



Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

A reduced SNP panel to trace gene flow across southern European wolf populations and detect hybridization with other *Canis* taxa

Stronen, Astrid Vik; Mattucci, Federica; Fabbri, Elena; Galaverni, Marco; Cocchiararo, Berardino; Nowak, Carsten; Godinho, Raquel; Ruiz-González, Aritz; Kusak, Josip; Skrbinšek, Tomaž; Randi, Ettore; Vlasseva, Albena; Mucci, Nadia; Caniglia, Romolo

Published in:
Scientific Reports

DOI (link to publication from Publisher):
[10.1038/s41598-022-08132-0](https://doi.org/10.1038/s41598-022-08132-0)

Creative Commons License
CC BY 4.0

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Stronen, A. V., Mattucci, F., Fabbri, E., Galaverni, M., Cocchiararo, B., Nowak, C., Godinho, R., Ruiz-González, A., Kusak, J., Skrbinšek, T., Randi, E., Vlasseva, A., Mucci, N., & Caniglia, R. (2022). A reduced SNP panel to trace gene flow across southern European wolf populations and detect hybridization with other *Canis* taxa. *Scientific Reports*, 12(1), Article 4195. <https://doi.org/10.1038/s41598-022-08132-0>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -



OPEN

A reduced SNP panel to trace gene flow across southern European wolf populations and detect hybridization with other *Canis* taxa

Astrid Vik Stronen^{1,2,3,13}✉, Federica Mattucci^{4,13}, Elena Fabbri⁴, Marco Galaverni⁵, Berardino Cocchiararo^{6,7}, Carsten Nowak^{6,7}, Raquel Godinho^{8,9}, Aritz Ruiz-González¹⁰, Josip Kusak¹¹, Tomaž Skrbinšek¹, Ettore Randi³, Albena Vlasheva¹², Nadia Mucci⁴ & Romolo Caniglia^{4,13}

Intra- and inter-specific gene flow are natural evolutionary processes. However, human-induced hybridization is a global conservation concern across taxa, and the development of discriminant genetic markers to differentiate among gene flow processes is essential. Wolves (*Canis lupus*) are affected by hybridization, particularly in southern Europe, where ongoing recolonization of historic ranges is augmenting gene flow among divergent populations. Our aim was to provide diagnostic canid markers focused on the long-divergent Iberian, Italian and Dinaric wolf populations, based on existing genomic resources. We used 158 canid samples to select a panel of highly informative single nucleotide polymorphisms (SNPs) to (i) distinguish wolves in the three regions from domestic dogs (*C. l. familiaris*) and golden jackals (*C. aureus*), and (ii) identify their first two hybrid generations. The resulting 192 SNPs correctly identified the five canid groups, all simulated first-generation (F1) hybrids ($0.482 \leq Q_i \leq 0.512$ between their respective parental groups) and all first backcross (BC1) individuals ($0.723 \leq Q_i \leq 0.827$ to parental groups). An assay design and test with invasive and non-invasive canid samples performed successfully for 178 SNPs. By separating natural population admixture from inter-specific hybridization, our reduced panel can help advance evolutionary research, monitoring, and timely conservation management.

During the past centuries, long-term isolation and demographic declines in southern European peninsular populations of wolves (*Canis lupus*) caused considerable genetic drift that is now apparent in their genome-wide profiles^{1–5}. However, the ongoing spatial and demographic recovery of most European wolf populations, mainly due to legal protection and increased prey and habitat availability^{6,7}, has prompted a gradual process of natural contact with dispersal and gene flow among discrete populations^{8–10}. This tendency is represented, for instance, by the wolf pack in Lessinia Regional Park in northern Italy, which was established in 2012 by a female from Italy

¹Department of Biology, Biotechnical Faculty, University of Ljubljana, Večna pot 111, 1000 Ljubljana, Slovenia. ²Department of Biotechnology and Life Sciences, Insubria University, via J. H. Dunant 3, 21100 Varese, Italy. ³Department of Chemistry and Bioscience, Aalborg University, Fredrik Bajers Vej 7H, 9220 Aalborg, Denmark. ⁴Unit for Conservation Genetics (BIO-CGE), Italian Institute for Environmental Protection and Research (ISPRA), Via Cà Fornacetta 9, 40064 Ozzano dell' Emilia, Bologna, Italy. ⁵Science Unit, WWF Italia, Via Po 25c, 00198 Rome, Italy. ⁶Wildlife Genetics Center, Senckenberg Research Institute and Natural History Museum Frankfurt, Clamecystrasse 12, 63571 Gelnhausen, Germany. ⁷LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Senckenberganlage 25, 60325 Frankfurt am Main, Germany. ⁸CIBIO/InBio – Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal. ⁹Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal. ¹⁰Department of Zoology and Animal Cell Biology, University of the Basque Country (UPV/EHU), Paseo de la Universidad, 7, 01006 Vitoria-Gasteiz, Spain. ¹¹Department of Biology, Faculty of Veterinary Medicine, University of Zagreb, Heinzelova 55, 10000 Zagreb, Croatia. ¹²Institute of Biodiversity and Ecosystem Research, Bulgarian Academy of Sciences, 2 Gagarin Street, 1113 Sofia, Bulgaria. ¹³These authors contributed equally: Astrid Vik Stronen, Federica Mattucci and Romolo Caniglia. ✉email: astrid.stronen@gmail.com

and an immigrant male from the Dinaric Mountains in southwestern Slovenia near the border with Croatia^{11,12}, followed by back-migration of their descendants to Slovenia detected during the national genetic wolf monitoring program¹³. Further examples are provided by the detection of Italian or Dinaric individuals re-sampled in the Iberian Peninsula¹⁴ and in southern Germany¹⁵.

However, despite growing numbers, several European wolf populations are now threatened by anthropogenic hybridization with free ranging domestic dogs (henceforth dogs, *C. l. familiaris*)^{14,16,17} and possibly with golden jackals (henceforth jackals, *C. aureus*)^{18,19}. Wolf and jackal ranges are showing an increasing degree of overlap in several countries^{20–22}, and hybridization between jackals and dogs has been confirmed²³. The ability to rapidly and accurately discriminate natural population admixture from inter-specific hybridization is therefore especially important in areas where active interventions may be needed to mitigate anthropogenic hybridization, while allowing or facilitating inter-population gene flow^{17,24–26}.

Earlier studies, prior to the development of genomic tools, provided key insights toward understanding population processes and wolf-dog hybridization based on traditional microsatellite loci (e.g., Ref. 27,28). However, although multi-allelic microsatellites have a higher per-locus information content than bi-allelic single nucleotide polymorphism (SNP) markers and may have benefits for applications involving short spatiotemporal scales²⁹, their use requires the calibration of results among laboratories when different datasets need to be merged. As an alternative, small panels of SNP markers, suitable also for non-invasive samples, have been developed for various species to address evolutionary and conservation questions (e.g., Ref. 30–34). Although a larger number of SNPs are needed for tasks such as individual identification, SNPs permit high-throughput analyses where results are directly comparable among laboratories without the need for calibration—a major advantage when timely results are needed for cross-boundary conservation management decisions^{32,33}.

von Thaden et al.³² performed comparative analyses for microsatellite and SNP data in wolves and European wildcats (*Felis silvestris*) where they noted that overall amplification rates were higher for SNPs than for microsatellites. However, new high-throughput sequencing approaches are providing important improvements also for microsatellite markers, as demonstrated for brown bears (*Ursus arctos*)³⁵. These recent developments can permit the integration of different marker types and maximize the information obtained from non-invasive monitoring efforts. Because SNPs also permit insights into selective processes^{36,37}, such markers can offer additional information for evolutionary research and may serve as indicators of responses to environmental change³⁸.

Genomic methods are thus essential tools in wildlife conservation and management, and provide opportunities to track allele frequency changes in adaptive and deleterious loci for improved population monitoring (reviewed in Ref. 39). SNP genotyping has also offered major advances for non-model species in fields such as biogeography⁴⁰ and conservation genomics⁴¹. Modern genomic platforms provide extensive and cost-effective screening of thousands of SNPs, from which small panels of highly ancestry-informative markers (AIMs) can be selected depending on the specific research or monitoring questions at hand. These panels can be applied in extensive monitoring of endangered taxa or priority management units through microfluidic or quantitative PCR techniques⁴², which allow cost-effective genotyping of dozens of samples and markers at a time, including low quality DNA from non-invasively collected materials^{32–34,43}, where multiple replicates can be performed when needed (see details in Ref. 32,33).

A first set of ancestry-informative markers (AIMs) to accurately detect wolf \times dog hybrids from non-invasively collected samples was recently identified by Harmoinen et al.¹⁵ and performed well for wolf populations across Europe. Their results confirmed the high genetic divergence among some southern European wolf populations, caused by protracted geographical isolation, demonstrated in earlier studies based on various genetic markers (e.g., Ref. 4,5,14,16,44). Strong wolf population substructure and the increasing gene flow among wolf populations could therefore make individual assignment more difficult^{3,17,27,44}. The 93-SNP panel proposed by Harmoinen et al.¹⁵ performs well for the distinct populations of Italy, Iberia, and the Dinaric region in detecting wolf \times dog hybrids. The identification of additional highly informative SNPs can nevertheless help to distinguish natural admixture between wolves from these genetically divergent populations (e.g., Ref. 12) from anthropogenic hybridization with dogs, a serious and increasing concern in all three regions^{25,45,46}, and possible inter-specific hybridization with jackals (e.g., Ref. 23). Building on the work of Harmoinen et al.¹⁵, we therefore selected an additional panel of highly informative SNPs. Our objectives were to: (i) identify dogs, jackals, and wolves from the Italian, Dinaric and Iberian populations in southern Europe and (ii) distinguish inter- and intra-specific first-generation hybrids (F1) and the first generation of backcrosses (BC1) with parental groups, to help prioritize conservation actions, especially in areas with high levels of anthropogenic hybridization^{17,24,47,48}. Two additional objectives were to (iii) assess the SNP panel for the presence of loci under possible selection, and (iv) evaluate whether the SNP panel exhibits any clear bias in reconstructing wolf population divergence patterns, compared to existing results based on genome-wide profiles. These additional two assessments were performed to obtain a better understanding of the identified SNPs and their representation of the investigated wolf populations.

