

DivjaLabs Ltd.

Aljaževa ulica 35a, SI-1000 Ljubljana, Slovenia

info@divjalabs.com

www.divjalabs.com

a University of Ljubljana Spinout Company



divjaLABS

#BiodiversityInSight

Single nucleotide polymorphism (SNP) analyses of canid samples from the Netherlands

Order number: 4017554 from Sep. 28th 2023

Client: Provincie Gelderland, Postbus 9090, 6800 GX Arnhem, Nederland

Authors: [REDACTED] 5.1.2e

Ljubljana, 22. 4. 2024

Suggested citation:

5.1.2e

2024. *Single nucleotide polymorphism (SNP) analyses of canid samples from the Netherlands*. Report – Order number: 4017554 from Sep. 28th 2023 from Provincie Gelderland, Nederland. Ljubljana. DivjaLabs Ltd.

Contents

Table of Figures	3
Introduction	4
Laboratory methods	4
Filtering of SNP profiles	5
Results	8
Discussion	15
References	16

Table of Figures

Figure 1: ADMIXTURE analyses with cross-validation (CV) error plot for K-values 1-10 including 127 canid profiles and 68,583 single nucleotide polymorphism (SNP) loci.	9
Figure 2: ADMIXTURE plot for K = 7 population clusters and n = 127 canid profiles, analysed with 68,583 single nucleotide polymorphism (SNP) loci	10
Figure 3: Principal component analysis of n = 127 canids with 68,583 SNP loci, using the colorplot-function	11
Figure 4: Principal component analysis of n = 112 canid samples with 68,583 SNP loci, using the colorplot-function	Error! Bookmark not defined.
Figure 5: Principal component analysis of n = 97 canids with 68,583 SNP loci, showing results per sampling group..	13
Figure 6: Principal component analysis of n = 81 canids with 96 SNP loci, showing results per sampling group..	14
Appendix 1A: ADMIXTURE plot for K = 6 population clusters and n = 127 canid profiles, analysed with 68,583 single nucleotide polymorphism (SNP) loci.	18
Appendix 1B: ADMIXTURE plot for K = 5 population clusters and n = 127 canid profiles, analysed with 68,583 single nucleotide polymorphism (SNP) loci.	19
Appendix 2A: Principal component analysis of n = 97 canids with 44,112 SNP loci, showing results per sampling group	20
Appendix 2B: Principal component analysis of n = 81 canids with 96 SNP loci, showing results per sampling group	21

Introduction

We performed analysis for $n = 11$ canid tissue samples submitted from the Netherlands. The samples were provided by the client, and the study was motivated by concerns around wolf-dog hybridization, which has been raised as an important long-term conservation concern for wolf populations across Europe (Salvatori et al. 2021).

Wolves have over the past decades recolonized large parts of their historical range, including parts of central and northern Europe (Hindrikson et al. 2017), a process involving long-distance dispersal, bottlenecks, and founder effects that have influenced population genetic structure (e.g., Andersen et al. 2015, Hulva et al. 2018, Jarausch et al. 2021, Szewczyk et al. 2019; 2021). Wolves were found to have returned to the Netherlands in 2015 (Lelieveld et al. 2016) and to Luxembourg in 2017 (Schley et al. 2021), where results from the latter study indicated immigration from the Alpine and Central European populations.

In the past decades, development in genomics have provided important advances for monitoring and research on wide-ranging wild species (e.g., Segelbacher et al. 2022). For domestic and related wild species with well-known genomes, there have been rapid developments in genomic resources that include high-density arrays with thousands of single nucleotide polymorphism (SNP) markers, providing new information on population genetic structure and publicly available datasets for species such as domestic dogs and wolves (e.g., Stronen et al. 2013, Vaysse et al. 2011).

Additionally, advancements in technologies for analysis of non-invasive sampling for wild species (e.g., De Barba et al. 2017, von Thaden et al. 2020) have permitted the development of reduced panels with discriminant markers for wolves, domestic dogs, and golden jackals (Harmoinen et al. 2021, Stronen et al. 2022).

The aim of this analyses was to genotype $n = 11$ canid samples from the Netherlands with the high-density SNP array and assess their ancestry, by comparing the genome-wide profile with relevant data from other canids. These include domestic dogs, golden jackals, and wolf population in other parts of Europe, with profiles from Germany, Poland, Italy, and beyond.

Laboratory methods

DNA was extracted from the tissue samples using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. Because of low DNA quality and quantity, we prepared three extracts for each sample and pooled them together, to obtain the required DNA concentration. Nevertheless, three samples did not obtain the optimal concentration for the analysis, and we ran out of the sample material (tissue). Samples were submitted to Eurofins Genomics (Eurofins Genomics Europe AgriGenomics Products & Services A/S, Galten, Denmark), for genotyping on the Illumina Canine HD Array with more than 170,000 SNP markers. All samples were genotyped in three replicates.

Filtering of SNP profiles

Genotyping results were highly successful for nine of the $n=11$ samples, with call rates (the number of SNPs successfully genotyped) over 99% (Table 1). However, two samples were less successful across all three replicates that had been submitted for genotyping per individual. Sample 11A was marginal, with the best sample having a call rate of approximately 0.88 (88%), whereas sample 10A was markedly lower, with the best result being just below 0.81 (81%). Accordingly, these two would typically have been screened out from further analyses but were retained to allow at least a partial comparison with other profiles from the Netherlands (hereafter NLD profiles) and reference populations.

Table 1. Individual call rate for single nucleotide polymorphism (SNP) genotypes for canid tissue samples from the Netherlands.

Sample number	Individual SNP call rate	Note
01A	0.994391	Best of x3 profiles
02A	0.994065	Best of x3 profiles
03A	0.994408	Best of x3 profiles
04A	0.994269	Best of x3 profiles
05A	0.994443	Best of x3 profiles
06A	0.994356	Best of x3 profiles
07A	0.994536	Best of x3 profiles
08A	0.994426	Best of x3 profiles
09A	0.994007	Best of x3 profiles
10A	0.808946	Best of x3 profiles (marginal)
11A	0.880420	Best of x3 profiles (poor)

Hence, we retained all of the profiles for filtering and population genetic analyses with other reference samples. However, individuals with lower call rates that display genomic signatures of being outliers will often remain unresolved, because it is not possible to determine whether they are true outliers or appear divergent because of missing data. In contrast, outliers with high-quality profiles can typically be explained by the relevant individuals being immigrants or their descendants, or because of hybridization (e.g., Andersen et al. 2015, Harmoinen et al. 2021).

