

# Genomes of the extinct Sicilian wolf reveal a complex history of isolation and admixture with ancient dogs

## Authors:

Marta Maria Ciucani<sup>1#</sup>, Jazmín Ramos-Madriral<sup>1,2</sup>, Germán Hernández-Alonso<sup>1,2</sup>, Alberto Carmagnini<sup>3</sup>, Sabhrina Gita Aninta<sup>3</sup>, Camilla Hjorth Scharff-Olsen<sup>1</sup>, Liam Thomas Lanigan<sup>1</sup>, Ilaria Fracasso<sup>4</sup>, Cecilie G. Clausen<sup>1,2</sup>, Jouni Aspi<sup>5</sup>, Ilpo Kojola<sup>6</sup>, Laima Baltrūnaitė<sup>7</sup>, Linas Balčiauskas<sup>7</sup>, Jane Moore<sup>8</sup>, Mikael Åkesson<sup>9</sup>, Urmas Saarma<sup>10</sup>, Maris Hindrikson<sup>10</sup>, Pavel Hulva<sup>11</sup>, Barbora Černá Bolříková<sup>12</sup>, Carsten Nowak<sup>13</sup>, Raquel Godinho<sup>14,15</sup>, Steve Smith<sup>16</sup>, Ladislav Paule<sup>17</sup>, Sabina Nowak<sup>18</sup>, Robert W. Mysłajek<sup>18</sup>, Sabrina Lo Brutto<sup>19,20</sup>, Paolo Ciucci<sup>21</sup>, Luigi Boitani<sup>21</sup>, Cristiano Vernesi<sup>4</sup>, Hans K. Stenøien<sup>22</sup>, Oliver Smith<sup>1</sup>, Laurent Frantz<sup>3,23</sup>, Lorenzo Rossi<sup>24</sup>, Francesco Maria Angelici<sup>25,26</sup>, Elisabetta Cilli<sup>27</sup>, Mikkel-Holger S. Sinding<sup>1,28</sup>, M. Thomas P. Gilbert<sup>1,2,29</sup>, Shyam Gopalakrishnan<sup>1,2,30#</sup>

## Affiliations:

- 1 Section for Evolutionary Genomics, The GLOBE Institute, University of Copenhagen, Copenhagen, Denmark
- 2 Center for Evolutionary Hologenomics, The GLOBE Institute, University of Copenhagen, Copenhagen, Denmark
- 3 School of Biological and Chemical Sciences, Queen Mary University of London, London, UK
- 4 Forest Ecology Unit, Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige (TN), Italy
- 5 Ecology and Genetics Research Unit, University of Oulu, Finland
- 6 Natural Resources Institute Finland, Rovaniemi, Finland
- 7 Nature Research Centre, Vilnius, Lithuania
- 8 Società Amatori Ciriaco dell'Etna, Modica (RG) Italy.
- 9 Swedish University of Agricultural Sciences, Grimsö Wildlife Research Station, Department of Ecology, Riddarhyttan, Sweden
- 10 Department of Zoology, Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia
- 11 Charles University, Department of Zoology, Faculty of Science, Prague 2, Czech Republic
- 12 Faculty of Tropical AgriSciences, Czech University of Life Sciences, Prague, Czech Republic
- 13 Center for Wildlife Genetics, Senckenberg Research Institute and Natural History Museum Frankfurt, Gelnhausen, Germany
- 14 CIBIO/InBIO, University of Porto, Vairão, Portugal
- 15 BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Vairão, Portugal
- 16 Konrad Lorenz Institute of Ethology, University of Veterinary Medicine, Vienna, Austria
- 17 Faculty of Forestry, Technical University, Zvolen, Slovakia
- 18 Department of Ecology, Institute of Functional Biology and Ecology, Faculty of Biology, University of Warsaw, Biological and Chemical Research Centre, Warszawa, Poland
- 19 Department of Biological, Chemical, and Pharmaceutical Sciences and Technology (STEBICEF), University of Palermo, Palermo, Italy
- 20 Museum of Zoology "P. Doderlein", SIMUA, University of Palermo, Palermo, Italy
- 21 Università di Roma La Sapienza, Dept. Biology and Biotechnologies "Charles Darwin", Roma, Italy
- 22 NTNU University Museum, Norwegian University of Science and Technology, Trondheim, Norway
- 23 Palaeogenomics Group, Department of Veterinary Sciences, Ludwig Maximilian University, Munich, Germany
- 24 Museo dell'Ecologia di Cesena, Cesena, Italy
- 25 FIZV, Via Marco Aurelio 2, Roma, Italy
- 26 National Center for Wildlife, Al Imam Faisal Ibn Turki Ibn Abdullah, Ulaishah, Saudi Arabia
- 27 Laboratory of Ancient DNA, Department of Cultural Heritage (DBC), University of Bologna, Italy
- 28 Department of Biology, University of Copenhagen, Denmark
- 29 University Museum, Norwegian University of Science and Technology, Trondheim, Norway
- 30 Bioinformatics, Department of Health Technology, Technical University of Denmark, Denmark

# Corresponding authors:

[ciucani@palaeome.org](mailto:ciucani@palaeome.org)

[shyam.gopalakrishnan@sund.ku.dk](mailto:shyam.gopalakrishnan@sund.ku.dk)

**Keywords:** *Canis lupus*, Sicilian wolf, historical DNA, island biodiversity, ancient dogs, hybridization, inbreeding

## Summary

The Sicilian wolf represented the only population of wolves living on a Mediterranean island until the first half of the twentieth century (1930s-1960s) <sup>1-7</sup>. Previous studies hypothesised that they remained isolated from mainland wolves from the end of the Last Glacial Maximum (LGM) <sup>8,9</sup>, until human persecutions led them to extinction <sup>1-7</sup>.

There are only seven known Sicilian wolf specimens from the 19th and 20th century preserved in museums in Italy and recent morphometric analyses assigned them to the new subspecies *Canis lupus cristaldii* <sup>10</sup>. To better understand the origins of the Sicilian wolf, and its relationship to other wolf populations, we sequenced four whole genomes (3.8×-11.6×) and five mitogenomes. We investigated the relationship between Sicilian wolves and other modern breeds to identify potential admixture. Furthermore, considering that the last land-bridge between Sicily and Italy disappeared after the LGM <sup>11</sup>, around 17 kya, we explored the possibility that the Sicilian wolf retained ancestry from ancient wolf and dog lineages. Additionally, we explored whether the long-term isolation might have affected the genomic diversity, inbreeding levels and genetic load of the Sicilian wolf.

Our findings show that the Sicilian wolves shared most ancestry with the modern Italian wolf population but are better modelled as admixed with European dog breeds, and shared traces of Eneolithic and Bronze age European dogs. We also find signatures of severe inbreeding and low genomic diversity at population and individual levels due to long-term isolation and drift, suggesting also low effective population size.

## Results and Discussion

We sampled the seven available museum specimens of Sicilian wolf (Sicilian wolf information presented in Table S1), and successfully recovered four nuclear genomes and five mitogenomes. Poor DNA preservation hindered recovery of whole genome sequences in three samples and mitochondria of two. Raw reads were mapped to the dog reference genome (CanFam 3.1) <sup>12</sup>, yielding average genomic depth of coverage ranging between 3.8× and 11.6× while the coverage on the mitochondrial DNA spanned between 19.7× and 1239.2× (Table S1). The analyses of the mapped reads and the damage pattern confirmed the authenticity of the historical DNA sequencing (Figure S1).

For comparative purposes we resequenced 33 modern wolf genomes (3.66×-41.9×) from across Europe. Further, we resequenced three Cirneco dell'Etna dogs (2×-2.6×), an old Sicilian hunting breed, that could have been in contact with Sicilian wolves. All modern samples were also mapped to the dog reference genome. To avoid biases introduced by ancient DNA damage, we limited all variant based analyses to transversion polymorphisms.

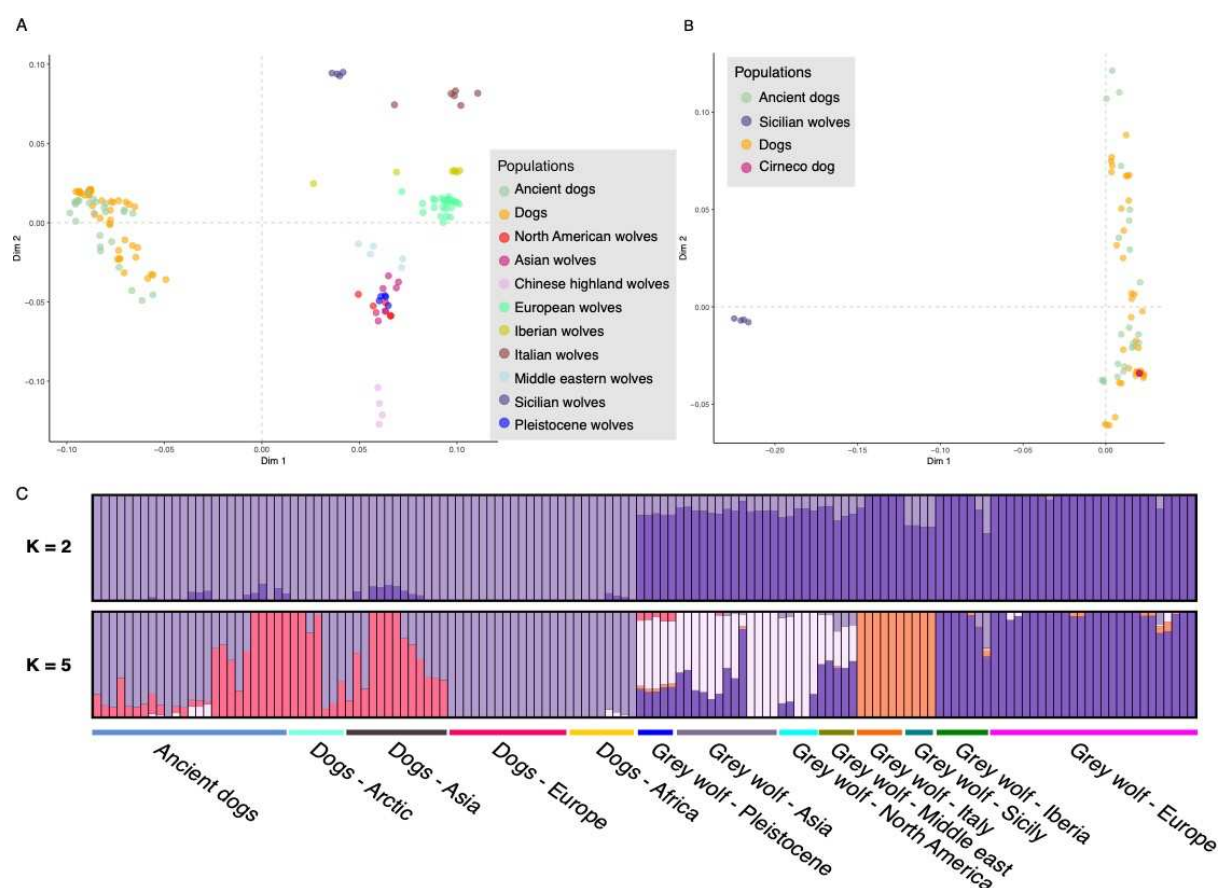
We determined the genetic sex of four Sicilian wolves, by comparing the average genomic coverage with the coverage of the X chromosome. The samples Sic1, Sic2 and Sic3 were identified as males (XY) while Sic7 was shown to be female (XX) (Table S1).

