-based method, estimates the most likely ancestry per SNP within windows based on reference allele frequencies (Sankararaman et al. 2008). For this method, Iberian wolves (N = 95) were considered as the admixed population, and Eurasian wolves (N = 55) and dogs (N = 120) were set as reference populations. We assumed 10 generations since the beginning of admixture, considering a generation time of 4.5 years (Mech et al. 2016). LAMP-ANC is not design for older admixture events, as additional generations lead to an overestimation of admixture proportions, especially when reference populations are closely related (Liu et al. 2013). We considered a genome-wide recombination rate of 9.7×10^{-9} (Campbell et al. 2016; Wong et al. 2010), a mixture proportion of 0.99:0.01 (based on global ancestry estimates), and an LD cut-off of $r^2 > 0.1$, as LAMP-ANC assumes unlinked markers. Additionally, we ran ELAI, which employs a two-layer hidden Markov model (HMM) for local ancestry inference without requiring phased data or a prior window size definition (Guan 2014). ELAI can also infer an unsampled reference population based on allele frequencies of the admixed population (Seixas et al. 2018). Given the absence of a true Iberian wolf reference population, we ran ELAI considering three reference populations (Eurasian wolves, dogs, and an unsampled population) with the number of upper-layer clusters set to -C 3 and lower-layer to -c 15 (five times the number of -C, as recommended). ELAI was run using the filtered dataset (~85K SNPs), performing 20 expectation-maximization (EM) steps, and testing several mixture generations, starting with 10 generations and incrementing in intervals of 100 up to 1000.

SNP-specific Δstatistics and validation of introgression

Local ancestry methods, while very informative, lack statistical significance for each identified ancestry block. To address this, we computed SNP-specific delta (Δ) statistics

(Tang et al. 2007) to find regions with a significantly elevated proportion of dog ancestry across individuals. The ∆statistics for parental population B at marker *m* is expressed as $q_B^m - q_{B,}$ where q_B^m denotes the mean ancestry at the marker *m* over all introgressed individuals, and q_B is the mean genome-wide ancestry across all individuals. Therefore, positive ∆ancestry values indicate core hotspot regions for introgression from population B across the genome of all individuals. We computed ∆ancestry for all chromosomes using local dog ancestry estimates determined by ELAI. To minimize the risk of false positives, we focused exclusively on the top genome-wide regions as candidates for further analysis, rather than applying a genome-wide threshold of two or three standard deviations (SD) from the mean as in previous canid studies (vonHoldt et al. 2016; Pilot et al. 2021). Additionally, we only considered regions as outliers if they were also detected by LAMP-ANC. A region on Chromosome 2 emerged as the top genome-wide outlier, representing a 12 SD from the mean genome-wide ∆ancestry. We then identified outlier SNPs within Chromosome 2 by applying a threshold of 3 SD from the mean chromosome ∆ancestry, resulting in a block spanning 78 SNPs (∆block).

To delve into the origin of the ∆block, we constructed a neighbor-joining (NJ) tree using SNPs within this region, based on pairwise genetic distances between Iberian wolves and village dogs from Iberia (the most abundant dogs across the wolf range). The NJ tree distances were estimated with Tassel 5 (Bradbury et al. 2007) as 1 – IBS (identity by state), where IBS represents the probability that alleles drawn at random from two individuals at the same locus are the same. The NJ was built also with Tassel 5 and drawn with FigTree v.1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). We also performed a PCA with PLINK, using the same set of 78 SNPs, to examine how individuals clustered. Additionally, a median-joining haplotype network for ∆block was constructed using POPART (Leigh and Bryant 2015) with reconstructed haplotypes of

Iberian wolves and village dogs from Iberia generated in PHASE v.2.1 (Stephens and Donnelly 2003) with the *LocusType* option set to *S* for bi-allelic SNPs.

Population genomics approach: historical samples

Historical samples and in-solution capture enrichment

To explore the historical prevalence of the ∆block on Chromosome 2 in Iberian wolves, we analyzed a set of 48 historical samples collected across the former range distribution in the Iberian Peninsula, spanning from 1912 to 2005 (**Fig. 1B**; **Supplemental Table S4**). This historical dataset, generated in parallel studies (Pacheco et al. 2022; Lobo et al. 2023), resulted from an in-solution target capture enrichment to obtain 100,000 genome-wide SNPs. The position of these SNPs was defined based on the coordinates available on the Canine HD BeadChip, ensuring its compatibility (full details about the methodology – bait design, capture experiment, sequencing, processing of raw reads and genotyping – can be found in the original publication). Only genotypes with a read depth above four were retained, and concordance rates with the Canine HD BeadChip genotypes were estimated (>99%, based on two contemporary samples). Only samples with more than 20,000 SNPs and missing data below 10% were kept, resulting in a final dataset with ~23,000 SNPs. The historical and contemporary datasets were merged, and filters for call rates per sample (> 0.85) and per SNP (> 0.90) were applied, and loci with MAF < 0.01 were removed, culminating in a combined dataset of ~18,000 SNPs. All the previous steps were performed in PLINK.

Local ancestry analysis in historical samples

The introgression analysis in historical samples was performed using the local ancestry method LAMP-ANC, following the methodology outlined for the contemporary dataset.

Although LAMP-ANC is not as accurate as LD-based methods, we choose it due to its recommendation when the genome-wide SNP density is not too high (Sankararaman et al. 2008; Pasaniuc et al. 2009). We ran the analysis on the merged dataset comprising historical and contemporary samples of Iberian wolves and dogs, encompassing ~18,000 SNPs. Reference populations comprised all dogs (N = 120) and contemporary Iberian wolves without any detected introgression signs on Chromosome 2 (N = 70). All the historical samples were considered as the admixed population. LAMP-ANC was run under the previously specified parameters. To mitigate potential biases from the limited number of loci in this dataset (Chr 2 = 623 SNPs), we subsampled the historical dataset, focusing solely on Iberian wolf samples carrying the ∆block on Chromosome 2 with at least 60,000 genome-wide SNPs (N = 13). This subset was merged with contemporary samples from Iberian wolves and dogs, and all the previously described filters were applied, resulting in a combined dataset of ~48,000 SNPs. A second LAMP-ANC run was then conducted, employing these 13 Iberian wolves as the admixed population, and maintaining all prior parameters. Outlier regions were identified for both LAMP-ANC runs using ∆ancestry. Subsequently, we constructed a NJ tree with all contemporary Iberian wolves, the subset of historical Iberian wolves (> 60K SNPs), and village dogs from Iberia, using SNPs within the ∆block (69 SNPs). Genetic distances and the NJ tree were estimated with Tassel 5, and the tree was drawn using FigTree.

Whole-genome sequencing approach

Whole-genome sequences and SNP calling

To attain the requisite high resolution for detecting small genomic dog blocks and confirming the presence of the ∆block in Chromosome 2 of Iberian wolves, we generated whole-genomes for 12 of the contemporary Iberian wolves used in the SNP chip dataset

(**Fig. 1B**; **Supplemental Table S3**). DNA extraction was performed using a Thermo Scientific™ KingFisher™ instrument according to the manufacturer's guidelines and resulting DNA fragments were fragmented in the Covaris LE220 plus Focused-ultrasonicator to achieve 350 bp fragment length. Fragmented DNA fragments were converted into BGISeq compatible double-stranded DNA sequencing libraries using the "single-tube" library building protocol BEST, as initially described by Carøe at al. (2018) and modified for BGISeq technology following Mak et al. (2017). Subsequently, libraries were pooled and sequenced on a DNBSEQ-G400 instrument, using a PE 150 bp chemistry to an average depth of 13.5x per sample at BGI-Europe, Copenhagen, Denmark. Sequence reads were processed and aligned using the BAM pipeline of PALEOMIX v.1.2.13.3 (Schubert et al. 2014). During the initial steps of the pipeline, low quality and N bases were trimmed from the reads, adapter sequences were removed, and overlapping read pairs were collapsed using AdapterRemoval v.2.2.0 (Schubert et al. 2016) with default parameters. Reads were mapped to the dog reference genome CanFam3.1 (Lindblad-Toh et al. 2005) using BWA v.0.7.16a (Li and Durbin 2009) with backtrack algorithm, considering a minimum base quality of 0 to ensure that all the reads were retained in the process. Reads were subsequently filtered for PCR duplicates using Picard MarkDuplicates (https://broadinstitute.github.io/picard/). In the final step of the pipeline, local realignment around indels was performed with IndelRealigner module of GATK v.3.8 (McKenna et al. 2010). Before SNP calling, we recalibrated the base quality scores (BQSR) using BaseRecalibrator/ApplyBQSR modules of the GATK v.4.1.8.1. Genotypes were then called for each sample using BCFtools v.1.10.2 (Li 2011) mpileup/call -m tools, with minimum Phred-scaled thresholds of 20 for base quality (BQ) and read mapping quality (MQ), and all the sites were emitted in the output. The genotypes were subsequently filtered using VCFtools (Danecek et al. 2011) to keep only

bi-allelic and autosomal SNPs that have passed previous filters, had less than 20% of missing data across the 12 samples, and were supported by at least four reads.