The raw data from the Illumina CanineHD array was filtered in the publicly available software PLINK 1.9 (Purcell et al. 2007, Chang et al. 2015, <https://www.cog-genomics.org/plink/>) according to the user manual and following approaches used for reference profiles from Stronen et al. (2013). The dataset was filtered with a per-SNP genotyping rate of 10 percent (plink command `--geno 0.10`) to retain the SNPs with the highest success rates. This was done with the triplicate profiles for all NLD individuals to provide a larger and more complete dataset for the filtering analyses, but we here removed the profiles with poorer success rates (from 10A and 11A, see Table 1) to avoid losing too many SNPs based on these two individuals.

We removed SNPs mapped to chromosome zero (unsuccessfully genotyped) and to the X and Y chromosome to retain autosomal SNPs, and next filtered for SNP loci in Hardy-Weinberg disequilibrium, using a low value as recommended by the PLINK authors and filtering with a threshold of 0.000001 (PLINK command `--hwe 10-6 'midp'`). The resulting data were screened with a minor allele frequency threshold of one percent (PLINK command `--maf 0.01`), and subsequently for loci in strong linkage disequilibrium, filtering SNPs with pairwise genotype associations ($r^2 > 0.8$) (PLINK command `--indep-pairwise 50 5 0.8`). We selected the best profile for each of the NLD samples based on the sample call rates and continued analyses with these profiles, which were merged with existing genome-wide profiles from relevant canid populations that might have influenced NLD profile ancestry. This dataset included profiles from golden jackals ($n=15$), domestic dogs ($n=12$), first-generation wolf-dog hybrids (WDH-F1, $n=4$), and $n = 96$ profiles from relevant wolf populations, as outlined in Table 2.

Table 2. Reference wolf profiles used in population genetic analyses of NLD samples. These are arranged in order from north to south. The Alpine population has ancestry from both Italian and Dinaric wolves, and both sources are included here.

Code	Number of profiles	Sources	Wolf population represented
FIN	$n = 8$	Finland	Northern Europe
PON	$n = 15$	Northern Poland	Baltic population
GER	$n = 9$	Germany	Central European population
CAR	$n = 10$	Southern Poland, Slovakia, Western Ukraine	Carpathian population
ITA	$n = 15$	Italy	Italian population
DIN	$n = 14$	Croatia, Slovenia	Dinaric population
BAL	$n = 15$	Bulgaria, Greece	Balkan population

The reference populations, which include profiles from Stronen et al. (2013) and other unpublished canid data, have all been converted to the CanFam3.1 version of the dog genome to provide a consistent dataset, because Stronen et al. (2013) used and published data with an earlier CanFam2 version of the dog genome. This reference dataset with CanFam3.1 coordinates therefore matches the CanFam3.1 genome coordinates of the canid profiles genotyped for this study.

We analysed the data with complementary methods, using approaches that have different assumptions and thus allow comparison of results to assess which findings are supported by multiple methods. These included principal component analyses (PCA) in the package *adegenet* (Jombart 2008) in R (R Core Team 2022). This approach has no underlying assumptions about equilibrium conditions in the analysed data, and one of its strengths is the ability to detect outlier profiles such as dispersing individuals.

Additionally, we used ADMIXTURE (Alexander et al. 2009), a maximum likelihood method developed for assessment of population genetic structure in larger genomic data sets, which has underlying population genetic assumptions and thus differ from the PCA. ADMIXTURE analyses were done for $K = 1-10$ populations, with 10 cross-validations (CV), 1000 bootstraps (B), using 24 cores (j) and a random seed generated from current time (-s).

From the array, we also extracted the loci used in the 96-SNP panel developed by Harmoinen et al. (2021) for the detection of wolf-dog hybridization. For the analysis of 96-SNP profiles, we retained profiles from relevant populations based on the genome-wide results (Table 2) and added new groups, or additional profiles to existing groups, as shown in Table 3.

Table 3. Reference profiles used in population genomic analyses of NLD samples with 96 SNP profiles. To improve resolution of visual outputs, only the most relevant wolf populations are included based on the genome-wide results. Unless otherwise noted, these canid profiles included were obtained from noninvasive monitoring in the Dinaric-Balkan area.

Code	Number of profiles	Sources	Population represented
DOG	n = 15	Unpublished noninvasive genetic monitoring results	Domestic dogs
Dbx2	n = 8	Unpublished noninvasive genetic monitoring results	Second generation backcrosses to dogs
WDH-F1	n = 8	Unpublished noninvasive genetic monitoring results and existing unpublished data	First-generation wolf-dog hybrids
Wbx2	n = 15	Unpublished noninvasive genetic monitoring results and existing unpublished data	Second generation backcrosses to wolves
PON	n = 15	Northern Poland	Baltic wolf population
GER	n = 9	Germany	Central European wolf population

Because the 96 SNP panel is suitable also for non-invasive data, it permitted analyses with additional unpublished reference profiles from faecal samples and other non-invasive sources such as saliva and hair, obtained during other canid monitoring projects and from analyses of livestock damages. This analysis with 96 discriminant SNPs therefore permitted comparison with additional canid samples collected in the Dinaric-Balkan area, previously genotyped and analysed in collaboration with Senckenberg Research Institute in Gelnhausen, Germany. Accordingly, WDH-F1, Wbx2 and Dbx2 canids listed in Table 3 have been assigned to these groups based on the Senckenberg laboratory's extensive reference sample collection and developed data analysis pipeline for assessment of canid profiles.

For the two samples that had lower quality genotypes (10A and 11A) we programmed an R script to make consensus genotypes using all three parallel repeated analyses. For the loci that had identical genotypes in at least two repeats, we used this genotype in the consensus. For the loci that didn't meet this criterion (i.e. that had different or missing genotypes) we set the consensus as "no data". This procedure should minimize the residual error in the consensus dataset.

Results

After filtering of the genomic data, we obtained a set of 77,312 SNPs for downstream analyses. These were merged with existing reference profiles and again filtered with a per-SNP genotyping filter of 90% to retain SNPs with high genotyping rates across the dataset. This resulted in a dataset set of $n = 127$ individuals with 68,583 (68K) SNPs.

ADMIXTURE analyses showed the highest support for $K = 7$ clusters (Figure 1), followed by $K = 6$ and $K = 5$. The $K = 6$ results clearly identified jackals (JAC), dogs (DOG), and Italian wolves (ITA), and indicated genetic structuring between wolves in northern and southern Europe (Figure 2, Appendix 1A, B). In northern Europe, the ancestry of the target NLD canid samples generally suggested influence from the Central European and Baltic wolf population (samples GER and PON, respectively).

