**iScience, Volume *26***

## Supplemental information

# The extinct Sicilian wolf shows a complex history of isolation and admixture with ancient dogs

**Marta Maria Ciucani, Jazmín Ramos-Madrigal, Germán Hernández-Alonso, Alberto Carmagnini, Sabhrina Gita Aninta, Xin Sun, Camilla Hjorth Scharff-Olsen, Liam Thomas Lanigan, Ilaria Fracasso, Cecilie G. Clausen, Jouni Aspi, Ilpo Kojola, Laima Baltrūnaitė, Linas Balčiauskas, Jane Moore, Mikael Åkesson, Urmas Saarma, Maris Hindrikson, Pavel Hulva, Barbora Černá Bolfíková, Carsten Nowak, Raquel Godinho, Steve Smith, Ladislav Paule, Sabina Nowak, Robert W. Mysłajek, Sabrina Lo Brutto, Paolo Ciucci, Luigi Boitani, Cristiano Vernesi, Hans K. Stenøien, Oliver Smith, Laurent Frantz, Lorenzo Rossi, Francesco Maria Angelici, Elisabetta Cilli, Mikkel-Holger S. Sinding, M. Thomas P. Gilbert, and Shyam Gopalakrishnan**

## SUPPLEMENTAL FIGURES

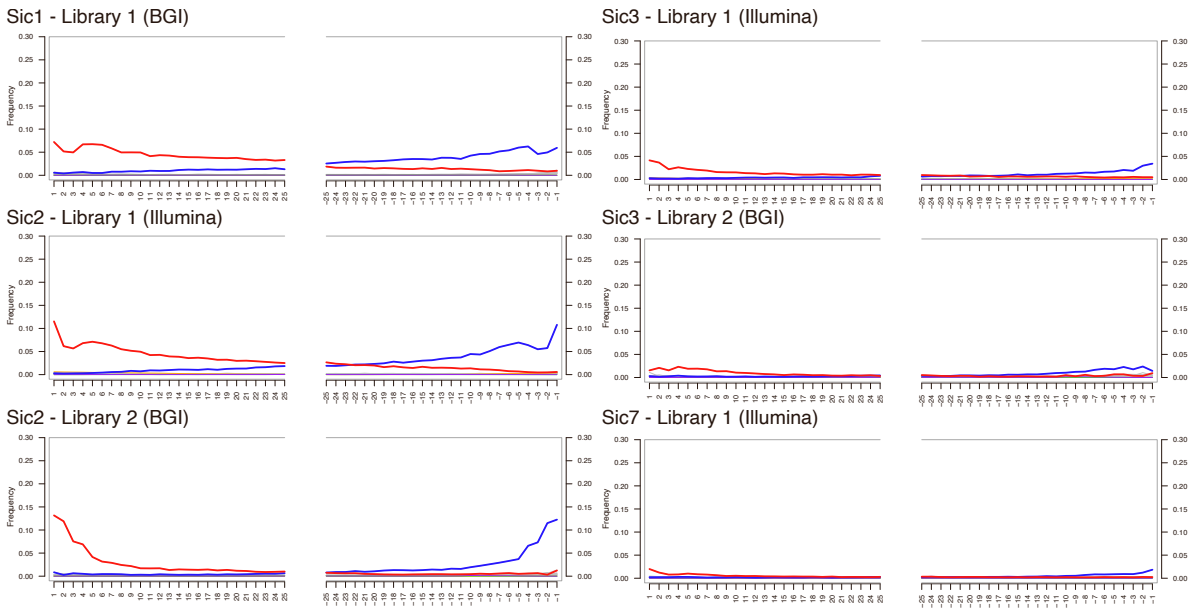


**Fig. S1: Sicilian wolves' misincorporation plots from mapDamage[1], related to Figure 1.** Nucleotide misincorporation patterns of the Sicilian wolf samples mapped to the dog reference genome (CanFam3.1) [2]. Two plots are shown for Sic2 and Sic3, one for the Illumina libraries and one for BGI libraries. For the samples Sic1 and Sic7 only the Illumina libraries plots are shown.

