CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Tropical AgriSciences

Detection of Genetic Structure and Genome Interactions of Domestic Forms and Wild Species of Canids

DOCTORAL THESIS

Prague 2022

Author: Ing. Milena Jindřichová
Chief supervisor: Mgr. Barbora Černá Bolfíková, Ph.D.
Second (specialist) supervisor: doc. RNDr. Pavel Hulva, Ph.D.

Declaration

I hereby declare that I have done this thesis entitled “Detection of genetic structure and genome interactions of domestic forms and wild species of canids” independently, all texts in this thesis are original, and all the sources have been quoted and acknowledged by means of complete references and according to Citation rules of the FTA.

In Prague, date
6.9.2022
.....
Ing. Milena Jindřichová

Acknowledgements

I would like to warmly thank the breeders of the club of the Czechoslovakian Wolfdog and all the dogs' owners for sample collection and permissions in the analysis of the data that allowed the realization of this study, volunteers of Friends of the Earth Czech Republic for collecting and providing wolf samples.

My warmest acknowledgement belongs to my supervisor Barbora Černá Bolfíková for her imperishable support and help during the years. I would like to thank my co-authors from Department of Zoology, Charles University in Prague and Department of Biology and Ecology of Ostrava University both in the Czech Republic and Laboratorio di Genetica, Istituto Superiore per la Protezione e Ricerca Ambientale (ISPRA), Ozzano Emilia in Italy.

This study developed on financial support of Internal Grant Agency, Czech University of Life Sciences Prague under numbers IGA 20155014, IGA 20165015, IGA 20175018 IGA 20195012, IGA 20205007, IGA 20223108andNorway Grants EHP-CZ02- OV-1-022-2015.

Abstract

The evolution of dogs represents the oldest domestication event. Throughout history, wild wolves interbred with domestic populations, however, recently, hybridization between wolves and dogs is considered as a threat to wild populations and should be prevented. Similarities of dogs' and wolves' genomes challenge the detection of hybridization; therefore, we took advantage of an artificial hybrid population to study its genome composition. The main aim was to detect population structure and evaluate the genomic differences and similarities to the parental populations. The model population is recognized under the name Czechoslovakian Wolfdogs (CSW), which is a dog breed originating from four Carpathian wolves and an unknown number of German Shepherds. The thesis is based on three published studies that are focused on genetic and genomic characterization of the hybrid breed and on the detection of hybrids between dogs and wolves in the wild. Specific aims of our studies were to characterize the genetic composition of CSW breed in relation to its parental populations and to apply genome-wide procedures describe genomic composition and to reconstruct the history of the CSW breed. Whereas the dogs´ reference genome is available, it was possible to detect candidate genes for wolf-like and dog-like phenotypic traits typical for CSW individuals, including commonly inherited disorders. To study genetic composition of CSW breed, 39 autosomal and4 y-linked microsatellite loci were used together with hypervariable control region of mitochondrial DNA (mtDNA). A range of population genetic analyses based on both empirical and simulated data were performed. A panel of 170k Single Nucleotide Polymorphisms (SNPs) was used to study the composition of CSW genome. The CSW breed was well distinguishable from its parental populations using microsatellites or SNPs, however, clearly sharing more genetic similarities with dogs than with wolves. Two recognized mtDNA haplotypes and two Y-linked haplotypes were of dog ancestry. The inbreeding coefficient was low despite the small population size. More than 300 genes inherited from wolf ancestors and more than 2,000 genes inherited from dog ancestors were identified. Studying genomic details of this breed is a first step to address topics like illegal trade with wildlife or illegal crossing of dogs with wolves.

Keywords: Czechoslovakian Wolfdogs, German Shepherds, microsatellites, mtDNA, genetic variability, SNPs, genome-wide differentiation

Contents

List of tables

List of figures

List of the abbreviations used in the thesis

bp	base pairs
CITES	Convention on International Trade in Endangered Species of Wild Fauna and Flora
COI	coefficient of inbreeding
COR	coefficient of relatedness
CSW	Czechoslovakian Wolfdog
CW	Carpathian wolf
FCI	Fédération Cynologique Internationale
FoE	Friends of the Earth
GBS	Genotyping by Sequencing
GS	German Shepherd
HAW	high-altitude wolves
HW	Himalayan wolf
IBD	identity-by-descent
ISPRA	Italian Institute for Environmental Protection and Research
LGM	Last Glacial Maximum
MCMC	Markov chain Monte Carlo
MeDIP	methylated DNA immunoprecipitation
mtDNA	mitochondrial DNA
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PNATE	Appennino Tosco–Emiliano National Park
QTP	Qinghai-Tibet Plateau
SNP	single nucleotide polymorphism
STR	Single Tandem Repeats
TW	Tibetan grey wolf
ybp	years before present

This Dissertation thesis is based on the following publications:

Randi E, Hulva P, Fabbri E, Galaverni M, Galov A, Kusak J, Bigi D, Černá Bolfíková B, **Smetanová M**, Caniglia R. 2014. Multilocus detection of wolf x dog hybridization in Italy, and guidelines for marker Selection. PLoS ONE **9**(1): e86409.

Smetanová M, Černá Bolfíková B, Randi E, Caniglia R, Fabbri E, Galaverni M, Kutal M, Hulva P. 2015. From Wolves to Dogs, and Back: Genetic Composition of the Czechoslovakian Wolfdog. PLoS ONE **10**:1–11.

Caniglia R, Fabbri E, Černá Bolfíková B, Hulva P, **Jindřichová M,** Stronen A, Dykyy I, Camatta A, Carnier P, Randi E, Galaverni M. 2018. Wolf outside, dog inside? The genomic make-up of the Czechoslovakian Wolfdog. BMC Genomics **19**:1–17.

1 Introduction

Domestication of wild plants and animals is a major evolutionary transition in human history (Zeder 2015) that is dated to start in the Late Pleistocene (Larson et al. 2014). Dog *(Canis familiaris)* was the first domesticated animal and it is the only fully domesticated member of canids. Simultaneously, it is considered as the oldest domesticated animal in the world (VonHoldt et al. 2010). The development of the mutualistic relationship between the wild wolves and humans was based on behavioural and social similarities. Numerous megafaunal extinctions occurred during the Pleistocene period and it might have resulted into extinction of wolves which are considered as source population for today's dogs. Dogs later spread to the most of the world, experiencing life in much the same terms as people do (Morey & Jeger 2017). Although current studies are still discussing the place and the time of dog's domestication, a study published by Larson et al. (2012) evidences domestication process to be started even before the beginning of agriculture era. Results using ancient dog genomes imply that indigenous populations of dogs were already present in Europe and East Asia during the Palaeolithic (Frantz et al. 2016).Recent studies estimate that dogs were domesticated via a commensal pathway (Zeder 2012). Whole process was affected by repeated admixtures of the wild and domestic lineages, and also it had been enriched by multiple founding events from the several independent populations of wolves (*Canis lupus*) (Vilà et al. 1997; Skoglund et al. 2015).

Among domestic animals, dogs are the most variable in terms of phenotype (Drake & Klingenberg 2010). Nowadays, almost 400 dog breeds exist. Each of them is unique in genetic profile and each breed has its own specific history (Parker et al. 2017). Many domestic dog breeds have been formed by intense artificial selection through the fixation of discrete mutations (Pollinger et al. 2005; Wayne & VonHoldt 2012). Variability of dogs is almost unbelievable. Dogs are very variable in their size (for example the very small breeds such as Chihuahuas and breeds of very big sizes such as St. Bernard dogs), in shape, colour and their use, although just one mutual ancestor connects all breeds (Morell 1997). Dog breeds have been selected for their feature, look and mainly for their different behaviour. The behavioural variation found today is staggering, with breeds specialized on, for example, herding, retrieving, guarding or hunting (Parker et al. 2017).

However, hybridization may still occur in the wild. Many events concerning mating between wild wolves and escaped or feral dogs have been detected (Randi et al. 2014). Artificial hybridizations exist as well. These artificial hybridizations are controlled and designed by humans and usually they are recognised as experimental crossbreeding. Targeted hybridization may lead to creation of the new dog breeds, for example Saarloos Wolfdog, the Lupo Italiano, the Kunming Wolfdog and the

Czechoslovakian Wolfdog (CSW) and CSW is the most widespread among breeds of wolfdog origin (Hartl & Jedlička 2002; Caniglia et al. 2018).

We decided to study CSW as a model breed that could bring more light in our understanding of hybridisation events due to the fact it has originated from artificial crossing of Carpathian wolves and German Shepherd dogs. Using different molecular markers, we wanted to compare genetic diversity of studied populations and to distinguish genome areas connected to the ancestor populations. Whereas the CSW breed is still young and not very numerous we believe the information obtained from our research evaluating CSW breed from genetic point of view will lead to maintaining the viability of the breed.

2 Literature review

2.1 Domestication of canids

Domestication is an ongoing process that has attracted many researchers for centuries. Domestication studies are based on archaeological and DNA/RNA sequencing technology that has been improved enormously in the last two decades. The results allow researchers to describe and understand the process leading to animals and plants domestication (Larson et al. 2012).

Domestication is an observable example of human induced evolution (Pilot et al. 2021). Process of domestication has been a complex and multi-staged event affected and controlled by human, leading to disparate appearances and behaviours of domesticates compared to their wild ancestors (Larson & Burger 2013; Pendleton et al. 2018). Studies from late nineties were discussing dogs' ancestry. Some opinions suggested that also other canids, for example jackal or coyote, might have influenced the formation of a dog (Wayne 1993). However, the use and rapid development of molecular techniques help to investigate phylogenetic relationships among species (Wayne 1993). The wolf (*Canis lupus*) was confirmed to be the only ancestor of modern dogs (Vilà et al. 1997; Perri et al. 2021).

Even though that the place of dogs´ domestication remains controversial, the study published by Verginelli et al. (2005) range from multiple dog-founding events to a single origin in East Asia. Other results using genomic SNP data show that Middle Eastern wolves were a determining source of genome diversity, although interbreeding with local wolf populations clearly occurred elsewhere in the early history of specific lineages (VonHoldt et al. 2010). On the other hand, scenario based on genetic data analyses published by Frantz et al. (2016) supports the opinion that domestication is a complex process where dogs continued to interbreed with wild wolves and the authors propose that there were at least two domestication centres. Study of Frantz et al. (2016) is based on DNA sequences of ancient dog discovered in Ireland and dated to approximately 4,800 years before present. Its genome was compared to 80 modern available full genome sequences and 605 modern dogs (including village dogs and 48 breeds). Results of this study demonstrate the split of dogs into two groups with obvious geographic pattern; Western Eurasian and East Asian dog populations (Figure 1). However, few breeds (for example Greenland sledge dogs or the Siberian husky) were poorly supported to accurate group, suggesting that these breeds probably possess mixed ancestry from both Western Eurasian and East Asian dog lineages (Frantz et al. 2016). The history of the origin of these two breeds (Greenland sledge dogs and the Siberian husky) was investigated in study of Ramos-Madrigal et al. (2021). Four Pleistocene canid samples (two sculls, humerus and puppy) were found in Siberia. Obtained sequences were mapped to the wolf reference genome. The results suggest

that the four samples belong to the extinct wolf lineages that occurred in Northeast Siberia from more than 50,000 to at least 14,100 years ago. These Pleistocene canids contributed to the ancestry of the Siberian husky and Greenland indigenous dog breeds (Wang et al. 2016; Ramos-Madrigal et al. 2021).

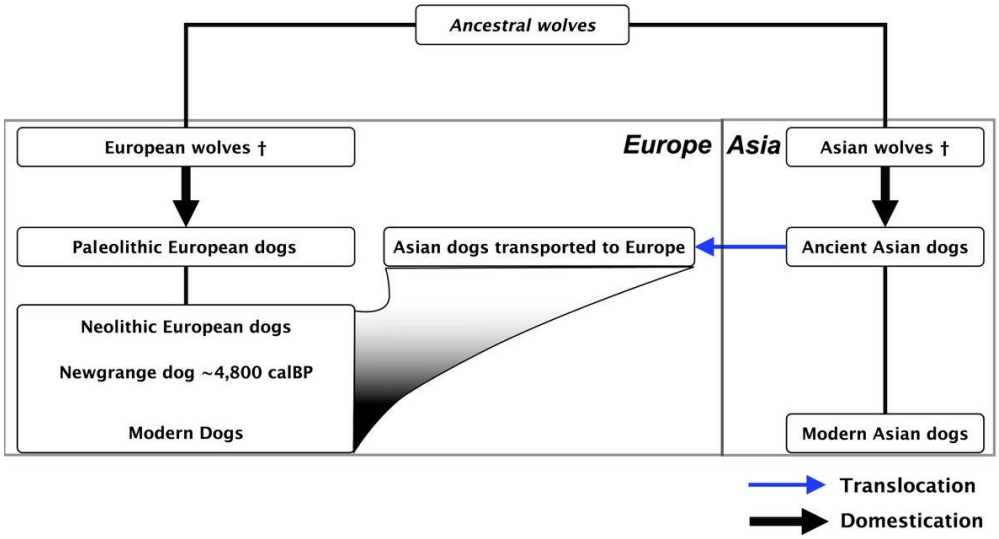


Figure 1: A suggested model of dog domestication under the dual-origin hypothesis. An original wolf population splits into East and West Eurasian wolves that were then domesticated independently before becoming extinct (as indicated by the † symbol) (Frantz et al. 2016). Number of different selective pressures associated with the domestication process is appertained to be the result of genetic responses to domestication, and might be linked to some other directly selected traits. In mammals, this so called "domestication syndrome" features lop ears, mottled coats, decreases in brain size, and changes in developmental rates all traits that may all be linked to strong selection for lowered reactivity to external stimuli (Wilkins et al. 2014; Zeder 2015)**.**

Not only the geographic location of domestication but also the timing of domestication remains controversial. Skoglund et al. (2015) found out that less than two-thirds of all loci in tested genomes (boxer and poodle) are more similar to each other than to a modern grey wolf. They assumed complete isolation without gene flow and the divergence time between grey wolves and modern European dogs dated to 3,500 generations before the present, corresponding with approximately 10,000 years ago (Skoglund et al. 2015). However, the most of the studies support the period of Pleistocene to be connected with the domestication of dogs. For example the study published by Lahtinen et al. (2021) describes the phenomenon of differences between dietary constraints of wolves and humans enabling dog domestication in rugged environments in the Late Pleistocene and at the same time suggests that the initial wolf domestication corresponds with the comparatively short glacial maximum at the last period of the ice age (Last Glacial Maximum (LGM), corresponding with Marine

Isotope Stage 2; see Figure 3), in that time the global ice volume reached its maximum (Lahtinen et al. 2021).

Study of Plassais et al. (2022) focused on the functional mutation associated with IGF1 gene and its role of an ancestral IGF1 allele in the propagation of modern canids. More than 1,400 genome sequences from 13 canid species, including both ancient and modern canids, were used in this study. A single variant in an antisense long non-coding RNA (IGF1-AS) was identified. This variant interacts with the IGF1 gene and together they create a duplex. The Plassais´s team of researchers focused their study on the new candidate SNP, rs22397284, that was identified to be associated with body size variation in both dogs and the other canid species. Seventy-five percent of sampled domestic dogs were homozygous for the C allele of rs22397284 and had a breed body size average < 15 kg. On the other hand, 75% of dogs homozygous for the T allele had a breed body mass average > 25kg. Trends of these results were confirmed in different poodle varieties (giant, standard, miniature and toy) and in three distinct schnauzer breeds (miniatures, standards and giants). Dog breeds of medium size and weight were showing heterozygosity (CT) in this mutation. The same trend was investigated in the body mass of many of the archaeological dogs (Bergström et al. 2020) and it correlated with the homozygosity of the C or T allele as well as heterozygous allele (Figure 2) (Plassais et al. 2022). Distribution of different body mass in ancient dogs correlates with Bergmann's rule. It means that populations and species of small size live in comparatively warmer climates while larger species and populations tend to live in colder climates (Gohli & Voje 2016).

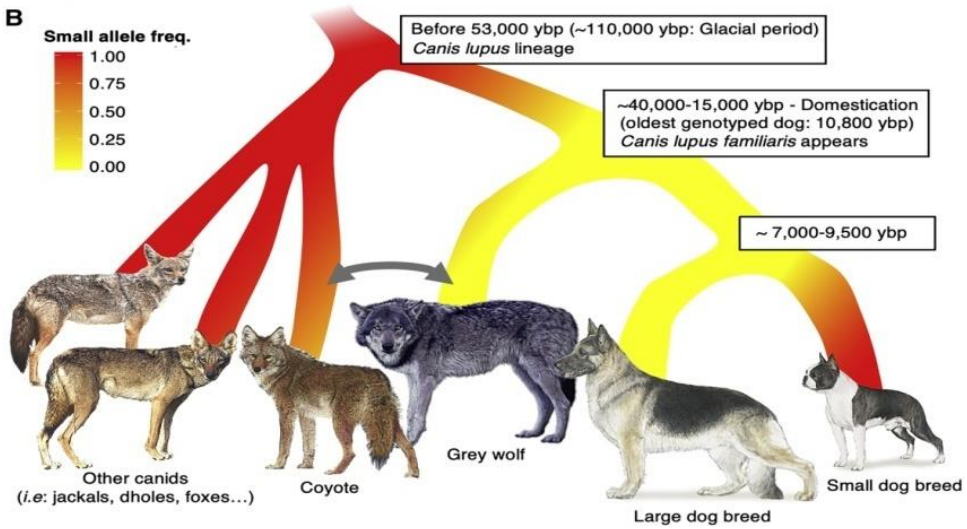


Figure 2: Canidae ancestor was likely small and carried the C allele (CFRNASEQ_AS_00037987). The Tallele (CFRNASEQ_AS_00037985) arose before 53,000 years before present (ybp) and generated bigger animals of wolves (*Canis lupus*). The ancestral C allele continues to exist in the grey wolf population, albeit at

a low frequency. Approximately 15,000 ybp, canine domestication likely began with large wolf-like dogs. Shortly thereafter, human selection of small canids with the ancestral C allele led to a preponderance of small modern domestic breeds. Grey arrow reflects actual hybridization observed between coyotes and wolves in eastern part of America.

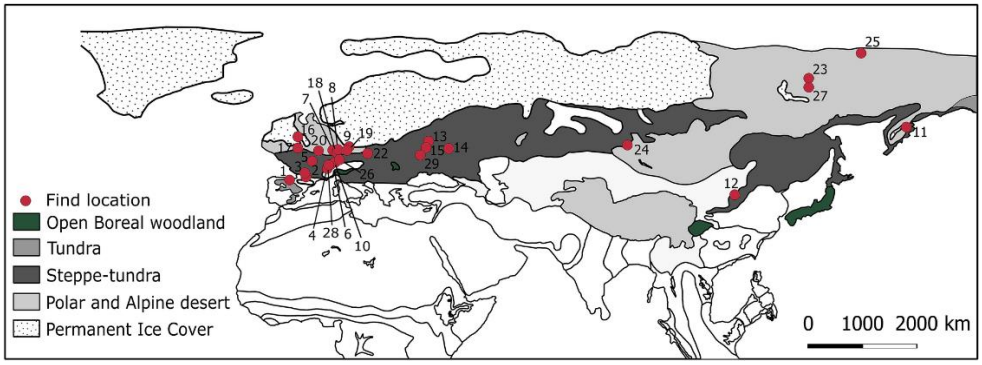


Figure 3: A map of vegetation zones during the Last Glacial Maximum and Palaeolithic dog remain discoveries (Lahtinen et al. 2021).

Study of Sundman et al. (2020) focuses on DNA methylation in brain (a piece of tissue from the medial prefrontal cortex of the left cerebral hemisphere), the most relevant tissue with respect to behaviour between wolf and dog breeds. To identify differences in methylation patterns, a team of Sundman used a novel approach combining genotyping by sequencing (GBS) and methylated DNA immunoprecipitation (MeDIP), which has already been used in chickens (*Gallus gallus*)(Pértille et al. 2017), therefore the method was optimized for wolf-dog study. However, this study investigates how epigenetic factors affect gene expression without altering the DNA sequence and because this process is dynamic, it can allow for plastic and adaptive responses to changes or challenges in the environment. Distinctive differences in DNA methylation in the brain of domesticated dogs and their ancestor species, the grey wolf, and between different breeds of dogs (beagles, boxers, German Shepherd dogs, Great Danes, Labrador retrievers, Pitbull terriers and Rottweilers), which reflects a more recent selection, were explored in this study. The results suggest that epigenetics has played an important role in the divergent selection during dog domestication and breed formation (Sundman et al. 2020).

Domestication of dogs is connected with human evolution and their mutual spread. It means they colonized new areas accordingly. The archaeological evidence documented presence of dogs in the Americas by at least 10,000 years ago. It isimplicatedthat dogs accompanied the early human groups who moved from northeast Asia across the Bering Land Bridge (Beringia) into the Americas (Perri et al. 2021).

2.2 Hybridization between dogs and wolves

Hybridization among canids and especially between dogs and wolves has interested scientists for decades. Development of sequencing methods allows the scientists to investigate how important role the gene flow between diverging lineages has played in evolution (Taylor & Larson 2019; Wang et al. 2020).Natural hybridization is frequently associated with several positive evolutionary outcomes (e.g., genetic rescue described in Brennan et al. 2014) (Donfrancesco et al. 2019). Nevertheless, understanding the processes that lead to hybridization of wolves and dogs is of scientific and management importance (Harmoinen 2020).

2.2.1 Ancient hybridization between dogs and wolves

Study of VonHoldt et al. (2016) looked into the ancient canids (wolves and coyotes) populations in the area of North America. They found that all North American wolves (*Canis lupus*) and coyotes (*Canislatrans*) have significant amounts of coyote ancestry. On the top of that, there was a strong correlation between the portion of coyote ancestry in wolf populations and the geographical latitude. Data also showed that genomes of eastern and red wolf (*Canis lupus lycaon* and *Canis lupus rufus*, respectively) contain significant contributions from grey wolves and coyotes to their ancestry and may be of hybrid origin. As expected, Eurasian wolves and dogs, which are allopatric to coyotes, do not have coyote ancestry (VonHoldt et al. 2016).

A study of Wang et al. (2020) focused on the canids of Tibetan region where origin and evolutionary relationships among the most enigmatic canid lineages remain unresolved. These are considered to be the high-altitude wolves (HAWs) of the Qinghai-Tibet Plateau (QTP). These HAWs are often divided into two groups, the Tibetan grey wolf (TW) and the Himalayan wolf (HW). HWs are distributed across the Trans-Himalayan region of Nepal, northern India in the Ladakh region of eastern Kashmir, Himachal Pradesh, and neighbouring regions, whereas TWs are found in the provinces of Gansu, Qinghai, and Tibet. Both of these wolf populations are critically small and vulnerable and HWs are currently protected by wildlife legislation in India and Nepal (Wang et al. 2020). The morphological work, already published in 1847 by Hodgson, described that the HAWs are distinct from the Eurasian wolves and are considered to be a distinct subspecies (Hodgson 1847), however, their relationships to other grey wolves still remain unclear. The comparison using mtDNA showed that HW cluster with TW in a clade that is distinct from other grey wolves (Werhahn et al. 2020). Analyses done by Fan et al. (2016) used whole genomes data and TW clustered with grey wolves from northern China in a clade that is sister to European wolves. This might be a consequence of an event around 25 thousand years ago when TW was derived from a lineage of Asian wolves and recolonized the QTP after divergence from North Chinese wolves (Fan et al. 2016). With the aim to find out the relationship

between TWs and HWs, a unique EPAS1 allele was examined. Nuclear genomes of both species, HWs and TWs were sequenced and compared with published genomes of Chinese wolves, dogs, and other canids including the golden jackal and the dhole. It was found out that TW and HW are closely related, and that both derived from admixture with Eurasian grey wolves, domestic dogs, and a now-extinct or unknown lineage of wolf-like canids. However, the team of Wang (2020) explored specifically the evolutionary origin of the EPAS1 allele, which was investigated to be shared by these lineages living on the Qinghai-Tibet Plateau and probably derived from the deeply divergent previously undescribed canid lineage (Wang et al. 2020).

The coat colour correlates with the ancient hybridization among canids in North America. There are coat colour frequencies differences that are related to whether open and forested types of habitat. Dark or black coat colour is in a relationship with *K* locus which lies on a different chromosome from Agouti and Melanocortin 1 receptor (*Mc1r*) that contribute to the colour and pigmentation in fish, birds, and many mammal species. Several noncoding segments distributed on *CBD103* were sequenced across 32 Arctic and 15 unrelated Yellowstone wolves, as well as in 12 domestic dogs (for example Akita, Basenji, Boxer, Bulldog, Doberman pinscher, Curly-coated retriever, Dalmatian, Great Dane, Labrador retriever, Poodle, and Portuguese water dog) in study of Anderson et al. (2009). The results show, the 3-bp deletion in *CBD103* is associated with black or melanistic coat colour in dogs, wolves as well as in coyotes in North America. From these results three possible evolutionary histories were suggested. First, the 3-bp deletion may be relatively old, having occurred in a canid ancestor more than 1 million years ago before the divergence of coyotes from wolves. On the other hand, the second 3-bp deletion may have occurred more recently in one of the species, followed by introgression into the others. Third, the 3-bp deletion may represent a mutational hotspot, having recurred independently in coyotes, wolves, and dogs (Anderson et al. 2009).

Study of Caniglia et al. (2013) focused on wolf-dog hybridization in Central Italy, Europe. The 3-bp deletion on *β-Defensin* locus together with the control region were used to distinguish wolf-dog hybrids among sampled groups of individuals of uncertain origin. The study confirmed a hybrid origin of the studied pack from Central Italy by dint of the admixture analyses of microsatellite genotypes and the results also showed the melanistic *β-Defensin* deletion in black-coated animals with admixed microsatellite genotypes. Although the results indicate an evident close relation between admixed genotypes and the occurrence of the melanistic K locus deletion, although they cannot definitely prove that the K^B allele entered in the Italian wolf population via hybridization (Caniglia et al. 2013).

2.2.2 Recent hybridization between dogs and wolves in Europe

In few past centuries the habitat loss and direct hunting by humans was a main coincidence of decreasing numbers of wolf populations worldwide especially in Europe (Pilot et al. 2014; de Groot et al. 2016; Hindrikson et al. 2017). Wolves are currently repopulating their original territories from which they had been eradicated by humans in the past (Chapron et al. 2014). Despite the fact that many people, many organizations and also governments of countries try to avoid the introgressive hybridization between domestic dogs (*Canis familiaris*) and wolves *(Canis lupus)*, it still represents a considerable case of anthropogenic hybridization in addition to this phenomenon is increasingly threatening the genomic integrity of wolf populations expanding into human-modified landscapes (Santostasi et al. 2021).

Hybridization between different canid species is perceived as negative. Introgression of alien genes may influence genetic integrity of the species, disrupt species-specific epistatic equilibria and local adaptations or even drive local populations either entire species to the genetic extinction (Brumfield 2010).The wolf-like canids are a closely related group of large carnivores whose chromosomes are stable in morphology and number (2n = 78) (Wayne 1993). Grey wolves and dogs are able to crossbreed. Many cross mating experiments have been done in captivity when males or females of different dog breeds were successfully crossed with wolf parallels, for example crossbreeding experiment run in Czechoslovakia (Hartl & Jedlička 2002), similar experiments were run in Italy (Talenti et al. 2018) or the Netherlands ("The Creation of Leendert Saarloos" 2011). Over the past thirty years the extensive hybridization between wild wolves and feral (free-ranging) dogs has been detected in Europe (Randi & Lucchini 2002; Verardi et al. 2006).

Many studies based on archaeological or genetic data have been published since then. Different studies are based on different molecular markers however all-purpose method leading to detect wolf hybridization is lacking. Analyses using microsatellites data offer only limited resolution due to the low number of markers showing distinctive allele frequencies between wolves and dogs (Harmoinen 2020).

Recent studies prefer using multi methods data analyses where microsatellite, mtDNA and SNPs markers are combined. These data especially coming from SNP markers allow assigning even third-generation backcrosses to wolves to the right category with very high accuracy reaching up to 92% (Harmoinen 2020).

Even though microsatellites offer only limited resolution due to the low number of markers, they show distinctive allele frequencies between wolves and dogs and they are still important markers being used (Harmoinen 2020).

A study of Santostasi et al. (2021) estimated population-wide prevalence of admixture in wild populations affected by anthropogenic hybridization of wolf population from the Appennino Tosco–Emiliano National Park (PNATE), in the

northern Apennines, Italy in years 2015-2016 (August 2015–May 2016). Extracted DNA was amplified at 12 unlinked autosomal microsatellites and a dominant 3-base pair (bp) deletion the b-defensin CBD103 gene described in Anderson et al. (2009). Results of all sampled individuals shared the typical Italian wolf mtDNA. Equally the Kb melanistic deletion was not detected in any sampled individual, however, four individuals shared a dog-derived Y holotype (Santostasi et al. 2021).

Thirty of European countries report stable wolf presence. The occurrence of hybrid individuals is mostly sporadic, but not impossible. Putative hybrid individual detection relies on confirmation through genetic analyses of individuals. Diagnostic tools have been improved and refined during years. Genetic analyses use variable numbers of autosomal microsatellite, small fragments ofmtDNA and SNPs to confirm suspected hybrids (Salvatori et al. 2020).

2.2.3 Anthropogenic hybridization

Hybridization facilitated by human impact and interference, either on purpose or accidentally, is considered to be an anthropogenic hybridization. This phenomenon may affect and result in the elimination of barriers between distinct populations, which may go through the processes of genetic admixing and loss of evolutionary adaptation. The vast majority of scientists and researchers agree that the anthropogenic hybridization is widely perceived as a threat for the conservation of biodiversity. Despite of this fact, relevant policy and management interventions have not been determined and remain highly convoluted (Allendorf et al. 2001; Donfrancesco et al. 2019).

Study of Donfrancesco et al. (2019) was carried out among the scientific community focusing on the wolf-dog anthropogenic hybridization situation across Europe. To receive investigative feedback from the scientific community, an anonymous, repetitious Delphi technique was chosen as the most suitable method. For the purpose of this study a clear distinction was drawn between three categories of free-ranging dogs (Boitani et al. 2007). The complete results of three rounds of the Delphi investigation were obtained from forty-two researchers. The majority of answers agreed on that the anthropogenic hybridization should always be mitigated. The admixed individuals should be always defined according to their genetic profile. Next agreement was about management strategies which should always be aimed at preserving the genetic integrity of the species, ensuring that evolutionary and ecological processes are maintained exempt from anthropogenic interference. When admixed individual is identified, it is not advisable to be managed by hunters or the general public and it is recommended to be managed by formal institutions only (Donfrancesco et al. 2019).

2.3 Czechoslovakian Wolfdog breed

Czechoslovakian Wolfdog (CSW) is very unique dog breed. It is one of few dog breeds originated in the Czech Republic, Czechoslovakia respectively, before year 1989. Dog breeding has got long and rich history and it is very popular even these days in our country.

Although CSW breed is very young, it became very popular not just in Central Europe but worldwide and the number of individuals has been increasing continuously. Nevertheless, the breed has never been studied before from genetic point of view. Therefore, we decided to apply molecular genetic markers to get to know more information about this extremely interesting dog breed.

CSW breed has been under strict breeding control since its creation, it means mating pairs are established and agreed by the main breeding advisor and breeding committee, therefore the results are very important and even practically useful for breeders. Obtained results will help to develop and keep viability of this low numerous breeds.

2.3.1 History of crossbreeding

The main aim of this military project was to select hard-working dogs for military purposes, to guard mountainous borders of the former Czechoslovakia during the Cold War, by improving their health, vitality, endurance and sensory abilities, as night vision (Smetanová et al. 2015).