Materials and methods

Dataset building and filtering procedures. We combined existing SNP genotypes from wolves, dogs, and jackals typed on the Illumina CanineHD BeadChip including over 170,000 SNPs^{2,49} with unpublished data from Europe. European wolves included individuals from population ranges broadly defined in previous SNP analyses^{1–3} as Dinaric-Balkan (n = 112), Iberia (n = 25), and Italy (n = 77). Dog profiles included village dogs from Italy (n = 30), Croatia (n = 11) and Bulgaria (n = 2), considered more likely to hybridize with wolves^{17,27,49} or jackals²³, and nine common dog breeds whose size is compatible with potential crossing with wolves (Bernese Mountain Dog—BMD, Border Collie—BoC, English Setter—ES, Gordon Setter—GoS, Greyhound—GRE, German Shepherd—GSh, Labrador Retriever—LRe, Rottweiler—Rtw, and Weimaraner—Wei), genotyped by the LUPA project^{50,51}. Jackal samples originated primarily from Bulgaria (n = 18), where possible hybridization

between wolves and jackals have been reported¹⁹ and included samples from Croatia ($n = 3$) where hybrids with dogs have been documented²³.

Canine genotypes mapped to the CanFam2 version of the dog genome were converted to the CanFam3 version with the UCSC liftOver tool (<http://genome.ucsc.edu/cgi-bin/hgLiftOver>) and imported into the SNP&Variant Suite 8.0.1 (hereafter SVS, Golden Helix Inc., Bozeman, MT). Nine of the SNPs included had previously been associated with three phenotypic traits (Table S1): black-coloured coat, white-coloured nails, and the presence of dewclaws, vestigial first toes on the hind legs^{49,52}. In wolves, these traits have been linked to wolf \times dog hybridization^{49,53,54}, although it is important to note that they can also be found in individuals that do not exhibit any other sign of dog introgression⁴⁹. These SNPs are henceforth described as phenotypic loci.

We performed an initial quality control of the dataset, by first applying a filter based on individual genotyping success and retaining profiles with $\geq 90\%$ success rate. Next, we screened the data with a per-SNP genotyping threshold of $> 90\%$. We then pruned the dataset for loci in strong linkage disequilibrium (LD) in SVS using the default settings: a LD-threshold of 0.5 and sliding windows of 50 SNPs. None of the nine phenotypic loci passed the initial filtering process, owing to missing values in the SNP data set. However, given their potential information value, we nevertheless selected three SNPs, one for each trait, for inclusion in the candidate panel (Table S1; selected loci in bold font). We next performed a preliminary assessment of all the canine profiles in ADMIXTURE 1.23⁵⁵ with $K = 3$ population clusters (roughly corresponding to jackals, dogs and wolves) to determine overall population structure. To exclude possible hybrids, we ran pairwise comparisons, each time with $K = 2$, for each of the three wolf populations and dogs, then wolves and jackals, and finally dogs and jackals. We retained only wolves and jackals assigned with $q_i \geq 0.95$ to their respective population clusters, and dogs assigned with $q_i \geq 0.90$, considering their higher within-group variability.

Marker selection. AIMS to identify admixture between dogs, jackals and the three regionally distinct wolf populations were selected by performing pairwise F_{ST} comparisons among the five canid groups and choosing loci with $F_{ST} \geq 0.90$, evenly distributed across the 38 autosomal chromosomes. We then selected the 45 best-performing loci from each pairwise comparison for an initial panel. From this list, we selected a reduced panel by prioritizing loci that contributed to differentiation for (a) wolves vs. jackals, (b) F_{ST} values across pairwise comparisons for the three regional wolf populations, (c) Iberian vs. Italian wolves, (d) Italian vs. Dinaric wolves, (e) dogs vs. Italian wolves, (f) dogs vs. Iberian wolves, (g) dogs vs. Dinaric wolves, and (h) wolves vs. dogs across all three regions.

Population structure analyses. We used the resulting SNPs in explorative multivariate analyses by performing a Discriminant Analysis of Principal Components (DAPC) in R 3.6.6⁵⁶ with the ADEGENET package 2.1.3⁵⁷ to detect patterns of genetic differentiation among groups and individuals⁵⁸. The “*find.clusters*” function was initially used to determine the best-supported number of genetic clusters using the Bayesian Information Criterion (BIC) running successive numbers (1–10) of K -means clusters of the individuals and the “*table.value*” function was used to graphically visualize the corresponding best clustering of the individuals. The optimal number of principal components (PCs) was identified using the “*optim.a.score*” function and the Discriminant Analysis (DA) was then run on the retained principal components using the “*dapc*” function. Finally, after selecting the optimal number of eigenvalues for the DA analysis, the DAPC results were graphically visualized with the “*scatter.dapc*” function and the assignment probability of individuals to each cluster were calculated and plotted using the “*assignplot*” function⁵⁹.

We used the same dataset to perform Bayesian clustering analyses in STRUCTURE 2.3.4⁶⁰. Five independent runs were performed for increasing values of K (the number of genetic clusters) from 1 to 10 using 500,000 Markov chain Monte Carlo (MCMC) iterations, after a burnin of 50,000 iterations, assuming no prior information (option *Usepopinfo* not activated), and choosing the “*Admixture*” and “*Independent Allele Frequency*” models⁵⁹. We used the highest rate of increase in the posterior probability $\text{LnP}(K)$ between consecutive values of K to estimate the most likely K -value, and then assessed the average (Q_i) and individual (q_i) proportions of membership in each cluster⁶¹. The software CLUMPP 1.1.1⁶² was used to concatenate the data from the five independent runs for each K value, and DISTRUCT 1.1⁶³ to graphically display the results.

Subsequently, we used HYBRIDLAB 1.0⁶⁴ to simulate jackal \times wolf, wolf \times dog, and wolf \times wolf hybrids. The five canid groups were used as reference populations, and we simulated pairwise crosses to generate 10 F1-hybrids and 10 BC1-individuals with the wild parental populations for each comparison (Table S2). The resulting profiles were analyzed in STRUCTURE as described above to evaluate the efficiency of the selected SNP panel in detecting hybrids. Finally, we ran ADEGENET (PC Analysis performed with the “*dudi.pca*” function) and STRUCTURE as described above with the 192 SNPs and a larger dataset of 191 individuals. This dataset included the five reference groups plus additional canids categorized as admixed or not admixed in earlier studies^{23,26,28,46,49,65} to test the discriminating power of the selected loci when applied to empirical data (comprising $n = 10$ Italian and $n = 9$ Dinaric wolves, $n = 4$ Italian wolf \times dog, $n = 4$ Dinaric wolf \times dog, $n = 4$ Iberian wolf \times dog, and $n = 2$ jackal \times dog hybrid individuals; admixed canids had been excluded from the marker selection step). To check the consistency of the inferred genetic structure, all clustering analyses were run again in ADMIXTURE to reassign each sample to its group of origin, assuming K -values from 1 to 10. The most likely number of clusters was identified based on the lowest cross-validation error⁵⁵.

Genetic variability analyses. The proportions of polymorphic (PL) and monomorphic (ML) loci observed and effective allele numbers (A_O and A_E), observed and unbiased expected heterozygosity (H_O and uH_E), numbers of private alleles (N_P), Probabilities of Identity among unrelated (PID) and among full sib (PID_{sibs}) individuals⁶⁶, F_{ST} values, number of migrants (N_M), and analysis of molecular variance (AMOVA) were

computed using GENALEX 6.502⁶⁷. The polymorphic information content (PIC) and the mean proportion of successfully genotyped loci (GL) were computed using CERVUS 3.0.3⁶⁸. Allelic richness (A_R), which corrects the observed number of alleles for differences in sample sizes, was computed with FSTAT 2.9.3.2⁶⁹, and values of the inbreeding coefficient Wright's F_{IS} were calculated using GENETIX 4.03⁷⁰.

Candidate loci under selection and population divergence times. We evaluated the candidate SNPs for the possible presence of loci under selection (see full details in Supplemental Note S1). Finally, we used the candidate SNP loci to model estimated divergence times (in generations) among populations using various demographic scenarios (Supplemental Note S2).

Fluidigm assay design, development and application. To verify the applicability of the resulting SNPs (see “Results”) in future genetic monitoring and conservation management, including their performance for non-invasively collected samples, we tested the SNPs on microfluidic 96 (samples) \times 96 (loci) Dynamic Arrays™ (Fluidigm Corp., South San Francisco, USA), following Harmoinen et al.¹⁵. The SNP panel consisted of a) 84 SNPtype™ genotyping assays previously developed by Harmoinen et al.¹⁵ (Panel 1 in Table S5) and b) 108 newly designed SNPtype™ assays of the selected SNPs from this study (Panel 2 and Panel 3 in Table S5). We applied all 192 SNPtype™ genotyping assays to reference individuals previously genotyped on the Illumina CanineHD BeadChip, encompassing A) 33 DNA samples from muscular tissues, including 4 dogs, 7 jackals, 7 Dinaric, 7 Iberian, and 7 Italian wolves, plus one red fox (*Vulpes vulpes*) to check for other possible cross-species amplifications. Additionally, we included B) 60 non-invasive DNA samples obtained from faecal swabs⁷¹ including 11 Italian wolves, 4 dogs, and 45 hybrids previously identified using traditional microsatellites²⁶. The samples in B) were collected during ongoing non-invasive genetic monitoring of the Italian wolf population. Finally, three blank controls were included. Invasive and non-invasive samples were genotyped 2–3 times and replicates were used to reconstruct consensus genotypes and calculate amplification success (henceforth AS) and error (Allelic Dropout, ADO and False Alleles, FA) rates in GIMLET 1.3.3⁷². Reliable Fluidigm consensus genotypes were then used to perform an additional Principal Coordinates Analysis (PCoA) using GENALEX 6.502⁶⁷.