## Population structure and admixture of the Sicilian wolf

We first investigated the placement of the Sicilian wolves among the global dog and wolf diversity by computing a multidimensional scaling (MDS) analysis. We included previously described

genomic groups of present-day wolves (Eurasian and North American), present-day and ancient dogs, and five ancient wolves that lived in Siberia during the Pleistocene<sup>13–15</sup>. The first dimension separated dogs from wolves while the second dimension split European wolves from the rest of the Eurasian, North American and Pleistocene wolves. The results reveal a pattern of genetic isolation and drift of the Sicilian wolves analysed in this study (Figure 1A) which are placed in the upper right quadrant of the plot, within the European wolf diversity, in proximity to the Italian wolves, but shifted towards the cluster of dogs (Figure 1A and B). When restricting the MDS analysis only on wolves, the Sicilian wolf is represented even more differentiated and drifted compared to the Italian wolves and the rest of the European diversity (Figure S2A).

We then performed an admixture analysis on the genomic data to determine the ancestry of the Sicilian wolf population using ADMIXTURE<sup>16</sup> (Figure 1C). When considering only the first two ancestral components (K=2) the Sicilian wolves are modelled as a mixture of wolf and dog components. Increasing the ancestral components up to five (K=5), the Sicilian wolves are represented by the same ancestral component as the Italian wolves. However, when considering seven ancestral components (K=7), there was further separation of the Sicilian wolves which were assigned to a different and unique component (blue) that remains consistent up to K=10 (Figure S2 B). These results confirmed that the genomic differentiation of the Sicilian wolves, supported also by the MDS plots, is consistent with the long-term geographic isolation and the extreme genetic drift that this population suffered in the past.



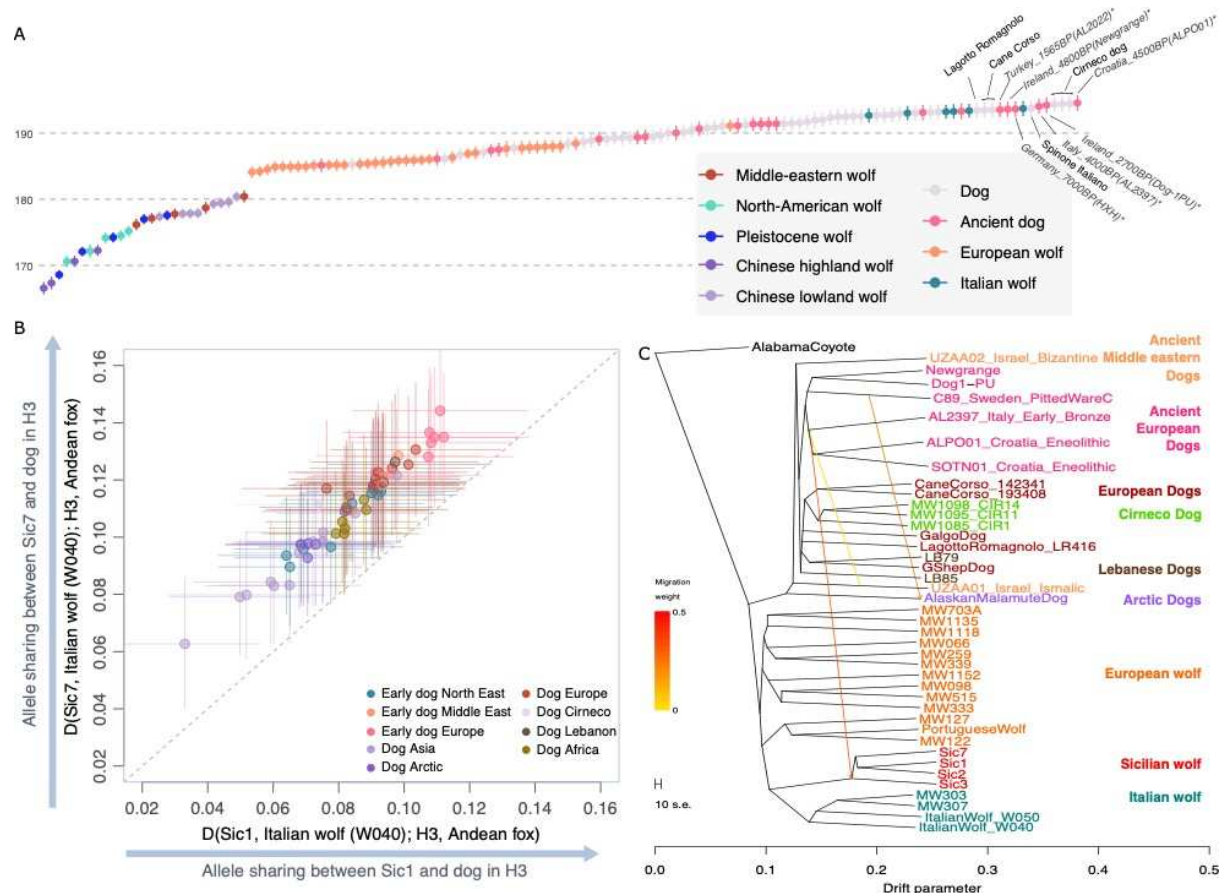
**Figure 1: Population structure of the Sicilian wolves retrieved from genomic data.**

A) Multidimensional scaling plot presenting the Sicilian wolves in the ancient and modern wolves and dogs diversity. B) Multidimensional scaling plot with the Sicilian wolves among the diversity of modern and ancient dogs. C) Admixture plots showing two and five ancestral components (Ks).

We investigated maternal lineage of the Sicilian wolves generating a mitochondrial phylogeny using RAxML-ng<sup>17</sup> including all the wolves and dogs in our dataset and using the coyote as the outgroup. The Maximum-Likelihood mitochondrial phylogeny is consistent with previous publications<sup>8,9</sup> and four out of the five Sicilian wolves (Sic1, Sic2, Sic3 and Sic7) cluster together in the phylogeny as a sister clade to wolves from Romania and Slovakia (Figure S2 D). They also fall within the same clade as wolves from Italy, Bulgaria, and Poland. The last sample, Sic4, falls within the dog mitochondrial diversity, as shown in a previous work based on the d-loop region<sup>8</sup>. Unfortunately, we did not recover sufficient nuclear data from Sic4 to further investigate its ancestry. However, the sample was labelled as *Canis lupus*, but morphological analysis identified dog features, suggesting a hybrid origin. Our mitochondrial phylogenetic analysis placing Sic4 within dogs, is in agreement with the hybrid origin or, alternatively, it might suggest a potential mislabelling.

### **The dual wolf-dog ancestry in the Sicilian wolves**

Previous analyses based on the mitochondrial DNA haplogroups<sup>8,9</sup> clustered the Sicilian wolves together with both Northern European/Siberian Pleistocene and South-eastern European modern wolves. To test these previous results on the nuclear genomes and identify which were closest populations to the Sicilian wolves we performed an *Outgroup f3* analysis using Admixtools<sup>18</sup>. Our results show that Northern European and Siberian Pleistocene wolves, together with East Eurasian, American and Middle Eastern wolves, share the least genetic drift with Sicilian wolves (Figure 2A and Figure S3A-D), indicating that they are the most distant populations. Instead, our results show that the Sicilian wolf is more closely related to modern Italian wolves, ancient European and modern Italian dogs (Figure 2A). In contrast, when comparing these results to the *f3*-statistics performed on Italian and Iberian wolves (Figure S3 E-F) we notice that the Sicilian wolf shares a similar or even higher level of drift than what was estimated for other individuals of the Italian and Iberian populations. These observations could be explained by the Sicilian wolf being highly differentiated due to long-term isolation.



**Figure 2: Relationship between Sicilian wolf and other populations/lineages (retrieved from genomic data).**

A) Outgroup  $f_3$ -statistics showing the genetic affinity of the Sicilian wolf (Sic1) with other dogs and wolves. The plot shows the Siberian Pleistocene, North American, Middle Eastern and Asian wolves to be the most distant population from the Sicilian wolves. European wolves show slightly higher  $f_3$  values. Italian wolves, European modern and ancient dogs show the highest  $f_3$  values identifying these populations as the closest to the Sicilian wolves. B)  $D$ -statistic in the form of  $D(\text{Sic1}, \text{Italian wolf (W040)}; \text{H3, Andean fox})$  - x axis - and  $D(\text{Sic7}, \text{Italian wolf (W040)}; \text{H3, Andean fox})$  - y-axes - computed using qpDstats of Admixtools. All tests resulted in significant Z scores ( $|Z| \geq 3.3$ ). C) Treemix output showing 3 migration edges. The first high weight arrow shows migration from southern European ancient dogs into the Sicilian wolf, a second migration edge goes into the Alaskan Malamute from a Swedish (C89) Mesolithic dog. The third migration edge brings a Middle eastern dog component into southern European ancient dogs. TreeMix residuals of the graph with three migration edges are shown in Figure S5.

Together with the peculiar placement of the Sicilian wolves in the nuclear phylogeny (Figure S2 C) and admixture analysis, the  $f_3$  results suggest the Sicilian wolf shares ancestry with both dogs and wolves.

Outgroup  $f_3$ , computed between every pair of Sicilian wolves and dogs indicate an affinity between Italian dogs such as Spinone, Cane Corso and Cirneco with Sicilian wolves (Figure 2A), suggesting shared ancestry. Of all the modern dog breeds included in the analyses, the Cirneco dog is the closest modern dog to all Sicilian wolves (Figure S3 A-D). We hypothesise that these results could be driven by recent dog introgression due to the numerous stray dogs present on the island.