One-year old female Carpathian wolf (CW), Brita, was received for an experimental crossing between wolf and dog. This female wolf was placed into the kennel Pohraniční stráže in Libějovice in southern Bohemia where whole experiment took place. She was stabled in the coop with kennel. Two male individuals of German shepherd (GS) breed were chosen for mating. They were presumed as good breeding founders. The first male was calm, the second was more aggressive, and both were very well-trained individuals of sable grey colour. In 1957 the mating was not successful because the female wolf was most of the time hidden in the kennel and she did not let the dogs mate with her. Next year, in 1958, the period when the wolf was on heat was assured. Twelfth day of being on heat the calm dog was placed together with the female wolf. She beat him hardly (Hartl & Jedlička 2002). Following day, the aggressive dog (Cézar z Březového háje) was let in the cage and the mating was successful. The mating was repeated every day till 20th day of her period. The gravidity was not well recognizable. Five offspring were born 61st day after first mating. Due to big aggressiveness of the female wolf, breeders were allowed to control and weigh the offspring in the age of 10 days. The difference in weight was -90g compared to GS offspring of the same age (Hartl & Jedlička 2002).

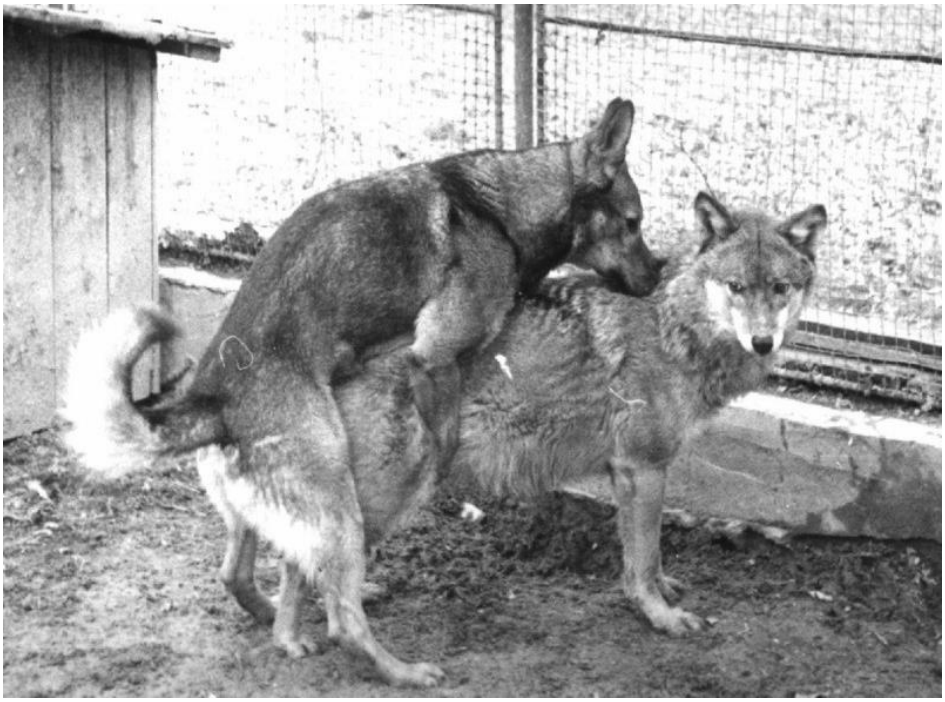


Figure 4: Mating of female wolf Brita and German shepherd Cézar z Březového háje (1958) (Hartl & Jedlička 2002).

The first findings of this experimental crossbreeding detected, that all F1 hybrids were more similar to wolves in phenotype either in behaviour when crossed female wolf and male GS whereas female GS and male wolf were crossed, juveniles were also similar to wolf ancestor, however, higher phenotype variability could be observed in the litter. Wolf was more dominant in both cases of crossing. F1 generation individuals and simultaneously hybrids of the next generations were fertile and could mate with dog and wolf, equally (Hartl & Jedlička 2002). Correlated selection responses lead to differences in phenotype are obtained in only a few generations (Sundman et al. 2020).

Mating was successively repeated and finally four filial generations were born after female wolf Brita and GS male Cézar z Březového háje (Figure 4) and other two filial generations were born after Brita and another GS male Kurt z Václavky (Hartl & Jedlička 2002).

The first attempt for registration of the new developing breed was rejected in that time due to the small number of individuals. In 1968 a new intervened crossbreeding passed in kennel Býchory. Argo, male of CW, mated with GS female Asta z SNB. Request for confirmation of a new breed named Czech wolfdog was again rejected in 1970 and later again in 1976. In that time more than hundred individuals were on duty of Czechoslovakian army and 56 animals were kept by private dog keepers. Breeding was later on moved close to Bratislava where experiments continued. In 1974, CW Šarik, was used for mating. Šarik was crossed with Xela z Pohraniční stráže, this female was already F3 hybrid from previous crossbreeding and with other

female of so called Czech wolfdog, Urta z Pohraniční stráže (Hartl & Jedlička 2002). Nevertheless, backcrosses with dogs provided individuals that were less aggressive in their behaviour and easier to train than F1 and F2 hybrids (Smetanová et al. 2015).

Club of breeders of Czechoslovakian Wolfdog was established 20.3.1982 in Brno. Name of new breed was confirmed there. The breeding program conception was approved. Regrettably Slovakian breeders had broken agreed rules during two following years and the entire 77% of litters were just after one male founder – Rep z Pohraniční stráže. This occasion evocated high relatedness between the majority of population (Hartl & Jedlička 2002).

The trend breeders followed was to select hybrid animals to keep the wolf-like phenotype and the dog-like behaviour, although animals with intermediate wolf-dog phenotypes were not removed in the first phase. Therefore the last official crossing was done in 1983 again in Libějovice (area of Southern Bohemia) when CW female Lejdy and very well trained GS male Bojar von Schotterhof mated (Hartl & Jedlička 2002). Unfortunately, only one of the offspring named Kazan z Pohraniční stráže was useful for service training. The rest of puppies were more similar to wolf in phenotype and also in behaviour – they were shy, hardly socialized, not so useful for people. Kazan was used frequently in breeding process (Šebková et.al.2008) and we might presume the population could be affected by founder effect again.

Currently, three decades after the CSW formal breed creation, the behavioural traits of the breed interest scientists (Maglieri, Sommese). A dog–human communication was tested among Czechoslovakian Wolfdogs, German Shepherds and Labrador Retrievers in study of Maglieri et al. (2019). Human-directed gazing was analysed during the 'solvable task' (in which dogs could obtain the food by manipulating the metal container) and the 'unsolvable task' (in which the container was fixed onto the plywood). German Shepherds and Labrador Retrievers showed a tendency to look back at humans, while Czechoslovakian Wolfdogs was the only group that did not show the tendency to gaze towards humans (Maglieri et al. 2019). A questionnaire examining the behaviour of Czechoslovakian Wolfdogs, German Shepherds and Labrador Retrievers was distributed among owners in Italy and the Czech Republic by Sommese et al. (2021). CSW individuals showed less stranger-directed fear than other two breeds and less non-social fear than German Shepherds. Males showed to be more aggressive than females in CSW breed and trained individuals showed reduced aggressive and separation-related behaviour. Authors conclude that CSW are more similar to ancient breeds (more wolf-like) for some behavioural traits and like modern breeds for other traits (Sommese et al. 2021).

2.3.2 CSW breed creation

In summary, two males and two females of Carpathian wolf were used for new breed creation. Since then, any new hybridization between wild wolves and GS is strictly prohibited (Hartl & Jedlička 2002) although there were few breeders who intended to new crossbreeding between these two populations again. Naděžda Šebková and František Hrach are experienced in artificial crossing between female wolf Lupina (*Canis lupus occidentalis*), born in 1993 in Brno Zoo and German shepherd dog Armin. Three puppies were born to female wolf, one male and two females (Figure 5)(Kutal & Rigg 2008). CSW breeders have said these two females of hybrid origin were included in breeding programme in Italy. This fact might had influenced the purity of CSW breed and cause many problems among breeders themselves.

Figure 5: F1 female hybrids (A, B) born in 2002 to female wolf Lupina (Kutal & Rigg 2008).

Finally, in 1994, the standard of CSW was confirmed in Helsinki. The CSW breed was classified to the 1.FCI (Fédération Cynologique Internationale) group (sheep dogs and cattle dogs) under the number 322. Country of origin is Czechoslovakia, although after the separation of federative republic patronage of the breed belongs to Slovakia ("CESKOSLOVENSKÝ VLCIAK" n.d.).

Annual report exemplifies, the number of CSW sires is 326 (July, 2020) (Čílová n.d.). During the years 2018 and 2019, 115 litters were born and 692 individuals were registered, 383 males and 309 females (Hurda n.d.).

3 Aim of this thesis

Taking the advantage of the known hybrid ancestry of the wolfdogs, the main goal of the study is to characterize in detail the genetic and genomic structure and composition of the Czechoslovakian Wolfdogs and wild wolf-dog hybrids, and to interpret the results in the context of domestication process and nature conservation. The population genetic analyses results may help with the breeding management.

Objectives Study 1: A study comparing different panels of genetic markers recognizing pure wolves, dogs or hybrid individuals in Italy (Appendix 1)

i) To evaluate the power of biparental and uniparental molecular markers;

ii) To identify presumptive wolf x dog hybrids;

Objectives Study 2: A study of genetic composition of Czechoslovakian Wolfdog breed (Appendix 2)

i) To evaluate and characterize, for the first time, the genetic diversity and structure of Czechoslovakian Wolfdog breed using microsatellite loci together with mtDNA and Y-linked loci;

Objectives: A genome study of Czechoslovakian Wolfdogs (Appendix 3)

ii) To identify wolf-derived and dog-derived ancestry of genomic regions using 170k SNPs.;

4 Materials and methods

4.1 Sampling

Three different studies were performed to reach the aims using different types of samples from different locations in Europe.

4.1.1 Ad focus of a study comparing different panels of genetic markers recognizing pure wolves, dogs or hybrid individuals in Italy

We genotyped 271 wolves, dogs and putative hybrids, collected from 1996 to 2011 in Italy, Croatia, the Czech Republic and Slovakia. We collected wolf samples from 3 populations: 1) 63 samples from Italy. These were collected from the entire wolf range in the Apennines (Fabbri et al. 2007); 2) 10 samples were of the western Carpathians origin ranging the eastern part of the Czech Republic and majority of Slovakia and 3) 26 samples were collected in three different Croatian regions (Dalmatia, Gorski kotar and Lika). All wolves had the typical wolf coat colour pattern and no apparent signal of morphological or genetic hybridization (Caniglia et al. 2013). We collected samples from 3 dog groups: 1) village dogs in Italy (DIT; n = 31), sampled from the north and central Apennines and not selected based on their coat colours; 2) an undescribed local dog breed, ‘‘Lupino del Gigante’’, bred in the northern Apennines and phenotypically similar to shepherd dogs, with variable grey, red, black, white and blue merle coats (named ‘‘Apennine dogs’’ in this study; DAP; n = 26); and 3) certified German Shepherd dogs bred in the Czech Republic (DCZ; n = 12). Samples of known or presumed hybrid origin were collected from 2 groups: 1) Czechoslovakian Wolfdogs (WDCZ; n = 73), 2) putative wild-living wolf x dog hybrids collected in Italy (HYIT, n = 30) and identified by their anomalous phenotypic traits (dog-like body shape, coat colour variations, presence of hind-leg spurs or white nails) or previous STR analyses (Randi & Lucchini 2002; Verardi et al. 2006; Caniglia et al. 2013). We obtained the tissue samples from found-dead wolves legally collected by officers on behalf of the Italian Institute for Environmental Protection and Research (ISPRA), the Czech Agency of Nature Conservation and Landscape Protection, and the Biology Department at Faculty of Veterinary Medicine, Zagreb University, Croatia. We obtained additional samples from legally hunted wolves in Croatia, according to quotas defined by the Croatian Commission for monitoring large carnivore populations and approved by the Croatian Ministry for Environmental and Nature Protection. No animal was sacrificed for the purposes of this study. Blood and saliva samples from dogs and Czechoslovakian Wolfdogs were collected by veterinaries that, according to Act 246/1992, sampled only animals in healthy conditions with permission and assistance of the owners and with all the possible efforts to minimise stress. We stored tissue and

blood samples at 220uC in 10 volumes of 95% ethanol, or in 2 volumes of a Tris/ SDS buffer, respectively. Saliva samples were dry-stored.

4.1.2 Ad focus of genetic composition of CSW study

Different types of samples, both invasive and non-invasive, were used in the study of genetic composition of CSWs. Sample collection started in 2010. In total, 79 individuals of CSW, 20 individuals of GS and 28 individuals of CW were analysed in the study evaluating and characterizing the CSW breed. Non-invasive cheek swab samples of CSW and GS were collected during dog shows in the Czech Republic. Only animals in healthy condition with permission and assistance of the owners were sampled, with every effort made to minimize their stress. Only one individual per litter was analysed to avoid biases in genetic variability measurements. Wolf samples consisted of25 non-invasive stool samples collected in the western Carpathians (N= 22), or obtained from the Prague Zoo (N= 3), and three tissue samples from Slovakia. Wolf stool samples were collected by Friends of the Earth organization (FoE CZ), which has been monitoring the wolf population in the Carpathian Mountains. There are no restrictions for the use of stool samples in the Czech Republic. In Slovakia, FoE CZ has permission to collect non-invasive samples of wolves, issued by Regional Office Trenčín, Department of Environment, No. OU-TN-OSZP1-2014/49/3475. The three tissue samples were derived from wolves that were legally culled during the open hunting season (November 1st–January 15th) in Slovakia within a quota set by the local authorities, in conformity with regulation No. 344/2009 Coll. The wolves were shot during individual patrols or collective hunts. The use of poisoned bait or leg-hold traps is strictly forbidden according to hunting law. All hunters had permission for hunting, and we confirmed that the culls were reported before quota fulfilment. No animals were sacrificed for the purposes of this study. Our laboratory has approval (No CZ 11712934) to storage and use of animal material according to § 48(1)(i) of Act No 166/1999 concerning veterinary care and amending certain related laws, as amended, pursuant to Article 17(1) of Regulation of the European Parliament and of the Council (EC) No 169/2009 and Article 27(1) of Commission Regulation (EU) No 142/2011. Tissue and stool samples were stored at −20°C in 10 volumes of 96% ethanol. Cheek swab samples were dry-stored.

4.1.3 Ad focus of a genome study of Czechoslovakian Wolfdogs

Blood samples of 12 unrelated CSWs and muscular tissue samples of 12 unrelated Carpathian wolves were used for the study mapping areas of CSWs´ genome. CSW blood samples were collected from 2003 to 2013 in the Czech Republic by veterinaries, from animals in healthy conditions, with the permission and assistance of the owners, minimizing any possible stress. No animal was sacrificed for the purposes

of this study. The dog owners also authorised the genetic data obtained from their animals to be used in this study, while maintaining their identity confidential. However, two owners did not give their permission to use the pedigree data associated to their dogs, therefore the individual pedigree-based analyses were based upon the 10 remaining CWDs. Wolf tissue samples were collected from eight Western Ukrainian, three Slovakian and one Polish wolves (Stronen et al. 2013), randomly sampled from different packs in order to avoid inbreeding or sampling bias and to be as much as possible representative of the Carpathian population. Tissues were collected, for purposes other than this project, from animals found dead or legally harvested by hunters with special permission under legal hunting quota limits. No ethics permit was required since wolf sample collection involved only dead animals. All samples were collected by specialized technician personnel.

4.2 DNA extraction

Genomic DNeasy Tissue Kit (Qiagen Inc., Hilden, Germany) was used to extract genomic DNA from tissue and PrestoTM Buccal Swab DNA Extraction Kit (Geneaid Biotech Ltd) was used to extract DNA from cheek swab samples. In the last step, genomic DNA was eluted to 100 μl of elution buffer in both cases. DNA from faeces was extracted using QIAamp Stool Mini Kit (Qiagen Inc, Hilden, Germany). DNA from faecal samples was extracted, amplified, and genotyped in separate rooms reserved for low-template DNA samples, under sterile ultraviolet laminar flow hoods, following a multiple-tube protocol including both negative and positive controls. Genomic DNA from samples was extracted according to the particular manuals.

The concentration of extracted DNA samples was measured on NanoDrop™ 2000/2000c spectrophotometers (Thermo Fisher). DNA from all used CSWs´, GSs´ and wolves´samples was extracted and measured in laboratory of Molecular genetics, FTA, ČZU.

4.3 Markers and genotyping

4.3.1 Ad focus of a study comparing different panels of genetic markers recognizing pure wolves, dogs or hybrid individuals in Italy

The pilot study where CSW genotypes were used to detect the hybridization between wild and domestic canids used variable panels of STR markers ranging from 12, 24 up to 39 (Table 1).

Fluorescently labelled primers, mainly dinucleotides and minority of tetranucleotides, were amplified in PCR (Polymerase Chain Reaction) using Bio-Rad thermal cycler T100. 12 STRs were originally used in a 10-year long non-invasive wolf

monitoring project in Italy (Caniglia et al. 2012), 12 STRs were used in a hybridization study focusing on Iberian Peninsula (Godinho et al. 2011) and 15 STRs were from the Finnzymes Canine multiplex kit (Finnzymes, Thermo Scientific Canine GenotypesTM).Amplifications were carried out in 10–20 µl reactions, using 1–2 µl DNA solution (containing c. 20–40 ng/µl of DNA). Negative (no DNA in PCR) and positive (samples with known genotypes) controls were used to detect laboratory contaminations. All samples were independently replicated twice to assess the occurrence of allelic dropout and false alleles. Four Y-STR by Sundqvist (2001) were amplified to identify parental haplotypes.

Total number of 39 autosomal STR markers and 1 sex linked marker *Amelogenin* were amplified in 4 PCR multiplexes using the Qiagen Multiplex PCR Kit (Qiagen, GmbH-Hilden, Germany) (Randi, 2014).

The hypervariable part of the mtDNA CR1 (350 bp) was amplified and sequenced according to Randi et al. (2000). A dominant 3-bp deletion (named KB or CBD103DG23) at the b-defensin CBD103 gene (the K-locus, that is connected to black coat colour) was genotyped following Caniglia et al. (2013).

The amplicons were analysed in an ABI DNA sequencer 3130XL (Applied Biosystems; Foster City, CA), using the software GENEMAPPER 4.0 for STRs and SEQSCAPE 2.5 for sequences. The mtDNA sequences were aligned using CLUSTAL W (Julie D.Thompson 2008) in BIOEDIT (Hall 1999). Identical haplotypes were collapsed using DNASP 5 (Librado & Rozas 2009) and blasted in GenBank. Allele binning and check for null STR alleles were performed in MICROCHECKER(Van Oosterhout et al. 2004) with an adjusted *P* value corresponding to α= 0.05 after Bonferroni correction (Rice 1989). The power ofthe STRs to identify each unique genotype was evaluated calculating the probability-of-identity values (PID and PIDsibs; (Waits et al. 2001)) in GENALEX 6.41 (Peakall & Smouse 2012).

Table 1: List of microsatellite markers

Locus	Chromosome	STR repeat size	Allele sizes (bp)	Dye label
AHTk211	CFA26	Dinucleotide	79-101	FAM
CXX279	CFA22	Dinucleotide	109-133	FAM
REN169O18	CFA29	Dinucleotide	150-170	FAM
INU055	CFA10	Dinucleotide	190-216	FAM
REN54P11	CFA18	Dinucleotide	222-244	FAM
AHT137	CFA11	Dinucleotide	126-156	HEX
REN169D01	CFA14	Dinucleotide	199-221	HEX
AHTh260	CFA16	Dinucleotide	230-254	HEX
AHTk253	CFA23	Dinucleotide	277-297	HEX
INU005	CFA33	Dinucleotide	102-136	NED

Locus	Chromosome	STR repeat size	Allele sizes (bp)	Dye label
INU030	CFA12	Dinucleotide	139-157	NED
FH2848	CFA2	Dinucleotide	222-244	NED
REN162C04	CFA7	Dinucleotide	192-212	PET
AHTh171	CFA6	Dinucleotide	215-239	PET
REN247M23	CFA15	Dinucleotide	258-282	PET
FH2004	CFA11	Tetranucleotide	104-202	PET
FH2088	CFA15	Dinucleotide	91-139	FAM
FH2096	CFA11	Tetranucleotide	86-110	HEX
FH2137	CFA3	Dinucleotide	140-192	HEX
CPH2	CFA32	Dinucleotide	88-106	NED
CPH8	CFA13	Dinucleotide	191-219	FAM
FH2079	CFA24	Tetranucleotide	246-282	FAM
CPH4	CFA15	Dinucleotide	130-155	NED
CPH5	CFA15	Dinucleotide	102-124	HEX
CPH12	CFA8	Dinucleotide	188-214	FAM
C09.250	CFA9	Dinucleotide	121-145	PET
C20.253	CFA20	Dinucleotide	90-120	NED
AHT132	CFA2	Dinucleotide	160-172	PET
C27.442	CFA27	Dinucleotide	158-172	HEX
FH2010	CFA24	Tetranucleotide	216-240	NED
PEZ1	CFA7	Tetranucleotide	99-131	HEX
PEZ5	CFA12	Tetranucleotide	95-119	PET
AHT103	CFA4	Dinucleotide	71-89	HEX
AHT111	CFA2	Dinucleotide	72-92	NED
FH2001	CFA23	Tetranucleotide	123-155	PET
C09.173	CFA9	Dinucleotide	100-118	FAM
C13.758	CFA13	Dinucleotide	220-244	NED
CPH9	CFA28	Dinucleotide	139-151	HEX
CPH14	CFA5	Dinucleotide	185-205	PET
Y-linked markers				
MSY34A	CFAY	Dinucleotide	160-190	NED
MSY41A	CFAY	Dinucleotide	90-150	HEX
MSY34B	CFAY	Dinucleotide	167-177	HEX
MSY41B	CFAY	Dinucleotide	109-137	NED
Sex-linked marker				
Amelogenin	CFAX	-	174-218	NED
K-locus	CFA16	Codon deletion	147-151	HEX

4.3.2 Ad focus of genetic composition of CSW study

The study of genetic variability of CSW breed was based on several genetic markers. Samples were thus genotyped based on the following: i) the Amelogenin gene (to sex individuals); ii) 39 autosomal microsatellites (to reconstruct individual genetic profiles) and four sex-linked microsatellites (Y-STR, to identify paternal haplotypes); and iii) the hypervariable part of the mtDNA control region (to determine maternal haplotypes). Genotyping of the Amelogenin gene, 39 autosomal and four Y-linked microsatellites was performed as described in Randi et al. (2014). Amplifications were replicated twice for tissue and salivary samples and from four to eight times for faecal material. Allele sizes were manually scored in GeneMarker v.1.85 (www. softgenetics.com) and binned using raw size in Autobin (http://www4.bordeaux-aquitaine. inra.fr/biogeco/Ressources/Logiciels/Autobin).

Genotyping errors such as large alleles dropout, stuttering or null alleles were tested through Markov chain Monte Carlo (MCMC) simulations of expected allele-size differences using 1000 randomizations in Micro-Checker (Van Oosterhout et al. 2004). The hypervariable domain of the mtDNA control region was amplified using polymerase chain reaction (PCR) according to Vilà et al. (1999). Sequences were aligned using CLUSTALW (Thompson et al. 2008) implemented in BIOEDIT [15]. Identical haplotypes were collapsed in DNASP 5 (Librado & Rozas 2009) and were compared with the GenBank database using the megablast algorithm.

4.3.3 Ad focus of a genome study of Czechoslovakian Wolfdogs

CWD and Carpathian wolf DNA samples were genotyped at c. 170k SNPs using the CanineHD BeadChip microarray (Illumina, Inc., San Diego, California, USA), following the Infinium HD Ultra Assay protocol and calling genotypes with GenomeStudio (http://www.illumina.com/documents/products/datasheets/datasheet_ genomestudio_software.pdf).

For comparative purposes, we then added publicly available genotypes from 355 dogs belonging to 30 breeds that were genotyped with the same 170k SNP microarray in the LUPA project, realized for the genetic mapping of a number of canine diseases (Lequarré et al. 2011; Vaysse et al. 2011). In particular, this dataset included also 12 German Shepherds that, thanks to their limited within-breed variation (Vaysse et al. 2011) and stable breeding practices, can represent a very good proxy of the original dog founders of the Czechoslovakian Wolfdog breed.

4.4 Molecular data analysis

4.4.1 Ad focus of a study comparing different panels of genetic markers recognizing pure wolves, dogs or hybrid individuals in Italy

The multilocus genotypes determined at 39 STRs in the complete data set (n= 271; 8 sampled groups: DIT, DAP, DCZ, WIT, WHR, WCZ, WDCZ and HYIT) were analyzed in GENALEX to estimate: 1) allele frequency by locus and population, observed (HO) and unbiased expected (UHE) heterozygosity, mean number of alleles per locus (Na) and the number of private alleles per population (Np); 2) AMOVA (analysis of molecular variance (Michalakis & Excoffier 1996) and Weir and Cockerham's average and pair-wise FST values (B. S. WEIR AND C. CLARK COCKERHAM 1984); 3) the frequency distributions of mtDNA CR1 and Y-STR haplotypes, and melanistic KB deletion. GENETIX 4.05 (Belkhir et al. 1996) was used to compute the fixation index FIS and to test for departures from Hardy-Weinberg and linkage equilibrium (HWLE) for each locus and population. A subset of the 24 most discriminating STRs was identified based on FST distances between wolves and dogs, and confirmed in WHICHLOCI analyses (Banks et al. 2003), performed using the ''allele frequency differential'' and the ''whichrun assignment'' methods (Shriver MD et al. 1997, Banks & Eichert 2000). A third marker subset included the 12 STRs used in the monitoring project of the Italian wolf population (Kerns et al. 2004; Candille et al. 2007).

Clustering and assignment testing were performed by: 1) a discriminant analysis of principal components computed by the ADEGENET package (DAPC (Jombart et al. 2010) in R; www.r-project.org), which maximizes the among-group divergence while minimizes the within-group variance, thus improving the discrimination of populations poorly differentiated as compared to standard principal component methods; 2) the Rannala and Mountain's (Rannala & Mountain 1997) assignment method in GENECLASS 2.0 (Piry et al. 2004); 3) the Bayesian clustering model (assuming HWLE in the genetic clusters) implemented in STRUCTURE 2.3 (Falush et al. 2003); 4) a non-Bayesian clustering procedure (that does not assume HWLE in the clusters) implemented in FLOCK 2.0 (Duchesne & Turgeon 2009). First, we used STRUCTURE to infer the optimal partition of the 8 sampled groups, assuming K from 1 to 12, with 2 independent runs for each K with 400 000 MCMC and discarding the first 40 000 burn-ins, using the ''admixture'' and independent allele frequency ''I'' models, and no prior information (option ''usepopinfo'' not activated). The DK statistics was used to identify the highest rate of increase in the posterior probability LnP(D) of the data between each consecutive K (Evanno et al. 2005). Based on the first STRUCTURE results, admixture analyses were performed again assuming 4 reference groups (DIT, DAP, DCZ and WIT) for the assignment of the putative Italian wolf x dog hybrids (HYIT), using 39, 24

and 12 STRs. STRUCTURE was run with Kfrom 1 to 8, with 400 000 MCMC and 40 000 burn- ins, with the option ‘‘usepopinfo’’ activated or not. In the former case, we assumed that reference wolves and dogs were a priori correctly identified and assigned to their own clusters (popflag= 1), while the putative hybrids were left to be assigned (popflag= 0). The estimated allele frequencies of the wolf and dog reference clusters were not affected by the allele frequencies of the other samples (option updatepfrompopflagonly activated). The software FLOCK implements a non-Bayesian clustering algorithm based on reiterated allocations that promises efficient partitioning of the admixed samples in groups of homogeneous genotypes, also if putative parental populations are not sampled, independent of any genetic model (i.e., HWLE is not assumed). FLOCK was used to partition samples DIT, DAP, DCZ, WIT and HYIT, with reference groups varying from 1 to 8, initial random choice of samples, 50 runs and 20 re-allocations per run (LOD threshold for allocation to reference groups= 0). Admixture inference may be difficult when model assumptions are not met and if small numbers ofmarkers are used (but also if the number of loci is large; (Corander & Marttinen 2006). For instance, when an unknown number of K parental populations must be inferred simultaneously to the admixture coefficients, both overfitting (too large K values) and false admixtures may results, particularly if the sampled populations diverged moderately (FST ,0.10; (Corander et al. 2008). Hence, false positives (error type I), namely individuals with false admixed ancestry, might arose by chance. In this study, we explored the risk of false admixtures using BAPS (Corander et al. 2008), which produces null distributions for the admixture expected by chance that are used to identify significant admixtures at a given p-value (Almudevar 2000). The power of the 39, 24 and 12 STRs to correctly detect a priori known parentals, hybrids and backcrosses was determined by simulations using HYBRIDLAB (Nielsen et al. 2006). We randomly selected 60 reference wolves and 60 reference dogs among WIT and DIT to generate 60 simulated genotypes in each of the following classes: first (F1) and second (F2) generation hybrids, first (BC1W, BC1D), second (BC2W, BC2D) and third (BC3W, BC3D) generation backcrosses with wolves and dogs, respectively. The simulated genotypes were then analyzed in STRUCTURE with the ‘‘admixture’’ and the ‘‘I’’ models, without prior population information. The proportion of individuals correctly assigned to each class led to define the appropriate threshold value to use in the admixture analyses. The software NEWHYBRIDS 1.1 (Anderson & Thompson 2002) was used to compute the posterior probability that each genotype belongs to each of the following 6 classes: wolf (W) and dog (D) parentals, F1 and F2, backcrosses of F1 with dogs (BC1D) and with wolves (BC1W). Posterior distributions were evaluated after 105 iterations of the Monte Carlo Markov chains, following a burn-in period of 104 iterations, without using any individual or allele frequency prior information, with

‘‘Jeffreys-like’’ or ‘‘Uniform’’ priors for mixing both proportions and allele frequencies.

4.4.2 Ad focus of genetic composition of CSW study

Genetic diversity measurements such as the mean number of different alleles per locus *(N_A)*, mean number of effective alleles per locus *(N_E)*, expected *(H_E)* and observed *(H_O)* heterozygosity, estimations of the inbreeding coefficient *(F_{IS})* and allelic richness with correction to equal sample size *(A_R)* were computed in FSTAT 2.9.3.2 [17]. FSTAT uses rarefaction to standardize sample size of allelic richness to the *N*ofthe smallest group in the data set, which is 20 in this study. The number of private alleles *(N_P)* was determined in GenAlEx 6.5 (Peakall & Smouse 2012). Deviations from Hardy-Weinberg equilibrium (HWE) and Linkage Equilibrium (LE) were tested in Gene- Pop 4.0 (Rousset 2008), using exact tests and MCMC simulations with 100 batches of1000 iterations. Factorial correspondence analysis (FCA) was performed in Genetix 4.05.2 (Belkihr et al. 2004). The Bayesian clustering method (Falush et al. 2003) implemented in the program STRUCTURE 2.3 (Pritchard et al. 2000) was used with an admixture model and correlated allele frequencies to detect substructure in the data, assign individuals to clusters and identify potentially admixed genotypes. The optimal number of clusters *(K)* was set by running the program from *K*=1 to *K*= 5, with 10 repetitions of 1,000,000 MCMC chain steps after a burn-in period of 100,000 steps for each *K*. STRUCTURE results were visualized in STRUCTURE HARVESTER (Earl & vonHoldt 2012) implementing the method of Evanno et al. (2005). Graphical output was performed in DISTRUCT 1.1 (Rosenberg 2004). Contemporary y effective population size *(N_e)* and 95% confidence intervals *(CI)* for CSW were estimated using, as single-sample estimator, a bias-corrected version of the linkage disequilibrium method (Waples & Do 2008) as implemented in the software NeEstimator v.2.0 (Do et al. 2014). This method uses multilocus diploid genotypes from a given population to obtain precise estimates of N_e with non-overlapping generations by using 10–20 microsatellite loci (5–10 alleles/locus) and samples of at least 25–50 individuals, if the effective population size is less than approximately 500 (Waples & Do 2010). NeEstimator was run using the 79 CSW and considering a P_{Crit} value (for screening out rare alleles) of 0.02, which was recommended as the value ensuring the most precise and less biased results when working with microsatellites (Do et al. 2014).

