Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Biodiversità ed Evoluzione

Ciclo XXVIII

Settore Concorsuale di afferenza: 05/B1 – Zoologia e Antropologia

Settore Scientifico disciplinare: BIO/05 - Zoologia

From Sanger to NGS: detecting MHC (Major Histocompatibility Complex) class II and OR (Olfactory Receptors) genetic variability in Italian wolves (*Canis lupus*) and relative canids

Presentata da: **Silvana Lapalombella**

Coordinatore Dottorato
Prof.ssa Barbara Mantovani

Relatore
Prof. Ettore Randi

Correlatore
Dott. Marco Galaverni

Esame finale anno 2016

PREFACE

Canids are the most known and fascinating family of the Placental mammal order of Carnivora. In particular the wolves (*Canis lupus*) always draw human attention, from folklore to scientific sciences, primarily because of their large diffusion in the northern hemisphere, and secondarily for their fascinating social system and complex evolutionary history. Wolves, indeed, are the ancestors of domesticated dogs, which evolved in deep association with humans.

From a scientific point of view biological and genetic studies on Canids have revolutionised the understanding of evolutionary biology and behavioural ecology. These animals possess an extraordinary ability to adapt and occupy a wide variety of ecosystems in every continent except Antarctica. However, their diffusion is extremely limited compared to the past, since they went through very dark periods of persecutions.

In Italy the wolf is the second largest predator after the brown bear (*Ursus arctos*), and is considered one of the most iconic species that represents Italian biodiversity. Even if wolves have been almost completely eradicated during the past century, they always have been part of human cultural heritage. Scientific community, local institutions and the European Commission started to support the protection of endangered large carnivore species and their habitats with practical conservation, restoration and management actions since almost 30 years. Although numbers of wild canids (wolves and jackals – *Canis aureus*) are actually on the rise in Europe, they still suffer habitat fragmentation caused by growing human depletion of natural habitats in favour of densely inhabited landscapes. Their trans boundary nature is in contrast with human activities rising long-standing conflicts with local communities, which often lead to events of direct persecution., We are far from guaranteeing a stable survival of these canids. As top predators they usually live at small densities, thus they are more prone to become endangered than more abundant species. In some countries, wild canids are still considered endangered or threatened and actions of survey, protection and overall management of these predators should continue, since they play important roles to maintain ecosystem functionality intact.

Thereafter, the conservation of jackal and wolf populations poses particular challenges since the management is further complicated by hybridization events with domestic dogs which poses scientific and ethical problems that still have not found a general solution.

Ecological, political, and economic approaches might be of primary importance, but also genetics and related considerations are crucial for the long-term persistence of these species. Thanks to the fast evolving field of molecular techniques, population genetic and genomic studies became a

useful and irreplaceable tool in conservation biology. In particular, these two topics – the genetics and conservation of wild canids – are the main subject of this thesis.

In this PhD thesis I will describe different aspects of conservation genetics and genomics of two wild *Canidae* species, the wolf (*Canis lupus*) and the golden jackal (*Canis aureus*), through the study of two of the most variable gene families: the Major Histocompatibility Complex genes (MHC), and Olfactory Receptors genes (OR). In order to perform these studies both Sanger and next generation sequencing (NGS) DNA techniques have been used. The background of the thesis is described in the "General introduction" with phylogeny, classification and evolutionary ecology of the *Canidae*, with a focus on the species *Canis lupus* and its main conservation concerns in Italy. Moreover, I will introduce the importance to perform genetic studies as tools for wild-life conservation and management, with a description of the framework of the principal historical and currently used molecular markers that had driven to develop MHC and OR sequencing projects.

The thesis is divided into two parts, "PART I – The MHC typing project" and "PART II – The OR genes typing project". A total of four scientific papers (already published or under revision) will be introduced and illustrated as result of three years of PhD activities at ISPRA's (Istituto Superiore per la Protezione e la Ricerca Ambientale), Laboratory of conservation genetics, in Ozzano dell'Emilia, and thankfully to a PhD fellowship granted by the Università di Bologna.

TABLES OF CONTENTS

GENERAL INTRODUCTION

PART I – THE MHC TYPING PROJECT

GENERAL INTRODUCTION

Phylogeny of the *Canidae*

The *Canidae* include 34 closely related species that diverged within the last 10 million years (Wayne *et al.* 1997). However, the origin of this family spans 50 million years with a series of successive radiations that led *Canidae* members to occupy a broad range of ecological niches. A phylogeny from morphological and paleontological perspective has been reconstructed in a review on by Wang *et al.* (2004). The family of *Canidae* is formed by three archaic major groups (subfamilies): *Hesperocyoninae*, *Borophaginae* and *Caninae* (two of these are represented by fossil forms only; Figure 1). The *Hesperocyon* is the basal member of *Hesperocyoninae,* the most ancient group among canids, from which originated to the two subfamilies *Borophaginae* and *Caninae*.

The subfamily *Caninae* started with *Leptocyon*, a small fox-sized ancestral species, firstly appeared in the early Oligocene (Orellan, 34–32 million years before present [Ma BP]) and persisted through the late Miocene (Clarendonian, 12–9 Ma BP). From this long-lived genus, during Miocene, originated *Vulpes* and *Canis* species. In particular in North America in the latest Miocene (Hemphillian, 9–5 Ma BP), the extinctions of all small borophagines let foxes occupy more niches. Later during the Pliocene *Vulpes* species diffused also in Eurasia, from an immigration event that resulted independent from that of the *Canis* clade.

In the medial Miocene, likely in North America appeared a jackal-size canid the *Eucyon* taxon, which species have appeared in Europe in late Miocene and by the early Pliocene in Asia, suggesting a dispersal of this form across the Bering Strait.

A sister group of *Eucyon* is the South American clade. This was probably originated by incursion of the grey fox (*Urocyon cinereoargentus*) and the extinct *Canis dirus* from North America. South American canids include: the maned wolf (*Chrysocyon brachyurus*), the bush dog (*Speothos venaticus*), the crab-eating fox (*Cerdocyon thous*), the small-eared dog (*Atelocynus microtis*) and the South American foxes (*Lycalopex* spp.).

While morphological and molecular evidence generally agrees that living South American canids belong to a natural group of their own (Wang *et al.* 2004; Lindblad-Toh *et al*. 2005; Bardeleben *et al*. 2005; Perini *et al.* 2010), the situation is different for the evolution of the genus *Canis*. Indeed, *Canis* in Eurasia was characterized by an extensive radiation and range expansion in the late Pliocene and Pleistocene, which led to numerous closely related species in Europe, Africa and Asia. Canid fauna invaded mid-latitude North America with the appearance of the grey wolf in the late Pleistocene (about 100,000 years BP; Wang *et al*. 2004).

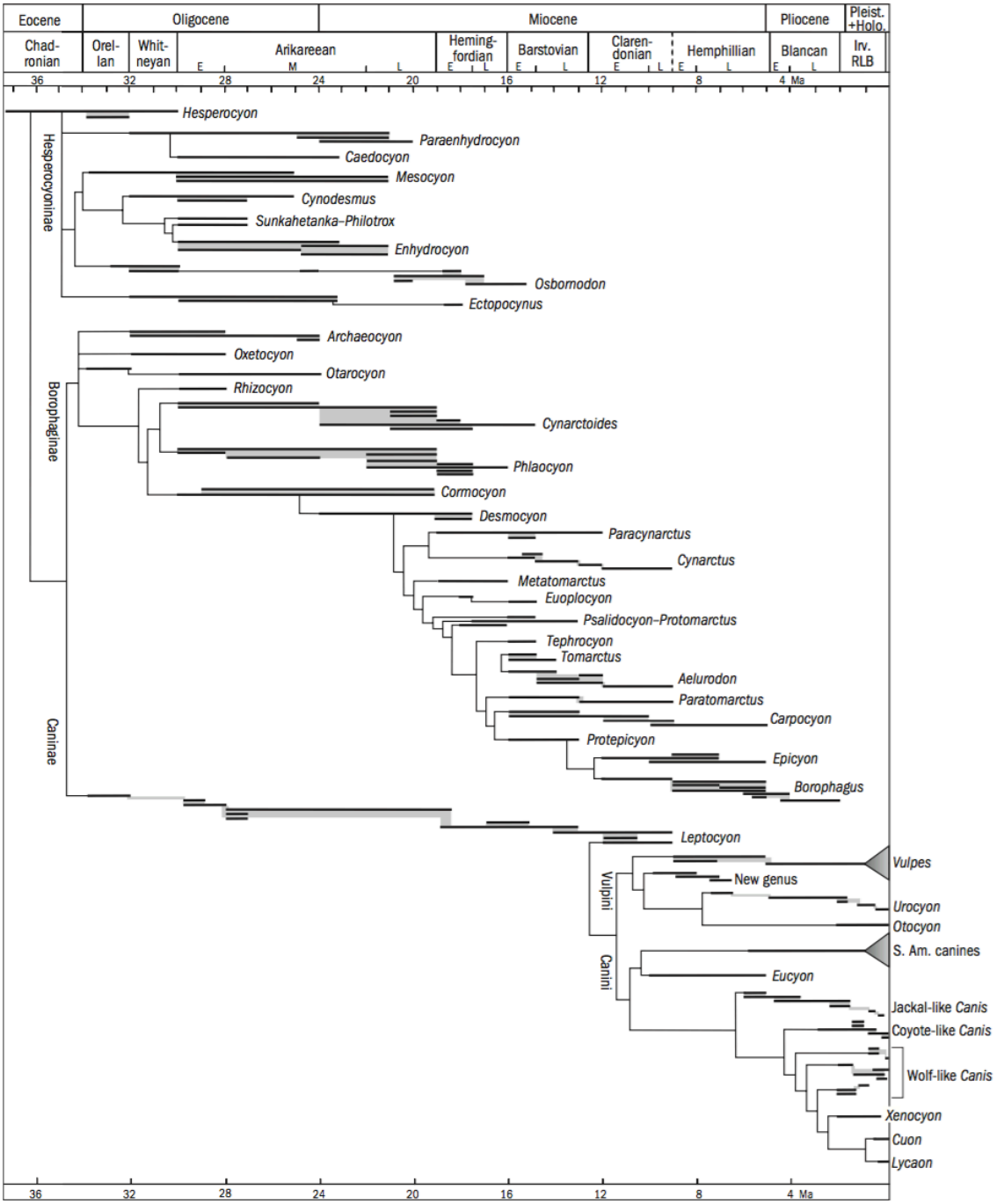


Figure 1

Simplified phylogenetic relationships of canids at the generic level from Xiaming *et al.* (2004). Species ranges are indicated by individual bars enclosed within grey rectangles, detailed relationships among species in a genus is not shown.

Canis species are not easily described by phylogenetic studies. Because of their Holarctic distribution, the presence of mixing events between New and Old world fauna, and weak patterns of intraspecific phenotypic differentiation, the resolution of the phylogenetic structure of such related taxa has always been challenging. Within the current *Canidae* three distinct phylogenetic groupings have been defined based on mtDNA (Wayne *et al.* 1987; 1997): (1) the fox-like canids, which include species closely related to the red fox (genus *Vulpes*); (2) the wolf-like canids including dog, wolf, coyote, Ethiopian wolf and three other species of jackals (genus *Canis*), as well as the African hunting dog (genus *Lycaon*) and the dhole (genus *Cuon*); (3) the South American canids including fox-sized canids.

The species - *Canis lupus* -

The grey wolf (*Canis lupus*) is a large carnivore belonging to the *Canidae* family. Studies based on mitochondrial DNA (mtDNA; Vila *et al.* 1999; Pilot *et al.* 2010) did not show clear worldwide phylogeographic structure in terms of monophyletic clades or large scale geographical structure, probably due to multiple expansions and contractions to refugia that wolf populations have experienced during the Ice Ages.

A complex evolutionary history was revealed by the presence of subpopulation structure related to different local environments (Carmichael *et al.* 2001; Geffen *et al.* 2004; Pilot *et al.* 2006) and then confirmed by the analysis of SNPs data (Figure 2) and linkage disequilibrium (LD) among wild and domestic canids (Gray *et al.* 2009; Pilot *et al.* 2014). Gray *et al.* (2009) showed that exists a clear genetic difference between Old World and New World populations (Figure 2 b), while investigating seven canids confirming also the genetic distance between wild species and domestic dogs, and the separation from foxes (Figure 2 a). When Old and New World populations were compared (Figure 2 c, d), populations with a known demographic decline or strongly inbred appeared to be separated from the others, according also to LD data. In general, LD extent is directly connected to the demographic history of a population, and low levels of LD have been observed in Alaskan, Canadian, and northern Quebec gray wolves, whose populations have remained large for a substantial period of time or have rapidly expanded, showing high variability and reduced population differentiation. Whereas the opposite situation have been observed among the more structured European populations.

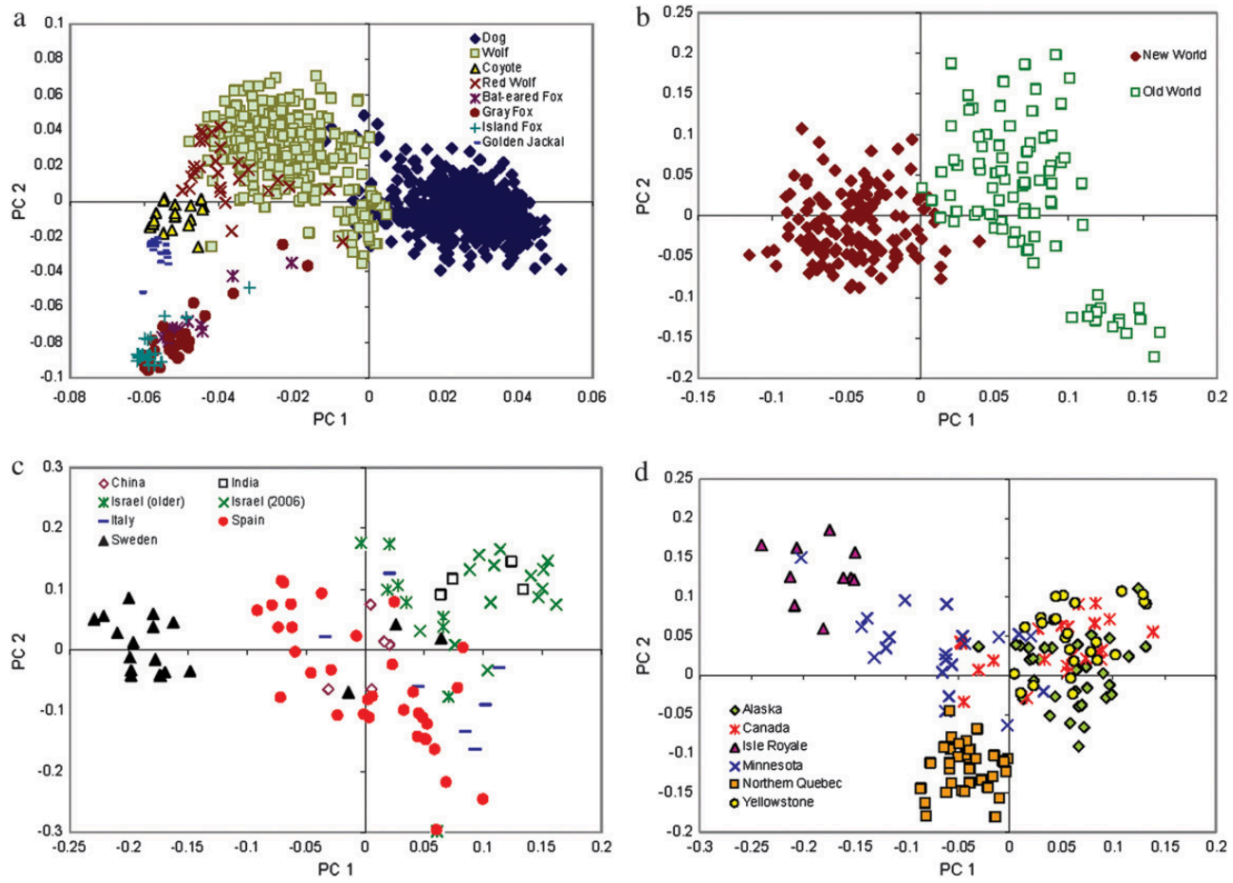


Figura 2

Principal component analysis of 106 SNP genotype data with and across species: (a) all species, (b) gray wolves, (c) Old World wolves, (d) New World wolves. From Gray *et al.* (2009)

In particular the Italian and Iberian were found to be the most differentiated wolf populations with high levels of LD likely due to ancient bottlenecks (vonHoldt *et al.* 2011; Pilot *et al.* 2014).

Several studies tried then to address the question of the of dog close relation with wolves, using different molecular markers such as nuclear loci (Bardeleben *et al.* 2005), intron and exon sequences (Lindblad-Toh *et al.* 2005) and genome-wide SNPs distribution (Vonholdt *et al.* 2010). The grey wolves and dogs are the most closely related taxa, and then the coyotes, golden jackals and Ethiopian wolves follow in close affiliation (Vonholdt *et al.* 2010). This phylogenetic proximity was recently further investigated to understand early evolutionary history and demographic events that shaped the genomes of wolves and dogs (Freedman *et al.* 2014). Two major bottlenecks were observed, one in dogs confirming previous findings of demographic declines caused by domestication (Lindblad-Toh *et al.* 2005; Gray *et al.* 2009), and another in wolves, that occurred soon after their divergence from dogs. This sharp decline in wolves was dated at ~8-25 kya, in coincidence with the first colonization of Eurasia by hunter-gatherers,

which likely could be the main cause of the bottleneck, since they were supposed to compete with wolves for prey (Fan *et al.* 2016). Furthermore, ancient traces of admixture between some eastern wolf lineages with other wild canids - Israeli wolf and golden jackal Chinese wolf and Dingo, Israeli wolf and Basenji - were found (Freedman *et al.* 2014; Koepfli *et al.* 2015). However, it is not excluded that these traces of admixture could came from ancestral wolves phylogenetically distinct from actual wolf lineages, and existing before the dog-wolf divergence happened approximately at 15 kya (Freedman *et al.* 2014; Fan *et al.* 2016). The genomic comparison of 24 wolves from Old and New world confirmed that the divergence between New and Old World wolves was the earliest branching event, then followed by the divergence of Old World wolves and dogs, suggesting that dogs were domesticated from an extinct Old world wolf population. A clear geographical structure within Old World wolves also emerged, with Middle Eastern wolves closer to European wolves, rather than to East Asian wolves, while the Tibetan wolf and Mexican wolf were identified as the most distinct populations in the Old World and New World, respectively (Figure 3; Fan *et al.* 2016).

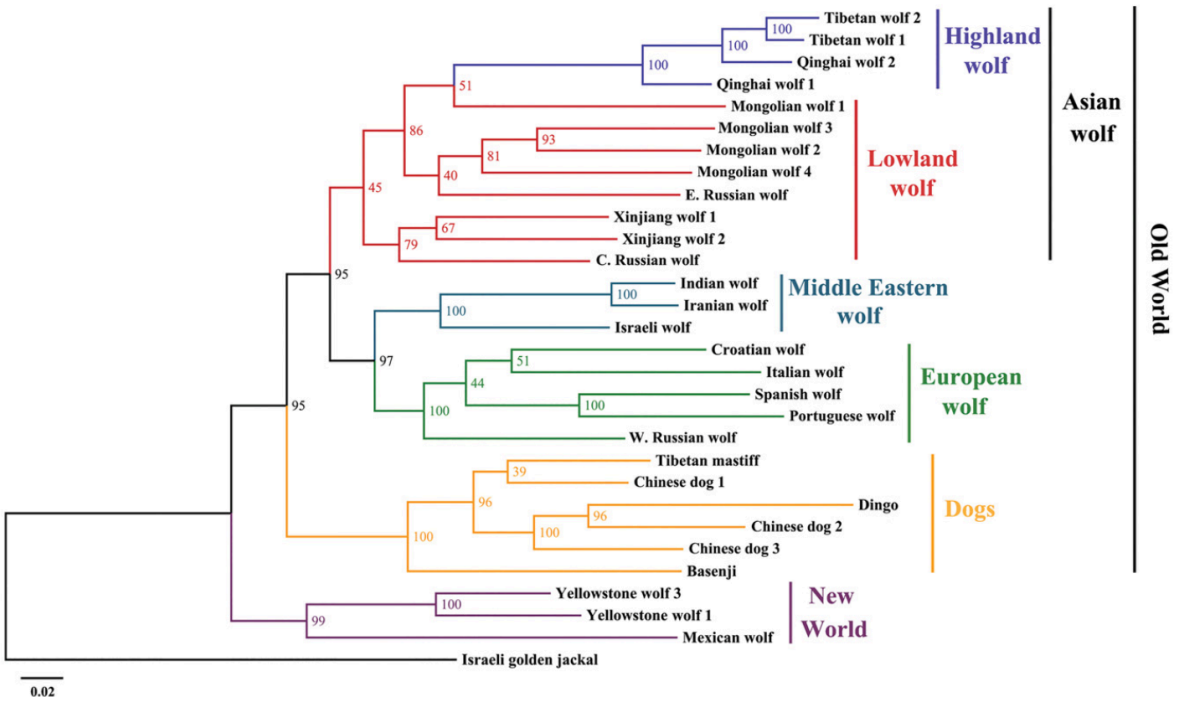


Figure 3

The maximum likelihood tree of genomic sequences from 30 wild canids. Numbers represent node support inferred from 100 bootstrap repetitions. The reference genome of the boxer was not included. The Israeli jackal is the outgroup. From Fan *et al.* (2016).

Wolves' phenotypes are variable in terms of size, colour and weight (Mech 1970). The height can vary from 0.6 to 0.95 meters at the shoulder and weight ranges from 20 to 62 kilograms. The most

remarkable dimensions can be found at high latitudes. The Italian wolves are smaller on average (25-35 kg; Ciucci and Boitani 1998), and present particular phenotypic characteristics such as a typical grey-brownish coat and a black stripe on the frontal part of the anterior legs. On the basis of these and other morphological characteristics Altobello (1921) proposed that Italian wolf should be recognized as the subspecies *Canis lupus italicus*. In the last decade taxonomic methods based both on morphometric studies (Nowak & Federoff 2002) and genetic analyses (Randi *et al.* 2000; Randi and Lucchini 2002; Lucchini et al. 2004; Vonholdt et al. 2011; Pilot *et al.* 2014), have later suggested that the Italian wolves were enough differentiated from other European population to support Altobello's classification (Boitani *et al.* 2003 b). However the colour of the fur varies depending on the season and on the different ecological adaptations to natural habitats (Musiani *et al.* 2007). In particular black wolves have been observed only in North America and Italy (Anderson *et al.* 2009; Caniglia *et al.* 2013; see paragraph "Species distribution and main conservation concerns" and "Molecular markers in Italian wolf conservation genetics").