Whereas this was the case also for the marginal profile 11A, the poor profile 10A had a profile where the possibility of a backcross to dogs could not be excluded. In contrast, the four reference WDH-F1 individuals (collected outside of this study) reflected their known origin, with the three first canids being F1 hybrids between Italian wolves and dogs, and the fourth individual being an F1 hybrid between a German wolf and a dog.

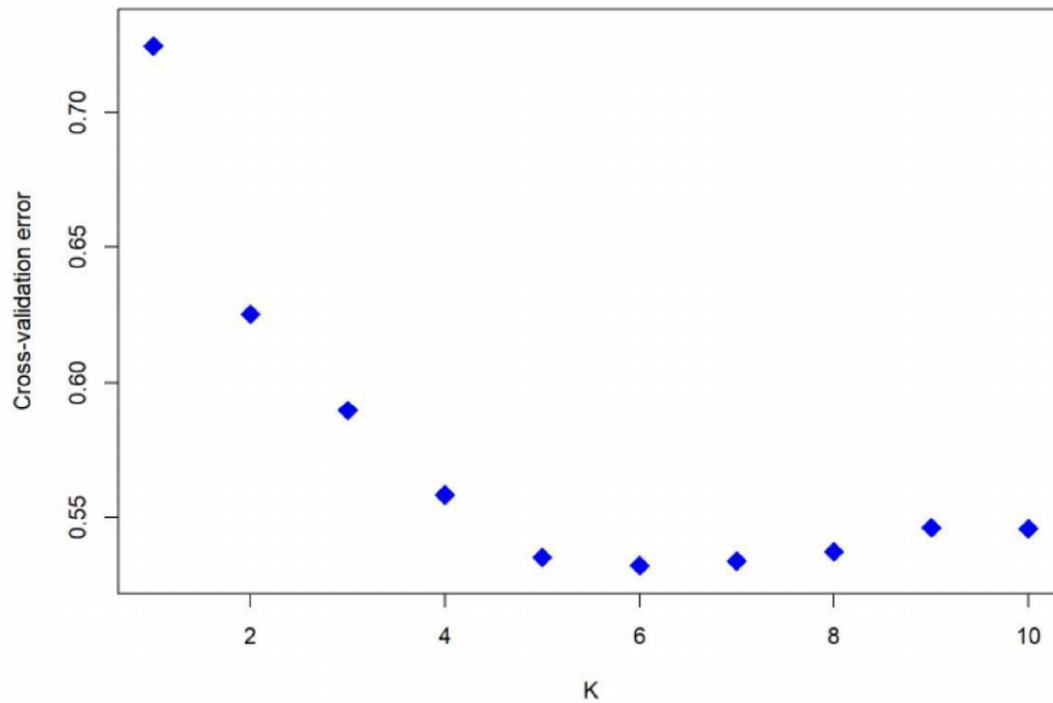


Figure 1. ADMIXTURE analyses with cross-validation (CV) error plot for K-values 1-10, including 127 canid profiles and 68,583 single nucleotide polymorphism (SNP) loci. The CV error markedly declines with each K-value until K = 5. The subsequent values for K from 5-8 are similar, although the highest support (lowest value) was shown for K = 7.

Neither ADMIXTURE or PCA supported any similarity with golden jackal profiles (Figures 2 and 3) and showed golden jackal profiles as highly distinct from other canids included in the analyses. We therefore removed these samples to better resolve the remaining profiles, which concurred with ADMIXTURE results in finding that Italian wolves were also clearly differentiated from other canids (Figure 4). We thus removed the Italian wolves to improve resolution of the remaining profiles, and the results placed the target NLD profiles together with Central European (German) and Baltic (Northern Polish) wolves (Figure 5). Based on the individual principal component (PC) scores, we also found the marginal profile 10A and the poor profile 11A to be somewhat divergent, with their scores on PC1 (the horizontal axis) closer to WDH-F1 and dog profiles.

For the analyses of 96 SNP profiles to improve resolution, we found that the two unresolved profiles had a higher amount of missing data also for this reduced set of SNPs. Whereas the nine high-quality samples were all missing less than 6% of the loci, 11A was missing 13.5% and 10A was missing 20.8%.

We subsequently repeated PCA with the consensus genotypes for samples 10A and 11A, using both a larger dataset (44K, Appendix 2A) and the discriminant 96 SNP markers (Appendix 2B). However, the consensus genotypes still had call rates below 90%, and despite the additional data filtering and analyses it was not possible to reliably resolve these two outlier profiles based on the available SNP data.