Interestingly, we also found stronger genetic affinity between the Sicilian wolves and ancient Eneolithic and Bronze Age dogs from Europe (~5000–3000 BP - eg. Croatian ALPO01, Italian AL2397, Irish Dog-1PU and Newgrange, see Table S2, Figure 2A) than with modern European dogs. In particular, the Sicilian wolf sample Sic1 shares the most drift with ancient European dogs followed by Italian modern dogs. A similar pattern was observed in the other Sicilian wolves (Sic2, Sic3 and Sic7) (Figure S3 A-D). This genetic affinity with ancient European dogs could be due to: 1) introgression with ancient dogs, or 2) the Sicilian wolf being closer to a now extinct ancient wolf lineage that contributed to ancient European dogs. This last hypothesis could be supported by other studies, based



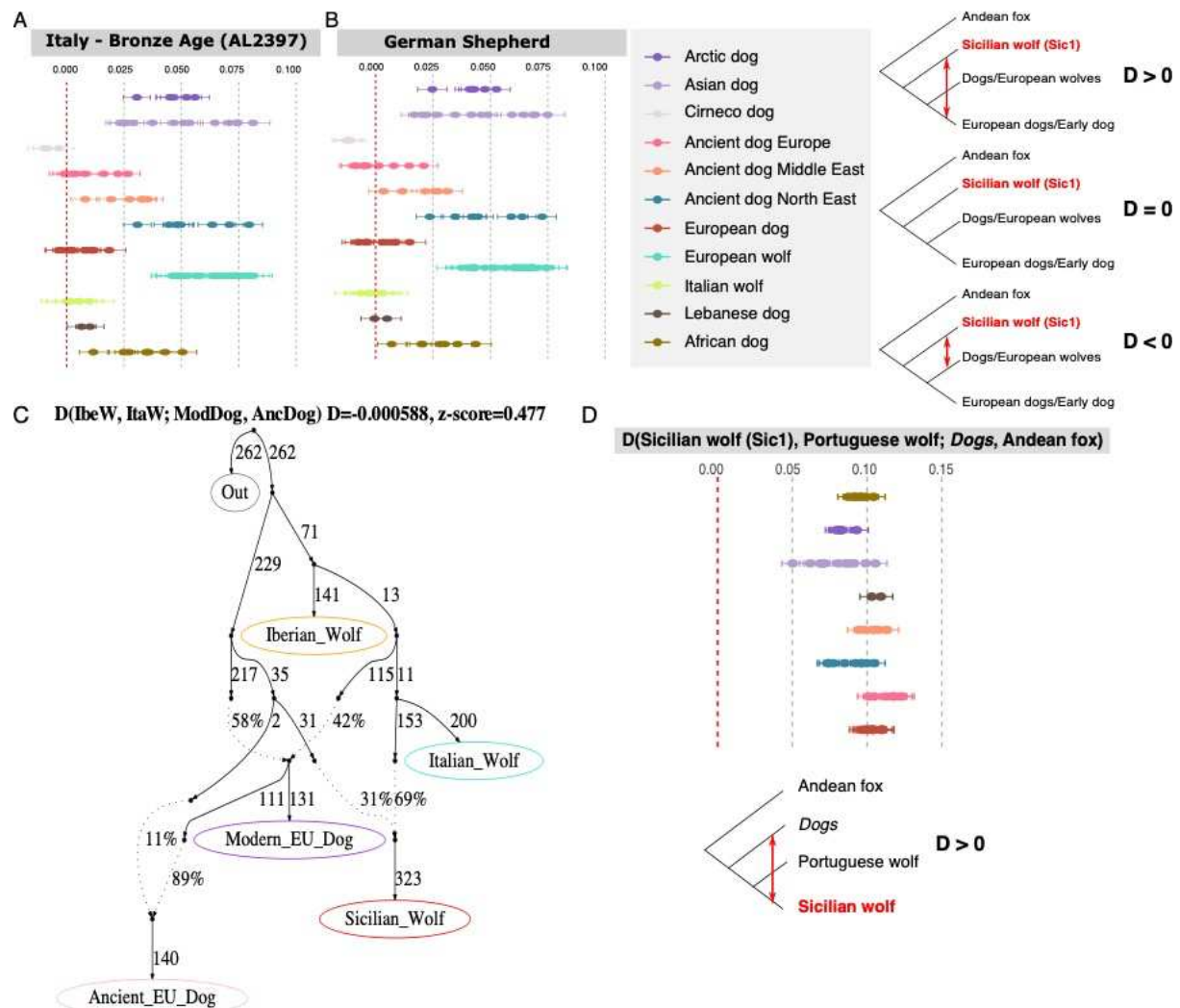
on mitochondrial DNA<sup>19,20</sup>, suggesting a genetic continuity with two Late Pleistocene wolves from Italy and dogs (ancient and modern). Therefore, to test these hypotheses, we computed *D-statistics* using qpDstats from Admixtool<sup>18</sup> to assess allele sharing between the Sicilian wolves, modern and ancient dogs, and wolves. Specifically, we computed all combinations of *D-statistics* of the form  $D(\text{Sicilian wolf, Italian wolf; Dogs, Andean fox})$ . We find the Sicilian wolf shares significantly more alleles with all dogs compared to the modern Italian wolf ( $|Z| \leq 3.33$ ). Furthermore, our results show that the ancient European dogs yield the largest values of *D* (Figure 2B and Figure S4A-B).

The outgroup *f3* and *D-statistics* results led us to focus our analyses on the investigation of the relationship between ancient dogs and Sicilian wolves. To disentangle the historical relationships between Sicilian wolves, modern and ancient dogs, we used TreeMix<sup>21</sup> to build a graph incorporating different numbers of admixture events. The TreeMix graph without migration edges is consistent with the nuclear phylogeny (Figure S2 C) placing the Sicilian wolves as basal to all modern and ancient dogs (Figure S5 A). Once migrations were incorporated into the analyses, the resulting graphs placed the Sicilian wolves as sister clade to the Italian wolves while drawing a high weight (weight = 0.322216) ancestry contribution from Croatian Eneolithic dogs going into the base of their clade (Figure 2C and S5).

Furthermore, since in the *D-statistics* tests  $D(\text{European ancient or modern dogs, Dogs/Wolves; Sicilian wolf, Andean fox})$ , the Sicilian wolf (Sic1) resulted to be equally distant to Italian wolves, ancient and modern European dogs (Figure 3A-B) we proceeded to model the ancestry proportions of the Sicilian wolf, taking into account gene flow from other populations. To do so we used qpGraph<sup>18</sup> (Figure 3C and Figure S4C) in which ancient and modern dogs, Sicilian, Italian and Iberian wolves were included. Consistent with results from the *f3*, *D-statistics* and TreeMix, this analysis indicates that the Sicilian wolf ancestry can be modelled as a mixture of ancient dogs (31%) and Italian wolves (69%).

Lastly, we investigated the directionality of gene flow between Sicilian wolves and dogs by estimating the *D-statistics* in the form  $D(\text{Sicilian wolf, Portuguese wolf, Dogs, Andean fox})$ . We expected all results to be positive if the Sicilian wolf carries dog ancestry. Conversely, if the gene flow goes from Sicilian wolf into specific dog breeds, we would expect only those to yield significant values. Our result consistently shows positive results for all the Sicilian wolves suggesting, therefore, gene flow from dogs into Sicilian wolves (Figure 3D).

Overall, the *D-statistics* and *f3-statistics* results show all four Sicilian wolves to have a similar ancestry pattern, where they carry a substantial dog ancestry component. A previous study on Sierra Morena wolves (*C. l. signatus*) found that they also carried a high dog ancestry proportion, up to a third of their genome<sup>22</sup>. This population shared similar characteristics as the Sicilian wolf, being a small, isolated and declining population in a human-modified landscape. Our results further suggest that these features could contribute to increased hybridization due to the combined effects of stray dogs and dwindling population sizes of wolves. This is not surprising considering that the samples used in this study represent the last individuals of Sicilian wolves and it is likely that the degree of hybridisation with modern dogs was significantly lower in the past. We can see that a high degree of hybridisation with modern dogs is also present in one of the last specimens of Japanese wolves (*C. l. hodophilax*) which became extinct at the same time as the Sicilian wolf<sup>23,24</sup>.



**Figure 3: Relationship between Sicilian wolf and dogs.**

A-B)  $D$ -statistics computed using qpDstats of Admixtools. A)  $D$ -statistic in the form of  $D(\text{Ancient dog (Italy AL2397), Wolves/Dogs; Sic1, Andean fox})$ . B)  $D$ -statistic in the form of  $D(\text{German Shepherd, Wolves/Dogs; Sic1, Andean fox})$ . C) qpGraph admixture graph modelling the Sicilian wolf as a mixture between the Italian wolf (69%) and ancient European dogs (31%). D)  $D$ -statistics investigating the directionality of gene flow between dogs and Sicilian wolves. The assumption behind this test is that if all the  $D$ -statistics of the  $D(\text{Sicilian wolf, Portuguese wolf; Dogs, Andean fox})$  are significant with a  $Z$ -value  $< 3$  the result is BABA meaning gene flow from dogs into Sicilian wolf. On the contrary if some tests were significant but negative resulting in ABBA that could be an indicator that the gene flow could have been from Sicilian wolf into certain dogs leading us to investigate the next  $D$  test.