Wolves are social living animals. Dominant individuals were usually called 'alpha', and have the leadership of the pack in activities such as the hunt, and long movements across the territory. Dominance status has been associated with the breeding status, which means that when multiple potential breeders are present, only the leaders have the possibility to form couples (the alpha pair). However, most of natural packs contain only a single breeding couple, thus the concept of "alpha pair" has been questioned (Mech 1999). In fact, all young wolves are potential breeders that usually will disperse from the natal group to form a pack on their own (Mech 1999). A pack is fundamentally a breeding unit that originates when a pair establishes in a territory and reproduces. It is generally made up by a mating pair, its yearling pups and by some other adults which are generally the offspring of the previous years remaining with the pack for a year or more, when new pups are born (Mech 1970). Some external individuals (the 'adoptees') coming by dispersal events can join the pack (Boitani *et al.* 2003 a) and when a mating member lacks, it can be substituted by another wolf of the same pack or by an adoptee wolf (Meier *et al*. 1995; Caniglia *et al*. 2014). The breeding pair share pack-leadership responsibilities (Mech 1999; Peterson *et al.* 2002). However different patterns of cooperation have been observed during travels, thus although dominant breeding wolves provided most leadership, sometime also subordinate wolves can help in leading, likely to reduce the energy expenditure of dominant individuals (Peterson *et al.* 2002).

The howling is the most widely known method of communication among wolves. Usually these animals tend to howl on the periphery of their territory to advise their presence at long distances and discourage foreign individuals to enter pack territory (Mech 1970). Howls become more frequent during the breeding season. Females enter in oestrus once a year following

hormones concentration induced by photoperiod and latitude (Kreeger 2003; Mech 2002). The mating period, indeed, can occur from January to April, in Italy generally in March, while northern populations tend to cycle later.

Gestation lasts for 56-68 days (Mech 1970) producing in average 4-6 pups, with some exceptions till reach 11 pups (Mech 1974). Dens are in holes, caves, pits, hollow logs, where pups reside for two months, during the lactation. When the sexual maturity is reached at 22–46 months of age and occasionally 10 months (Mech 1970, 1974), individuals leave the packs and undergo short or on long-distance dispersal, even across suboptimal habitats (217 km in Italy; Ciucci *et al*. 2009). Using this strategy wolves can prevents inbreeding and promote natural selection and cross-breeding.

The pack-territory size can vary according to the presence of neighbouring packs, to prey density and to geographical and morphological features of the landscape, but also to human disturbance (Boitani 2000). Preys are usually constituted by wild ungulates but events of depredations on livestock can occur (Milanesi *et al.* 2015). Nevertheless, wolves are opportunistic feeders and can hunt small mammals or eat vegetation or dead animals. The territory of the pack can ranges from 80 to 2.500 km^2 in North America, from 100 to 500 km^2 in Europe (Boitani 2000), and from 20 to 300 km^2 in Italy (Apollonio 2004; Ciucci and Boitani 1998; Caniglia *et al.* 2014).

In addition to vocalizations, wolves communicate through complex body language, which comprehends body carriage, postures of tail and ears, and facial expressions. A large combination of these coded signals can convey to conspecific emotional states and intentions, which underline also the hierarchical relationships.

Another crucial form of communication is the olfaction. Odours are secreted by scent glands and can identify each single wolf, its health conditions, and its social and reproductive status (Mech 1999; Peterson *et al.* 2002). These glands are present all over the body, especially at the base of the tail or among toes (Harrington and Asa 2003). Breeding wolves scent-mark frequently, with both faeces and urine (Mech 1999; Peterson *et al.* 2002). Defecation are also often deposed along frequently used paths or at important crossroads (Barja *et al.* 2005), likely to ensure the pack orientation in the territory and preventing outsider pack members to traverse an occupied terrain. Moreover, these scent markers are often completed by ground-scratching, which is the main way to depose the scent of the inter-toes glands.

Species distribution and main conservation concerns

In origin wolf distribution covered the entire northern hemisphere (Boitani et al. 2003 a) occupying a variety of habitats from Arctic tundra to Arabian desert (Mech 1970). Spread throughout the entire Palearctic, wolf populations may have experienced times of historical isolation in the period after last glacial maximum (LGM 20,000-140,000 years ago). In Europe the Iberian, Italian and Balkan peninsulas have been identified as the three main glacial refuge areas during the Pleistocene (Randi 2003; vonHoldt *et al.* 2011). In particular, wolves in the Apennines which show genetic and morphological peculiarities could have been at least partially isolated from any other wolf population in Europe for some thousand years (Lucchini *et al.* 2004) and then survived at the LGM in refuge areas south of the Alps.

Back to more recent times the wolf could spread out in all Europe, but it has been always hunted and persecuted since Medieval times. In addition to active persecution wolf populations were condemned to a slow decline by human expansion, deforestation and decrease of natural prey. At the end of 18th century wolves were still present in all European countries with the exception of the British islands. In the last century, few surviving and fragmented European populations were subjected to an even stronger contraction.

In 1970 in Italy only few cores of the historical population survived in wild remote areas in the central and southern Apennines. Considering its rapid decline in the first decades of 20th century, Italian government agreed to recognize the wolf as a positive presence that should be protected in order to preserve the structure and function of natural ecosystems. Wolves, indeed, as a top predator lead the ecological cascades that contribute to maintain balance in natural prey and smaller predator densities, and also in the conservation of plant communities (Sergio *et al.* 2008).

The combination of protected areas and legal protection, with high dispersal rates and large individual territories (Ciucci *et al.* 2009) led wolves to recolonize most of their historical range in Italy from the Apennines to the Alps (Lucchini *et al.* 2002; Marucco *et al.* 2009), as well as in many other European countries (Chapron et al. 2014; Figure 4).

It was observed that despite the high potential rates of dispersal and gene flow, wolf populations may not mix even for long periods (Lucchini *et al.* 2004). However, due to rapid population expansion, for the first time after centuries, in Italy we might be able to observe the genetic effects of breeding events among individuals belonging to different European populations. In western Alps Italian wolf population is rapidly recovering its lost range (Lucchini *et al.* 2002; Fabbri *et al.* 2007) till reaching French boundaries, and in the east of Italy there is already a breeding pair formed by an Italian female and a male from Slovenia who abandoned his native

pack travelling through Austrian territories and then reaching Italy after more than 800km (Fabbri *et al.* 2013).

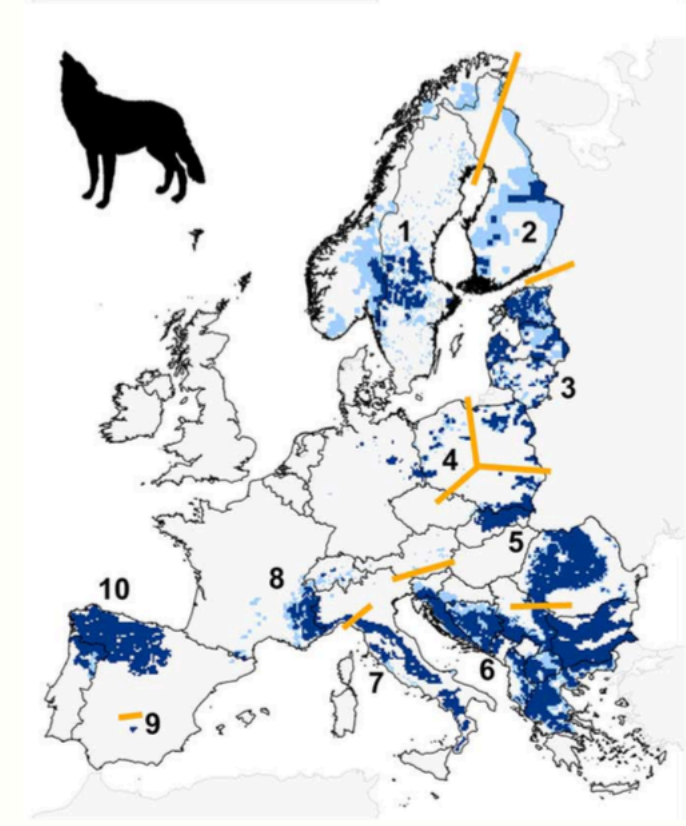


Figure 4

Distribution of wolf population in Europe in 2011 from (Chapron *et al.* 2014). Dark blue cells indicate areas of permanent occurrence, and light blue cells indicate areas of sporadic occurrence. Orange lines indicate boundaries between populations Numbers refer to populations: 1 Scandinavian, 2 Karelian, 3 Baltic, 4 Central European, 5 Carpathian, 6 Dinaric-Balkan, 7 Italian peninsula, 8 Alpine, 9 NW Iberian, 10 Sierra Morena.

Wolf presence in anthropized landscapes raises conflicts with human activities due to real or feared livestock depredation. Even though wolves are protected in most European countries, in some of which controlled hunting is also allowed, illegal or incidental killing is widespread, and wolf conservation remains problematic (Linnell *et al.* 2007).

Wolf is protected under international laws. The Bern Convention on Conservation of Wildlife and Natural Habitats (1979) includes wolves in appendix II (Strictly Protected Species), forbidding catching, killing, possessing and trading the species. Moreover, the Convention on International Trade of Endangered Species of Fauna and Flora (CITES, Washington 1973) strictly protects several wolf populations (the ones from Bhutan, India, Nepal and Pakistan), which can be included in two appendixes. In Appendix I species threatened with extinction, which are or may be affected by trade are included, while in Appendix II are listed the species that are not necessarily

threatened with extinction in the present, but that may become so, unless trade in specimens of such species is subject to strict regulation. Finally, there is the European Council Directive 92/43/EEC on the conservation of natural habitats and of wild fauna and flora (HABITAT).

Awareness campaigns since more than 40 years, and the introduction of refunds on depredation events (Boitani et al. 2010), did not change the diffused bad feeling of local populations towards this carnivore (Linnell *et al.* 2002), and poaching after depredations remains one of the major causes of wolf mortality.

Habitat fragmentation and human activities especially clear-cutting clearance of forest for agriculture, and development in rural areas, including roads, houses and tourist facilities still remain a source of vulnerability for wildlife. Roads in particular can both restrict the movements and the access to preys, promoting isolation and causing accidental killings.

To further worsen the situation, due to their close relationship, wolves and dogs can successfully mate and hybridize in captivity and in the wild. Hybridisation is occurring in several areas of the European range: in Italy (Randi & Lucchini 2002; Verardi *et al.* 2006a; Caniglia *et al.* 2014), Iberian Peninsula (Godinho *et al.* 2011), Scandinavia (Vilà *et al.* 2003), Germany (Andersen *et al.* 2015), Latvia (Andersone *et al.* 2002; Hindrikson *et al.* 2012), and Bulgaria (Vila 1997; Randi *et al.* 2000) creating a serious concern to the genetic integrity of the wolf populations. Hybridization may occur frequently when two genetically distinct populations co-occur in the same geographic area, and one population is fragmented and the other is overabundant. In Italy, despite the demographic recovery of wolf population, free-ranging dogs still largely outnumber wolves. Interactions with free ranging dogs may occur more often in the delicate phase of population's expansion or when wolves densities are low (Hindrikson *et al.* 2013). Few individuals, once left the pack, can meet and potentially mate with their domestic counter-part at the edges of their distribution (Godinho *et al.* 2011). The extent of hybrid diffusion is poorly known and the biological consequences of introgression are often controversial. Furthermore official guidelines to manage hybrids do not exist yet (Allendorf et al. 2001). Hybridization has the potential to produce morphological, physiological and behavioural changes in captive and wild-living canids (Mengel 1971; Thurber & Peterson 1991; Larivière & Crête 2007), and eventually lead to the origin of a new *taxon*, as it was seen for the red wolf (*Canis rufus*; vonHoldt et al. 2011). The presence of anomalous morphological characters such as black coat colour or dewclaws, has been observed in some wolf-like canids in Italy (Caniglia *et al.* 2013; Randi *et al.* 2014). Both these traits are supposed to have been introduced in the Italian wolf population via hybridization. The vestigial first toe on the hind legs are common in some dog breeds, but they have never been detected in wolves (Ciucci *et al.* 2003). The black coat is diffused in North America but it was suggested that

the introduction of this phenotype should be imputable to ancient hybridization events with domestic dogs (Anderson *et al.* 2009; Ollivier *et al.* 2013).

Conservation genetics

Detecting and interpreting the existing patterns of genetic variation within and between canid populations has long been one of the major aims of population genetic studies. In the last decades, the increasing availability of genetic data from worldwide sets of canid populations has offered new powerful tools for population geneticists to ask new questions about domestication processes, hybridization, and phylogenetic relationships. These studies are necessary in order to identify the effects of contemporary genetic structures on long-term periods and guarantee the survival of a species, particularly when it is endangered or threatened in human activities. Furthermore, those investigations could allow to identify cases of reduced effective population size, restricted gene flow, limited heterozygosity, but also inbreeding, past bottlenecks and hybridization or gene introgression, all factors that could seriously affect the population viability.

Although wolves in Italy are recovering from past bottleneck, the species is still considered endangered. Conservation and management of large carnivores are quite expensive and require ecological and genetic information on the species, which are difficult to collect for elusive predators like wolves, which cannot be directly counted especially in densely forested mountain habitats. Therefore, through traces such as scats, hairs and urine, non-invasive genetic sampling is used to estimate the abundance of wolves. The ISPRA genetic laboratory in Ozzano dell'Emilia collaborated with different regions in Italy for 12 years of non-invasive genetic monitoring (Caniglia *et al.* 2011, 2014; Fabbri *et al.* 2012) analysing ~ 20,000 of non invasive samples and ~ 2,000 invasive samples from found dead animals provided by Italian sanitary institutions, local monitoring projects and parks. Thanks to these collaborations, estimations of density (Galaverni et al. 2015; Figure 5), demographic fluctuations, gene flow, variability, migration, hybridization and introgression with domestic dogs of Italian wolf population were obtained (Randi *et al.* 2000, 2014; Verardi *et al.* 2006b; Fabbri *et al.* 2007, 2012, 2013; Caniglia *et al.* 2014).

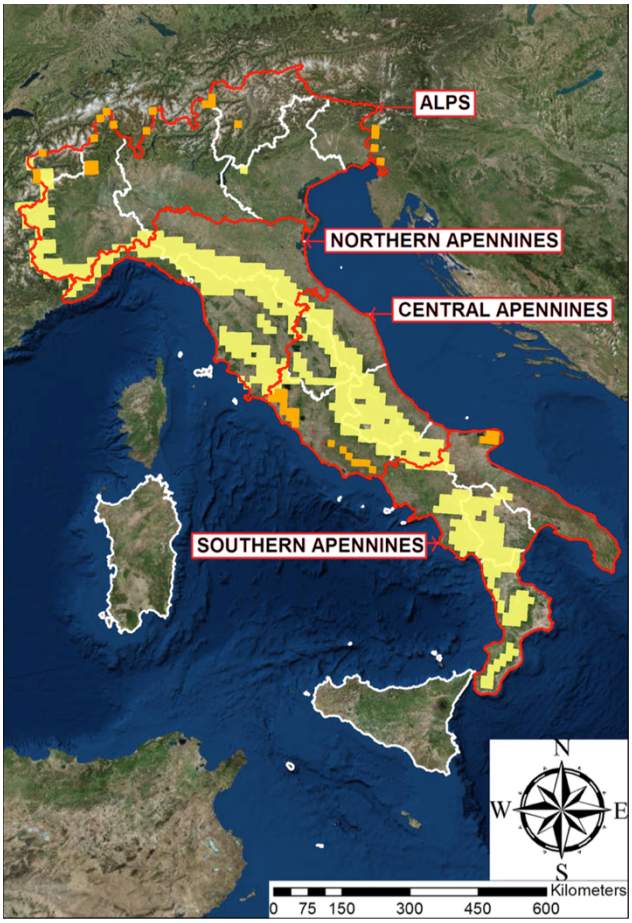


Figure 5

Wolf (*Canis lupus*) presence and distribution in Italy, obtained from data contained in the report on species of community interest compiled by ISPRA for the National Biodiversity Strategy referred to 2013 (Genovesi *et al.* 2014). Stable presence is indicated in light yellow, occasional presence in light orange. Geographical sectors are indicated and surrounded by red lines. From Galaverni *et al.* (2015).

Molecular markers in Italian wolf conservation genetics

In the past ten years mtDNA and neutral-behaving microsatellites, tandem repeats of 1–6 nucleotides found at high frequency in the nuclear genomes of most taxa also known as variable number tandem repeats (VNTR) or single tandem repeats (STRs), were the most used genetic markers to assess the basic genetic variables in animal and plant populations (Li et al. 2002). Although mtDNA evolves 5–10 times faster than single-copy nuclear DNA (scnDNA), it records few traces of contemporary events. Thus, mtDNA can be used to resolve taxonomic uncertainties. The mitochondrial control region (CR) is the major non coding region in animal mtDNA molecule. A 350 base pair of CR1 mtDNA was analysed and became diagnostic for its characteristic mutations in the identification of the unique Italian wolf haplotype W14 (Randi *et al.* 2000; Pilot *et al.* 2010).

VNTRs evolve 100–1000 times faster than scnDNA and provide a powerful tool to analyse recent and contemporary events. Microsatellites are high variable molecular markers that enable the recognition of the unique multi-locus genotype that each individual possess among many samples and even with a small number of loci. Therefore through microsatellites it is possible to provide estimates of relatedness of individuals. In species with particular demographic history due to high rate of inbreeding or long-term isolation or bottlenecks, these markers may be also diagnostic to differentiate rate of migration and admixture patterns. The best choice to describe variability at those species is to use microsatellites loci with high mutation rates, which result in high standing allelic diversity. Firstly, 12 canine microsatellite loci were identified and used in the Italian wolf population (Randi & Lucchini 2002) to estimate the species membership, social structure and kin affiliations (Caniglia *et al.* 2011, 2014; Fabbri *et al.* 2012). As hybrids identification became a major concern in conservation management, a higher number of markers were needed to increase the power of the assignment of presumptive wolf x dog individuals. Thus the panel of highly polymorphic microsatellites was expanded till joining 39 loci together with the use of uni-parental markers (Randi *et al.* 2014).

STRs are present in the Y chromosome (e.g. MS34A, MS34B, MSY41A and MS41B), and can generate haplotypes, which have different frequencies in wolves and dogs (Iacolina *et al.* 2010). Subsequently, these markers are usually genotyped to identify Y-haplotypes detectable in the population, helping to complete genealogies or species identification. Among Italian wolves only two haplotypes have been described (Fabbri *et al.* 2013; Randi *et al.* 2014).

A three-nucleotide deletion found in a gene coding for the β-defensin CBD103 (also known as the K locus) is considered responsible the melanistic mutation of coat colour in dogs (Candille *et al.* 2007). With the occurrence of first black wolves sighted in Italy, the K locus has also been analysed (Caniglia *et al.* 2013), describing the presence of the deletion in individuals having admixed wolf *x* dog microsatellites genotypes, in agreement to the domestic dog introgression theory of the trait (Hedrick 2009; Anderson *et al.* 2009; Ollivier *et al.* 2013).

SNPs represent the most widespread source of sequence variation within genomes. Although they are mostly biallelic and less informative than microsatellites, if they are used in high numbers as in microarray DNA chips, they are considered the most valuable choice in terms of cost and effects. As seen for mtDNA and multilocus microsatellite genotypes, also SNPs analysis showed a unique population structure for the Italian wolves that clusters separately from all other European populations (Stronen *et al.* 2013).

Multi-gene families

A multi-gene family is a group of genes that evolved from a common ancestor. These genes encode for proteins with similar function and as consequence of multiple duplication events can share sequence identity over their full length, or limited to a specific domain. Examples of multi-gene families include the major histocompatibility complex (MHC; Hughes and Nei 1989; Kelley, Walter, and Trowsdale 2005) and interferons (Samarajiwa *et al.* 2009), actins and tubulins (Engel *et al.* 1982; Muller 2005; Gunning *et al.* 2015), haemoglobins (Hardison 2012) and immunoglobulins (Ota and Nei 1994; Ota et al.2000), protein kinases (Robinson et al. 2000), and G-coupled proteins such as olfactory receptor genes (Buck and Axel 1991; Buck 1992; Niimura 2012).

Studies on these genes often showed evidence of positive selection and ecological niche adaptation, such was the case of olfactory receptor genes in mammals (Nei & Rooney 2005; Hayden *et al.* 2010). For this reason, they are suitable markers for evolutionary comparative and adaptive variation studies, useful in conservation genetics to evaluate the degree to which demographic declines in endangered populations can reduce functional genetic diversity and therefore impact on the population long-term viability.

With the advent of next generation technologies, it became more affordable to perform the sequencing of multiple loci compared to traditional Sanger sequencing, giving the possibility to deepen our knowledge on multi-gene families.

Major Histocompatibility Complex (MHC) genes

The MHC is a gene family responsible for the presentation of peptides for immune recognition, a role that is integral to the normal functioning of the immune system and essential for the survival of individuals. Because of their well-known role in tissue transplantation in humans and mice, these genes have been studied for over two decades and nearly every approach has been taken to examine the effects of their sequence polymorphism. Thereafter, this highly polymorphic complex also became a case of study in many natural populations, for its significant contribution to understand how the different species adapt to their environment through time and space (Bernatchez & Landry 2003). Moreover, the study of the MHC gives the possibility to investigate expectedly low genetic diversity in small isolated populations and likely predict how these populations can successfully cope with future environmental changes, such in the case of the Italian wolf population studies presented in this work.