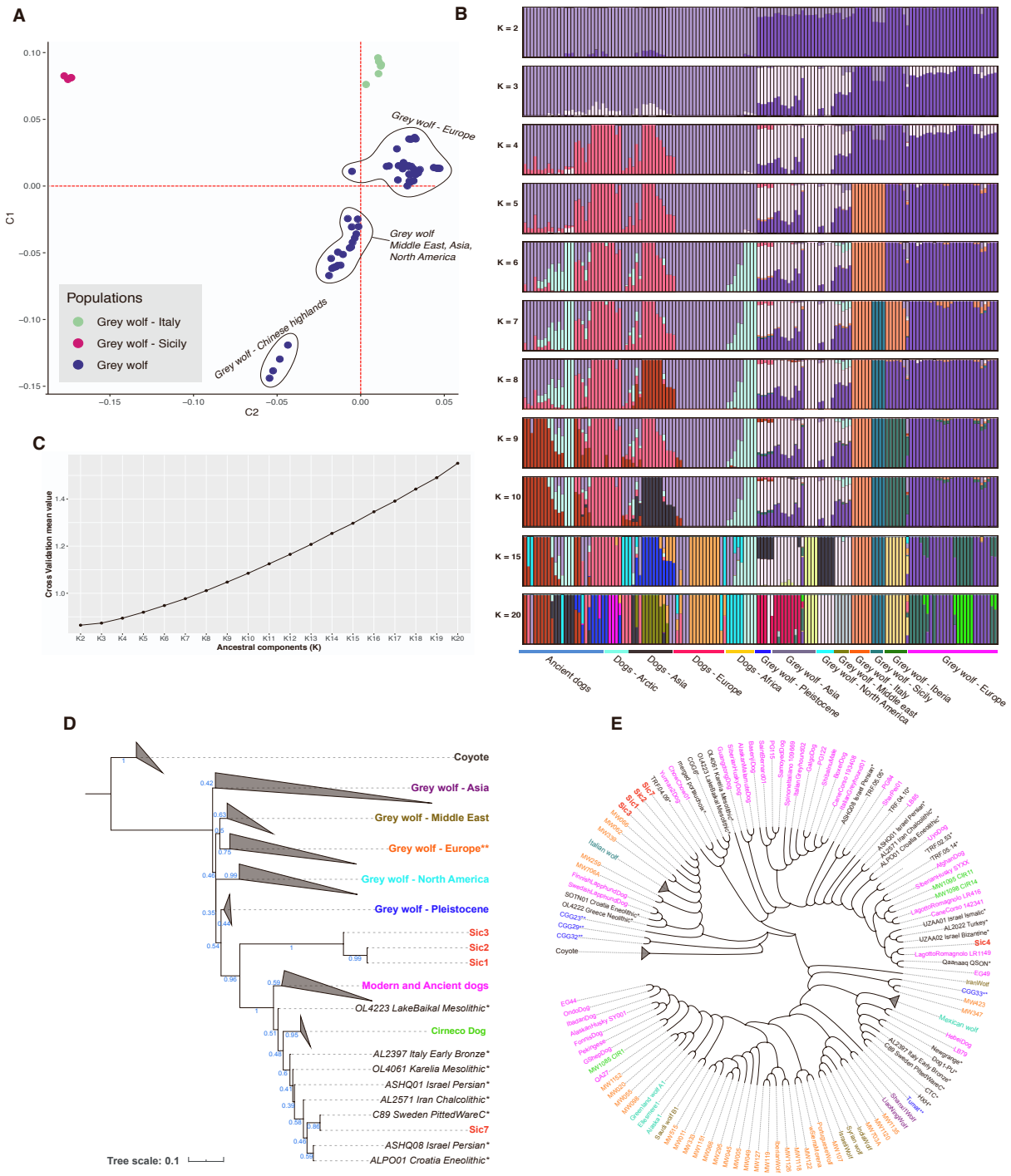


**Fig. S2: Population structure and nuclear and mitochondrial phylogenies of the Sicilian wolves, related to Figure 1 and Figure 2.**

(**A**) Multidimensional Scaling (MDS) analysis restricting to wolves. Grey wolves from Eurasia and North America are represented in purple while Italian and Sicilian wolves are represented in light green and pink, respectively.