Results

Marker selection. Our initial filtering of the data retained 98,004 unlinked SNPs (including 3 phenotypic SNPs) available for a dataset of 158 canids, namely 52 dogs, 21 jackals (18 from Bulgaria and 3 from Croatia), and 30 Italian, 30 Dinaric, and 25 Iberian wolves. Following the selection process for highly discriminant loci, we retained an initial panel of 198 candidate SNPs (Table S3). These included 24 loci fixed in jackals, 13 fixed between dogs and Italian wolves, 4 fixed between dogs and Iberian wolves, 34 fixed between Italian and Iberian wolves and 7 fixed between Italian and Dinaric wolf populations (Table S3). From the preliminary panel of 198 SNPs, we then selected a reduced panel of 105 candidate SNPs (Table S4), comprising: (a) 15 loci discriminant for wolves and jackals (14 of which were fixed in jackals), (b) 37 SNPs common across F_{ST} tests for regional wolf populations, (c) 23 SNPs fixed between Italian and Iberian wolves, (d) 2 SNPs fixed between Italian and Dinaric wolves, (e) 13 SNPs fixed between dogs and Italian wolves, (f) 2 SNPs fixed between dogs and Iberian wolves, (g) 3 highly discriminant SNPs between dogs and Dinaric wolves and (h) 10 SNPs highly discriminant for wolf-dog tests in all the three geographic regions. These 105 SNPs were added to the three phenotypic loci (black-coloured coat, white-coloured nails, and the presence of dewclaws; Table S1) and to 84 of the 93 unlinked SNPs selected by Harmoinen et al.¹⁵ based on F_{ST} values between genome-wide profiles of European wolves and 58 dog breeds. Nine of the 93 SNPs from Harmoinen et al.¹⁵ did not pass the filtering process in our study and were removed. The total dataset for further analyses therefore comprised 158 canids and 192 SNPs highly informative across groups (Table S5).

Population structure. The K -means clustering analysis identified the best-supported number of genetic clusters at $K=5$, which showed the lowest BIC value (Fig. 1a). At $K=5$ all individuals clustered in their five original sample groups, which clearly corresponded to the five inferred genetic clusters (Fig. 1b). Discriminant analysis was performed using the five clusters defined by the K -means procedure, retaining the first 100 PCs during the data transformation step, which explained more than 99.5% of the total variance in the data, and four principal eigenvalues. The DAPC clearly identified the five original canid groups with the first component, explaining most of the genetic variability (c. 66%) and distinguished dogs from all of the wild canids, which were visibly separated along the second axis, describing c. 19% of the genetic diversity (Fig. 1c). Furthermore, the discriminant functions based on DAPC correctly assigned all individuals to their *a-priori* genetic clusters defined by the K -means analyses with individual membership probabilities always > 0.999 (Fig. 1d).

Multivariate analyses were strongly supported by the Bayesian clustering procedures implemented in STRUC-TURE that showed increasing rates in the estimated posterior probability $\text{LnP}(K)$ of the clusters until $K=5$ (Fig. 2a). For all K -values we observed very low standard deviations among different runs of the same K , with an average variation of only 0.00017 (± 0.00014 SD) in individual coefficient values (q_i) among runs (Fig. 2a). At $K=2$ (Fig. 2b), corresponding to the first main increase in $\text{LnP}(K)$, dogs (mean estimated membership of population to the assigned cluster $Q_1=0.999$) were clearly separated from the other four taxa (mean $Q_2=0.999$). At $K=3$, dogs ($Q_1=0.998$) clustered separately from Italian wolves ($Q_2=1.000$) and from the other wild canids ($Q_3=0.990$). At $K=4$, dogs were included in cluster 1 ($Q_1=0.997$), Italian wolves in cluster 2 ($Q_2=0.999$), jackals in cluster 3 ($Q_3=0.999$), and Dinaric wolves together with Iberian wolves in cluster 4 ($Q_4=0.994$). At $K=5$ (Fig. 2b and Table 1), corresponding to the optimal number of genetic clusters and consistent with the phylogenetic and geographic subdivision of the samples, all five canid groups were correctly allocated in their own categories with dogs assigned to cluster 1 ($Q_1=0.996$), Italian wolves to cluster 2 ($Q_2=0.999$), jackals to cluster 3 ($Q_3=0.999$),

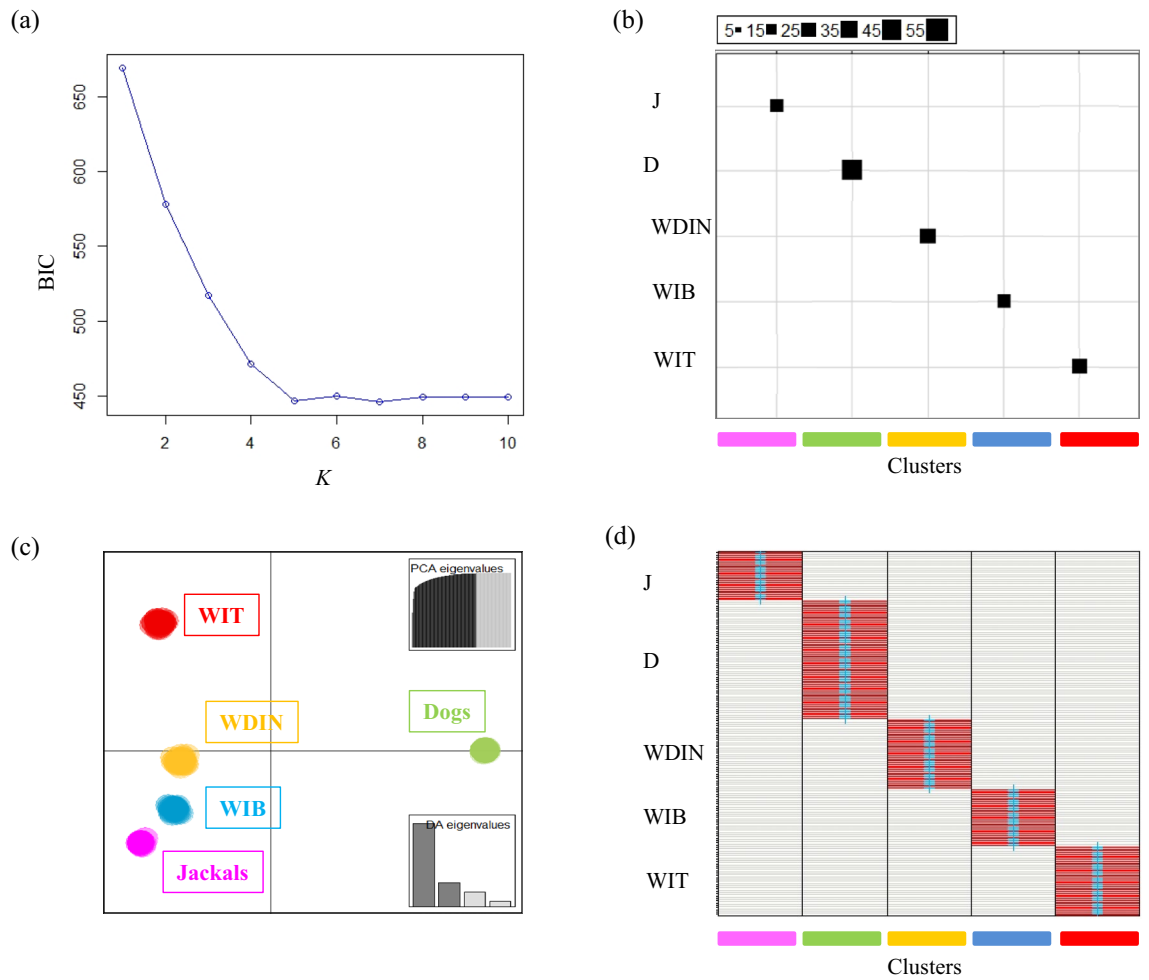


Figure 1. Explorative multivariate analyses performed with ADEGENET 2.1.3⁵⁷ on the 158 canid individuals typed at 192 SNPs. **(a)** Bayesian Information Criterion (BIC) run with successive numbers (from 1 to 10) of *K*-means clusters. The function “*table.value*” was used to graphically visualize the corresponding best clustering of the individuals. **(b)** Plot showing that the analysed individuals clustered in their five original sample groups which corresponded to the five inferred genetic clusters (J: Jackals, purple bar; D: domestic dogs, green bar; WDIN: Dinaric wolves, yellow bar; WIB: Iberian wolves, blue bar; WIT: Italian wolves, red bar). **(c)** Discriminant analysis of principal component (DAPC) scatterplot showing the genetic distribution among dogs (left side of PC-I) and wild canids (right side of PC-I, which explained 66% of the total genetic variability. PC-II (explaining 19% of the total genetic variability) shows the distinctions among the Italian wolves WIT (top of the graphic) and the other wolf (WIB: Iberian wolves, WDIN: Dinaric wolves) and jackal populations. See also the DA and PCA eigenvalue histogram inserted on the lower and upper right side, respectively. **(d)** Assignment plot based on DAPC: all individuals were assigned to their *a-priori* genetic clusters defined by the *K*-means analyses with individual membership probabilities >0.999.

Dinaric wolves to cluster 4 ($Q_4 = 0.995$) and Iberian wolves to cluster 5 ($Q_5 = 0.998$). For $K > 5$, the $\text{LnP}(K)$ reached a plateau and no further interpretable substructures were observed in the data (Fig. 2a). Notably, the substructure identified at the optimal genetic subdivision ($K = 5$) with the 192 SNPs was highly coherent with results obtained through an additional STRUCTURE analysis performed with the entire panel of 98,004 unlinked SNPs (Fig. S1a).