Olfactory receptor (OR) genes

Olfactory Receptor genes are the largest multi-gene family present in vertebrate genomes. Acting as a filter to perceive, understand and communicate with the environment and other living organisms, these genes are of capital importance for survival especially in natural environments. Despite they were discovered a decade ago, ORs still challenge researchers in the pursuit of defining their origins and the evolutionary forces shaping their characterization as functional or not functional. Furthermore, even if ORs have been studied in many organisms, little is still known about their function and variability in canids. Nearly no studies have been performed on olfactory repertoire in wolves. Due to their closest genetic and phylogenetic relation among canids, dogs and wolves represent an ideal model to investigate the mechanisms underlying possible different patterns of adaptive genomic variation.

Demographic events can change the direction of evolution and selection. In dogs the creation of breeds has been demonstrated to be responsible for peculiar genetic signatures (Robin et al.2009), therefore it could be interesting to investigate if also in the wolf OR gene repertoire it is possible to retrieve signs of ancient and recent demographic declines. Moreover, the characterization of wolves OR genes is necessary to better understand the underlying biology of olfaction in those animals, and may help disentangling doubts about different olfaction ability in wolves in comparison with the well-documented capacities of dogs.

From Sanger to NGS

DNA sequencing is used within conservation biology to reveal the genetic information of organisms. In 1977 when Fred Sanger and Alan R. Coulson (Sanger *et al.* 1977) published their discovery on a novel DNA sequencing technique, biology and genetics changed forever. Their method indeed, provided an improvement in sequencing complete genes and later entire genomes, overcaming the analogous biochemical techniques developed by Maxam and Gilbert published in the same year (Maxam & Gilbert 1977). The chemistry of Sanger's method reduced the use of toxic chemicals and radioisotopes, using fluorescently labelled dideoxynucleotides (ddNTPs), which during the extension step of the polymerase chain reaction (PCR), are stochastically incorporated at the end of each duplicated fragment of DNA. The end fluorescent

labelled fragments are then read during a high-resolution electrophoresis in a capillary-based polymer gel (Swerdlow *et al.* 1991).

The Sanger method became within these last 30 years almost exclusively the only semi-automated method used for DNA sequencing. The first genomes to be assembled were produced by Sanger method, such as the human genome sequence, completed in draft form in 2001 (Lander *et al.* 2001; Venter *et al.* 2001), and thereafter came the genome sequences of several model organisms (Chinwalla *et al.* 2002; Gibbs *et al.* 2004; Sequencing & Consortium 2005).

However, despite excellent accuracy and reasonable read length (~ 1000 bp), the Sanger method is very low throughput and is expensive for genomic applications. For these reasons the National Human Genome Research Institute (NGHRI) created a 70 million dollar DNA sequencing (International Human Genome Sequencing Consortium, 2004) technology initiative aimed at achieving a $1,000 human genome in 10 years (Schloss 2008). This posed the bases for a new input in DNA sequencing technologies.

The pace at which new industrial competitors developed new sequencing chemistry and platforms have been astonishing in the last ten years. Each platform has its own strategy determining quality, quantity, and biases of the sequencing results, and thus changing their utility for different applications. All these technologies are called, in a generalist way, the next generation sequencing (NGS) technologies. In DNA sequencing a new paradigm emerged, made of three basic phases: library preparation, clonal amplification and cyclical rounds of massively parallel sequencing. The high-throughput sequencing (HTS) was born.

The first phase is the library preparation, which should guarantee the most genomic coverage with the least amount of sequencing. The quality of sequencing data often depends upon the quality of the sequenced material, which is actually prepared with different library preparation protocols. A library is the mix of fragment of DNA (or RNA) appropriately modified for HTS. The first step consists in the DNA fragmentation that can be obtained by enzymatic reaction or by physical break. The fragmentation is followed by the end-repair and adapters ligation step, which is indeed necessary for the immobilization of the molecules on a solid surface for the final parallel sequencing procedure. The last step is the size selection, during which molecules of chosen length are separated from larger or free adapters molecules using solid-phase reversible immobilization (SPRI) beads, or for a more precise selection by gel extraction. The selected molecules are then duplicated by several approaches of clonal amplification, including in situ polonies (Mitra & Church 1999), emulsion PCR (Dressman *et al.* 2003; Williams *et al.* 2006) or bridge PCR (Adessi *et al.* 2000; Fedurco *et al.* 2006). These amplifications are crucial, since may introduce serious biases, such as duplications, underrepresentation or even complete loss of

fragments with high AT- or GC content, due to different polymerase efficiency in amplifying GC-neutral than GC-rich or AT-rich fragments. These biases could be overcome by limiting the number of amplification cycles, when possible, or using a high-fidelity polymerase enzyme (Quail *et al.* 2012).

There are different types of libraries that can be used. Most of the libraries can be prepared aiming at sequencing either single fragments, paired-end and mate-pair libraries. Paired-end libraries are formed by short DNA fragments, usually a few hundred base pairs long and separated by two ends, whereas in mate-pair libraries the fragment included by the two ends is usually several thousand base-pairs long. The amount and type of libraries needed for a specific NGS application depend on the minimum depth of coverage needed to achieve the desired sensitivity and specificity. These aspects are connected to the size, number and repeat content of the target, the number of samples, the platform's read length and sequencing error rates. Indeed, each sequencing technology, according to manufacturers, declares to achieve different coverage depths, which depends on the number of reads per run and the read length.

Therefore, a combination of technologies and experimental protocols may often be appropriate for different specific projects (Babik *et al.* 2009; Galan *et al.* 2010; Kircher & Kelso 2010), of course following the available resources in terms of time, money and bioinformatic tools. An important consideration about the costs is the type of library preparation and the number of samples employed, since some protocols indeed could be expensive if applied to high number of samples. Unfortunately, biases in sample preparation, sequencing, genomic alignment and assembly can result in regions of the genome that lack coverage (gaps) and in regions with much higher coverage than theoretically expected. For example, high-quality assemblies can be reached combining the advantages of high-depth short-read sequencing with those of lower-depth but longer-read sequencing, useful to complete gaps.

Giving the same amount of targets is further possible to choose between running a single sample at a higher coverage versus running multiple samples at a lower coverage. Generally speaking, the higher is the coverage, the higher is the accuracy that could be reached, for example in polymorphic sites identification. It could also be possible to pool the samples from multiple individuals in the same run, choosing if we need individual information or not. To address such needs, some platforms allow also the addition of sample specific barcode sequences (tags; Parameswaran *et al.* 2007) that can be added before or during the library preparation. This approach is usually applied to parallel sequence multiple samples in the same run, giving the possibility to reconstruct the sequence of every specimen using bioinformatics tools, but at the same time obtaining information about the inter-individual variability. In this case, the limiting

factor is usually the number of tags that can be provided by the manufacturer, their cost and the total amount of coverage that can be reached by a single run.

Pros and cons of NGS

NGS can provide several hundred sequences per amplicon in only a fraction of the time and effort that it would cost to acquire the same amount of data using Sanger methods (NGS is 100-1000 times faster than Sanger sequencing; Kircher and Kelso 2010).

There are several advantages in using NGS: the working time, and only recently the lowering cost of this techniques promise to let them be accessible to individual investigators rather than only to major genome centres. This progress is significantly scaling-up the number of genetic markers, allowing for genome wide approaches that enhance the power and resolution of applications and improve the reliability of conclusions. An important advantage comes for the sequencing of highly polymorphic gene families, where the detection of false alleles or chimeras can change final results. With NGS is possible to detect artefact sequences, which should occur on a broader number of sequences at lower frequencies than true allele sequences (Babik *et al.* 2009; Lighten *et al.* 2014), and thus remove them from the dataset through strict allele validation steps (Galan *et al.* 2010).

The chance to sequence individuals using high-throughput, scalable, low-cost, technology constitutes an important step in our ability to connect phenotypic information to health and disease, improving our ability to diagnose and predict outcomes of diseases for individual patients. However passing from 'personalized genomes' to 'personalized medicine' also raises some ethical problems (Wheeler *et al.* 2008). Should patients receive information about their individual sequence? There are still significant limitations in the correct interpretation of personal genomes but still the impact of the clinical information on the patients is huge.

Finally another important advantage especially in ecological studies, is the possibility to use small amounts of genetic material, which makes NGS suitable for analyses of endangered species when non-invasive sampling is needed.

Anyway, there are also some limitations in NGS. As mentioned above, apparently these methods are more prone to genotyping errors, which can deeply impact on correct identification of high polymorphic sites, if a strict validation pipeline is not applied (Nielsen *et al.* 2011; Pabinger *et al.* 2013; Ekblom & Wolf 2014). All current high-throughput technologies have an average error rate that is considerably higher than the typical 1/10,000 to 1/100,000 observed for high-quality Sanger sequences (Kircher & Kelso 2010).

NGS generates terabytes of raw data, making management of data and analysis problematic with the necessity of intensive bioinformatic support. Giving this, even if sequencing costs are more accessible, it is difficult for many research laboratories to successfully conduct NGS projects because investment in data management, analysis tools, and formed personnel is needed. Even for larger, experienced genome centres this aspect remains an ever-increasing challenge for the ongoing use of NGS. Smaller research groups may still find a profitable solution in using sequencing services or cloud computing (O'Driscoll, Daugelaite, and Sleator 2013) on demand, which means, that it is possible to pay for a limited time the use of networks, servers, storage, applications and services. Of course, this is a compromise between significant outlays to maintain and sustain a complete NGS network and investing money each time for a single project using facilities from external services. However, this strategy may not be advantageous for long periods or many projects.

Moreover, each NGS platform requires bioinformatics tools that should be able to handle with different data format, length of reads, and specific type of errors. For this reason there are many tools (Shendure & Ji 2008) for almost all functions: quality scoring, mapping and alignment, de novo assembly or reference-based assembly (Zhang *et al.* 2011), base-calling and/or genetic variation detection (such as SNV, Indel), genome annotation, and much more utilities for data analysis and processing. As time goes by, communities of developers and expert users grow around each tool, to freely improve bioinformatics knowledge with tutorials and explanations reachable by common users. However, these costs time and benefits of NGS will not be achieved until bioinformatics will be able to maximally simplify and standardize protocols, and since extremely high-performance computing will not be available for everyone.

Applications of next-generation sequencing

The most known application is the whole genome sequencing (WGS). Generally, only a single individual is sequenced, with the purpose to identify genome sequences that may be associated with disease, or are predictive of response to medication (Wheeler et al. 2008). Nevertheless, sometimes (like in the HUGO project) the genome represents a 'consensus' of a number of pooled samples (International Human Genome Sequencing Consortium 2004). In WGS the candidate genome is sequenced up to a sufficient coverage useful to identify individual comprehensive polymorphism (Single Nucleotide Variants; SNVs) and mutation discovery. This information can then be applied at a population level to discover markers such as microsatellites that should be used in population genetics studies or in Genome-Wide Association Studies

(GWAS). GWAS addresses the need to perform genotype-phenotype correlations at the genomic level, for example to identify putative genes whose mutational status contribute to the phenotype, examine changes in gene expression or evaluate changes in genome-wide methylation patterns (Meaburn & Schulz 2012). However, this is not possible when the reference genome, or a genome of a close relative species is not available, in this case a *de novo* assembly process is required, e.g. the case of the de novo assembly of the giant panda (*Ailuropoda melanoleuca*) genome (Li *et al.* 2010). Further, WGS is also used for the mapping of structural rearrangements, which may include copy number variation, balanced translocation breakpoints (Chen *et al.* 2008) and chromosomal inversions.

However, sometimes is neither practical nor necessary to sequence entire genomes, since it could be possible to focus only on defined regions of the genome. By decreasing the number of targets will increase the coverage, and more valuable information can be obtained. One of these targeted regions can be exomes, as in whole-exome sequencing (WES) where only the coding regions of the genome are sequenced. But also multiple independent loci, or long stretches of genomic DNA can be targeted and re-sequenced for targeted polymorphism and mutation discovery (Dahl *et al.* 2007).

After WGS, RNA sequencing (RNA-seq) allows the detection and the quantification of expressed transcripts in a biological sample and is currently the most common application of NGS. In non-model species this is used to develop molecular markers, create targeted sequencing assays or to construct microarrays for gene expression profiling, transcript annotation, discovery of transcribed SNPs or somatic mutations, and quantification of gene expression and alternative splicing. Gene expression profiling using microarrays (Kammenga *et al.* 2007) is a method for measuring gene expression by hybridising complementary DNA (cDNA), synthetized from a mRNA template, to a chip with a large number of microscopic spots with DNA oligonucleotides. This approach has mainly been restricted to model species with previous genome information. Another method of expression profiling is by digital transcriptomics (Murray *et al.* 2007), where the representation of specific sequences derived from deep cDNA sequencing is proportional to the amount of RNA from the gene in the original sample (t Hoen *et al.* 2008). An example of how different assays and NGS methods can be used is the ENCODE project (Encode Project Consortium 2004), which has the main purpose to report all DNA regulatory elements, including those acting at the protein and RNA levels, and those that control cells and circumstances in which a gene is active. Regulatory elements are typically investigated through DNA hypersensitivity assays, assays of DNA methylation, and immunoprecipitation (IP) of proteins that interact with DNA and RNA, i.e., modified histones,

transcription factors, chromatin regulators, and RNA-binding proteins, followed by sequencing.

Genomic approaches can also be applied to highly fragmented DNA from ancient material (e.g. from museum specimens; Bi *et al.* 2013), or for metagenomic sequencing for the study of physiology and ecology of environmental microorganisms (Handelsman 2004), or for the discovery of infectious and commensal flora, like in the Human Microbiome Project (Human Microbiome Project Consortium 2012) which has used NGS to characterize the diversity and types of bacteria and viruses present in the human body (Gevers *et al.* 2012). Therefore, in the future more applications of NGS will of course be employed, and the complexity with which these applications are performed will parallel grow.

The role of high throughput sequencing in conservation genetics and genomics

HTS ensures the possibility to increase the number of genetic markers by which conservation genetics, with limited surveys of neutral marker variation, traditionally used to address demographic, environmental, and genetic issues (Selkoe & Toonen 2006). Indeed, despite its long history, the relevance of neutral genetic variation to predict and describe the fitness of individuals remains debated (Väli *et al.* 2008).

Although variation in few neutral markers might conceptually predict levels of detrimental and adaptive variation (Hedrick 2001), empirical studies suggest that neutral variation does not correlate closely with detrimental and adaptive variation (Reed & Frankham 2003) and underestimates population differentiation in quantitative traits (Merila & Crnokrak 2001). Hence, small fraction of the genome sampled for a limited number of neutral markers might lead to misconceptions about levels of detrimental and adaptive variation in endangered species.

In these years, the rapid progress of cost-efficient NGS methods has facilitated the development of so-called "-omics" increasing the scale and dimensions of accessible molecular information for evolutionary and conservation biology studies, opening the gates to new opportunities in genomewide scans for neutral, detrimental and adaptive variation.

An alternative to classic microsatellites and mitochondrial markers is to sample higher amount of SNPs throughout animal and plant genome. SNPs array are already used for demographic and admixture inferences (vonHoldt *et al.* 2011; Novembre & Ramachandran 2011; Stronen *et al.* 2013, 2015), and also for association mapping within candidate regions of adaptive interest (Jarvis *et al.* 2012).

While neutral genetic variation is shaped by interaction of mutation, genetic drift, recombination and migration, the selective forces shape adaptive variation. One of the most

important application of genomic approach to conservation studies is the ability to identify adaptive loci (Allendorf *et al.* 2010), allowing to describe genetic changes that accompany local adaptations and the way in which these changes may influence fitness and habitats requirements of populations (Crandall *et al.* 2000). This knowledge will assist in defining conservation units in the wild (Manel *et al.* 2010) and may help evaluate the potential of populations to respond to changing environments (Hoffmann & Sgrò 2011).

Multigene families are often used to infer adaptive variation levels, utilizing sequence polymorphism within and divergence between species. With these genes in particular, high-throughput, high-quality sequencing and assembly can be quite challenging, particularly when a species lacks a reference genome (Galan *et al.* 2010; Ekblom & Galindo 2011). Previous studies at the MHC using 454 sequencing have further highlighted the challenges involved in large multigene sequencing, proposing different evaluation and quality control protocols (Babik *et al.* 2009; Galan *et al.* 2010; Zagalska-Neubauer *et al.* 2010; Sommer *et al.* 2013). However even if a standard protocol has not been validated and accepted (Lighten *et al.* 2014). In the future the genotyping of multigene families would of course profit of the increase of the amplicon sequencing depth using NGS methods together with new unanimous model-based approaches.

In conclusion, the results previously obtained by conservation genetics approaches have often aided management of rare and endangered species, thus it could be difficult to push institutions to make the step from conservation genetics to conservation genomics, especially when the latter remains still very expensive for small conservation projects. However, conservation genomics surveys programs should be developed to combine detailed demographic analyses based on neutral marker data with the application of genomic approaches, especially to infer the levels of detrimental and adaptive variation in threatened natural populations. Furthermore, it may be helpful to use genome wide comparative approach to compare significant differences in variation between endangered and closely related non-endangered taxa, to identify changes in levels of negative or adaptive variation. Finally, the genome wide study of patterns of variation may allow the identification of selective forces changing population allele frequencies of variation, which can be crucial in conservation biology, since changes in population demography can change the status of some variants from neutral to adaptive or detrimental.

AIMS OF THE THESIS

Variation at neutral loci cannot provide direct information on selective processes involved in the interaction between individuals and their environment, nor the capacity for future adaptive changes. Studying the genetic structure of canid populations and their spatial distribution of genetic variation has helped increasing our knowledge on this field and it also allows crucial decisions in the management of those species.

In this thesis the use of different molecular investigation tools from ten years old sequencing Sanger method to next generation sequencing (NGS) have been used to sequence several loci belonging to two of the largest gene families ever described in mammals, with the aim to describe present patterns of genetic variability within and among canids groups.

The results obtained from conservation genetics studies that combined neutral markers and genes under selection such as three MHC class II loci are presented in the PART I of the thesis "The MHC genes typing project" with the main aim to help ameliorate evaluations about the conservation status of the studied wild species. However, each paper addresses different needs in conservation genetics knowledge.

The main objective of PAPER I (Galaverni *et al.* 2013) was to analyse population variability and discuss the evolutionary hypotheses, i.e. long bottleneck and isolation, which could have caused a loss of variability at MHCs in the Italian wolf population. In PAPER II (Galov *et al.* 2015) the main aim was to use a genetic multilocus approach to identify population-specific patterns of variation that can help in the discrimination of putative hybrid individuals found in the Croatian jackal population. Finally, in the PAPER III (Galaverni *et al.* 2015) we wanted to give an insight into the main behavioural process guided by innate MHC genetic identity involved in mate choice selection, trying to identify the presence of assortative or dissassortative mating preferences within Italian wolves.

In the PART II of the thesis "The olfactory receptor genes typing project" the research has been addressed on ORs to perform a comparative study between domestic dogs and wolves described in the PAPER IV (Lapalombella *et al.* in prep). The olfactory system, like the immune system, works interacting with the environment, and thus evolves at pace with its changes. Since different species live in different niches (with different immune and olfactory requirements), selective pressures should modulate ORs in conjunction with speciation, domestication and breeds creation. Degenerate primers and a gradient PCR method were used to parallel sequence multiple genes. Based on this approach we tried to obtain a picture of how many genes could be sequenced by our method among the ~1,000 genes previously described in dog olfactory

repertoire (Galibert *et al.* 2011). Subsequently, we aimed to identify population-specific ORs, and describe polymorphisms to determine presence of significant divergences between Italian wolves and domestic dogs. This also allowed us to test the effects driven by selective forces associated with demography to identify which genes are under neutrality and which have been selectively driven to divergence in the two groups.

REFERENCES

Adessi C, Matton G, Ayala G *et al.* (2000) Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic acids research*, **28**, E87.

Allendorf FW, Hohenlohe P a, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics*, **11**, 697–709.

Altobello, G. 1921. Mammiferi IV, Carnivori. In: Fauna d'Abruzzo e Molise. Campobasso, Italy, pp 38–45.

Andersen LW, Harms V, Caniglia R *et al.* (2015) Long-distance dispersal of a wolf, Canis lupus, in northwestern Europe. *Mammal Research*.

Anderson TM, vonHoldt BM, Candille SI *et al.* (2009) Molecular and evolutionary history of melanism in North American gray wolves. *Science (New York, N.Y.)*, **323**, 1339–43.

Andersone Ž, Lucchini V, Ozoliņš J (2002) Hybridisation between wolves and dogs in Latvia as documented using mitochondrial and microsatellite DNA markers. *Mammalian Biology - Zeitschrift für Säugetierkunde*, **67**, 79–90.

Apollonio M (2004) Wolves in the Casentinesi Forests: insights for wolf conservation in Italy from a protected area with a rich wild prey community. *Biological Conservation*, **120**, 249–260.

Babik W, Taberlet P, Ejsmond MJ, Radwan J (2009) New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Molecular ecology resources*, **9**, 713–9.

Bardeleben C, Moore RL, Wayne RK (2005) A molecular phylogeny of the Canidae based on six nuclear loci. *Molecular phylogenetics and evolution*, **37**, 815–31.

Barja I, Miguel FJ De, Bárcena F (2005) Faecal marking behaviour of Iberian wolf in different zones of their territory. , **54**, 21–29.

Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of evolutionary biology*, **16**, 363–77.

Bi K, Linderoth T, Vanderpool D *et al.* (2013) Unlocking the vault: Next-generation museum population genomics. *Molecular Ecology*, **22**, 6018–6032.

Boitani L. 2003a. In Mech D.L. and L. Boitani (eds). Wolves: behavior, ecology and conservation. The University of Chicago Press, Chicago, USA, pp. 317-340.

Boitani, L., Lovari, S., & Vigna Taglianti, A. 2003b. "Fauna d'Italia Mammalia III Carnivora-Artiodactyla." Calderoni, Bologna, Italy.

Boitani, L. (2000). Action plan for the conservation of wolves in Europe (Canis lupus) (No. 18-113). Council of Europe.

Boitani L, Ciucci P, Raganella-Pelliccioni E (2010) Ex-post compensation payments for wolf predation on livestock in Italy: a tool for conservation? *Wildlife Research*, **37**, 722.

Buck LB (1992) The olfactory multigene family. *Current opinion in neurobiology*, **2**, 282–8.

Buck L, Axel R (1991) A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell*, **65**, 175–187.

Candille SI, Kaelin CB, Cattanach BM *et al.* (2007) A -defensin mutation causes black coat color in domestic dogs. *Science (New York, N.Y.)*, **318**, 1418–23.