4.4.3 Ad focus of a genome study of Czechoslovakian Wolfdogs

4.4.3.1 Data Filtering

The genotypes from these 379 individuals were filtered in the SNP&Variant Suite 8.0.1 (SVS, Golden Helix Inc., Bozeman, MT) discarding samples and SNPs with callrates ≤ 95% and all loci mapping on chromosomes X and Y (quality-pruned dataset). Genotypes were further filtered to discard loci in linkage disequilibrium (LD) by PLINK 1.07 (Purcell et al. 2007), using the dog option in order to manage the correct number of chromosomes and removing SNPs with pairwise genotypic associations r2 > 0.2 calculated along sliding windows of 50 SNPs (LD-pruned dataset).

4.4.3.2 Summary statistics, assignment and admixture tests

A pairwise F_{ST} matrix of genetic distance (B. S. WEIR AND C. CLARK COCKERHAM 1984) among groups, values of observed heterozygosity (H_o) and the inbreeding coefficient (F) within groups were estimated from the quality-pruned dataset in SVS. To visualize the distribution of genotypes in the genetic space, an exploratory principal component analysis (PCA; (Novembre & Stephens 2008)) was performed in SVS using the quality-pruned dataset and the additive genetic model (Price et al. 2006). We then ran assignment tests in ADMIXTURE 1.23(Alexander et al. 2009) on the LD-pruned dataset of CSWs, Carpathian wolves and German Shepherds, assuming K values from 1 to 5, to assign each sample to its population of origin and to evaluate the level of admixture in CSWs. The most likely number of clusters was identified based on the lowest cross-validation error (Alexander et al. 2009) and results were plotted in R 3.0.2 (www.r-project.org).

A more accurate reconstruction of the parental pro- portions of ancestry in CSWs was achieved by the PCA-based admixture deconvolution approach implemented in PCADMIX 1.0 (Brisbin et al. 2012; Lawson et al. 2018), which was run with blocks of 10 consecutive, non-overlapping SNPs. For each CSW, we calculated the average genome-wide proportion of blocks assigned to each reference population. We then compared it to the percentage of wolf ancestry estimated from the CSW pedigrees with the software BREEDMATE PEDIGREE EXPLORER (www.breedmate.com).

5 Results

5.1 Ad focus of a study comparing different panels of genetic markers recognizing pure wolves, dogs or hybrid individuals in Italy

The 39 STRs were polymorphic, except CPH5 and PEZ5 (monomorphic in dogs from The Czech Republic - DCZ), showing from 5 (at locus FH2096) to 23 (at locus FH2137) alleles per locus. The number of alleles and private alleles was higher in dogs than in wolves (Table Add1-1). Heterozygosity varied from Ho=0.46 - UHe= 0.48 in WIT to Ho= 0.69 - UHe= 0.71 in DIT. DIT, WIT, HYIT and WHR were not in HWE, showing significantly positive FIS values. We found no null and false alleles, and no occurrence of allelic dropout. Values of PID and PIDsibs were very low, and all genotypes were unique (Table Add1-1). The proportions of significant pairwise correlations among loci were low (from 0.7% in DCZ to 3.0% in WCZ), indicating no departures from LE. We found a total of 17 Y-STR haplotypes (Table Add1-2). WHR and DIT were the most variable groups. HYIT showed 2 haplotypes (YH17 and YH26) shared with WIT, plus haplotype YH5, shared with dogs and WDCZ, and the private haplotype YH32. There were 19 mtDNA CR1 haplotypes in total (Table Add1-3). DIT had the highest number of haplotypes (8). All WIT had the diagnostic W14 haplotype, that was found in 26/30 (87%) of HYIT. HYIT showed also haplotypes D15 (1), shared with both DIT and DAP, and W16 (3) that was previously identified in Bulgarian wolves (Randi et al. 2000). We detected the KB melanistic deletion only in samples from Italy, with similar KB/K+ heterozygote frequencies in DIT (0.20), DAP (0.31) and HYIT (0.23). Genetic diversity at autosomal and uniparental markers was significantly (P,0.001) partitioned among the 8 groups, with FST(phiPT) = 0.25 (39 STRs), 0.52 (Y-STRs) and 0.49 (mtDNA CR1). Pairwise FST varied deeply among groups (min FST= 0.01 between WIT and HYIT; max FST =0.42 between WIT and DCZ), and among loci (min FST = 0.09 at locus FH2001; max FST= 0.45 at locus U253). The 24 wolf-dog most divergent STRs, identified by both single-locus FST and WHICHLOCI selections, were: C20.253, CPH9, CPH4, RE247M23, CPH12, AHTh260, INU030, AHT103, CPH2, CPH14, AHTk253, C27.442, CPH5, FH2010, AHTk211, AHT132, C13.758, C09.173, AHT111, AHTh171, REN169D01, INU055, FH2848 and AHT137. Wolf-dog average FST computed using 24 STRs (0.31) was higher than with 39 STRs (0.25) or 12 STRs (0.25). A DAPC plot obtained using 39 STRs showed that all groups were sharply distinct except the partially overlapping Italian wolves and hybrids (Fig. Add1-1). Multivariate distances among groups decreased progressively using 39, 24 or 12 STRs, but wolves and dogs were more distant with the most divergent 24 STRs. Two individuals, the most probable F1 and F2 in STRUCTURE and NEWHYBRIDS

analyses (see Table Add1-4), were roughly intermediate between Italian wolves and dogs (see Fig. Add 1-1B).

Table Add 1-1: Genetic variability estimated at 39 autosomal microsatellite loci (STR) and at the KB melanistic deletion on the b-defensin CBD103 gene in the wolf, dog and putative hybrid sampled groups used in this study.

Group*	*n*[a]	Na/Np[b]	Ho[c]	UHe[d]	F_{IS}[e]	%LE[f]	PID[f]	PIDsib[g]	K^{-}/K^{+h}	K^{B}/K^{-h}	K^{B}/K^{Bh}
DIT	31	7.1/35	0.69	0.71	0.07*	2.4	2.0E-39	9.1E-16	21 (0.70)	8 (0.20)	2 (0.07)
DAP	26	5.0/4	0.63	0.64	0.03	2.0	2.5E-31	1.8E-13	17 (0.65)	8 (0.31)	1 (0.04)
DCZ	12	3.3/0	0.50	0.48	−0.03	0.7	5.4E-20	1.8E-09	12 (1.00)	0	0
WIT	63	4.0/4	0.46	0.48	0.06*	1.3	1.1E-21	3.8E-10	63 (1.00)	0	0
WCZ	10	4.1/6	0.58	0.66	0.13*	3.0	7.8E-29	6.2E-13	10 (1.00)	0	0
WHR	26	5.6/18	0.68	0.70	0.03	1.3	4.2E-35	1.0E-14	26 (1.00)	0	0
WDCZ	73	4.7/12	0.54	0.54	0.00	1.6	8.2E-24	4.7E-11	73 (1.00)	0	0
HYIT	30	5.3/5	0.53	0.57	0.08*	1.9	2.6E-26	8.5E-12	23 (0.77)	7 (0.23)	0

*DIT = Village dogs from Italy; DAP = Apennine dogs; DCZ = German Shepherd dogs from Czech and Slovakian republics; WIT = Wolves from Italy; WCZ = Wolves from Carpathian Mountains; WHR = Wolves from Croatia; WDCZ = Czechoslovakian wolfdogs; HYIT = putative wolf *x* dog hybrids from Italy.
[a]*n* = sample size.
[b]Na = average number of alleles per STR locus; Np = total number of private STR alleles in each sampled group.
[c]Ho = average observed heterozygosity over 39 autosomal STRs.
[d]UHe = average expected heterozygosity (unbiased) over 39 autosomal STRs.
[e]F_{IS} = deviation from Hardy-Weinberg equilibrium (* $P<0.01$).
[f]%LE = proportion of significant correlations ($P=0.05$, Bonferroni corrected) among 39×39 pairwise STR comparisons.
[g]PID and PIDsib = Hardy-Weinberg probability-of-identity among unrelated and full sib individuals in the sampled groups, computed using 39 autosomal STRs.
[h]number and frequency (in parenthesis) of genotypes at the *β-defensin CBD103* gene: K^{-}/K^{+} = homozygotes wild-type (no deletion); K^{B}/K^{-} = heterozygotes for the K^{B} melanistic deletion; = K^{B}/K^{B} = homozygotes for the K^{B} melanistic deletion.
doi:10.1371/journal.pone.0086409.t001

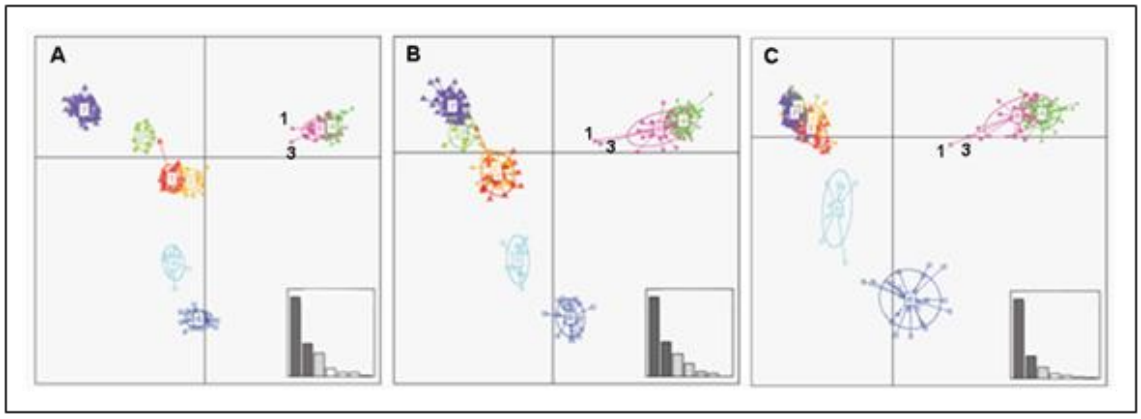


Figure Add 1-1: Discriminant analysis of principal components (DAPC) of wolf, dog and wolf *x* dog hybrids genotyped with 39 (A), 24 (B) and 12 (C) autosomal microsatellites.

(Sampling groups: 1) village dogs sampled in Italy (DIT; *n*=31); 2) "Lupino del Gigante" dogs from Italy (DAP; *n*=26); 3) German Shepherd dogs from Czech Republic (DCZ; *n*=12;); 4) wolves in Italy (WIT; *n*=63); 5) wolves in Czech and Slovak republics (WCZ; *n*=10); 6)wolves in Croatia (WHR; *n*=26); 7) certified Czechoslovakian Wolfdogs (WDCZ; *n*=73); and 8) putative wolf *x* dog hybrids (HYIT; *n*=30) collected in Italy and identified by their anomalous phenotypic traits (dog-like body shape, coat colour variations, presence of hind-leg spurs or white nails), or by previous microsatellite analyses. Black numbers indicate the most probable F1 (sample

n. 1) and F2 (sample n. 3) individuals as determined by Structure and NewHybrids analyses. The first principal component PC I (abscissa) explains 51.48%, 49.96% and 63.65% of the total genetic variance shown by genotypes determined at 39, 24 and 12 microsatellites, respectively. The corresponding second principal component PC II (ordinate) explains 21.25%, 21.93% and 18.19% of the total genetic variance. The inserts (low right corners) indicate the proportion of genetic variability explained by the first 6 eigenvalues.)

Table Add 1-2: Distribution of the Y-linked microsatellite haplotypes in the wolf, dog and putative hybrid sampled groups.

Y haplotype	DIT	DAP	DCZ	WIT	WCZ	WHR	WDCZ	HYIT	S-2001[a]	I-2010[b]
YH01							19		J	H4
YH05	6	14	6				10	2	L	H3
YH06	6									
YH08						3			C	–
YH09						1			–	–
YH11					2	3			G	–
YH16					1	1			K	–
YH17				27				13	–	H1
YH20						6			–	–
YH24	1								–	–
YH26				4				2	Q	H2
YH27	1								–	–
YH28	1								–	–
YH31					2	1			–	–
YH32								1	–	–
YH33						1			I	–
YH34	3	1							–	
Total males	18	15	6	31	5	16	29	18		
Total haplotypes	6	2	1	2	3	7	2	4		
Private haplotypes	4	0	0	0	0	4	1	1		

[a]S-2001 = haplotype identifications as named in Sundqvist et al. [41].
[b]I-2010 = haplotype identifications as named in Iacolina et al. [39].
doi:10.1371/journal.pone.0086409.t002

Table Add 1-3: Distribution of the mtDNA CR1 haplotypes in the wolf, dog and putative hybrid sampled groups.

CR*	DIT	DAP	DCZ	WIT	WCZ	WHR	WDCZ	HYIT
D01[a]	6	3						
D05[a]	1	2						
D06[a]			1					
D08[a]	3							
D09[a]	3	5						
D10[a]	2							
D13[a]			2				47	
D14[a]	13	3	3					
D15[a]	2	13						1
D16[a]	1							
D17[a]			1					
D18[a]			4				22	
H6[b]					2			
H14[b]					6			
W1[c]						13		
W2[c]						1		
W6[c]						12		
W14[a]				63				26
W16[a]								3
Total samples	31	26	11	63	8	26	69	30
Total haplotypes	8	5	5	1	2	3	2	3
Private haplotypes	3	0	2	0	2	3	0	1

*mtDNA CR1 haplotypes identifications as named in Randi et al. [40][a], Pilot et al. [90][b] and Gomerčić et al. [98][c].

doi:10.1371/journal.pone.0086409.t003

Table Add 1-4: Identifications of the 30 putative wolf *x* dog hybrid samples used in this study.

						Structure *popinfo* = 0[g]			Structure *popinfo* = 1[h]			GeneClass[i]			Flock[j]	Baps[k]	NH[l]	
ID[a]	Y[b]	Ph[c]	CR[d]	Y-STR[e]	K[f]	39	24	12	39	24	12	39	24	12	39	39	39	ID[m]
1	1997	Dog-like	D15	YH17	+	0.439	0.472	0.405	0.448	0.468	0.382	99.9	100	98.4	HY	HY	F1	F1
2	1999	Spur	W14	YH17	+	0.690	0.700	0.914	0.684	0.706	0.866	100	100	98.7	HY	HY	F2-BC	F2
3	2007	unknown	W14	*female*	B	0.625	0.459	0.602	0.629	0.470	0.600	100	100	100	HY	HY	F2	F2
4	1999	Spur; dark coat	W14	YH17	B	0.886	0.934	0.960	0.860	0.890	0.862	99.9	99.8	52.7	HY	W	BC	BC
5	2001	White nails	W14	*female*	+	0.760	0.833	0.957	0.759	0.811	0.853	100	100	99.1	HY	HY	BC	BC
6	2006	White nails	W14	*female*	+	0.833	0.900	0.959	0.818	0.842	0.884	100	100	95.1	HY	HY	BC	BC
7	2011	Brown coat	W14	YH17	+	0.855	0.809	0.839	0.846	0.798	0.794	100	100	99.9	HY	HY	BC	BC
8	2007	Spur	W14	*female*	+	0.795	0.774	0.989	0.772	0.754	0.971	100	100	17.9	HY	HY	BC	BC
9	2007	Dark coat	W16	YH26	+	0.995	0.987	0.989	0.984	0.963	0.968	2.5	71.1	94.3	HY	W	W	BC
10	2006	wild-type	W14	YH17	+	0.972	0.946	0.979	0.919	0.886	0.920	99.9	99.9	99.8	HY	W	W	BC
11	2011	wild-type	W14	YH32	+	0.661	0.592	0.686	0.646	0.591	0.629	100	100	100	HY	HY	BC	BC
12	2012	wild-type	W14	*female*	+	0.882	0.971	0.993	0.871	0.930	0.982	100	97.0	23.1	HY	W	BC	BC
13	2012	unknown	W14	YH05	+	0.837	0.740	0.958	0.826	0.791	0.929	100	100	66.6	HY	HY	BC	BC
14	2009	unknown	W14	YH05	+	0.988	0.979	0.977	0.971	0.956	0.956	95.9	44.9	87.4	W	W	W	BC
15	2011	wild-type	W14	YH17	+	0.940	0.977	0.992	0.906	0.941	0.980	99.9	99.6	7.0	W	W	W-BC	BC
16	2006	unknown	W14	*female*	+	0.977	0.977	0.958	0.951	0.947	0.901	97.9	68.4	99.5	W	W	W	BC
17	2002	Black coat	W14	YH17	B	0.997	0.996	0.988	0.993	0.989	0.963	6.3	2.4	75.3	W	W	W	IG
18	2009	Black coat	W14	*female*	B	0.998	0.997	0.993	0.995	0.992	0.980	0.0	0.0	9.2	W	W	W	IG
19	2009	Dog-like; dark	W14	YH17	B	0.996	0.995	0.993	0.989	0.982	0.982	1.3	43.1	36.3	W	W	W	IG
20	2000	Black coat	W14	*female*	B	0.997	0.996	0.991	0.993	0.988	0.978	6.5	10.0	46.9	W	W	W	IG
21	2012	White nails	W14	YH17	+	0.998	0.997	0.994	0.994	0.991	0.985	0.0	8.2	4.3	W	W	W	IG
22	2011	unknown	W16	*female*	+	0.998	0.997	0.993	0.994	0.991	0.978	0.0	0.0	54.5	W	W	W	IG
23	2010	unknown	W16	YH26	+	0.995	0.993	0.986	0.979	0.964	0.935	98.6	99.0	99.0	W	W	W	IG
24	2006	unknown	W14	YH17	B	0.995	0.997	0.989	0.986	0.992	0.964	13.4	0.0	30.2	W	W	W	IG
25	2007	wild-type	W14	YH17	+	0.997	0.997	0.990	0.989	0.986	0.962	0.0	6.4	12.5	W	W	W	FP
26	2006	wild-type	W14	*female*	+	0.995	0.994	0.978	0.980	0.969	0.905	9.6	47.5	99.4	W	W	W	FP
27	2007	wild-type	W14	YH17	+	0.996	0.995	0.990	0.988	0.980	0.958	8.8	25.2	84.4	W	W	W	FP
28	1997	wild-type	W14	*female*	+	0.997	0.996	0.981	0.980	0.966	0.867	2.6	63.4	95.6	W	W	W	FP
29	2010	wild-type	W14	YH17	+	0.997	0.996	0.992	0.988	0.979	0.962	11.4	46.1	97.3	W	W	W	FP
30	2006	wild-type	W14	YH17	+	0.996	0.993	0.979	0.984	0.979	0.944	4.1	27.2	77.9	W	W	W	FP

[a]ID = sample identification number.
[b]Y = year of sampling.
[c]Oh = phenotype traits.
[d]CR = mtDNA control region haplotypes.
[e]Y-STR = Y-linked STR haplotypes detected in males.
[f]*K* = melanistic deletion at the *β-defensin CBD103* gene:+ = homozygote wild-type (no deletion), B = heterozygote for the K^B melanistic deletion.
[g]and [h] Structure = individual proportion of assignment in Structure admixture analyses to the Italian wolf cluster with 39, 24 and 12 microsatellites, with option usepopinfo not activated (*popinfo* = 0) or activated (*popinfo* = 1).
[i]GeneClass = probability of assignment to a distinct cluster (admixed genotypes) with 39, 24 and 12 microsatellites performed in GeneClass.
[j]Flock = assignment obtained through the non-Bayesian clustering procedure implemented in Flock to an Italian wolf (W) or hybrid (HY) cluster.
[k]Baps = assignment to an Italian wolf (W) or an admixed (HY) cluster with 39 microsatellites as inferred using Baps.
[l]NH = assignment to parental Italian wolf (W), F1, F2 or first generation backcross (BC) genotypic classes obtained with NewHybrids.
[m]ID = final identification of each sample as a likely F1, F2, backcross (BC), introgressed (IG) or false admixed (FP) genotype, based on a qualitative consensus of all th probabilistic admixture analyses.
doi:10.1371/journal.pone.0086409.t004

The probability of the data reached a plateau at ΔK= 4–6, with minimum LnP(D) values at K=6 in STRUCTURE analyses performed with 39 STRs and 8 groups (Figure Add1-2). At K=4 wolves and dogs were split into 4 clusters: dogs, WDCZ, WIT, WCZ plus WHR. At K=5 and K=6 the 3 wolf groups (WIT, WCZ and WHR) were assigned to 3 distinct clusters. WIT was not admixed while 14/30 (47%) of the putative hybrids showed signals of Italian wolf x dog admixture with qi values ranging from 0.509 to 0.953. The main contributions to admixture derived from WIT, DIT and DAP (Table S2). There was no apparent contribution from the 2 non-Italian wolf populations, with the exception of one sample that also showed the private Y-haplotype YH32. FLOCK

results with K=7 and 8 were concordant: the 3 wolf groups, DWCZ and HYIT were correctly assigned to different groups while the 3 dog groups were not separated.

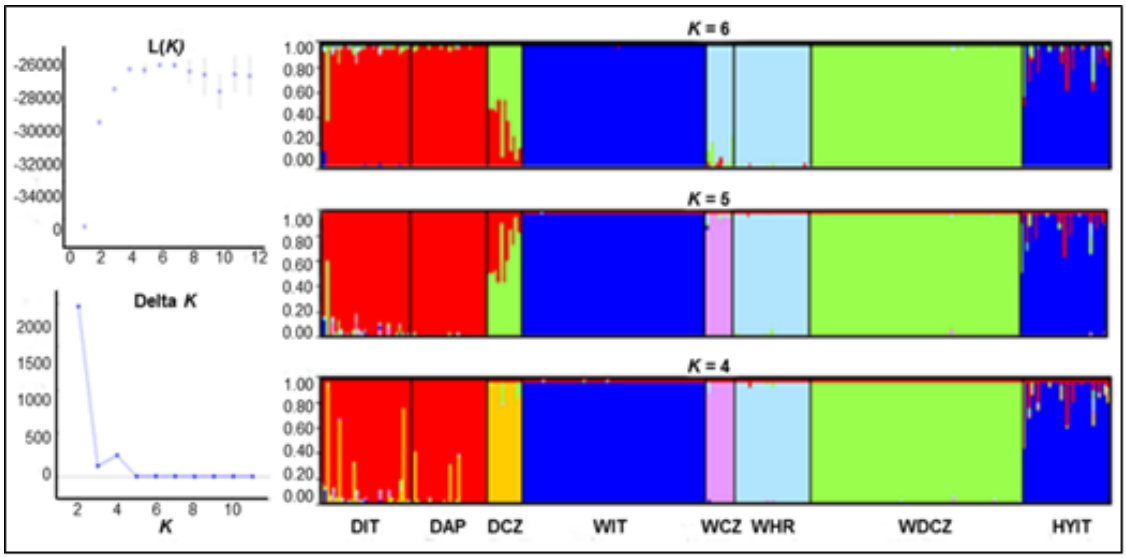


Figure Add 1-2: Structure analyses performed to infer the optimal partition of 8 sampled groups (A): DIT=village dogs in Italy; DAP=Apennine dogs; DCZ=German Shepherd; WIT=wolves in Italy; WCZ=wolves in Czech and Slovak republics; WHR=wolves in Croatia; WDCZ=Czechoslovakian Wolfdogs; HYIT=putative wolf *x* dog hybrids collected in Italy; (genotyped at 39 autosomal microsatellites).

The posterior probability Ln(*K*) of the data and the statistics Δ*K* were used to identify the optimal *K*=4 (averages of 2 independent runs). Plots of individual assignment probability to each inferred cluster are shown (B) for optimal *K*=4, 5 and 6. Structure was run assuming *K* from 1 to 12, with 400 000 MCMC and discarding the first 40 000 burn-ins, using the "*admixture*" and independent allele frequency "*I*" models, and no prior information (option "*usepopinfo*" not activated).

We compared the efficiency of the 39, 24 and 12 STRs to assign HYIT to their most likely parental groups (DIT, DAP, DCZ and WIT). ΔK stabilized at K=3–4 in STRUCTURE analyses (Fig. Add1-3). All WIT were assigned to their own cluster with qi 0.993, with the exception of one sample. In contrast, 13 (43%) to 15 (50%) of the 30 HYIT genotypes showed detectable signals of admixture with qi values ranging from 0.405 to 0.988. The other 15 samples did not show signals of admixture at their STR genotypes. The main contributions to admixture derived from the Italian wolves (cluster 4 in Table S3-add 1). STRUCTURE run with *popflag*=1 for wolves and dogs showed the same results (Table Add1-3), while FLOCK did not split WIT from HYIT and did not detect admixtures (not shown).

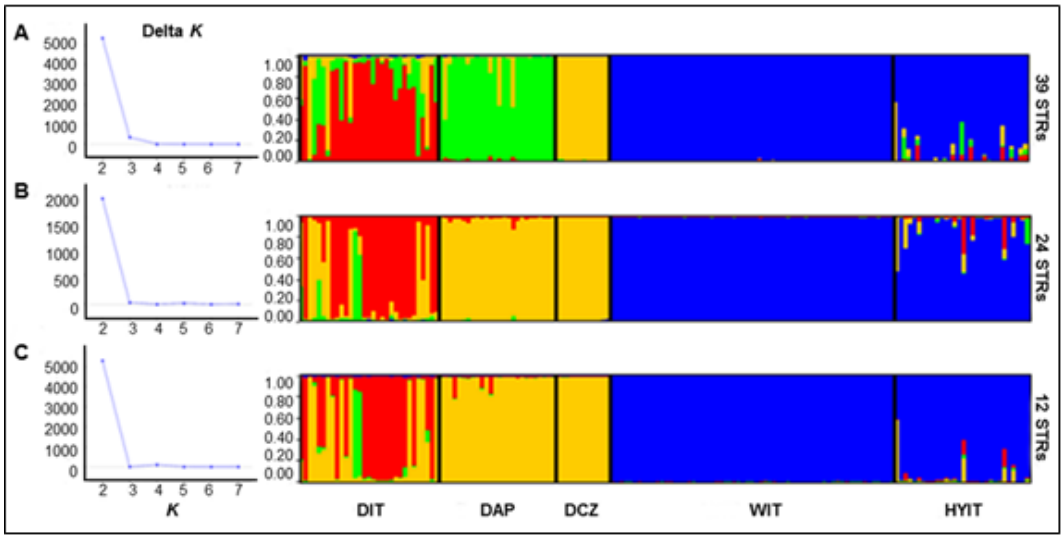


Figure Add 1-3: STRUCTURE analyses performed on the putative Italian wolf xdog hybrids (HYIT), assuming 4 reference groups (DIT, DAP, DCZ and WIT), at 39 (A), 24 (B) and 12 (C) microsatellites. STRUCTURE was run with K from 1 to 8 (left side: values of DK; (Evanno et al. 2005)), with 400 000 MCMC and 40 000 burn-ins, with option ‘‘usepopinfo’’ not activated.

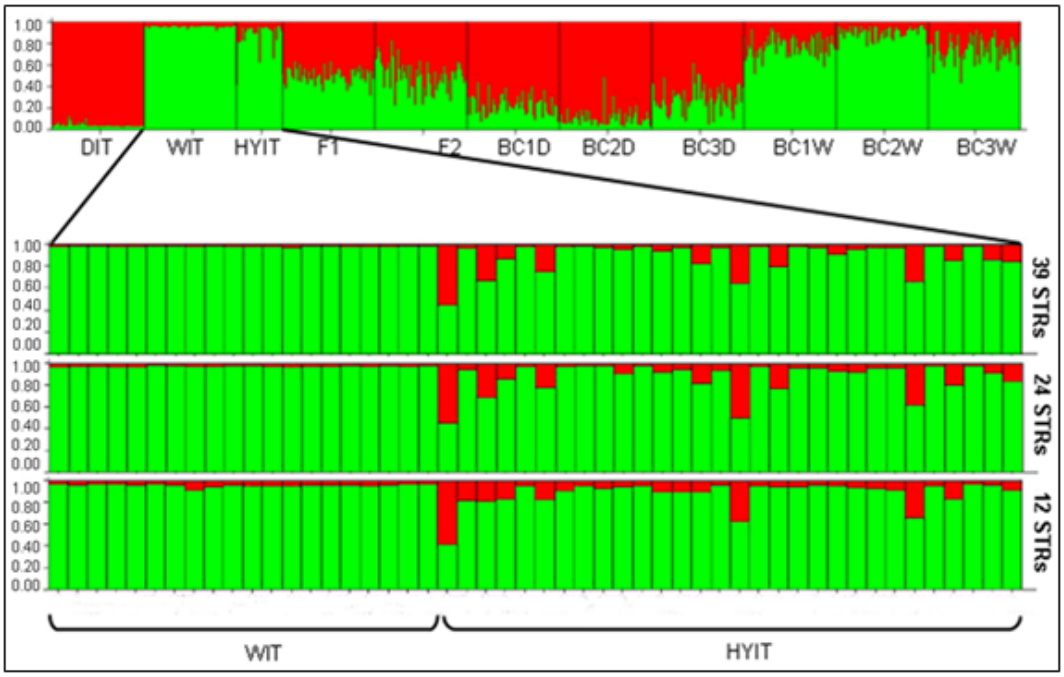


Figure Add 1-4: STRUCTURE analyses of empirical (DIT, WIT and HYIT) and HYBRIDLAB-simulated genotypes identified using 39 microsatellites. F1 and F2 between wolf and dogs; BC1= first, and BC2= second backcross with dogs (D) or wolves (W); BC3D and BC3W=F2 backcrossed with dogs or wolves, respectively. STRUCTURE was run with K=2; admixture and I models, popflag = 0. Details of the individual proportion of admixture in the Italian wolves (WIT) and putative hybrid (HYIT), genotyped with 39 (top), 24 (mid) or 12 (bottom) microsatellites are showed.

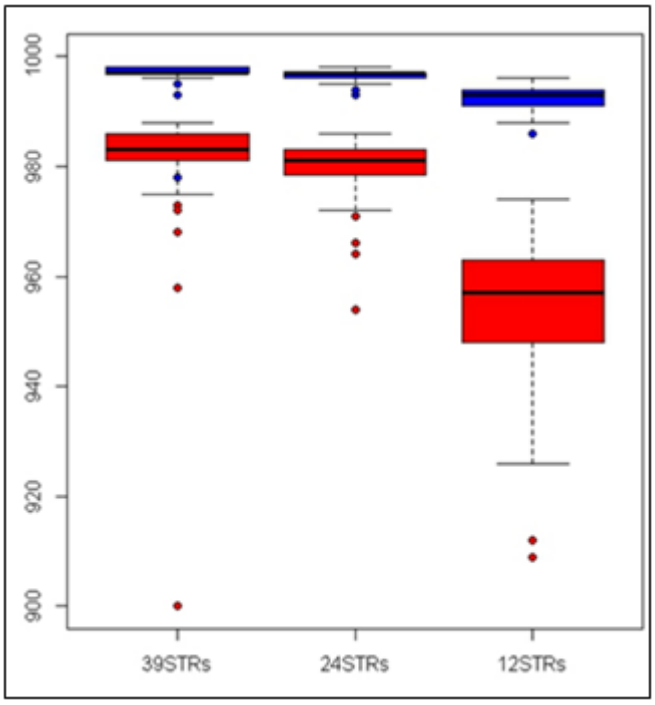


Figure Add 1-5: Average proportion of membership (q_i, upper boxplots) of wolves from Italy (WIT) to the wolf cluster and lower boundary of their 90% credibility intervals (CI; lower boxplots), computed on genotypes at 39, 24 or 12 microsatellites

The 5 genotypic classes simulated in HYBRIDLAB were correctly identified by STRUCTURE (K=2; Fig. Add1-4) with 39, 24 or 12 STRs. All simulated F1, F2 and BC1 were correctly assigned while c. 20% of the BC2 were confused with parental dogs or wolves. Decreasing the number of loci yielded decreasing values of the average proportion of membership in dogs (Qi = 0.973, 0.968 and 0.958) and wolves (Qi = 0.985, 0.980 and 0.960 with 39, 24 and 12 STRs respectively) due to increasing background noise in both wolves and dogs (Fig. Add1-4). Consequently, the 90% confidence interval (CI) values broadened, thus increasing the uncertainty of the assignments, particularly when STRUCTURE was run with 12 STRs (Fig. Add1-5). The risk of false positives (false admixed individuals) was inversely proportional to the number of STRs as indicated by BAPS results: admixture analyses based on 100 simulations for spurious admixture coefficients yielded 9 (30%), 8 (23%) and only 5 (17%) significantly admixed individuals (P= 0.05) with 39, 24 and 12 STRs, respectively.