Bayesian clustering procedures run with the five canid groups plus their simulated first two generations of hybrids clearly confirmed that $K = 5$ corresponded to the optimal number of genetic clusters in the dataset with all five parental groups correctly assigned to their own cluster with $0.989 \leq Q_i \leq 0.998$ (Fig. 3 and Table 2). As expected, all simulated F1 hybrids had intermediate Q_i assignment values ($0.482 \leq Q_i \leq 0.512$) between their respective parental groups (Fig. 3 and Table 2) and all BC1 showed Q_i assignment values to the parental wild groups they were backcrossed with ranging from 0.723 to 0.827 (Fig. 3 and Table 2). Finally, both multivariate discriminant analyses and Bayesian clustering procedures that were run with the additional 33 canid genotypes clearly demonstrated the high discriminating power of the identified loci when applied to empirical data. These correctly and unambiguously identified all wolf (WDIN $q_i \geq 0.992$ and WIT $q_i \geq 0.999$), all known wolf x dog ($0.441 \leq q_i \leq 0.847$) and jackal x dog ($0.486 \leq q_i \leq 0.744$) hybrid individuals (Fig. 4a,b and Table S6). Bayesian clustering and assignment analyses run in STRUCTURE were highly concordant with the Maximum-likelihood clustering and assignment procedures performed in ADMIXTURE using both 192-SNP and 98,004-SNP parental

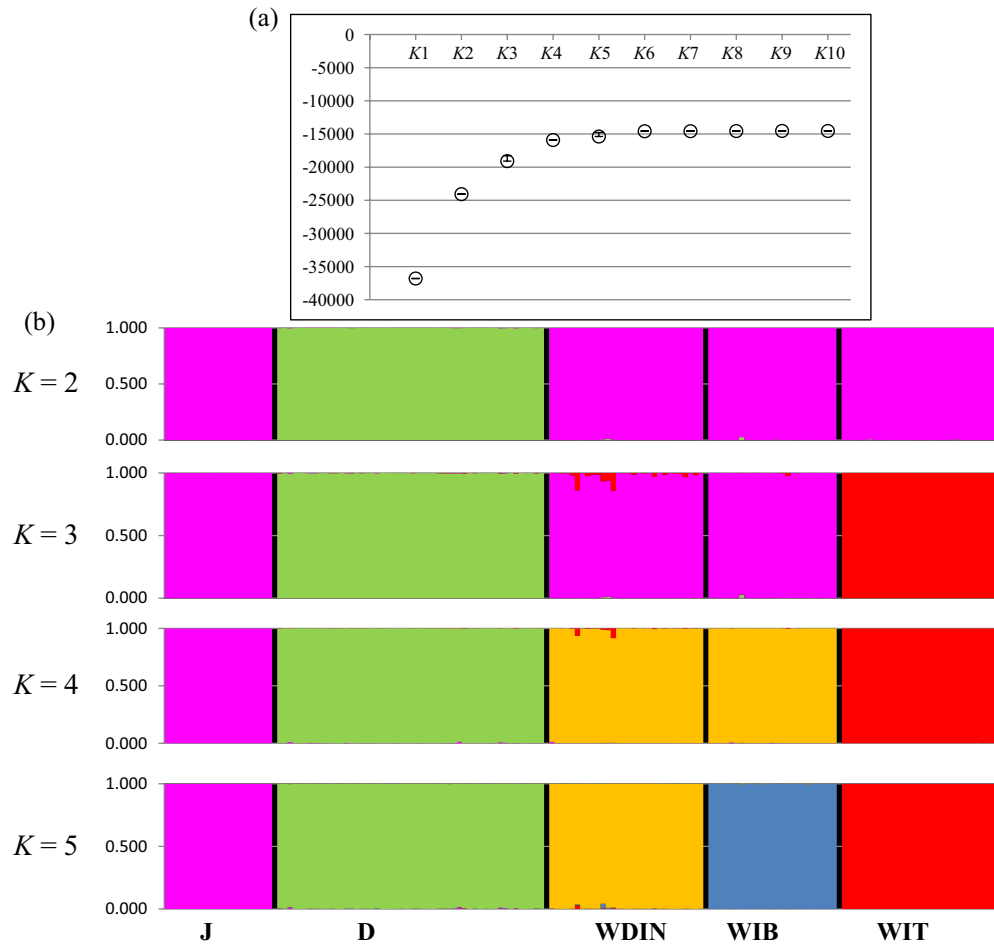


Figure 2. Results from the Bayesian assignment procedures. **(a)** Estimated posterior probability $\ln P(K)$ and corresponding standard deviations of the K clusters from 1 to 10. **(b)** Bar plots for the 158 individual q_i -values obtained through assignment with the 192-SNP genotypes. Each individual is represented by a vertical line partitioned into coloured segments, whose length is proportional to the individual coefficients of membership (q_i) in the jackal, dog and wolf clusters inferred by Bayesian assignment analyses performed in STRUCTURE2.3.4⁶⁰, assuming K from 2 to 5, no prior information (option *Usepopinfo* not activated) and choosing the “Admixture” and “Independent Allele Frequency” models. J: Jackals; D: dogs; WDIN: Dinaric wolves; WIB: Iberian wolves; WIT: Italian wolves. Bar plots were obtained concatenating data from the five independent runs using CLUMPP 1.1.1⁶² and graphically displayed using DISTRUCT 1.1⁶³.

Populations	N	Q_J (90% CI)	Q_D (90% CI)	Q_{WDIN} (90% CI)	Q_{WIB} (90% CI)	Q_{WIT} (90% CI)
Jackals	21	0.999 (0.995–1.000)	0.000 (0.000–0.001)	0.000 (0.000–0.001)	0.000 (0.000–0.001)	0.000 (0.000–0.001)
Domestic dogs	52	0.001 (0.000–0.009)	0.996 (0.977–1.000)	0.001 (0.000–0.004)	0.001 (0.000–0.005)	0.001 (0.000–0.004)
Dinaric wolves	30	0.001 (0.000–0.004)	0.000 (0.000–0.002)	0.995 (0.975–1.000)	0.003 (0.000–0.012)	0.002 (0.000–0.008)
Iberian wolves	25	0.001 (0.000–0.003)	0.000 (0.000–0.001)	0.001 (0.000–0.004)	0.998 (0.988–1.000)	0.000 (0.000–0.001)
Italian wolves	30	0.000 (0.000–0.002)	0.000 (0.000–0.001)	0.000 (0.000–0.001)	0.000 (0.000–0.001)	0.999 (0.994–1.000)

Table 1. Average membership proportions Q_i of the five canid groups in their original sampling group (in bold) with 90% confidence intervals (CI), estimated from the Bayesian assignment analyses performed in STRUCTURE 2.3.4⁶⁰, assuming $K=5$ clusters and using the “Admixture” and “Independent allele frequencies” models as parameter settings. Average Q_i were obtained concatenating the data from the five independent runs using CLUMPP 1.1.1⁶².

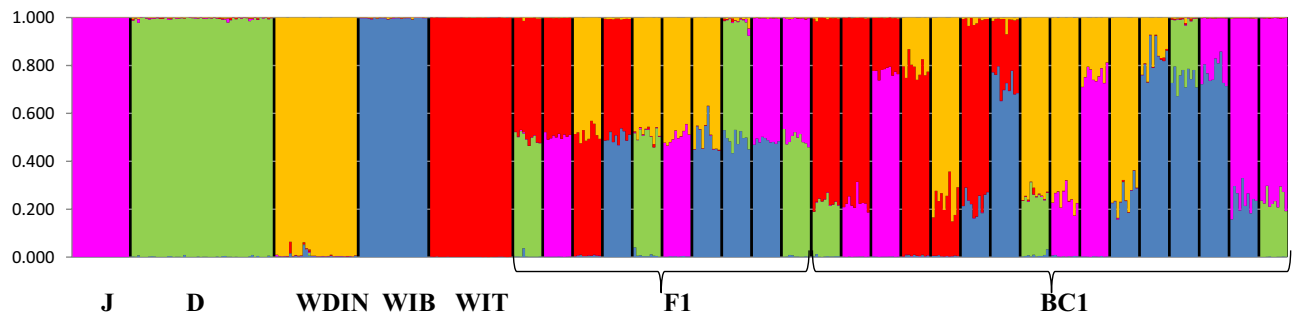


Figure 3. Bar plots for the 158 individual q_T -values obtained through STRUCTURE 2.3.4⁶⁰ assignment for the canid reference populations with the 192-SNP genotypes, assuming $K=5$ clusters: jackals (J), dogs (D), Dinaric wolves (WDIN), Iberian wolves (WIB), Italian wolves (WIT), and 10 first-generation (F1) hybrid and first-generation backcross (BC1) genotypes simulated for each pairwise combination of canid groups. Bar plots were obtained concatenating data from the five independent runs using CLUMPP 1.1.1⁶² and graphically displayed using DISTRUCT 1.1⁶³.

population genotypes (Fig. S1b,c,d,e and Table S7, S8). Previous genome-wide analyses have indicated substructure between Dinaric and Balkan wolves² and earlier studies from Bulgaria have reported wolf-dog and possibly also wolf-jackal hybridization¹⁹. We therefore performed additional tests with 10 wolf profiles from Bulgaria. These analyses indicated that F1-hybrids and BC1-individuals from the Balkan region can also be distinguished based on the identified loci (results not shown).

Genetic variability. The proportions of polymorphic loci ranged from about 8% in jackals to more than 80% in dogs, with intermediate values (30%–45%) in the three wolf populations (Table 3). The highest rates of A_O , A_E , A_R , PIC, H_O and uH_E were observed in dogs and the lowest values were seen in jackals (Table 3). Among the wolf populations, Italian wolves showed the lowest values for the variability indexes, followed by Iberian and Dinaric wolves (Table 3). AMOVA showed that more than 81% of the total genetic diversity was significantly ($P < 0.001$) partitioned among the five groups. All pairwise F_{ST} values (Table 4) were highly statistically significant ($P < 0.0001$) and aligned with the very low corresponding effective F_{ST} -based migration rates (Table 4). The PID and PID_{SIBS} values for the 192 loci, and the probability of finding multiple individuals with the same genotype, were low (Table 3).

Candidate loci under selection and population divergence times. With a false discovery rate threshold of 0.05, BAYESCAN⁷³ identified $n=3$ putative outlier SNPs, one for each pairwise comparison between wolf populations, mapped on chromosomes 1, 8, and 25, respectively (See Supplemental Note S1, Fig. S2 and Fig. S3 for details). For wolf population divergence times, ABC simulations showed the highest support for scenario number 7 (sequential population splitting with subsequent bottlenecks), which performed better than the other models (Supplemental Note S2).