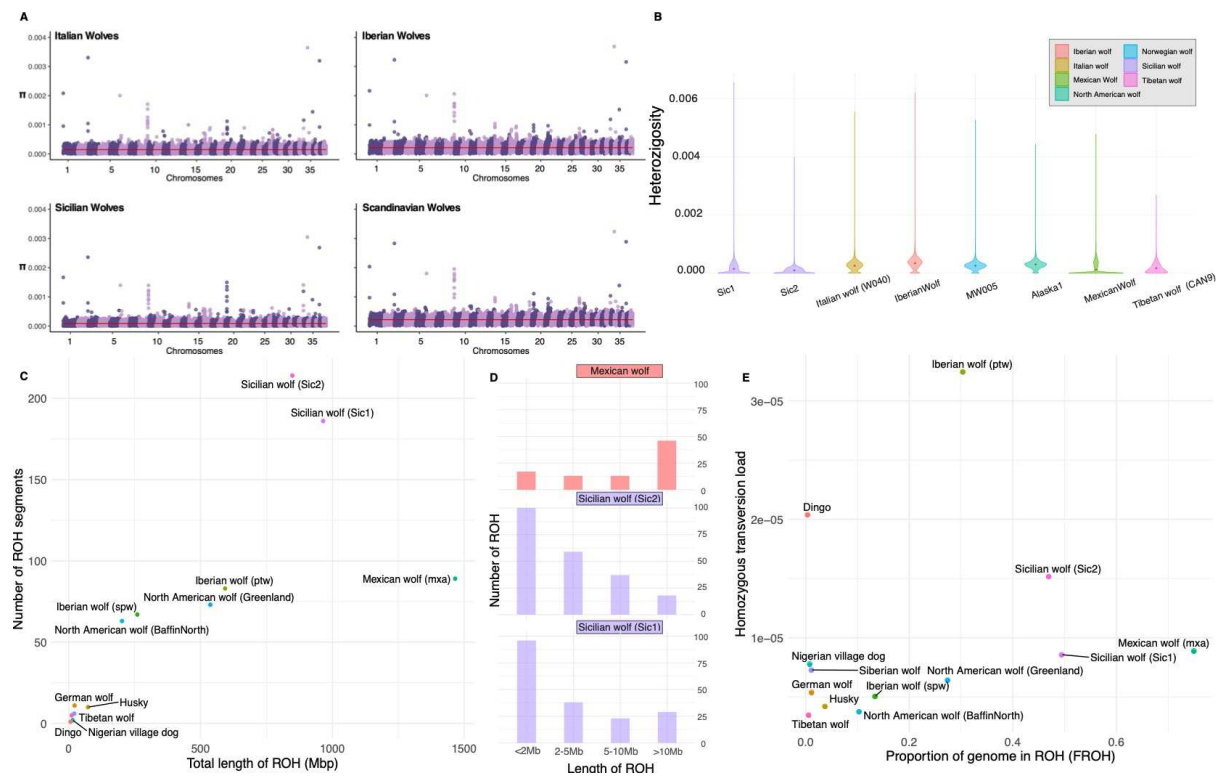
### The effects of island isolation on the Sicilian wolves

We then explored the insular effect and the consequences of the population decline in the last few hundred years on the genomic diversity of the Sicilian wolf population.

Compared to other populations, we find that the Sicilian wolves have a low population level nucleotide diversity ( $\pi$ ) and at the individual level this is confirmed by very low heterozygosity (Figure 4 A-B). These results are expected in populations that are particularly affected by founder effect, long-term isolation and small effective population sizes<sup>25-28</sup>.

Furthermore, we find that the Sicilian wolves have a mean length of runs of homozygosity (ROHs) of  $\sim 5\text{Mb}$ , with a high proportion of their genome in ROHs, *ca* 40% and 35% for Sic1 and Sic2 respectively. While the proportion of the genome with long runs of homozygosity ( $\text{ROH} \geq 2\text{Mb}$ ) is still particularly elevated (*ca* 32%), the number of short and medium size ROH segments is higher than the ones observed in other highly and recently inbred population such as the Mexican wolf (*C. l. baileyi*)

(Figure 4D). The Sicilian wolves also contain a relatively higher mutation load with respect to their inbreeding extent (Figure 4E). Overall, these results indicate that the Sicilian wolf population already at the end of the 19th century was suffering from the effect of small population size leading to mating between closely related individuals. However, short ROH segments could indicate that the inbreeding happened in the past, rather than towards the time of extinction, as a result of perhaps previous persecutions.



**Figure 4: Island isolation effect**

A) Population level nucleotide diversity of Sicilian, Italian, Iberian and Scandinavian wolves. The chromosomes and the nucleotide diversity values are shown on the x and y axis respectively. The red line represents the mean diversity value. B) Heterozygosity in sliding windows (without including repetitive regions) for different wolf samples show reduced levels of heterozygosity in the Sicilian wolves Sic1 and Sic2. C) Individual runs of homozygosity represented as the number of ROH segments on the y-axis and total length of ROH (Mbp) on the x-axis. D) Histogram of the number of small, medium and long ROH segments. The Sicilian wolves (Sic1 and Sic2) show remarkably high numbers of ROH, particularly short ones. E) Genetic load represented as the number of transversion sites that are homozygous is shown on the y-axis compared to the proportion of the genome in ROH (FROH). The Sicilian wolves seem to have a higher transversion load compared to other wolf populations.

To conclude, our study showed that the Sicilian wolf population was strongly affected by long-term insular isolation with particularly low population and individual levels of diversity and a high proportion of the genome in ROH.

The Sicilian wolves also shared common ancestry with modern Italian wolves and dogs. The similarity between Italian and Sicilian wolves, and the evidence for long-term isolation (i.e. long-term inbreeding), raised the question about the timing of colonisation of Sicily by the ancestor of the Sicilian wolf. In the past, due to the climatic changes during the Pleistocene, Sicily was occasionally connected to Italy through a land bridge on the Strait of Messina. This patch of land disappeared, *ca* 17 kya, at the end of the last glacial period, due to increased sea levels, and since then the Strait has always been characterised by strong currents prohibiting the crossing in the past<sup>11</sup>. Therefore, even if wolves have proven to be able to swim<sup>29</sup> it is unlikely that a substantial number of Italian wolves colonised Sicily by sea after the end of the LGM.



Our results, showing the Sicilian wolf closest to the Italian wolves, support that around the time of Sicilian colonisation, the founder population present in the Italian Peninsula was already differentiated from the other European wolf populations (including the Iberian and Balkan wolves)<sup>30</sup>. Additionally, previous studies, based on the whole nuclear genome, suggested that contemporary wolf populations shared a common ancestor between 32 kya and 13 kya<sup>13,14,31</sup>. We suggest that the last common ancestor between the Italian and the Sicilian wolves colonised Sicily through the land bridge between 25 kya and 17 kya<sup>9</sup>.

Furthermore, from our analyses we detected an excess of allele sharing between Sicilian wolves and European dogs, particularly with ancient Eneolithic and Bronze dogs. This indicates that Sicilian wolves sequenced could be the result of admixture, likely limited in the island, with ancient European dogs in the past, and potentially with additional modern dog admixture up to the time of extinction.

This study shows that the Sicilian wolf population survived until the 20th century preserving traces of the ancient dog lineage that was widespread in Europe until 3-4000 years ago<sup>32</sup>. Among all other extant and extinct European wolf populations that have been genetically characterised until now, the Sicilian wolf is unique in showing such a strong signal of genetic affinity with ancient European dogs.

### Acknowledgements

This work was supported by ERC Consolidator Award 681396 Extinction Genomics, DNR143 Center for Evolutionary Hologenomics, and the Norwegian Environment Agency (project 18088069). We would like to thank Luca Sineo for helping collect the Sicilian wolf samples and Domenico Tricomi, together with the owners of the Cirneco dell'Etna dogs for their collaboration in this project. We are grateful to the curators of the museums who allowed us to sample the historical Sicilian wolves: Paolo Agnelli of the Museum of Natural History of the University of Florence, zoology section "La Specola", Fabio Lo Bono of the Civic Museum "Baldassare Romano" in Termini Imerese (PA), Enrico Bellia of the Museum of Zoology "P. Doderlein", SIMUA, Palermo (PA) and Ferdinando Maurici and Fabio Lo Valvo from Regional Museum of Terrasini (PA). We also thank the staff of the Danish National High-throughput Sequencing Centre and BGI Denmark for their support in data generation and the Danish National Supercomputer for Life Sciences at the DTU National Lifescience Center at Technical University of Denmark (DTU) for facilitating the data analysis process. The National Science Centre, Poland supported S. Nowak (grant No. 2020/39/B/NZ9/01829) and R.W. Mysłajek (grant No. 2019/35/O/NZ8/01550). G.H-A. is supported by the Consejo Nacional de Ciencia y Tecnología from Mexico (CONACyT 576734). R.Godinho is supported by the Portuguese Foundation for science and Technology (DL57/2016/CP1440).

## METHOD DETAILS

### Description of the historical Sicilian samples

The sample Sic1 (cat number C11875) was a petrous bone from a wolf skull preserved at the Museum of Natural History, Section of Zoology ‘La Specola’, University of Florence, Florence, Italy. The sample still has the original tag from the 19th century, and it belongs to a wolf killed in 1883 at Vicari (PA). The sample Sic2 (cat number AN/855) is represented by a petrous bone from a wolf skull belonging to a juvenile individual from 1879 preserved at the Museum of Zoology ‘Pietro Doderlein’ of the University of Palermo, Palermo, Italy. The sample Sic3 (cat number M/18) is a mounted specimen of an adult male. This individual was killed in Sicily, presumably around 1870-1880 as documented by an old picture in the museum records. The specimen is preserved in the Museum of Zoology ‘Pietro Doderlein’ of the University of Palermo, Palermo, Italy. The individual Sic4 (cat number 9) is represented by a hide labelled as an adult wolf that was shot in 1924 at Bellalampo (PA), and it is preserved in the Regional Museum of Terrasini (Palermo), Italy. The last sample, Sic7 (cat number NA) is a hide dated around 1880-1920 and is preserved at the Museum of Termini Imerese (Palermo), Italy.

### Data generation for the historical Sicilian samples

The Sicilian wolf samples were processed under strict clean laboratory conditions at the GLOBE Institute, University of Copenhagen. DNA extractions were performed by following the silica-based protocol described in Dabney et al (2013) <sup>33</sup> for the bone samples and Campos & Gilbert <sup>34</sup> in case of keratin samples (hides and claws). Both bone and keratin samples were digested using 1 mL digestion buffer. The extracts were purified using modified PB buffer (Qiagen), washed twice with PE buffer (Qiagen) and eluted twice in 20 µL of buffer EB - with 10 minutes of incubation time at 37°C. The concentration of each extract was checked on a Qubit Fluorometer (ThermoFisher Scientific) in ng/µL and on the Tapestation (Agilent Technologies) for concentration and fragment size.

BGI libraries for the samples Sic1 and Sic4 were constructed following Carøe et al. <sup>35</sup> and Mak et al. <sup>36,37</sup> while BEST Illumina libraries for Sic2, Sic3 and Sic7 were built following Carøe et al. using Illumina adapters <sup>35</sup>. Libraries were prepared using up to 32 µL of DNA in a final reaction volume of 80 µL.