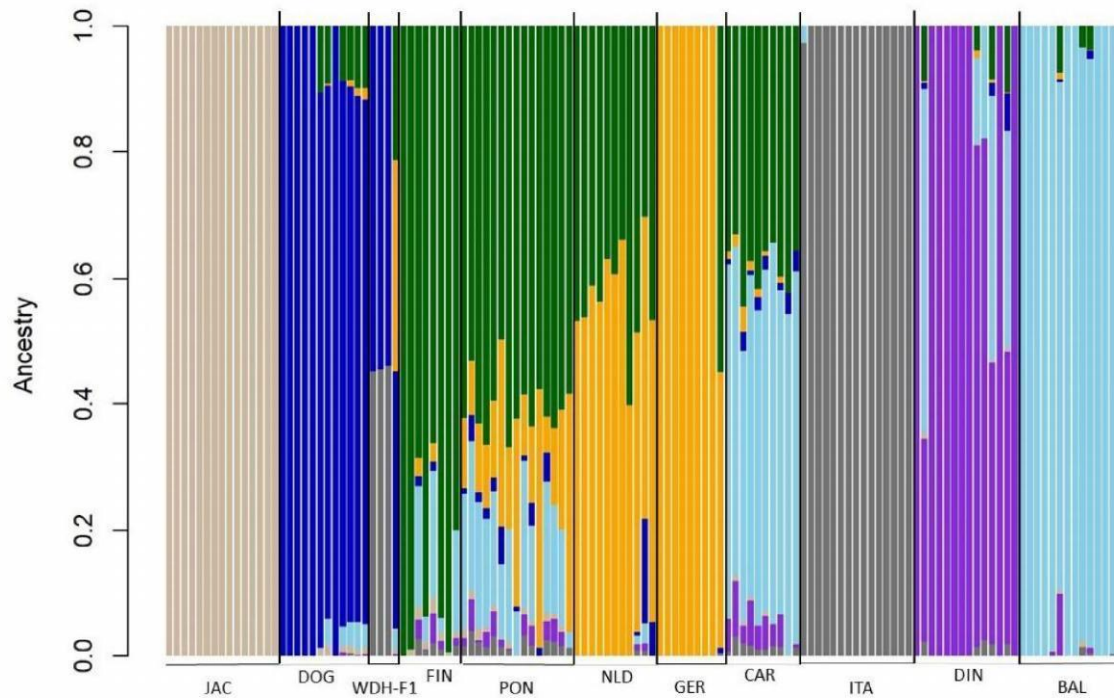


Figure 2. ADMIXTURE plot for $K = 7$ population clusters and $n = 127$ canid profiles, analysed with 68,583 single nucleotide polymorphism (SNP) loci. The population codes are: JAC (golden jackal), DOG (domestic dog), WDH-F1 (known first-generation wolf-dog hybrid), FIN (wolves from Finland), PON (wolves from Northern Poland), NLD (canid profiles from the Netherlands), GER (wolves from Germany), CAR (wolves from the Carpathians; Western Ukraine, Southern Poland, Slovakia), ITA (wolves from Italy), DIN (wolves from the Dinaric region; Slovenia and Croatia), BAL (Balkan region; Bulgaria and Greece).

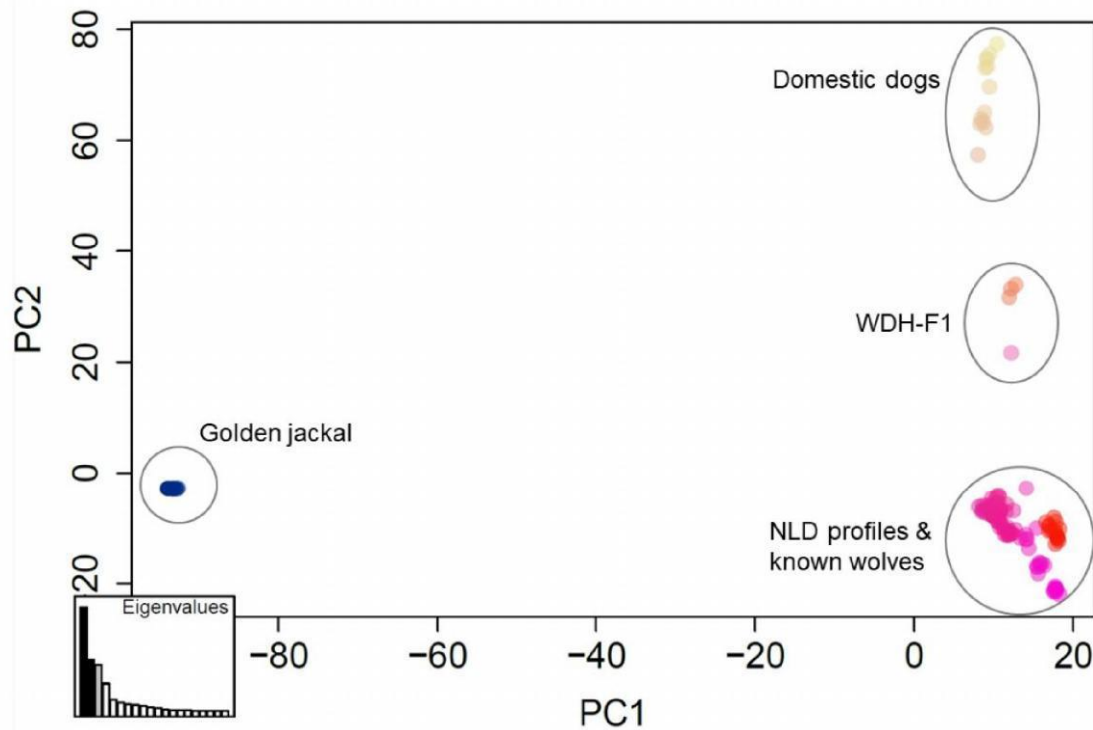


Figure 3. Principal component analysis of $n = 127$ canids with 68,583 SNP loci, using the colorplot-function. Distance and colours represent genetic diversity, and profiles with more differentiated colours and separated by a larger distance have more divergent genotypes. The results show four major groups: golden jackals, domestic dogs, first generation wolf-dog hybrids (WDH-F1) and one group of the NLD profiles and all known wolves. The first principal component (PC1) represents 14.8% of the variation, and the second (PC2) represents 7.7%.

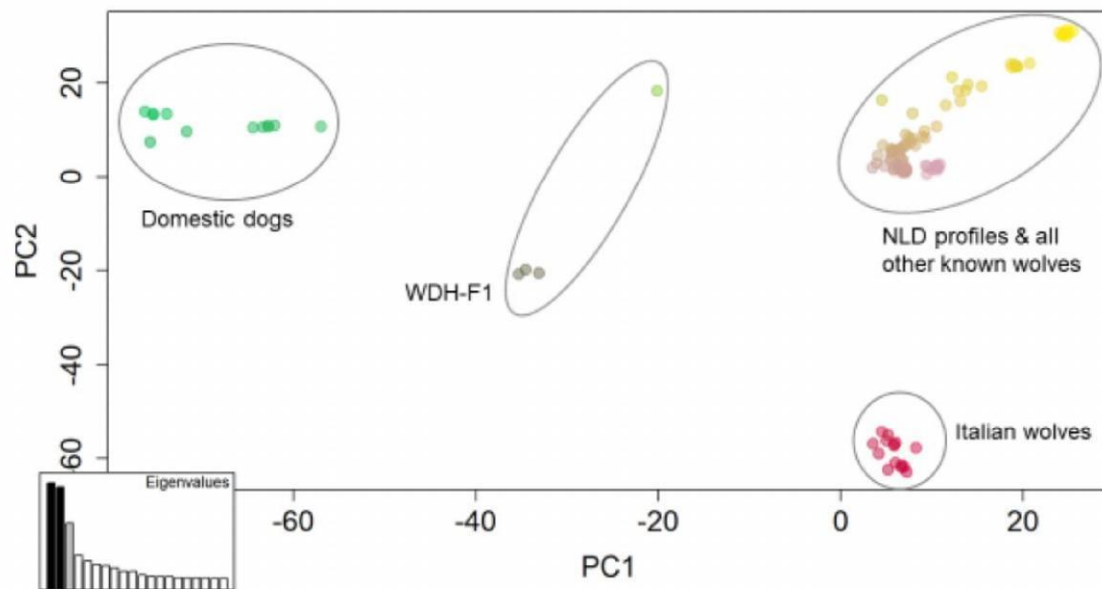


Figure 4. Principal component analysis of $n = 112$ canid samples with 68,583 SNP loci, using the colorplot-function. Distance and colours represent genetic diversity, and profiles with more differentiated colours and separated by a larger distance have more divergent genotypes. The results showed four major groups: domestic dogs, known first generation wolf-dog hybrids (WDH-F1), Italian wolves, and the NLD canids plus all other known wolves. Within the WDH-F1 group, the uppermost sample is a F1-hybrid between a dog and a Central European wolf, whereas the lower three are F1-hybrids between dogs and Italian wolves. The first principal component (PC1) represents 8.8% of the variation, and the second (PC2) represents 8.5%.