(**B**) Admixture proportions assuming from 2 to 10 ancestry components (K=10) estimated using ADMIXTURE [3] and the pseudo-haploid SNP panel. Each vertical bar represents a different genome, and the proportions of each colour show the estimated ancestry proportions.
(**C**) Cross-validation errors for the ADMIXTURE results in B.
(**D**) Nuclear phylogeny of wolves and dogs using coyotes as outgroup. The phylogeny was computed by ASTRAL-III [4] on 1000 trees generated with IQtree [5] v.2.1.2 using 1000 bootstrap replicates and *ModelFinder Plus* [6] to identify the best evolutionary model for each region. The Sicilian wolves are represented in bold red, while the star (*) represents ancient samples. Modern and ancient dogs are represented in pink, Pleistocene wolves in blue, North American wolves in turquoise, European wolves in orange, Middle Eastern wolves in sand and Asian wolves in purple. * Note that the Italian wolves are collapsed within the European wolf clade.
(**E**) Mitochondrial phylogeny of wolves and dogs using the coyote as an outgroup. The Sicilian wolf is shown in red. The label of each sample is coloured based on its group or geographic location according to the legend in Fig. S2 D.

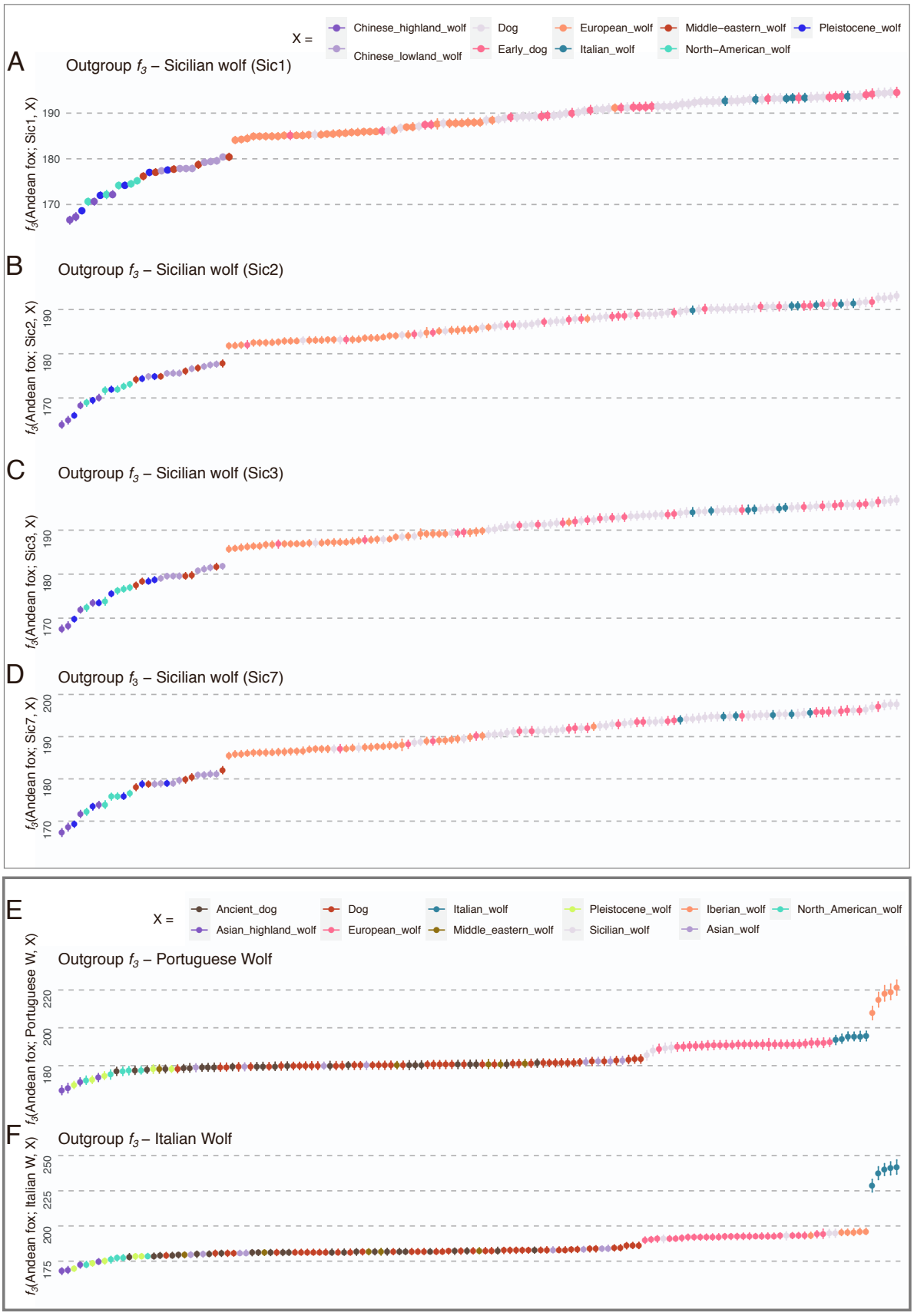


**Fig. S3: Outgroup f3 for Sicilian, Italian and Iberian wolves, related to Figure 2.**

(**A-D**) Outgroup $f_3$-statistics of the form $f_3$(Andean fox; Sicilian wolf, X), where X represents all samples in the dataset (each of the points coloured according to the legend on the top). $F_3$-statistics were estimated for each of the four Sicilian wolves (Sic1, Sic2, Sic3 and Sic7). The Andean fox was used as an outgroup. Italian wolves, European modern and ancient dogs show the highest $f_3$-statistics, therefore these populations are the closest to the Sicilian wolf. (**E-F**) Outgroup $f_3$-statistics of the for $f_3$(Andean fox; Iberian/Italian wolf, X), where X represents all samples in the dataset. $F_3$-statistics for the Iberian (Portuguese) and Italian wolf were used as comparison for the Sicilian wolf. Points show the estimated $f_3$-statistics for each of the tests. Error bars correspond to 3.3 standard errors estimated through a block jackknife approach.