To this dataset, we added a set of 27 publicly available whole-genome sequences from several worldwide canids (14 gray wolves, including two Iberian wolves, 11 dogs, one Golden jackal, and one Andean fox) (**Supplemental Table S3**). Genotypes from the 27 canids were extracted from a VCF file containing 91 million variants and 722 canid genomes created by Plassais et al. (2019), available on NCBI (accession number: PRJNA448733). This VCF file was then filtered as described above and merged with the VCF file containing the genotypes of the 12 Iberian wolves, using the BCFtools merge tool. In the final dataset, sites with less than 10% missing and with a minor allele frequency greater or equal to 1% were retained, resulting in 11.7 million SNPs ("variants only" VCF). Finally, we estimated missing and depth statistics across all individuals using VCFtools, ensuring that all individuals had less than 10% missing data and an average coverage above 17× across the 11.7 million SNPs.

Patterns of allele sharing and phylogenetic reconstruction

We first explored the genome-wide mixture proportion between wolves and dogs by conducting ADMIXTURE analysis for K = 2, replicating the approach used for the SNP chip dataset. To validate the ∆block on Chromosome 2, we ran ELAI considering three reference populations (Eurasian wolves, N = 10; dogs, N = 11; and an unsampled population), while treating all Iberian wolves as admixed (N = 14), following the previously described parameters.

To quantify the fraction of introgression, we computed f_d statistics (Martin et al. 2015) along Chromosome 2 (420,698 SNPs) in non-overlapping 500 kb windows, each containing at least 100 SNPs, using the *ABBABABAwindows.py* script from the

genomics_general package, available at https://github.com/simonhmartin/genomics_general. This test was conducted based on the following phylogenetic configuration: P1 – Eurasian wolves (N = 10), P2 – Iberian wolves with the ∆block (N = 6, as defined by ELAI), P3 – dogs (N = 11), and the Andean fox serving as outgroup. f_d statistics will measure the excess shared variation between P3 and P2 that is not shared with P1, excluding the hypothesis of incomplete lineage sorting. Significantly introgressed regions were identified as those above the 99th percentile.

To unveil potential discordances in tree topology between introgressed regions and the expected species tree, we reconstructed population trees for Chromosome 2 and the top 500 kb outlier region from the window-based f_d statistics. Population trees were constructed in TreeMix v.1.13 (Pickrell and Pritchard 2012), defining seven major groups: Iberian wolves with the ∆block (N = 6), Iberian wolves without the ∆block (N = 8), Eurasian wolves (N = 10), North American wolves (N = 2), dogs (N = 11), one Golden jackal, and an Andean fox set as the root. Allelic frequencies for each group were estimated with PLINK, and the output file was converted into the TreeMix input with *plink2treemix.py* script provided in TreeMix. Maximum likelihood trees were constructed by running TreeMix in blocks of 200 SNPs with 200 bootstrap replicates. For the Chromosome 2 tree, one migration event was incorporated into the analysis. TreeMix was run 100 times and the trees were summarized with SumTrees of the DendroPy library (Sukumaran and Holder 2010) and visualized with FigTree. Individual-based NJ trees were also reconstructed using genome-wide SNPs (11.7 million) and SNPs for Chromosome 2 (420,698 SNPs). In both cases, SNPs were concatenated to generate pseudo FASTA alignments, and the genetic distances and the NJ trees were estimated using Tassel 5 and visualize on FigTree.

Genetic sequence divergence and haplotype reconstruction

To examine the genetic similarity between Iberian wolves carrying the ∆block, dogs, and other Eurasian wolves, we initially estimated F_{ST} levels between these groups along Chromosome 2 in non-overlapping windows of 500 kb using VCFtools. Levels of differentiation between the ∆block and the genomic background (defined as the entire Chromosome 2 excluding the introgressed block) were compared using a Welch two-sample *t*-test in R. To identify the haplotype(s) of the ∆block, genotypes on Chromosome 2 were phased using Beagle v.5.1 (Browning and Browning 2007) with the aid of the dog recombination map (retrieved from: https://github.com/cflerin/dog_recombination). HYBRIDCHECK v.1.0 (Ward and van Oosterhout 2016) was then employed to screen the haplotypes for blocks of high sequence similarity in 1 kb windows. Pseudo sequences of Chromosome 2 were created by concatenating phased SNPs for each individual. Pairwise comparisons were performed in sets of triplets (Iberian wolf with the ∆block – Eurasian wolf – Portuguese village dog) for each of the six individuals carrying the ∆block.

To distinguish between ancestral shared polymorphism and introgression, we estimate the mean pairwise sequence divergence (d_{XY}) between haplotypes of Iberian wolves with the ∆block, Eurasian wolves, and dogs from the Iberian Peninsula. To ensure specificity, we used village dogs from Iberia as they likely represent the closest source for the donor of introgression. Considering that d_{XY} analysis requires invariant sites, we considered that all the genotyped invariant sites that were not present in the combined VCF ("variants only") were homozygous for the reference, using the *--missing-to-ref* option from BCFtools. Sites in a 6 Mb region (Chr 2: 49-55.5 Mb) surrounding the midpoint of the 500 kb outlier region from the window-based f_d statistics were then

extracted, resulting in 5.4 million sites. Flanking regions were treated as the genomic background. Haplotypes within this 6 Mb region were reconstructed using Beagle v.5.1, and d_{XY} was calculated using the *popgenWindows.py* script from genomics_general package in 10 kb non-overlapping windows with at least 100 sites, using the -f flag set to *haplo* to handle haploid data. Welch's *t*-test was applied to compare sequence divergence between the genomic background and outlier regions. Density distributions were calculated using the R package *sm* (Bowman and Azzalini 2021). Genotypes from outlier regions in d_{XY} analysis were manually inspected using Tassel to identify signatures of allelic shared variation between dogs and Iberian wolves with the ∆block that were not observed with other wolves (e.g., major allele in dogs was only present in Iberian wolves with the ∆block).

Probability of maintaining the introgressed haplotype from shared ancestral variation

After identifying a common 100 kb haplotype in all Iberian wolves with the ∆block and dogs, we calculated the probability of maintaining the introgressed haplotype due to ancestral polymorphism. We followed the approach established by Huerta-Sánchez et al. (2014) and Miao et al. (2017), which follows a Gamma distribution:

$$p = 1 - pgamma\,(m, shape = 2, rate = 1/L)$$

Here, *L* represents the expected length of a shared ancestral haplotype and is defined as $L = 1/(r \times t)$, where *r* is the recombination rate per generation per bp, and *t* is the time since the divergence between wolves and dogs. We considered a divergence range of 40-14 Kya (Skoglund et al. 2015; Fan et al. 2016; Frantz et al. 2016; Perri et al. 2021; Bergström et al. 2022). To account for specific variability in recombination rates within the introgressed haplotype, we used the Chromosome 2 recombination map inferred for

dogs by Campbell *et al.* (2016). Given the limited genomic resolution across the 100 kb haplotype, we estimated the recombination rate for the entire ∆block to be ~0.37 cM/Mb. For comparison, we also performed the analysis using the sex-averaged recombination rate for Chromosome 2 reported by Campbell *et al.* (2016) at 0.84 cM/Mb. We report p-values for divergence estimates of 14,000, 27,000, and 40,000 years.