STRUCTURE results with *popinfo* =0 and 39 STRs showed that, at the threshold q_i=0.985, all the Italian wolves were assigned to the same cluster with one exception. Sixteen (53%) of the 30 HYIT were identified as admixed (Table Add1-4). STRUCTURE results with 24 and 12 STRs (and *popinfo* =0) yielded 15 (50%) and 10 (33%) admixed HYIT at thresholds q_i = 0.980 and 0.960, respectively. At the same thresholds, STRUCTURE results obtained with *popinfo* =1 yielded 20 (67%), 23 (77%)

and 21 (70%) admixed genotypes with 39, 24 and 12 STRs respectively. STRUCTURE assignments were consistent for the 16 admixed genotypes showing the lowest qi values. These genotypes were also identified by GENECLASS, with the exception of sample n. 9. FLOCK identified as admixed 13 of these genotypes with 39 STRs, which were, however, assigned to the WIT cluster if analysed with 24 or 12 STRs. Only 9 of these genotypes were identified as admixed by BAPS (Table Add1-4). NEW HYBRIDS with 39 STRs showed that: 1) all dogs had posterior probability $P>0.999$ to be ''parentals'', except 4 samples; 2) all Italian wolves had posterior probability $P= 1.000$ to be ''parentals'', with one exception (see STRUCTURE results above); 3) among HYIT there was one F1 (n. 1), 2 F2 (n. 2 and n. 3), and 9 backcrosses (Table Add1-4). The other 18 samples (60%) were assigned to the parental wolf population.

Twelve of the 16 admixed genotypes also showed one or more anomalous phenotypic traits, the K^{β}deletion, variant mtDNA or Y-STR haplotypes (Table Add1-4). These 16 genotypes were finally identified as F1 (n. 1), F2 (n. 2 and 3) or backcrosses (BC; n. 4 to n. 16). Eight genotypes (n. 17 to 24) did not show any admixture signal at their STR genotypes but showed phenotypic anomalies, the K^{β} allele or the mtDNA haplotype W16 and were finally identified as introgressed (Table Add1-4). The remaining 6 samples (n. 25 to 30) were identified as presumptive hybrids only in STRUCTURE analyses with 12 STRs.

5.1.1 Ad focus of genetic composition of CSW study

The alignment of 79 mtDNA sequences from CSW individuals showed the occurrence of only two distinct haplotypes. Twenty-two dogs carried CSWA and 57 carried CSWB haplotypes, which differed by six mutations from each other [GenBank: KJ776748 and KJ776749]. Analysis in GenBank showed that both mtDNA haplotypes found in the CSW were shared with other domestic breeds but not with wolves.

Y-linked microsatellite variability analyses showed the presence of only two haplotypes in CSW, one shared with GS and one private (Table Add2-1).

Table Add 2-1: Distribution of the Y-linked microsatellite haplotypes as named by Randi et al. (2014). For all haplotypes, the alleles of each locus are listed.

Y-haplotypes	MSY34A	MSY34B	MSY41A	MSY41B	Population GS (11)	Population CSW (32)	Population CW (12)
YH01	168	177	113	118	0	21	0
YH05	170	175	113	126	**11**	**11**	**0**
YH11	172	175	113	126	0	0	5
YH16	174	173	113	122	0	0	1
YH31	174	173	113	126	0	0	5
YH49	172	175	113	124	0	0	1
Private haplotypes					0	1	4

GS—German Shepherds, CSW—Czechoslovakian Wolfdogs, CW—Carpathian wolves. In parentheses the number of individuals is reported for each group.

doi:10.1371/journal.pone.0143807.t001

All the samples, including the 25 non-invasive samples, provided distinct multilocus genotypes at autosomal microsatellite loci. The biparental microsatellite dataset did not show any significant presence of genotyping errors after Bonferroni corrections. All 39 autosomal microsatellites were polymorphic in CSW with a total of188 alleles. Mean N_A across all loci was 4.82 and ranged from 2 to 8 alleles per locus. The total number of N_P was 20, and none of the private alleles were shared with CW or GS. Average F_{IS} was 0.004, mean H_O= 0.5420, mean H_E = 0.5409 and meanA_R= 3.751 in CSW (Table Add2-2). While CW carry the highest number of N_P (N_P = 60) and their A_R is the highest among the studied groups (A_R = 4.626), they have the largest differences between H_E (H_E = 0.6404) and H_O (H_O = 0.691); thus, their F_{IS} (F_{IS} = 0.069) is slightly elevated compared to GS and CSW (Table Add2-2).

Table Add 2-2: Genetic variability in the three analysed groups at 39 autosomal microsatellite loci.

Group	N	N_A	N_E	P_N	A_R	H_O	H_E	F_{IS}
CSW	79	4.82	2.39	20	3.751	0.5420	0.5409	0.004
GS	20	3.90	2.26	11	3.709	0.5026	0.4921	0.005
CW	28	5.08	3.13	63	4.626	0.6091	0.6404	0.069

CSW—Czechoslovakian Wolfdogs, GS—German Shepherds, CW—Carpathian wolves.
Number of analyzed individuals (N), mean number of alleles across all studied loci (N_A), mean number of effective alleles per locus (N_E), total number of private alleles (P_N), mean allelic richness corrected by sample size (A_R), observed (H_O) and expected (H_E) heterozygosity, inbreeding coefficient (F_{IS}).

doi:10.1371/journal.pone.0143807.t002

Czechoslovakian Wolfdogs did not show significant deviations from HWE and LE. The best-supported number of clusters in STRUCTURE was K= 2, separating pure wolves from both dog breeds (mean estimated membership of population to the assigned cluster (Q_i) was 0.993). However, at K= 3, all individuals were correctly assigned to their own breed (CSW or GS) or wolf clusters with q_i values > 0.80 (mean Q_i was 0.981) and any internal substructure was detected among CSW (Figure Add2-1). Details of Bayesian analysis in STRUCTURE are shown in the appendix (Table Add2-S1). Factorial correspondence analysis clearly separated the three studied groups. The position of the CSW individuals in the plot is not intermediate between the parental groups of GS and CW, but it is closer to GS (Figure Add2-2).

The scenario obtained using BOTTLESIM showed a slight reduction of variability indexes through generations, as expected because of inbreeding and drift due to bottleneck demography of the breed. Specifically, we observed a reduction in the heterozygosity: H_O from 0.6944 to 0.5865 and H_E from 0.6095 to 0.5812. Such a reduction also appears when genetic variability is compared to the real CSW population analysed (see S2 Table) as CSW showed a higher N_A (N_A = 4.74 vs. 3.61) but a lower N_E (N_E = 2.38 vs. 2.69) and heterozygosity values (H_O = 0.54 vs. 0.59, H_E = 0.54 vs. 0.58) compared with simulated data (25th generation).

When we used the 339 simulated genotypes obtained after 25 generations of mating in NEESTIMATOR, the effective CSW population sizes we observed, considering P_{Crit} = 0 and P_{Crit} = 0.02, were *Ne* = 94.4 (95% CI: 87.8–101.6) and *Ne* = 92.8 (95% CI: 86.0–100.2), respectively, which are values, that are slightly higher than those obtained with real data. This difference could suggest a change in the maintenance of the genetic variation because of the non-random mating applied in the origin of the breed.

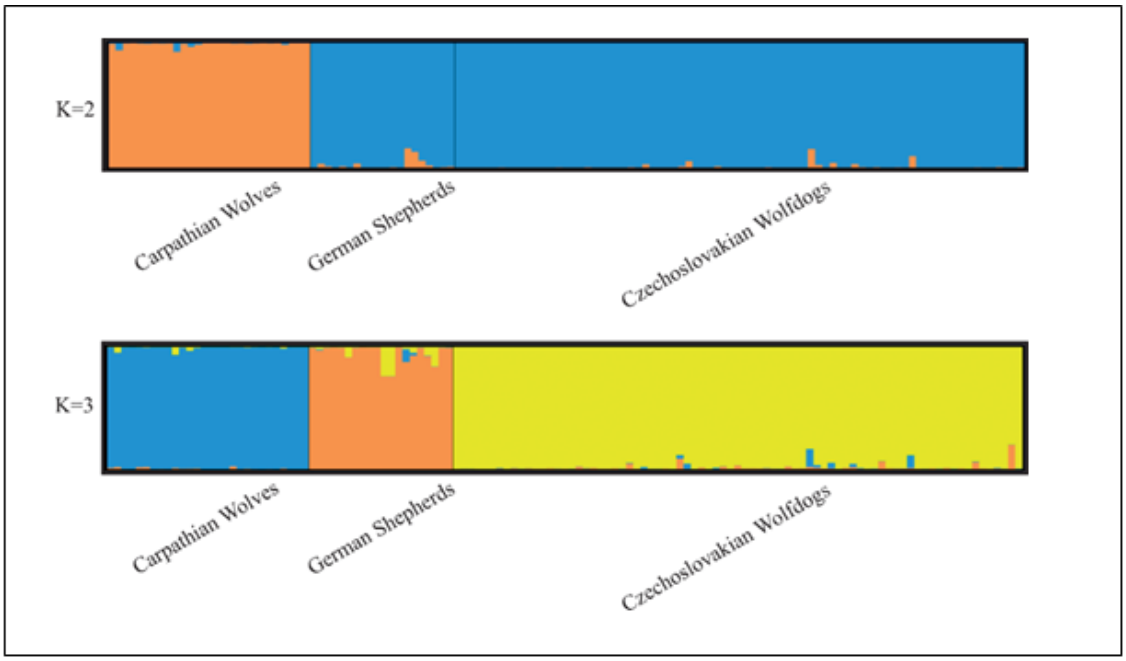


Figure Add 2-1: Bayesian clustering analysis of the three populations obtained by STRUCTURE.

Each individual is represented by one vertical bar that is divided into segments representing the proportion of memberships to the respective populations. The results are displayed for two (*K* = 2) and three (*K* = 3) suggested clusters.

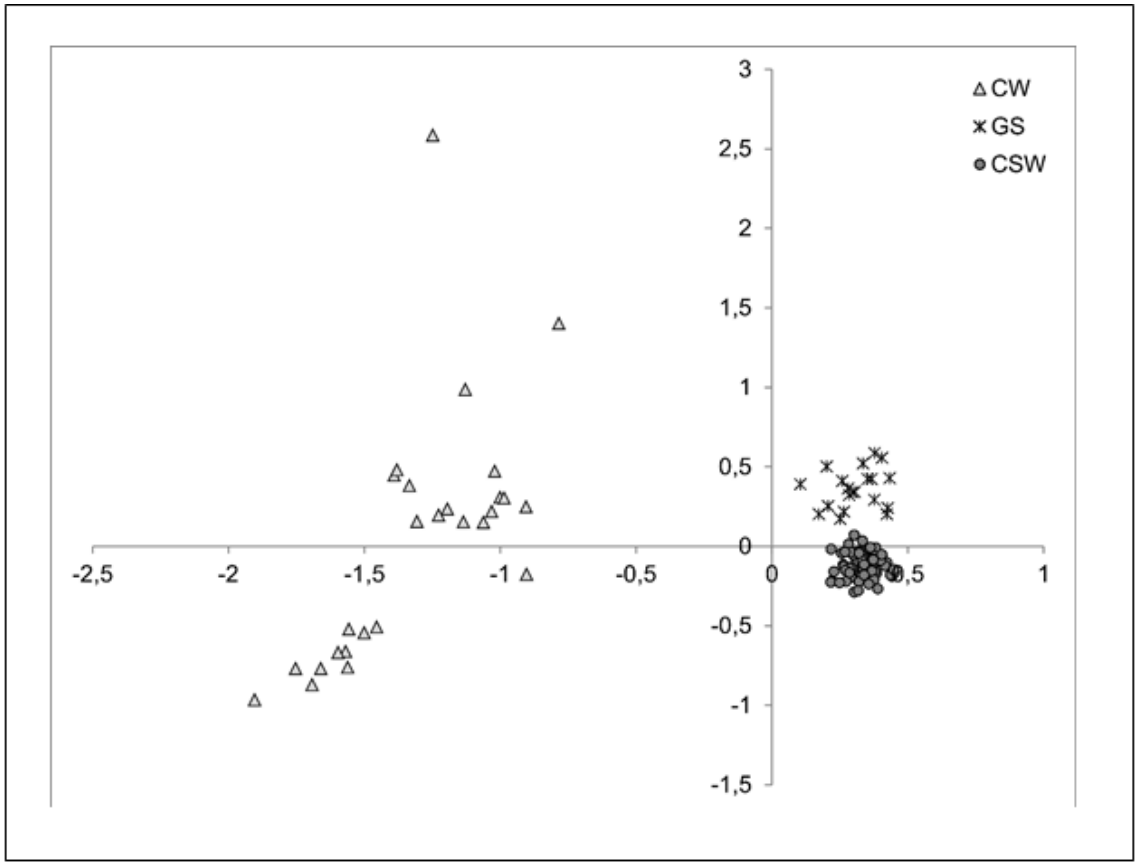


Figure Add 2-2: A two-dimensional plot of the factorial correspondence analysis performed in Genetix. CW = Carpathian wolves, GS = German Shepherds, CSW = Czechoslovakian Wolfdogs. NeEstimator showed a high concordance between the results obtained considering P_{Crit} = 0 and P_{Crit} = 0.02, from which the effective CSW population size was *Ne* = 76.5 (95% CI: 68.2–86.5) and *Ne* = 82.9 (95% CI: 72.3–96.4), respectively.

Table Add 2-3: S1 Table: Parameters of analysis in STRUCTURE for K1-K5.

	Mean LnP(D)	DK	Q
K1	-12903.500000		1.000000
K2	-11000.480000	2953.848686	0.993050
K3	-10441.380000	926.210010	0.981481
K4	-10847.940000	15.772914	0.971835
K5	-10052.540000		0.975886

5.1.2 Ad focus of a genome study of Czechoslovakian Wolfdogs

After removing loci mapping on chromosomes X and Y and following genotyping and quality cleaning steps per- formed in SVS, both per sample and per locus, we retained the 379 samples that were all successfully genotyped with call rate > 0.99 at 126,848 autosomal SNPs (73%, hereafter referred to as the 126k dataset). These samples included the 12 CWDs and the 12 Carpathian wolves, plus the 12 German Shepherds and the additional 343 dog genotypes from 30 breeds obtained from the LUPA project dataset. A subset of 57,020 SNPs (33%) was retained after LD pruning at threshold r^2 = 0.2 (the 57k dataset). Finally, a smaller set of 9,063 SNPs (5.2%) was obtained after discarding all sites with any missing data (the 9k dataset).

In a pairwise F_{ST} matrix of the genetic distances among groups (Supplementary files: Figure S1) computed from the 126k dataset, CWDs were relatively divergent from Carpathian wolves (F_{ST} = 0.33) but, as expected, the breed least differentiated from German Shepherds (F_{ST} =0.19). We found considerable genome-wide variability within groups (Supplementary files: Fig. S2a). Overall, heterozygosity was generally higher in dogs (H_o = 0.265 ± 0.032) than in wolves (H_o = 0.231 ± 0.025). However, a direct comparison between wolves and dogs should be treated with caution due to the possible ascertainment bias from the SNP array, mostly designed on dogs, although it is expected to be minimal when considering closely related taxa (VonHoldt et al. 2011). CWDs showed heterozygosity levels (H_o =0.249) lower than most breeds but, as expected, slightly higher than in German Shepherds (H_o = 0.234, p-values < 0.05 ; t-test) and also than Carpathian wolves (Ho = 0.231, *p*-values < 0.05 ; t-test), which showed values coincident with those described in other wolf studies based on SNP chip genotyping (H_o =0.210-0.240; (Vaysse et al. 2011; VonHoldt et al. 2011; Pilot et al. 2018)).

In an exploratory PCA performed considering CWDs and their parental populations (Fig. 1), the first two axes of the PCA clearly discriminated the three groups, explaining more than the 90% of the whole genetic variability, with Czechoslovakian Wolfdogs plotted along the first axis (which explains 68% of variability) between wolves and dogs, though closer to the latter in accordance to the history of the breed. When we considered the whole 126k dataset (Supplementary files: Figure S3), Czechoslovakian Wolfdogs were located intermediate between German Shepherds and Carpathian wolves along the PC1 axis, which explained more than 30% of the entire genetic variability, and well separated from the other dog breeds overall. Along axis 2, CWDs and German Shepherds clustered close to one another, likely for the higher number of individuals sharing common genetic components compared to those belonging to other breeds, as it occurs when regrouping these same taxa in a neighbour-joining tree (Vaysse et al. 2011).

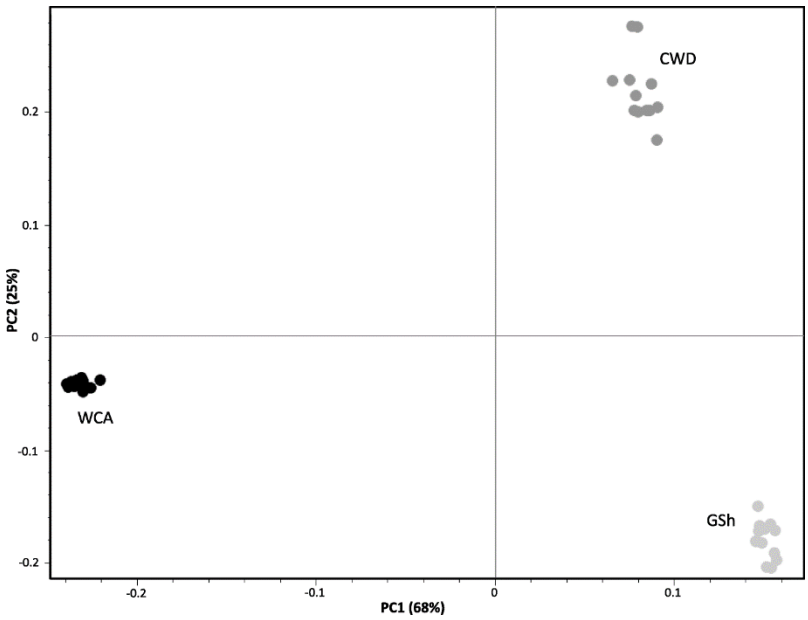


Figure Add 3-1: PC1 vs. PC2 results from an exploratory principal component analysis (PCA) computed in SVS on the 126k SNP dataset and including Carpathian wolves (WCA; black dots), German Shepherds (GSh; light grey dots), and Czechoslovakian Wolfdogs (CWD; dark gray dots). The two axes are not to scale, in order to better distinguish individuals along PC2.

Results from ADMIXTURE, run with the 57k dataset and including only CWD, Carpathian wolf and German Shepherd genotypes, showed that the first main decrease in CV error was observed at K=2(Fig. 2a), when Carpathian wolves (mean estimated membership of population to the assigned cluster Q_1 = 1.00) were clearly separated from the two dog breeds (Fig. 2b), which clustered together (mean Q_2 = 0.987), although several CWDs (Q_2 = 0.975) presented limited but clear traces of wolf components (individual qi ranging from 0.940 to 1.00). However, the optimal number of genetic clusters corresponded to K =3(Fig. 2c), when CWDs (Q_3 = 0.994) were clearly separated from both Carpathian wolves (Q_1 = 1.00) and German Shepherds (Q_2 =0.995). In CWDs, the average genome-wide proportion of blocks assigned by PCA_{DMIX} to the reference wolf population was 0.30±0.03, with individual assignment values ranging from 0.27 to 0.34, significantly higher (p-values = 1.75×10^{-10}; t-test) than the mean proportion of membership to the wolf cluster (q_w) estimated from ADMIXTURE at K =2. Conversely, PCA_{DMIX} assignment values were not significantly different (p-values = 0.09, t-test) from the percentage of wolf ancestry estimated from the pedigrees, whose mean proportion was 0.28±0.01, with individual scores ranging from 0.27 to 0.30 (Fig. 3a).

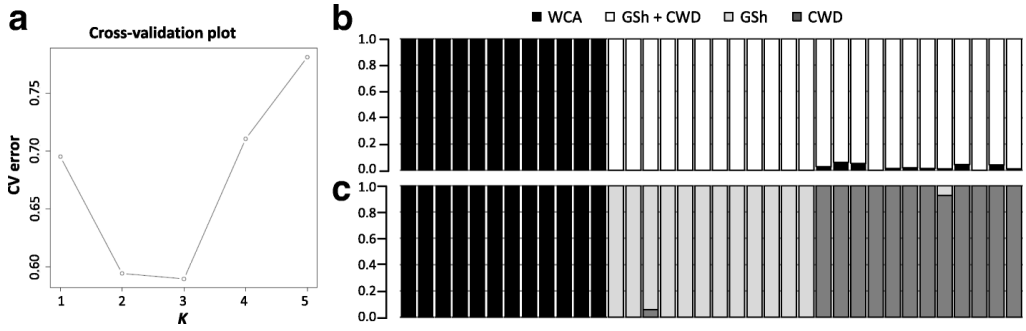


Figure Add 3-2: Admixture results obtained running the 57k SNP dataset with *K* from 1 to 5 and including genotypes from Carpathian wolves (WCA), German Shepherds (GSh) and Czechoslovakian Wolfdogs (CWD). a Cross validation plot showing the most likely number of genomic clusters. b Admixture results at *K* = 2 show how Carpathian wolves are clearly separated from the two dog groups that cluster together. c Admixture results at *K* = 3 show that the three groups are well differentiated from one another.

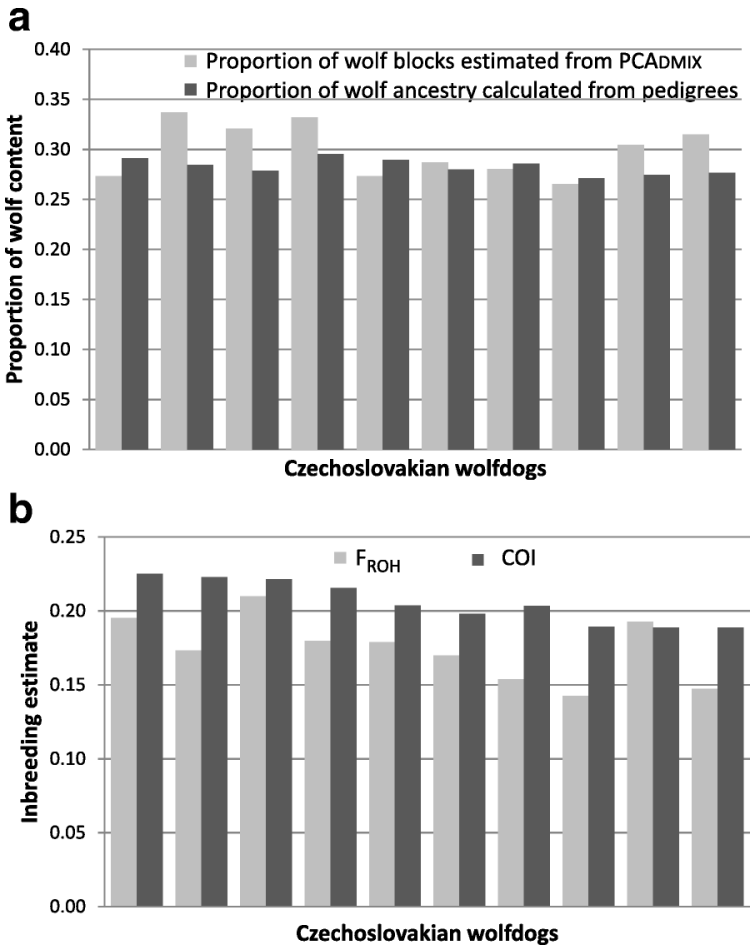


Figure Add 3-3: Wolf ancestry proportions and inbreeding rates.

a Comparison between individual wolf proportions estimated from the analysis of blocks of 10 consecutive, non-overlapping SNPs performed in PCAdmix (in light grey)

and individual wolf ancestry rates obtained from pedigrees using BreedMate Pedigree Explorer (in dark grey).

b Comparison between the individual frequency of ROHs (F_{ROH}), calculated in SVS as the proportion of ROHs on the genome length spanned by the analysed SNPs (in light grey), and the individual Wright's inbreeding coefficient (COI) estimated from the pedigrees with the software U-WGI (in dark grey).

Analysing the whole 126k dataset, CWDs showed a mean number of ROHs (117 ± 33), intermediate between that of German Shepherds (124 ± 16) and that of Carpathian wolves (71 ± 31) (Fig. 4a). As expected according to recent the history of the breed, which allowed a very short time for recombination to break up segments that were identical-by-descent, CWDs showed a mean ROH length (3.234 ± 400 kb) longer than both German Shepherds (2.971 ± 501 kb) and Carpathian wolves (2.699 ± 1.398 kb) (Fig. 4b). This was due to the fact that, although the mode of the ROH length in CWDs and German Shepherds was similar (with most of their ROHs around 2000 kb-long), and much longer than in Carpathian wolves (about 1000 kb), CWDs also showed a second peak of ROHs of 7000 kb length, suggesting that inbreeding events also occurred in the few generations after the breed creation (Fig. 4c).

CWDs showed a mean value of the inbreeding coefficient F_{ROH} (0.17 ± 0.02) similar to German Shepherds (0.16 ± 0.02; *p*-value = 0.10; *t*-test) but significantly higher than Carpathian wolves (0.08 ± 0.03; *p*-value < 0.05; *t*-test) with individual, F_{ROH} values ranging from 0.14 to 0.21 (Fig. 3b). F_{ROH} was significantly correlated with the inbreeding coefficient estimated from the genotype information *F* (R^2> 0.395; *p*< 0.01; Additional file 2:Figure S2b,c)and also with the pairwise coefficient of inbreeding calculated on the basis of pedigree data (COI), that ranged from 0.19 to 0.23 (R^2> 0.369; *p*< 0.01; Additional file 4:FigureS4).

Looking at identity-by-descent (IBD) between individuals, the highest mean values of pairwise IBD scores (*p*-values < 0.05; *t*-test), as expected according to the low number of founders used in the first steps of the breed creation, were observed in CWDs (0.477 ± 0.049, ranging from 0.426 to 0.738), followed by German Shepherds (0.362 ± 0.054, ranging from 0.000 to 0.451) and then by Carpathian wolves (0.112 ± 0.034, ranging from 0.000 to 0.403). The IBD values found in CWDs were highly concordant (R^2 = 0.584; *p*<0.01) with the coefficients of relatedness (COR) estimated from the pedigrees (mean 0.431 ± 0.040, ranging from 0.380 to 0.607), though the pairwise scores between individuals detected from the two approaches in some cases showed marked differences (Fig. 5).

The mean LD in CWDs was intermediate (r^2 = 0.26) between German Shepherds (r^2 = 0.30) and Carpathian wolves (r^2 = 0.13). Similarly, the LD decreased to values of

r^2< 0.10 at a smaller distance in Carpathian wolves (18 kb) than in CWDs (76 kb) and German Shepherds (110 kb; Additional file 5: Figure S5).

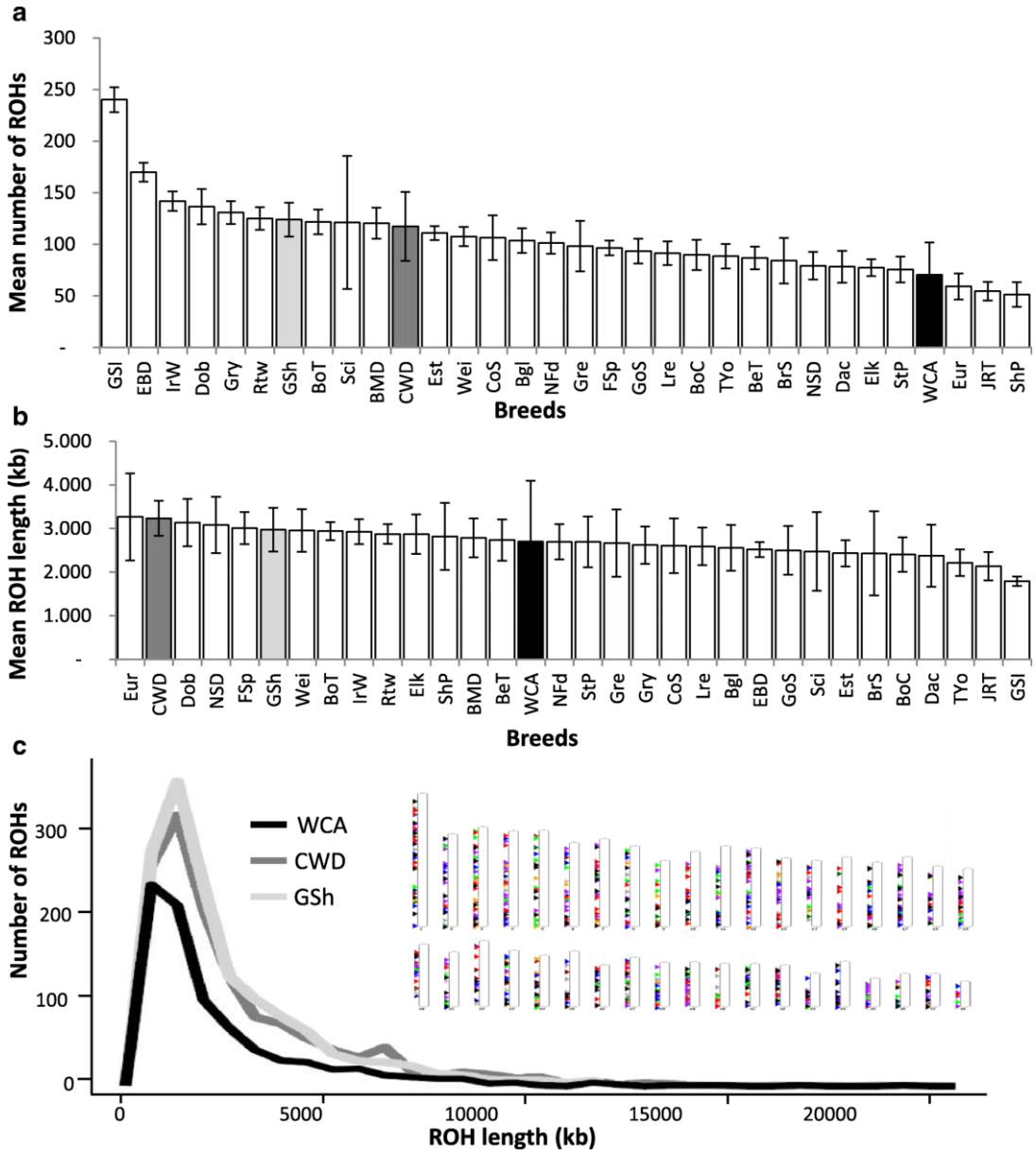


Figure Add 3-4: Runs of homozygosity (ROH) analysis. a Mean number of ROHs per breed. Czechoslovakian Wolfdogs (CWD) show a mean number of ROHs intermediate between values from parental populations. German Shepherds (GSh) are closer to the breeds with the highest values whereas Carpathian wolves (WCA) to breeds with the lowest values. Bars indicate standard deviations. b Mean ROH length (kb) per breed. The mean length of ROHs in Czechoslovakian Wolfdogs (CWD) is wider than parental populations suggesting a high recent inbreeding rate. Bars indicate standard deviations. c Distribution of ROH lengths in the three groups. Carpathian wolves (WCA; black line) show most of ROHs of 1000 kb length whereas German Shepherds (GSh; light grey line) and Czechoslovakian Wolfdogs (CWD; dark grey line) exhibit similar

patterns, both with most of ROHs around 2000 kb length. However, Czechoslovakian Wolfdogs also show a second peak of ROHs of about 7000 kb length suggesting a stronger inbreeding in more recent generations. Bar plots indicate the 38 Czechoslovakian Wolfdog autosomal chromosomes which show a quite uniformly distributed number of ROHs.