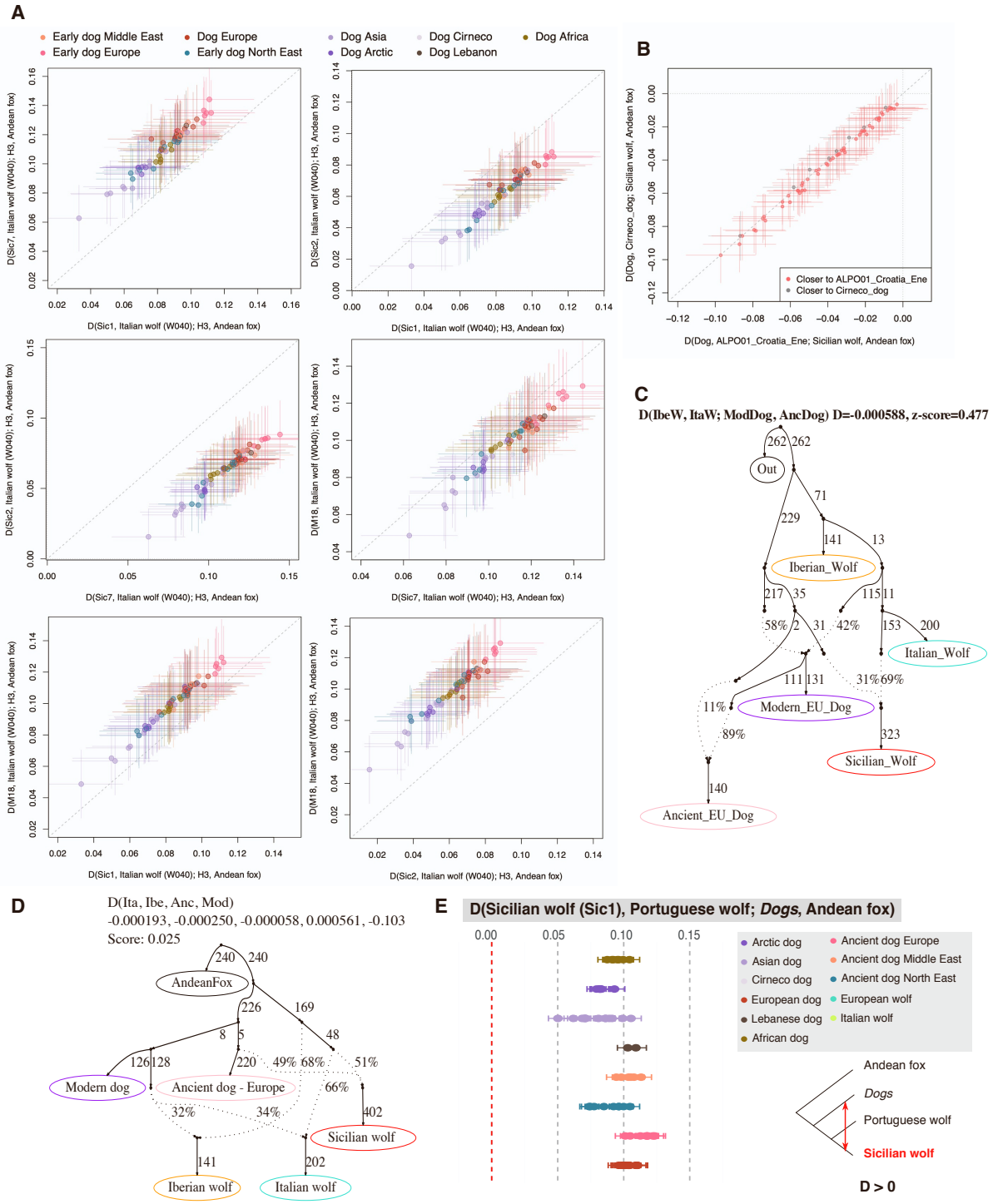


**Fig. S4: Testing for gene flow between the Sicilian wolf and dogs, related to Figure 2.**

(**A**) *D*-statistics testing for gene flow between the Sicilian wolf and dogs. The x- and y-axes show *D*-statistics of the form of *D*(Sicilian wolf, Italian wolf, H3, Andean fox), where H3 represents all dogs in the dataset. In each of the plots, we compare the results of obtained from different Sicilian wolf genomes (*e.g.* in the top left the x-axis corresponds to *D*(Sic1, Italian wolf, H3, Andean fox) and the y-axis *D*(Sic7, Italian wolf, H3, Andean fox)). Each point corresponds to a different test

where H3 varies. Error bars correspond to 3.3 standard errors estimated through a block jackknife approach. All tests yielded significant results (z-score > 3.3) suggesting gene flow between the Sicilian wolf and the corresponding dog in each test. A dotted line indicates the identity line, with points above the line showing tests where the Sicilian wolf in the x-axi had higher affinity to H3 compared to the Sicilian wolf in the y-axis.

(**B**) *D*-statistic test of the form D(*Dogs*, Croatia Eneolithic dog (ALP01); Sicilian wolf, Andean fox) on the x-axis, and D(*Dogs*, Cirneco dog; Sicilian wolf, Andean fox) on the y-axis. Each point shows the results of a different test where the dog varies. Error bars correspond to 3.3 standard errors estimated through a block jackknife approach.

(**C-D**) $F_4$-statistics-based admixture graphs modelling the ancestry of the Sicilian wolf, Italian wolf, Iberian wolf, ancient European dogs, modern dogs and the Andean fox as an outgroup. Graph in panel C shows the best fitting graph obtained using qpBrute, while panel D shows the best fitting graph obtained using a 'base-graph' approach as described in the methods section.

(**E**) *D*-statistics of the form D(Sicilian wolf, Portuguese wolf, *Dogs*, Andean fox). Error bars correspond to 3.3 standard errors estimated through a block jackknife approach. All tests yielded significant results (z-score > 3.3) suggesting gene flow between the Sicilian wolf and the corresponding dog in each test.