Individual-based forward simulation in SLiM

To assess the adaptive potential of the introgression, we implemented individual-based forward simulations in SLiM 3 (Haller and Messer 2019). The main goal was to investigate whether introgressed variants could have evolved under neutrality and to estimate the amount of selection required to recover the observed level of admixture in the empirical data. We simulated the entire Chromosome 2 of the dog reference genome CanFam3.1, along with its local estimates of recombination (Campbell et al. 2016), to capture realistic dynamics of the genomic background. Our model simulates two divergent lineages, representing dogs and wolves, that experience migration. We simulated the wolf lineage with an effective population size (Ne) of 3,000, following Silva et al. (2020). Unique SNPs were introduced every 10 kb in the dog lineage across the entire chromosome to track their ancestry upon introgression into the wolf genome.

Initially, we simulated a neutral model across a range of scenarios for the timing and amount of introgression, considering parameters such as the number of migration events (single, one every generation, or one every 100 generations) and the number of migrants (one, 1%, 2% or 4% of the wolf Ne). For each set of parameters simulations ran for 1,200 generations, with migration starting at generation 10, and 100 replicates were performed. Subsequently, we simulated different adaptive scenarios to investigate whether non-neutral models were a better fit to the empirical data. Given that neutral scenarios with higher and more frequent migration consistently led to a signal of greater

chromosome-wide admixture than observed in the empirical data, we conducted the adaptive scenario using the parameters that reflected realistic dynamics in the neutral model (i.e., one migration event with a single migrant). In the adaptive simulations, we introduced positively selected alleles (referred to as “introgressed alleles”) inside the 100 kb introgressed haplotype in the dog genome. To test the impact of linkage and the genomic architecture of introgressed alleles, we simulated a range (1-6) of SNPs inside the introgressed haplotype. To investigate the magnitude of selection required for introgression to take place, we varied the selection (s) and the number of adaptive SNPs, calculating the combined selection by summing selection coefficients across adaptive SNPs as $\sum s$. Additionally, we investigated the mode of selection by varying the dominance coefficients (additive h = 0.5, dominant h = 1, or overdominance h = 2). Lastly, we assessed the true admixture proportion of ancestry-tracking SNPs and the frequency of introgressed SNPs inside the introgressed haplotype 100 kb, considering the estimated average time for introgression (see Results).

Estimating time of dog introgression

We estimated the time since introgression of the 100 kb haplotype in Iberian wolves using STARTMRCA (Smith et al. 2018), an Markov chain Monte Carlo (MCMC) based method that leverages LD decay between the selected allele and nearby sites, along with new mutations in the introgressed haplotype. This method calculates the time to the most recent common ancestor (TMRCA), assuming the haplotype has been subjected to positive selection. It requires a panel of reference haplotypes without the selected allele with which the selected haplotype has recombined following introgression. For this, we used the haplotypes of Iberian wolves without the introgression to represent the reference haplotypes. We centered the analysis on Chr 2: 52,449,886 (central position in

the introgressed 100 kb haplotype) and included 1 Mb of upstream and downstream sequence (Chr 2: 51,449,886-53,449,886), as recommended. This resulted in a genomic region comprising 1.8 M sites. We used the recombination map from Campbell *et al.* (2016) to infer the local recombination rate for the entire 2 Mb region, estimating it at ~0.49 cM/Mb. TMRCA estimates were also calculated using the sex-averaged recombination rate for Chromosome 2, reported as 0.84 cM/Mb (Campbell et al. 2016; Wong et al. 2010). Additionally, we tested two widely used mutation rates for dogs and wolves, specifically 4×10^{-9} and 4.5×10^{-9} per base pair per generation (Skoglund et al. 2015; Koch et al. 2019). Ten independent MCMC chains were run, each with 50,000 iterations and a standard deviation of 20 for the proposal distribution. The final 10,000 iterations from each chain were used to generate posterior TMRCA distributions. Estimates of TMRCA were then converted to time in years using a generation time of 4.5 years per generation (Mech et al. 2016). The results were visualized in violin plots using the R package *vioplot* v.0.3.5 (https://github.com/TomKellyGenetics/vioplot) in R.

Data access

The SNP genotypes of all contemporary samples generated in this study have been submitted to The Open Science Framework (OSF; https://osf.io/) under DOI 10.17605/OSF.IO/NSZ9K. All raw whole-genome sequencing data generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA1078274.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

This research was funded by the Portuguese Foundation for Science and Technology, FCT, under the projects PTDC/BIA-EVF/2460/2014, DivProtect/0012/2021, and UIDP/50027/2020. DL, RG, PS and CP were supported by FCT (PhD grant PD/BD/132403/2017 and research contract in project DivProtect/0012/2021 to DL; research contract 2022.07926.CEECIND to RG; research contract in project PTDC/BIA-EVL/31902/2017 to PS; and PhD grant PD/BD/135026/2017 to CP). CVO was funded by the Earth and Life Systems Aliance (ELSA), Norwich Research Park. JVLB was supported by the Spanish Ministry of Economy, Industry and Competitiveness (RYC-2015-18932; CGL2017-87528-R AEI/FEDER EU), and by a GRUPIN research grant from the Regional Government of Asturias (AYUD/2021/51314). JA was supported by FEDER funds through the Operational Programme for Competitiveness Factors - COMPETE (POCI-01-0145-FEDER-029115) through FCT project PTDC/BIA-EVL/29115/2017. MTPG acknowledges ERC Consolidator Award 681396 “Extinction Genomics”, Danish National Research Foundation award DNRF143, and Norwegian Environment Agency project 18088069 for funding. GHA is supported by the national Council of Science and Technology in Mexico (CONACYT) Grant CVU 576734. We acknowledge the Portuguese Institute for Nature Conservation and Forestry (ICNF) for providing wolf samples in Portugal. Samples from wolves in Spain were collected by technical staff and wildlife agents of Consejería de Medio Ambiente del Principado de Asturias, Consellería de Medio Ambiente de la Junta de Galicia, Consejería de Medio Ambiente de la Junta de Castilla y León, Consejería de Agricultura, Medio Ambiente y Desarrollo Rural de la Junta de Castilla-La Mancha, Gobierno de Cantabria and Parque Nacional Picos de Europa. Our thanks extend to all the collaborators and veterinarians whose assistance facilitated the collection of dog samples in Portugal and Spain. We are

also indebted to the Museu Nacional de História Natural e da Ciência (Lisbon, Portugal), Museo Nacional de Ciencias Naturales (Madrid, Spain) and Estación Biológica de Doñana (Seville, Spain) for granting us access to the natural history collections of Iberian wolves. Our acknowledgments are extended to the owners of private collections who shared historical wolf samples. Furthermore, we express our gratitude to Sofia Mourão and Diogo Lima for laboratory and technical assistance, respectively. Finally, we are also indebted to the Associate Editor and two anonymous referees for all their helpful suggestions to improve the manuscript. This is scientific paper no. 33 from the Iberian Wolf Research Team (IWRT).

Authors' contribution

RG, NF, and DL designed the study. RG coordinated the project. RG, DL and MTPG generated data. JVLB and LL coordinated the sample collection efforts in Spain. DL, GP, DC, and CP conducted laboratory work. DL performed data analysis under the guidance of RG, PS and CVO. HEM performed SLiM simulations with the help of CVO. CP and GHA provided technical support. JA and PS helped with bioinformatic scripting. DL, CVO, and RG wrote the paper with input from all the other authors.

References

Aguillon SM, Dodge TO, Preising GA, Shumer M. 2022. Introgression. *Current Biology* **32**: R855-R873.