The appropriate number of cycles for the amplification were determined using Mx3005 qPCR (Agilent Technologies) in which 1 µL of SYBRgreen fluorescent dye (Invitrogen, Carlsbad, CA, USA) was loaded in 20 µL indexing reaction volume using also 1 µL of template, 0.2 mM dNTPs (Invitrogen), 0.04 U/µL AmpliTaq Gold DNA polymerase (Applied Biosystems, Foster City, CA, USA), 2.5 mM MgCl<sub>2</sub> (Applied Biosystems), 1X GeneAmp® 10X PCR Buffer II (Applied Biosystems), 0.2 µM forward and reverse primers mixture <sup>36</sup>, and 16.68 µL AccuGene molecular biology water (Lonza, Basel, CH). qPCR cycling conditions were 95°C for 10 minutes, followed by 40 cycles of 95°C for 30 seconds, 60°C for 30 seconds, and 72°C for 45 seconds.

The libraries were amplified using PfuTurbo Cx HotStart DNA Polymerase (Agilent Technologies) and Phusion® High-Fidelity PCR Master Mix with HF buffer (New England Biolabs Inc). The amplification was performed in 50 µL PCR reactions that contained:

- 14 µL of purified library, 0.1 µM of each forward (BGI 2.0) and custom made reverse BGI primers, 2x Phusion® High-Fidelity PCR Master Mix with HF buffer and 8.6 µL AccuGene molecular biology water.
- 10 µL of purified library, forward and reverse primers, 2.5U/L PfuTurbo Cx HotStart DNA Polymerase, 0.4 BSA, 1 µL of dNTPs (25 µM), 1 µL of Buffer 10X and 30.6 µL AccuGene molecular biology water.

PCR cycling conditions for libraries amplified using Phusion® High-Fidelity PCR Master Mix with HF buffer were: initial denaturation at 98°C for 45 seconds followed by 18 to 20 cycles of 98°C

for 20 seconds, 60°C for 30 seconds, and 72°C for 20 seconds, and a final elongation step at 72°C for 5 minutes. Amplified libraries were then purified using 1.5x ratio of SPRI beads to remove adapter dimers and fragments smaller than 90-100bp and eluted in 50 µL of EB (Qiagen) buffer after 10 minutes incubation at 37°C. The samples Sic1 and Sic4 were sequenced using BGISEQ platform while the samples Sic2, Sic3 and Sic7 were pooled together and sequenced on Novaseq6000 Illumina platform, S2 flow-cell PE50.

## **Data generation for modern samples**

### ***Cirneco dell'Etna buccal swab samples - DNA Sample Collection, Storage and Extraction***

SK-1S Isohelix buccal swabs for non-invasive DNA collection were used to sample three Cirneco dell'Etna individuals (MW1085\_CIR1, MW1095\_CIR11, MW1098\_CIR14). The dogs were sampled by their owners at least 30 minutes after eating and to avoid contamination gloves and face masks were used. The swabs were inserted into the dogs' mouth, rubbed on their cheek for *ca* 1 minute and placed back into the collector tube together with the SGC-50 Isohelix Dri-capsule to preserve and stabilize the DNA at room temperature during shipping. Once received in the Modern DNA labs of the GLOBE Institute (University of Copenhagen) the samples were stored at -20°C.

Buccal swab samples were placed in 2 mL Eppendorf tubes and were extracted using a modified version of the DNeasy® Blood and Tissue Kit (Qiagen). In each tube were added 380 µL of ATL Buffer and 20 µL of Proteinase K (Roche). The samples were placed in a thermomixer for 1 hour at 56°C. The lysate was transferred in a different tube and 400 µL of AL Buffer and 400 µL of 96% Ethanol were added. The extraction reaction was then spun down in a DNeasy Mini column and the filter was washed with 500 µL of Buffer AW1 and 500 µL of Buffer AW2. The DNA was eluted twice using 50 µL of AE Buffer directly onto the DNeasy Mini column membrane.

### ***Modern Italian wolves tissue samples - DNA Sample Collection, Storage and Extraction***

The Italian wolf samples were collected around 2010 and 2012 from road-killed individuals populating the Simbruini Mountain Range Regional Park and National Park of Abruzzo in the Central Apennines (See Table S2). The muscle samples were stored in ethanol at -20°C and subsequently processed in the "Modern DNA laboratory" at the Fondazione Edmund Mach (FEM). Small pieces of tissue of around 25 mg were extracted using the DNeasy Blood and Tissue Kit (Qiagen) with an overnight digestion at 56°C. The elution was conducted at the GLOBE Institute (University of Copenhagen) using two washes of 50 µL of AE buffer, with 10 minutes of incubation at 37°C. Until the elution, samples were stored at -20°C inside the DNeasy Mini spin columns.

### ***Modern wolves - DNA Sample Collection, Storage and Extraction***

Modern wolf tissue and blood samples from several locations in Europe (Table S2) were extracted using Thermo Scientific KingFisher instrument and following the manufacturer's protocol. The samples were then checked for concentration with a Qubit Fluorometer (ng/µL). All the extracts - with the exception of the modern Italian wolves and the Cirneco dell'Etna - were sent to BGI Copenhagen for library build and sequenced on 1/8 of a lane each on DNBSEQ PE150.

### ***Library build, amplification and sequencing of modern Italian wolves and Cirneco dogs***

Extracts were fragmented in the Covaris LE220 plus Focused-ultrasonicator with the parameters set for getting 350-bp fragment length. The extracts were diluted to obtain 100 ng concentration and BGI libraries for the Italian wolves and Cirneco dogs were constructed following Carøe et al.<sup>35</sup> and Mak et al.<sup>36</sup> using 10 µM adaptors. Libraries were purified using MinElute columns using PE buffer (Qiagen) and eluted in 60 µL of EB buffer.

The PCR mixture for the Italian wolf libraries consisted on: 20  $\mu$ L of purified library, 0.2  $\mu$ M of forward and reverse BGI primers, 2.5 U/ $\mu$ L PfuTurbo Cx HotStart DNA Polymerase, 10  $\mu$ L of Buffer 10X, 0.08 mg/mL BSA, 0.5 mM of dNTPs (25  $\mu$ M) and 61.2  $\mu$ L AccuGene molecular biology water (Lonza, Basel, CH). The Cirneco dog libraries were amplified using 20  $\mu$ L of purified library, 0.2  $\mu$ M of forward and reverse BGI primers, 0.05 U/ $\mu$ L of AmpliTaq Gold (Thermo Fisher Scientific, USA), 0.4 mg/mL BSA, 0.2 mM of dNTPs (25  $\mu$ M), 10  $\mu$ L of Buffer 10X, 2.5mM of MgCl<sub>2</sub> and 50.2  $\mu$ L AccuGene molecular biology water in a total volume of 100  $\mu$ L reaction.

The amplification of the Italian wolves was performed as follows: initial denaturation at 95°C for 2 minutes followed by 10 to 12 cycles of 95°C for 30 seconds, 60°C for 30 seconds, and 72°C for 110 seconds, and a final elongation step at 72°C for 10 minutes. PCR cycling conditions for Cirneco dog libraries were: initial denaturation at 95°C for 10 minutes followed by 12 cycles of 95°C for 20 seconds, 60°C for 30 seconds, and 72°C for 45 seconds, and a final elongation step at 72°C for 5 minutes. The Italian wolves and Cirneco dogs were sequenced on 1/8 of a lane each on MGISEQ2000 PE150 and DNBSEQ PE150 respectively.

### Data processing and historical DNA authentication

Short sequencing reads from modern, historical, and ancient samples were mapped to the dog reference genome (CanFam3.1)<sup>12</sup> using Paleomix v.1.2.13 pipeline<sup>38</sup>. First, adaptor sequences were removed using AdapterRemoval 2.0<sup>39</sup> with default settings and BWA v0.7.12 *backtrack*<sup>40</sup> was used to perform the alignment of reads to the dog genome (bwa seed was disabled) setting the minimum base mapping quality to 0 to retain all the reads in this step. In further computational steps reads with mapping quality lower than 20 or 30 will be discarded. Picard MarkDuplicates v2.9.1<sup>41</sup> was used to filter out PCR duplicates and GATK v4.1.0<sup>42,43</sup> was used to perform the indel-realignment step.

For the historical Sicilian samples in this study the post-mortem DNA profiles and misincorporation patterns were performed through mapDamage2.0<sup>44</sup>.

### Dataset

The dataset designed for this study is represented by 154 samples (Table S2), including one Andean fox (*Lycalopex culpaeus*)<sup>45</sup>, three golden jackals (*Canis aureus*)<sup>13,14,46</sup>, four coyotes (*Canis latrans*)<sup>46,47</sup>, four African golden wolf (*Canis lupaster*)<sup>46,48,49</sup>, one Ethiopian wolf (*Canis simensis*)<sup>46</sup>, 62 modern wolves (*Canis lupus*) from Eurasia and North America<sup>13,14,22,46,50–53</sup>, 45 modern dogs<sup>12,14,45,51–59</sup> (see Table S2 for more information), 24 ancient dogs (*Canis lupus familiaris*)<sup>32,51,60–63</sup>, five Pleistocene wolves<sup>15,51</sup> and five newly sequenced Sicilian wolves (*Canis lupus cristaldii*). In this dataset 33 modern wolves and three Cirneco dogs were newly sequenced as part of this study. See Table S2 for information regarding sample ID, location, coverage, project number and publication for each sample used.

### Variant calling

For each modern sample at each genomic site, we sampled a random read using ANGSD v0.931<sup>64</sup> (-doHaploCall 1) from the reads with a minimum mapping quality of 30 (-minMapQ 30) and bases with minimum quality of 20 (-minQ 20). The following parameters were used in the command line: -minMinor 1 -maxMis 10 -skipTriallelic 1 -doMajorMinor 2 -C 50 -baq 1 -remove\_bads 1 -only\_proper\_pairs 1. The output file was converted in Plink files (tped and tfam) using haploToPlink tool in ANGSD and used to create a list of variable sites to call in ancient and historical samples and outgroups. Also, in this case ANGSD with -doHaploCall option was used to call the haplotypes for the list of sites (option -s) that were variable across the modern samples. Transitions were discarded (-rmTrans 1) to reduce the aDNA derived error in the historical samples included in our dataset. The

modern, ancient, and historical samples were merged and converted in Plink files. The resulting dataset consisted of 74.5 million SNPs that we were able to merge.