Fluidigm assay design, development and application. We successfully designed primer pairs and assays for all the new 108 SNPs that satisfied the recommended parameters (minimum separation distance of 100 base pairs), avoiding regions with large repeats or otherwise difficult for the interpretation of results. After amplifying the 93 DNA samples 2–3 times at the 192 SNPs, we discarded 14 loci among the newly-designed SNPs (corresponding to 13% of panel 2–3, and 7% of the 192 SNPs) that failed amplifications in all reactions (Tables S3, S4, and S5), resulting in 178 SNPs (Fig. 5). Genotyping success for the remaining 178 SNPs was high across samples and markers, and only one non-invasively sampled dog with 96% missing data was removed from the analyses. All of the remaining wolf, hybrid, and dog samples were successfully genotyped at $\geq 95\%$ of the tested loci, showing an average amplification success rate of 98% and an average allelic dropout rate of 4.4% across loci (Table S5). As expected, invasive samples showed the highest AS (99%) and the lowest error (0.7%) rates (Table S5), and their genotypes fully matched the original profiles from the Illumina CanineHD BeadChip. Interestingly, non-invasive samples showed only slightly lower AS (97%) and only slightly higher error (3.7%) rates across loci (Table S5). Cross-species amplification tests also produced valid genotypes for the other canid species with jackals showing AS=97%, ADO=1% and FA=0, and red foxes showing AS=90%, ADO=0.5% and FA=0 (Table S5). Notably, Panel 1 confirmed its high performance (AS=99%, ADO=0.6% and FA=0 for tissue samples and AS=98%, ADO=6.1% and FA=0 for faecal samples) for genotyping of various DNA sources. The newly designed and tested Panels 2 and 3 also performed well for genotyping both invasive (AS=98%, ADO=0.8% and FA=0) and non-invasive samples (AS=96%, ADO=1.6% and FA=0) on the Fluidigm platform (Table S5). A PCoA on invasive and non-invasive samples genotyped with the 178 SNPs confirmed the previous assignments. The different canid groups were clearly separated, with non-invasive genotypes clustering with profiles from their respective reference groups (Fig. S4). Admixed individuals were placed between parental groups, with recent admixture resulting in intermediate positions and older backcrosses positioned closer to wild parental groups (Fig. S4). Importantly, re-analyses with 178 of 192 SNPs did not affect the overall results. Outcomes from multivariate, Bayesian and Maximum-likelihood assignment and clustering analyses in

	Populations	N	Q_i (90% CI)		Q_D (90% CI)		Q_{WDIN} (90% CI)		Q_{WIB} (90% CI)		Q_{WIT} (90% CI)	
Parentals	Jackals	21	0.998	(0.990–1.000)	0.000	(0.000–0.002)	0.001	(0.000–0.003)	0.001	(0.000–0.005)	0.000	(0.000–0.002)
	Domestic dogs	52	0.002	(0.000–0.014)	0.992	(0.965–1.000)	0.002	(0.000–0.013)	0.002	(0.000–0.015)	0.002	(0.000–0.011)
	Dinaric wolves	30	0.001	(0.000–0.008)	0.001	(0.000–0.005)	0.989	(0.957–1.000)	0.006	(0.000–0.030)	0.003	(0.000–0.014)
	Iberian wolves	25	0.001	(0.000–0.006)	0.000	(0.000–0.003)	0.002	(0.000–0.013)	0.996	(0.978–1.000)	0.001	(0.000–0.005)
	Italian wolves	30	0.000	(0.000–0.003)	0.000	(0.000–0.003)	0.001	(0.000–0.004)	0.001	(0.000–0.004)	0.998	(0.989–1.000)
Simulated F1	D-WIT	10	0.003	(0.000–0.017)	0.492	(0.431–0.553)	0.004	(0.000–0.025)	0.006	(0.000–0.028)	0.496	(0.431–0.559)
	J-WIT	10	0.503	(0.431–0.574)	0.000	(0.000–0.003)	0.002	(0.000–0.011)	0.002	(0.000–0.012)	0.493	(0.422–0.565)
	WIT-WDIN	10	0.002	(0.000–0.009)	0.001	(0.000–0.004)	0.485	(0.395–0.574)	0.005	(0.000–0.031)	0.508	(0.422–0.594)
	WIT-WIB	10	0.001	(0.000–0.006)	0.000	(0.000–0.003)	0.006	(0.000–0.039)	0.501	(0.419–0.580)	0.492	(0.414–0.569)
	D-WDIN	10	0.003	(0.000–0.018)	0.502	(0.437–0.567)	0.482	(0.398–0.555)	0.009	(0.000–0.054)	0.004	(0.000–0.027)
	J-WDIN	10	0.500	(0.407–0.593)	0.000	(0.000–0.002)	0.496	(0.401–0.589)	0.003	(0.000–0.020)	0.001	(0.000–0.006)
	WDIN-WIB	10	0.001	(0.000–0.008)	0.000	(0.000–0.003)	0.496	(0.367–0.627)	0.501	(0.370–0.630)	0.001	(0.000–0.008)
	D-WIB	10	0.007	(0.000–0.037)	0.488	(0.424–0.552)	0.010	(0.000–0.058)	0.492	(0.408–0.564)	0.004	(0.000–0.023)
	J-WIB	10	0.512	(0.419–0.604)	0.000	(0.000–0.002)	0.003	(0.000–0.015)	0.485	(0.391–0.579)	0.001	(0.000–0.004)
D-J	10	0.501	(0.436–0.564)	0.490	(0.430–0.552)	0.003	(0.000–0.021)	0.004	(0.000–0.025)	0.002	(0.000–0.010)	
Simulated BC1	D-WIT-WIT	10	0.002	(0.000–0.012)	0.224	(0.175–0.275)	0.004	(0.000–0.029)	0.004	(0.000–0.027)	0.766	(0.710–0.816)
	J-WIT-WIT	10	0.226	(0.167–0.290)	0.001	(0.000–0.003)	0.002	(0.000–0.010)	0.002	(0.000–0.010)	0.770	(0.707–0.829)
	J-WIT-J	10	0.772	(0.709–0.831)	0.000	(0.000–0.002)	0.002	(0.000–0.011)	0.002	(0.000–0.012)	0.224	(0.166–0.286)
	WIT-WDIN	10	0.002	(0.000–0.010)	0.001	(0.000–0.003)	0.213	(0.137–0.291)	0.005	(0.000–0.035)	0.780	(0.705–0.849)
	WIT-WDIN	10	0.001	(0.000–0.007)	0.000	(0.000–0.003)	0.763	(0.680–0.838)	0.004	(0.000–0.025)	0.232	(0.158–0.311)
	WIT-WIB-WIT	10	0.001	(0.000–0.008)	0.000	(0.000–0.003)	0.013	(0.000–0.081)	0.226	(0.130–0.303)	0.760	(0.689–0.826)
	WIT-WIB-WIB	10	0.001	(0.000–0.008)	0.000	(0.000–0.003)	0.014	(0.000–0.070)	0.723	(0.638–0.798)	0.261	(0.192–0.334)
	D-WDIN-WDIN	10	0.003	(0.000–0.017)	0.247	(0.194–0.304)	0.737	(0.661–0.798)	0.008	(0.000–0.051)	0.005	(0.000–0.027)
	J-WDIN-WDIN	10	0.240	(0.161–0.325)	0.000	(0.000–0.003)	0.754	(0.666–0.835)	0.004	(0.000–0.026)	0.001	(0.000–0.010)
	J-WDIN-J	10	0.756	(0.674–0.833)	0.000	(0.000–0.003)	0.240	(0.162–0.323)	0.003	(0.000–0.017)	0.001	(0.000–0.006)
	WDIN-WIB-WDIN	10	0.001	(0.000–0.009)	0.000	(0.000–0.004)	0.746	(0.619–0.864)	0.250	(0.132–0.377)	0.002	(0.000–0.013)
	WDIN-WIB-WIB	10	0.001	(0.000–0.008)	0.000	(0.000–0.003)	0.167	(0.054–0.286)	0.827	(0.711–0.927)	0.004	(0.000–0.027)
	D-WIB-WIB	10	0.002	(0.000–0.014)	0.243	(0.191–0.298)	0.005	(0.000–0.030)	0.746	(0.682–0.803)	0.004	(0.000–0.026)
	J-WIB-WIB	10	0.226	(0.148–0.311)	0.000	(0.000–0.002)	0.002	(0.000–0.012)	0.771	(0.685–0.850)	0.001	(0.000–0.004)
	J-WIB-J	10	0.758	(0.675–0.834)	0.000	(0.000–0.002)	0.002	(0.000–0.014)	0.239	(0.162–0.322)	0.001	(0.000–0.004)
D-J	10	0.758	(0.704–0.807)	0.237	(0.189–0.288)	0.002	(0.000–0.013)	0.002	(0.000–0.012)	0.001	(0.000–0.008)	

Table 2. Average membership proportions Q_i in their original sampling group (in bold) with 90% confidence intervals (CI) for the five canid groups and the first two generations of simulated hybrids, estimated from the Bayesian assignment analyses performed in STRUCTURE 2.3.4⁶⁰, assuming $K=5$ clusters and using the “Admixture” and “Independent allele frequencies” models as parameter settings. Data comprise the 192-SNP genotypes of the parental populations (jackals, dogs, Dinaric, Iberian and Italian wolves), their first-generation hybrids (F1) and the first backcross (BC1) of simulated admixed genotypes. J = jackal, D = dog, W = wolf, IT = Italy, DIN = Dinaric, IB = Iberia. Average Q_i were obtained concatenating the data from the five independent runs using CLUMPP 1.1.1⁶².

ADEGENET, STRUCTURE and ADMIXTURE with the five reference populations and 178-SNP genotypes (Fig S5a; Table S9 and S12), simulated (Fig S5b; Tables S10 and S13) and other empirical data (Fig S5c,d; Table S11), were always significantly correlated ($R \geq 0.997$; $P < 0.0001$ for all pairwise combinations) and not significantly different ($t \geq 0.005$, $P \geq 0.977$; t -tests for all pairwise combinations) from the results obtained with 192 SNPs.

Discussion

We employed a reduced panel of 192 SNPs to reliably discriminate between anthropogenic hybridization and natural gene flow among populations in five canid groups, centered on wolves in southern Europe. Of these SNPs, 178 were successfully amplified on the Fluidigm platform, and provided the same results as the 192 SNPs for analysis of population structure. Results from empirical data and simulated profiles suggest that the selected marker set is well-suited for this purpose, and our analyses unambiguously identified all F1-hybrids and all BC1-individuals. Our results have shown that the panel performs well, also for non-invasively collected samples, and the reduced set of SNPs presented by Harmoinen et al.¹⁵ and in this study can therefore provide informative results for monitoring, research, and conservation management.