Albizuri S, Grandal-d'Anglade A, Maroto J, Oliva M, Rodríguez A, Terrats N, Palomo A, López-Cachero FJ. 2021. Dogs that Ate Plants: Changes in the Canine Diet During the Late Bronze Age and the First Iron Age in the Northeast Iberian Peninsula. *J World Prehistory* **34**: 75–119.

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

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in

unrelated individuals. *Genome Res* **19**: 1655–1664.

Allendorf FW, Leary RF, Spruell P, Wenburg JK. 2001. The problems with hybrids: setting conservation guidelines. *Trends Ecol Evol* **16**: 613–622.

Anderson TM, vonHoldt BM, Candille SI, Musiani M, Greco C, Stahler DR, Smith DW, Padhukasahasram B, Randi E, Leonard JA, et al. 2009. Molecular and Evolutionary History of Melanism in North American Gray Wolves. *Science* **323**: 1339–1343.

Benazzo A, Trucchi E, Cahill JA, Maisano Delser P, Mona S, Fumagalli M, Bunnefeld L, Cornetti L, Ghirotto S, Girardi M, et al. 2017. Survival and divergence in a small group: The extraordinary genomic history of the endangered Apennine brown bear stragglers. *Proc Natl Acad Sci* **114**: E9589–E9597.

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

Bernal-Wormull JL, Moreno A, Bartolomé M, Arriolabengoa M, Pérez-Mejías C, Iriarte E, Osácar C, Spotl C, Stoll H, Cacho I, et al. 2023. New insights into the climate of northern Iberia during the Younger Dryas and Holocene: The Mendukilo multi-speleothem record. *Quat Sci Rev*, **305**: 108006.

Boitani L. 2003. Wolf conservation and recovery. In: *Wolves: behaviour, ecology, and conservation*. eds L.D. Mech and L. Boitani. pp. 317–340. University of Chicago Press, Chicago.

Bono JM, Olesnicky EC, Matzkin LM. 2015. Connecting genotypes, phenotypes and fitness: harnessing the power of CRISPR/Cas9 genome editing. *Mol Ecol* **24**: 3810-3822.

Bowman A, Azzalini A. 2021. R package 'sm': nonparametric smoothing methods (version 2.2-6.0) URL http://www.stats.gla.ac.uk/~adrian/sm

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**: 2633–2635.

Browning SR, Browning BL. 2007. Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *Am J Hum Genet* **81**: 1084–1097.

Campbell CL, Bhérer C, Morrow BE, Boyko AR, Auton A. 2016. A Pedigree-Based Map of Recombination in the Domestic Dog Genome. *G3 Genes|Genomes|Genetics* **6**: 3517–3524.

Carøe C, Gopalakrishnan S, Vinner L, Mak SST, Sinding MHS, Samaniego JA, Wales N, Sicheritz-Pontén T, Gilbert MTP. 2018. Single-tube library preparation for degraded

DNA. *Methods Ecol Evol* **9**: 410–419.

Catagnano V. 2016. Aproximación morfométrica y paleogenética al estudio de la variabilidad de canis l. familiaris en la península ibérica desde el neolítico hasta época romana y su contextualización en el ámbito del mediterráneo occidental. Universitat Autònoma de Barcelona.

Chapron G, Kaczensky P, Linnell JDC, von Arx M, Huber D, Andrén H, López-Bao JV, Adamec M, Álvares F, Anders O, et al. 2014. Recovery of large carnivores in Europe's modern human-dominated landscapes. *Science* **346**: 1517–1519.

Clavero M, García-Reyes A, Fernández-Gil A, Revilla E, Fernández N. 2022. Where wolves were: setting historical baselines for wolf recovery in Spain. *Animal Conservation* **26**: 239-249.

Cui Y, Wang F, Zhang D, Huang J, Yang Y, Xu J, Gao Y, Ding H, Qu Y, Zhang W, et al. 2022. Estrogen-responsive gene *MAST4* regulates myeloma bone disease. *Journal of bone and mineral research* **37**: 711-723.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

Dennehy E, Llaneza L, López-Bao JV. 2021. Contrasting wolf responses to different paved roads and traffic volume levels. *Biodivers Conserv* **30**: 3133–3150.

Eriksson A, Manica A. 2012. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *Proc Natl Acad Sci* **109**: 13956–13960.

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

Ferreira MS, Thurman TJ, Jones MR, Farelo L, Kumar AV, Mortimer SME, Demboski JR, Mills LS, Alves PC, Melo-Ferreira J, et al. 2023. The evolution of white-tailed jackrabbit camouflage in response to past and future seasonal climates. *Science* **379**: 1238-1242.

Fijarczyk A, Dudek K, Niedzicka M, Babik W. 2018. Balancing selection and introgression of newt immune-response genes. *Proc. R. Soc. B* **285**: 201801819.

Frantz LAF, Bradley DG, Larson G, Orlando L. 2020. Animal domestication in the era of ancient genomics. *Nat Rev Genet* **21**: 449–460.

Frantz LAF, Mullin VE, Pionnier-Capitan M, Lebrasseur O, Ollivier M, Perri A, Linderholm A, Mattiangeli V, Teasdale MD, Dimopoulos EA, et al. 2016. Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science* **352**:1228–1231.

Gel B, Serra E. 2017. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33**: 3088–3090.

Geza E, Mugo J, Mulder NJ, Wonkam A, Chimusa ER, Mazandu GK. 2019. A comprehensive survey of models for dissecting local ancestry deconvolution in human genome. *Brief Bioinform* **20**: 1709–1724.

Godinho R, Llaneza L, Blanco JC, Lopes S, Álvares F, García EJ, Palacios V, Cortés Y, Javier T, Ferrand N. 2011. Genetic evidence for multiple events of hybridization between wolves and domestic dogs in the Iberian Peninsula. *Mol Ecol* **20**: 5154–5166.

Gómez-Sánchez D, Olalde I, Sastre N, Enseñat C, Carrasco R, Marques-Bonet T, Lalueza-Fox C, Leonard JA, Vilà C, Ramírez O. 2018. On the path to extinction: Inbreeding and admixture in a declining grey wolf population. *Mol Ecol* **27**: 3599–3612.

Gopalakrishnan S, Sinding MS, Ramos-Madrigal J, Niemann J, Castruita JAS, Vieira FG, Carøe C, de Manuel Montero M, Kuderna L, Serres A, et al. 2018. Interspecific gene flow shaped the evolution of the genus *Canis*. *Current Biology* **28**: 1-9.

Grant PR, Grant BR. 2019. Hybridization increases population variation during adaptive radiation. *Proc Natl Acad Sci* **116**: 23216–23224.

Guan Y. 2014. Detecting Structure of Haplotypes and Local Ancestry. *Genetics* **196**: 625–642.

Haller BC, Messer PW. 2019. SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Mol Biol Evol* **36**: 632–637.

Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, Liang Y, Yi X, He M, Somel M, et al. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**: 194–197.

Instituto Nacional de Estatística Portugal. 2017. A Península Ibérica em números 2016. In *Península Ibérica em números*, INE 2017, Lisboa.

Jimenez MD, Bangs EE, Boyd DK, Smith DW, Becker SA, Ausband DE, Woodruff SP, Bradley EH, Holyan J, Laudon K. 2017. Wolf dispersal in the Rocky Mountains, Western United States:1993-2008. *Journal of Wildlife Management* **81**: 581-592.

Jones MR, Mills LS, Alves PC, Callahan CM, Alves JM, Lafferty DJR, Jiggins FM, Jensen JD, Melo-Ferreira J, Good JM. 2018. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* **360**: *1355-1358*.

Kim P, Park J, Lee D, Mizuno S, Shinohara M, Hong CP, Jeong Y, Yun R, Park H, Park S, et al. 2022. Mast4 determines the cell fate of MSCs for bone and cartilage development. *Nature Communications* **13**:3960.

Koch EM, Schweizer RM, Schweizer TM, Stahler DR, Smith DW, Wayne RK, Novembre J.