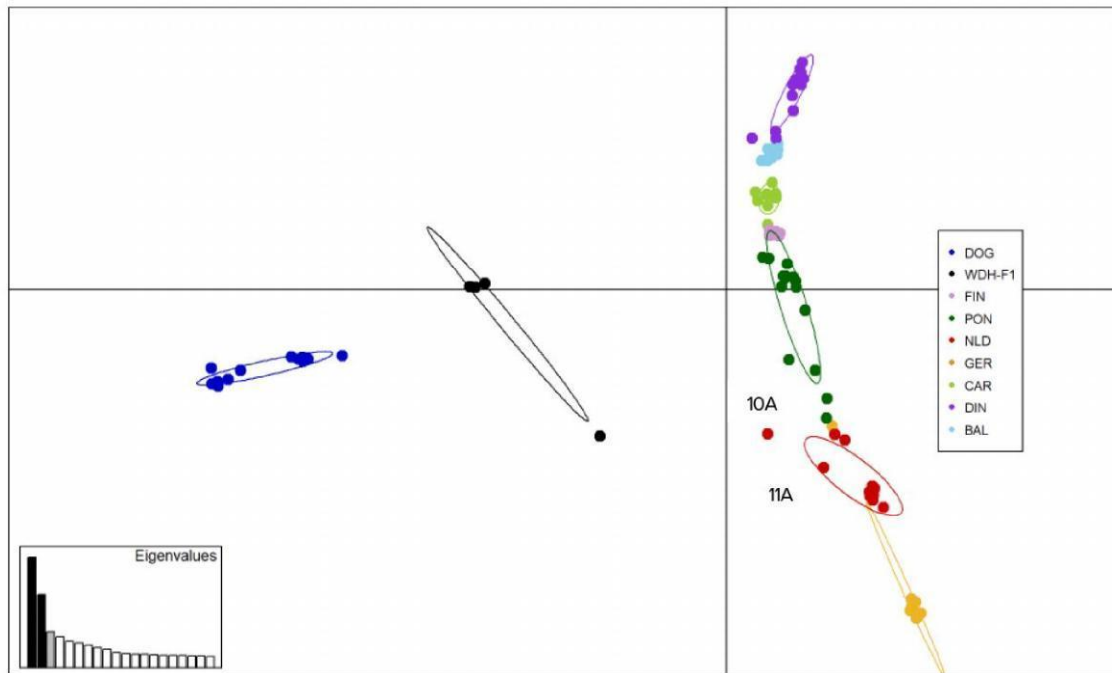


Figure 5. Principal component analysis of $n = 97$ canids with 68,583 SNP loci, showing results per sampling group. The first principal component (PC1) represents 10.4% of the variation, and the second (PC2) represents 6.8%. The population codes are: DOG (domestic dog), WDH-F1 (known first-generation wolf-dog hybrid), FIN (wolves from Finland), PON (wolves from Northern Poland), NLD (canid profiles from the Netherlands), GER (wolves from Germany), CAR (wolves from the Carpathians; Western Ukraine, Southern Poland, Slovakia), ITA (wolves from Italy), DIN (wolves from the Dinaric region; Slovenia and Croatia), BAL (Balkan region; Bulgaria and Greece). The results show NLD canid profiles grouped with wolves from Germany (Central European wolf population) and Northern Poland (Baltic wolf population). The marginal (11A) and poor (10A) profiles are placed closer to WDH-F1 canids, with 10A located furthest to the left among NLD canids.

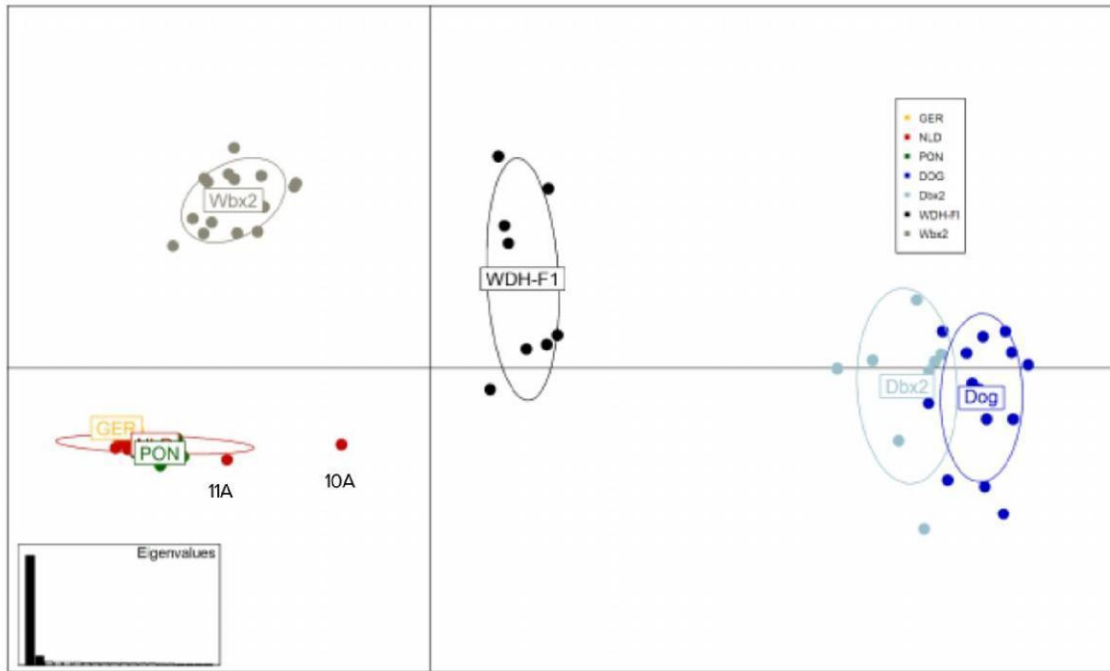


Figure 6. Principal component analysis of $n = 81$ canids with 96 SNP loci, showing results per sampling group. The first principal component (PC1) represents 61.4% of the variation, and the second (PC2) represents 5.3%. The population codes are: DOG (domestic dog), Dbx2 (second-generation backcross to dogs), WDH-F1 (known first-generation wolf-dog hybrid), Wbx2 (second-generation backcross to wolves), PON (wolves from Northern Poland), NLD (canid profiles from the Netherlands), GER (wolves from Germany). The results show NLD canid profiles grouped with wolves from Germany (Central European wolf population) and Northern Poland (Baltic wolf population). The marginal (11A) and poor (10A) profiles are placed closer to WDH-F1 canids.