Marker selection and the power of discriminant loci. Our results based on empirical data and simulated profiles indicated that the identified SNPs clearly differentiate F1-hybrids and accurately detect BC1-individuals within and across canid species, showing no type I (non-admixed individuals erroneously identified

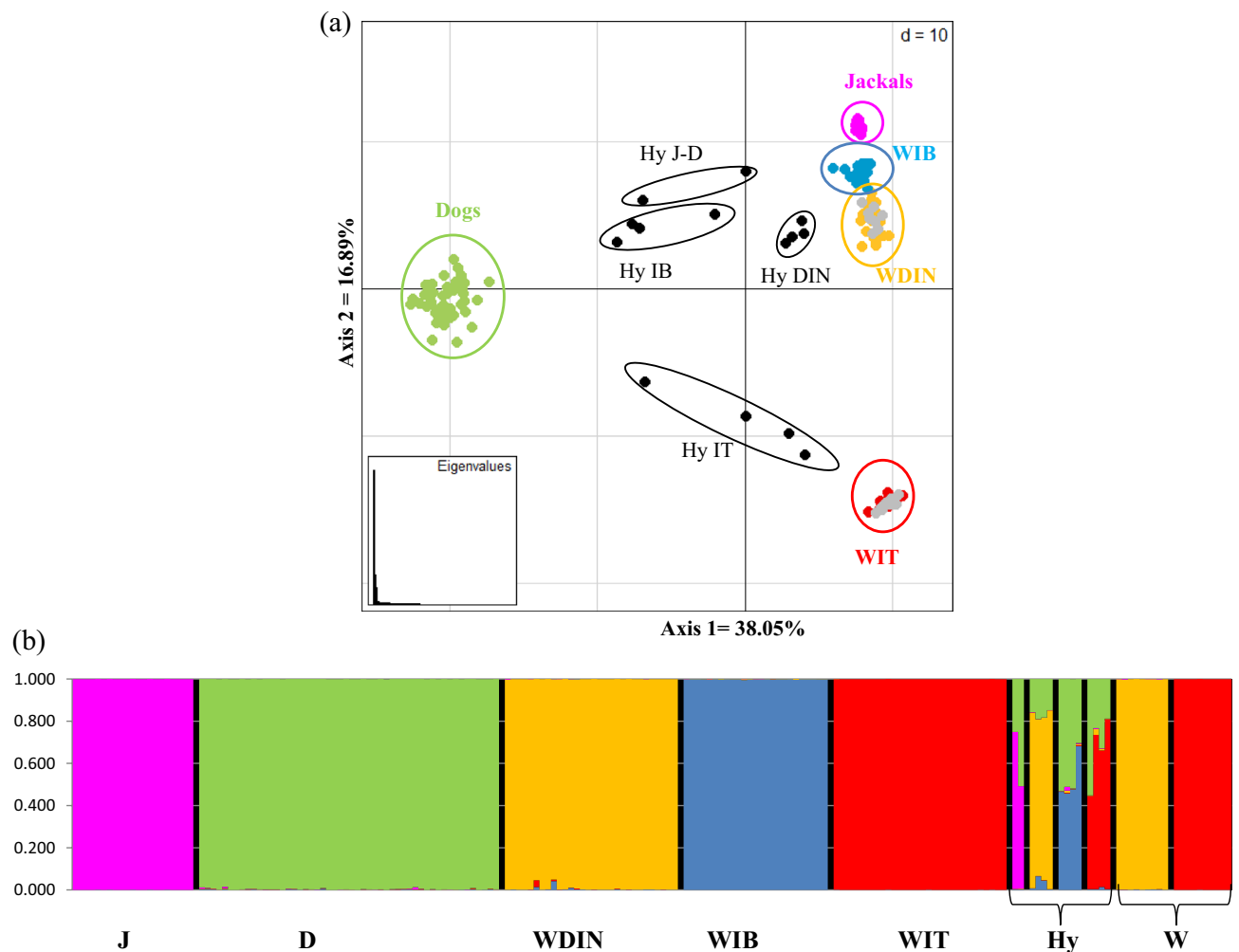


Figure 4. Multivariate discriminant analyses (PC performed with the “*dudi.pca*” function) (a) in the ADEGENET 2.3.4⁵⁷ package and (b) with a Bayesian assignment approach in STRUCTURE 2.3.4⁶⁰ (assuming $K = 5$ and the “Admixture” and “Independent allele frequencies” models) obtained analysing the 192-SNP genotypes of the five reference populations (J: Jackals; D: Dogs; WDIN: Dinaric; WIB: Iberian; WIT: Italian wolves) together with 33 additional canid genotypes categorized as non-admixed (W, grey dots) or admixed (Hy, black dots) in earlier studies ($n = 10$ Italian and $n = 9$ Dinaric wolves, $n = 4$ Italian wolf \times dog (HyIT), $n = 4$ Dinaric wolf \times dog (HyDIN), $n = 4$ Iberian wolf \times dog (HyIB), and $n = 2$ jackal \times dog individuals (HyJ-D)). Bar plots were obtained concatenating data from the five independent runs using CLUMPP 1.1.1⁶² and graphically displayed using DISTRUCT 1.1⁶³.

as admixed animals) nor type II (admixed individuals falsely identified as non-admixed animals) errors²⁶. The identified loci incorporate the panel of 93 SNPs developed for microfluidic genotyping by Harmoinen et al.¹⁵, who found almost identical genotypes in a test of 30 invasive and non-invasive samples from the same wolves. As a complement to the European wolf-dog hybrid panel presented by Harmoinen et al.¹⁵, we therefore propose the selected SNPs for a more detailed evaluation of population admixture *versus* hybridization in southern European populations, including possible events of jackal-hybridization^{19,23}.

The presence of fixed alleles in jackals, dogs, and between wolf populations contributed to the discriminant power of certain loci, which can be particularly valuable as diagnostic markers for monitoring purposes. Where incomplete profiles are obtained from non-invasive samples, the results from a few such markers can be enough to allow species identification⁷⁴, which is highly relevant for tracking the expanding distribution of jackals. Their rapid range expansion across Europe²² and within our study area, recently illustrated by the confirmation of jackal presence south of the Po River in northern Italy (<http://gojage.blogspot.com/2021/01/all-roads-lead-to-rome-european-golden.html>), highlights the urgent need to account for this species in future canid monitoring and management plans. Our results indicated that the selected SNPs perform well for identification of F1-hybrids and BC1-individuals for all canid group combinations. Detecting the presence and distribution of these early-generation hybrids is typically a higher priority than that of individuals with lower proportions of dog ancestry, to prioritize scarce conservation resources²⁶. The findings from our empirical and simulated data therefore indicate that our panel can reliably identify the animals that are most essential for monitoring and conservation management (e.g., the LIFE MIRCO-Lupo project, <http://www.lifemircolupo.it/>).

Populations (N)	GL (%)	PL (%)	ML (%)	A _O	A _E	A _R	N _P	PIC	H _O	uH _E	F _{IS}	P	PID	PID _{sibs}
Jackals (21)	99.581	8.334	91.666	1.083 (0.020)	1.039 (0.012)	1.076	2	0.019	0.024 (0.007)	0.023 (0.007)	-0.035	0.240	1.343 × 10 ⁻⁰⁴	1.039 × 10 ⁻⁰²
Domestic dogs (52)	99.712	88.545	11.455	1.885 (0.023)	1.481 (0.022)	1.859	31	0.239	0.264 (0.010)	0.296 (0.011)	0.110	<0.001 ***	1.005 × 10 ⁻⁵⁰	4.279 × 10 ⁻²⁶
Dinaric wolves (30)	99.861	46.352	53.648	1.464 (0.036)	1.156 (0.020)	1.383	0	0.081	0.098 (0.012)	0.099 (0.011)	0.010	0.348	2.776 × 10 ⁻¹⁸	3.828 × 10 ⁻⁰⁹
Iberian wolves (25)	99.793	35.946	64.054	1.359 (0.035)	1.114 (0.017)	1.322	0	0.063	0.065 (0.009)	0.076 (0.010)	0.143	<0.001 ***	2.002 × 10 ⁻¹³	3.855 × 10 ⁻⁰⁷
Italian wolves (30)	99.913	32.818	67.182	1.328 (0.034)	1.092 (0.017)	1.245	0	0.047	0.058 (0.010)	0.057 (0.009)	-0.003	0.439	2.428 × 10 ⁻¹⁰	1.247 × 10 ⁻⁰⁵

Table 3. Genetic variability at 192 SNPs in jackals, dogs, and wolves from the Dinaric, Iberian and Italian populations. Values in parentheses denote standard errors. N = sample size; GL = percentage of mean proportion of successfully genotyped loci; PL = proportion of polymorphic loci; ML = proportion of monomorphic loci; A_O = number of observed alleles; A_E = number of expected alleles; A_R = allelic richness; N_P = number of private alleles; PIC = polymorphic information content; H_O = observed heterozygosity; uH_E = unbiased expected heterozygosity; F_{IS} = estimates of departure from Hardy–Weinberg equilibrium; P = probability of obtaining, by chance, F_{IS} values higher than those observed after 10,000 random permutations (*P < 0.05, **P < 0.01, ***P < 0.001); PID = probability of identity; PID_{sibs} = probability of identity among full sibs.

Populations	Jackals	Domestic dogs	Dinaric wolves	Iberian wolves	Italian wolves
Jackals	–	0.119	0.091	0.070	0.033
Domestic dogs	0.677	–	0.141	0.136	0.110
Dinaric wolves	0.734	0.640	–	0.257	0.087
Iberian wolves	0.781	0.648	0.493	–	0.058
Italian wolves	0.884	0.695	0.741	0.813	–

Table 4. Matrix of pairwise F_{ST} values (below the diagonal) and the corresponding estimates of historical gene flow N_M (above the diagonal) among the sampled populations computed using 192-SNP genotypes. All F_{ST} values were highly statistically significant ($P < 0.0001$).

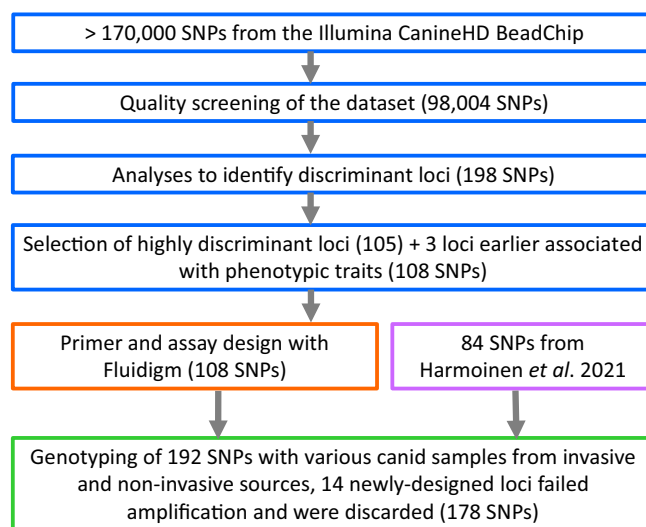


Figure 5. Flowchart describing the SNP selection process for design of the reduced marker panel.