2019. De novo mutation rate estimation in wolves of known pedigree. *Mol Biol Evol* **36**: 2536–2547.

Lee S, Park J, Lee D, Orsu K, Kim P, Mizuno S, Lee M, Kim H, Harada H, Takahashi S, et al. 2021. *Mast4* knockout shows the regulation of spermatogonial stem cell self-renew via the FGF2/ERM pathway. *Cell Death & Differentiation* **28**: 1441-1454.

Leigh JW, Bryant D. 2015. popart: full-feature software for haplotype network construction. *Methods Ecol Evol* **6**: 1110–1116.

Lequarré A-S, Andersson L, André C, Fredholm M, Hitte C, Leeb T, Lohi H, Lindblad-Toh K, Georges M. 2011. LUPA: A European initiative taking advantage of the canine genome architecture for unravelling complex disorders in both human and dogs. *Vet J* **189**: 155–159.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**: 2987–2993.

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

Li H, Wang X, Zhang R, Liu L, Zhu H. 2024. Generation of golden goldfish *Carassius auratus* via *tyrosinase* geen targeting by CRISPR/Cas9. *Aquaculture* **30**: 740594.

Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ, Zody MC, et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**: 803–819.

Liu Y, Nyunoya T, Leng S, Belinsky SA, Tesfaigzi Y, Bruse S. 2013. Softwares and methods for estimating genetic ancestry in human populations. *Hum Genomics* **7**: 1.

Llaneza L, García EJ, Palacios V, Sazatornil V, López-Bao JV. 2016. Resting in risky environments: the importance of cover for wolves to cope with exposure risk in human-dominated landscape. *Biodivers Conserv* **25**: 1515–1528.

Llaneza L, López-Bao JV, Sazatornil V. 2012. Insights into wolf presence in human-dominated landscapes: the relative role of food availability, humans and landscape attributes. *Divers Distrib* **18**: 459–469.

Lobo D, López-Bao JV, Godinho R. 2023. The population bottleneck of the Iberian wolf impacted genetic diversity but not admixture with domestic dogs: A temporal genomic approach. *Mol Ecol* **32**: 5986–5999.

Mak SST, Gopalakrishnan S, Carøe C, Geng C, Liu S, Sinding M-HS, Kuderna LFK, Zhang W, Fu S, Vieira FG, et al. 2017. Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing. *Gigascience*

6: gix049.

Martin SH, Davey JW, Jiggins CD. 2015. Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. *Mol Biol Evol* **32**: 244–257.

Martins-de-Souza D, Guest PC, Mann DM, Roeber S, Rahmoune H, Bauder C, Kretzschmar H, Volk B, Baborie A, Bahn S. 2012. Proteomic Analysis Identifies Dysfunction in Cellular Transport, Energy, and Protein Metabolism in Different Brain Regions of Atypical Frontotemporal Lobar Degeneration. *J Proteome Res* **11**: 2533–2543.

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

Mech LD, Barber-Meyer SM, Erb J. 2016. Wolf (Canis lupus) Generation Time and Proportion of Current Breeding Females by Age. *PLoS One* **11**: e0156682.

Mech LD, Boitani L. 2003. *Wolves: behavior, ecology, and conservation*. eds. L.D. Mech and L. Boitani. University of Chicago Press.

Miao B, Wang Z, Li Y. 2017. Genomic analysis reveals hypoxia adaptation in the tibetan mastiff by introgression of the gray Wolf from the tibetan plateau. *Mol Biol Evol* **34**: 734–743.

Morales-González, Fernández-Gil A, Quevedo M, Revilla E. 2021. Patterns and determinants of dispersal in grey wolves (*Canis lupus*). *Biol Rev* **97**: 466-480.

Murray DL, Smith DW, Bangs EE, Mack C, Oakleaf JK, Fontaine J, Boyd D, Jiminez M, Niemeyer C, Meier TJ, et al. 2010. Death from anthropogenic causes is partially compensatory in recovering wolf populations. *Biol Conserv* **143**: 2514-2524.

Newsome TM, Fleming PJS, Dickman CR, Doherty TS, Ripple WJ, Ritchie EG, Wirsing AJ. 2017. Making a New Dog? *Bioscience* **67**: 374–381.

Nores C, López-Bao JV. 2022. Historical data to inform the legal status of species in Europe: An example with wolves. *Biol Conserv* **272**: 109639.

Pacheco C, Lobo D, Silva P, Álvares F, García EJ, Castro D, Layna JF, López-Bao JV, Godinho R. 2022. Assessing the performance of historical skins and bones for museomics using wolf specimens as a case study. *Front Ecol Evol* **10**.

Pacheco C, López-Bao JV, García EJ, Lema FJ, Llaneza L, Palacios V, Godinho R. 2017. Spatial assessment of wolf-dog hybridization in a single breeding period. *Sci Rep* **7:42475**.

Pasaniuc B, Sankararaman S, Kimmel G, Halperin E. 2009. Inference of locus-specific ancestry in closely related populations. *Bioinformatics* **25**: i213–i221.

1009 Pereira P, Fandos Esteruelas N, Nakamura M, Rio-Maior H, Krofel M, Di Blasio A, Zoppi S,
1010 Robetto S, Llaneza L, García E, et al. 2022. Hair cortisol concentration reflects the life
1011 cycle and management of grey wolves across four European populations. *Sci Rep* **12**:
1012 5697.
1013 Perri AR, Feuerborn TR, Frantz LAF, Larson G, Malhi RS, Meltzer DJ, Witt KE. 2021. Dog
1014 domestication and the dual dispersal of people and dogs into the Americas. *Proc Natl*
1015 *Acad Sci* **118**:e2010083118.
1016 Pickrell JK, Pritchard JK. 2012. Inference of Population Splits and Mixtures from Genome-
1017 Wide Allele Frequency Data. *PLoS Genet* **8**: e1002967.
1018 Pilot M, Greco C, VonHoldt BM, Jędrzejewska B, Randi E, Jędrzejewski W, Sidorovich VE,
1019 Ostrander EA, Wayne RK. 2014. Genome-wide signatures of population bottlenecks
1020 and diversifying selection in European wolves. *Heredity* **112**: 428–442.
1021 Pilot M, Moura AE, Okhlopkov IM, Mamaev N V., Manaseryan NH, Hayrapetyan V, Kopaliani
1022 N, Tsingarska E, Alagaili AN, Mohammed OB, et al. 2021. Human-modified canids in
1023 human-modified landscapes: The evolutionary consequences of hybridization for grey
1024 wolves and free-ranging domestic dogs. *Evol Appl* **14**: 2433–2456.
1025 Plassais J, Kim J, Davis BW, Karyadi DM, Hogan AN, Harris AC, Decker B, Parker HG,
1026 Ostrander EA. 2019. Whole genome sequencing of canids reveals genomic regions
1027 under selection and variants influencing morphology. *Nat Commun* **10**: 1489.
1028 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de
1029 Bakker PIW, Daly MJ, et al. 2007. PLINK: A Tool Set for Whole-Genome Association
1030 and Population-Based Linkage Analyses. *Am J Hum Genet* **81**: 559–575.
1031 R Development Core Team. 2017. R: A language and environment for statistical computing.
1032 *Vienna, Austria*.
1033 Rhymer JM, Simberloff D. 1996. Extinction by hybridization and introgression. *Annu Rev*
1034 *Ecol Syst* **27**: 83–109.
1035 Rio-Maior H, Nakamura M, Álvares F, Beja P. 2019. Designing the landscape of coexistence:
1036 Integrating risk avoidance, habitat selection and functional connectivity to inform large
1037 carnivore conservation. *Biol Conserv* **235**: 178–188.
1038 Rosenzweig BK, Pease JB, Besansky NJ, Hahn MW. 2016. Powerful methods for detecting
1039 introgressed regions from population genomic data. *Mol Ecol* **25**: 2387-2397.
1040 Sankararaman S, Sridhar S, Kimmel G, Halperin E. 2008. Estimating Local Ancestry in
1041 Admixed Populations. *Am J Hum Genet* **82**: 290–303.
1042 Sastre N, Vilà C, Salinas M, Bologov VV, Urios V, Sánchez A, Francino O, Ramírez O. 2011.
1043 Signatures of demographic bottlenecks in European wolf populations. *Conserv Genet*

12: 701–712.