Discussion

The genotyping results were very successful for nine of the eleven samples and provided a high-quality output. Two samples were less successful, but we nevertheless analysed all profiles and included them in assessments with 68K and 96 SNPs to resolve their ancestry as much as possible. The nine successful samples consistently grouped with earlier wolf profiles from Germany and Northern Poland, and our results are consistent with these samples originating from wolves. For the two remaining profiles with marginal (11A) and poor (10A) genotyping results, our findings were less conclusive.

Canid 11A shows signs of being an outlier but generally groups with Central European and Baltic wolves, suggesting these two populations are dominant in its ancestry. The analyses with 96 discriminant SNP markers also supported this result. This sample did not cluster in the direction of second-generation backcrosses to dogs (Dbx2), although it should be noted that the latter samples are not collected in northern Europe, but in the Dinaric-Balkan area. Canid 10A showed some similarities to the 11A profile but the results from this sample with extensive missing data also suggested the possible presence of dog ancestry. However, the Illumina array is developed for dogs with SNPs selected to be most representative for this group, whereby genetic diversity results are generally higher for dogs than for wolves (Harmoinen et al. 2021). This could have biased our results, and we therefore consider sample 10A to be inconclusive.

For the nine successful samples that clearly clustered with wolves, it nevertheless important to note that the regional reference profiles we had available were collected more than a decade ago. Given the dynamic situation for wolves in the study region, and the likelihood that these older German samples reflect the founder effect detected in this population (Jarausch et al. 2021), it is therefore not possible to give a more precise assessment of the NLD samples' ancestry. Although the Baltic and Central European populations are genetically distinct (Szewczyk et al. 2021), results from central Europe show long-distance dispersal connecting different populations (e.g., Hulva et al. 2017, Stiftung KORA 2023). However, our results are consistent with earlier findings of dispersal between the Central European and Baltic populations (Andersen et al. 2015) whereby one German profile shows high affinity to the wolves sampled in Poland.

Results for the high-quality genotypes did not provide any support for introgression from other canids, and the extensive testing done with wolves from across Europe (Harmoinen et al. 2021, Stronen et al. 2022) suggest that we should have been able to detect such introgression at least up to and including the second-backcross generation regardless of wolf ancestry from different European populations. However, for the two less successful genotypes (lower quality samples), additional analyses, ideally with new genomic profiles based on high-quality DNA samples, would be needed to make a more complete assessment. Hybridization is a concern for wolf populations across Europe (Salvatori et al. 2020) and hybridization and introgression require continuous monitoring, which is being aided by rapidly advancing methods in genotyping of DNA from noninvasive sources.

References

- Alexander, D.H. et al. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19, 1655–1664.
- Andersen, L.W. et al. (2015). Long-distance dispersal of a wolf, *Canis lupus*, in northwestern Europe. *Mammal Research* 60, 163–168.
- Chang, C.C. et al. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4.
- De Barba, M. et al. (2017). High-throughput microsatellite genotyping in ecology: improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Molecular Ecology Resources* 17, 492–507.
- Harmoinen, J. et al. (2021). Reliable wolf-dog hybrid detection in Europe using a reduced SNP panel developed for non-invasively collected samples. *BMC Genomics* 22, 473.
- Hindrikson, M. et al. (2017). Wolf population genetics in Europe: a systematic review, meta-analysis and suggestions for conservation and management. *Biological Reviews* 92, 1601–1629.
- Hulva, P. et al. (2018). Wolves at the crossroad: Fission-fusion range biogeography in the Western Carpathians and Central Europe. *Diversity and Distributions* 24, 179–192.
- Jarausch et al. (2021). How the west was won: genetic reconstruction of rapid wolf recolonization into Germany's anthropogenic landscapes. *Heredity* 127, 92–106.
- Jombart, T. (2008). Adegnet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405.
- Lelieveld, G. et al. (2016). The first proof of the recent presence of wolves in The Netherlands. *Lutra* 59: 23–31.
- Purcell, S. et al. (2007) PLINK: A Tool Set for Whole-Genome and Population-Based Linkage Analyses. *American Journal of Human Genetics*, 81. <https://www.cog-genomics.org/plink/>.
- R Core Team. (2022). R: A language and environment for statistical computing. In R Foundation for Statistical Computing. <https://www.R-project.org/>
- Schley, L. et al. (2021). First wolves in Luxembourg since 1893, originating from the Alpine and Central European populations. *Mammalia* 85, 193–197.
- Salvatori, V. et al. (2020). European agreements for nature conservation need to explicitly address wolf-dog hybridization. *Biological Conservation* 248, 108525.
- Segelbacher, G. et al. (2022). New developments in the field of genomic technologies and their relevance to conservation management. *Conservation Genetics* 23, 217–242.
- Stiftung KORA., (2023). Longest known dispersal of wolf in Europe. Available from <https://www.kora.ch/en/news/longest-known-dispersal-of-wolf-in-europe--551> [Accessed 14.04.2024].

Stronen, A. V. et al. (2013). North-south differentiation and a region of high diversity in European wolves. PLoS ONE 8(10), e76454.

Stronen, A. V. Et al. (2022). A reduced SNP panel to trace gene flow across southern European wolf populations and detect hybridization with other *Canis* taxa. Scientific Reports 12, 4195

Szewczyk, M. et al. (2019). Dynamic range expansion leads to establishment of a new, genetically distinct wolf population in Central Europe. Scientific Reports 9, 19003.

Szewczyk, M. et al. (2021). Genetic support for the current discrete conservation unit of the Central European wolf population. Wildlife Biology, wlb.00809 (2021). <https://doi.org/10.2981/wlb.00809>.