Although the original panel of 93 SNPs from Harmoinen et al.¹⁵ was developed and successfully tested with non-invasive samples, and the additional loci presented in this study also performed well with non-invasive sources of DNA, we underline that the proposed panel may need future modifications given the highly dynamic situation for canids in our study area. In view of the recent gene flow between Italian Alpine and Dinaric wolves¹², the detection of dog-jackal hybridization²³ and the increasing overlap seen in wolf and jackal distributions

throughout Europe²¹, it will also be necessary to monitor for possible changes in the spatial and temporal distribution of allele frequencies in the selected SNPs²⁶. However, ongoing international collaborations in population monitoring, as illustrated by the LIFE WolfAlps EU project (<http://www.lifewolfalps.eu/>), will provide such information for wolves and other wide-ranging carnivore species. Accordingly, the opportunities for data sharing offered by new genomic approaches will contribute toward keeping these tools as useful and updated as possible.

Population structure. All five *a priori* sampling groups emerged as well-separated genetic clusters based on multivariate and Bayesian clustering analyses of the identified SNPs. As expected, dogs represented the most divergent group, and Italian wolves were also clearly differentiated from the other wolves, as observed in earlier analyses of SNP profiles^{1–3} and supported by earlier findings from analyses of mitochondrial DNA and microsatellite markers^{11,75,76}. Moreover, pairwise values of genetic distances between populations align with previous findings^{1–3} for southern European wolf populations. Notably, jackals emerged as a distinct group only at $K=4$ population clusters, which is likely a reflection of the ascertainment bias in the SNP panel originally developed for dogs¹.

Genetic variability. The lower diversity values for Italian wolves, followed by Iberian wolves and Dinaric wolves, accord with earlier SNP analyses of European wolf populations^{1–3}. It is nonetheless important to note that the reported levels of diversity differ among analyses, and that the *a priori* selection of loci for the purposes of our study likely influenced the results. The SNP array that formed the basis of our study is centered on genetic variation in dogs^{1,77}, magnifying their genetic diversity while underestimating that of wolves and the more distantly related jackals¹. Furthermore, our selection of markers focused on SNPs with high F_{ST} -values. For these reasons, direct comparisons of genetic variability among taxa should be treated with caution. Future developments based on whole-genome data may provide additional new markers less affected by ascertainment bias. However, the selection of loci from the existing—and commonly used—SNP panel facilitates the use of already-published canid profiles from earlier projects as reference populations, which could increase the utility of the panel.

Notwithstanding, we report various diversity parameters for comparative purposes, which might also inform future efforts for different canid groups or other taxa. Relevant examples are the values for PID and PID_{sibs} , where values calculated across 21 jackals ($PID = 1.343 \times 10^{-04}$, $PID_{sibs} = 1.039 \times 10^{-02}$) were substantially higher than those calculated for the 30 Italian wolves ($PID = 2.428 \times 10^{-10}$, $PID_{sibs} = 1.247 \times 10^{-05}$) that represented the 2nd-highest values. Nonetheless, our selected SNPs provide essential information, given that the initial 93-SNP panel by Harmoinen et al.¹⁵ did not distinguish among wild canids (i.e., among wolves, jackals, and red foxes). Although a higher number of SNPs are required to obtain PID values similar to standard microsatellite panels^{32,43}, careful testing have shown that such panels can still work well across broad geographic regions and divergent populations¹⁵.

Candidate loci under selection and population divergence times. We detected a limited number of loci under possible selection (Supplemental Note S1). Although earlier SNP analyses have suggested possible environmental selection in European wolves^{3,78}, future research with more samples and genomic data are needed to evaluate these findings, including the potential presence of false positives among our results. Our population divergence modeling (Supplemental Note S2) indicated sequential population splits and bottlenecks broadly consistent with whole-genome results⁵, although our results showed very wide confidence intervals and could be partly affected by ascertainment bias, mutation rates, marker numbers, and secondary gene flow^{3,5,79}.

Fluidigm assay design, development and application. Our results indicate that the selected SNPs are well-suited for conservation management, including projects based on non-invasively collected samples. When tested on standard platforms such as the microfluidic arrays, most of the candidate markers performed efficiently and reliably, showing high amplification success and low error rates on various DNA sources from different canid groups. Well-differentiated allele frequencies allowed us to distinguish among the different canid groups and their first generation hybrids. Importantly, we were also able to distinguish the aforementioned groups from the red fox, whose DNA may be present in non-invasively collected samples such as livestock damage cases.

Conclusion and recommendations. We found that the identified SNPs can reliably distinguish natural population admixture from inter-specific hybridization in five canid groups comprising dogs, jackals, and the long-divergent Iberian, Italian and Dinaric wolf populations, up to the first generation of backcrosses, using various sources of DNA, including non-invasively collected materials. New genomic tools provide opportunities to integrate data from different laboratories without the need for calibration, which is essential for wide-ranging species such as wolves. The sequencing of microsatellite markers³⁵, which is now in progress for wolves (Skrbinšek et al., *unpublished data*), may also offer opportunities to combine hypervariable microsatellites with SNP loci possibly associated with adaptive genetic variation. The selected SNPs can provide timely identification of dispersers and ensure that such important individuals are not erroneously classified as hybrids, with the subsequent risk of management decisions that could be harmful for conservation. The proposed panel will also facilitate the monitoring of jackals, which are rapidly expanding their ranges across Europe, and contribute to research, monitoring, and conservation management of wild European canids.

Received: 12 May 2021; Accepted: 1 March 2022

Published online: 09 March 2022

References

1. von Holdt, B. M. *et al.* A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. *Genome Res.* **21**, 1294–1305 (2011).
2. Stronen, A. V. *et al.* North-south differentiation and a region of high diversity in European wolves. *PLoS ONE* **8**(10), e76454 (2013).
3. Pilot, M. *et al.* Genome-wide signatures of population bottlenecks and diversifying selection in European wolves. *Heredity* **112**, 428–442 (2014).
4. Fan, Z. *et al.* Worldwide patterns of genomic variation and admixture in gray wolves. *Genome Res.* **26**, 1–11 (2016).
5. Silva, P. *et al.* Genomic evidence for the Old divergence of Southern European wolf populations. *Proc. R. Soc. B* **287**, 20201206 (2020).
6. Chapron, G. *et al.* Recovery of large carnivores in Europe's modern human-dominated landscapes. *Science* **346**, 1517–1519 (2014).
7. Ripple, W. J. *et al.* Status and ecological effects of the world's largest carnivores. *Science* **343**, 1241484 (2014).
8. Fabbri, E. *et al.* From the Apennines to the Alps: colonization genetics of the naturally expanding Italian wolf (*Canis lupus*) population. *Mol. Ecol.* **16**, 1661–1671 (2007).
9. Hulva, P. *et al.* Wolves at the crossroads: fission-fusion range biography in the Western Carpathians and Central Europe. *Divers. Distrib.* **24**, 179–192 (2018).
10. Szewczyk, M. *et al.* Dynamic range expansion leads to establishment of a new, genetically distinct wolf population in Central Europe. *Sci. Rep.* **9**, 19003 (2019).
11. Fabbri, E. *et al.* Genetic structure of expanding wolf (*Canis lupus*) populations in Italy and Croatia, and the early steps of the recolonization of the Eastern Alps. *Mamm. Biol.* **79**, 138–148 (2014).
12. Ražen, N. *et al.* Long-distance dispersal connects Dinaric-Balkan and Alpine grey wolf (*Canis lupus*) populations. *Eur. J. Wildl. Res.* **62**, 137–142 (2016).
13. Bartol, M., Boljte, B., Černe, R., Jelenčič, M., Jonozovič, M., Konec, M., Kos, I., Kraševc, R., Krofel, M., Kuralt, Ž., Potočnik, H. & Skrbinšek, T. Spremljanje varstvenega stanja volkov v Sloveniji v letih 2017/2020. Drugo delno poročilo – poročilo za sezono 2017/2018. 41 (2018).
14. Hindrikson, M. *et al.* Wolf population genetics in Europe: a systematic review, meta-analysis and suggestions for conservation and management: wolf population genetics in Europe. *Biol. Rev.* **92**, 1601–1629 (2017).
15. Harmoinen, J. *et al.* Reliable wolf-dog hybrid detection in Europe using a reduced SNP panel developed for non-invasively collected samples. *BMC Genomics* **22**, 473 (2021).
16. Randi, E. Genetics and conservation of wolves *Canis lupus* in Europe. *Mammal Rev.* **41**, 99–111 (2011).
17. Pilot, M. *et al.* Widespread, long-term admixture between grey wolves and domestic dogs across Eurasia and its implications for the conservation status of hybrids. *Evol. Appl.* **11**, 662–680 (2018).
18. Freedman, A. H. *et al.* Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet.* **10**(1), e1004016 (2014).
19. Moura, A. E. *et al.* Unregulated hunting and genetic recovery from a severe population decline: the cautionary case of Bulgarian wolves. *Conserv. Genet.* **15**, 405–417 (2014).
20. Fabbri, E. *et al.* Genetic structure and expansion of golden jackals (*Canis aureus*) in the north-western distribution range (Croatia and eastern Italian Alps). *Conserv. Genet.* **15**, 187–199 (2014).
21. Krofel, M., Giannatos, G., Čirovič, D., Stoyanov, S. & Newsome, T. Golden jackal expansion in Europe: a case of mesopredator release triggered by continent-wide wolf persecution?. *Hystrix* **28**, 9–15 (2017).
22. Spassov, N. & Acosta-Pankov, I. Dispersal history of the golden jackal (*Canis aureus moreoticus* Geoffroy, 1835) in Europe and possible causes of its recent population explosion. *Biodiv. Data J.* **7**, e34825 (2019).
23. Galov, A. *et al.* First evidence of hybridization between golden jackal (*Canis aureus*) and domestic dog (*Canis familiaris*) as revealed by genetic markers. *R. Soc. Open Sci.* **2**, 150450 (2015).
24. Donfrancesco, V. *et al.* Unravelling the scientific debate on how to address wolf-dog hybridisation in Europe. *Front. Ecol. Evol.* **7**, 175 (2019).
25. Salvatori, V., Godinho, R., Braschi, C., Boitani, L. & Ciucci, P. High levels of recent wolf x dog introgressive hybridization in agricultural landscapes of central Italy. *Eur. J. Wildl. Res.* **65**, 73 (2019).
26. Caniglia, R. *et al.* A standardized approach to empirically define reliable assignment thresholds and appropriate management categories in deeply introgressed populations. *Sci. Rep.* **10**, 2862 (2020).
27. Randi, E. *et al.* Multilocus detection of wolf x dog hybridization in Italy, and guidelines for marker selection. *PLoS ONE* **9**(1), e86409 (2014).
28. Godinho, R. *et al.* Real-time assessment of hybridization between wolves and dogs: combining non-invasive samples with ancestry informative markers. *Mol. Ecol. Res.* **15**, 317–328 (2015).
29. Putnam, A. I. & Carbone, I. Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecol. Evol.* **4**, 4399–4428 (2014).
30. Norman, A. J. & Spong, G. Single nucleotide polymorphism-based dispersal estimates using noninvasive sampling. *Ecol. Evol.* **5**, 3056–3065 (2015).
31. Kraus, R. H. S. *et al.* A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in Europe based on noninvasively collected samples. *Mol. Ecol. Resour.* **15**, 295–305 (2015).
32. von Thaden, A. *et al.* Assessing SNP genotyping of noninvasively collected wildlife samples using microfluidic arrays. *Sci. Rep.* **7**, 10768 (2017).
33. von Thaden, A. *et al.* Applying genomic data in wildlife monitoring: Development guidelines for genotyping degraded samples with reduced single nucleotide polymorphism panels. *Mol. Ecol. Resour.* **20**, 662–680 (2020).
34. Giangregorio, P., Norman, A. J., Davoli, F. & Spong, G. Testing a new SNP-chip on the Alpine and Apennine brown bear (*Ursus arctos*) populations using non-invasive samples. *Conserv. Genet. Resour.* **11**, 355–363 (2019).
35. De Barba, M. *et al.* High-throughput microsatellite genotyping in ecology: improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Mol. Ecol. Resour.* **17**, 492–507 (2017).
36. Werhahn, G. *et al.* The unique genetic adaptation of the Himalayan wolf to high-altitudes and consequences for conservation. *Glob. Ecol. Conserv.* **16**, e00455 (2018).
37. Zimmerman, S. J., Aldridge, C. L. & Oyler-McCance, S. J. An empirical comparison of population genetic analyses using microsatellite and SNP data for a species of conservation concern. *BMC Genomics* **21**, 382 (2020).
38. Jones, M. R. *et al.* Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* **360**, 1355–1358 (2018).
39. Hohenlohe, P. A., Funk, W. C. & Rajora, O. P. Population genomics for wildlife conservation and management. *Mol. Ecol.* **30**, 62–82 (2020).
40. McDevitt, A. *et al.* Next-generation phylogeography resolves post-glacial colonization patterns in a widespread carnivore, the red fox (*Vulpes vulpes*), in Europe. *Molecular Ecology* **31**, 993–1006 (2022).