Caniglia R, Fabbri E, Cubaynes S *et al.* (2011) An improved procedure to estimate wolf abundance using non-invasive genetic sampling and capture–recapture mixture models. *Conservation Genetics*, 53–64.

Caniglia R, Fabbri E, Galaverni M, Milanesi P, Randi E (2014) Noninvasive sampling and genetic variability, pack structure, and dynamics in an expanding wolf population. *Journal of Mammalogy*, **95**, 41–59.

Caniglia R, Fabbri E, Greco C *et al.* (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.

Carmichael LE, Nagy J a., Larter NC, Strobeck C (2001) Prey specialization may influence patterns of gene flow in wolves of the Canadian Northwest. *Molecular Ecology*, **10**, 2787–2798.
Chapron G, Kaczensky P, Linnell JDC *et al.* (2014) Recovery of large carnivores in Europe's modern human-dominated landscapes. *Science*, **346**, 1517–1519.
Chen W, Kalscheuer V, Tzschach A *et al.* (2008) Mapping translocation breakpoints by next-generation sequencing. *Genome Research*, **18**, 1143–1149.
Chinwalla AT, Cook LL, Delehaunty KD *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.
Ciucci P, Boitani L (1998) Wolf and Dog Depredation on Livestock in Central Italy. *Wildlife Society Bulletin (1973-2006)*, **26**, 504–514.
Ciucci P, Lucchini V, Boitani L, Randi E (2003) Dewclaws in wolves as evidence of admixed ancestry with dogs. *Canadian Journal of Zoology*, **81**, 2077–2081.
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.
Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution*, **15**, 290–295.
Dahl F, Stenberg J, Fredriksson S *et al.* (2007) Multigene amplification and massively parallel sequencing for cancer mutation discovery. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 9387–92.
Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B (2003) Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 8817–22.
Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, **107**, 1–15.
Ekblom R, Wolf JBW (2014) A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, **7**, 1026–1042.
Engel J, Gunning P, Kedes L (1982) Human cytoplasmic actin proteins are encoded by a multigene family. *Molecular and cellular biology*, **2**, 674–684.
Fabbri E, Caniglia R, Kusak J *et al.* (2013) Genetic structure of expanding wolf (Canis lupus) populations in Italy and Croatia, and the early steps of the recolonization of the Eastern Alps. *Mammalian Biology - Zeitschrift für Säugetierkunde*.
Fabbri E, Caniglia R, Mucci N *et al.* (2012) Comparison of single nucleotide polymorphisms and microsatellites in non-invasive genetic monitoring of a wolf population. *Archives of Biological Sciences*, **64**, 321–335.
Fabbri E, Miquel C, Lucchini V *et al.* (2007) From the Apennines to the Alps: colonization genetics of the naturally expanding Italian wolf (Canis lupus) population. *Molecular ecology*, **16**, 1661–71.
Fan Z, Silva P, Gronau I *et al.* (2016) Worldwide patterns of genomic variation and admixture in gray wolves. *Genome Research*, **26**, 163–173.
Fedurco M, Romieu A, Williams S, Lawrence I, Turcatti G (2006) BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic acids research*, **34**, e22.
Freedman AH, Gronau I, Schweizer RM *et al.* (2014) Genome sequencing highlights the dynamic early history of dogs. *PLoS genetics*, **10**, e1004016.
Galan M, Guivier E, Caraux G, Charbonnel N, Cosson J-F (2010) A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC genomics*, **11**, 296.
Galaverni M, Caniglia R, Fabbri E, Milanesi P, Randi E (2016) One, no one, or one hundred thousand: how many wolves are there currently in Italy? *Mammal Research*, **61**, 13–24.

Galibert F, Quignon P, Hitte C, André C (2011) Toward understanding dog evolutionary and domestication history. *Comptes rendus biologies*, **334**, 190–6.
Geffen E, Anderson MJ, Wayne RK (2004) Climate and habitat barriers to dispersal in the highly mobile grey wolf. *Molecular Ecology*, **13**, 2481–2490.
Genovesi P, Angelini P, Bianchi E, Dupré E, Ercole S, Giacanelli V, Ronchi F, Stoch F (2014). Specie e habitat di interesse comunitario in Italia: distribuzione, stato di conservazione e trend. Serie Rapporti 194/2014, ISPRA.
Gevers D, Knight R, Petrosino JF *et al.* (2012) The Human Microbiome Project: A Community Resource for the Healthy Human Microbiome. *PLoS Biology*, **10**, 6–11.
Gibbs R a, Weinstock GM, Metzker ML *et al.* (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, **428**, 493–521.
Godinho R, Llaneza L, Blanco JC *et al.* (2011) Genetic evidence for multiple events of hybridization between wolves and domestic dogs in the Iberian Peninsula. *Molecular ecology*, **20**, 5154–66.
Gray MM, Granka JM, Bustamante CD *et al.* (2009) Linkage disequilibrium and demographic history of wild and domestic canids. *Genetics*, **181**, 1493–505.
Gunning PW, Ghoshdastider U, Whitaker S, Popp D, Robinson RC (2015) The evolution of compositionally and functionally distinct actin filaments. *Journal of cell science*, **128**, 2009–2019.
Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and molecular biology reviews : MMBR*, **68**, 669–685.
Hardison RC (2012) Evolution of hemoglobin and its genes. *Cold Spring Harbor perspectives in medicine*, **2**, a011627.
Harrington, F. H. & Asa, C. S. 2003: Wolf communication. In: Wolves: Behavior, ecology, and conservation (Mech, D. & Boitani, L., eds). University of Chicago Press, Chicago, pp. 66—103.
Hayden S, Bekaert M, Crider T a. *et al.* (2010) Ecological adaptation determines functional mammalian olfactory subgenomes. *Genome Research*, **20**, 1–9.
Hedrick PW (2001) Conservation genetics: Where are we now? *Trends in Ecology and Evolution*, **16**, 629–636.
Hedrick PW (2009) Wolf of a different colour. *Heredity*, **103**, 435–436.
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 (S-O Kolokotronis, Ed,). *PLoS ONE*, **7**, e46465.
Hindrikson M, Remm J, Männil P *et al.* (2013) Spatial genetic analyses reveal cryptic population structure and migration patterns in a continuously harvested grey wolf (Canis lupus) population in north-eastern Europe. *PloS one*, **8**, e75765.
Hoffmann A a, Sgrò CM (2011) Climate change and evolutionary adaptation. *Nature*, **470**, 479–485.
Hughes a. L, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: vidence for overdominant selection. *Proceedings of the National Academy of Sciences of the United States of merica*, **86**, 958–62.
Human Microbiome Project Consortium T (2012) Structure, function and diversity of the healthy human microbiome. *Nature*, **486**, 207–214.
Iacolina L, Scandura M, Gazzola A *et al.* (2010) Y-chromosome microsatellite variation in Italian wolves: A contribution to the study of wolf-dog hybridization patterns. *Mammalian Biology - Zeitschrift für Säugetierkunde*, **75**, 341–347.
Jarvis JP, Scheinfeldt LB, Soi S *et al.* (2012) Patterns of ancestry, signatures of natural selection, and genetic association with stature in Western African pygmies. *PLoS genetics*, **8**, e1002641.
Kammenga JE, Herman M a., Ouborg NJ, Johnson L, Breitling R (2007) Microarray challenges in ecology. *Trends in Ecology and Evolution*, **22**, 273–279.

Kelley J, Walter L, Trowsdale J (2005) Comparative genomics of major histocompatibility complexes. *Immunogenetics*, **56**, 683–695.
Kircher M, Kelso J (2010) High-throughput DNA sequencing - Concepts and limitations. *BioEssays*, **32**, 524–536.
Koepfli K-P, Pollinger J, Godinho R *et al.* (2015) Genome-wide Evidence Reveals that African and Eurasian Golden Jackals Are Distinct Species. *Current Biology*, **25**, 2158–2165.
Lander ES, Linton LM, Birren B *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
Larivière S, Crête M (2007) The size of eastern coyotes (Canis latrans): a comment. *Journal of mammalogy*, **74**, 1072–1074.
Li R, Fan W, Tian G *et al.* (2010) The sequence and de novo assembly of the giant panda genome. *Nature*, **463**, 311–317.
Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: Genomic distribution, putative functions and mutational mechanisms: A review. *Molecular Ecology*, **11**, 2453–2465.
Lighten J, van Oosterhout C, Bentzen P (2014) Critical review of NGS analyses for de novo genotyping multigene families. *Molecular Ecology*, **23**, 3957–3972.
Lindblad-Toh K, Wade CM, Mikkelsen TS *et al.* (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, **438**, 803–819.
Linnell JDC, Andersen R, Andersone Ž *et al.* (2002) The fear of wolves: a review of wolf attacks on humans. *Norw. Inst. Nature Res. Oppdragsmelding*, **731**, 1–65.
Linnell JDC, Salvatori V, Boitani L (2007) Guidelines for Population Level Management Plans for Large Carnivore in Europe. A Large Carnivore Initiative for Europe report prepared for the European Commission. Final draft May 2007.
Lucchini V, Fabbri E, Marucco F *et al.* (2002) Noninvasive molecular tracking of colonizing wolf (Canis lupus) packs in the western Italian Alps. *Molecular ecology*, **11**, 857–68.
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.
Manel S, JOOST S, EPPERSON BK *et al.* (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology*, **19**, 3760–3772.
Marucco F, Pletscher DH, Boitani L *et al.* (2009) Wolf survival and population trend using non-invasive capture recapture techniques in the Western Alps. *Journal of Applied Ecology*, **46**, 1003–1010.
Maxam a M, Gilbert W (1977) A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **74**, 560–564.
Meaburn E, Schulz R (2012) Next generation sequencing in epigenetics: insights and challenges. *Seminars in cell & developmental biology*, **23**, 192–9.
Mech D (1999) Alpha status , dominance , and division of labor in wolf packs. , **1203**, 1196–1203.
Mech, L.D. (1970). The wolf: the ecology and behavior of an endangered species. Natural History Press, Doubleday Publishing Co., NY, USA.
Mech, L.D. (1974). Canis lupus. Mammalian Species **37**, 1–6.
Mech, D. (1999). Alpha Status , Dominance , and Division of Labor in Wolf Packs **1203**, 1196–1203.
Mech, L.D. (2002) Breeding season of wolves, Canis lupus, in relation to latitude. Canadian Field Naturalist **116**, 139– 140.
Meier TJ, Burch JW, Mech LD, Adams LG (1995) Pack structure and genetic relatedness among wolf packs in a naturally-regulated population. In: Ecology and Conservation of Wolves in a Changing World (eds Carbyn LN, Fritts SH and Seip DR), pp. 293–302. Canadian Circumpolar Institute, Edmonton, AB.
Mengel RM (1971) A study of dog-coyote hybrids and implications concerning hybridization in canis. *Journal of mammalogy*, **52**, 316–336.
Merila J, Crnokrak P (2001) Comparison of genetic differentiation at marker loci and quantitative

traits. *Journal of Evolutionary Biology*, **14**, 892–903.
Milanesi P, Caniglia R, Fabbri E *et al.* (2015) Non-invasive genetic sampling to predict wolf distribution and habitat suitability in the Northern Italian Apennines: implications for livestock depredation risk. *European Journal of Wildlife Research*.
Mitra RD, Church GM (1999) In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic acids research*, **27**, e34.
Muller J (2005) Sequence and Comparative Genomic Analysis of Actin-related Proteins. *Molecular Biology of the Cell*, **16**, 5736–5748.
Murray D, Doran P, MacMathuna P, Moss AC (2007) In silico gene expression analysis--an overview. *Molecular cancer*, **6**, 50.
Musiani M, Leonard J a, Cluff HD *et al.* (2007) Differentiation of tundra/taiga and boreal coniferous forest wolves: genetics, coat colour and association with migratory caribou. *Molecular ecology*, **16**, 4149–70.
Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annual review of genetics*, **39**, 121–152.
Nielsen R, Paul JS, Albrechtsen A, Song YS (2011) Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, **12**, 443–451.
Niimura Y (2012) Olfactory Receptor Multigene Family in Vertebrates: From the Viewpoint of Evolutionary Genomics. *Current Genomics*, **13**, 103–114.
Novembre J, Ramachandran S (2011) Perspectives on human population structure at the cusp of the sequencing era. *Annual review of genomics and human genetics*, **12**, 245–74.
Nowak RM, Federoff NE (2002) The systematic status of the Italian wolf Canis lupus. *Acta Theriologica*, **47**, 333–338.
O'Driscoll A, Daugelaite J, Sleator RD (2013) "Big data", Hadoop and cloud computing in genomics. *Journal of Biomedical Informatics*, **46**, 774–781.
Ollivier M, Tresset A, Hitte C *et al.* (2013) Evidence of coat color variation sheds new light on ancient canids. *PloS one*, **8**, e75110.
Ota T, Nei M (1994) Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Molecular biology and evolution*, **11**, 469–482.
Ota T, Sitnikova T, Nei M (2000) Evolution of vertebrate immunoglobulin variable gene segments. *Current topics in microbiology and immunology*, **248**, 221–45.
Pabinger S, Dander A, Fischer M *et al.* (2013) A survey of tools for variant analysis of next-generation genome sequencing data. *Briefings in bioinformatics*.
Parameswaran P, Jalili R, Tao L *et al.* (2007) A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Research* , **35** , e130.
Perini FA, RUSSO CAM, SCHRAGO CG (2010) The evolution of South American endemic canids: a history of rapid diversification and morphological parallelism. *Journal of Evolutionary Biology*, **23**, 311–322.
Peterson RO, Jacobs AK, Drummer TD, Mech LD, Smith DW (2002) Leadership behavior in relation to dominance and reproductive status in gray wolves , Canis lupus. , **1412**, 1405–1412.
Pilot M, Branicki W, Jędrzejewski W *et al.* (2010) Phylogeographic history of grey wolves in Europe. *BMC Evolutionary Biology*, **10**, 104.
Pilot M, Greco C, vonHoldt BM *et al.* (2014) Genome-wide signatures of population bottlenecks and diversifying selection in European wolves. *Heredity*, **112**, 428–42.
Pilot M, Jedrzejewski W, Branicki W *et al.* (2006) Ecological factors influence population genetic structure of European grey wolves. *Molecular Ecology*, **15**, 4533–4553.
Project. TE (ENCyclopedia ODE (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science (New York, N.Y.)*, **306**, 636–640.
Quail MA, Otto TD, Gu Y *et al.* (2012) Optimal enzymes for amplifying sequencing libraries. *Nat Meth*, **9**, 10–11.
Randi E (2003) Conservation genetics of carnivores in Italy. *Comptes Rendus Biologies*, **326**, 54–

60.
Randi E, Hulva P, Fabbri E *et al.* (2014) Multilocus Detection of Wolf x Dog Hybridization in Italy, and Guidelines for Marker Selection (G Larson, Ed,). *PLoS ONE*, **9**, e86409.
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*, 31–45.
Randi E, Lucchini V, Christensen MF *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.
Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. *Conservation Biology*, **17**, 230–237.
Robinson DR, Wu Y-M, Lin S-F (2000) The protein tyrosine kinase family of the human genome. , 5548–5557.
Samarajiwa SA, Forster S, Auchettl K, Hertzog PJ (2009) INTERFEROME: the database of interferon regulated genes. *Nucleic Acids Research*, **37**, D852–D857.
Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, **74**, 5463–7.
Schloss JA (2008) How to get genomes at one ten-thousandth the cost. *Nat. Biotechnol.*, **26**, 1113–1115.
Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, **9**, 615–629.
Sequencing TC, Consortium A (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, **437**, 69–87.
Sergio F, Caro T, Brown D *et al.* (2008) Top Predators as Conservation Tools: Ecological Rationale, Assumptions, and Efficacy. *Annual Review of Ecology, Evolution, and Systematics*, **39**, 1–19.
Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nature biotechnology*, **26**, 1135–45.
Sommer S, Courtiol A, Mazzoni CJ (2013) MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. *BMC genomics*, **14**, 542.
Stronen AV, Je B, Pertoldi C *et al.* (2015) Genome-wide analyses suggest parallel selection for universal traits may eclipse local environmental selection in a highly mobile carnivore. , 1–16.
Stronen A V, Jędrzejewska B, Pertoldi C *et al.* (2013) North-South Differentiation and a Region of High Diversity in European Wolves (Canis lupus). *PloS one*, **8**, e76454.
Swerdlow H, Zhang JZ, Chen DY *et al.* (1991) Three DNA sequencing methods using capillary gel electrophoresis and laser-induced fluorescence. *Analytical chemistry*, **63**, 2835–41.
t Hoen P a C, Ariyurek Y, Thygesen HH *et al.* (2008) Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Research*, **36**, e141.
Thurber J, Peterson R (1991) Changes in body size associated with range expansion in the coyote (Canis latrans). *Journal of Mammalogy*, **72**, 750–755.
Väli Ü, Einarsson A, Waits L, Ellegren H (2008) To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology*, **17**, 3808–3817.
Venter JC, Adams MD, Myers EW *et al.* (2001) The sequence of the human genome. *Science (New York, N.Y.)*, **291**, 1304–1351.
Verardi a, Lucchini V, Randi E (2006a) Detecting introgressive hybridization between free-ranging domestic dogs and wild wolves (Canis lupus) by admixture linkage disequilibrium analysis. *Molecular ecology*, **15**, 2845–55.
Verardi a, Lucchini V, Randi E (2006b) Detecting introgressive hybridization between free-ranging domestic dogs and wild wolves (Canis lupus) by admixture linkage disequilibrium analysis. *Molecular ecology*, **15**, 2845–55.

Vila C (1997) Multiple and Ancient Origins of the Domestic Dog. *Science*, **276**, 1687–1689.
Vila C, Amorim I, Leonard J *et al.* (1999) Mitochondrial DNA phylogeography and population history of the grey wolf canis lupus. *Molecular ecology*, **8**, 2089–103.
Vilà C, Walker C, Sundqvist a-K *et al.* (2003) Combined use of maternal, paternal and bi-parental genetic markers for the identification of wolf-dog hybrids. *Heredity*, **90**, 17–24.
vonHoldt BM, Pollinger JP, Earl D a *et al.* (2011) A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. *Genome Research*, **21**, 1294–1305.
Vonholdt BM, Pollinger JP, Lohmueller KE *et al.* (2010) Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature*, **464**, 898–902.
Wang, X., Tedford, R. H., Van Valkenburgh, B., & Wayne, R. K. (2004). Evolutionary history, molecular systematics, and evolutionary ecology of Canidae. Biology and conservation of wild canids (DW Macdonald and C. Sillero-Zubiri, eds.). Oxford University Press, Oxford, United Kingdom, 39-54.
Wayne RK, Geffen ELI, Girman DJ *et al.* (1997) MOLECULAR SYSTEMATICS OF THE CANIDAE jackal- , and foxlike carnivores that includes gentina . The first appearance of South American canids followed the immigration of North American mammals into South America during the early Pleistocene after the geolog. , **46**, 622–653.
Wayne RK, Nash WG, O'Brien SJ (1987) Chromosomal evolution of the Canidae. *Cytogenetic and Genome Research*, **44**, 134–141.
Wheeler D a, Srinivasan M, Egholm M *et al.* (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature*, **452**, 872–6.
Williams R, Peisajovich SG, Miller OJ *et al.* (2006) Amplification of complex gene libraries by emulsion PCR. *Nat Meth*, **3**, 545–550.
Zagalska-Neubauer M, Babik W, Stuglik M *et al.* (2010) 454 sequencing reveals extreme complexity of the class II Major Histocompatibility Complex in the collared flycatcher. *BMC evolutionary biology*, **10**, 395.
Zhang J, Chiodini R, Badr A, Zhang G (2011) The impact of next-generation sequencing on genomics. *Journal of genetics and genomics = Yi chuan xue bao*, **38**, 95–109.

PART I – The MHC genes typing project

Introduction to the Major Histocompatibility Complex (MHC)

I briefly introduced in the "General introduction" of this thesis the growing interest in using the Major Histocompatibility Complex (MHC) loci as molecular markers to deepen our knowledge on the genetic variability of the Italian wolf population. In this first part of the thesis I will present the main scientific background of this multi-gene family to introduce the results obtained by three studies that we performed on two canid populations (Italian wolf and Croatian jackal).

MHC genes were firstly described in humans in the 1950s in studies on skin graft rejection (Billingham and Medawar 1951; 1953). The initial studies by Medawar et al. were then applied in the 1980s in mice (Pierson et al. 1989; Rosenberg and Singer 1992) clarifying the MHC role in immunological activities of self and non self detection. Since 20 years, markers supposed to be under neutral selection (SNPs, microsatellites and mtDNA control region sequences) have been employed in population genetics to discover the amount of variability and perform phylogenetic reconstructions on wild species. Nonetheless, emerging evidences of MHCs high variability demonstrated that these genes could be used more efficiently than neutral markers to detect signals of selection at smaller evolutionary time scales. Thereafter, more and more studies to describe levels of polymorphism at MHCs have been done with the indirect purpose to provide also measures of the immunological fitness of a population.

Structure and function

The MHC comprehends a set of genes that plays a pivotal role in self/nonself peptide recognition and trigger the T-cell mediated specific immune response. MHC class I molecules are found in all cells and present intracellular pathogen peptides to CD8+ T lymphocytes (T cells), primarily in response to viral infections. Inversely, MHC class II molecules are only found on specialized immune cells, for example macrophages, and present extracellular pathogen peptides to CD4+ T cells after invasion by bacteria and fungi. Both types of MHC proteins are heterodimers, incorporating three α-chain domains and a single β-microglobulin peptide in class I molecules, and two α- and two β-chains in class II molecules.

The combination of folds and pockets between specific portions of the α1-α2, or the α1-β1 chains constitute the peptide-binding region (PBR), where peptides belonging to pathogens are received and exposed outside the cellular membrane in order to start the immune response. A third group (class III), although less well-studied, is implicated in other immune functions and consists in

several proteins of the complement system. Class III are also includes cytokines with roles of immune signalling, and heat shock proteins for protecting cells from thermal stresses.