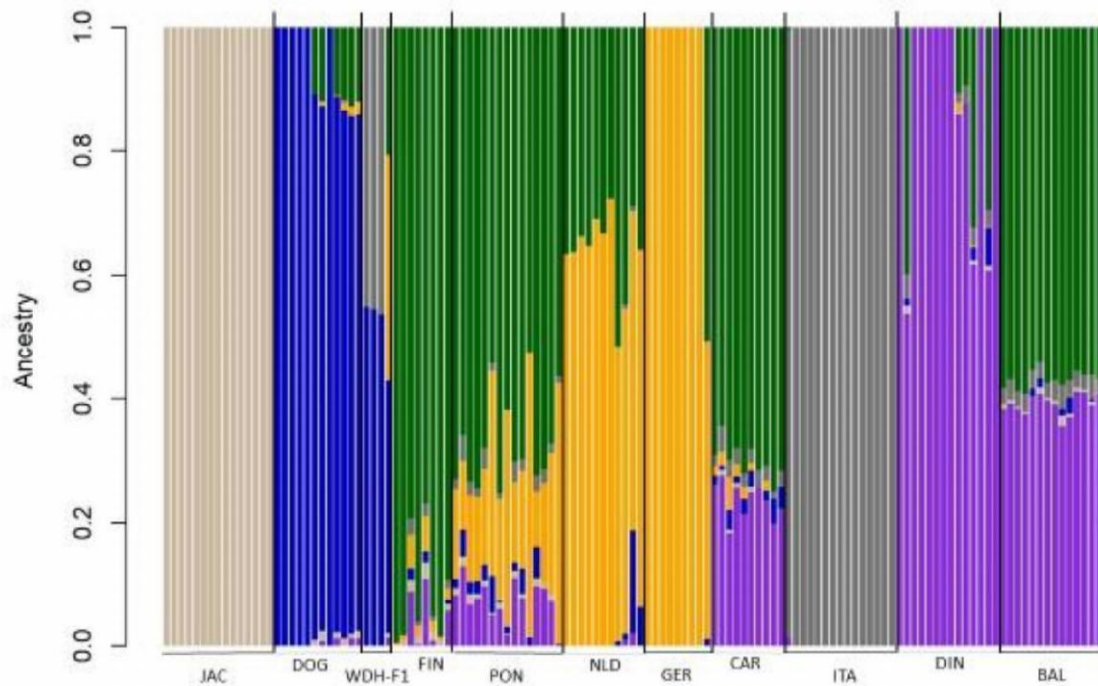
Vaysse, A. et al. (2011). Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. PLoS Genetics 7(10), e1002316.

von Thaden, A. et al. (2020). Using genomic data for applied wildlife genetics: a guideline for developing standardized SNP panels for genotyping of degraded samples. Molecular Ecology Resources 20, 662–680.

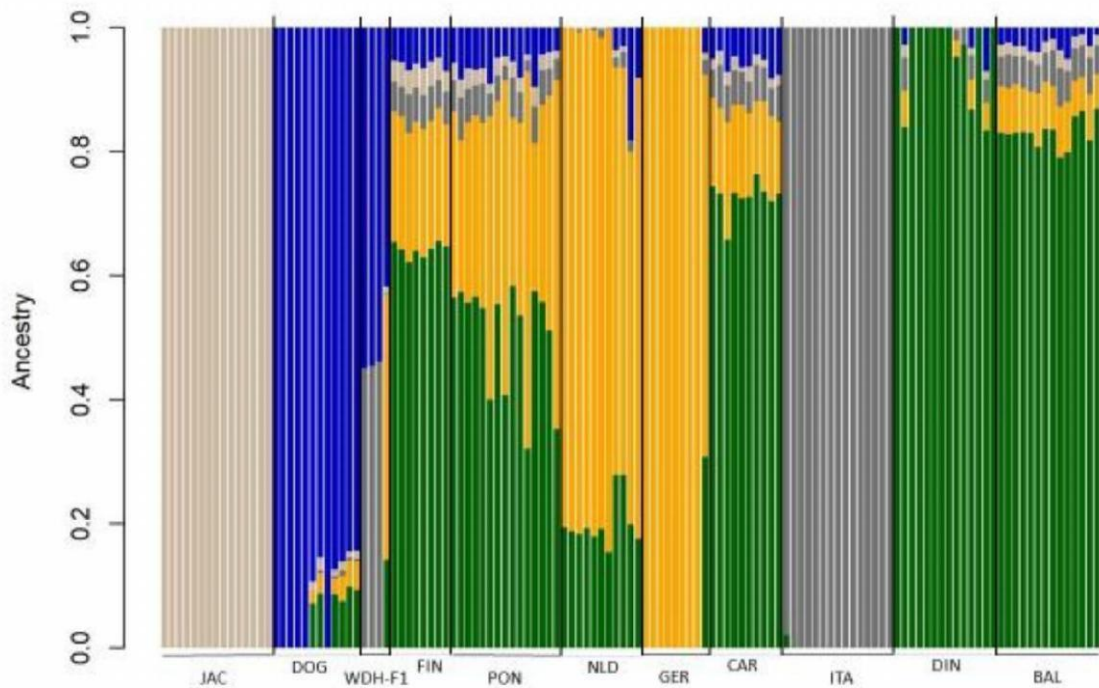
Acknowledgements

We acknowledge the use of unpublished reference profiles from Bulgarian jackals and wolves from Croatia, Germany, and Italy. Additionally, we thank our abovementioned colleagues at ISPRA, Italy, for conversion of genomic data from CanFam2 to CanFam3.1 reference coordinates.