Plink v1.9<sup>65</sup> was used to prune the panel from minimum allele frequency below 5% (-maf 0.05), missing data (N) and filter out all the variants with missing call rates below 25% (-geno 0.25). Variants in linkage disequilibrium (LD) within 10kb window size were also removed (--indep-pairwise 10kb 2 0.5). The remaining panel contained 2.3 million SNPs and it was used (or a subset of it) for the following analyses and named “haploid SNPs panel”.

## MDS

The SNPs panel created with ANGSD in the previous section that consist of 153 samples and a total of 2.3 million transversion sites was used to generate a multidimensional scaling plot (MDS) by estimating first the IBS (pairwise distance) matrix between samples using Plink v1.9<sup>65</sup>. The MDS plots were generated including all the wolf and dog samples, one including all the samples, one with only wolves and another one with only dogs and the Sicilian wolves.

## Admixture

The haploid SNPs panel including wolves and dogs was used as input in the ADMIXTURE<sup>16</sup> analysis to estimate the ancestry components on a subset of samples in the dataset. The outgroups were not included in this analysis. ADMIXTURE was run on 2 to 10 ancestry components (K2 to K10), and, for each K, 50 independent replicates were performed. The replicate with the best likelihood value was chosen for each K. Pong<sup>66</sup> was used to visualise the admixture plots.

## Mitochondrial and nuclear phylogenies

We built a mitochondrial phylogeny using all the samples in the dataset (n = 131) using the coyote as the outgroup. To generate the fasta files we extracted the chrMT from the bam file using Samtools<sup>67</sup>, and built a majority consensus sequence using ANGSD (-dofasta 2). The mitochondrial sequences for each sample were combined and aligned using MUSCLE<sup>68</sup>, and a ML tree was built with RAXML-ng<sup>17</sup> using the evolutionary model GTR+G and 500 bootstrap replicates. The resulting tree was visualised using FigTree v.1.4.4<sup>69</sup>.

To build the nuclear phylogeny on all the modern and ancient individuals in the dataset, we first used ANGSD through which we generated a consensus sequence for each genome using the dog reference genome (CanFam3.1)<sup>12</sup> as reference and sampling each base randomly (-dofasta1). Bases with base quality lower than 20 and reads with mapping quality lower than 20 were discarded (-minQ 20 -minmapq 20). The minimum coverage for each individual was set to 3x (-setminDepthInd 3), and the following additional filters were used: -doCounts 1 -remove\_bads 1 -uniqueOnly 1 -baq 1 -C 50. We then selected only from autosomal chromosomes 1000 random regions, each 5000bp long, from the dog reference genome using BEDTools<sup>70</sup> *random* with the following parameters: -l 5000 -n 1000. We then used Samtools<sup>67</sup> to select from each sample the regions of interest and combined them into a multi-sequence alignment (MSA) fasta file. Each MSA files were used in IQ-TREE<sup>71</sup> v.2.1.2 to reconstruct the phylogeny using 1000 bootstrap replicates (-B 1000) with UFBoot2<sup>72</sup>, 1000 bootstrap replicates for SH-aLRT (-alrt 1000) and *ModelFinder Plus*<sup>73</sup> to identify the best evolutionary model for each region. The resulting 1000 trees were then concatenated in a single file and used as input in Astral-III<sup>74</sup> to estimate the nuclear phylogeny. FigTree<sup>69</sup> v.1.4.4 tool was used to visualise the species tree estimated with Astral-III.

## Outgroup three population test (Outgroup *f3*-statistics)

We calculated *outgroup f3* using qp3pop tool in Admixtools<sup>18</sup> v.5.1 to assess the shared genetic drift between the reference population A and B since the separation from the outgroup Andean fox (A,



*B*; *AndeanFox*). A negative *f3-statistics* value with the correspondence Z-value less than -3 indicates that the population A is close to the population B. The higher the value the closer the 2 populations are to one another.

#### Four-population test (*D-statistics*)

We computed *D-statistics* on the haploid dataset using qpDstats in Admixtools<sup>18</sup> v.5.1 to study how modern and ancient wolves and dogs are related to the Sicilian wolves.

The following forms of *D-statistics* were tested:

1. D(Sicilian wolf, Italian wolf; *Wolf/Dog*, Andean fox) (Figure 2B and Figure S4A);
2. D(ancient European breed, *Wolf/Dog*; Sic1, Andean fox) (Figure 3A);
3. D(modern European dogs, *Wolf/Dog*; Sic1, Andean fox) (Figure 3B);
4. D(*Dogs*, Cirneco dog; Sicilian wolf, Andean fox) (Figure S4B);
5. D(*Dogs*, Croatian Eneolithic dog (ALP01); Sicilian wolf, Andean fox) (Figure S4B);
6. D(Sicilian wolf, Portuguese wolf; *Dogs*, Andean fox) to determine which wolf populations have a closer relationship to dogs (Figure 3D). We assumed that if all d-stats are significant and positive the gene flow could be from dogs into Sicilian wolves. If some tests are not significant it could indicate that the gene flow is directed into certain dogs.

In all the D-statistic tests the outgroup (Andean fox) is placed in H4 therefore if the value of the *D-statistics* is different from 0 it means that there is allele sharing with one population from H1 or H2 to the population in H3. Specifically, if the value is negative with a Z-value smaller than -3 ( $|Z| < -3$ ) the result will be ABBA, implying allele sharing between populations H2 and H3. Instead, if the d value is positive – with a Z-value greater than 3 ( $|Z| > 3$ ) – the results will be the opposite and the allele sharing will be between H1 and H3 (BABA).

#### TreeMix

The software TreeMix<sup>21</sup> v1.13 was used to evaluate the admixture event from dogs into Sicilian wolves to understand, in combination with *f3* and *D-statistics*, the dog ancestral component into wolves. TreeMix was run on a subset of the haploid dataset including the four Sicilian wolves, Italian wolves, Iberian and other European wolves and ancient and modern dogs for a total of 37 samples. We restricted the analysis to sites without missing data resulting in a final dataset of 160,233 transversion sites. The program was run ten times (10 replicates) for each migration event (m) ranging from 0 to 4. The optional parameters -noss, -global and -k 500 were used and the tree was rooted using a coyote (Alabama coyote). The trees with the best likelihood for each migration event were selected using R<sup>75</sup> and plotted using the plotting\_funcs.R provided by Treemix software. The topology of these clades remains overall constant from the 1st up to 4 migration events, and none of the inferred migrations involves the Cirneco dog and the Sicilian wolves.

#### Admixture graphs

To model the relationships between the Sicilian wolf, modern wolves and dogs we estimated an admixture graph using qpGraph<sup>18</sup>. We included samples from relevant groups: Italian wolf (MW303, MW307, ItalianWolf\_W050, ItalianWolf\_W040), Iberian wolf (PortugueseWolf, MW122, MW127), modern dog (BoxerDog, GalgoDog, GShepDog), ancient dog (ALP001\_Croatia\_Eneolithic, SOTN01\_Croatia\_Eneolithic, AL2397\_Italy\_Early), Sicilian wolf (Sic1, Sic7, Sic2) and the Andean fox. First, we identified samples that could be grouped into each of the relevant groups using qpWave<sup>18</sup>. By comparing two lists of populations (left and right populations) qpWave identifies the minimum number of migrations required from the right populations to define the ancestry of the left populations. As ‘left populations’ we tested all possible pairs of Italian wolves, Iberian wolves, modern European

dogs, ancient dogs and Sicilian wolves. For the ‘right populations’ we used all other samples, except the ones included among the ‘left populations’ and using the Andean fox as fixed outgroup. qpWave was run using the ‘allsnps=YES’ option. Pairs of samples that were consistent with a single migration ( $p\text{-value} \geq 0.05$ ) were grouped into a population for the admixture graphs: Italian wolf (MW303, MW307 and ItalianWolf\_W040), Iberian wolf (PortugueseWolf, MW122, MW127), ancient dog (ALPO01\_Croatia\_Eneolithic and AL2397\_Italy\_Early), modern European dog (GalgoDog and GShepDog) and Sicilian wolf (Sic1 and Sic7).

To efficiently explore the possible admixture graph models we used two approaches: qpBrute<sup>63,76</sup> and a ‘base graph’ approach as described in Ramos-Madrigal<sup>15</sup>. For the qpBrute approach we created a configuration file for qpBrute and performed a heuristic search of the admixture graph space. For the ‘base graph’ approach we use *admixturegraph* R library<sup>77</sup> to list all possible graphs including the Italian wolf, Iberian wolf, ancient dog, and modern European dog, and fixed the Andean fox as an outgroup. We tested all possible graphs including 0, 1 and 2 migration edges. We then included the Sicilian wolf to the two graphs that fitted our data in terms of the Z-score of the worst fitting *F*-statistic and with a score that was not significantly different<sup>78</sup>. The Sicilian wolf was included both as a non-admixed lineage and as an admixed lineage. We show the best graph from each approach from the graphs that fitted our data (Figures 3C and S4C).

### Nucleotide diversity and Heterozygosity in sliding windows

The genome-wide nucleotide diversity ( $\pi$ ) was estimated for four wolf populations using 4 Sicilian wolves, 5 Italian wolves, 4 Scandinavian wolves and 6 Iberian wolves. VcfTools 0.1.16<sup>79</sup> was run on the haploid SNPs panel with a window of 100kbp. Rstudio<sup>80,81</sup> and ggplot<sup>82</sup> were used to visualise the data for each population.

We used ANGSD to estimate the heterozygosity of each sample, by calculating the folded site frequency spectrum (SFS) on the autosomal chromosomes. A saf.idx file based on individual genotype likelihoods using GATK (-GL 2) was generated for each bam file (doSaf 1 -fold 1) and the transitions (rmtrans 1) and reads with quality score and bases with mapping quality lower than 20 (-minQ 20 -minmapq 20) were excluded. The dog reference genome was used both as reference and as ancestral (-ref and -anc options) and the repeat regions were masked using a repeat mask file (-sites). The chromosomes were partitioned into overlapping windows of size 1 Mb with a step size of 500 kb using the BEDTools<sup>70</sup> windows tool. Windows shorter than 1Mb at the end of the chromosomes were discarded. The SFS for each window was estimated using the realSFS utility tool provided in ANGSD and subsequently the final heterozygosity per window was calculated as the ratio of heterozygous sites/total sites. Rstudio, ggplot2 and dplyr<sup>83</sup> were used to visualise the heterozygosity level at each window in the form of a violin plot for each individual.