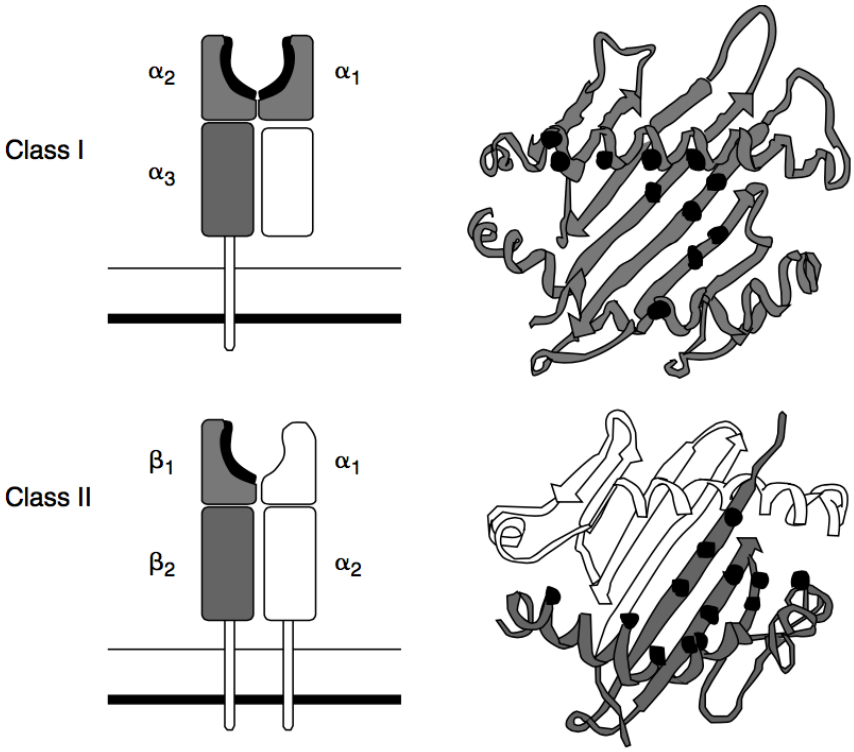


Figure 1

Schematic representation of MHC molecules class I and class II from (Höglund 2009). From the left, the molecules are seen from the side with the cell surfaces at the bottom. Antigen-binding sites are shown by the black areas and the approximate positions of α and β chains are indicated. To the right the molecules are shown from the top with the antigen-binding sites in black.

Genetic organization

Among vertebrates, the MHC shows variation in size and organization (Beck and Trowsdale 2000; Kelley, Walter, and Trowsdale 2005). Chromosome breaks, inversions, and centromere invasion occurred in the MHC region during evolution. In humans, the HLA (Human Leukocyte Antigen) complex contains 421 loci (Horton et al. 2004) in the centromeric region of chromosome 6. Genes from this region can produce 3 class I proteins (HLA-A, B, C), more than six class II proteins (DP, DM, DOA, DOB, DQ, e DR) and class III proteins too. In domestic chicken the classical region (BF/BL) is much smaller (about 20 genes) and is therefore sometimes referred to be as the minimal

essential (Kaufman et al. 1999). In dogs MHC is referred to as the to dog leukocyte antigen (DLA) family, composed of two super-contigs. This division predates to the ancestral split (55 Mya) of carnivores between felids and canids. Two-thirds of canine MHCs are located near the centromere border of canine chromosome 12, and the remaining ones can be found on the telomeric region of canine chromosome 35, in proximity to a few olfactory genes too (Yuhki et al. 2007).

In this thesis we will focus attention on the MHC class II genes, in particular on their exon 2, which encodes for the most variable part of the protein, the PBR. The genomic region that includes class II MHC loci is similar in humans and dogs, and contains several homologous genes that also share similar nomenclature between HLA and DLA. Nearly all mammalian MHC class II regions retain DP, DQ, and DR genes, whereas DP tends to become non-functional as observed in dog DLA, cat FLA, and equine MHCs (Yuhki et al. 2007). Like its human counterpart, DLA-DQB1 is considered the most polymorphic MHC gene.

Variability, evolution and selection forces

The MHC as other multi-gene families have been characterized, in their evolution, by duplication events (bird and death evolutionary mechanism; Nei, Gu, and Sitnikova 1997) from an ancestral proto-MHC that appeared before the emergence of the adaptive immune system (AIS) in the jawed vertebrates (gnathostomes; Flajnik and Kasahara 2010). The hypothesis for the existence of an ancestral MHC is supported by the presence of MHC-like regions in non-vertebrates, and also by the presence of three paralogous regions of MHCs in humans (Martínez-Borra and López-Larrea 2012). Different levels of evolution by duplication have been found in both class I and II MHC proteins (Hughes and Nei 1989; Klein 1987), but the reason for this difference are not well understood. Two hypothesis of “block duplication” –describing one 1R (Ohno 1970) and two rounds 2R (Dehal and Boore 2005) of duplications occurred in the vertebrate genome (Figure 2) - were proposed to explain the MHC paralogous regions found in vertebrates. This mechanism allowed the accumulation of redundant gene copies, which can evolve following selective pressures generating a potent system of defence.

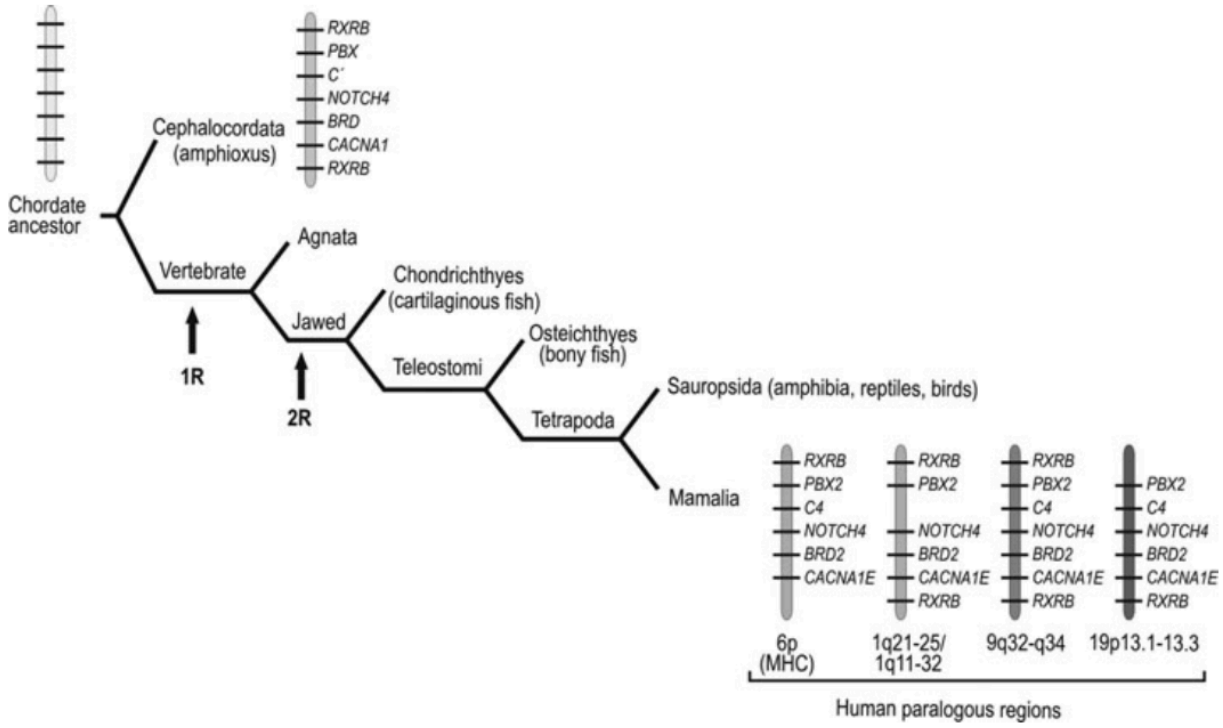


Figure 2

Large scale duplications during vertebrate evolution. The MHC paralogous regions suggest that two large-scaled duplications (indicated with arrows) of the genome took place after the divergence of cephalochordates and vertebrates and before the jawed vertebrate radiation. Highly conserved genes with at least three copies in the MHC paralogous regions are indicated. (From Martínez-Borra and López-Larrea 2012).

The MHC genes act as a mediating interface between organisms and pathogens, and their ability to bind wide arrays of pathogens is ensured by a high allelic sequence variation in the PBR coding region (Potts and Wakeland 1990). Extraordinary nucleotide polymorphism observed at MHC genes has been frequently described as being controlled by balancing selection.

Balancing selection results mainly in the maintenance of large numbers of alleles in populations, and several studies demonstrated that at the PBR more non-synonymous than synonymous substitutions can be found (Sommer 2005), which cannot be explained by a higher mutation rate but likely as a consequence of balancing selection (Bernatchez and Landry 2003).

To proof balancing selection in the sampled generation, different methods such as observed deviations from Hardy–Weinberg equilibrium, Mendelian expectations, or expectations about random associations (Garrigan and Hedrick 2003) could be used. Conversely, when looking at MHC evolution over evolutionary times the most common approach is to examine the ratio of non-synonymous to synonymous substitutions (d_n / d_s) in the coding sequences. Indeed, under neutral

theory, the rate of synonymous nucleotide substitution (d_S) is predicted to be equal ($d_S = d_n$) to the rate of non- synonymous substitution (d_n), since non-synonymous substitutions changing the amino acid composition are likely to be deleterious and thus be targeted by natural selection.

Commonly three main hypotheses have been proposed to explain different directionality of balancing selection: i) the frequency-dependent selection, ii) the fluctuating selection and iii) the heterozygote advantage.

In frequency-dependent selection, individuals bearing common alleles are likely to be more susceptible to diseases, whereas specific or rare alleles can confer resistance (Clarke and Kirby 1966) for a limited period of time. This equilibrium is indeed granted to last until the frequency of rare alleles increases in the host population, forcing parasites to mutate for their survival. It is believed that hosts and parasites co-evolve in time and space, in a continuous competition for survival. Therefore, since both host and parasites fluctuate in numbers and fitness, this lead to a hypothetically endless evolutionary race. This mechanism is called pathogen-mediated evolution.

Environmental factors are important too, but they are not sufficient for the maintenance of such high polymorphism level (Hedrick 1986). In the fluctuating selection hypothesis (Hill 1991), selection is driven by host-pathogens fluctuations not in response to rare-allele frequency, as in the frequency dependent selection, but in response to ecological changes, which alternatively produce different combinations of MHC alleles.

The heterozygote advantage hypothesis is also called over-dominance (Doherty and Zinkernagel 1975). In this hypothesis heterozygote individuals may be favoured over the homozygote ones, because their higher variability allows to cope with a larger number of pathogens (Garrigan and Hedrick 2003).

There is not consensus on which of those hypotheses is more important (Bernatchez and Landry 2003) and furthermore since these selective pressures may operate solely or in concert with other neutral forces (Spurgin and Richardson 2010) it is usually difficult to define which roles they may play in maintaining MHC diversity.

Balancing selection in MHC has also greatly enhanced the persistence of allelic diversity over extremely long time periods. This phenomenon is called the “trans-species polymorphism” (Klein 1987) and is usually observed when MHC alleles among different species are more similar than alleles within species (Garrigan and Hedrick 2003).

Moreover, there is a relation between the high level of MHC polymorphism and sexual selection (Bernatchez and Landry 2003). The hypothesis is that sexual selection can be driven by MHC-dependent mating preferences (Penn and Potts 1999). These preferences are finalized to

obtain a mate pair between individuals who are seeking for good genes to transmit to their progeny. Generally this leads to MHC-disassortative mating preferences where individuals retaining low levels of MHC diversity mate with individuals having higher diversity (Penn and Potts 1999; Ilmonen et al. 2009) and vice versa (Griggio *et al.* 2011), in order to increase the resistance of their progeny against parasites.

This resistance it predicted by both MHC-heterozygote advantage theory and by the "moving target theory", where the new MHC haplotypes generated provide a new target for pathogen to be reached forcing them to evolve for survival. Nonetheless, it is also possible to observe assortative mating preferences, when both partners with good or intermediate immunogenic make up search for the mate, that also has high or intermediate MHC diversity (Tregenza and Wedell 2000; Eizaguirre *et al.* 2009). Such strategy could help to avoid disruption of local adaptations but also leads to fixation of some traits.

Another mechanism linked to MHC and sexual selection is the inbreeding avoidance. This a mechanism by which mating preferences are driven to avoid partners which share the same MHC alleles, mostly because those individuals could likely be related. In sexual selection, indeed, it is a common behaviour for individuals to tend avoiding to mate with relatives or with individuals that share high genetic similarity, because this can ensure to skip the accumulation of deleterious mutations, and an excess of homozygosity in their offspring. MHC-based disassortative mating behaviour is therefore expected to be stronger if species are more exposed to the risk of inbreeding (Tregenza and Wedell 2000; Jamieson *et al.* 2009).

MHC-derived mate preferences have been described in humans (Wedekind et al. 1995; Chaix, Cao, and Donnelly 2008), mice (Egid & Brown 1989 (Ninomiya and Brown 1995) and sticklebacks (Reusch et al. 2001; Milinski *et al.* 2005). Moreover, olfactory-based mate choice has also been shown to be MHC-linked (Jacob et al. 2002; Ilmonen *et al.* 2009; P. a Brennan and Kendrick 2006; Ã, Robson, and Waterhouse 2006). MHC is in fact likely perceived via olfaction or pheromone detection (P. A. Brennan and Zufall 2006), and may influences the pleasantness of the perceived odours (Janeš *et al.* 2010).

MHC in *Canis*

Since more than 90% of synteny is maintained between dog and human genome, many genetic systems determining human diseases are also present in dogs (Lindblad-Toh *et al.* 2005). Together with mice, dogs are considered a model in transplantation biology and a great number of

studies were performed to describe variability at their MHC (Kennedy *et al.* 1998, 2001, 2002; Angles, Kennedy, and Pedersen 2005; Wagner *et al.* 1996; Runstadler, Angles, and Pedersen 2006). All these genetic investigations revealed that DLA system alleles have a great variability and while many alleles are common to almost all dog breeds, haplotype signatures specific to certain breeds were also identified, somethimes showing severely limited variability. Higher differences were identified across than within breeds, a fact that was also supported by high levels of linkage disequilibrium later discovered in dog genome (Lindblad-Toh *et al.* 2005), which determine the drop of allele frequencies and fixation for several genes.

Given the genetic relationships across breeds described above, it is likely that the same risk allele would be carried in multiple breeds (Lindblad-Toh *et al.* 2005). By comparing risk-associated haplotypes in multiple breeds, it should be possible to substantially narrow the region containing the gene under risk. Therefore, many studies focused also to identify possible correlations between dog breeds and genetic factors influencing resistance or susceptibility for specific diseases such as: narcolepsy in Labrador retrievers and Doberman pinschers (Wagner *et al.* 2000), leishmaniasis (Quinnell *et al.* 2003), hypothyroid disease in Doberman pinscher (Kennedy *et al.* 2006), canine transmissible venereal tumor (Murgia *et al.* 2006), necrotizing meningoencephalitis (NME) in Pugs (Greer *et al.* 2010), chronic superficial keratitis in German Shepherd (Jokinen *et al.* 2011).

MHC studies growth in importance also in the management of natural populations, especially in small endangered, fragmented populations such as several natural populations of canids in the world. Some researches focused on describing genetic diversity (Berggren and Seddon 2005; Berggren and Seddon 2008; Arbanasić *et al.* 2013), or monitoring genetic variation after a bottleneck (Seddon and Ellegren 2002, 2004; Marsden *et al.* 2009; Bollmer, Vargas, and Parker 2007; Niskanen *et al.* 2014), and genetic variation in severely endangered populations as in the case of the Mexican wolf (*C. l. baileyi*) (Hedrick, Lee, and Parker 2000), Ethiopian wolves (*C. l. simensis*; Kennedy *et al.* 2011), and African wild dog (*Lycaon pictus*; Marsden *et al.* 2009, 2012).

In the family of Canids despite different events of speciation and different patterns of selection and recombination, a widely diffused trans-species polymorphism in MHC genes has been documented. A portion of variability, thus, is shared with domestic dogs, which thus partially retain the same polymorphisms observed in wolves, especially at DQA and DRB loci (Berggren and Seddon 2005). The larger variety of DLA alleles found in dogs and the high diversity of DLA alleles observed in Asian dogs than in European ones, has been proposed to be consistent with the hypothesis of the Asian origin of dogs from a large founding population (Niskanen *et al.* 2013). However, high DLA polymorphism in dogs could also have been caused by a combination of balancing selection and backcrossing events to wolves (Wayne and Bridgett 2012; Vilà, Seddon,

and Ellegren 2005; Vonholdt *et al.* 2010) also confirmed by findings of ancient admixture traces between dog and wolf (Freedman *et al.* 2014; Koepfli *et al.* 2015). In closely related species, coyote (*Canis latrans*) and red wolf (*Canis rufus*), whose admixed origin has been recently clarified in Vonholdt *et al.* 2011, most of the alleles found in the latter were also present in the former (Hedrick, Lee, and Garrigan 2002).

Different levels of variability have been described among Eurasian wolves. Despite few exceptions - in the Scandinavian population a reduced number of allele per locus was described (Seddon and Ellegren 2004) -, even after several decades of direct persecution, reasonable levels of variation have been maintained in most of the studied populations (Arbanasić *et al.* 2013; Niskanen *et al.* 2014). The majority of these studies were in agreement with the classic theory of balancing selection maintaining polymorphisms, even if in the studied populations different levels of variations at neutral loci were observed after periods of genetic isolation or bottlenecks.

INTRODUCTION TO PAPER I

MHC Variability in an Isolated Wolf Population in Italy

During the last two decades, neutral markers have been frequently employed to quantify the genetic diversity of wild populations and the results from such studies have often provided the basis for management recommendations. However, emerging evidences from MHC studies showed that patterns of variation and divergence in adaptive traits are not always associated with concomitant variation in neutral markers (Aguilar *et al.* 2004; Väli *et al.* 2008).

Within-species genetic variation at MHC loci can either be similar to that at neutral loci or, because of past balancing selection, exceed the neutral variation. Especially in small and endangered populations the basic knowledge of MHC variability can help to corroborate concerns about population viability in a human dominated landscape.

In this article we described variability and traces of selection found at three class II MHC loci in the Italian wolf population. To complete the genetic sight on the general level of variability of this population, we also typed 12 STR loci, 4 Y-linked microsatellites, a melanistic deletion at the β-defensin CB103 gene (involved in black coat color expression) and mitochondrial DNA control region.

PAPER I

Galaverni, Marco, Romolo Caniglia, Elena Fabbri, **Silvana Lapalombella**, and Ettore Randi. 2013. "**MHC Variability in an Isolated Wolf Population in Italy**." Journal of Heredity 104 (5): 601–612. doi:10.1093/jhered/est045.

Authors' Contributions:

Authors' contributions are listed in alphabetical order. Conceived and designed the experiments: RC EF MG ER. Performed the experiments: MG SL. Analysed the data: MG SL. Contributed reagents/ materials/analysis tools: EF RC ER. Drafted the paper: RC MG SL ER. All authors revised the manuscript and gave final approval for publication.

Journal of Heredity 2013:104(5):601–612
doi:10.1093/jhered/est045
Advance Access publication July 24, 2013

MHC Variability in an Isolated Wolf Population in Italy

Marco Galaverni, Romolo Caniglia, Elena Fabbri, Silvana Lapalombella, and Ettore Randi

From the Laboratorio di Genetica, Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), Ozzano dell'Emilia, Bologna, Italy (Galaverni, Caniglia, Fabbri, Lapalombella, and Randi); Department of Evolutionary Biology, University of Bologna, via Selmi 3, 40126 Bologna, Italy (Galaverni and Lapalombella); and the Department 18/Section of Environmental Engineering, Aalborg University, Aalborg, North Jutland, Denmark (Randi).

Address correspondence to Marco Galaverni, Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), Via Cà Fornacetta 9, 40064, Ozzano dell'Emilia, Bologna, Italy, or e-mail: marcogalaverni@iol.it.

Data deposited at Dryad: http://dx.doi.org/10.5061/dryad.15r7f

Abstract

Small, isolated populations may experience increased extinction risk due to reduced genetic variability at important functional genes, thus decreasing the population's adaptive potential. The major histocompatibility complex (MHC), a key immunological gene cluster, usually shows high variability maintained by positive or balancing selection in response to challenges by pathogens. Here we investigated for the first time, the variability of 3 MHC class II genes (DRB1, DQA1, and DQB1) in 94 samples collected from Italian wolves. The Italian wolf population has been long isolated south of the Alps and is presently recovering from a recent bottleneck that decreased the population to less than 100 individuals. Despite the bottleneck, Italian wolves show remarkable MHC variability with 6–9 alleles per locus, including 2 recently described alleles at DRB1. MHC sequences show signatures of historical selective pressures (high d_N/d_S ratio, $\omega > 1.74$) but no evidence of ongoing selection. Variation at the MHC genes and 12 background microsatellite loci were not apparently affected by the recent bottleneck. Although MHC alleles of domestic dog origin were detected in 8 genetically admixed individuals, these alleles were rare or absent in nonadmixed wolves. Thus, despite known hybridization events between domestic dogs and Italian wolves, the Italian wolf population does not appear affected by deep introgression of domestic dog MHC alleles.

Key words: *bottleneck, Canis lupus, hybridization, major histocompatibility complex, MHC phylogenetics, natural selection*

During the past few centuries, wolves (*Canis lupus*) in Italy were threatened by direct human persecution and decline of their natural prey (Breitenmoser 1998). By the late 1960s, less than 100 individuals survived in remote areas of central and southern Apennine (Zimen and Boitani 1975). Reduced wolf genetic diversity at the mitochondrial (mtDNA) and nuclear DNA levels has been attributed to the recent population bottleneck or to historical isolation south of the Alps (Lucchini et al. 2004). Recent studies have found that heterozygosity at autosomal microsatellites (short tandem repeats or STRs) and single-nucleotide polymorphism (SNP) markers in Italian wolves was approximately 25% lower than that in other wolf populations (Fabbri et al. 2007, Sastre et al. 2011) and a unique mtDNA control region (CR) haplotype has also been detected (W14; Randi et al. 2000). Italian wolves are thus genetically distinct from all other *C. lupus* populations worldwide (vonHoldt et al. 2011; their Figure 3). However, erosion of genetic diversity may be offset by hybridization with free-ranging domestic dogs (*Canis lupus familiaris*; Boitani 1984, Randi 2008). Admixture analyses of neutral molecular markers (autosomal STR and mtDNA CR) identified approximately 4–7% hybrid genotypes among wolves in Italy (Randi 2008), some of which also showed anomalous phenotypic traits characteristic of domestic dogs (e.g., vestigial first toes on the hind legs, white nails, black coats; Ciucci et al. 2003; Caniglia et al. 2013a). Similarly, a melanistic β-defensin deletion causing black coats could have been introduced in North American and Italian wolf populations via hybridization with dogs (Anderson et al. 2009; Caniglia et al. 2013a). In contrast, black individuals were rarely observed elsewhere in Europe (Godinho et al. 2011).