Sazatornil V, Rodríguez A, Klaczek M, Ahmadi M, Álvares F, Arthur S, Blanco JC, Borg BL, Cluff D, Cortés Y, et al. 2016. The role of human-related risk in breeding site selection by wolves. *Biol Conserv* **201**: 103–110.

Schubert M, Ermini L, Sarkissian C Der, Jónsson H, Ginolhac A, Schaefer R, Martin MD, Fernández R, Kircher M, McCue M, et al. 2014. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat Protoc* **9**: 1056–1082.

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

Schweizer RM, Durvasula A, Smith J, Vohr SH, Stahler DR, Galaverni M, Thalmann O, Smith DW, Randi E, Ostrander EA, et al. 2018. Natural Selection and Origin of a Melanistic Allele in North American Gray Wolves. *Mol Biol Evol* **35**: 1190–1209.

Seixas FA, Boursot P, Melo-Ferreira J. 2018. The genomic impact of historical hybridization with massive mitochondrial DNA introgression. *Genome Biol* **19**: 91.

Silva P, Galaverni M, Ortega-Del Vecchyo D, Fan Z, Caniglia R, Fabbri E, Randi E, Wayne R, Godinho R. 2020. Genomic evidence for the Old divergence of Southern European wolf populations. *Proc R Soc B Biol Sci* **287**: 20201206.

Silva P, López-Bao JV, Llaneza L, Álvares F, Lopes S, Blanco JC, Cortés Y, García E, Palacios V, Rio-Maior H, et al. 2018. Cryptic population structure reveals low dispersal in Iberian wolves. *Sci Rep* **8**: 14108.

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

Smith J, Coop G, Stephens M, Novembre J. 2018. Estimating Time to the Common Ancestor for a Beneficial Allele. *Mol Biol Evol* **35**: 1003–1017.

Smith J, Kronforst MR. 2013. Do Heliconius butterfly species exchange mimicry alleles? *Biol Lett* **9**: 20130503.

Stephens M, Donnelly P. 2003. A Comparison of Bayesian Methods for Haplotype Reconstruction from Population Genotype Data. *Am J Hum Genet* **73**: 1162–1169.

Stronen AV, Jędrzejewska B, Pertoldi C, Demontis D, Randi E, Niedziałkowska M, Borowik T, Sidorovich VE, Kusak J, Kojola I, et al. 2015. Genome-wide analyses suggest parallel selection for universal traits may eclipse local environmental selection in a highly mobile carnivore. *Ecol Evol* **5**: 4410–4425.

Strupp M, Maul S, Konte B, Hartmann AM, Giegling I, Wollenteit S, Feil K, Rujescu D. 2020.

A Variation in FGF14 Is Associated with Downbeat Nystagmus in a Genome-Wide Association Study. *Cerebellum* **19**: 348–357.

Suarez-Gonzalez A, Lexer C, Cronk QCB. 2018. Adaptive introgression: a plant perspective. *Biol Lett* **14**: 20170688.

Sukumaran J, Holder MT. 2010. DendroPy: a Python library for phylogenetic computing. *Bioinformatics* **26**: 1569–1571.

Tang H, Choudhry S, Mei R, Morgan M, Rodriguez-Cintron W, Burchard EG, Risch NJ. 2007. Recent Genetic Selection in the Ancestral Admixture of Puerto Ricans. *Am J Hum Genet* **81**: 626–633.

Taylor SA, Larson EL. 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nat Ecol Evol* **3**: 170-177.

Tereso JP, Bettencourt AMS, Ramil-Rego P, Teira-Brión A, López-Dóriga I, Lima A, Almeida R. 2016. Agriculture in NW Iberia during the Bronze Age: A review of archaeobotanical data. *J Archaeol Sci Reports* **10**: 44–58.

Torres RT, Ferreira E, Rocha RG, Fonseca C. 2017. Hybridization between wolf and domestic dog: First evidence from an endangered population in central Portugal. *Mamm Biol* **86**: 70–74.

Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Rosengren Pielberg G, Sigurdsson S, Fall T, Seppälä EH, Hansen MST, Lawley CT, et al. 2011. Identification of Genomic Regions Associated with Phenotypic Variation between Dog Breeds using Selection Mapping. *PLoS Genet* **7**: e1002316.

Vilà C, Wayne RK. 1999. Hybridization between wolves and dogs. *Conserv Biol* **13:** 195–198.

vonHoldt BM, Kays R, PoWayne RK. 2016. Admixture mapping identifies introgressed genomic regions in North American canids. *Mol Ecol* **25**: 2443-2453.

Wang MS, Wang S, Li Y, Jhala Y, Thakur M, Otecko NO, Si JF, Chen HM, Shapiro B, Nielsen R, et al. 2020. Ancient hybridization with an unknown population facilitated high-altitude adaptation of canids. *Mol Biol Evol* **37**: 2616–2629.

Ward BJ, van Oosterhout C. 2016. hybridcheck: software for the rapid detection, visualization and dating of recombinant regions in genome sequence data. *Mol Ecol Resour* **16**: 534–539.

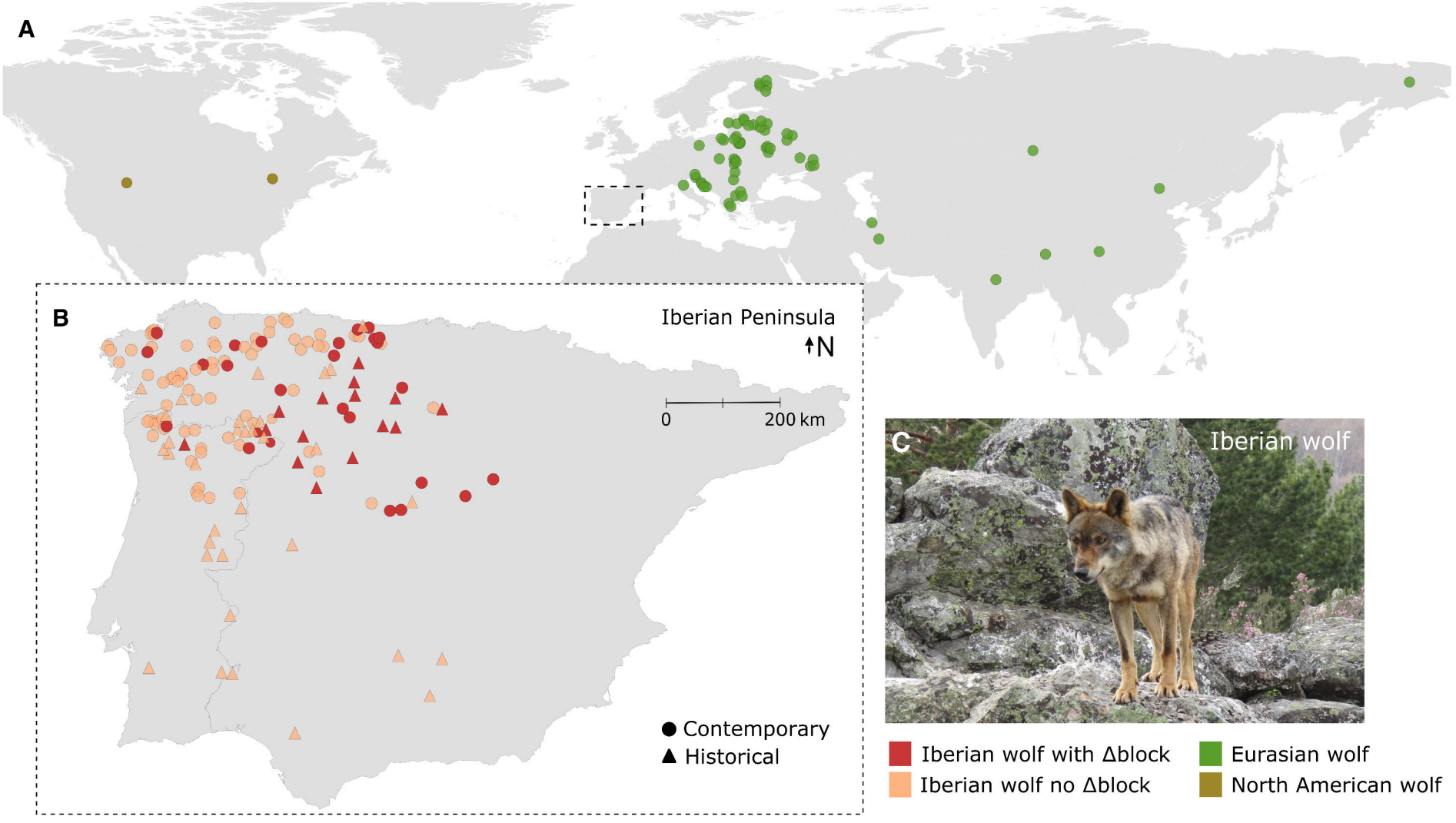
Wilkins AS, Wrangham RW, Fitch WT. 2014. The “Domestication Syndrome” in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics* **197**: 795–808.