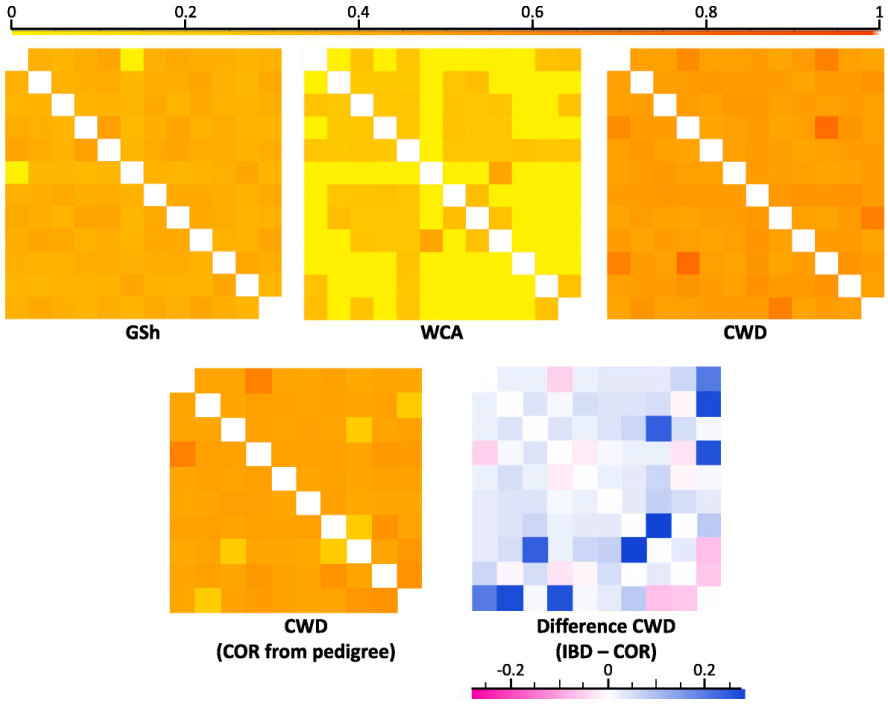


Figure Add 3-5: Relatedness analyses. Chromatograms represent pairwise Isolation-by-distance (IBD) scores between Czechoslovakian Wolfdog (CWD), Carpathian wolf (WCA) and German Shepherd (GSh) individuals computed using SVS and CWD coefficient of relatedness (COR) estimated from their pedigrees using the software BreedMate Pedigree Explore. Interestingly, a comparison between the two approaches shows marked differences in some Czechoslovakian Wolfdogs.

The demographic trajectory estimated from LD well- reflected the history of the breed, which experienced a continuous population decline begun 20 generations ago, thus in the late 1950s', ranging from a maximum of 418 individuals in 1959 to a minimum of 21 individuals in 2010 (Fig. 6). The only four growth peaks in N_E were observed in periods corresponding to the deliberate crossings with wolves performed for the creation of the breed, plus another moderate one in more recent times not matching any registered crossing.

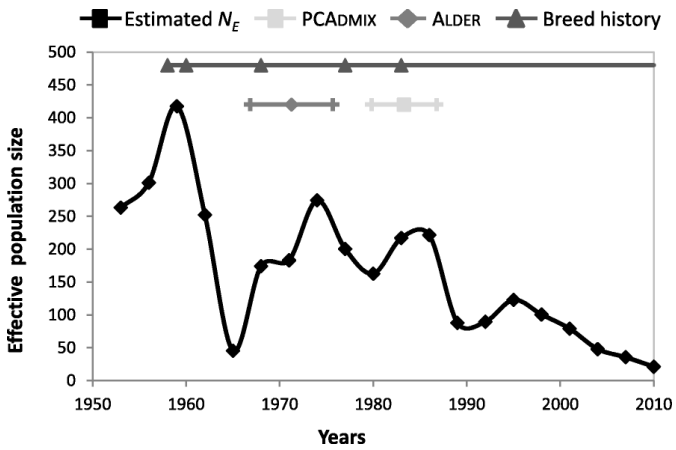


Figure Add 3-6: Estimates of demographic trends. The effective population size N_E estimated from LD (squares on black line) shows a decreasing trend over time, though it shows four growth peaks that are concordant with the deliberate crossings with wolves that occurred in the history of the breed (triangles on the dark grey line). The temporal distribution of the admixture events deduced from PCAdmix (squares on light grey horizontal bars) and the time intervals reconstructed by Alder (diamonds on grey horizontal bars) are also described. Square, triangle and diamond symbols represent mean values whereas vertical sticks represent confidence intervals.

The software A_{LDER}(Loh et al. 2013) identified significant admixture between the parental populations (p-values = 1.0×10^{-17}) in our CWDs, with successful decay rates (meaning that both the parentals could have been fully sampled; (Loh et al. 2013)). Hybridization was estimated to have occurred about 12.91 ± 1.47 generations before sampling, which, assuming a wolf generation time of 3 years (Skoglund et al. 2011), corresponded to a period ranging from 1967 to 1976, centred around 1971 (Fig. 6).

Results from PCA_{DMIX}, used to estimate individual admixture times, showed that the individual number of switches from German Shepherd to Carpathian wolf ancestry blocks ranged from 165 to 367 (mean value 196 ± 55), indicating that the admixture likely occurred from 7.8 to 10.1 generations before individual sampling. Considering the same value of 3 years per generation (Skoglund et al. 2011), when converted into years these values indicated that the oldest individual hybridization event likely traced back to 1975, whereas the most recent one traced to 1990, highlighting slightly more recent times than those provided by the software A_{LDER}.

The analysed CWDs revealed a complex genomic mosaic of wolf and dog ancestry, as reconstructed by PCA_{DMIX} (Additional file 6: Figure S6).

From the 10-SNP blocks found to be fixed for wolf or dog haplotypes in all CWDs by PCA_{DMIX}, we identified 14 “wolf-like” blocks, including 31 protein-coding genes significantly enriched for metabolic and enzymatic processes and for HP categories related to aortic and renal disorders, and 1784 “dog-like” blocks, including 2238

annotated protein-coding genes, significantly enriched for GO categories mainly related to brain and heart development (Table 1 and Additional file 7: Tables S1a-S1d).

Table Add 3-1: Subset of wolf-like (a) and dog-like (b) outlier genes detected in Czechoslovakian Wolfdogs analysed in this study which have been previously described in the canid literature.

Gene name	Methods	Chr	Start (bp)	End (bp)	Reference	Association
a						
CRHBP	BayeScan	3	29,726,312	29,738,001	[60]	Social behavior and maternal aggression
NPHP4	PCAdmix, GO	5	59,805,955	59,936,808	[64]	Bone and retinal disorder
ENO1	PCAdmix	5	62,301,164	62,312,161	[57]	Related to mRNA transcript variants, genes responsible for bone and cartilage tissues
ASTN2	BGC	11	70,248,612	70,977,896	[57]	Related to mRNA transcript variants, genes responsible for bone and cartilage tissues
PCDH15	BGC, GO	26	33,962,360	34,571,935	[63]	Vision regulation and hearing abilities
BMP3	BGC	32	5,207,833	5,231,966	[58]	Morphological features: paws and bones
b						
ARID1B	PCAdmix	1	46,370,636	46,799,104	[68]	Cellular responses, DNA repair
URI1	BGC, GO	1	121,528,137	121,612,185	[57]	DNA-binding
RPE65	F_{ST} SNP	6	76,887,399	76,911,146	[64, 67]	Dog diseases (Leber congenital amaurosis)
EPAS1	PCAdmix, GO	10	48,551,410	48,634,643	[77]	Environmental adaptation
ASCC3	PCAdmix, GO	12	58,592,025	58,932,720	[68]	Cellular responses, DNA repair
GRIK2	PCAdmix, GO	12	59,590,231	59,992,091	[68]	Lipid metabolism
SMARCD3	PCAdmix, GO	16	15,279,418	15,289,275	[77]	Muscle cell differentiation, heart morphogenesis
ZMAT4	F_{ST} SNP, PCAdmix	16	24,561,867	24,889,045	[57]	DNA-binding
ADAM9	PCAdmix , GO	16	26,410,907	26,551,122	[64]	Dog diseases (cone-rod dystrophy)
STRN	F_{ST} SNP	17	29,273,978	29,365,239	[78]	Dog diseases (arrhythmogenic right ventricular cardiomyopathy)
MGST2	PCAdmix	19	3,067,163	3,070,563	[68]	Cellular responses, DNA repair
NOCT	PCAdmix, GO	19	3,589,720	3,607,191	[65]	Circadian rhythms, body weight and digestion
SLC7A11	PCAdmix	19	4,289,915	4,371,635	[68]	Lipid metabolism
CNTN5	PCAdmix	21	1,128,048	1,614,989	[20]	Nervous system differentation
OXT	PCAdmix	24	18,193,429	18,194,002	[69]	Learning and memory processes
CBDs	PCAdmix	24	20,614,030	20,971,219	[68]	Immune system
DEFB119	PCAdmix	24	20,905,210	20,918,355	[68]	Immune system
HM13	PCAdmix, GO	24	21,026,827	21,067,920	[68]	Cellular responses, DNA repair
RALY	ROH	24	23,211,141	23,262,511	[46]	Coat color
ASIP	ROH	24	23,354,642	23,393,918	[20, 46]	Coat color, social behavior
NCOA6	ROH	24	23,802,887	23,866,792	[68]	Co-activation of several hormone-dependent receptors
ACSS2	ROH	24	23,928,670	23,972,633	[68]	Lipid metabolism
TMEM132D	F_{ST} SNP, PCAdmix	26	2,074,728	2,662,470	[68]	Oligodendrocyte differentiation, metabolism
CUX2	PCAdmix, GO	26	8,730,082	8,996,271	[68]	DNA-binding
SEZ6L	PCAdmix, GO	26	19,889,395	20,079,319	[70]	Social behavior
ARVCF	BGC, PCAdmix, GO	26	29,314,144	29,534,294	[70]	Polydactyly and morphological features
COMT	PCAdmix, GO	26	29,360,372	29,366,006	[70]	Social behavior (aggression and attention regulation)
PCDH15	BayeScan, PCAdmix, F_{ST} SNP, GO	26	33,962,360	34,571,935	[63]	Polydactyly and morphological features, vision and hearing abilities, communication and behavior
BMPR1B	PCAdmix, GO	32	17,819,265	17,978,113	[66]	Polydactyly and morphological features
UNC5C	F_{ST} SNP, PCAdmix, GO	32	17,987,785	18,332,959	[66]	Tumor suppression
BANK1	PCAdmix, GO	32	23,281,315	23,603,279	[67]	Regulation processes of calcium ions
TGIF1	PCAdmix	32	32,950,116	32,950,934	[66]	Nervous system differentation
IGF2BP2	PCAdmix	34	18,368,131	18,522,156	[75, 76]	Lipid metabolism
MARCH7	BGC	36	5,499,129	5,531,823	[43]	Cellular responses, DNA repair
NHEJ1	PCAdmix, GO	37	25,633,562	25,719,307	[80]	Dog diseases (Collie eye anomaly)
SLC4A3	PCAdmix	37	26,136,624	26,149,312	[64, 79]	Dog diseases (progressive retinal atrophy)

When we considered ROHs that were shared by all Czechoslovakian Wolfdogs, we identified a genomic region of about 15 Mb on Chr24 that was always assigned as dog-derived by PCA_{DMIX}. This region hosted 29 annotated protein-coding genes, including the coat colour regulating genes *ASIP* and *RALY* (Vaysse et al. 2011; Dreger et al. 2013), and genes significantly enriched for a high number of HP categories linked to amino acid metabolism (Table 1 and Additional file 7:TablesS2a-S1b).

Based on the lowest F_{ST} between Czechoslovakian Wolfdogs and Carpathian wolves, we identified 15 wolf-like SNPs and one 10-SNP block on chr24 that hosted 1 gene included in significantly enriched GO and HP categories principally related to regulation of catabolic processes, response to external stimulus, locomotory and

learning disability (Table 1 and Additional file 7: Tables S3a-S3b; S4a-S4b). When we considered the lowest F_{ST} between Czechoslovakian Wolfdogs and German Shepherds, we identified 241 dog-like SNPs and 9 dog-like blocks of 10 consecutive SNPs that included 25 annotated protein-coding genes, significantly enriched for BP category mainly related to palate development and GO categories principally related to regulation of ion transmembrane transport (Table 1 and Additional file 7: Tables S3c-S3d; S4c-S4d).

BGC results detected 78 SNPs with an excess of wolf ancestry (significantly negative values of α) and 62 SNPs with an excess of dog ancestry (significantly positive values of α), with overall higher absolute values in the latter (Additional file 8: Figure S7a). The 50-kb regions surrounding the SNPs with excess of wolf ancestry contained 109 coding genes enriched for HP categories mainly related to cerebral atrophy (Table 1 and Additional file 7:Tables S5a-S5b). Conversely, regions surrounding the SNPs with excess of dog ancestry contained 79 protein-coding genes that were mostly enriched for a GO biological process related to granulocyte regulation, and HP categories linked to earlobe morphology and skeletal, aortic or parathyroid disorders (Table 1 and Additional file 7: Tables S5c-S1d).

Finally, comparing CWDs with German Shepherds, $B_{AYE}S_{CAN}$ identified 29 outlier SNPs with positive α values (suggestive of diversifying selection) hosted in regions including 29 protein-coding genes, significantly enriched for GO categories mainly linked to biological processes such as maternal aggressive behaviour and corticotropin secretion, and HP categories principally related to abnormal proportions of face and hands (Table 1 and Additional file 7: Tables S6a-S6b). When we compared CWDs to Carpathian wolves, BAYESCAN identified 7 outlier SNPs with positive α values that were hosted in regions including 7 annotated protein-coding genes, significantly enriched for GO categories mostly linked to tRNA regulation. (Table 1 and Additional file 7: Tables S6c-S6d).

6 Discussion

In a study comparing different panels of genetic markers recognizing pure wolves, dogs or hybrid individuals in Italy were all genotyped wolf, dog and hybrid samples variable at autosomal and uniparental markers. Italian wolves showed the lowest genetic diversity, probably as a consequence of long-lasting genetic isolation south of the Alps and a recent bottleneck (Lucchini et al. 2004), clustering separately from the other studied populations. Wolf populations in eastern Europe, though experienced less dramatic bottlenecks or had persistent gene flow with neighbouring populations (VonHoldt et al. 2011), exhibited partial signals of isolation in Mediterranean, Balkan and perhaps Carpathian glacial refuges during Pleistocene (Pilot et al. 2010). Post-glacial recolonization determined a complex population mosaic, which has been further shaped by restrictions to gene flow due to local prey specialization, or by random drift due to recent anthropogenic fragmentation (Pilot et al. 2006). Consequently, wolf genetic diversity in Europe is geographically partitioned and populations are genetically identifiable (VonHoldt et al. 2011). Genetic divergence among wolf populations and between wolves and dogs provides the basis for a wealth of molecular markers that can be used in assignment and admixture analyses (Vonholdt et al. 2010).

Admixed wolf genotypes may originate in consequence of intentional or accidental escapes of non-indigenous wolves from captivity, or by crossbreeding with dogs. In this study we did not detect consistent signals of admixture between Italian and other wolves. All the presumed hybrids clustered very close to - or partially overlapped with - the Italian wolves and showed no obvious connection to any other group. An exception was the strongly admixed sample n. 11, which was partially assigned to the WIT, WHR and WDCZ clusters. This sample was collected in the northern Apennines during 2011, showed the Italian wolf W14 mtDNA haplotype and a private YH32 Y-haplotype that was not found in any other wolf or dog group analysed in this study(Table Add1-4). Ancestry of this sample with dogs, non-Italian wolves or WDCZ cannot be excluded, although the haplotype YH32 was not found in the WDCZ samples. Most of the HYIT showed the Italian wolf haplotypes YH17 and YH26, but 2 individuals had the haplotype YH05, shared with DIT and DAP and that was also detected in WDCZ. The position of WDCZ in the DAPC plots indicates a higher proportion of dog genome. This is in agreement with the origin of the breed that was established in the 50ies by crossing 4 Carpathian wolf founders with German Shepherd dogs. Backcrossing with German Shepherds continued together with artificial selection. The standard of the current breed was approved by the Federation Cynologique Internationale in 1989.The genotypes of WDCZ in the DAPC were not intermediate relative to their parental populations (DCZ and WCZ), probably in consequence of strong founder effect, persistent low effective population size and genetic drift.

Moreover, captive wolfdogs experienced artificial selection designed to keep dog behaviour while preserving wolf-like phenotypic traits including coat colour, sensory abilities and endurance, with possible hitchhiking effects on linked neutral loci. Unofficial recurrent and more recent crossbreeding of wolf-like dogs with wolves by individual breeders may continue to generate hybrids with variable phenotypes and behaviours (Hope 1994). Hybrids may be aggressive, difficult to control and have chances to survive in nature crossbreeding with wolves. Genetic analyses of larger sample sizes are needed to identify local hybridization events, but, with the exception of the uncertain origin of sample n. 11 and haplotype YH05, the available evidence led to exclude that the WDCZ are a widespread source of hybridization with wolves in Italy, pointing out a main contribution of village and other dogs.

The K^B melanistic deletion was detected only in DIT, DAP and HYIT. A different melanistic mutation at the *MC1R* gene is known to determine black coats specifically in German Shepherd dogs (Kerns et al. 2004; Candille et al. 2007) and explains also why WDCZs do not have the K^B deletion. Wolf samples collected in Croatia and in the Carpathians were all wild-type grey and the K^B deletion was not expected. We do not know the origin of the K^B haplotype in the Italian wolves, if via hybridization with black dogs or by a spontaneous mutation at the *β-defensin CBD103* gene. The K^B deletion was already present in ancient canids in Europe over 10 000 years ago, probably entered in North American wolf populations through ancient hybridization with dogs, and was also found in a melanistic pack of hybrid origin in Italy (Anderson et al. 2009; Caniglia et al. 2013; Ollivier et al. 2013). However, these findings are still controversial and cannot be generalized. Results of admixture analyses reported in our study are not unequivocal: 2 black-coated individuals showed strong signals of admixture at their multilocus STR genotypes and the concomitant presence of the melanistic K^B deletion, while other 5 black-coated individuals did not show any signal of admixture. These animals, which could originate from past hybridization no longer detectable at their 39 STR genotypes, deserve additional investigations. The origin of the K^B deletion in dog or wolf ancestors could be ascertained by sequencing the flanking haplotypes in Italian wolves and village dogs. Signals of past hybridizations may become detectable using genomic data and haplotype block reconstructions (VonHoldt et al. 2011; Miller et al. 2012; Feulner et al. 2013). Melanistic phenotypes in wolves and dogs can be determined also by epistatic interactions among other and still undescribed mutations (Kaelin & Barsh 2013). It is noteworthy that one Italian dark-coated backcross did not show the K^B deletion (sample n. 9 in Table Add1-4), suggesting that mutations at other structural or regulatory genes may add complexities to the expression of melanistic coat colour variations in wolves. A lack of samples or the absence of hybridization may explain why melanistic wolves were not found in other European countries (but see Godinho et al. 2011). We found

that 50% of our putative hybrids showed unequivocal signals of Italian wolf x dog admixture at STRs. Seven of them also showed morphological anomalies, and 4 had the K^B deletion or mtDNA CR1 and Y-STR haplotypes not found in the Italian wolves. Nine genotypes yielded weaker STR admixture signals, but showed dark coats, white nails, the K^B deletion or variant mtDNA CR1 and Y-STR haplotypes. Hence, 24 (80%) of the putative hybrids showed combinations of variant phenotypic and genetic traits suggesting admixed origins. The remaining 6 samples were identified as presumptive hybrids only by STRUCTURE analyses with 12 STRs, which however might produce false positives. The putative hybrids were not randomly collected and the admixed individuals are not representative of the frequency of hybridization in the Italian wolf population. The frequency of hybridization should be estimated by extensive sampling through the entire wolf distribution range. A well-planned country-wide program of wounded or found-dead wolf sampling would provide additional, but likely biased information, because the probability to encounter dead or wounded wolves is expected to vary in the heterogeneous landscapes used by wolves (Ciucci et al. 2007). Moreover, carcasses of introgressed individuals can be confused with dogs and not collected (Godinho et al. 2011). Instead, exhaustive sampling collection throughout the wolf distribution range can be obtained by long-term non-invasive genetic monitoring programs (Caniglia et al. 2013). Hybridization in wolves seems to be prevalently asymmetric, originating by female wolves mating with male dogs. The vast majority of admixed wolf genotypes described so far showed wolf mtDNA haplotypes(Randi 2008, 2011; Godinho et al. 2011), with a few exceptions (Muñoz-Fuentes et al. 2010; Hindrikson et al. 2012). Mating of male wolves x female dogs, however, could occur because young males disperse frequently and are expected to explore and colonize new areas more rapidly than females (Fabbri et al. 2007; Ciucci et al. 2009). In this study, sample n. 1, an F1 hybrid identified by all the admixture analyses and confirmed by the allelic composition of all its STR loci, showed dog-like body shape and the mtDNA haplotype D15 that is shared only with Italian village and Apennine dogs, indicating a female dog parental. Moreover, we identified 3 backcrosses that shared the same mtDNA haplotype W16 so far detected only in Bulgarian wolves (Randi et al. 2000). These samples were collected from carcasses found in 3 geographically distant areas of the northern Apennines during different years (2007, 2010 and 2011). Theoretically, they might originate from the same or a few related packs. However, their average Queller and Goodnight'srelatedness r= 0.076 was significantly lower (P<0.001, t-test) than average r estimated with the same panel of 12 STRs in 26 Italian wolf packs with known pedigrees (r= 0.3906±0.106 (Caniglia et al. 2013)), suggesting independent crossbreeding events. Thus, hybridization of wolves in Italy was not strictly patrilineal. Two other backcrosses had the Y-haplotype YH05, which was found in dogs sampled in Italy and in WDCZ. These samples were collected in 2 distant areas (central and

southern Apennines) in 2009 and 2012. Probably they were not closely related (average *r*= - 0.128) and have originated from 2 independent hybridization events.

Simulations showed that *c.* 48 STRs with F_{ST}>0.10–0.15 are needed to significantly improve the reliability of backcross identifications (Vähä & Primmer 2006). VonHoldt et al. (2013) demonstrated that even 100 highly diagnostic SNPs cannot efficiently discriminate second generation wolf *x* dog backcrosses. Thus, estimating the minimum number of markers to identify backcrosses is still an open issue. The outcomes of our admixture analyses computed using 39, 24 and 12 STRs were not as straightforward as expected. The estimate of admixed individuals did not increase using more loci, and a naive assumption that larger panels of markers should lead to identify more admixed individuals was not fulfilled. The 24 most discriminating STRs were equally or more efficient than the full set of 39 STRs. Individual assignments were consistent for the 16 genotypes with the lowest q_i values in the Italian wolf cluster, which were also identified by GENECLASS, independently on the assumptions embedded in the algorithms implemented in the different software. The assignments of the other genotypes were less consistent, and variable outcomes were obtained using 12 STRs. Some genotypes had disproportionally high q_i values (particularly running STRUCTURE with *popflag*=0; e.g., n. 5 and 12) and could represent false negatives. Other samples showed disproportionally low q_i values (e.g., n. 25, 26, 27, 28, 29 and 30) and could be false positives. These results highlight 2 related issues that were often neglected in other studies:

1) HYBRIDLAB simulations showed how the power to correctly identify known (simulated) hybrids and backcrosses changes with the number of markers: the larger is the number of STRs, the higher is the threshold. Decreasing the number of markers decreases the average proportion of membership in reference clusters due to increasing background noise. Consequently, the width of CI values and the individual assignment uncertainty will increase. The risk of false admixed individuals is inversely proportional to the number of STRs, as indicated by BAPS simulations. The use of more markers allows to apply higher qi thresholds, reducing uncertainty and the risk of false positives. Therefore, each study should plan adequate power analyses to identify the appropriate thresh- olds, whereas the adoption of threshold used in other studies might not guarantee optimal assignments.

2) The number of markers used in admixture analyses is not important per se, but the discriminating power of markers deeply affects the results of the assignments. The power of markers can be approximated by single-locus FST values between reference sample groups or assessed by other computational approaches (Banks et al. 2003; Rosenberg et al. 2003). Selecting the most discriminating STRs will reduce the costs (both manpower and chemicals) of genetic assays, the rate of genotyping errors intrinsically associated to each additional marker, and the risk of false positives. The 24

most discriminating STRs selected in our study only partially overlapped the most discriminating STRs identified by Godinho et al. (2011), indicating that the selection of loci should be performed based on local populations' data sets. A similar approach was suggested by Hindrikson et al. (2012), for mtDNA and Y haplotype identification.

Although genomic platforms promise extensive screening of thousands of SNPs, practical and financial constraints still limit their applications in conservation genetics. Genetic monitoring of carnivores is still based on the genotyping of limited numbers of STRs, often in DNA molecules extracted from non-invasively collected samples. In other cases, tissue samples are collected from found dead animals, which often produce degraded DNA. Genotyping large numbers of STRs will probably continue to be problematic in practical conservation genetics, due to risks of false alleles and allelic dropout in molecular identifications of low-content DNA. For the same reasons, genotyping large numbers of SNPs is still unpractical in non-invasive genetics. Selecting the minimum number of informative autosomal STRs, plus informative mtDNA and Y-linked markers will remain the most viable strategy in the near future.

In study analysing the genetic composition of Czechoslovakian Wolfdogs, we obtained estimates of genetic diversity parameters, including mean number of alleles per locus (N_A), mean number of effective alleles per locus (N_E), number of private alleles (N_P), allelic richness (A_R), expected (H_E) and observed (H_O) heterozygosity, inbreeding coefficient (F_{IS}), and Hardy-Weinberg equilibrium (HWE) deviations, in the recent Czechoslovakian Wolfdog breed that were compared to those obtained from the breed's ancestral founder populations, the wild Carpathian wolves and the domestic German Shepherd dogs. Our results show a high proportion of dog genome in CSW, which is in agreement with the origin of the breed and data from previous studies (Leroy et al. 2009). For example, the position of CSW in the FCA is closer to GS, as expected as a consequence of the 25 years of backcrossing. The lack of wolf mtDNA haplotypes indicates the loss of these variants during the lineage sorting acting in the wolfdog pedigree. The same evidence also seems to be associated with the Y- chromosome, as we found that none of the two Y-haplotypes detected in the 32 CSW males we analysed for this study were shared with CW. Using the same panel of four Y-linked microsatellites, Čílová et al. (2011) found that some of the CSW males still carry one Y-haplotype of wolf origin, although they did not report the frequency of this haplotype in the CSW population. The different CSW and reference CW samples used in the analyses could be the reason why we did not detect Y-haplotypes derived from wolves. Moreover, the moderate number of microsatellite markers used in our analyses may not reflect the whole genome variability of the breed.

Analyses in BOTTLESIM indicated that neutral processes acting in small populations, such as the founder effect and genetic drift, have changed the genetic composition of CSW. This is in agreement with the relatively low number of founding

individuals of this breed. The loss of genetic variation may cause the overrepresentation of some deleterious alleles in many dog breeds including CSW and GS (Wilbe et al. 2010), causing genetic diseases such as dwarfism (Voorbij et al. 2014) and degenerative myelopathy (Zeng et al. 2014). On the other hand, populations ofmoderate size may still possess some degree of evolutionary potential (Hayes et al. 2003). Among proximate mechanisms responsible for the positive effects of bottleneck, purging of deleterious alleles was also described (Glémin 2003). However, the FIS value was relatively low in CSW compared with other purebred dog breeds (Huson et al. 2010). Recent simulation and empirical studies showed that diverse life history traits, including mating patterns and overlap of generations, may influence the effect of bottleneck on diversity patterns (Hoban 2014). The loss of genetic variation may also be caused by rapid post-bottleneck recovery (Hundertmark & van Daele 2010). However, as the life history traits in CSW are similar to other breeds and the population size remained relatively low during the breed's history, the compensation for the expected founder effect and genetic drift could be ascribed to the effects of outbreeding that are related to the introgression of wolf alleles. The high variability expected in hybrids, deduced from assumptions of Mendelian inheritance, could cause some hybrids to be far from parental phenotypic optima and start evolutionary trajectories that are divergent from the parental forms (Rieseberg et al. 1999). This may also explain the position of CSW in FCA, which is not at the centre between both parental populations.

The position of the breed in cluster analysis may also be influenced by other factors. After obtaining admixed genotypes between shepherd dogs and their wild ancestors, the population of captive Wolfdogs experienced artificial selection that aimed to keep wolf-like phenotypes while preserving dog behaviour and the inherent preference of other characteristics connected with domestication. These traits are associated with many genomic regions, related to, for example, morphology, physiology and behaviour (Axelsson et al. 2013). If some of the examined neutral loci were linked to these genes under selection, their allele frequency could be changed by selective sweep or background selection (Barton 2000). For example, coat colour, which is in strong artificial selection in dogs including CSW, is controlled by more than 300 known genetic loci and 150 known genes in mammals (Cieslak et al. 2011). Considering physiological traits, comparison of wolfand dog genomes provided evidence that dogs, for example, adapted to a starch-rich diet by a higher production of amylase and maltase-glucoamylase (Axelsson et al. 2013). Last but not least, evolutionary novelties accompanying domestication also include complex changes in social behaviour (Belyaev 1979), which in dogs enabled social interactions with humans including cooperation, social learning and communication (Miklósi et al. 2003; Axelsson et al. 2013; Nagasawa et al. 2015). Controllability was one of the key factors targeted during the breeding process of CSW.

The fast-growing number of registered Czechoslovakian Wolfdogs worldwide demonstrates the elevated economical value of this breed and the need of a deeper comprehension of the genetic bases of its morphological and behavioural traits, as well as of the causative mutations of some common diseases. In this study we provide the most complete genomic description of the breed to date by genotyping 12 individuals at 170k SNPs and comparing their genome-wide diversity to samples as representative as possible of their parental populations (Carpathian wolves and German Shepherds) and to genomic profiles from 30 other common breeds publicly available from the LUPA project (Lequarré et al. 2011; Vaysse et al. 2011).

From a preliminary genomic screening, based on pairwise F_{ST} values, multivariate and assignment procedures, CWDs appeared highly differentiated from all the other analysed breeds and were also well-distinguished from both parental populations. In particular, despite our limited sampling, the Bayesian clustering analysis performed in ADMIXTURE revealed the presence of three optimal clusters clearly separating CWDs from both parental populations, consistent with previous findings based on a few autosomal microsatellites (Leroy et al. 2009; Randi et al. 2014; Bigi et al. 2015; Smetanová et al. 2015). Compared with the LD-based approach of ADMIXTURE (K =2), the PCA-based admixture deconvolution approach implemented in PCADMIX (Brisbin et al. 2012), which reflects the ancestry proportions of an individual better than ADMIXTURE(Lawson et al. 2018), identified larger wolf components (> 25%) in the genome of the analysed CWDs. These proportions compared well with the pedigree-based estimates, confirming that such a haplotype block-based approach is an appropriate and reliable tool to assess real admixture proportions from genomic data (Lawson et al. 2018).

Our results on the observed genome–wide heterozygosity levels in CWDs were consistent with other studies, based on different types and number of markers (Leroy et al. 2009; Vaysse et al. 2011; Bigi et al. 2015). In particular, values of autosomal heterozygosity in our small sample of CWDs were slightly higher than those observed in the parental populations, consistent with the recent admixture occurred in the creation of the breed (Leroy et al. 2009; Randi et al. 2014; Smetanová et al. 2015) that is still visible in the large genomic regions hosting both dog and wolf haplotype blocks, thus representing islands of high heterozygosity, even after c. 30 generations since the breed foundation and c. 11 generations since the last official outcrossing, contrasting the expected decay in heterozygosity due to inbreeding.