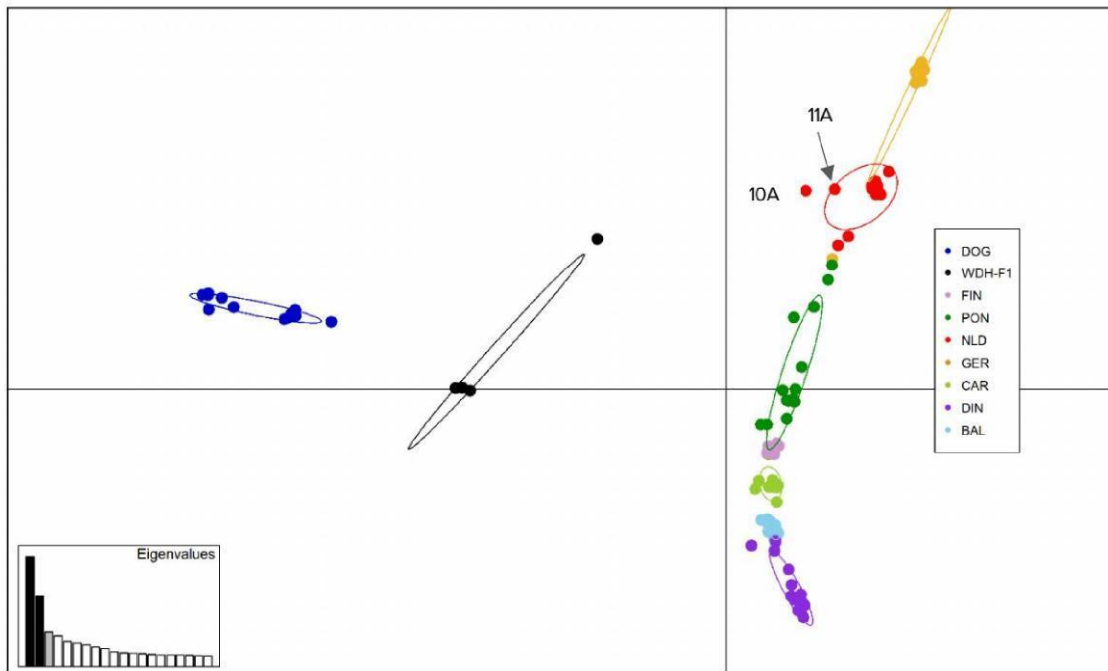
Appendices



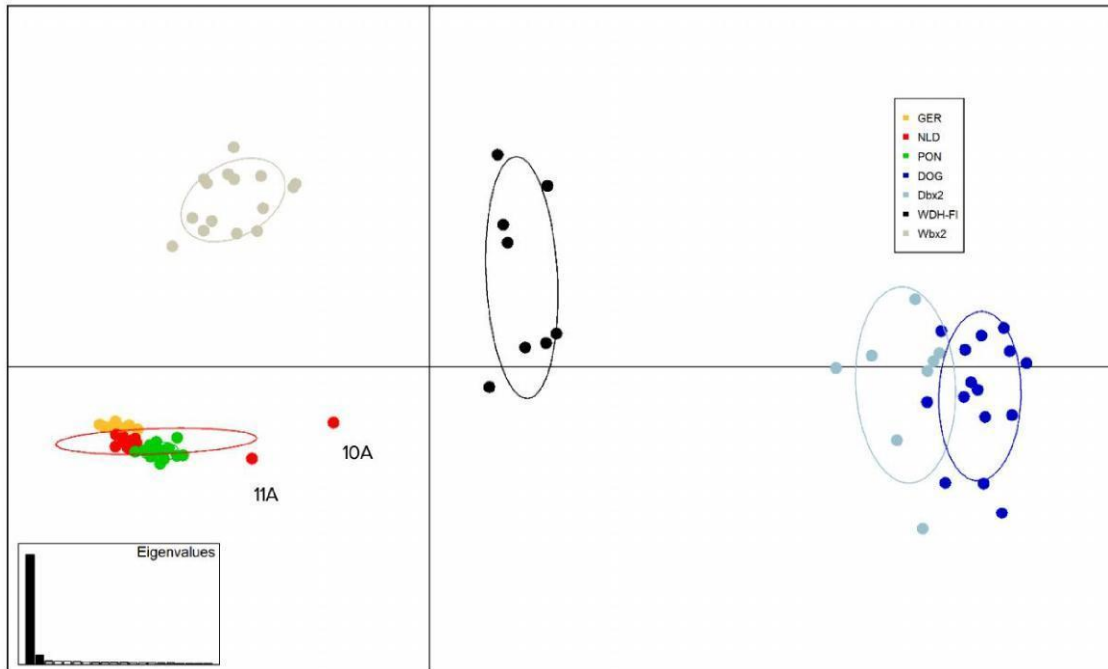
Appendix 1A. ADMIXTURE plot for $K = 6$ population clusters and $n = 127$ canid profiles, analysed with 68583 single nucleotide polymorphism (SNP) loci. The population codes are: JAC (golden jackal), DOG (domestic dog), WDH-F1 (known first-generation wolf-dog hybrid), FIN (wolves from Finland), PON (wolves from Northern Poland), NLD (canid profiles sampled in the Netherlands), GER (wolves from Germany), CAR (wolves from the Carpathians; western Ukraine, southern Poland, Slovakia), ITA (wolves from the Italy), DIN (wolves from the Dinaric region; Slovenia and Croatia), BAL (Balkan region; Bulgaria and Greece).



Appendix 1B. ADMIXTURE plot for $K = 5$ population clusters and $n = 127$ canid profiles, analysed with 68583 single nucleotide polymorphism (SNP) loci. The population codes are: JAC (golden jackal), DOG (domestic dog), WDH-F1 (known first-generation wolf-dog hybrid), FIN (wolves from Finland), PON (wolves from Northern Poland), NLD (canid profiles sampled in the Netherlands), GER (wolves from Germany), CAR (wolves from the Carpathians; western Ukraine, southern Poland, Slovakia), ITA (wolves from the Italy), DIN (wolves from the Dinaric region; Slovenia and Croatia), BAL (Balkan region; Bulgaria and Greece).



Appendix 2A. Principal component analysis of $n = 97$ canids with 44,112 SNP loci, showing results per sampling group. The first principal component (PC1) represents 10.6% of the variation and PC2 represents 6.8%. The population codes are: DOG (domestic dog), WDH-F1 (known first-generation wolf-dog hybrid), FIN (wolves from Finland), PON (wolves from Northern Poland), NLD (canid profiles from the Netherlands), GER (wolves from Germany), CAR (wolves from the Carpathians; Western Ukraine, Southern Poland, Slovakia), ITA (wolves from Italy), DIN (wolves from the Dinaric region; Slovenia and Croatia), BAL (Balkan region; Bulgaria and Greece). The results show NLD canid profiles grouped with wolves from Germany (Central European wolf population) and Northern Poland (Baltic wolf population). The marginal (11A) and poor (10A) profiles are placed closer to WDH-F1 canids, with 10A located furthest to the left among NLD canids.



Appendix 2B. Principal component analysis of $n = 81$ canids with 96 SNP loci, showing results per sampling group. The first principal component (PC1) represents 61.4% of the variation, and the second (PC2) represents 5.2%. The population codes are: DOG (domestic dog), Dbx2 (second-generation backcrosses to dogs), WDH-F1 (known first-generation wolf-dog hybrid), Wbx2 (second-generation backcrosses to wolves), PON (wolves from Northern Poland), NLD (canid profiles from the Netherlands), GER (wolves from Germany). The results show NLD canid profiles grouped with wolves from Germany (Central European wolf population) and Northern Poland (Baltic wolf population). The marginal (11A) and poor (10A) profiles are placed closer to WDH-F1 canids.