41. Ekblom, R. *et al.* Genome sequencing and conservation genomics in the Scandinavian wolverine population. *Conserv. Biol.* **32**, 1301–1312 (2018).
42. von Holdt, B. M. *et al.* Identification of recent hybridization between gray wolves and domesticated dogs by SNP genotyping. *Mamm. Genome* **24**, 80–88 (2013).
43. López-Bao, J. V. *et al.* Consistent bear population DNA-based estimates regardless molecular markers type. *Biol. Conserv.* **248**, 108651 (2020).
44. Silva, P. *et al.* Cryptic population structure and evidence of low dispersal in the Iberian wolf. *Sci. Rep.* **8**, 14108 (2018).
45. Pacheco, C. *et al.* Spatial assessment of wolf-dog hybridization in a single breeding period. *Sci. Rep.* **7**, 42475 (2017).
46. Kusak, J. *et al.* Wolf-dog hybridization in Croatia. *Vet. Arhiv* **88**, 375–395 (2018).
47. McFarlane, S. E. & Pemberton, J. M. Detecting the true extent of introgression during anthropogenic hybridization. *Trends Ecol. Evol.* **4**, 315–326 (2019).
48. Salvatori, V. *et al.* European agreements for nature conservation need to explicitly address wolf-dog hybridisation. *Biol. Conserv.* **248**, 108525 (2020).
49. Galaverni, M. *et al.* Disentangling timing of admixture, patterns of introgression, and phenotypic indicators in a hybridizing wolf population. *Mol. Biol. Evol.* **34**, 2324–2339 (2017).
50. Lequarré, A. S. *et al.* LUPA: A European initiative taking advantage of the canine genome architecture for unravelling complex disorders in both human and dogs. *Vet. J.* **189**, 155–159 (2011).
51. Vaysse, A. *et al.* Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet.* **7**(10), e1002316 (2011).
52. Caniglia, R. *et al.* Black coats in an admixed wolf x dog pack is melanism an indicator of hybridization in wolves?. *Eur. J. Wildl. Res.* **59**, 543–555 (2013).
53. Ciucci, P., Reggioni, W., Maiorano, L. & Boitani, L. Long-distance dispersal of a rescued wolf from the Northern Apennines to the Western Alps. *J. Wildl. Mngmt.* **73**, 1300–1306 (2009).
54. Anderson, T. M. *et al.* Molecular and evolutionary history of melanism in North American gray wolves. *Science* **323**, 1339–1343 (2009).
55. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
56. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2020).
57. Jombart, T. ADEGENET: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
58. Miller, J. M., Cullingham, C. I. & Peery, R. M. The influence of a priori grouping on inference of genetic clusters: simulation study and literature review of the DAPC method. *Heredity* **125**, 269–280 (2020).
59. Jombart, T., Devillard, S. & Balloux, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94 (2010).
60. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
61. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587 (2003).
62. Jakobsson, M. & Rosenberg, N. A. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801–1806 (2007).
63. Rosenberg, N. A. DISTRUCT: A program for the graphical display of population structure. *Mol. Ecol. Notes* **4**, 137–138 (2004).
64. Nielsen, E. E., Back, L. A. & Kotlicki, P. HybridLab (version 1.0): a program for generating simulated hybrids from population samples. *Mol. Ecol. Notes* **6**, 971–973 (2006).
65. Godinho, R. *et al.* Genetic evidence for multiple events of hybridization between wolves and domestic dogs in the Iberian Peninsula. *Mol. Ecol.* **20**, 5154–5166 (2011).
66. Waits, L. P., Luikart, G. & Taberlet, P. Estimating the probability of identify among genotypes in natural populations: cautions and guidelines. *Mol. Ecol.* **10**, 249–256 (2001).
67. Peakall, R. & Smouse, P. E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28**, 2537–2539 (2012).
68. Kalinowski, S. T., Taper, M. L. & Marshall, T. C. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* **16**, 1099–1106 (2007).
69. Goudet, J. Fstat (Version 1.2): a computer program to calculate F-statistics. *J. Hered.* **86**, 485–486 (1995).
70. Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N. & Bonhomme, F. GENETIX 4.05, logiciel sous Windows TM pour la génétique des Populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier, France (1996–2004).
71. Velli, E. *et al.* Ethanol versus swabs: what is a better tool to preserve faecal samples for non-invasive genetic analyses?. *Hystrix It. J. Mamm.* **30**, 24–29 (2019).
72. Valière, N. GIMLET: a computer program for analysing genetic individual identification data. *Mol. Ecol. Notes* **2**, 377–379 (2002).
73. Foll, M. & Gaggiotti, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977–993 (2008).
74. Stronen, A. V. *et al.* “Passive surveillance” across species with cross-amplifying molecular markers: the potential of wolf (*Canis lupus*) genetic monitoring in tracking golden jackal (*C. aureus*) colonization and hybridization. *Hystrix It. J. Mamm.* **31**, 74–76 (2020).
75. Lucchini, V., Galov, A. & Randi, E. Evidence of genetic distinction and long-term population decline in wolves (*Canis lupus*) in the Italian Apennines. *Mol. Ecol.* **13**, 898–902 (2004).
76. Pilot, M. *et al.* Phylogeographic history of grey wolves in Europe. *BMC Evol. Biol.* **21**, 10–104 (2010).
77. Gopalakrishnan, S. *et al.* The wolf reference genome sequence (*Canis lupus lupus*) and its implications for *Canis* spp. population genomics. *BMC Genomics* **18**, 495 (2017).
78. Stronen, A. V. *et al.* Genome-wide analyses suggest parallel selection for universal traits may eclipse local environmental selection in a highly mobile carnivore. *Ecol. Evol.* **5**, 4410–4425 (2015).
79. Montana, L. *et al.* Combining phylogenetic and demographic inferences to assess the origin of the genetic diversity in an isolated wolf population. *PLoS ONE* **12**(5), e0176560 (2017).

Acknowledgements

We thank S. Czarnomska, A. Galov, J. Harmoinen, E. Velli, D. Battilani, P. Aragno, P. Genovesi, and the Mammal Research Institute, Polish Academy of Sciences, for their assistance. We are also grateful to two anonymous reviewers for their constructive feedback that greatly improved our manuscript. Funding was provided to ISPRA by the Italian Ministry of Environment (MATM; Direzione Tutela della Natura) and Regione Emilia Romagna (Assessorato Agricoltura) within a multi-year collaborative project to genotype and monitor the Italian wolf

population. AVS was supported by a senior postdoctoral fellowship from Insubria University, Italy. RG was supported by a research contract from the Portuguese Foundation for Science and Technology.

Author contributions

R.C., M.G., F.M., E.R., and A.V.S. conceived, designed and planned the study. R.G., J.K., E.R., A.R.-G., T.S., and A.V. collected samples and data. B.C. and C.N. designed the SNP array and performed the lab work. R.C., B.C., E.F., M.G., F.M., N.M., C.N., and A.V.S. performed data analyses. R.C., E.F., M.G., F.M., and A.V.S. wrote the manuscript, which was revised and approved by all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-08132-0>.

Correspondence and requests for materials should be addressed to A.V.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022