On the contrary, the lower heterozygosity observed in Carpathian wolves, which was expected to be higher than in dogs for genomic sequences (Wang et al. 2016), should be treated with caution, since it could be partially attributable to a possible ascertainment bias linked to the original SNP chip design, mostly based on dog variation (VonHoldt et al. 2011; Gopalakrishnan et al. 2017), although such event is

unlikely for closely related taxa diverging less than one million years (VonHoldt et al. 2011). However, our estimates of observed heterozygosity in Carpathian wolves well compare with those from other Central-Eastern European wolf populations reported in previous studies using the same SNP chip approaches (VonHoldt et al. 2011; Stronen et al. 2013; Pilot et al. 2018) and certainly did not affect the ability of our methods to discriminate between wolf-like and dog-like haplotype blocks in CWDs. The analysis of ROHs allowed us to better reconstruct the breed history and clarify its dynamics. Czechoslovakian Wolfdogs showed a higher number of long ROHs (>5 Mb) than the progenitors, reflecting the recent inbreeding events (Ferenčaković et al. 2013; Kim et al. 2015; Iacolina et al. 2016) that occurred during and after the origin of the breed. Moreover, coherently with the low number of founders utilized in the breed creation, CWDs showed inbreeding coefficient values (FROH)higher than both parental populations (Smetanová et al. 2015), and also higher values of relatedness between individuals, on average. Though a direct comparison between genomic dataand pedigree information should be treated with caution given the different methodologies these two types of computations rely on (Kardos et al. 2015), estimates of inbreeding levels calculated from the frequency of homozygosity regions (FROH) were comparable with those calculated from the coefficient of inbreeding (COI) derived from the available pedigree data. Such a concordance confirms the reliability of several proxies in identifying inbreeding, which is crucial for breeders since matings among closely related individuals can affect their offspring fitness due to the increased probability of deleterious alleles being expressed in their phenotypes. Conversely, in several cases the coefficient of relatedness (COR) between individuals estimated from the pedigrees underestimated the IBD (identity-by-descent) scores determined from genetic profiles. Such discrepancies could be due to the higher ability of genome-wide methods to identify random segregation effects compared to pedigree-based methods (Kardos et al. 2015), or to the uncertainties of pedigree records, in which breeders might deliberately not report some crossings between related individuals, since the possible negative effects on health could reduce the marketability of dogs (Iacolina et al. 2016), even if this latter possibility appears very unlikely given the strict breeding control operated by the military during the early years of breed establishment.

Therefore, genomic reconstructions represent a useful tool to implement carefully planned mating strategies among breeders in order to predict and contrast possible deleterious effects such as lethal genetic disorders, reduction of fertility, and lower adaptive potential (Bjelland et al. 2013; Stronen et al. 2017). For these reasons, genomic pairwise IBD values and ROH-based metrics could provide breeders with additional information that could be evaluated for the selection of lineages to reduce the levels of inbreeding per generation, taking into account not only the blood lines but also the stochastic effects of recombination (McQuillan et al. 2008; Kardos et al. 2015).

Our genome-wide characterization allowed us to verify the timing of the admixture in the cohort of the analysed CWDs, which compared well with the key steps of the breed selection, namely the repeated insertion of wolf alleles that officially continued until 1983. When applying A_{LDER}, hybridization was estimated to have occurred from 1967 to 1976, roughly corresponding to the mid- point of the known crossing events, whereas PCA_{DMIX} better identified the most recent ones. These findings show that genomic-based dating methods can be effective and complementary in tracing recent hybridization events both in hybrid breeds such as CWD and in wild-living populations (Galaverni et al. 2017).

The N_E trends estimated from the LD patterns showed that, despite the growing number of registered individuals, N_E overall declined from the breed origin to the present. This decreasing trend is likely due to the progressive artificial selection and to the so-called "popular sire effect", namely the overrepresentation of the genetic contribution of popular dogs (e.g. small number of winner individuals at dog shows) in subsequent generations of the breed (Dreger et al. 2016). Conversely, N_E fluctuations, with four main peaks around years 1959, 1968, 1974 and 1986, are consistent with the official wolf x dog registered crossings (1960, 1968, 1974 and 1983). However, we unexpectedly detected an additional slight increase in N_E around 1995, which could be due to the genetic contribution from a distinct lineage (Dreger et al. 2016) of CWDs (e.g. from the Slovakian to the Czech lineage), or might be the signal of an undeclared wolf contribution that occurred after the official breed recognition. Should this second hypothesis be confirmed, it would value genomic investigations also as a tool to identify illegal crossings of wild species protected under the CITES Convention with commercialised domestic breeds (Ouborg et al. 2010; Boscari et al. 2014). Nonetheless, this overall, fast decline in N_E did not erode all the additional variation provided by the wolf founders, since the heterozygosity levels appear to be still currently slightly higher in the analysed CWDs than in German Shepherds.

Looking at the genomic landscape of Czechoslovakian Wolfdogs, PCA_{DMIX} results showed a variegated chromosomal ancestry mosaic, ranging from fully dog-derived to mostly wolf-like regions. A gene search based on ancestry-outlier regions obtained from multiple methods, which was possible thanks to the availability of the well-annotated dog reference genome, allowed us to identify more than 300 genes with an excess of wolf ancestry and more than 2000 genes with an excess of dog ancestry in Czechoslovakian Wolfdogs compared to random expectations.

The key wolf-like genes we identified were mainly related to body size and shape traits, which could explain the overall morphological similarity of CWDs with wolves. In particular, we detected two wolf-excess genes, *ASTN2* and *ENO1*, which were described in the human genome to be adjacent to loci putatively responsible for bone and cartilage tissue production and that were earlier found to be under selection in

European wolves (Pilot et al. 2014). Another 9 wolf-like genes were related to key morphological features, such as prominent occiput (*ITCH*) and prominent nasal bridge (*CLIP1, WDPCP*), narrow face (*AP4M1, CLIP1*), short ears (*CAMTA1*), narrow and small mouth (*KCNAB2, CAMTA, AP4M1, CLIP1*), pointed chin(*CLIP1, AP4M1*), strong facial musculature (*CLIP1, AP4M1, HNRNPA2B1*), robust paws and bones (*AGGF1, BMP3*; (Schoenebeck et al. 2012)), all typical of the breed. However, other wolf-excess genes were described to be associated with communication and behaviour. In particular, *CRHBP*, coding for Corticotropin Releasing Hormone Binding Protein, is a gene expressed during pregnancy (Mastorakos & Ilias 2003), involved in the anomalous maternal aggressive behaviour against puppies observed both in mice and in Australian Working Kelpie female dogs (Arnott et al. 2015). Such peculiar behaviour is well known also in Czechoslovakian Wolfdogs, where mothers killing their offspring shortly after parturition have often been observed (A. Camatta, personal communication). *PCDH15* has been identified as a candidate gene related to echolocation in mammals(Le Guédard et al. 2007; Parker et al. 2013)and has been described to be under selection in different ecological contexts in wolves (Schweizer et al. 2016). Similarly, other wolf-excess genes were related to cardiac (*KCNAB2, WDPCP*), pancreatic (*PLCG2*), bone and retinal (*NPHP4*) disorders that have been widely described in a number of dog breeds (Miyadera et al. 2012), but not yet in wolves, and that could provide a higher resistance of CWDs to such disorders compared to German Shepherds.

Conversely, a number of behavioural traits desired by the breeders could be hosted in a large set of dog-like genes, often involved in brain development, which has been demonstrated to be a pivotal target of domestication (Li et al. 2013). In particular, two genes were related to neural differentiation and formation of the nervous system (*TGIF1* and *CNTN5*; (Vaysse et al. 2011; Pfahler & Distl 2015)) and the gene *TMEM132D* was involved in oligodendrocyte differentiation that was previously identified in dog and wolf selection scans (Ostrander 2012; Pilot et al. 2014; Freedman et al. 2016). Similarly, we identified a number of dog-like genes playing important roles in learning and memory processes, such as *OXT*, which can affect canine cognition, tolerance, adaptation and maternal behaviour (Seifi Moroudi et al. 2014), in vision and hearing abilities, such as *PCDH15* (Schweizer et al. 2016), and in the regulation of circadian rhythms, body weight and digestion, such as NOCT (Li et al. 2013; Freedman et al. 2016)which could be crucial in adapting the physiological activity of CWDs to that of their human owners.

Interestingly, we also detected genes described to be correlated with sociability: *COMT*, a gene involved in dopamine catalysis and in regulating aggressive behaviours and attention in many breeds, and *SEZ6L*, both mapping on chr26 and described as significantly associated with the time dogs spend in close proximity of humans (Persson

et al. 2016), reinforcing the hypothesis that the transformation of negative defensive reactions toward humans into positive responses could have been a primary step in early dog domestication (Persson et al. 2016) and that deliberate artificial selection on tameness may been have further reinforced (Li et al. 2013). Direct or indirect artificial selection for tameness or sociability played a key role on the evolution of a number of other domesticated and wild taxa: the possibility of a strong artificial selection on tameness was demonstrated also in the rat (*Rattus norvegicus*; (Albert et al. 2009)) and in the red jungle fowl (*Gallus gallus*; (Bélteky et al. 2017)), showing that a number of other traits were influenced by their sole selection on tameness, as already revealed in the keystone study by Balyaev and colleagues on silver foxes (Trut et al. 2009), leading to the concept of General Domestication Syndrome to indicate a set of phenotypic traits common to a number of domesticated species (Wilkins et al. 2014). However, the long-lasting presence of human-dominated landscapes can indirectly affect the genetic bases of tameness also in wild-living populations, such as the Apennine brown bear (*Ursus arctos marsicanus*), which shows reduced aggressive behaviours compared to other populations reflected in a unique genomic signature (Benazzo et al. 2017).

Another set of dog-excess genes were involved in the regulation of calcium ions (*BANK1*), in the co-activation of several hormone-dependent receptors (*NCOA6*) and in DNA-binding (*CUX2, URI1, ZMAT4*), which were also identified to be under selection in previous canid studies (Ostrander 2012; Pilot et al. 2014; Freedman et al. 2016).

Additionally, we identified other dog-like genes known to be involved in immune functions, such as those coding for the immunity-related beta-defensins (*CBDs* and *DEFB119*) and those responsible for cellular responses and DNA repair (*ARID1B,ASCC3, HM13, MGST2, MARCH7*), and tumour suppression (*UNC5C*), that were identified to be hosted in key-differentiating regions for dog domestication (Freedman et al. 2016). We detected another four dog-excess genes, *IGF2BP2*(Chase et al. 2009; Boyko 2011)and *SLC7A11, ACSS2, GRIK2*(Freedman et al. 2016), which were related to lipid metabolism and to the synthesis of energy that could indicate the importance of dietary modifications during the domestication process, especially during the phase of breed creation (Freedman et al. 2016).We also found two genes (*ASIP* and *RALY*) involved in regulating coat coloration by the synthesis of the yellow pigment known as pheomelanin, that could confer the typical colour to the breed (Vaysse et al. 2011; Dreger et al. 2013). Interestingly, recent evidences demonstrated that variations in *ASIP*, found to be under selection also in ancient Asian dog breeds (Yang et al. 2017) ,can influence social behaviour too, most likely through its antagonistic effects on melanocortin receptors or α-melanocortin stimulating hormone (Freedman et al. 2016; Yang et al. 2017), confirming that morphological and behavioural characteristics in canids can be strongly linked (Trut et al. 2009; Stone et al. 2016). However, we identified also a series of dog-excess genes previously described in the literature to be

linked to a number of common dog disorders such as arrhythmogenic right ventricular cardiomyopathy (*STRN;*(Cattanach et al. 2015)), progressive retinal atrophy (*SLC4A3*;(Downs et al. 2011; Miyadera et al. 2012)), Collie eye anomaly (*NHEJ1*;(Parker et al. 2007)), cone-rod dystrophy (*ADAM9*; (Miyadera et al. 2012)), and canine Leber congenital amaurosis, previously known as congenital stationary night blindness (*RPE65*;(Miyadera et al. 2012)), that are typical of the parental population German Shepherd and that could be retained during the strong artificial selection that occurred during the CWD creation.

6.1 Synthesis

Nowadays, the role of dogs has transformed dramatically, various roles as guardians, herders, hunters are not prioritized any longer and companion role of dogs is more important for humans. This modification corresponds with the environmental transformation in which humans and dogs co-inhabit(Sykes et al. 2020). Approximately four hundred dog breeds exist in present, all variable in their use, body size and phenotype. All breeds originated during the long lasting pressures exposed to human-related selection (Morrel 1997; Pedersen et al. 2013; Parker et al. 2017).

Dog genome has been completely sequenced in 2005 (Ostrander & Wayne 2005). Since that time many studies have been done and a large scale of connectedness has been investigated. The aim of this study was to assess a genetic diversity of a young low numerous breed under conditions of human managed breeding programmes.

Czechoslovakian Wolfdog breed is a unique dog breed which was created in Central Europe (in that time Czechoslovakian republic) within twenty years of originally experimental crossbreeding of domesticated dog breed (German Shepherd) and wild wolves (*Canis lupus lupus*). Only two wolf-dog breeds are accepted by the Fédération Cynologique Internationale (FCI) as wolf hybrid breeds. The Czechoslovakian Wolfdog breed and Saarloos wolfdog, both belonging to the FCI Group 1. Currently, the CSW has become one of the most popular wolf-like phenotype dog breeds in the world accepted by the FCI (Moravčíková et al. 2021). Number of stallion sire individuals has been gradually increasing during the years. Current number of stallion sirs is 265 (information from February 2022) (Čílová n.d.).

Czechoslovakian Wolfdog breed has not been examined from a genetic point of view before. Therefore, a preliminary study that would compare CSW breed to its parental species was necessary to be done. Microsatellite markers were found to be a suitable markers and method to study and distinguish different dog breeds (Zajc et al. 1997). Microsatellite markers were used in several studies analysing different dog breeds, for example study of Bracco Italiano (Ciampolini et al. 2011) analysing a single breed, or on the other hand a study comparing and recognizing seven breeds - cairn terriers, Dachshunds, flat-coated retrievers, continental toy spaniels, Siberian huskies, Vizslas and whippets, where microsatellite markers were a sufficient method (Schelling et al. 2005).Most dog breeds are closed populations originated from only low number of founders and during the process of their emergence not unfrequently exposure strong selection pressure (Leroy et al. 2009; Mäki 2010). A genetic variability of an Italian dog breed Segugio Italiano and the two related breeds Segugio Maremmano and Segugio dell' Appennino was analysed byPallotti et al. (2017). Even though less microsatellite loci were used compared to this study (Smetanová et al. 2015) both low breeds are similar population size. All alleles were polymorphic in both breeds, which was highly expected in CSW due to its origin. Both breeds are managed by non-random mating

which most likely affect the reduction in heterozygosity. Therefore, it is important to publish the results of similar studies among breeders, who can thus adapt their breeding management(Smetanová et al. 2015; Pallotti et al. 2017).A set of 39 autosomal and 4 Y-linked microsatellites was performed as described in Randi et al. (2014) for the study analysing Czechoslovakian Wolfdog breed. Using the microsatellite markers, we were able to distinguish the Czechoslovakian Wolfdogs from its parental species. Unexpected finding was done when Y-linked microsatellite variability was detected. Only two haplotypes were found in CSW breed, one shared with GS and one private. This information might be very useful when potential wolf-dog hybrids are analysed. A dog Y-haplotype, indicating a past introgression of dog genes, was detected in sampled wolf population in Dalmatia, Croatia and thanks to that the hybrid individuals were found (Kusak et al. 2018).

The hypervariable part of the mtDNA control region was sequenced to determine maternal haplotypes. In the data set of CSWs the occurrence of only two distinct haplotypes occurred. Although the CSW breed is still young, mtDNA sequences of two female wolf founders were not detected. This ascertaining was further supported by information from breeders who did not detect any preserved wolf maternal lineage in pedigrees. Study by Hulva et al. (2018) researched wolf population in Western Carpathians. A low variation was demonstrated among sampled individuals using mtDNA, mainly two haplotypes predominated. Despite the fact that Carpathian wolves and CSW do not share concordant haplotypes, this information may still help during putative hybrid recognition (Smetanová et al. 2015; Hulva et al. 2018).

Another wolfdog breeds is Lupo Italiano, which most likely originated by crossing German Shepherd Dog and an Apennine wolf in 1960s.Even in this case, no shared haplotype was identified between Lupo Italiano and its wolf ancestors (Talenti et al. 2018).By the fact that the CSW is genetically distinguishable from its ancestors it can be used as a reference dog in cases of hybrid detection. Although only two, five respectively, CSW individuals were sampled in the study of Harmoinen (2020). Pure wolves, wolf-dog hybrids, wolfdog breeds (Saarloos wolfdog and CSW) and pure dog breeds were all recognized and allocated to separate group. With regard to the hybrid origin, Czechoslovakian Wolfdog might be considered as a unique and useful breed to study and to understand the wolf-dog hybridization in the wild at least in Central European region. Several studies describing hybrid detection in the wild was released. This phenomenon is usually caused by the occurrence of feral dogs in the wild(Randi 2008; Godinho et al. 2011; Hindrikson et al. 2012).

Although the microsatellites are distributed throughout the whole genome, they offer only limited resolution (Harmoinen 2020). Therefore, in our most recent study (Caniglia et al. 2018)we decided to use SNPs that describe more specific sequences in the genome. Different dog breeds are in the interest and focus of scientists. Braques

Françaises are very popular hunting dogs widely used and bred in France, however, rarely seen elsewhere. Mastrangelo et al. (2018) studied the genomic diversity of this breed, simultaneously it was the first genomic exploration of this low numbered breed.

All used markers (microsatellites, mtDNA or SNPs) show a high proportion of a dog genome in CSW, which might be expected because of only four wolf founders and a strong selection to dog-like behaviour (Hartl & Jedlička 2002).The results show clear evidences about Y-linked and mtDNA haplotypes, these data can help when individuals from the wild or from the captivity need to be identified and assigned to appropriate group. This was applied to a practical use during a very serious case in Italy in 2017 when hundreds of wolf-dogs were illegally sold. This deliberate and illegal crossbreeding between CSW and wolves was done in order to improve the wolf-like phenotype to impress the judges during dog shows. However, breeding domesticated and wild animals is now against the law in Italy, and since wolves are a protected species, animals up to fourth-generation cannot be kept in captivity without proper authorisation from CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora)(Daniela Piccoli n.d.; The UK Wolf Conservation Trust 2017). In Western countries dogs and cats are the most popular pets among humans. With this popularity, however, illegal processes may relate. Information from our research might be helpful for the detecting and fact-finding in such criminal cases.

7 Conclusions

In a study comparing different panels of genetic markers recognizing pure wolves, dogs or hybrid individuals in Italy, we show that the 24 most divergent microsatellites (largest wolf-dog F_{ST} values) were equally or more informative than the panel of 39 loci. A smaller panel of 12 microsatellites increased risks to identify false admixed individuals. The frequency of F1 and F2 was lower than backcrosses or introgressed individuals, suggesting hybridization already occurred some generations in the past, during early phases of wolf expansion from their historical core areas. Empirical and simulated data indicated the identification of the past generation backcrosses is always uncertain, and a larger number of ancestry-informative markers is needed.

In a study evaluating genetic composition of the CSW breed, we have described the impact ofprocesses such as hybridization betweendogs and wolves, population bottlenecks and artificial selection on the genetic diversity patterns in the Czechoslovakian Wolfdog. It seems that hybridization with the dog's wild ancestors may have compensated for the loss ofgenetic variation caused by bottleneck demography. The phase of artificial selection applied to mosaic wolfx dog hybrid genomes was unique for many reasons, especially the following: i) it was very fast compared to the natural domestication process, which possibly lasted thousands of years, and therefore, selection pressures were presumably intense; and ii) it included direct selection for morphological traits, whereas during the first domestication phase, the main selective pressures presumably acted on behavioural (or physiological) traits, with phenotypic variants originating as by-products of these changes (Belyaev 1979). To conclude, the mode and direction of selection acting on particular traits were different compared to the domestication process (e.g., preferring phenotypes of wild ancestors), which enables us to study particular modules of domestication in different contexts. Due to these unique evolutionary pathways, wolfdogs represent interesting models for studying the out- comes of interactions between socialized and wild canid genomes and the role of processes accompanying animal adaptations to the human environment.

Our study provides the first genome-wide characterization of the Czechoslovakian Wolfdog, highlighting how the breed, despite the declared low number of founders, currently shows relatively high levels of heterozygosis thanks to its hybrid ancestry. Our genome-wide approach confirmed to be a valid method in reconstructing the breed history and dating its dynamics, to assess the actual wolf ancestry proportions of single individuals, as well as their relatedness. Therefore, it could provide a valid instrument also for forensic applications in order to unmask possible trades of individuals sold as purebreds but that originated from illegal crossings with wild wolves, which would be difficult to identify through multivariate and

Bayesian assignment procedures based on a limited number of loci or on their morphology alone. Moreover, our gene search approach, made possible by the availability of a well-annotated reference genome, allowed us to identify a first set of genes whose expression and interaction would likely determine the typical wolf-like appearance of the breed. Interestingly, most of the genes associated with brain functions, behaviour, metabolism and disorders we detected are clearly dog-derived, as expected in a breed that, despite its recent hybrid origins, mostly shows typical dog-like phenotypes. The best example is represented by the *COMT* gene, which has been described as the candidate genefor sociability in dogs (Persson et al. 2016) and only its dog alleles have been retained in the gene pool of the analysed CWDs. However, finding the causal mutations for single traits needs further research, in particular for polygenic traits (van Rooy et al. 2014). Future genotyping of a larger number of individuals with certified pedigrees from different lineages sampled worldwide will contribute to a deeper comprehension of many genetic, morphological, and behavioural characteristics of this breed. The optimization of a small and rapid marker panel, for example of 96 SNPs, including also mutations for common diseases or particular behaviours, could help to monitor the health of all the commercialized captive-born individuals and to allow their genomic identification, contrasting unreported crossings and illegal trading of wild wolves.

8 References

Albert FW et al. 2009. Genetic architecture of tameness in a rat model of animal domestication. Genetics **182**:541–554.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Research **19**:1655–1664.

Allendorf FW, Leary RF, Spruell P, John K. Wenburg. 2001. The upper palaeolithic of Britain: a study of man and nature in the late ice age. TRENDSin Ecology & Evolution **16**:613–622.

Almudevar A. 2000. Exact Confidence Regions for Species Assignment Based on DNA Markers. The Canadian Journal of Statistics / La Revue Canadienne de Statistique **28**:81.

Anderson EC, Thompson EA. 2002. A model-based method for identifying species hybrids using multilocus genetic data. Genetics **160**:1217–1229.

Anderson TM et al. 2009. Molecular and Evolutionary History of Melanism in North American Gray Wolves. Science **323**:277–313.

Arnott ER, Peek L, Early JB, Pan AYH, Haase B, Chew T, McGreevy PD, Wade CM. 2015. Strong selection for behavioural resilience in Australian stock working dogs identified by selective sweep analysis. Canine Genetics and Epidemiology **2**:1–6. ??? Available from ???

Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, Liberg O, Arnemo JM, Hedhammar Å, Lindblad-Toh K. 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature **495**:360–364.

B. S. WEIR AND C. CLARK COCKERHAM. 1984. ESTIMATING F-STATISTICS FOR THE ANALYSIS OF POPULATION STRUCTURE. Evolution **13**:5823–5832.

Banks MA, Eichert W. 2000. WHICHRUN (version 3.2): a computer program for population assignment of individuals based on multilocus genotype data. Journal of Heredity **91**: 87–89.

Banks MA, Eichert W, Olsen JB. 2003. Which genetic loci have greater population assignment power? Bioinformatics **19**:1436–1438.

Barton NH. 2000. Genetic hitchhiking. Philosophical Transactions of the Royal Society B: Biological Sciences **355**:1553–1562.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F. 1996. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire génome, populations, interactions, CNRS UMR **5000**: 1996–2004.

Bélteky J, Agnvall B, Jensen P. 2017. Gene expression of behaviorally relevant genes in the cerebral hemisphere changes after selection for tameness in Red Junglefowl. PLoS ONE **12**:1–12.

Belyaev DK. 1979. Destabilizing selection as a factor in domestication. Journal of heredity **70**:301–308.

Benazzo A et al. 2017. Survival and divergence in a small group: The extraordinary genomic history of the endangered Apennine brown bear stragglers. Proceedings of the National Academy of Sciences of the United States of America **114**:E9589–E9597.

Bergström A et al. 2020. Origins and genetic legacy of prehistoric dogs. Science **370**:557–564.

Bigi D, Marelli SP, Randi E, Polli M. 2015. Genetic characterization of four native Italian shepherd dog breeds and analysis of their relationship to cosmopolitan dog

breeds using microsatellite markers. Animal **9**:1921–1928. Elsevier. Available from http://dx.doi.org/10.1017/S1751731115001561.

Bjelland DW, Weigel KA, Vukasinovic N, Nkrumah JD. 2013. Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. Journal of Dairy Science **96**:4697–4706. Elsevier. Available from http://dx.doi.org/10.3168/jds.2012-6435.

Boitani L, Paolo Ciucci, Alesia Ortolani. 2007. The behavioural biology of dogs. Pages 147–165 in Per Jensen, editor. The Behavioural Biology of Dogs. Available from www.cabi.org.

Boscari E, Barmintseva A, Pujolar JM, Doukakis P, Mugue N, Congiu L. 2014. Species and hybrid identification of sturgeon caviar: A new molecular approach to detect illegal trade. Molecular Ecology Resources **14**:489–498.

Boyko AR. 2011. The domestic dog: man's best friend in the genomic era. Genome Biology **12**.

Brennan AC, Woodward G, Seehausen O, Muñoz-Fuentes V, Moritz C, Guelmami A, Abbott RJ, Edelaar P, Edelaar P. 2014. Hybridization due to changing species distributions: adding problems or solutions to conservation of biodiversity during global change? Consult the copyright statement on the inside front cover for non-commercial copying policies. Evolutionary Ecology Research:475–491. Available from http://www.evolutionary-ecology.com/open/v16n06/ffar2905.pdf.

Brisbin A, Bryc K, Byrnes J, Zakharia F, Omberg L, Degenhardt J, Reynolds A, Ostrer H, Mezey JG, Bustamante CD. 2012. Pcadmix: Principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. Human Biology **84**:343–364.

Brumfield RT. 2010. Speciation genetics of biological invasions with hybridization. Molecular Ecology **19**:5079–5083.

Candille SI, Kaelin CB, Cattanach BM, Yu B, Thompson DA, Nix MA, Kerns JA, Schmutz SM, Millhauser GL, Barsh GS. 2007. A β-defensin mutation causes black coat color in domestic dogs. Science **318**:1418–1423.

Caniglia R et al. 2018. Wolf outside, dog inside? The genomic make-up of the Czechoslovakian Wolfdog. BMC Genomics **19**:1–17.

Caniglia R, Fabbri E, Cubaynes S, Randi E. 2012. An improved procedure to estimate wolf abundance using non-invasive genetic sampling and capture – recapture mixture models:53–64.

Caniglia R, Fabbri E, Greco C, Galaverni M, Manghi L, Boitani L, Sforzi A, Randi E. 2013. Black coats in an admixed wolf × dog pack is melanism an indicator of hybridization in wolves? European Journal of Wildlife Research **59**:543–555.

Cattanach BM, Dukes-McEwan J, Wotton PR, Stephenson HM, Hamilton RM. 2015. Paper: A pedigree-based genetic appraisal of Boxer ARVC and the role of the Striatin mutation. Veterinary Record **176**:492.

CESKOSLOVENSKÝ VLCIAK. (n.d.). Available from http://www.fci.be/en/nomenclature/CZECHOSLOVAKIAN-WOLFDOG-332.html (accessed August 26, 2020).

Chapron G et al. 2014. 27 Jasna Jeremić, 28 Klemen Jerina, 29 Gesa Kluth, 30 Felix Knauer, 2 Ilpo Kojola, 31 Ivan Kos, 29 Miha Krofel, 29 Jakub Kubala, 32 Saša Kunovac, 33 Josip Kusak, 5 Miroslav Kutal, 34,35 Olof Liberg, 1 Aleksandra Majić, 29 Peep Männil, 36 Ralph Manz, 4 Eri:25. Available from http://www.ncdc.noaa.gov/paleo/study/17455%0Ahttp://doi.pangaea.de/10.1594/PANGAEA.839454%0Ahttp://www.ncdc.noaa.gov/paleo/study/17455%0Ahttp://do

i.pangaea.de/10.1594/PANGAEA.839454.
Chase K, Jones P, Martin A, Ostrander EA, Lark KG. 2009. Genetic mapping of fixed phenotypes: Disease frequency as a breed characteristic. Journal of Heredity **100**:S37–S41.
Ciampolini R, Cecchi F, Bramante A, Casetti F, Presciuttini S. 2011. Genetic variability of the Bracco Italiano dog breed based on microsatellite polymorphism. Italian Journal of Animal Science **10**:267–270.
Cieslak M, Reissmann M, Hofreiter M, Ludwig A. 2011. Colours of domestication. Biological Reviews **86**:885–899.
Čílová D. (n.d.). Seznam chovných psů. Available from http://www.cswolfdog.cz/index.php/chov/seznam-chovnych-psu-tl (accessed September 2, 2020).
Cilova D, Vejl P, Sebkova N, Castkova M, Jurkovicova P, Kadlecova V. 2011. Microsatellite analysis of X and Y gonosome variability in the czechoslovakian and saarloos wolfdog breeds. Journal of Veterinary Behavior **6**:60–61. Elsevier Inc. Available from http://dx.doi.org/10.1016/j.jveb.2010.09.054.
Ciucci P, Chapron G, Guberti V, Boitani L. 2007. Estimation of mortality parameters from (biased) samples at death: Are we getting the basics right in wildlife field studies? A response to Lovari et al. (2007) [1]. Journal of Zoology **273**:125–127.
Ciucci P, Reggioni W, Maiorano L, Boitani L. 2009. Long-Distance Dispersal of a Rescued Wolf From the Northern Apennines to the Western Alps. Journal of Wildlife Management **73**:1300–1306.
Corander J, Marttinen P. 2006. Bayesian identification of admixture events using multilocus molecular markers. Molecular Ecology **15**:2833–2843.
Corander J, Marttinen P, Sirén J, Tang J. 2008. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. BMC bioinformatics **9**:539.
Daniela Piccoli. (n.d.). Hundreds of wolf-dog hybrids illegally sold in Italy. Available from https://www.thelocal.it/20170116/hundreds-of-wolf-dog-hybrids-seized-in-italy/.
de Groot GA et al. 2016. Decades of population genetic research reveal the need for harmonization of molecular markers: The grey wolf Canis lupus as a case study. Mammal Review **46**:44–59.
Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR. 2014. NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Molecular Ecology Resources **14**:209–214.
Donfrancesco V et al. 2019. Unravelling the scientific debate on how to address wolf-dog hybridization in Europe. Frontiers in Ecology and Evolution **7**:1–13.
Downs LM, Wallin-Håkansson B, Boursnell M, Marklund S, Hedhammar Å, Truvé K, Hübinette L, Lindblad-Toh K, Bergström T, Mellersh CS. 2011. A frameshift mutation in Golden Retriever dogs with progressive retinal atrophy endorses SLC4A3 as a candidate gene for human retinal degenerations. PLoS ONE **6**.
Drake AG, Klingenberg CP. 2010. Large-scale diversification of skull shape in domestic dogs: Disparity and modularity. American Naturalist **175**:289–301.
Dreger DL, Parker HG, Ostrander EA, Schmutz SM. 2013. Identification of a mutation that is associated with the saddle tan and black-and-tan phenotypes in Basset Hounds and Pembroke Welsh Corgis. Journal of Heredity **104**:399–406.
Dreger DL, Rimbault M, Davis BW, Bhatnagar A, Parker HG, Ostrander EA. 2016. Whole-genome sequence, SNP chips and pedigree structure: Building demographic profiles in domestic dog breeds to optimize genetic-trait mapping. DMM Disease

Models and Mechanisms **9**:1445–1460.
Duchesne P, Turgeon J. 2009. FLOCK: a method for quick mapping of admixture without source samples. Molecular Ecology Resources **9**: 1333–1344.
Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources **4**:359–361.
Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology **14**:2611–2620.
Excoffier L, Foll M, Petit RJ. 2009. Genetic consequences of range expansions. Annual Review of Ecology, Evolution, and Systematics **40**:481–501.
Fabbri E et al. 2007. From the Apennines to the Alps: Colonization genetics of the naturally expanding Italian wolf (Canis lupus) population. Molecular Ecology **16**:1661–1671.
Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics **164**:1567–1587.
Fan Z et al. 2016. Worldwide patterns of genomic variation and admixture in gray wolves. Genome Research **26**:163–173.
Ferenčaković M, Hamzić E, Gredler B, Solberg TR, Klemetsdal G, Curik I, Sölkner J. 2013. Estimates of autozygosity derived from runs of homozygosity: Empirical evidence from selected cattle populations. Journal of Animal Breeding and Genetics **130**:286–293.
Feulner PGD et al. 2013. Genome-wide patterns of standing genetic variation in a marine population of three-spined sticklebacks. Molecular Ecology **22**:635–649.
Frantz LAF et al. 2016. Genomic and archaeological evidence suggests a dual origin of domestic dogs. Science **352**:1228–1231.
Freedman AH et al. 2016. Demographically-Based Evaluation of Genomic Regions under Selection in Domestic Dogs. PLoS Genetics **12**:1–23.
Galaverni M, Caniglia R, Pagani L, Fabbri E, Boattini A, Randi E. 2017. Disentangling timing of admixture, patterns of introgression, and phenotypic indicators in a hybridizing Wolf population. Molecular Biology and Evolution **34**:2324–2339.
Glémin S. 2003. How Are Deleterious Mutations Purged? Drift Versus Nonrandom Mating. Evolution **57**:2678.
Godinho R, Llaneza L, Blanco JC, Lopes S, Álvares F, García EJ, Palacios V, Cortés Y, Talegón J, Ferrand N. 2011. Genetic evidence for multiple events of hybridization between wolves and domestic dogs in the Iberian Peninsula. Molecular Ecology **20**:5154–5166.
Gohli J, Voje KL. 2016. An interspecific assessment of Bergmann's rule in 22 mammalian families. BMC Evolutionary Biology **16**:1–12. BMC Evolutionary Biology. Available from http://dx.doi.org/10.1186/s12862-016-0778-x.
Gopalakrishnan S et al. 2017. The wolf reference genome sequence (Canis lupus lupus) and its implications for Canis spp. population genomics. BMC Genomics **18**:1–11. BMC Genomics.
Harmoinen J. 2020. Reliable Wolf-Dog Hybrid Detection in Europe Using a Reduced SNP Panel Developed for Non-Invasively Collected Samples. Research Square:1–30. Available from https://doi.org/10.21203/rs.3.rs-113866/v1.
Hartl IK, Jedlička J. 2002. Československý Vlčák. LOBA, Brno. Available from http://www.vlci-legenda.com/download/csv.pdf.