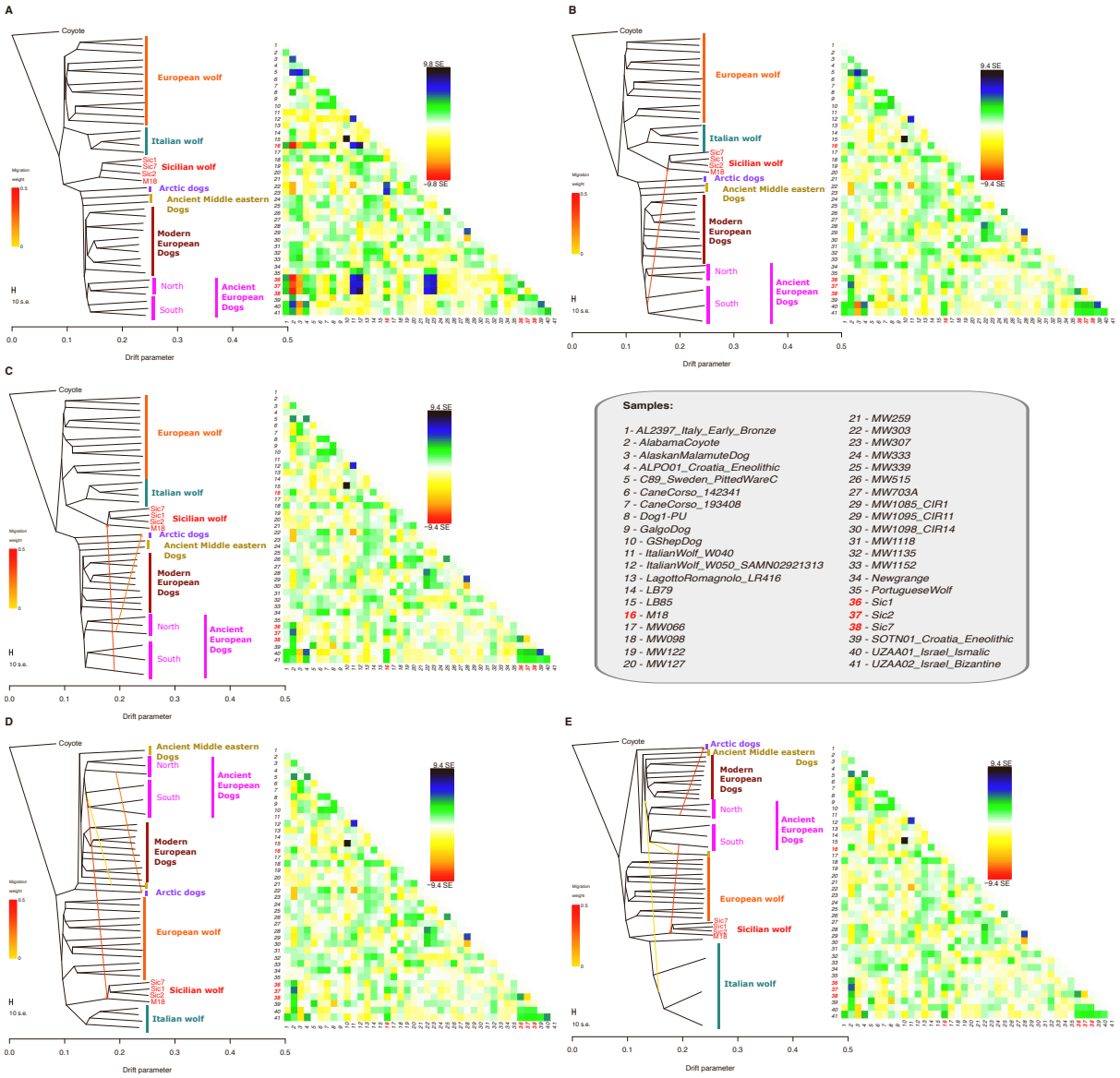


**Fig. S5: TreeMix [7] results including 0 up to 4 migration events, related to Figure 2.**

**A-E**) TreeMix admixture graph estimating from 0 to 4 migration edges and their corresponding residuals. A) TreeMix tree without migration edges shows the Sicilian wolves basal to modern and ancient dogs. B) TreeMix tree with one migration edge shifts the Sicilian wolves position as sister to the Italian wolves and a first arrow (with high weight: 0.322216) departs from the two Croatian Eneolithic dogs into the base of the Sicilian wolves. C) TreeMix tree with two migration edges: a second arrow from the Swedish Mesolithic dog (C89) goes into the Alaskan Malamute. D) TreeMix tree with three migration edges: the third migration edge brings a Middle eastern component from a ~900BP Israelian dog into Southern European ancient dogs

(Italy and Croatia). E) TreeMix tree showing four migration edges: the fourth arrow goes from two Italian wolf samples (MW303 and MW307) to the base of modern dogs.