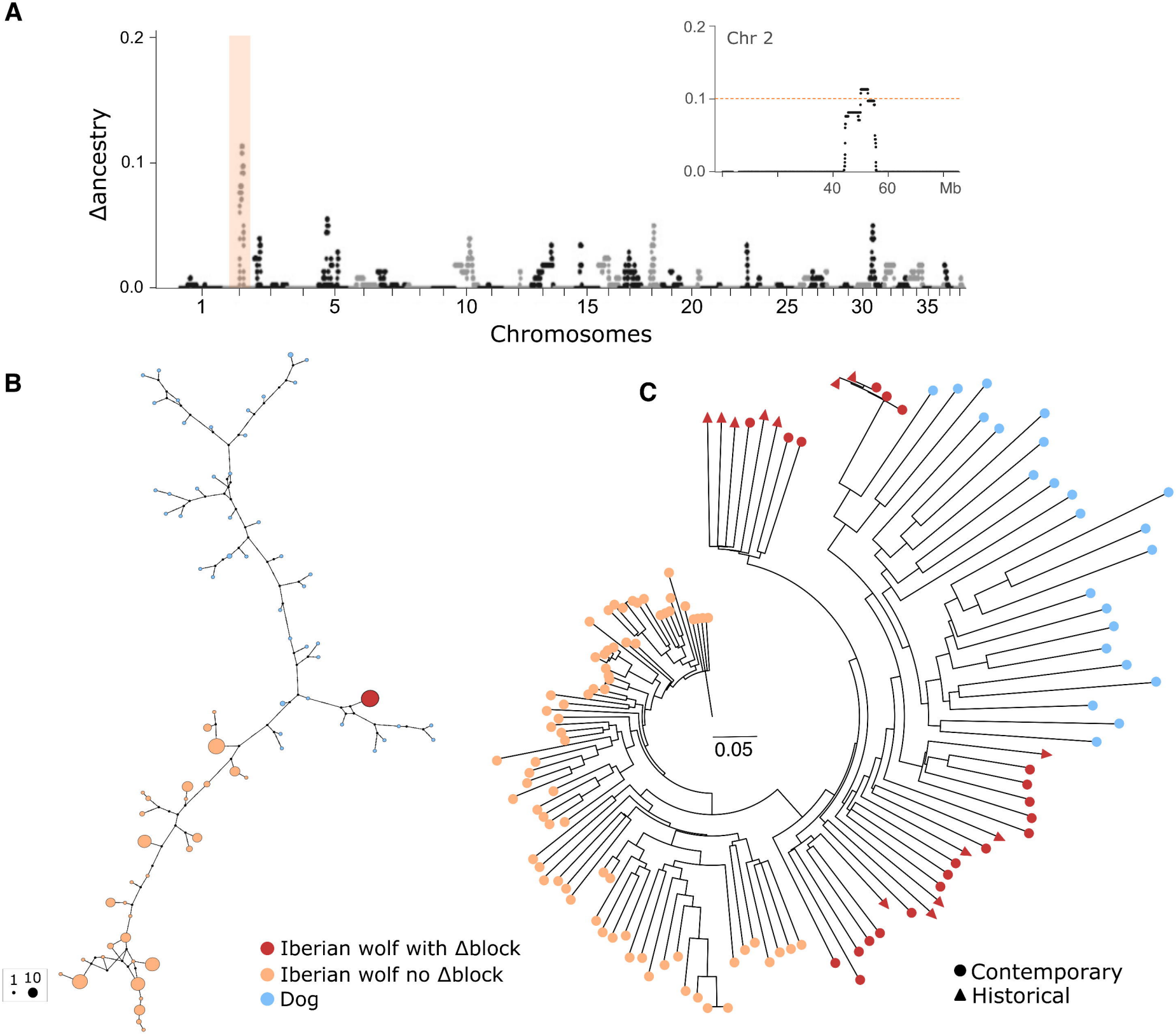
Wong AK, Ruhe AL, Dumont BL, Robertson KR, Guerrero G, Shull SM, Ziegle JS, Millon L

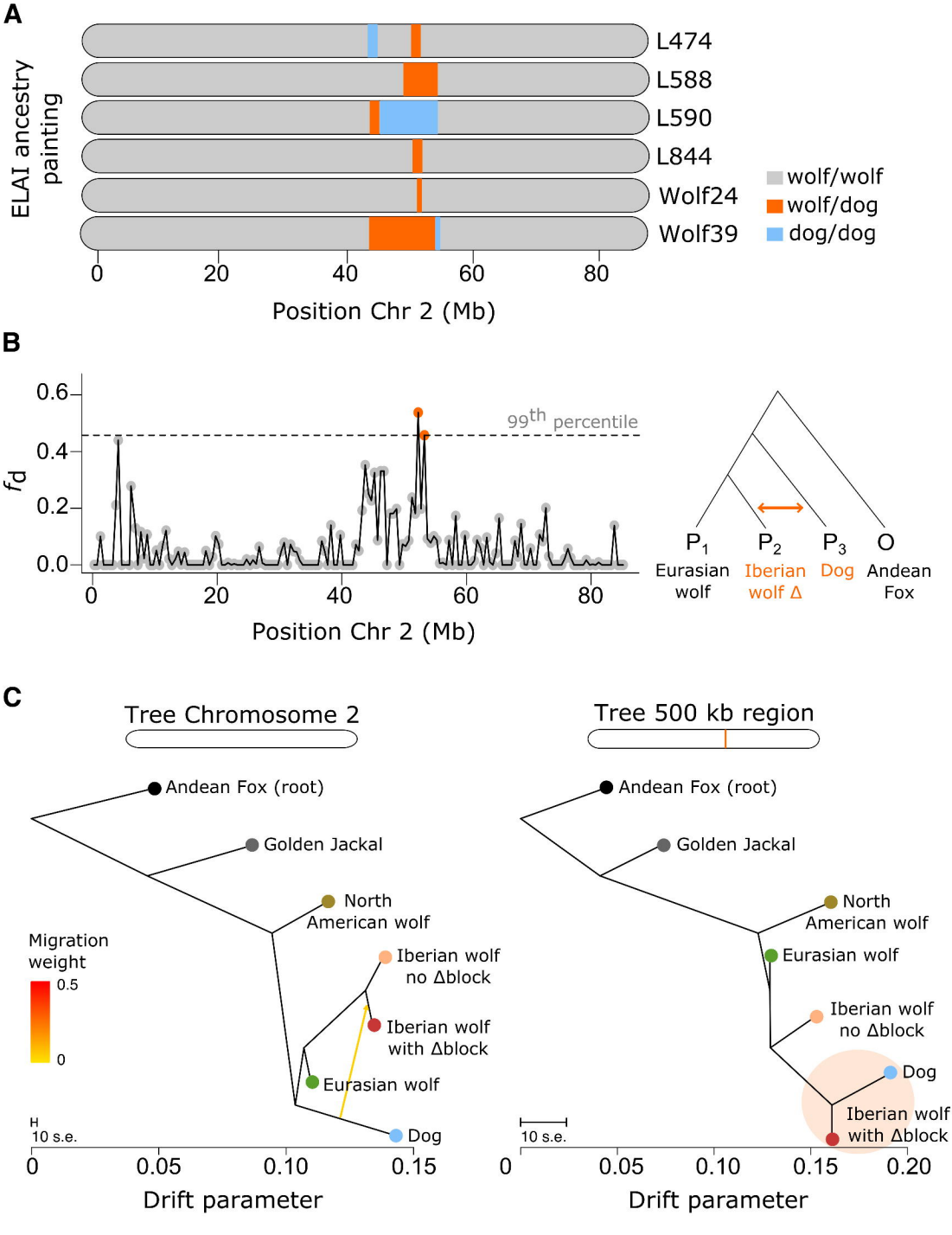
V., Broman KW, Payseur BA, et al. 2010. A Comprehensive Linkage Map of the Dog Genome. *Genetics* **184**: 595–605.