Hayes BJ, Visscher PM, McPartlan HC, Goddard ME. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Research **13**:635–643.

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.

Hindrikson M, Männil P, Ozolins J, Krzywinski A, Saarma U. 2012. Bucking the Trend in Wolf-Dog Hybridization: First Evidence from Europe of Hybridization between Female Dogs and Male Wolves. PLoS ONE **7**:1–12.

Hoban S. 2014. An overview of the utility of population simulation software in molecular ecology. Molecular Ecology **23**:2383–2401.

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.

Hundertmark KJ, van Daele LJ. 2010. Founder effect and bottleneck signatures in an introduced, insular population of elk. Conservation Genetics **11**:139–147.

Hurda P. (n.d.). Zpráva o chovu za období 2018-2019. Available from http://www.cswolfdog.cz/images/Konference_2020/Ploha 11 Zprva o chovu za obdob 2018-2019.pdf (accessed September 2, 2020).

Huson HJ, Parker HG, Runstadler J, Ostrander EA. 2010. A genetic dissection of breed composition and performance enhancement in the Alaskan sled dog. BMC Genetics **11**.

Iacolina L, Stronen A V., Pertoldi C, Tokarska M, Nørgaard LS, Muñoz J, Kjærsgaard A, Ruiz-Gonzalez A, Kaminski S, Purfield DC. 2016. Novel graphical analyses of runs of homozygosity among species and livestock breeds. International Journal of Genomics **2016**.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics **11**: 94.

Julie D.Thompson DGH and TJG. 2008. CLUSTAL W (improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice). Encyclopedia of Genetics, Genomics, Proteomics and Informatics **22**:376–377.

Kaelin CB, Barsh GS. 2013. Genetics of pigmentation in dogs and cats. Annual Review of Animal Biosciences **1**:125–156.

Kardos M, Luikart G, Allendorf FW. 2015. Measuring individual inbreeding in the age of genomics: Marker-based measures are better than pedigrees. Heredity **115**:63–72. Nature Publishing Group. Available from http://dx.doi.org/10.1038/hdy.2015.17.

Kerns JA, Newton J, Berryere TG, Rubin EM, Cheng JF, Schmutz SM, Barsh GS. 2004. Characterization of the dog Agouti gene and a nonagouti mutation in German Shepherd Dogs. Mammalian Genome **15**:798–808.

Kim ES, Sonstegard TS, Van Tassell CP, Wiggans G, Rothschild MF. 2015. The relationship between runs of homozygosity and inbreeding in Jersey cattle under selection. PLoS ONE **10**:1–17.

Kusak J, Fabbri E, Galov A, Gomerčić T, Arbanasić H, Caniglia R, Galaverni M, Reljić S, Huber D, Randi E. 2018. Wolf-dog hybridization in Croatia. Veterinarski Arhiv **88**:375–395.

Kutal M, Rigg R. 2008. Perspectives of wolves in Central Europe: Proceedings from the conference held on 9th April 2008.

Lahtinen M, Clinnick D, Mannermaa K, Salonen JS, Viranta S. 2021. Excess protein enabled dog domestication during severe Ice Age winters. Scientific Reports **11**:1–5. Nature Publishing Group UK. Available from https://doi.org/10.1038/s41598-020-78214-4.

Larson G et al. 2012. Rethinking dog domestication by integrating genetics, archeology, and biogeography. Proceedings of the National Academy of Sciences of the United States of America **109**:8878–8883.

Larson G et al. 2014. Current perspectives and the future of domestication studies. Proceedings of the National Academy of Sciences of the United States of America **111**:6139–6146.

Larson G, Burger J. 2013. A population genetics view of animal domestication. Trends in Genetics **29**:1–9. Elsevier Ltd. Available from http://dx.doi.org/10.1016/j.tig.2013.01.003 (accessed August 26, 2020).

Lawson DJ, van Dorp L, Falush D. 2018. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nature Communications **9**:1–11. Springer US. Available from http://dx.doi.org/10.1038/s41467-018-05257-7.

Le Guédard S, Faugère V, Malcolm S, Claustres M, Roux AF. 2007. Large genomic rearrangements within the PCDH15 gene are a significant cause of USH1F syndrome. Molecular Vision **13**:102–107.

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

Leroy G, Verrier E, Meriaux JC, Rognon X. 2009. Genetic diversity of dog breeds: Between-breed diversity, breed assignation and conservation approaches. Animal Genetics **40**:333–343.

Li Y, Von Holdt BM, Reynolds A, Boyko AR, Wayne RK, Wu DD, Zhang YP. 2013. Artificial selection on brain-expressed genes during the domestication of dog. Molecular Biology and Evolution **30**:1867–1876.

Librado P, Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25**:1451–1452.

Loh PR, Lipson M, Patterson N, Moorjani P, Pickrell JK, Reich D, Berger B. 2013. Inferring admixture histories of human populations using linkage disequilibrium. Genetics **193**:1233–1254.

Lucchini V, Galov A, Randi E. 2004. Evidence of genetic distinction and long-term population decline in wolves (Canis lupus) in the Italian Apennines. Molecular Ecology **13**:523–536.

Maglieri V, Prato-Previde E, Tommasi E, Palagi E. 2019. Wolf-like or dog-like? A comparison of gazing behaviour across three dog breeds tested in their familiar environments. Royal Society Open Science **6**.

Mäki K. 2010. Population structure and genetic diversity of worldwide Nova Scotia Duck Tolling Retriever and Lancashire Heeler dog populations. Journal of Animal Breeding and Genetics **127**:318–326.

Mastorakos G, Ilias I. 2003. Maternal and fetal hypothalamic-pituitary-adrenal axes during pregnancy and postpartum. Annals of the New York Academy of Sciences **997**:136–149.

Mastrangelo S, Filippo B, Auzino B, Marco R, Spaterna A, Ciampolini R. 2018. Genome-wide diversity and runs of homozygosity in the “Braque Français, type Pyrénées” dog breed. BMC Research Notes **11**:11–16. BioMed Central. Available

from https://doi.org/10.1186/s13104-017-3112-9.
McQuillan R et al. 2008. Runs of Homozygosity in European Populations (DOI:10.1016/j.ajhg.2008.08.007). American Journal of Human Genetics **83**:658. Available from http://dx.doi.org/10.1016/j.ajhg.2008.10.009.
Michalakis Y, Excoffier L. 1996. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics **142**:1061–1064.
Miklósi A, Kubinyi E, Topál J, Gácsi M, Virányi Z, Csányi V. 2003. A Simple Reason for a Big Difference: Wolves Do Not Look Back at Humans, but Dogs Do. Current Biology **13**:763–766.
Miller W et al. 2012. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. Proceedings of the National Academy of Sciences of the United States of America **109**.
Miyadera K, Acland GM, Aguirre GD. 2012. Genetic and phenotypic variations of inherited retinal diseases in dogs: The power of within- and across-breed studies. Mammalian Genome **23**:40–61.
Moravčíková N, Kasarda R, Židek R, Vostrý L, Vostrá-Vydrová H, Vašek J, Čílová D. 2021. Czechoslovakian Wolfdog genomic divergence from its ancestors canis lupus, german shepherd dog, and different sheepdogs of european origin. Genes **12**:1–18.
Morey DF, Jeger R. 2017. From wolf to dog: Late Pleistocene ecological dynamics, altered trophic strategies, and shifting human perceptions. Historical Biology **29**:895–903. Taylor & Francis. Available from http://dx.doi.org/10.1080/08912963.2016.1262854.
Morrel V. 1997. The origin of dogs: running with the wolves. Science **276**:1647–1648.
Muñoz-Fuentes V, Darimont CT, Paquet PC, Leonard JA. 2010. The genetic legacy of extirpation and re-colonization in Vancouver Island wolves. Conservation Genetics **11**:547–556.
Nagasawa M et al. 2015. Oxytocin-Gaze Positive Loop and the Coevolution of Human-Dog Bonds Common Variants Spanning PLK4 Are Associated With Mitotic-Origin Aneuploidy in Human Embryos. Science **348**:450–451.
Nielsen EE, Bach LA, Kotlicki P. 2006. HYBRIDLAB (version 1.0): A program for generating simulated hybrids from population samples. Molecular Ecology Notes **6**:971–973.
Novembre J, Stephens M. 2008. Interpreting principal component analyses of spatial population genetic variation. Nature Genetics **40**:646–649.
Ollivier M et al. 2013. Evidence of Coat Color Variation Sheds New Light on Ancient Canids. PLoS ONE **8**.
Ostrander EA. 2012. Both Ends of the Leash — The Human Links to Good Dogs with Bad Genes. New England Journal of Medicine **367**:636–646.
Ostrander EA, Wayne RK. 2005. The canine genome. Genome Research **15**:1706–1716.
Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW. 2010. Conservation genetics in transition to conservation genomics. Trends in Genetics **26**:177–187.
Pallotti S, La Terza A, De Cosmo A, Pediconi D, Pazzaglia I, Nocelli C, Renieri C. 2017. Genetic variability of the short-haired and rough-haired Segugio Italiano dog breeds and their genetic distance from the other related Segugio breeds. Italian Journal of Animal Science **16**:531–537. Informa Healthcare USA, Inc. Available from https://doi.org/10.1080/1828051X.2017.1317221.
Parker HG, Dreger DL, Rimbault M, Davis BW, Mullen AB, Carpintero-Ramirez G,

Ostrander EA. 2017. Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development. Cell Reports **19**:697–708. ElsevierCompany. Available from http://dx.doi.org/10.1016/j.celrep.2017.03.079.

Parker HG, Kukekova A V., Akey DT, Goldstein O, Kirkness EF, Baysac KC, Mosher DS, Aguirre GD, Acland GM, Ostrander EA. 2007. Breed relationships facilitate fine-mapping studies: A 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds. Genome Research **17**:1562–1571.

Parker J, Tsagkogeorga G, Cotton JA, Liu Y, Provero P, Stupka E, Rossiter SJ. 2013. Genome-wide signatures of convergent evolution in echolocating mammals. Nature **502**:228–231. Nature Publishing Group. Available from http://dx.doi.org/10.1038/nature12511.

Peakall R, Smouse PE. 2012. GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics **28**:2537–2539.

Pedersen N, Liu H, Theilen G, Sacks B. 2013. The effects of dog breed development on genetic diversity and the relative influences of performance and conformation breeding. Journal of Animal Breeding and Genetics **130**:236–248.

Pendleton AL, Shen F, Taravella AM, Emery S, Veeramah KR, Boyko AR, Kidd JM. 2018. Comparison of village dog and wolf genomes highlights the role of the neural crest in dog domestication. BMC Biology:1–21. BMC Biology. Available from https://doi.org/10.1186/s12915-018-0535-2 (accessed August 26, 2020).

Perri AR, Feuerborn TR, Frantz LAF, Larson G, Malhi RS, Meltzer DJ, Witt KE. 2021. Dog domestication and the dual dispersal of people and dogs into the Americas. Proceedings of the National Academy of Sciences of the United States of America **118**:1–8.

Persson ME, Wright D, Roth LSV, Batakis P, Jensen P. 2016. Genomic regions associated with interspecies communication in dogs contain genes related to human social disorders. Scientific Reports **6**:1–9. Nature Publishing Group. Available from http://dx.doi.org/10.1038/srep33439.

Pértille F, Brantsæter M, Nordgreen J, Coutinho LL, Janczak AM, Jensen P, Guerrero-Bosagna C. 2017. DNA methylation profiles in red blood cells of adult hens correlate with their rearing conditions. Journal of Experimental Biology **220**:3579–3587.

Pfahler S, Distl O. 2015. Effective population size, extended linkage disequilibrium and signatures of selection in the rare dog breed Lundehund. PLoS ONE **10**:1–14.

Pilot M et al. 2021. Human-modified canids in human-modified landscapes: The evolutionary consequences of hybridization for grey wolves and free-ranging domestic dogs. Evolutionary Applications:1–24.

Pilot M, Branicki W, Jędrzejewski W, Goszczyski J, Jędrzejewska B, Dykyy I, Shkvyrya M, Tsingarska E. 2010. Phylogeographic history of grey wolves in Europe. BMC Evolutionary Biology **10**.

Pilot M, Greco C, Vonholdt BM, Jędrzejewska B, Randi E, Jędrzejewski W, Sidorovich VE, Ostrander EA, Wayne RK. 2014. Genome-wide signatures of population bottlenecks and diversifying selection in European wolves. Heredity **112**:428–442.

Pilot M, Greco C, vonHoldt BM, Randi E, Jędrzejewski W, Sidorovich VE, Konopiński MK, Ostrander EA, Wayne RK. 2018. Widespread, long-term admixture between grey wolves and domestic dogs across Eurasia and its implications for the conservation status of hybrids. Evolutionary Applications **11**:662–680.

Pilot M, Jedrzejewski W, Branicki W, Sidorovich VE, Jedrzejewska B, Stachura K, Funk SM. 2006. Ecological factors influence population genetic structure of European grey wolves. Molecular Ecology **15**:4533–4553.

Piry S, Alapetite A, Cornuet J-M, Paetkau D, Baudouin L. 2004. GENECLASS 2: a software for genetic assignment and first-generation migrant detection. Journal of Heredity **95**: 536–539.

Plassais J et al. 2022a. Natural and human-driven selection of a single non-coding body size variant in ancient and modern canids. Current Biology **32**:889-897.e9.

Plassais J et al. 2022b. Natural and human-driven selection of a single non-coding body size variant in ancient and modern canids. Current biology : CB:1–9. Available from http://www.ncbi.nlm.nih.gov/pubmed/35090588.

Pollinger JP, Bustamante CD, Fledel-Alon A, Schmutz S, Gray MM, Wayne RK. 2005. Selective sweep mapping of genes with large phenotypic effects. Genome Research **15**:1809–1819.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics **38**:904–909.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics **155**:945–959.

Purcell S et al. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics **81**:559–575.

Ramos-Madrigal J et al. 2021. Genomes of Pleistocene Siberian Wolves Uncover Multiple Extinct Wolf Lineages. Current Biology **31**:198-206.e8.

Rannala B, Mountain JL. 1997. Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences, USA **94**: 9197–9201.

Randi E. 2008. Detecting hybridization between wild species and their domesticated relatives. Molecular Ecology **17**:285–293.

Randi E. 2011. Genetics and conservation of wolves Canis lupus in Europe. Mammal Review **41**:99–111.

Randi E, Hulva P, Fabbri E, Galaverni M, Galov A. 2014. Multilocus Detection of Wolf x Dog Hybridization in Italy, and Guidelines for Marker Selection. PLoS ONE **9**:86409. Available from www.plosone.org.

Randi E, Lucchini V. 2002. Detecting rare introgression of domestic dog genes into wild wolf (Canis lupus) populations by Bayesian admixture analyses of microsatellite variation. Conservation Genetics **3**:31–45.

Randi E, Lucchini V, Christensen MF, Mucci N, Funk SM, Dolf G, Loeschcke V. 2000. Mitochondrial DNA variability in Italian and east European wolves: Detecting the consequences of small population size and hybridization. Conservation Biology **14**:464–473.

Rice WR. 1989. Analysing Tables of Statistical Tests. Evolution **43**:223–225. Available from http://projecteuclid.org/euclid.ss/1177011945%0Ahttp://journals.sagepub.com/doi/10.1177/0003122413496252.

Rieseberg LH, Archer MA, Wayne RK. 1999. Transgressive segregation, adaptation and speciation. Heredity **83**:363–372.

Rosenberg NA. 2004. DISTRUCT: A program for the graphical display of population structure. Molecular Ecology Notes **4**:137–138.

Rosenberg NA, Li LM, Ward R, Pritchard JK. 2003. Informativeness of Genetic Markers for Inference of Ancestry. American Journal of Human Genetics **73**:1402–

1422.
Rousset F. 2008. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources **8**:103–106.
Salvatori V et al. 2020. European agreements for nature conservation need to explicitly address wolf-dog hybridisation. Biological Conservation **248**:108525. Elsevier. Available from https://doi.org/10.1016/j.biocon.2020.108525.
Santostasi NL, Gimenez O, Caniglia R, Fabbri E, Molinari L, Reggioni W, Ciucci P. 2021. Estimating Admixture at the Population Scale: Taking Imperfect Detectability and Uncertainty in Hybrid Classification Seriously. Journal of Wildlife Management:1–16.
Schelling C, Gaillard C, Dolf G. 2005. Genetic variability of seven dog breeds based on microsatellite markers. Journal of Animal Breeding and Genetics, Supplement **122**:71–77.
Schoenebeck JJ et al. 2012. Variation of BMP3 Contributes to Dog Breed Skull Diversity. PLoS Genetics **8**:1–12.
Schweizer RM, VonHoldt BM, Harrigan R, Knowles JC, Musiani M, Coltman D, Novembre J, Wayne RK. 2016. Genetic subdivision and candidate genes under selection in North American grey wolves. Molecular Ecology **25**:380–402.
Seifi Moroudi R, Masoudi AA, Vaez Torshizi R, Zandi M. 2014. Identification of learning and memory genes in canine; promoter investigation and determining the selective pressure. Molecular Biology Reports **41**:8019–8030.
Shriver MD, Smith MW, Jin L, Marcini A, Akey JM. 1997. Ethnic-affiliation estimation by use of population-specific DNA markers. American Journal of Human Genetics **60**: 957.
Skoglund P, Ersmark E, Palkopoulou E, Dalén L. 2015. Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. Current Biology **25**:1515–1519. Elsevier Ltd. Available from http://dx.doi.org/10.1016/j.cub.2015.04.019.
Skoglund P, Götherström A, Jakobsson M. 2011. Estimation of population divergence times from non-overlapping genomic sequences: Examples from dogs and wolves. Molecular Biology and Evolution **28**:1505–1517.
Smetanová M, Bolfíková BČ, Randi E, Caniglia R, Fabbri E, Galaverni M, Kutal M, Hulva P. 2015. From Wolves to Dogs, and Back: Genetic Composition of the Czechoslovakian Wolfdog. PLoS ONE **10**:1–11.
Sommese A, Valsecchi P, Pelosi A, Prato-Previde E. 2021. Comparing behavioural characteristics of Czechoslovakian Wolfdogs, German shepherds and Labrador retrievers in Italy and the Czech Republic. Applied Animal Behaviour Science **237**:105300. Elsevier B.V. Available from https://doi.org/10.1016/j.applanim.2021.105300.
Stone HR, McGreevy PD, Starling MJ, Forkman B. 2016. Associations between domestic-dog morphology and behaviour scores in the dog mentality assessment. PLoS ONE **11**:1–14.
Stronen A V. et al. 2013. North-South Differentiation and a Region of High Diversity in European Wolves (Canis lupus). PLoS ONE **8**:1–9.
Stronen A V. et al. 2017. Genetic rescue of an endangered domestic animal through outcrossing with closely related breeds: A case study of the Norwegian Lundehund. PLoS ONE **12**:1–18.
Sundman AS, Pértille F, Coutinho LL, Jazin E, Guerrero-Bosagna C, Jensen P. 2020. DNA methylation in canine brains is related to domestication and dog-breed

formation. PLoS ONE **15**:1–20.
Sundqvist. 2001. Y chromosome haplotyping in Scandinavian wolves (Canis lupus) based on microsatellite markers. Molecular Ecology **10**:1959–1966.
Sykes N et al. 2020. Addressing Global Challenges:1–18.
Talenti A et al. 2018. Studies of modern Italian dog populations reveal multiple patterns for domestic breed evolution. Ecology and Evolution **8**:2911–2925.
Taylor SA, Larson EL. 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. Nature Ecology and Evolution **3**:170–177. Springer US. Available from http://dx.doi.org/10.1038/s41559-018-0777-y.
The Creation of Leendert Saarloos. 2011.:182–184.
The UK Wolf Conservation Trust. 2017. Wolf Print. Illegality breeds contempt: the hybridisation of wolves and Czechoslovakian wolf-dogs in Italy:1–31.
Trut L, Oskina I, Kharlamova A. 2009. Animal evolution during domestication: The domesticated fox as a model. BioEssays **31**:349–360.
Vähä JP, Primmer CR. 2006. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. Molecular Ecology **15**:63–72.
Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes **4**:535–538.
van Rooy D, Arnott ER, Early JB, McGreevy P, Wade CM. 2014. Holding back the genes: limitations of research into canine behavioural genetics. Canine Genetics and Epidemiology **1**:7.
Vaysse A et al. 2011. Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. PLoS Genetics **7**.
Verardi A, Lucchini V, Randi E. 2006. Detecting introgressive hybridization between free-ranging domestic dogs and wild wolves (Canis lupus) by admixture linkage disequilibrium analysis. Molecular Ecology **15**:2845–2855.
Verginelli F, Capelli C, Coia V, Musiani M, Falchetti M, Ottini L, Palmirotta R, Tagliacozzo A, De Grossi Mazzorin I, Mariani-Costantini R. 2005. Mitochondrial DNA from prehistoric canids highlights relationships between dogs and South-East European wolves. Molecular Biology and Evolution **22**:2541–2551.
Vilà C, Amorim IR, Leonard JA, Posada D, Castroviejo J, Petrucci-Fonseca F, Crandall KA, Ellegren H, Wayne RK. 1999. Mitochondrial DNA phylogeography and population history of the grey wolf Canis lupus. Molecular Ecology **8**:2089–2103.
Vilà C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, Crandall KA, Lundeberg J, Wayne RK. 1997. Multiple and ancient origins of the domestic dog. Science **276**:1687–1689.
Vonholdt BM et al. 2010. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. Nature **464**:898–902. Nature Publishing Group. Available from http://dx.doi.org/10.1038/nature08837.
VonHoldt BM et al. 2011. A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. Genome Research **21**:1294–1305.
VonHoldt BM, Cahill JA, Fan Z, Gronau I, Robinson J, Pollinger JP, Shapiro B, Wall J, Wayne RK. 2016. Whole-genome sequence analysis shows that two endemic species of North American Wolf are admixtures of the coyote and gray Wolf. Science Advances **2**:1–14.
vonHoldt BM, Pollinger JP, Earl DA, Parker HG, Ostrander EA, Wayne RK. 2013. Identification of recent hybridization between gray wolves and domesticated dogs

by SNP genotyping. Mammalian Genome **24**:80–88.
Voorbij AMWY, Leegwater PA, Kooistra HS. 2014. Pituitary Dwarfism in Saarloos and Czechoslovakian Wolfdogs is Associated with a Mutation in LHX3. Journal of Veterinary Internal Medicine **28**:1770–1774.
Waits LP, Luikart G, Taberlet P. 2001. Waits LP , Luikart G , Taberlet P . Estimating the probability of identity among genotypes in natural populations : Cautions and guidelines . Mol Ecol 10 : 249-256 Estimating the probability of identity among genotypes in natural populations : cautions an. Molecular Ecology **10**:249–256.
Wang GD et al. 2016. Out of southern East Asia: The natural history of domestic dogs across the world. Cell Research **26**:21–33. Nature Publishing Group.
Wang MS et al. 2020. Ancient hybridization with an unknown population facilitated high-altitude adaptation of canids. Molecular Biology and Evolution **37**:2616–2629.
Waples RS, Do C. 2008. LDNE: A program for estimating effective population size from data on linkage disequilibrium. Molecular Ecology Resources **8**:753–756.
Waples RS, Do C. 2010. Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: A largely untapped resource for applied conservation and evolution. Evolutionary Applications **3**:244–262.
Wayne RK. 1993, June 1. Molecular evolution of the dog family. Elsevier Current Trends.
Wayne RK, Vonholdt BM. 2012. Evolutionary genomics of dog domestication. Mammalian Genome **23**:3–18.
Werhahn G et al. 2020. Himalayan wolf distribution and admixture based on multiple genetic markers. Journal of Biogeography **47**:1272–1285.
Wilbe M et al. 2010. Genome-wide association mapping identifies multiple loci for a canine SLE-related disease complex. Nature Genetics **42**:250–254. Nature Publishing Group. Available from http://dx.doi.org/10.1038/ng.525.
Wilkins AS, Wrangham RW, Tecumseh Fitch W. 2014. The “domestication syndrome” in mammals: A unified explanation based on neural crest cell behavior and genetics. Genetics **197**:795–808.
Yang H et al. 2017. The origin of chow chows in the light of the East Asian breeds. BMC Genomics **18**:1–13. BMC Genomics.
Zajc I, Mellersh CS, Sampson J. 1997. Variability of canine microsatellites within and between different dog breeds. Mammalian Genome **8**:182–185.
Zeder MA. 2012. Pathways to animal domestication. Page Biodiversity in Agriculture: Domestication, Evolution, and Sustainability.
Zeder MA. 2015. Core questions in domestication research. Proceedings of the National Academy of Sciences of the United States of America **112**:3191–3198.
Zeng R et al. 2014. Breed distribution of SOD1 alleles previously associated with canine degenerative myelopathy. Journal of Veterinary Internal Medicine **28**:515–521.

9 Appendices

9.1 Supplementary materials

Table S1-add 1. Description of the genotyped autosomal (CFA) and Y-linked (CFAY) microsatellites (STR), *Amelogenin* and *β-defensin CBD103* (*K*-locus) genes, and the hypervariable part of the mtDNA control-region (mtDNA CR1).

Locus	Chromosome	STR repeat size	Allele sizes (bp)	Dye label	Multiplex	Reference	Genomic context[a]	Gene location / Nearest gene[a]	Source[a]
AHTk211	CFA26	Dinucleotide	79-101	FAM	MF	[1]	Inter-genic	*LHX5*	HO
CXX279	CFA22	Dinucleotide	109-133	FAM	MF	[2]	Inter-genic	*CLYBL*	HO
REN169O18	CFA29	Dinucleotide	150-170	FAM	MF	[3]	Inter-genic	*CAP1*	HO
INU055	CFA10	Dinucleotide	190-216	FAM	MF	Finnzymes	Inter-genic	*ETAA1*	HO
REN54P11	CFA18	Dinucleotide	222-244	FAM	MF	[3]	Intronic	*MAGI2*	HO
AHT137	CFA11	Dinucleotide	126-156	HEX	MF	[4]	Inter-genic	*UBTD2*	HO
REN169D01	CFA14	Dinucleotide	199-221	HEX	MF	[3]	Inter-genic	*ABCB5*	HO
AHTh260	CFA16	Dinucleotide	230-254	HEX	MF	[5]	NA	NA	NA
AHTk253	CFA23	Dinucleotide	277-297	HEX	MF	[1]	Inter-genic	*RPS2P32*	HO
INU005	CFA33	Dinucleotide	102-136	NED	MF	Finnzymes	Intronic	*PARP9*	HO
INU030	CFA12	Dinucleotide	139-157	NED	MF	Finnzymes	Intronic	*GRIK2*	HO
FH2848	CFA2	Dinucleotide	222-244	NED	MF	[5]	Inter-genic	*LINC00710*	HO
REN162C04	CFA7	Dinucleotide	192-212	PET	MF	[3]	Intronic	*DNM3*	HO
AHTh171	CFA6	Dinucleotide	215-239	PET	MF	[5]	Intronic	*SRM3*	HO
REN247M23	CFA15	Dinucleotide	258-282	PET	MF	[3]	Inter-genic	*LIN7A*	HO
FH2004	CFA11	Tetranucleotide	104-202	PET	M1	[6]	Intronic	*PTPRD*	RS, HO
FH2088	CFA15	Dinucleotide	91-139	FAM	M1	[6]	Inter-genic	*FHDC1*	HO
FH2096	CFA11	Tetranucleotide	86-110	HEX	M1	[6]	Inter-genic	*DDX43*	HO
FH2137	CFA3	Dinucleotide	140-192	HEX	M1	[6]	Inter-genic	*CHD1*	HO
CPH2	CFA32	Dinucleotide	88-106	NED	M1	[7]	Intronic	*SCD5*	RS, HO
CPH8	CFA13	Dinucleotide	191-219	FAM	M1	[7]	Inter-genic	*LOC1720*	HO

FH2079	CFA24	Tetranucleotide	246-282	FAM	M2	[6]	Inter-genic	*PTNP1*	HO
CPH4	CFA15	Dinucleotide	130-155	NED	M2	[7]	Intronic	*ANKS1B*	HO
CPH5	CFA15	Dinucleotide	102-124	HEX	M2	[7]	Inter-genic	*ATAD2B*	HO
CPH12	CFA8	Dinucleotide	188-214	FAM	M2	[7]	Intronic	*NPAS3*	RG, HO
C09.250	CFA9	Dinucleotide	121-145	PET	M2	[2]	Intronic	*NXN*	HO
C20.253	CFA20	Dinucleotide	90-120	NED	M2	[2]	Intronic	*FHIT*	RS, HO
AHT132	CFA2	Dinucleotide	160-172	PET	M3	N. Holmes	Inter-genic	*CALML3*	HO
C27.442	CFA27	Dinucleotide	158-172	HEX	M3	[2]	Intronic	*SLC11A2*	HO
FH2010	CFA24	Tetranucleotide	216-240	NED	M3	[6]	Intronic	*PLK1S1*	RG, HO
PEZ1	CFA7	Tetranucleotide	99-131	HEX	M3	[8]	NA	NA	NA
PEZ5	CFA12	Tetranucleotide	95-119	PET	M3	[8]	Inter-genic	*FBXL4*	HO
AHT103	CFA4	Dinucleotide	71-89	HEX	M4	[4]	Inter-genic	*RANBP3L*	HO
AHT111	CFA2	Dinucleotide	72-92	NED	M4	[4]	Intronic	*IL22RA1*	HO
FH2001	CFA23	Tetranucleotide	123-155	PET	M4	[6]	Inter-genic	*ARHGEF26*	HO
C09.173	CFA9	Dinucleotide	100-118	FAM	M4	[2]	Intronic/Exonic	*ABCA5*	RS, HO
C13.758	CFA13	Dinucleotide	220-244	NED	M4	[9]	Inter-genic	*FER1L6*	HO
CPH9	CFA28	Dinucleotide	139-151	HEX	M4	[7]	Inter-genic	*ZFHX4-AS1*	HO
CPH14	CFA5	Dinucleotide	185-205	PET	M4	[7]	NA	NA	NA
MSY34A	CFAY	Dinucleotide	160-190	NED	M5	[10]	NA	NA	NA
MSY41A	CFAY	Dinucleotide	90-150	HEX	M5	[10]	NA	NA	NA
MSY34B	CFAY	Dinucleotide	167-177	HEX	M5	[10]	NA	NA	NA
MSY41B	CFAY	Dinucleotide	109-137	NED	M5	[10]	NA	NA	NA
Amelogenin	CFAX	-	174-218	NED	MF	Finnzymes	Intronic	*AMELX*	RS,HO
K-locus	CFA16	Codon deletion	147-151	HEX	M2	[11, 12]	Esonic/Intronic	*CBD103*	RG,RS,HO
mtDNA CR1	mtDNA	-	350	-	-	[13]	Control-region	*CYTOCHROME b*	RS

[a] Genomic context, gene location, source. Primer locations along the reference dog genome (CanFam3.1 assembly) have been identified either via the UniSTS database at NCBI (http://www.ncbi.nlm.nih.gov/genome/sts/), or by in-silico PCR via the USCS genome browser (http://genome.ucsc.edu/). The UCSC genome browser was also used to identify the genomic context of each STR, which were classified as inter-genic, intronic or exonic, in relation to the annotated dog genes available from RefGene (RG), RefSeq (RS), or to the presence of human orthologous transcripts (HO). NA = locus not mapped in the reference dog chromosomes (CanFam3.1 assembly).