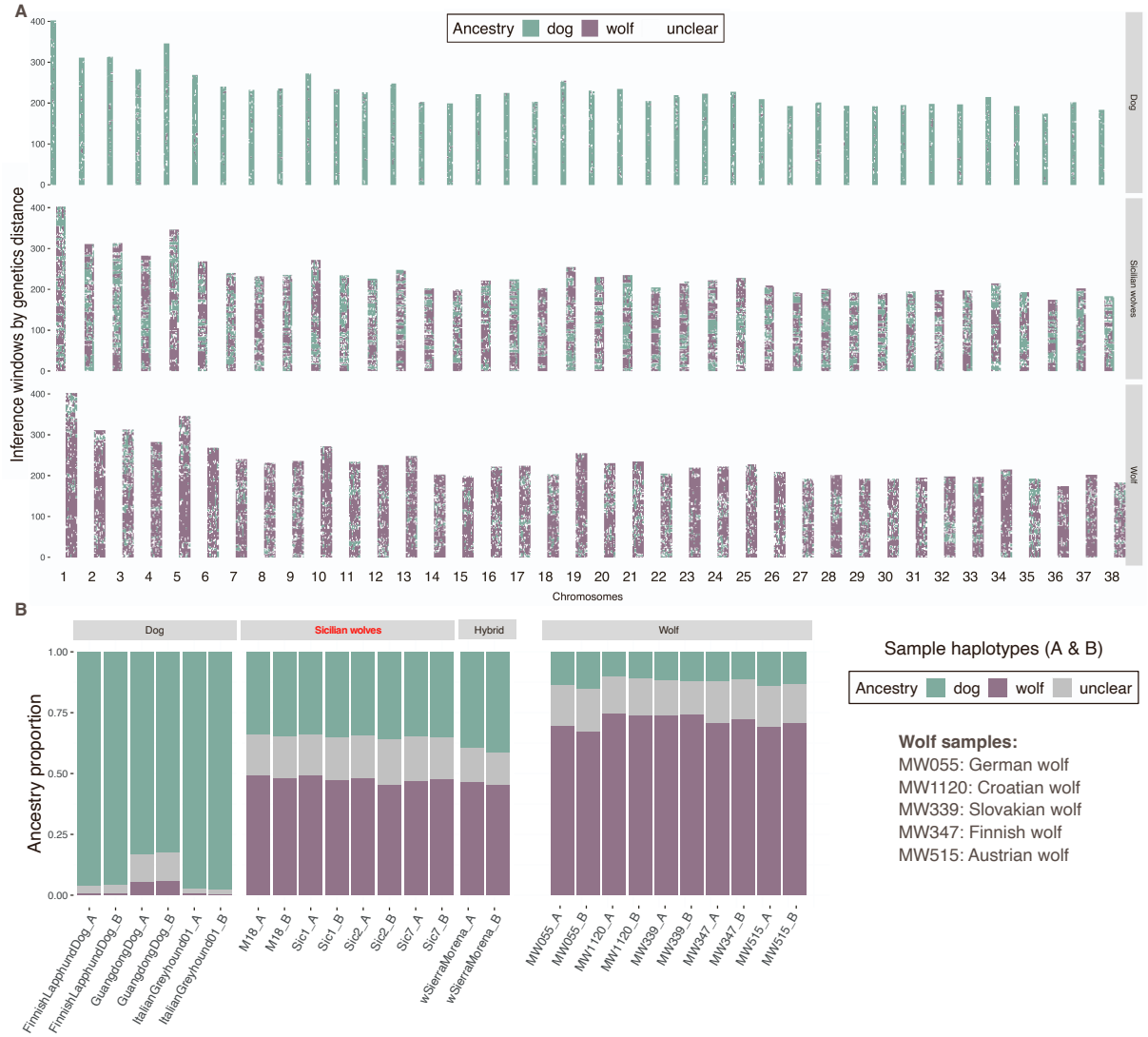


**Fig. S6: Local ancestry inferred using PCAdmix on wolves and dogs, related to Figure 2 and Figure 3.**

A) PCAdmix plot showing the local ancestry inferred in windows of size 0.01cM for each chromosome and three groups: dogs, Sicilian wolves and wolves. Windows associated with dog, wolf or unclear ancestry are coloured in green, purple and grey, respectively.

B) Summarised ancestry proportions of the genomes of the Sicilian wolves, dogs and European wolves from PCAdmix inference. Regions associated with dog ancestry are

coloured in green while regions associated with wolf ancestry are coloured in purple while unclear regions are represented in grey.

**Table S1. Supplemental table containing information of the Sicilian wolf samples used in this study, related to Figure 1.**

Columns: samples ID, museum ID, date, museum of origin, average read length of mapped reads, average genomic coverage, mitochondrial DNA coverage, chrX coverage, chrX/nuclear genome coverage ratio, genetic sex, endogenous content (%) and clonality (%). N.A.: Not available. PE: Paired-end reads. SE: Single end reads.

| Sample ID | Museum ID | Date | Museum of origin/ Sample Information | Sequenced reads | Mapped filtered reads **** | Read length | Avg. genomic coverage | mtDNA coverage | ChrX coverage | ChrX/ nuclear coverage ratio | Genetic Sex | Endo. (%) | Clonality (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sic1 | C11875 | 1883 | Museum of Natural History, Section of Zoology 'La Specola' | 791,813,357 (PE) | 307,966,094 | 82.47 | 11.6 | 174.32 | 7.83 | 0.675 | Male | 21.65 | 17,92** |
| Sic2 | AN/855 | 1879 | Museum of Zoology 'Pietro Doderlein' | 336,276,750 (PE) | 207,178,656 | 51.17 | 5.39 | 585.91 | 3.66 | 0.679 | Male | 21.69 | 59,55** |