Previous studies have examined the genetic structure of wolves in Italy by genotyping putatively neutral molecular markers (Randi et al. 2000; Fabbri et al. 2007; Iacolina et al. 2010; Scandura et al. 2011). However, it is well known that the dynamics of functional genes may be very different and that patterns of genetic diversity at neutral or quantitative trait loci in small populations vary according to the interplay between

drift and natural selection (Ejsmond and Radwan 2011; Aguilar et al. 2004). Some small isolated populations host more genetic diversity than expected from neutral models because of avoidance of active inbreeding (Vilà et al. 2003; Geffen et al. 2011) or due to positive natural selection pressures on functional genes (Bernatchez and Landry 2003, Spurgin and Richardson 2010). In an exemplary case, the San Nicolas Island fox (*Urocyon littoralis dickeyi*) has been identified as an extreme case of genetic monomorphism, measured using neutral markers (Goldstein et al. 1999), but it shows a high degree of variation at the DRB1 and DQB1 loci, suggesting that strong balancing selection can maintain variability at functional regions despite strong bottleneck events (Aguilar et al. 2004). Alternatively, for populations with small effective size, genetic drift can overwhelm selective pressures, leading to decreased genetic diversity in functional regions (Bollmer et al. 2011). For highly fragmented African wild dog populations (*Lycaon pictus*), past declines have led to extremely low variability at 2 major histocompatibility complex (MHC) loci (Marsden et al. 2009), thus exposing this species to increased risk of extinction.

The MHC multigene cluster controls a variety of immune response functions (Klein 1986, Ploegh and Watts 1998). MHC genes are among the most variable in vertebrate genomes, often showing exceptionally high heterozygosity compared with neutral markers. Multiple selection models have been proposed to explain the observed polymorphisms and evolutionary dynamics of the MHC in vertebrates (Bernatchez and Landry 2003; van Oosterhout 2009). However, identifying the processes of pathogen-mediated selection that have shaped the MHC structure in populations is never trivial (Spurgin and Richardson 2010).

Canid MHC studies have primarily focused on domestic dogs. The dog leukocyte antigens cluster (DLA; Angles et al. 2005, Yuhki et al. 2007) includes more than 100 genes, grouped into 3 major subfamilies (classes I, II, and III) according to their structure and function (Wagner et al. 1999; Yuhki et al. 2007). DLA class II genes DQA1, DQB1, and DRB1 were found to be highly polymorphic in multiple canid species (i.e., Wagner et al. 1996; Francino et al. 1997; Kennedy et al. 1998, 1999a, 1999b, 2005; Angles et al. 2005; Runstadler et al. 2006; Fliegner et al. 2008). MHC variability is sometimes limited in inbred dog breeds (Angles et al. 2005), whereas village dogs maintain high variability (Runstadler et al. 2006). Numerous diseases in dogs and in the endangered Ethiopian wolf (*C. simensis*, Kennedy et al. 2011) were associated with specific MHC alleles or with lack of heterozygosity (Quinnell et al. 2003; Kennedy et al. 2006; It et al. 2010; Barber et al. 2011; Jokinen et al. 2011). Correlations between MHC heterozygosity and resistance to pathogens are also described in the highly endangered Mexican wolf (*C. lupus baileyi;* Hedrick et al. 2003). Wild-living gray wolves often display high MHC variation (Kennedy et al. 2007; Arbanasic et al. 2013), with some alleles shared with dogs and coyotes (*C. latrans*). Signatures of balancing selection are observed in wild populations of red (*C. rufus*, Hedrick et al. 2002) and gray wolves (*C. lupus*, Berggren and Seddon 2005, 2008). These findings indicate that even limited variation at the MHC loci can be essential for the survival of species threatened by small population numbers. Nonetheless, bottlenecks, fragmentation, and genetic drift can mask evidence of selection (Seddon and Ellegren 2004).

In this study, we explored the influences of historical bottlenecks and gene introgression on the genetic variability of the Italian wolf population, comparing patterns of polymorphism at the MHC with background STR variation. We predicted several outcomes for this study. In a case of shared demographic history and absence of selection, putatively neutral STR loci and functional MHC class II genes should display equivalent levels of genetic diversity. Alternatively, positive natural selection may have acted to maintain genetic variation at functional MHC loci. Finally, genetic diversity could be introduced into the Italian wolf population via hybridization and introgression with domestic dogs, as may have occurred in the case of the causal mutation of the black coat color (Anderson et al. 2009) in North American wolves. Therefore, we sought for dog-derived MHC alleles in a sample of wild-living Italian wolves composed of apparently purebred wolves, wolf × dog admixed individuals, and putative wolves showing the melanistic β-defensin deletion or anomalous phenotypic traits (e.g., vestigial first toes on the hind legs, black coats) but without any detectable signal of admixture at the autosomal STR loci.

Methods

Samples

In this study, we used 94 DNA samples obtained from wild-living wolves or putative wolf × dog hybrids of both genders sampled in Italy. Samples were collected from 3 categories of individuals that included the following: 1) genetically and phenotypically pure wolves ($n = 65$); 2) wolf × dog admixed individuals ($n = 16$), as detected by admixture analyses of their multilocus STR genotypes (Verardi et al. 2006; Caniglia et al. 2013a); and 3) wolves (or hybrids) showing the melanistic β-defensin deletion or anomalous phenotypic traits (dewclaws, white nails, or black coats; $n = 13$) but that did not show any detectable signal of admixture at their multilocus genotypes (Randi and Lucchini 2002; Ciucci et al. 2003; Caniglia et al. 2013a). Muscle tissue was collected from opportunistically found mortalities, primarily roadkill or poached animals (Caniglia et al. 2013a). Tissue was stored at −20 °C in 10 volumes of 95% ethanol. Additionally, blood was obtained from wounded or live-trapped individuals (Ciucci et al. 2009, Galaverni et al. 2012). DNA was extracted using the Qiagen DNeasy Blood and Tissue Kits (QIAGEN). Phenotypic information was recorded, including the presence of morphological abnormalities (dewclaw, black or darker-than-usual coat color, and white nails). Samples were from 4 geographic regions encompassing the wolf distribution across Italy, including the Alps (A) and northern (nAp), central (cAp) and southern Apennine (sAp).

MtDNA Sequencing and Microsatellite Genotyping

DNA was amplified and sequenced at the 350-bp region of the mtDNA CR, which contains diagnostic mutations for the identification of the Italian wolf haplotype W14 (Randi et al. 2000).

Samples were genotyped at 12 canine autosomal STRs that were selected for their high polymorphism in the Italian wolf population (FH2004, FH2079, FH2088, FH2096, FH2137, CPH2, CPH4, CPH5, CPH8, CPH12, C09.250, and C09.253). These loci yield unique individual genotypes with a probability of identity (PID) equaling 3.2×10^{-10}, and an expected PID among full-sib dyads, $PID_{sibs} = 1.1 \times 10^{-4}$ in the Italian wolf population (Caniglia et al. 2013a). When unknown, the gender was determined by restriction fragment length polymorphism (RFLP) analysis of polymerase chain reaction (PCR)–amplified fragments (PCR–RFLP) of diagnostic ZFX/ZFY sequences. Paternal haplotypes in males were identified by genotyping 4 Y-linked microsatellites: MS34A, MS34B, MS41A, and MS41B (Iacolina et al. 2010). The samples were assayed for a β-defensin melanistic mutation (a 3-bp deletion at the CBD103 gene, also named the *K*-locus) that induces a black coat color in wolves and dogs (Candille et al. 2007). Negative and positive controls were used in each PCR. Locus-specific PCR conditions and additional details on the analyzed loci can be found in Caniglia et al. (2013a). PCR products were analyzed in an automated sequencer ABI 3130XL (Foster City, CA), using the software SEQSCAPE version 2.5 for sequences and GENEMAPPER version 4.0 for microsatellites.

Assignment Tests

The software STRUCTURE version 2.2 (Falush et al. 2003) was used to assign the 94 samples to reference wolves (n = 154) or village dogs (n = 116) selected from a large database of Italian wolf and dog genotypes (the *Canis* database at the Instituto Superiore per la Protezione e la Ricerca Ambientale; Caniglia et al. 2013b). Reference wolves did not show any detectable phenotypic or genetic signal of hybridization. Reference village dog samples were collected from rural areas in Italy (Caniglia et al. 2013a). We ran STRUCTURE with a burn-in period of 10^4 iterations, followed by 5 repetitions of 10^5 iterations, independent of any prior nongenetic information, selecting the "admixture" (each individual may have ancestry in more than 1 parental population) and the "I" models (independent allele frequencies) with the population flag option activated. The optimal number of populations was set at $K = 2$, the value that maximized the posterior probability of the data (according to Randi and Lucchini 2002, Verardi et al. 2006). We then assessed the average proportion of membership (Q_i) of the sampled populations to the inferred clusters. The threshold for the individual assignment was set at $q_i = 0.95$, as determined from the minimum values observed in the reference wolves. Wide credibility interval (CI) of the q_i values could indicate admixture (in absence of missing data; Falush et al. 2003); thus, we also set a threshold for the lower limit of 90% CI as 0.79. Individuals that showed a proportion of membership higher than the threshold were entirely assigned to the wolf cluster as pure wild-type wolves (Wt); individuals showing values of q_i or CI less than 0.95 and 0.79, respectively, were considered admixed (H). Independent of microsatellite-based assignments, samples that showed mtDNA haplotypes different from W14 (Randi et al. 2000) or Y chromosome haplotypes different from those described in the Italian wolf population (Iacolina et al. 2010, Caniglia et al. 2013a) were also considered of hybrid origin and assigned to group H. Individuals genetically assigned to the wolf cluster but showing anomalous phenotypic features were assigned to a third group (Ph).

DLA Genotyping

We amplified the second exon of the MHC class II genes DRB1, DQA1, and DQB1 in the 94 Italian wolf or admixed samples, using primers DRB1F (5' - ccg tcc cca cag cac att tc - 3') and DRB1R (5' - tgt gtc aca cac ctc agc acc a - 3'; Hedrick et al. 2002, after Kennedy et al. 1998); DQAin1 (5' - taa ggt tct ttt ctc cct ct - 3') and DQAin2 (5' - gga cag att cag tga aga ga - 3'); DQB1B (5' - ctc act ggc ccg gct gtc tc - 3') and DQBR2 (5' - cac ctc gcc gct gca acg tg - 3'; Kennedy et al. 2006, after Wagner et al. 1996). Each of these primers is intronic and locus specific. Amplifications were carried out in a 10-µl mix, including 2 µl genomic DNA solution, 1 µl bovine serum albumin (2%), and 0.2 µl of each primer (10 µM) plus 0.25 units Taq, at conditions specific for each primer pair. PCR products were purified with Exo/SAP and sequenced in both directions using BigDye Terminator 1.1, according to the manufacturer's protocol. Sequences were analyzed in an automated sequencer ABI 3130XL with the software SEQSCAPE version 2.5, using the sequences DLA-DRB1*03101 (AF336108.1), DLA-DQA1*014012 (AJ316220.1) and DLA-DQB1*05601 (FM246843.1) as references.

Genetic Variability and Phylogenetics

MHC genotypes were phased in DNASP version 5.10 (Librado and Rozas 2009) using PHASE (Stephens and Donnelly 2003), with the "recombination" model (−MR0) and 1000 iterations after 100 burn-ins. Unlike similar software, PHASE is able to cope with triallelic sites that are commonly found in MHC sequences. When the probability of reconstruction of the alleles was lower than 0.9 and with multiple combinations of alleles being possible, the sample was discarded. The alleles were then matched via BLASTn at the National Center for Biotechnology Information (Johnson et al. 2008) with those available in GenBank for all the species of the genus *Canis*, which were downloaded and aligned in GENEIOUS version 5 (Drummond et al. 2011). We also included the sequences available in the Immuno Polymorphism-MHC Database (IPD; http://www.ebi.ac.uk/ipd/mhc/dla/index.html; Robinson et al. 2010). Sequences that matched along the analyzed regions but showed different accession numbers were grouped and a single name was used, respecting the rules defined in the official International Society for Animal Genetics (ISAG) reports (Kennedy et al. 1999a, 1999b, 2001; Robinson et al. 2003; Ellis et al. 2006). Alleles were accepted if they matched previously described ones; otherwise, if they were observed in homozygosis and in at least 2 different samples, they were considered as potential new alleles and were submitted to the DLA Nomenclature Committee, then to GenBank. Multilocus haplotypes were also reconstructed, following the subtractive method described

in Kennedy et al. (2007). Haplotype reconstruction was then confirmed computationally in PHASE (Berggren and Seddon 2008) by concatenating the gene sequences prior to the phasing step and applying the recombination model with 2 hotspots (–MR2) corresponding to the boundaries between adjacent genes (DRB1/DQA1/DQB1).

MEGA version 5 (Tamura et al. 2011) was used to reconstruct the phylogenetic relationships of all the sequences available for each gene, using a Neighbor-Joining method, with 5000 bootstrap replicates based on the Kimura 2-parameter substitution model (Berggren and Seddon 2008). As outgroups for each gene, we included 2 corresponding MHC sequences of macaque (1 from *Macaca fascicularis* and 1 from *M. mulatta*). When present, gaps were excluded from pairwise comparisons. Similarly, we reconstructed single-locus haplotype networks in NETWORK version 4.6.1, using the median-joining method with values of $\varepsilon = 10$.

For both microsatellites and MHC genes, the number of alleles, allele frequencies (AF) by population and by locus, and the observed (H_o) and expected (H_e) heterozygosities were assessed in GENALEX version 6.4 (Peakall and Smouse 2005). *F*-statistics and departures from Hardy–Weinberg equilibrium (HWE) were evaluated after 1000 permutations in GENALEX and verified in ARLEQUIN version 3.5 (Excoffier et al. 2005). In order to identify the specific sites that were responsible for the largest effects on the latter metrics, we also tested the departure from HWE at every SNP, considering each variable site as a single marker. Linkage disequilibrium (LD) between markers was assessed in GENEPOP version 4.2 (Rousset 2008), Web version (http://genepop.curtin.edu.au), after 1000 iterations. DNASP version 5 was used to compute the number of segregating sites, haplotype (H_d), and nucleotide diversity (π) for each MHC gene. Rather than F_{ST}, we used Jost's *D* (Jost 2008), calculated in SMOGD (Crawford 2010), to overcome the differences in marker types and variability when comparing the levels of differentiation between pure and admixed individuals at MHC and background microsatellites.

Selection and Neutrality Analyses

The average pairwise ratio (d_N/d_S) was calculated in DNASP. In MEGA, we calculated a codon-based test of neutrality with the Nei–Gojobori counting method, in which the significance of the difference between d_N and d_S was assessed for each gene through a 1-tailed *t*-test after 500 bootstrap replicates. However, different selective pressures probably act on specific portions of a gene. Therefore, we evaluated the d_N/d_S ratio on a single-codon basis, as implemented by the software CODEML in PAML (Yang 2007). CODEML was run under the M2a, M3, and M8 models, to identify the codons showing d_N/d_S (ω) values significantly higher than 1—suggesting positive selection—and also for comparison under the models M1a, M0, and M7, respectively. Significance of the model comparison was tested by a Likelihood Ratio Test (Anisimova et al. 2003); single codons were considered to be under positive selection when this probability was higher than 0.95 under all models (M2a, M3, and M8) by both the Bayes Empirical Bayes and the Naïve Empirical Bayes tests (for models M2a and M8).

Average observed heterozygosity at both STR loci and MHC genes was compared with that expected by an Ewens–Watterson statistics of heterozygosity, implemented in BOTTLENECK version 1.2.02 (Cornuet and Luikart 1996), under the following assumptions: 1) an infinite allele mutation model (IAM); and 2) a 2-phase mutation model (TPM) with 90% single-step mutations. In populations where a recent bottleneck occurred, as for the Italian wolf population (Lucchini et al. 2004; Fabbri et al. 2007), allele number (*k*) decreases faster than gene diversity (H_e, or Hardy–Weinberg heterozygosity) at polymorphic loci. This discrepancy leads to an observed gene diversity that is higher than the expected equilibrium gene diversity (H_{eq}), which can be computed from the observed number of alleles under the assumption of a constant-sized population (Cornuet and Luikart 1996). Thus, the test calculates the difference (DH) between the observed and expected heterozygosity values and divides it by the standard deviation (SD) of gene diversity, retrieving the corresponding *P* values after simulating 1000 iterations per locus. The significance of the results was evaluated by a Wilcoxon test and a mode-shift test was also computed.

Results

Genotyping and Sequencing Success

We obtained complete and reliable genotypes at the 12 autosomal STR in all 94 samples. However, 20 samples provided low-quality sequences at 1 or more MHC loci due to poor DNA isolation and/or storage conditions and so were discarded. We therefore obtained reliable sequences at the 3 MHC loci in 74 out of 94 samples (79%), which were retained for subsequent analyses.

Sample Assignment to the Italian Wolf Population

Results from STRUCTURE analyses led to the assignment of 48 (65%) of these samples to the reference wolf cluster, with q_i greater than 0.95 and the lower limit of the 90% CI greater than 0.79. These samples also showed the Italian wolf mtDNA W14 haplotype and the most frequent Y chromosome haplotypes in the Italian wolf population (haplotypes H1 and H2; Iacolina et al. 2010). Although these samples were genetically identified as Italian wolves, they showed either the typical Italian wolf Wt phenotype ($n = 38$) or unusual phenotypic traits, such as black coat, white nails, or dewclaw ($n = 10$, named "Ph"). The other 26 samples (35%) showed both q_i values less than 0.95 and lower limit of 90% CI less than 0.79 and thus they were identified as admixed (and labeled "H," independent of their phenotypes).

MHC Genetic Variability

The 3 MHC loci were polymorphic with 9 (DRB1), 6 (DQA1), and 8 (DQB1) alleles across the 74 genotyped samples (Table 1). Both H_o and H_e were highest at DQB1 and lowest at DQA1. The effective number of alleles (N_e) was largest at DRB1 (Supplementary Table S1). The number of segregating sites was 43 in DRB1, 8 in DQA1,

Table 1 Official names and frequencies (*f*) of alleles found at each locus in the Italian wolf population, with corresponding GenBank names and accession numbers (AN), and the canid populations where they were described to date

			GenBank		Total (*n* = 74)		Wt (*n* = 38)		Ph (*n* = 10)		H (*n* = 26)	
Gene	Nomenclature	Taxa	Name	AN	2*n*	*f*	2*n*	*f*	2*n*	*f*	2*n*	*f*
DRB1												
1	DLA-DRB1*12801	Wit	12801	JX206799	56	0.38	28	0.37	12	0.60	16	0.31
2	DLA-DRB1*092013	Wit	092013	JX206798	40	0.27	28	0.37	6	0.30	6	0.12
3	DLA-DRB1*03601	We,Wa	03601	AF336110.1	21	0.14	11	0.14	1	0.05	9	0.17
4	DLA-DRB1*02001	D	D20	U58684.1	10	0.07	3	0.04	0	0.00	7	0.13
5	DLA-DRB1*03202	Wa,Rw	03901	AF343740.1	6	0.04	3	0.04	1	0.05	2	0.04
6	DLA-DRB1*01501	D	DRB1-W	DQ056281.1	5	0.03	0	0.00	0	0.00	**5**	**0.10**
7	DLA-DRB1*03701	Wa	03701	AF343738.1	5	0.03	2	0.03	0	0.00	3	0.06
8	DLA-DRB1*00101	D	DRB1-U; DRB1-Q	DQ056278.1; DQ056274.1	3	0.02	0	0.00	0	0.00	**3**	**0.06**
9	DLA-DRB1*092011	Wa	09201	AM408904.1	2	0.01	1	0.01	0	0.00	1	0.02
DQA1												
1	DLA-DQA1*005011	We,Wa,D,Wm	DQA3	U44787.1	101	0.68	58	0.76	18	0.90	25	0.48
2	DLA-DQA1*012011	We,Wa,D,C	01201	AF343734.1	21	0.14	11	0.14	1	0.05	9	0.17
3	DLA-DQA1*00401	Wa,D	DQA4	U44788.1	11	0.07	3	0.04	0	0.00	8	0.15
4	DLA-DQA1*00201	We,Wa,D	DQA9	U75455.1	6	0.04	3	0.04	1	0.05	2	0.04
5	DLA-DQA1*00601	We,Wa,D	DQA6	U44790.1	6	0.04	1	0.01	0	0.00	5	0.10
6	DLA-DQA1*00101	We,Wa,D,C,Wm	DQA2	U44786.1	3	0.02	0	0.00	0	0.00	**3**	**0.06**
DQB1												
1	DLA-DQB1*03901	We	03901	AY126651.1	56	0.37	29	0.38	12	0.60	15	0.29
2	DLA-DQB1*00701	D,Wm,We,Wa	DQB4	AF043149.1	45	0.31	29	0.38	6	0.30	10	0.19
3	DLA-DQB1*03501	We,Wa,D	03501	AJ311107.1	22	0.15	12	0.16	1	0.05	9	0.17
4	DLA-DQB1*01303	D,Wm,We,Wa	DQB7	AF043152.1	9	0.06	2	0.03	0	0.00	7	0.13
5	DLA-DQB1*02901	We	02901	AY126648.1	6	0.04	3	0.04	1	0.05	2	0.04
6	DLA-DQB1*00301	D	DQB6	AF043151.1	5	0.03	0	0.00	0	0.00	**5**	**0.10**
7	DLA-DQB1*00201	D	DQB3	AF043148.1	3	0.02	0	0.00	0	0.00	**3**	**0.06**
8	DLA-DQB1*02002	Wa,D	DQB19	AF043164.1	2	0.01	1	0.01	0	0.00	1	0.02

We = European wolf; Wa = North American wolf; Wm = Mexican wolf; Rw = red wolf; D = dog; C = coyote; Wit = Italian wolf, from present study only). *n* indicates the number of individuals, 2*n* the number of chromosomes carrying a given allele, all over the population and by group (Wt = wild type; Ph = atypical phenotype; H = admixed wolves). The H group also includes 3 wolf × dog first-generation hybrids with a known origin. Alleles that are private to a group are highlighted in bold in the corresponding column.

and 39 in DQB1. The average nucleotide diversity was $\pi = 0.049$ in DRB1, 0.008 in DQA1, and 0.048 in DQB1. All alleles matched previously published MHC sequences, with the exception of 2 DRB1 alleles, which were the most frequent in the Italian wolf population. Because each allele was found in homozygosity across multiple individuals, they met the ISAG and DLA nomenclature committee criteria (Kennedy et al. 2001; Ellis et al. 2006). These sequences also matched recently proposed alleles and thus received the official names DLA-DRB1*092013 and DLA-DRB1*12801 (Kennedy L, personal communication) and were submitted to GenBank (accession numbers JX206798, JX206799). DLA-DRB1*092013 differed at a single nucleotide site (a C/A mutation at nucleotide 60 in our alignment) from the already known allele DRB1*092011 found in North American wolves (Kennedy et al. 2007). DLA-DRB1*12801 also showed a single nucleotide difference in relation to Calu-DRB1*13, already described in European wolves (Seddon and Ellegren 2002), with a G/A mutation at site 255 in our alignment. DRB1*092011 was found in 40 wolves from the southern and central Apennine, whereas Calu-DRB1*13 was never found in the Italian wolf samples.