References

1. Thomas R, Holmes N, Fischer P, Dickens H, Breen M, et al. (1997) Eight canine microsatellites. Animal Genetics 28: 153-154.
2. Ostrander EA, Sprague GF, Rine J (1993) Identification and characterization of dinucleotide repeat (CA) markers for genetic mapping in dog. Genomics 16: 207-213.
3. Guyon R, Lorentzen TD, Hitte C, Kim L, Cadieu E, et al. (2003) A 1-Mb resolution radiation hybrid map of the canine genome. Proceedings of the National Academy of Sciences 100: 5296-5301.
4. Holmes N, Humphreys S, Binns M, Holliman A, Curtis R, et al. (1993) Isolation and characterization of microsatellites from the canine genome. Animal Genetics 24: 289-292.
5. Breen M, Jouquand S, Renier C, Mellersh CS, Hitte C, et al. (2001) Chromosome-specific single-locus FISH probes allow anchorage of an 1800-marker integrated radiation-hybrid/linkage map of the domestic dog genome to all chromosomes. Genome Research 11: 1784-1795.
6. Francisco L, Langsten A, Mellersh C, Neal C, Ostrander E (1996) A class of highly polymorphic tetranucleotide repeats for canine genetic mapping. Mammalian Genome 7: 359-362.
7. Fredholm M, Winterø A (1995) Variation of short tandem repeats within and between species belonging to the Canidae family. Mammalian Genome 6: 11-18.
8. Neff MW, Broman KW, Mellersh CS, Ray K, Acland GM, et al. (1999) A second-generation genetic linkage map of the domestic dog, *Canis familiaris*. Genetics 151: 803-820.
9. Mellersh CS, Langston AA, Acland GM, Fleming MA, Ray K, et al. (1997) A linkage map of the canine genome. Genomics 46: 326-336.

10. Sundqvist AK, Ellegren H, Olivier M, Vila C (2001) Y chromosome haplotyping in Scandinavian wolves (*Canis lupus*) based on microsatellite markers. Molecular Ecology 10: 1959-1966.
11. Candille SI, Kaelin CB. Cattanach BM, Yu B, Thompson DA, Nix MA, Kerns J, Schmutz SM, Millhauser GL, and Barsh GS (2007) A *β-defensin* mutation causes black coat color in domestic dogs. Science 318: 1418-1423.
12. Caniglia R, Fabbri E, Greco C, Galaverni M, Manghi L, Boitani L, Sforzi A, Randi E(2013) Black coats in an admixed wolf × dog pack: is melanism an indicator of hybridization in wolves? European Journal of Wildlife Research 59: 543-555.
13. Randi E, Lucchini V, Christensen MF, Mucci N, Funk SM, et al. (2000) Mitochondrial DNA variability in Italian and East European wolves: Detecting the consequences of small population size and hybridization. Conservation Biology 14: 464-473.

Table S2-add 1. Values of the average proportions of membership of dogs (DIT, DAP and DCZ), wolves (WIT, WCZ, WHR), Czechoslovakian wolfdogs (WDCZ) and putative hybrids (HYIT) from Italy in K = 4 clusters computed with STRUCTURE (39 autosomal STRs, admixture and I models, popflag = 0).

Group	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
DIT	0.940	0.010	0.017	0.033
DAP	0.990	0.001	0.002	0.007
DCZ	0.337	0.001	0.001	0.661
WIT	0.001	0.997	0.001	0.001
WCZ	0.016	0.004	0.914	0.066
WHR	0.003	0.004	0.989	0.003
WDCZ	0.002	0.001	0.002	0.994
HYIT	0.057	0.906	0.007	0.029

Table S3-add 1. Admixture analyses in dogs (DIT, DAP and DCZ), wolves (WIT) and putative hybrids (HYIT) from Italy. Values of the average proportions of membership of each sampled group in K = 4 clusters computed with STRUCTURE.

	STRUCTURE *Popinfo* inactive[a]				STRUCTURE *Popinfo* active[b]			
Group	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
DIT	0.632	0.201	0.161	0.006	0.863	0.044	0.093	0.000
DAP	0.009	0.912	0.077	0.002	0.002	0.970	0.028	0.000
DCZ	0.002	0.005	0.992	0.001	0.000	0.000	1.000	0.000
WIT	0.001	0.001	0.001	0.997	0.000	0.000	0.000	1.000
HYIT	0.026	0.030	0.041	0.903	0.036	0.031	0.038	0.891

[a]STRUCTURE analyses performed using 39 STRs, *admixture* and *I* models, *popflag* = 0 (*Popinfo* inactive).
[b]STRUCTURE analyses performed using 39 STRs, *admixture* and *I* models, *popflag* = 1 (*Popinfo* active).

Text S1-add 1. Description of laboratory methods with details on primers and PCR profiles for all the genotyped markers

The microsatellites in the Finnzymes Canine multiplex kit (Finnzymes, Thermo Scientific Canine Genotypes™) and the *Amelogenin* locus were amplified in a single multiplex PCR reaction (MF) using an Applied Biosystems Thermal Cycler (ABI GeneAmp® PCR System9700) with the following thermal profile:98°C/3 min, 98°C/15 sec, 60°C/90 sec, 72°C/30 sec (30-40 cycles), followed by a final extension step at 72°C for 5 min. The amplifications were carried out in a 20 µl total PCR volume, including 2 µl of DNA solution from saliva samples, or 1 µl of DNA solution from muscle and blood samples, corresponding to *c.* 20 – 40 ng of DNA, 10 µl of Finnzymes Canine Genotypes™ Panel 1.1 Master Mix (which included an optimized buffer containing MgCl2, deoxynucleoside triphosphates (dNTP) and Phusion™ Hot Start DNA Polymerase with an activity of 0.05 U/µl), and 10 µl of Finnzymes Canine Genotypes™ Panel 1.1 Primer Mix (including forward and reverse primers for the 19 markers).

The autosomal and Y-linked STR loci were amplified in other 5 multiplexed primer mixes (M1, M2, M3, M4, M5) using the Qiagen Multiplex PCR Kit (Qiagen Inc, Hilden, Germany), an ABI GeneAmp® PCR System9700,and the following thermal profile:94°C/15 min, 94°C/30 sec, 57°C/90 sec, 72°C/60 sec (30 cycles), followed by a final extension step at 72°C for 5 min. Amplifications were carried out in 10 µl total volume, including 2 µl of DNA solution from saliva samples, or 1 µl of DNA solution from muscle and blood samples, 5 µl Qiagen Multiplex PCR mix, 1 µl

QiagenQ solution, 0.4 µM deoxynucleotide triphosphates (dNTP), from 0.1 µl to 0.4 µl of 10 µMprimer mix (forward and reverse) and RNase-free water up to the final volume.

The 3-bp deletion (named K^Bor $CBD103^{\Delta G23}$) at the *β-defensin CBD103* gene (the *K*-locus) was genotyped following Caniglia et al. [1],in 10 µl PCR volumes including 1 µl or 2 µl of DNA solution and 0.3 pmol of the primers CBD103_ΔG23F (TCCGGCACGTTCTGTTTT, 6-FAM) and CBD103_ΔG23R (TTCGGCCAGTGGAAGAAC) that amplify a fragment of 190/193 bp. PCR conditions were: 94°C/2 min, 94°C/15 sec, 55°C/15 sec, 72°C/30 sec (40 cycles), plus one final cycle at 72°C for10 min.

The mtDNA control-region was amplified in 10 µl PCR volumes, including 1 µl or 2 µl of DNA solution, 0.3 pmol of the primers L-Pro and H350 (Randi et al. [2]), using the following thermal profile: 94°C/2 min, 94°C/15 sec, 55°C/15 sec, 72°C/30 sec (40 cycles), followed by a final extension at 72°C for 5 min. PCR products were purified using the exonuclease/shrimp alkaline phosphatase procedure (Exo-Sap; Amersham) and sequenced in both directions using the ABI Big Dye Terminator kit with the following steps: 96°C/10 sec, 55°C/5 sec, 60°C/4 min of final extension (25 cycles).

References

1. Caniglia R, Fabbri E, Greco C, Galaverni M, Manghi L, Boitani L, Sforzi A, Randi E(2013) Black coats in an admixed wolf × dog pack: is melanism an indicator of hybridization in wolves? European Journal of Wildlife Research 59: 543-555.
2. Randi E, Lucchini V, Christensen MF, Mucci N, Funk SM, et al. (2000) Mitochondrial DNA variability in Italian and East European wolves: Detecting the consequences of small population size and hybridization. Conservation Biology 14: 464-473.

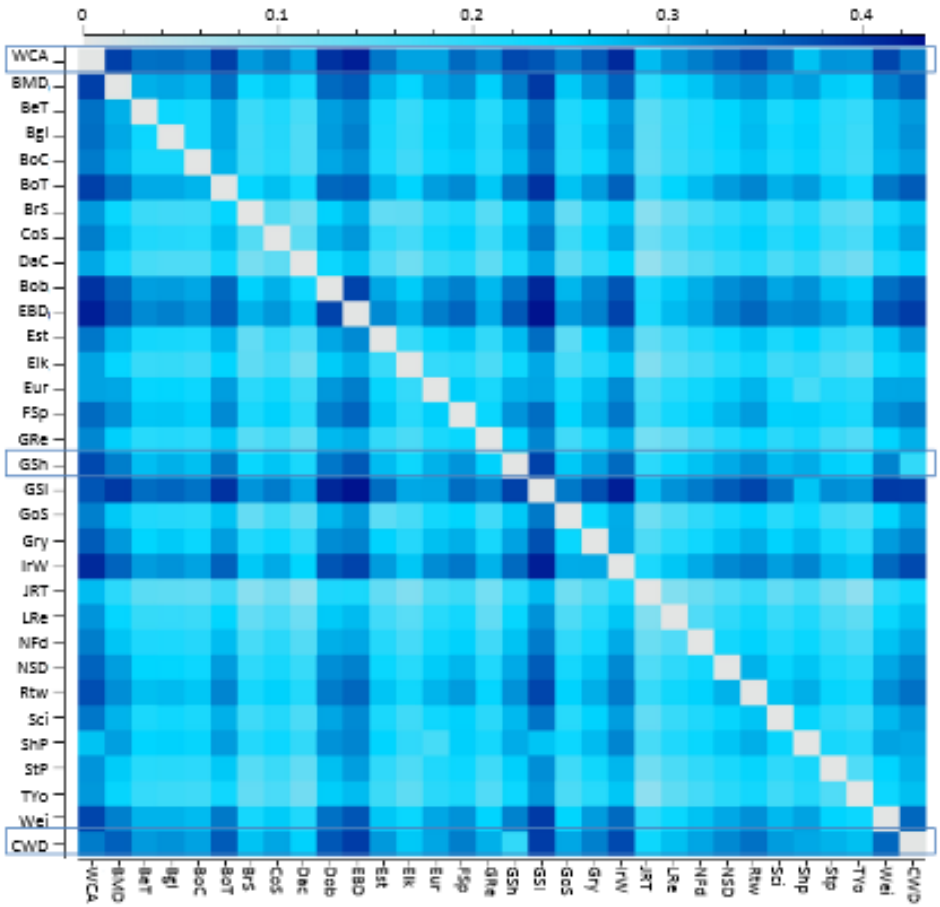


Figure S1-add 3. FST heat plot matrix of the genetic distances among groups computed from the 126k dataset in SVS. The most distant breed to Carpathian wolves (WCA) is the English Bulldog (EBD) while the closest one is the ancient breed Shar-Pei (ShP). As expected the least differentiated breed from the Czechoslovakian Wolfdog (CWD) is the German Shepherd (GSh).

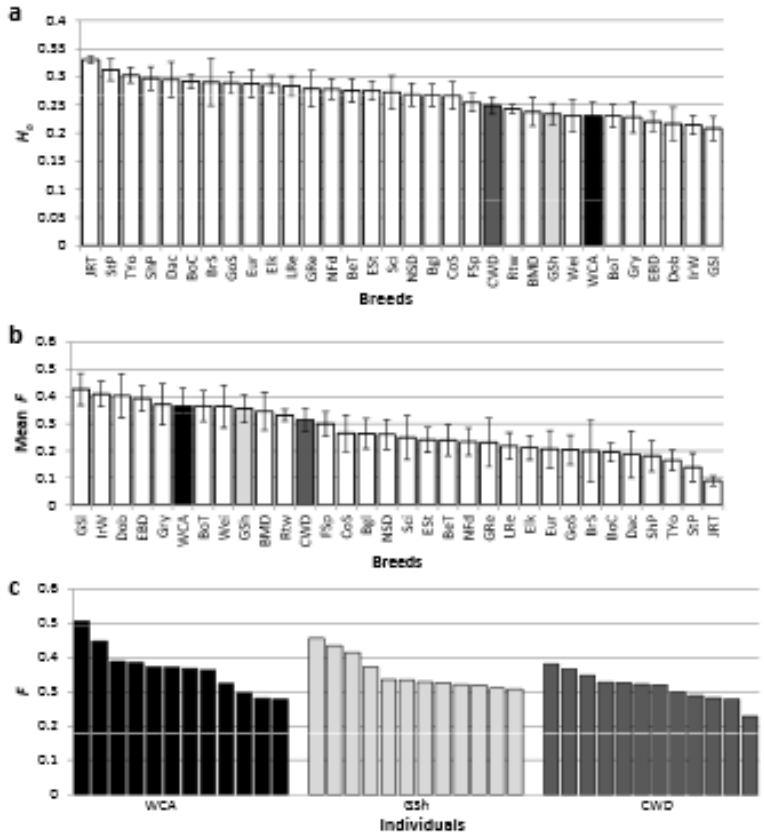


Figure S2-add 3: Genetic variability indexes computed in SVS using the 126k SNP dataset. a Mean values of observed heterozygosity (H_o) within groups. Czechoslovakian Wolfdogs (in dark gray) show higher levels of heterozygosity than parental populations (Carpathian wolves in black and German Shepherds in light grey), as expected from the recent crossings that originated the breed, but lower than most breeds. Bars indicate standard deviations. b Plots of the mean inbreeding coefficient *F* per breed. Czechoslovakian Wolfdogs show a mean *F* value intermediate among the other breeds but lower than both parental populations. c: from left to right: individual *F* values for Carpathian wolf (black histograms), German Shepherd (light grey histograms) and Czechoslovakian Wolfdog (dark gray histograms) groups. Bars indicate standard deviations.

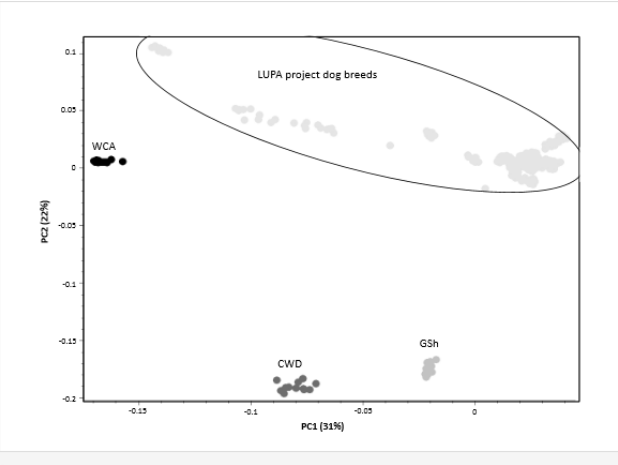


Figure S3-add 3. PC1 vs. PC2 results from an exploratory principal component analysis (PCA) computed in SVS on the 126k SNP dataset and including dogs from 30 pure breeds (extrapolated from the available LUPA project dataset; top side of the graph, in grey inside the circle), Carpathian wolves (WCA; black dots to the left), German Shepherds (GSh; light grey dots in the bottom), and Czechoslovakian Wolfdogs (CWD; dark gray dots in the bottom). The two axes are not to scale, in order to better distinguish individuals along PC2.

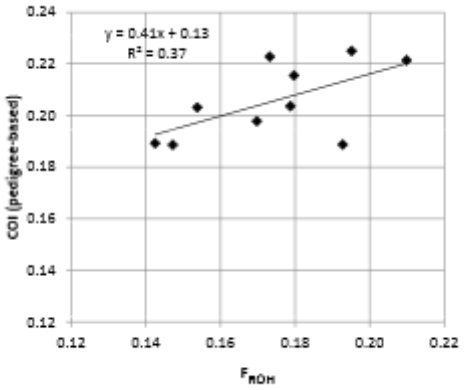


Figure S4-add 3. Comparison between the individual frequency of ROHs (F_{ROH}), calculated in SVS as the proportion of ROHs on the genome length spanned by the analysed SNPs (on the horizontal axis), and the individual Wright's inbreeding coefficient (COI), estimated from the pedigrees with the software U-WGI (on the vertical axis). The two inbreeding indexes are significantly ($p < 0.01$) correlated.

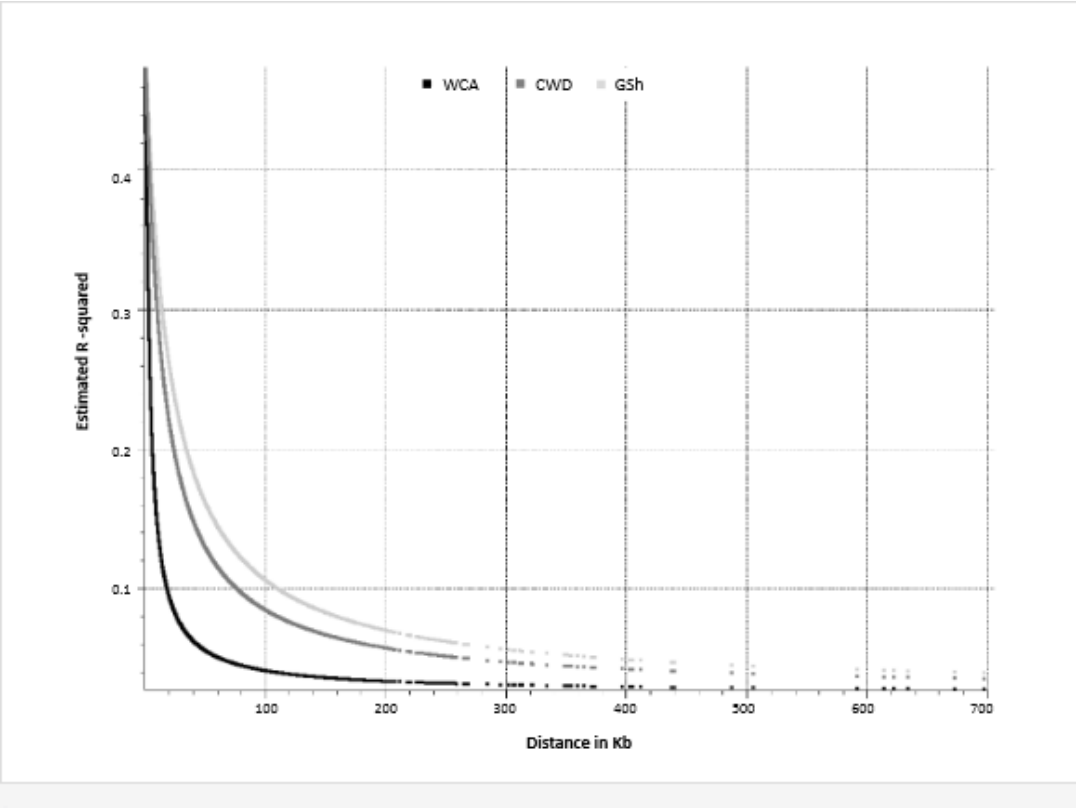
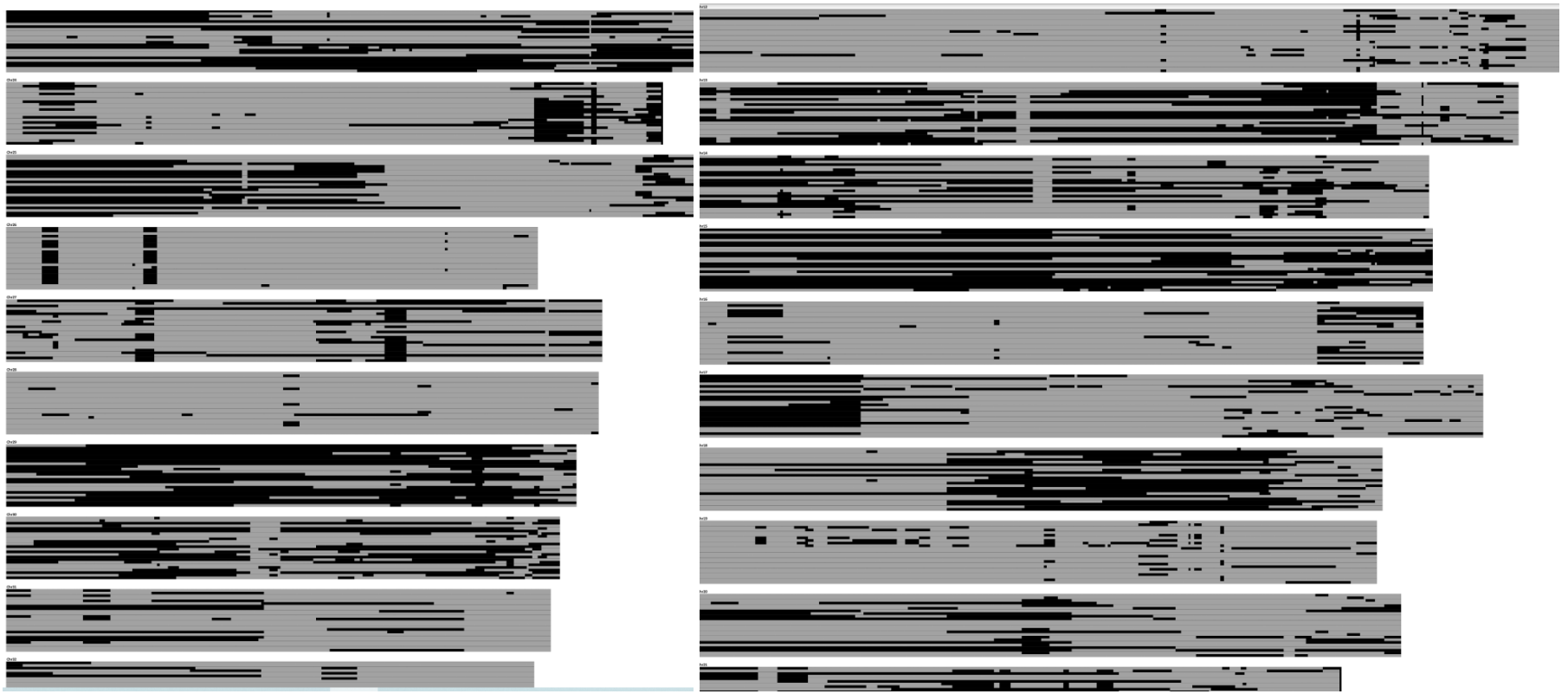


Figure S5-add 3. Linkage disequilibrium (LD) decay plot. The vertical axis indicates the mean Estimated R-squared (r^2), and the horizontal axis indicates the distance in kb at which LD decays.



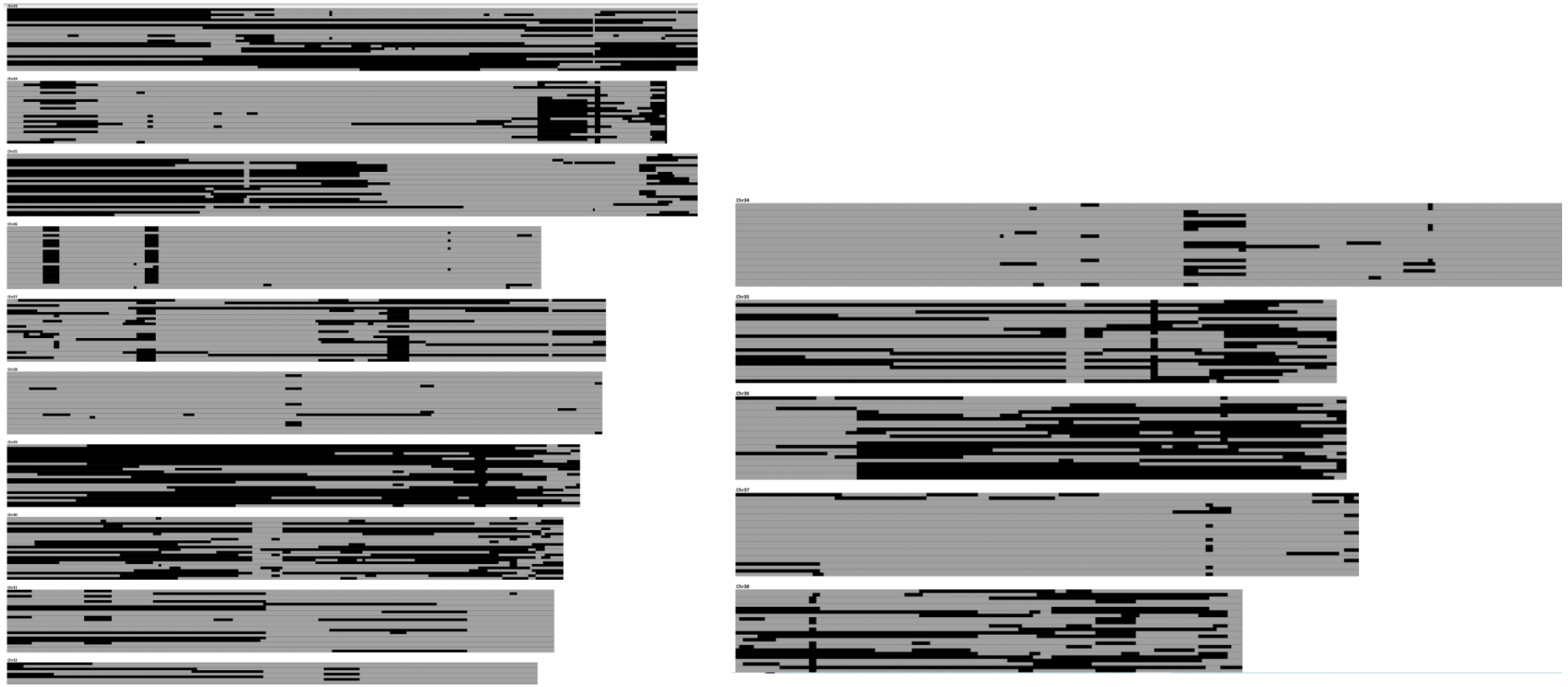


Figure S6-add 3. Graphical representation, for each chromosome of each analysed Czechoslovakian Wolfdog, of the ancestry components identified by PCAdmix based on the analysis of 10-SNP haplotype blocks. Each horizontal bar represents the two homologous chromosomes of an individual showing in black the genomic regions assigned as wolf-like and in light grey those assigned as dog-like.

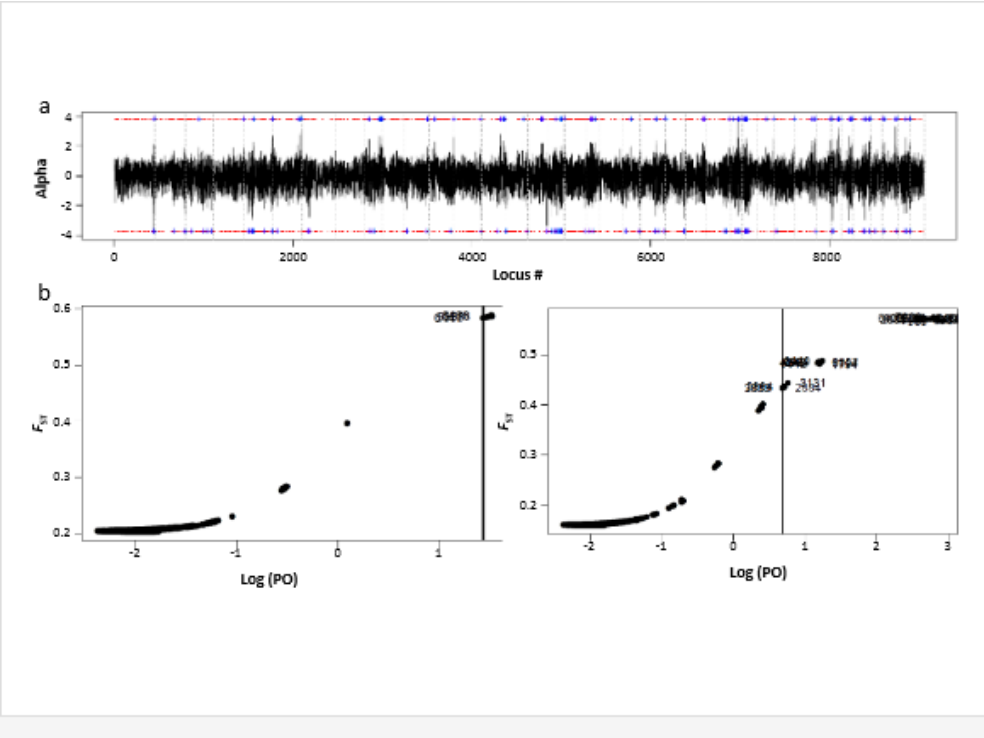


Figure S7-add 3.

a BGC *alpha* parameter outlier SNPs. Values lower than 0 indicate excess of wolf alleles, values higher than 0 indicate excess of dog alleles. BGC significant outliers are indicated by blue crosses (top or bottom 1% of the empirical distribution of values) and by red dots (95% credibility intervals of 10,000 iterations not including 0).

b BayeScan outlier SNPs detected comparing differences in allele frequency between Czechoslovakian Wolfdogs and German Shepherds (right) and between Czechoslovakian Wolfdogs and Carpathian wolves (left). The vertical axis indicates mean F_{ST} values between populations, and the horizontal axis indicates the logarithm of posterior odds (log(PO)). The vertical line indicates the log(PO) value corresponding to the false discovery rate threshold of 0.05.