### ROHan

To assess the extent of recent inbreeding, we estimated the length and abundance of segments of runs of homozygosity (ROH) using ROHan on two genomes of Sicilian wolf (Sic1 and Sic2) with minimum 5x following the recommended coverage for ancient genomes<sup>84</sup>. Before running ROHan, we created a bed file containing mappable regions in the reference dog genome (CanFam3.1) to account for divergence between the genomes of the wolf and the dog. Then, a file listing only the autosomal chromosomes was created to limit our analysis to the autosomal regions. Afterwards, to get the number of segregating sites expected in the ROH segments (--rohmu) required by ROHan, we estimated background heterozygosity from X chromosomes of male canids which found an average of  $4 \times 10^{-5}$  segregating sites. We also observed the deamination profile of the ancient genomes using bam2prof (included in ROHan) and determined that for most of the ancient bam files the damage levels were off

at -q 20. The bam2prof results were then used on the ROHan call for the Sicilian wolf genomes (--deam5p and --deam3p).

We calculated the number of segments (NROH), mean length (LROH), and total sum of runs of homozygosity (SROH) in the autosomal region of each sample from the .hmmrohl output of ROHan using an R script. The genomic coefficient of inbreeding was calculated in the same script by dividing SROH with the total number of validated sites by ROHan available on the .hEst file. The presented plot only includes the middle mid-point estimates from the three mid-point estimates provided by ROHan (min, mid, and max).

## Transversion load

To assess the abundance of deleterious mutations in the wolves genome, we used the conservation scores from SIFT, following the pipeline of Greer (2020)<sup>85</sup>. More specifically, we obtained the conservation scores from VEP Ensembl, selected several taxa to obtain ancestral allele of positions, and used this along with the genotype probability values from each sample to calculate mutation load. We consider mutation load as the sum of conservation scores where there are homozygous transversion weighted by the genotype probability of that site divided by the sum of genotype probability of homozygous transversion and total the whole k positions in the genome:

$$L_C = \frac{\sum_{i=1}^k P_i(G_{hom_{tv}}) \cdot C_i}{\sum_{i=1}^k P_i(G_{hom_{tv}}) + P_i(G_{hom_{anc}})}$$

As the SIFT score for deleterious allele tend to approach 0 and the score is per allele, we calculated load by selecting two kinds of homozygous derived genotypes that are transversions from ancestral alleles, marked by  $G_{hom_{tv1}}$  and  $G_{hom_{tv2}}$  with their respective SIFT scores  $S_1$  and  $S_2$ .

$$L_S = \frac{\sum_{i=1}^k (1 - P_i(G_{hom_{tv1}})S_{i,1}) + (1 - P_i(G_{hom_{tv2}})S_{i,2})}{\sum_{i=1}^k P_i(G_{hom_{tv1}}) + P_i(G_{hom_{tv2}}) + P_i(G_{hom_{anc}})}$$

## References

1. Jhala, Y., Boitani, L., and Phillips, M. (2018). IUCN Red List of Threatened Species: *Canis lupus*. IUCN Red List of Threatened Species.
2. David Mech, L., and Boitani, L. (2007). *Wolves: Behavior, Ecology, and Conservation* (University of Chicago Press).
3. Zimen, E., and Boitani, L. (1975). Number and distribution of wolves in Italy. *Z. Säugetierkunde* 40, 102–112.
4. Salvatori, V. (2005). Report on the Conservation Status and Threats for Wolf (*Canis Lupus*) in Europe (Council of Europe).
5. Dufresnes, C., Miquel, C., Remollino, N., Biollaz, F., Salamin, N., Taberlet, P., and Fumagalli, L. (2018). Howling from the past: historical phylogeography and diversity losses in European grey wolves. *Proc. Biol. Sci.* 285.
6. Sastre, N., Vilà, C., Salinas, M., Bologov, V.V., Urios, V., Sánchez, A., Francino, O., and Ramírez, O. (2011). Signatures of demographic bottlenecks in European wolf populations. *Conserv. Genet.* 12, 701–712.
7. Sarà, M. (1999). Il “Catalogo dei Mammiferi della Sicilia” rivisitato. *Soc. messinese di Storia Patria, Messina*.
8. Angelici, F.M., Ciucani, M.M., Angelini, S., Annesi, F., Caniglia, R., Castiglia, R., Fabbri, E., Galaverni, M., Palumbo, D., Ravegnini, G., et al. (2019). The Sicilian Wolf: Genetic Identity of a Recently Extinct Insular Population. *Zoolog. Sci.* 36, 189–197.
9. Reale, S., Randi, E., Cumbo, V., Sammarco, I., Bonanno, F., Spinnato, A., and Seminara, S. (2019). Biodiversity lost: The phylogenetic relationships of a complete mitochondrial DNA genome sequenced from the extinct wolf population of Sicily. *Mamm. Biol.* 98, 1–10.
10. Angelici, F.M., and Rossi, L. (2018). A new subspecies of grey wolf (*Carnivora Canidae*), recently extinct, from Sicily, Italy. *Bollettino del Museo Civico di Storia Naturale di Verona* 42, 3–15.
11. Antonioli, F., Presti, V.L., and Morticelli, M.G. (2016). Timing of the emergence of the Europe–Sicily bridge (40–17 cal ka BP) and its implications for the spread of modern humans. *Geological Society*.
12. Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas, E.J., 3rd, Zody, M.C., et al. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438, 803–819.
13. Fan, Z., Silva, P., Gronau, I., Wang, S., Armero, A.S., Schweizer, R.M., Ramirez, O., Pollinger, J., Galaverni, M., Del-Vecchio, D.O., et al. (2016). Worldwide patterns of genomic variation and admixture in gray wolves. *Genome Res.* 26, 163–173.
14. Freedman, A.H., Gronau, I., Schweizer, R.M., Ortega-Del Vecchio, D., Han, E., Silva, P.M., Galaverni, M., Fan, Z., Marx, P., Lorente-Galdos, B., et al. (2014). Genome Sequencing Highlights the Dynamic Early History of Dogs. *PLoS Genet.* 10, e1004016.
15. Ramos-Madrigal, J., Sinding, M.-H.S., Carøe, C., Mak, S.S.T., Niemann, J., Samaniego Castruita, J.A., Fedorov, S., Kandyba, A., Germonpré, M., Bocherens, H., et al. (2021). Genomes

- of Pleistocene Siberian Wolves Uncover Multiple Extinct Wolf Lineages. *Curr. Biol.* *31*, 198–206.e8.
16. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* *19*, 1655–1664.
17. Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* *35*, 4453–4455.
18. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. *Genetics* *192*, 1065–1093.
19. Ciucani, M.M., Palumbo, D., Galaverni, M., Serventi, P., Fabbri, E., Ravegnini, G., Angelini, S., Maini, E., Persico, D., Caniglia, R., et al. (2019). Old wild wolves: ancient DNA survey unveils population dynamics in Late Pleistocene and Holocene Italian remains. *PeerJ* *7*, e6424.
20. Koupadi, K., Fontani, F., Ciucani, M.M., Maini, E., De Fanti, S., Cattani, M., Curci, A., Nenzioni, G., Reggiani, P., Andrews, A.J., et al. (2020). Population Dynamics in Italian Canids between the Late Pleistocene and Bronze Age. *Genes* *11*.
21. Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* *8*, e1002967.
22. Gómez-Sánchez, D., Olalde, I., Sastre, N., Enseñat, C., Carrasco, R., Marques-Bonet, T., Lalueza-Fox, C., Leonard, J.A., Vilà, C., and Ramírez, O. (2018). On the path to extinction: Inbreeding and admixture in a declining grey wolf population. *Mol. Ecol.* *27*, 3599–3612.
23. Niemann, J., Gopalakrishnan, S., Yamaguchi, N., Ramos-Madriral, J., Wales, N., Gilbert, M.T.P., and Sinding, M.-H.S. (2021). Extended survival of Pleistocene Siberian wolves into the early 20th century on the island of Honshū. *iScience* *24*, 101904.
24. Gojobori, J., Arakawa, N., Xiayire, X., Matsumoto, Y., Matsumura, S., Hongo, H., Ishiguro, N., and Terai, Y. (2021). The Japanese wolf is most closely related to modern dogs and its ancestral genome has been widely inherited by dogs throughout East Eurasia. *bioRxiv*, 2021.10.10.463851.
25. Dussex, N., van der Valk, T., Morales, H.E., Wheat, C.W., Díez-del-Molino, D., von Seth, J., Foster, Y., Kutschera, V.E., Guschanski, K., Rhie, A., et al. (2021). Population genomics of the critically endangered kākāpō. *Cell Genomics* *1*, 100002.
26. Ciucani, M.M., Jensen, J.K., Sinding, M.-H.S., Smith, O., Lucenti, S.B., Rosengren, E., Rook, L., Tuveri, C., Arca, M., Cappellini, E., et al. (2021). Evolutionary history of the extinct Sardinian dhole. *Curr. Biol.* *31*, 5571–5579.e6.
27. Robinson, J.A., Räikkönen, J., Vucetich, L.M., Vucetich, J.A., Peterson, R.O., Lohmueller, K.E., and Wayne, R.K. (2019). Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. *Sci Adv* *5*, eaau0757.
28. Kardos, M., Åkesson, M., Fountain, T., Flagstad, Ø., Liberg, O., Olason, P., Sand, H., Wabakken, P., Wikenros, C., and Ellegren, H. (2018). Genomic consequences of intensive inbreeding in an isolated wolf population. *Nat Ecol Evol* *2*, 124–131.
29. Darimont, C.T., and Paquet, P.C. (2002). Gray wolves, *Canis lupus*, of British Columbia's Central and North Coast: distribution and conservation assessment. *Canadian Field Naturalist*



116, 416–422.