Values of observed heterozygosity were lower than expected at all loci across groups. The average fixation index was not significantly different from 0 ($F_{IS} = 0.033$, $P = 0.19$; exact test in Arlequin). The genotype frequencies were significantly different from that based on HWE at locus DRB1 ($P = 0.03$, Exact Test in Arlequin; $P = 0.004$, chi-square test in GenAlEx). The nucleotides mostly responsible for HWE departures were in positions 60, 96, 156, and 158 ($P < 0.05$, chi-square test in GenAlEx), which also showed significant LD ($P < 0.05$ for nucleotide 60 vs. 96, $P < 0.01$ for all other combinations). Most of them (60, 156, and 158) also had significantly high F_{IS} values ($F_{IS} > 0.2$, $P < 0.05$), similar to 2 additional DRB1 nucleotides (12 and 65). A single nucleotide was out of HWE at DQB1 (Site 155, $P = 0.03$). Allele frequencies at the MHC were variable among the 3 groups (Table 1) and significantly different between H and Wt ($F_{ST} = 0.046$, $P = 0.003$), as well as between H and Ph ($F_{ST} = 0.088$, $P = 0.001$). Moreover, 5 low-frequency alleles were detected only in admixed individuals (Group H). Within groups, departures from HWE were significant only in Wt, mainly at Site 60 of the DRB1 sequence ($P < 0.01$), which corresponds to the private mutation in the DRB1*092013 allele.

Reconstruction of the multilocus MHC haplotypes revealed the presence of 13 combinations of alleles, confirmed both by the subtractive approach and by PHASE (Table 2). The frequency of the 3 most common haplotypes (Nos. 1, 2, and 3 in Table 2) accounted for approximately 80% of the total. Further, 3 (private) haplotypes were present at low frequency only in the H group and 2 in the Wt group (Table 2). Haplotype frequencies were variable among the 3 groups and significantly different between H and Wt (F_{ST} = 0.036, P = 0.008), as well as between H and Ph (F_{ST} = 0.076, P = 0.012). The number of haplotypes was higher in the Apennine (n = 10 in nAp, n = 10 in cAp, and n = 7 in sAp) than in the Alps (n = 3; Figure 1, Supplementary Table S2). However, the observed frequency distribution partially reflects the uneven sample size, as confirmed by the significant correlation between $\log_{(\text{haplotype number})}$ and sample size (R^2 = 0.97). Moreover, wolves sampled in the Apennine also carried haplotypes of presumed dog derivation (particularly in cAp, n = 3).

The reconstructed phylogenies at each of the 3 MHC loci for the species of *Canis* showed that alleles found in the Italian wolf population are dispersed throughout the trees, not clustering in any specific clade (Supplementary Figure S1). The 2 newly described alleles at DRB1 were, respectively, basal (DRB1*12801) and terminal (DRB1*092013) to the closest ones described in previous studies (Supplementary Figure S1a). The network reconstruction confirmed the dispersion of the wolf alleles throughout a relatively unstructured topology (data not shown).

Selection and Neutrality Tests

The average d_N/d_S values were higher than 1 at each locus (Table 3). The $d_N - d_S$ statistics computed in MEGA was significant at all loci ($P < 0.05$) and highest at DQB1. CODEML results indicate that the models accounting for positive selection (M2a, M3, and M8; $\omega > 1$) explained the d_N/d_S values significantly better than the corresponding ones (M1a, M0, and M7) assuming neutral ($\omega = 1$) or negative ($\omega = 0$) selection (Table 4). Model M2a, which includes 3 classes of ω values (0, 1, and estimated from the data), suggested that sites under positive selection (with $\omega > 1$) are 20% at DRB1, 11% at DQA1, and 18% at DQB1, whereas the majority of sites (67%, 89%, and 64%, respectively) are under negative selection. M2a fits the data significantly better than M1a (which only assumes the site classes $\omega_0 = 0$ and $\omega_1 = 1$) at all loci. Model M3 (discrete), which assumes 3 site classes (with ratios inferred from the data), also suggested that the majority of sites (64–89%) are under negative selection (with $\omega_0 < 0.01$) at all loci but that the remaining sites are under positive selection (with ω_1 and ω_2 being higher than 1). M3 fits the data significantly better than M0 (which only assumes a single ω value). Model M8 (which assumes a beta distribution of ω ranging from 0 to 1, plus an extra class estimated from data) also suggests that 20%, 11%, and 18% of sites at DRB1, DQA1, and DQB1, respectively, are under diversifying selection and fits the data significantly better than model M7 (which only assumes a beta distribution of ω between 0 and 1). In particular, M3 was the model that best fit the data at all loci, outperforming other models especially at DRB1 and DQB1. Even applying strict testing criteria, several codons showed signals of positive selection, mostly matching sites predicted as peptide-binding regions (PBRs) in their human homologues (Brown et al. 1988, 1993). Among the loci, DQB1 showed the highest number of sites (n = 11) possibly affected by positive selection, compared with 8 in the more variable DRB1 and 5 at DQA1 (Table 4). These codons included all the sites that have been observed to significantly depart from HWE at DRB1 and DBQ1, except for the synonymous variant differentiating DRB1*092013 from DRB1*092011.

Table 2 Haplotype counts ($2n$) and frequencies (f) across the population and by group

		Total (n = 74)		Wt (n = 38)		Ph (n = 10)		H (n = 26)	
Haplotype	Nomenclature (DRB1 / DQA1 / DQB1)	$2n$	f	$2n$	f	$2n$	f	$2n$	f
1	DRB1*12801/DQA1*005011/DQB1*03901	54	0.36	27	0.36	12	0.60	15	0.29
2	DRB1*092013/DQA1*005011/DQB1*00701	39	0.26	27	0.36	6	0.30	6	0.12
3	DRB1*03601/DQA1*012011/DQB1*03501	21	0.14	11	0.14	1	0.05	9	0.17
4	DRB1*02001/DQA1*00401/DQB1*01303	9	0.06	2	0.03	0	0.00	7	0.13
5	DRB1*03202/DQA1*00201/DQB1*02002	6	0.04	3	0.04	1	0.05	2	0.04
6	DRB1*03701/DQA1*005011/DQB1*00701	5	0.03	2	0.03	0	0.00	3	0.06
7	DRB1*01501/DQA1*00601/DQB1*00301	4	0.03	0	0.00	0	0.00	**4**	**0.08**
8	DRB1*00101/DQA1*00101/DQB1*00201	3	0.02	0	0.00	0	0.00	**3**	**0.06**
9	DRB1*092011/DQA1*00601/DQB1*02002	2	0.01	1	0.01	0	0.00	1	0.02
10	DRB1*12801/DQA1*005011/DQB1*00701	2	0.01	1	0.01	0	0.00	1	0.02
11	DRB1*01501/DQA1*00401/DQB1*00301	1	0.01	0	0.00	0	0.00	**1**	**0.02**
12	DRB1*02001/DQA1*00401/DQB1*03901	1	0.01	**1**	**0.01**	0	0.00	0	0.00
13	DRB1*092013/DQA1*005011/DQB1*03501	1	0.01	**1**	**0.01**	0	0.00	0	0.00

The 2 most common haplotypes in the Italian wolf population include the 2 newfound alleles at DRB1; one of them is also present in a low-frequency combination but is always associated with the most common DQA1 allele. Haplotypes that are private to a group are highlighted in bold in the corresponding column. Haplotypes 3 and 6 have also been reported in North American wolves (Kennedy et al. 2007). Haplotype 8, private to the H group, has been identified as the most common in purebred dogs (Kennedy et al. 2002). Haplotypes 9–13 should be treated with caution, because they have been observed in this study less than twice and never previously reported.

The mean values of observed and expected heterozygosities and the number of alleles per locus at the MHC were higher, although not significantly ($P > 0.05$, t-test), than the ones averaged for the 12 STR markers across groups (Supplementary Table S3). The average fixation index was lower at the MHC than at the STR loci. Both expected and observed heterozygosities was higher at the MHC than at the STRs, and this difference was more marked in the admixed individuals ($P = 0.04$ for H_e; $P = 0.06$ for H_o, 1-tailed t-test) than in the Wt and Ph groups (Supplementary Table S4). The differentiation between genetically pure (Wt + Ph) and admixed wolves (H), as calculated from Jost's D statistics across loci, was higher at the MHC ($D_{est} = 0.123$) than at the STRs ($D_{est} = 0.025$). The Ewens–Watterson statistics (Supplementary Table S5) showed that the heterozygosity levels at the STRs in the population were higher than expected under the IAM (Wilcoxon test, 1 tail for H_o excess $P = 0.01$) but not under the TPM (Wilcoxon test, 1 tail for H_o excess $P = 0.36$), which is the most appropriate model to describe STR mutations (Luikart et al. 1998). Conversely, at the MHC, we did not find any trace of significant excess under both models (Wilcoxon test, 1 tail for H_o excess, $P = 0.12$ under the IAM, and $P = 1.00$ under the TPM). When looking at the allele frequency spectrum, we found a higher proportion of rare alleles (frequency < 0.1) at the MHC than at STRs (0.65 vs. 0.48 across the population, and 0.56 vs. 0.35 in Wt wolves), but once again, these allele frequencies were not significantly different from a normal L-shaped distribution.

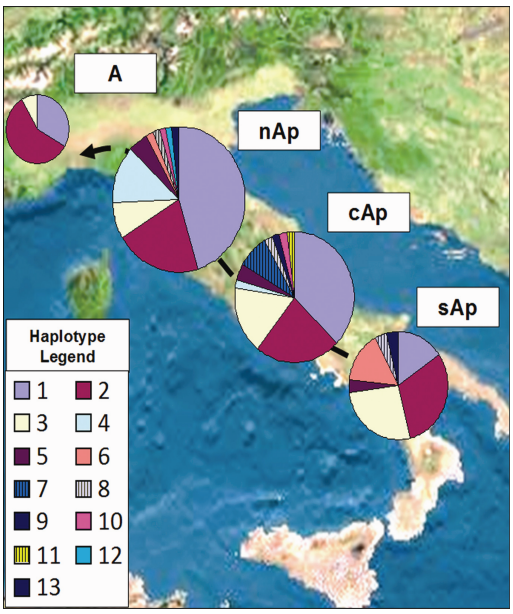


Figure 1. Distribution of MHC haplotypes in the Italian wolf population (which includes individuals showing anomalous phenotypic traits or genetic traces of admixture), split by the geographic origin of the samples: A = Alps ($n = 6$); nAp = Northern Apennine ($n = 31$); cAp = Central Apennine ($n = 24$); sAp = Southern Apennine ($n = 13$). Haplotypes private to admixed individuals (with likely dog origin via hybridization) are indicated by vertical bars on the background (haplotypes 7, 8, and 11).

Discussion

In this study, we investigated the variability of 3 MHC class II loci (DRB1, DQA1, and DQB1) in the Italian wolf population, which has been affected by long-term isolation and a recent population decline (Lucchini et al. 2004). Wolves in Italy are now expanding at a fast-growing pace, recolonizing many areas of their former distribution range (Fabbri et al. 2007). However, hybridization with feral dogs has been repeatedly documented (Randi and Lucchini 2002; Verardi et al. 2006; Caniglia et al. 2013a). Although the use of putatively neutral markers (mtDNA control region and autosomal microsatellites) revealed apparently limited frequency of hybridization and backcrossings (Randi and Lucchini 2002, Verardi et al. 2006), past-generation introgression of functional genetic variants could be underestimated (Anderson et al. 2009). Our findings show that, despite the demographic declines, the Italian wolf population retained high levels of MHC variability, both in number of alleles (from 6 to 9 at DRB1, DQA1, and DQB1 loci) and multilocus haplotypes ($n = 13$), comparable with more abundant populations in stable mutation-drift equilibrium (Seddon and Ellegren 2002). The alleles retained in the Italian wolf population correspond to more than 50% of the alleles described in the overall European or North American wolf populations (Seddon and Ellegren 2002; Kennedy et al. 2007). As a comparison, the highly endangered Mexican wolf population only shows 5 DRB1 alleles (Hedrick et al. 2000), whereas the Swedish population, which probably

Table 3 Distribution of Synonymous (SynDif) and Nonsynonymous differences (NsynDif), their proportions (d_S;d_N) relative to the total number of Synonymous (SynPos) and Nonsynonymous sites (NsynPos), and their ratios (d_N/d_S), as calculated in DnaSP

Locus	SynDif	SynPos	d_S	NsynDif	NsynPos	d_N	d_N/d_S	$d_N - d_S$	P
DRB1	2.74	62.50	0.05	15.34	204.50	0.08	1.74	1.823	0.035
DQA1	0.00	56.50	0.00	3.53	189.50	0.02	N/A	2.441	0.008
DQB1	2.24	64.98	0.04	16.61	202.02	0.09	2.45	3.039	0.001

The probability values (P) of the difference $d_N - d_S$, based on the Nei and Gojobori method implemented in MEGA (Tamura et al. 2011), have been reconstructed after 500 bootstrap replicates.

Table 4 Likelihood of the selection models tested in CodeML

		DRB1	DQA1	DQB1
Models' likelihood				
Model 0	One-ratio	−728.755	−390.862	−707.094
Model 1a	Nearly neutral	−694.025	−392.665	−673.630
Model 2a	Positive selection	**−682.410**	**−385.461**	**−656.114**
Model 3	Discrete	**−670.534**	**−385.459**	**−652.942**
Model 7	Beta	−694.143	−392.999	−673.637
Model 8	Beta & ω>1	**−682.517**	**−385.461**	**−656.135**
Likelihood-ratio test				
M2a versus M1a (4 df)	*D*	23.230	14.409	35.032
	P	9.0E−06	7.4E−04	2.5E−08
M3 versus M0 (2 df)	*D*	116.441	10.804	108.304
	P	3.1E−24	2.9E−02	1.7E−22
M8 versus M7 (2 df)	*D*	23.253	15.077	35.003
	P	8.9E−06	5.3E−04	2.5E−08
Codons under positive selection		**6, 8, 27, 32, 52,** 58, **62, 81**	5, 20, 50, 64, 71, **77**	**6, 8, 23,** 24, **25, 32**, 52, 62, **66, 80, 84**

All the values are indicated as log-likelihood; for each gene, the 3 models that better fit the data are shown in bold. The models have been compared also by a pairwise comparison based on a likelihood-ratio test (rows 9–11), with *D* being twice the log-likelihood difference between the methods; its probability value is based on the chi-square distribution expected with the number of degrees of freedom (df) indicated, corresponding to the differences in free parameters among the models. Codons (row 12) were assumed to be under positive selection only when the probability (*P*) of ω being higher than 1 was greater than 0.95, by both the Bayes Empirical Bayes (for the models M2a and M8) and the Naïve Empirical Bayes (for M2a, M3, and M8) tests. The codons that are likely to correspond to the peptide-binding regions (PBRs) are indicated in bold.

originated from a very limited (<5) number of founders (Seddon and Ellegren 2004), shows 5, 4, and 4 alleles at the same loci, respectively (which decrease to 2 at each locus when excluding nonbreeding immigrants). The neighboring Croatian wolf population hosts an even higher number of alleles per locus (13, 7, and 11, respectively; Arbanasic et al. 2013). Croatian wolves are connected with larger populations in the Balkans, but gene flow with the Italian Apennine wolves during the past century is very unlikely. Wolves in the Italian Alps show approximately half of the observed or expected MHC haplotypes compared with wolves in the Apennine. This is explained by the recent colonization of the Alps by a small (<16) number of founders (Fabbri et al. 2007). Five alleles were detected only in admixed individuals (group H). All of these have previously been described only in dogs (Sarmiento et al. 1990; Sarmiento et al. 1993; Wagner et al. 1996; Kennedy et al. 2005), with the exception of DLA-DQA1*00101, which has been reported in other wolf populations and is shared with dogs and coyotes (Sarmiento et al. 1993; Hedrick et al. 2002). On the other hand, 1 allele (DLA-DRB1*02001) previously only described in dogs (Wagner et al. 1996) was found in the admixed H group and also in 3 Italian Wt wolves. It should be noted that the samples used in this study were not selected randomly and so the number of admixed genotypes is not proportional to their frequency in the Italian wolf population.

The phylogenetic trees did not show any clustering of the alleles found in the Italian wolf population, suggesting ancient coalescence of the MHC haplotypes. This is concordant with the trans-species polymorphisms described for all class II MHC loci (Seddon and Ellegren 2002; van Oosterhout 2009). Long-term consequences of positive selection left detectable molecular signatures at the MHC in the Italian wolves. The high values of the d_N/d_S ratio, significantly better explained by models that allow for positive selection, are clear signatures of strong historical selective pressure driving the MHC variation. Such traces can be observed at both gene-wide and codon-specific levels, with the highest number of codons found to be under selection at the DQB1 locus. Ongoing natural selection might generate deviations from HWE, as observed at the locus DRB1. In particular, 4 nucleotides were responsible for this skew, 3 of which belong to codons under positive selection (CodeML model analyses). All of these are included within 2 of the DRB1 Hyper Variable Domains (Marsden et al. 2009), and 2 nucleotides specifically matched the same putative PBR site (codon 52). The remaining nucleotide corresponds to the single synonymous mutation differentiating 1 of the 2 newly described alleles (DRB1*092013) from its closest sequence, possibly suggesting a recent mutation in the derived state that has not yet reached HWE. However, in the case of ongoing selection, heterozygosity excess should be expected, whereas we observe both an overall heterozygosity deficit and high F_{IS} values at departing nucleotides. Moreover, deviations from HWE could be explained by a Wahlund effect, because at background loci, the pairwise F_{ST} is significant among all geographic locations ($P < 0.05$), except for cAp versus sAp (see also Fabbri et al. 2007). Therefore, the results do not allow us to clearly document traces of ongoing selection.

Balancing selection, leading to an excess of alleles with similar frequencies (i.e., for symmetrical overdominance), and bottlenecks, leading to a loss of rare alleles, are expected to skew allele frequencies toward a uniform distribution and an excess of heterozygosity. Moreover, balancing selection can maintain more variability at functional loci than at neutral markers (Aguilar et al. 2004), although, in small populations, the effects of selection to maintain variability may be overwhelmed by genetic drift (Ejsmond and Radwan 2011).

In our case, the Ewens–Watterson test could not clearly document any deviation from neutrality, either at STRs or at MHC. However, computer simulations carried out by Garrigan and Hedrick (2003) showed that in a population of approximately 100 individuals—the number of individuals estimated to have survived the Italian wolf population bottleneck—the time needed to gain statistical significance in similar tests can be as long as 30 generations. Assuming a generation time of 3 years in wolves, this corresponds to approximately 90 years, thus preceding the population bottleneck that occurred in the 1960s. Therefore, we do not have enough statistical power to detect traces of a bottleneck at the neutral loci nor balancing selection at the MHC. Similar conclusions were drawn by Seddon and Ellegren (2004) in a study on northern European wolf populations. Moreover, the mating schemes of the species could also bias the observed heterozygosity patterns. In wolves, mate choice has only been studied in relation to inbreeding avoidance at STRs (Geffen et al. 2011). Thus, the possible existence of reproductive schemes different from disassortative mating (Galaverni et al. in preparation) could also result in values of heterozygosity lower than those expected under either drift or balancing selection.

Differences in heterozygosity between MHC and STR loci were not significant but more marked in the putatively admixed wolf × dog individuals, suggesting that genetic differentiation among parental populations is higher at the MHC than at STRs, as confirmed by estimates of Jost's *D* between genetically pure and admixed individuals. Similarly, private MHC alleles have been found in the admixed group, whereas wolves showing atypical phenotypic traits, such as dark coat color (Anderson et al. 2009, Randi 2011), only possessed alleles also found in the Wt wolves. These findings confirm the assignment based on the STRs, although a limited number of neutral markers can be inefficient in detecting past hybridization events or gene introgression (Randi 2011). Typically, these events can be reliably identified only up to the second-generation backcrosses using the same panel of STRs (Caniglia et al. 2013a). A single DRB1 allele (DRB1*02001), described only in dogs thus far, was found in 3 Wt wolves and in admixed individuals. Its distribution could be explained in 2 ways: 1) the allele is present both in wolves and dogs but has not been described so far in the former; 2) the allele is dog specific and has been retained in the wolf population after gene introgression not detected at other neutral markers, such as mtDNA CR or STRs. In this study, Y chromosome haplotypes could not be investigated because all 3 Wt individuals were female wolves. Therefore, a greater number of markers will be needed to better discriminate the 2 hypotheses.