| Sic3 | M/18 | 1870-1880 | Museum of Zoology 'Pietro Doderlein' | 146.462,246 (PE) 73,2491,992 (SE) | 160,919,870 | 49.37 | 4.24 | 1239,19 | 3 | 0.708 | Male | 22.31 | 70.87** |
| Sic4*** | TSA / 9 | 1924 | Regional Museum of Terrasini (PA), Italy | N.A. | N.A. | 39.71 | N.A. | 19,73 | N.A. | N.A. | N.A. | 13,46 | 96,02 |
| Sic5* | TSB | N.A. | Regional Museum of Terrasini (PA), Italy | N.A. | N.A. | 46,15 | 0.0005 | N.A. | N.A. | N.A. | N.A. | 0,18 | 59,45 |
| Sic6* | M/17 | N.A. | Museum of Zoology 'Pietro Doderlein' | N.A. | N.A. | 41.47 | 0.024 | N.A. | N.A. | N.A. | N.A. | 19,7 | 74,34 |
| Sic7 | / | ~1880-1920 | Termini Imerese | 1,499,951,713 (PE) | 32,143,102 | 41.27 | 3.8 | 99.46 | 5.17 | 1.361 | Female | 14.09 | 83,12** |

* These samples were not deep sequenced due to poor DNA preservation or high clonality levels. We did not use the screening data generated for these samples.

** The clonality is estimated after deep sequencing

*** This sample was labelled *Canis lupus* but morphologically described to have "strongly anomalous features, thus possibly attributable to a feral dog or a hybrid.

**** Reads mapped to the dog (CanFam3) reference genome after removing PCR duplicates and reads with mapping quality below 20.

## SUPPLEMENTAL REFERENCES


1. 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.

2. 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.

3. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.

4. 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.

5. 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.

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

7. Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. *8*, e1002967.