30. Montana, L., Caniglia, R., Galaverni, M., Fabbri, E., Ahmed, A., Bolfiková, B.Č., Czarnomska, S.D., Galov, A., Godinho, R., Hindrikson, M., et al. (2017). Combining phylogenetic and demographic inferences to assess the origin of the genetic diversity in an isolated wolf population. *PLoS One*.
31. Silva, P., Galaverni, M., Ortega-Del Vecchyo, D., Fan, Z., Caniglia, R., Fabbri, E., Randi, E., Wayne, R., and Godinho, R. (2020). Genomic evidence for the Old divergence of Southern European wolf populations. *Proc. Biol. Sci.* 287, 20201206.
32. Bergström, A., Frantz, L., Schmidt, R., Ersmark, E., Lebrasseur, O., Girdland-Flink, L., Lin, A.T., Storå, J., Sjögren, K.-G., Anthony, D., et al. (2020). Origins and genetic legacy of prehistoric dogs. *Science* 370, 557–564.
33. Dabney, J., Meyer, M., and Paabo, S. (2013). Ancient DNA Damage. *Cold Spring Harb. Perspect. Biol.* 5, a012567–a012567.
34. Campos, P.F., and Gilbert, M.T.P. (2019). DNA Extraction from Keratin and Chitin. *Methods Mol. Biol.* 1963, 57–63.
35. Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S.S.T., Sinding, M.H.S., Samaniego, J.A., Wales, N., Sicheritz-Pontén, T., and Gilbert, M.T.P. (2018). Single-tube library preparation for degraded DNA. *Methods Ecol. Evol.* 9, 410–419.
36. Mak, S.S.T., Gopalakrishnan, S., Carøe, C., Geng, C., Liu, S., Sinding, M.-H.S., Kuderna, L.F.K., Zhang, W., Fu, S., Vieira, F.G., et al. (2017). Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing. *Gigascience* 6, 1–13.
37. Sirén, K., Mak, S.S.T., Melkonian, C., Carøe, C., Swiegers, J.H., Molenaar, D., Fischer, U., and Gilbert, M.T.P. (2019). Taxonomic and Functional Characterization of the Microbial Community During Spontaneous in vitro Fermentation of Riesling Must. *Front. Microbiol.* 10, 697.
38. Schubert, M., Ermini, L., Der Sarkissian, C., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M.D., 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.
39. Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* 9, 88.
40. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
41. Toolkit, P. (2018). Picard Toolkit. Broad Institute.
42. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498.
43. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, 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.
44. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., and Orlando, L. (2013).

mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29, 1682–1684.

45. Auton, A., Rui Li, Y., Kidd, J., Oliveira, K., Nadel, J., Holloway, J.K., Hayward, J.J., Cohen, P.E., Gready, J.M., Wang, J., et al. (2013). Genetic recombination is targeted towards gene promoter regions in dogs. *PLoS Genet.* 9, e1003984.
46. Gopalakrishnan, S., Sinding, M.-H.S., Ramos-Madrugal, J., Niemann, J., Samaniego Castruita, J.A., Vieira, F.G., Carøe, C., de Manuel Montero, M., Kuderna, L., Serres, A., et al. (2019). Interspecific Gene Flow Shaped the Evolution of the Genus *Canis*. *Curr. Biol.* 29, 4152.
47. vonHoldt, B.M., Cahill, J.A., Fan, Z., Gronau, I., Robinson, J., Pollinger, J.P., Shapiro, B., Wall, J., and Wayne, R.K. (2016). Whole-genome sequence analysis shows that two endemic species of North American wolf are admixtures of the coyote and gray wolf. *Sci Adv* 2, e1501714.
48. Koepfli, K.-P., Pollinger, J., Godinho, R., Robinson, J., Lea, A., Hendricks, S., Schweizer, R.M., Thalmann, O., Silva, P., Fan, Z., et al. (2015). Genome-wide Evidence Reveals that African and Eurasian Golden Jackals Are Distinct Species. *Curr. Biol.* 25, 2158–2165.
49. Liu, Y.-H., Wang, L., Xu, T., Guo, X., Li, Y., Yin, T.-T., Yang, H.-C., Hu, Y., Adeola, A.C., Sanke, O.J., et al. (2018). Whole-Genome Sequencing of African Dogs Provides Insights into Adaptations against Tropical Parasites. *Mol. Biol. Evol.* 35, 287–298.
50. Zhang, W., Fan, Z., Han, E., Hou, R., Zhang, L., Galaverni, M., Huang, J., Liu, H., Silva, P., Li, P., et al. (2014). Hypoxia adaptations in the grey wolf (*Canis lupus chanco*) from Qinghai-Tibet Plateau. *PLoS Genet.* 10, e1004466.
51. Sinding, M.-H.S., Gopalakrishnan, S., Ramos-Madrugal, J., de Manuel, M., Pitulko, V.V., Kuderna, L., Feuerborn, T.R., Frantz, L.A.F., Vieira, F.G., Niemann, J., et al. (2020). Arctic-adapted dogs emerged at the Pleistocene-Holocene transition. *Science* 368, 1495–1499.
52. Wang, G.-D., Zhai, W., Yang, H.-C., Fan, R.-X., Cao, X., Zhong, L., Wang, L., Liu, F., Wu, H., Cheng, L.-G., et al. (2013). The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat. Commun.* 4, 1860.
53. Wang, G.-D., Zhai, W., Yang, H.-C., Wang, L., Zhong, L., Liu, Y.-H., Fan, R.-X., Yin, T.-T., Zhu, C.-L., Poyarkov, A.D., et al. (2016). Out of southern East Asia: the natural history of domestic dogs across the world. *Cell Res.* 26, 21–33.
54. Wiedmer, M., Oevermann, A., Borer-Germann, S.E., Gorgas, D., Shelton, G.D., Drögemüller, M., Jagannathan, V., Henke, D., and Leeb, T. (2015). A RAB3GAP1 SINE Insertion in Alaskan Huskies with Polyneuropathy, Ocular Abnormalities, and Neuronal Vacuolation (POANV) Resembling Human Warburg Micro Syndrome 1 (WARBM1). *G3* 6, 255–262.
55. Marsden, C.D., Ortega-Del Vecchyo, D., O'Brien, D.P., Taylor, J.F., Ramirez, O., Vilà, C., Marques-Bonet, T., Schnabel, R.D., Wayne, R.K., and Lohmueller, K.E. (2016). Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc. Natl. Acad. Sci. U. S. A.* 113, 152–157.
56. Metzger, J., Nolte, A., Uhde, A.-K., Hewicker-Trautwein, M., and Distl, O. (2017). Whole genome sequencing identifies missense mutation in MTBP in Shar-Pei affected with Autoinflammatory Disease (SPAID). *BMC Genomics* 18, 348.
57. Becker, D., Minor, K.M., Letko, A., Ekenstedt, K.J., Jagannathan, V., Leeb, T., Shelton, G.D., Mickelson, J.R., and Drögemüller, C. (2017). A GJA9 frameshift variant is associated with polyneuropathy in Leonberger dogs. *BMC Genomics* 18, 662.

58. Decker, B., Davis, B.W., Rimbault, M., Long, A.H., Karlins, E., Jagannathan, V., Reiman, R., Parker, H.G., Drögemüller, C., Corneveaux, J.J., et al. (2015). Comparison against 186 canid whole-genome sequences reveals survival strategies of an ancient clonally transmissible canine tumor. *Genome Res.* 25, 1646–1655.
59. Plassais, J., Kim, J., Davis, B.W., Karyadi, D.M., Hogan, A.N., Harris, A.C., Decker, B., Parker, H.G., and Ostrander, E.A. (2019). Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat. Commun.* 10, 1489.
60. Botigué, L.R., Song, S., Scheu, A., Gopalan, S., Pendleton, A.L., Oetjens, M., Taravella, A.M., Seregély, T., Zeeb-Lanz, A., Arbogast, R.M., et al. (2017). Ancient European dog genomes reveal continuity since the Early Neolithic. *Nat. Commun.* 8.
61. Frantz, L.A.F., Mullin, V.E., Pionnier-Capitan, M., Lebrasseur, O., Ollivier, M., Perri, A., Linderholm, A., Mattiangeli, V., Teasdale, M.D., Dimopoulos, E.A., et al. (2016). Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science* 352, 1228–1231.
62. Feuerborn, T.R., Carmagnini, A., Losey, R.J., Nomokonova, T., Askeyev, A., Askeyev, I., Askeyev, O., Antipina, E.E., Appelt, M., Bachura, O.P., et al. (2021). Modern Siberian dog ancestry was shaped by several thousand years of Eurasian-wide trade and human dispersal. *Proc. Natl. Acad. Sci. U. S. A.* 118.
63. Ní Leathlobhair, M., Perri, A.R., Irving-Pease, E.K., Witt, K.E., Linderholm, A., Haile, J., Lebrasseur, O., Ameen, C., Blick, J., Boyko, A.R., et al. (2018). The evolutionary history of dogs in the Americas. *Science* 361, 81–85.
64. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15, 356.
65. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575.
66. Behr, A.A., Liu, K.Z., Liu-Fang, G., Nakka, P., and Ramachandran, S. (2016). pong: fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics* 32, 2817–2823.
67. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
68. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
69. Rambaut, A. (2012). FigTree v. 1.4.0. <http://tree.bio.ed.ac.uk/software/figtree/>.
70. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
71. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* 37, 1530–1534.
72. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* 35, 518–522.
73. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermini, L.S. (2017).

- ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* *14*, 587–589.
74. Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* *19*, 153.
75. Computing, R., and Others (2013). R: A language and environment for statistical computing. R Core Team.
76. Liu, L., Bosse, M., Megens, H.-J., Frantz, L.A.F., Lee, Y.-L., Irving-Pease, E.K., Narayan, G., Groenen, M.A.M., and Madsen, O. (2019). Genomic analysis on pygmy hog reveals extensive interbreeding during wild boar expansion. *Nat. Commun.* *10*, 1–9.
77. Leppälä, K., Nielsen, S.V., and Mailund, T. (2017). admixturegraph: an R package for admixture graph manipulation and fitting. *Bioinformatics* *33*, 1738–1740.
78. Lipson, M., and Reich, D. (2017). A Working Model of the Deep Relationships of Diverse Modern Human Genetic Lineages Outside of Africa. *Mol. Biol. Evol.* *34*, 889–902.
79. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. *Bioinformatics* *27*, 2156–2158.
80. Allaire, J. (2012). RStudio: integrated development environment for R. Boston, MA.
81. RStudio Team (2020). RStudio: Integrated Development Environment for R.
82. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag New York).
83. Hadley Wickham, R.F., Henry, L., Müller, K., and Others (2018). dplyr: A grammar of data manipulation. Version 0. 7 6.
84. Renaud, G., Hanghøj, K., Korneliussen, T.S., Willerslev, E., and Orlando, L. (2019). Joint Estimates of Heterozygosity and Runs of Homozygosity for Modern and Ancient Samples. *Genetics* *212*, 587–614.
85. Greer, D. (2020). Extinction Genomics: The Case of Grey Wolves Year, [https://github.com/DebGreer/MSc\\_Project](https://github.com/DebGreer/MSc_Project) (Queen Mary University of London).