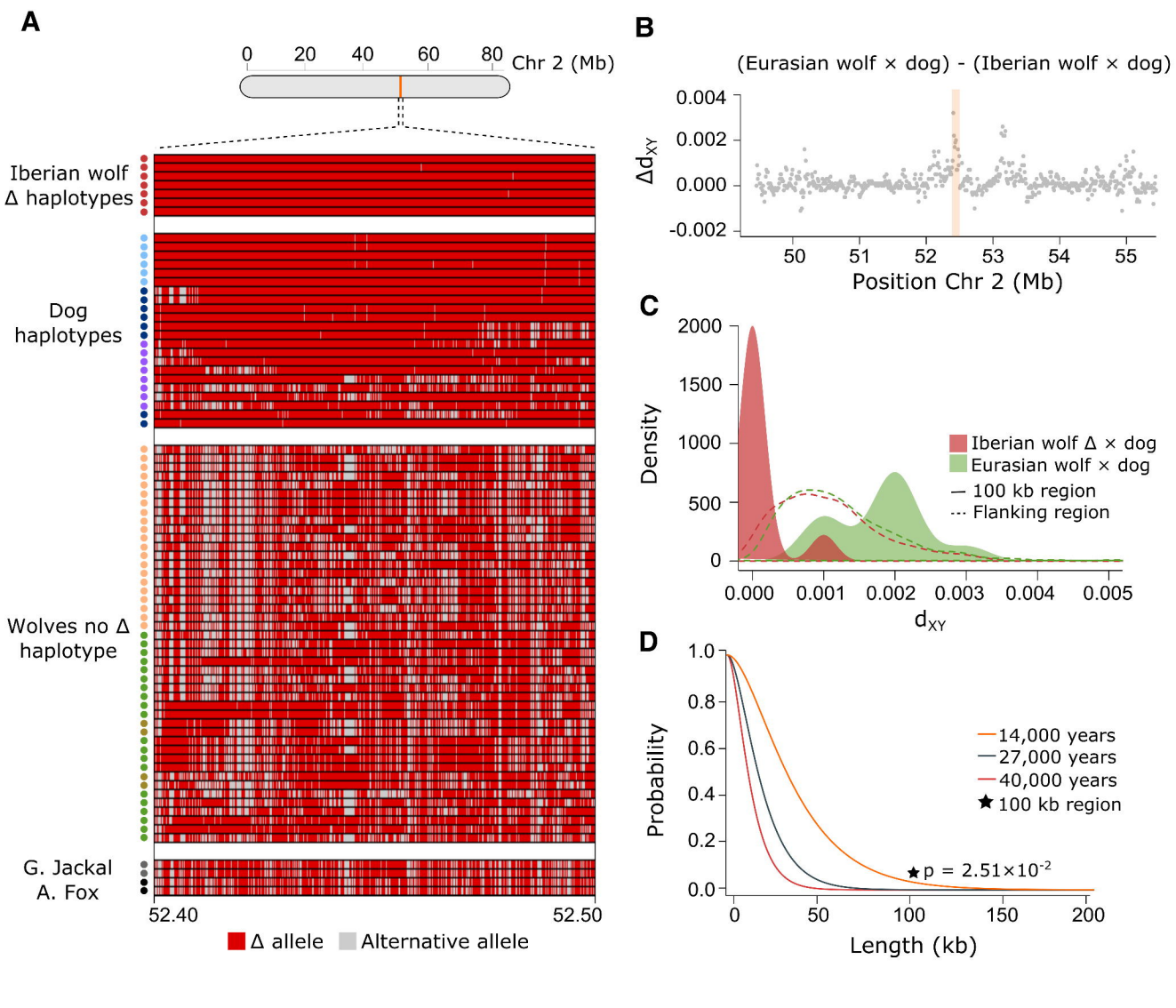
Zhang X, Xiao N, Cao Y, Peng Y, Lian A, Chen Y, Wang P, Gu W, Xiao B, Yu J, et al. 2023. *De novo* variants in *MAST4* related to neurodevelopmental disorders with developmental delay and infantile spasms: Genotype-phenotype association. *Front Mol Neurosci* **16**:1097553.

Zheng Y, Janke A. 2018. Gene flow analysis method, the D-statistic, is robust in a wide parameter space. *BMC Bioinformatics* **19**: 10.


A
B
Iberian Peninsula
N
0
200 km
Contemporary
Historical
C
Iberian wolf
Iberian wolf with Δblock
Iberian wolf no Δblock
Eurasian wolf
North American wolf


A
Δancestry
0.2
0.1
0.0
1
5
10
15
20
25
30
35
Chromosomes
Chr 2
0.2
0.1
0.0
40
60
Mb
B
1 10
Iberian wolf with Δblock
Iberian wolf no Δblock
Dog
C
0.05
Contemporary
Historical


A
ELAI ancestry painting
L474
L588
L590
L844
Wolf24
Wolf39
wolf/wolf
wolf/dog
dog/dog
0
20
40
60
80
Position Chr 2 (Mb)
B
f_d
0.6
0.4
0.2
0.0
99th percentile
0
20
40
60
80
Position Chr 2 (Mb)
P1
Eurasian wolf
P2
Iberian wolf Δ
P3
Dog
O
Andean Fox
C
Tree Chromosome 2
Andean Fox (root)
Golden Jackal
North American wolf
Iberian wolf no Δblock
Iberian wolf with Δblock
Eurasian wolf
Dog
Migration weight
0.5
0
H
10 s.e.
0
0.05
0.10
0.15
Drift parameter
Tree 500 kb region
Andean Fox (root)
Golden Jackal
North American wolf
Eurasian wolf
Iberian wolf no Δblock
Dog
Iberian wolf with Δblock
10 s.e.
0
0.05
0.10
0.15
0.20
Drift parameter


A
0
20
40
60
80
Chr 2 (Mb)
Iberian wolf
Δ haplotypes
Dog
haplotypes
Wolves no Δ
haplotype
G. Jackal
A. Fox
52.40
52.50
Δ allele
Alternative allele
B
(Eurasian wolf × dog) - (Iberian wolf × dog)
0.004
0.002
0.000
-0.002
ΔdXY
50
51
52
53
54
55
Position Chr 2 (Mb)
C
2000
1500
1000
500
0
Density
Iberian wolf Δ × dog
Eurasian wolf × dog
100 kb region
Flanking region
0.000
0.001
0.002
0.003
0.004
0.005
dXY
D
1.0
0.8
0.6
0.4
0.2
0.0
Probability
14,000 years
27,000 years
40,000 years
100 kb region
p = 2.51×10-2
0
50
100
150
200
Length (kb)

Ancient dog introgression into the Iberian wolf genome may have facilitated adaptation to human-dominated landscapes

Diana Lobo, Hernan E Morales, Cock Van Oosterhout, et al.

Genome Res. published online February 14, 2025
Access the most recent version at doi:10.1101/gr.279093.124