Conclusions

Although thoroughly studied, the role of natural selection in shaping MHC variation is still controversial (Bernatchez and Landry 2003, Sutton et al. 2011). When decreased MHC variation is compounded by population isolation, past bottlenecks, and loss of genetic diversity or inbreeding, the long-term population viability for these species is questionable (Radwan et al. 2010). However, direct correlations between MHC variability and fitness traits (e.g., parasite load) have been seldom demonstrated (Wegner et al. 2003; but see Hedrick et al. 2003; Kennedy et al. 2011). Our study describes the variability at important functional genes in the isolated Italian wolf population, dispelling doubts about a dramatic loss of variation that could threaten its future survival. We found traces of historical selection, but we could not detect clear signals of ongoing selection. Results also showed that the assessment of wolf or dog private MHC class II alleles and haplotypes should be used in addition to traditional neutral markers to improve the identification of past-generation hybrids.

Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/

Funding

Italian Ministry of Environment; Emilia-Romagna Region (Dr M. L. Bargossi, Dr M. L. Zanni); University of Bologna (doctoral fellowship to M.G.).

Acknowledgments

We would like to thank Ana Galov (University of Zagreb) and Pedro Silva (Centro de Investigação em Biodiversidade e Recursos Genéticos, Porto) for their useful comments on the manuscript, Laurel Klein Serieys (University of California, Los Angeles) for both comments and language revision, and Hisayo Kishi (University of Kyoto) for her collaboration during laboratory work. We also thank all the numerous collaborators who helped to collect Italian wolf samples, the associated editor, and 3 anonymous referees who positively helped to improve an early version of the manuscript.

References

Aguilar A, Roemer G, Debenham S, Binns M, Garcelon D, Wayne RK. 2004. High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. Proc Natl Acad Sci U S A. 101:3490–3494.

Anderson TM, vonHoldt BM, Candille SI, Musiani M, Greco C, Stahler DR, Smith DW, Padhukasahasram B, Randi E, Leonard JA, et al. 2009. Molecular and evolutionary history of melanism in North American gray wolves. Science. 323:1339–1343.

Angles JM, Kennedy LJ, Pedersen NC. 2005. Frequency and distribution of alleles of canine MHC-II DLA-DQB1, DLA-DQA1 and DLA-DRB1 in 25 representative American Kennel Club breeds. Tissue Antigens. 66:173–184.

Anisimova M, Nielsen R, Yang Z. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics. 164:1229–1236.

Arbanasić H, Huber Đ, Kusak J, Gomerčić T, Hrenović J, Galov A. 2013. Extensive polymorphism and evidence of selection pressure on major histocompatibility complex DLA-DRB1, DQA1 and DQB1 class II genes in Croatian grey wolves. Tissue Antigens. 81:19–27.

Barber RM, Schatzberg SJ, Corneveaux JJ, Allen AN, Porter BF, Pruzin JJ, Platt SR, Kent M, Huentelman MJ. 2011. Identification of risk loci

for necrotizing meningoencephalitis in Pug dogs. J Hered. 102 (Suppl 1):S40–S46.

Berggren KT, Seddon JM. 2005. MHC promoter polymorphism in grey wolves and domestic dogs. Immunogenetics. 57:267–272.

Berggren KT, Seddon JM. 2008. Allelic combinations of promoter and exon 2 in DQB1 in dogs and wolves. J Mol Evol. 67:76–84.

Bernatchez L, Landry C. 2003. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J Evol Biol. 16:363–377.

Boitani L. 1984. Genetic considerations on wolf conservation in Italy. Boll Zool. 51:37–41.

Bollmer JL, Ruder EA, Johnson JA, Eimes JA, Dunn PO. 2011. Drift and selection influence geographic variation at immune loci of prairie chickens. Mol Ecol. 20:4695–4706.

Breitenmoser U. 1998. Large predators in the Alps: the fall and rise of man's competitors. Biol Conserv. 83:279–289.

Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC. 1993. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature. 364:33–39.

Brown JH, Jardetzky T, Saper MA, Samraoui B, Bjorkman PJ, Wiley DC. 1988. A hypothetical model of the foreign antigen binding site of class II histocompatibility molecules. Nature. 332:845–850.

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

Caniglia R, Fabbri E, Greco C, Galaverni M, Manghi L, Boitani L, Sforzi A, Randi E. 2013a. Black coats in an admixed wolf × dog pack: is melanism an indicator of hybridization in wolves? Eur J Wildl Res. Advance Access published February 21, 2013, doi:10.1007/s10344-013-0703-1

Caniglia R, Fabbri E, Mastrogiuseppe L, Randi E. 2013. Who is who? Identification of livestock predators using forensic genetic approaches. Forensic Sci Int Genet. 7:397–404.

Ciucci P, Lucchini V, Boitani L, Randi E. 2003. Dewclaws in wolves as evidence of admixed ancestry with dogs. Canad J Zool. 81:2077–2081.

Ciucci P, Reggioni W, Maiorano L, Boitani L. 2009. Long-distance dispersal of a rescued wolf from the Northern Apennines to the Western Alps. J Wildl Manag. 73:1300–1306.

Cornuet JM, Luikart G. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics. 144:2001–2014.

Crawford NG. 2010. smogd: software for the measurement of genetic diversity. Mol Ecol Resour. 10:556–557.

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, et al. 2011. Geneious v5.4 [Internet]. Geneious, Auckland, New Zealand; [cited 2011 Sept 1]. Available from: http://www.geneious.com/

Ejsmond MJ, Radwan J. 2011. MHC diversity in bottlenecked populations: a simulation model. Conserv Genet. 12:129–137.

Ellis SA, Bontrop RE, Antczak DF, Ballingall K, Davies CJ, Kaufman J, Kennedy LJ, Robinson J, Smith DM, Stear MJ, et al.; ISAG/IUIS-VIC Comparative MHC Nomenclature Committee. 2006. ISAG/IUIS-VIC Comparative MHC Nomenclature Committee report, 2005. Immunogenetics. 57:953–958.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online. 1:47–50.

Fabbri E, Miquel C, Lucchini V, Santini A, Caniglia R, Duchamp C, Weber JM, Lequette B, Marucco F, Boitani L, et al. 2007. From the Apennines to the Alps: colonization genetics of the naturally expanding Italian wolf (*Canis lupus*) population. Mol Ecol. 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.

Fliegner RA, Holloway SA, Lester S, McLure CA, Dawkins RL. 2008. Evaluation of the class II region of the major histocompatibility complex of the greyhound with the genomic matching technique and sequence-based typing. Tissue Antigens. 72:131–136.

Francino O, Amills M, Sánchez A. 1997. Canine Mhc DRB1 genotyping by PCR-RFLP analysis. Anim Genet. 28:41–45.

Galaverni M, Palumbo D, Fabbri E, Caniglia R, Greco C, Randi E. 2012. Monitoring wolves (*Canis lupus*) by non-invasive genetics and camera trapping: a small-scale pilot study. Eur J Wildl Res. 58:47–58.

Garrigan D, Hedrick PW. 2003. Detecting adaptive molecular polymorphism: lessons from the MHC. Evolution. 57:1707–1722.

Geffen E, Kam M, Hefner R, Hersteinsson P, Angerbjörn A, Dalèn L, Fuglei E, Norèn K, Adams JR, Vucetich J, et al. 2011. Kin encounter rate and inbreeding avoidance in canids. Mol Ecol. 20:5348–5358.

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. Mol Ecol. 20:5154–5166.

Goldstein DB, Roemer GW, Smith DA, Reich DE, Bergman A, Wayne RK. 1999. The use of microsatellite variation to infer population structure and demographic history in a natural model system. Genetics. 151:797–801.

Hedrick PW, Lee RN, Buchanan C. 2003. Canine parvovirus enteritis, canine distemper, and major histocompatibility complex genetic variation in Mexican wolves. J Wildl Dis. 39:909–913.

Hedrick PW, Lee RN, Garrigan D. 2002. Major histocompatibility complex variation in red wolves: evidence for common ancestry with coyotes and balancing selection. Mol Ecol. 11:1905–1913.

Hedrick PW, Lee RN, Parker KM. 2000. Major histocompatibility complex (MHC) variation in the endangered Mexican wolf and related canids. Heredity. 85:617–624.

Iacolina L, Scandura M, Gazzola A, Cappai N, Capitani C, Mattioli L, Vercillo F, Apollonio M. 2010. Y-chromosome microsatellite variation in Italian wolves: a contribution to the study of wolf-dog hybridization patterns. Mamm Biol. 75:341–347.

It V, Barrientos L, López Gappa J, Posik D, Díaz S, Golijow C, Giovambattista G. 2010. Association of canine juvenile generalized demodicosis with the dog leukocyte antigen system. Tissue Antigens. 76:67–70.

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. 2008. NCBI BLAST: a better web interface. Nucleic Acids Res. 36(Suppl 2):W5–W9.

Jokinen P, Rusanen EM, Kennedy LJ, Lohi H. 2011. MHC class II risk haplotype associated with canine chronic superficial keratitis in German Shepherd dogs. Vet Immunol Immunopathol. 140:37–41.

Jost L. 2008. $G_{(ST)}$ and its relatives do not measure differentiation. Mol Ecol. 17:4015–4026.

Kennedy LJ, Angles JM, Barnes A, Carmichael LE, Radford AD, Ollier WE, Happ GM. 2007. DLA-DRB1, DQA1, and DQB1 alleles and haplotypes in North American Gray Wolves. J Hered. 98:491–499.

Kennedy LJ, Angles JM, Barnes A, Carter SD, Francino O, Gerlach JA, Happ GM, Ollier WE, Thomson W, Wagner JL. 2001. Nomenclature for factors of the dog major histocompatibility system (DLA), 2000: second report of the ISAG DLA Nomenclature Committee. Anim Genet. 32:193–199.

Kennedy LJ, Barnes A, Happ GM, Quinnell RJ, Bennett D, Angles JM, Day MJ, Carmichael N, Innes JF, Isherwood D, et al. 2002. Extensive interbreed, but minimal intrabreed, variation of DLA class II alleles and haplotypes in dogs. Tissue Antigens. 59:194–204.

Kennedy LJ, Barnes A, Short A, Brown JJ, Lester S, Seddon J, Fleeman L, Francino O, Brkljacic M, Knyazev S, et al. 2007. Canine DLA diversity: 1. New alleles and haplotypes. Tissue Antigens. 42:272–288.

Kennedy LJ, Carter SD, Barnes A, Bell S, Bennett D, Ollier WE, Thomson W. 1998. Nine new dog DLA-DRB1 alleles identified by sequence-based typing. Immunogenetics. 48:296–301.

Kennedy LJ, Carter SD, Barnes A, Bell S, Bennett D, Ollier B, Thomson W. 1999a. DLA-DRB1 polymorphisms in dogs defined by sequence-specific oligonucleotide probes (SSOP). Tissue Antigens. 53:184–189.

Kennedy LJ, Carter SD, Barnes A, Bell S, Bennett D, Ollier B, Thomson W. 1999b. Interbreed variation of DLA-DRB1, DQA1 alleles and haplotypes in the dog. Vet Immunol Immunopathol. 69:101–111.

Kennedy LJ, Huson HJ, Leonard J, Angles JM, Fox LE, Wojciechowski JW, Yuncker C, Happ GM. 2006. Association of hypothyroid disease in Doberman Pinscher dogs with a rare major histocompatibility complex DLA class II haplotype. Tissue Antigens. 67:53–56.

Kennedy LJ, Quarmby S, Fretwell N, Martin AJ, Jones PG, Jones CA, Ollier WE. 2005. High-resolution characterization of the canine DLA-DRB1 locus using reference strand-mediated conformational analysis. J Hered. 96:836–842.

Kennedy LJ, Randall DA, Knobel D, Brown JJ, Fooks AR, Argaw K, Shiferaw F, Ollier WE, Sillero-Zubiri C, Macdonald DW, et al. 2011. Major histocompatibility complex diversity in the endangered Ethiopian wolf (*Canis simensis*). Tissue Antigens. 77:118–125.

Klein J. 1986. Natural history of the major histocompatibility complex. 1st ed. New York: John Wiley & Sons.

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

Lucchini V, Fabbri E, Marucco F, Ricci S, Boitani L, Randi E. 2002. Noninvasive molecular tracking of colonizing wolf (*Canis lupus*) packs in the western Italian Alps. Mol Ecol. 11:857–868.

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

Luikart G, Allendorf FW, Cornuet JM, Sherwin WB. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. J Hered. 89:238–247.

Marsden CD, Mable BK, Woodroffe R, Rasmussen GSA, Cleaveland S, McNutt JW, Emmanuel M, Thomas R, Kennedy LJ. 2009. Highly endangered African wild dogs (*Lycaon pictus*) lack variation at the major histocompatibility complex. J Hered. 100(Suppl 1):S54–S65.

Peakall ROD, Smouse PE. 2005. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes. 6:288–295.

Ploegh H, Watts C. 1998. Antigen recognition. Curr Opin Immunol. 10:57–58.

Quinnell RJ, Kennedy LJ, Barnes A, Courtenay O, Dye C, Garcez LM, Shaw MA, Carter SD, Thomson W, Ollier WE. 2003. Susceptibility to visceral leishmaniasis in the domestic dog is associated with MHC class II polymorphism. Immunogenetics. 55:23–28.

Radwan J, Biedrzycka A, Babik W. 2010. Does reduced MHC diversity decrease viability of vertebrate populations? Biol Conserv. 143:537–544.

Randi E. 2008. Detecting hybridization between wild species and their domesticated relatives. Mol Ecol. 17:285–293.

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

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. Conserv Genet. 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. Conserv Biol. 14:464–473.

Robinson J, Mistry K, McWilliam H, Lopez R, Marsh SG. 2010. IPD - the Immuno Polymorphism Database. Nucleic Acids Res. 38(Suppl 1):D863–D869.

Robinson J, Waller MJ, Parham P, de Groot N, Bontrop R, Kennedy LJ, Stoehr P, Marsh SG. 2003. IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. Nucleic Acids Res. 31:311–314.

Rousset F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol Ecol Resour. 8:103–106.

Runstadler JA, Angles JM, Pedersen NC. 2006. Dog leucocyte antigen class II diversity and relationships among indigenous dogs of the island nations of Indonesia (Bali), Australia and New Guinea. Tissue Antigens. 68:418–426.

Sarmiento UM, DeRose S, Sarmiento JI, Storb R. 1993. Allelic variation in the DQ subregion of the canine major histocompatibility complex: II. DQB. Immunogenetics. 37:148–152.

Sarmiento UM, Sarmiento JI, Storb R. 1990. Allelic variation in the DR subregion of the canine major histocompatibility complex. Immunogenetics. 32:13–19.

Sastre N, Vilà C, Salinas M, Bologov VV, Urios V, Sánchez A, Francino O, Ramírez O. 2011. Signatures of demographic bottlenecks in European wolf populations. Conserv Genet. 12:701–712.

Scandura M, Iacolina L, Capitani C, Gazzola A, Mattioli L, Apollonio M. 2011. Fine-scale genetic structure suggests low levels of short-range gene flow in a wolf population of the Italian Apennines. Eur J Wildl Res. 57:949–958.

Seddon JM, Ellegren H. 2002. MHC class II genes in European wolves: a comparison with dogs. Immunogenetics. 54:490–500.

Seddon JM, Ellegren H. 2004. A temporal analysis shows major histocompatibility complex loci in the Scandinavian wolf population are consistent with neutral evolution. Proc Biol Sci. 271:2283–2291.

Spurgin LG, Richardson DS. 2010. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc Biol Sci. 277:979–988.

Stephens M, Donnelly P. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet. 73:1162–1169.

Sutton JT, Nakagawa S, Robertson BC, Jamieson IG. 2011. Disentangling the roles of natural selection and genetic drift in shaping variation at MHC immunity genes. Mol Ecol. 20:4408–4420.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28:2731–2739.

van Oosterhout C. 2009. A new theory of MHC evolution: beyond selection on the immune genes. Proc Biol Sci. 276:657–665.

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. Mol Ecol. 15:2845–2855.

Vilà C, Sundqvist AK, Flagstad Ø, Seddon J, Björnerfeldt S, Kojola I, Casulli A, Sand H, Wabakken P, Ellegren H. 2003. Rescue of a severely bottlenecked wolf (*Canis lupus*) population by a single immigrant. Proc Biol Sci. 270:91–97.

vonHoldt BM, Pollinger JP, Earl DA, Knowles JC, Boyko AR, Parker H, Geffen E, Pilot M, Jedrzejewski W, Jedrzejewska B, et al. 2011. A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. Genome Res. 21:1294–1305.

Wagner JL, Burnett RC, Storb R. 1999. Organization of the canine major histocompatibility complex: current perspectives. J Hered. 90:35–38.

Wagner JL, Burnett RC, Works JD, Storb R. 1996. Molecular analysis of DLA-DRBB1 polymorphism. Tissue Antigens. 48:554–561.

Wegner KM, Reusch TB, Kalbe M. 2003. Multiple parasites are driving major histocompatibility complex polymorphism in the wild. J Evol Biol. 16:224–232.

Yang Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.

Yuhki N, Beck T, Stephens R, Neelam B, O'Brien SJ. 2007. Comparative genomic structure of human, dog, and cat MHC: HLA, DLA, and FLA. J Hered. 98:390–399.

Zimen E, Boitani L. 1975. Number and distribution of wolves in Italy. Z. Säugetierkunde. 40:102–112.

Received April 11, 2013; First decision June 4, 2013; Accepted June 13, 2013

Corresponding Editor: Robert Wayne

Supplementary materials PAPER I

Galaverni,_SupMat,_Tables.pdf Wolf MHC variability JHered

Table S1. Number of alleles (Na), effective alleles (Ne = 1/(∑ pi2)), Shannon's Information Index (I = -1* ∑ (pi * Ln (pi))), observed (Ho = No. of Hets / N), expected (He = 1 - ∑ pi2) and unbiased expected (UHe = (2N / (2N-1)) * He) heterozygosity, and Fixation Index (F = (He - Ho) / He), where pi is the frequency of the ith allele in the *n* Italian wolves analyzed.

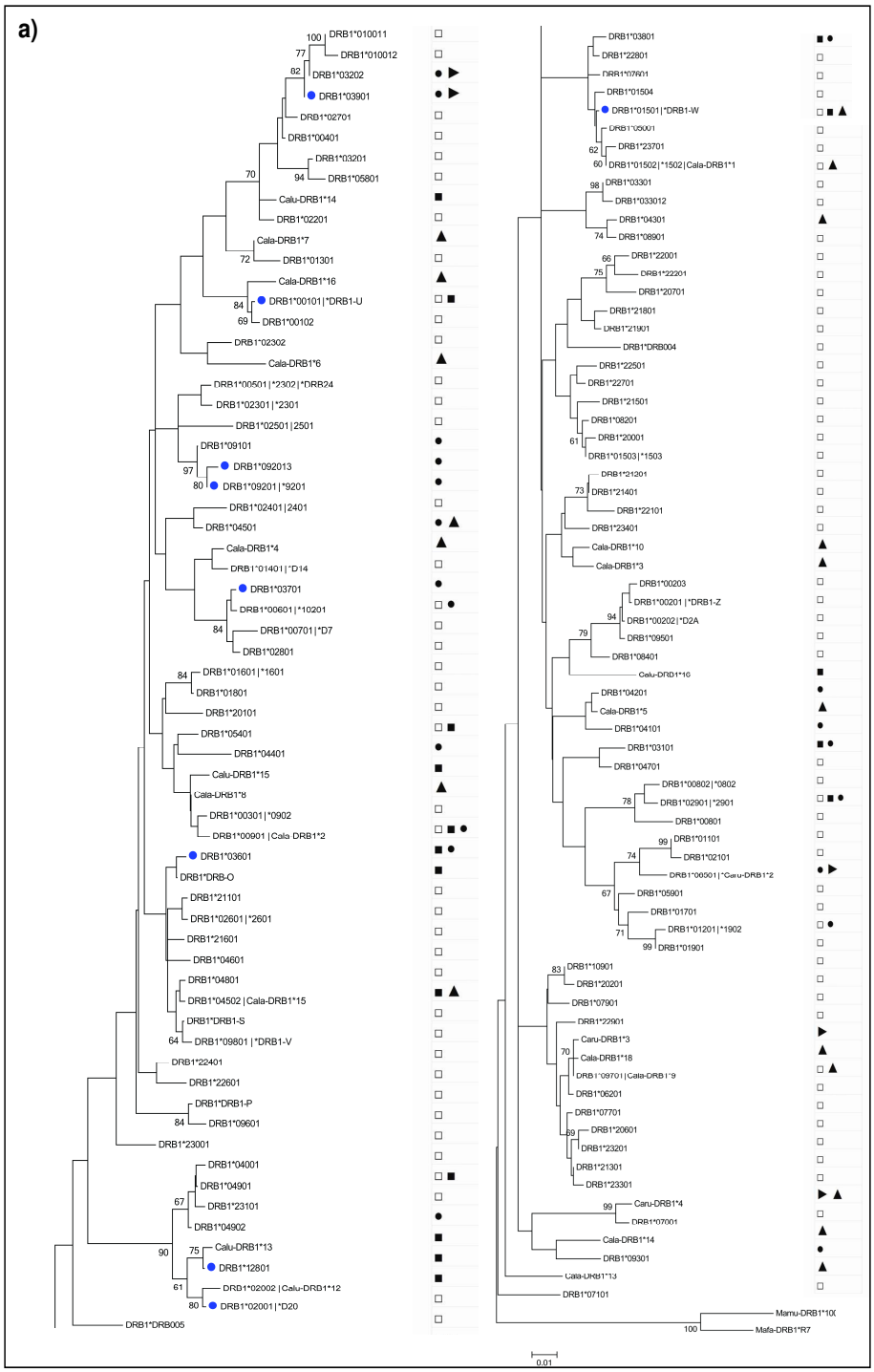
Locus	*n*	Na	Ne	I	Ho	He	UHe	F
DRB1	74	9	4.074	1.677	0.689	0.755	0.760	0.087
DQA1	74	6	2.020	1.070	0.486	0.505	0.508	0.037
DQB1	74	8	3.777	1.565	0.716	0.735	0.740	0.026

Table S2. Haplotype distribution among geographic groups. Haplotypes that have been found only in admixed individuals H are highlighted in bold; haplotypes private to one geographic group are in italics. However, numbers should be treated with caution given the low sample sizes, especially for the Alpine (A) group.

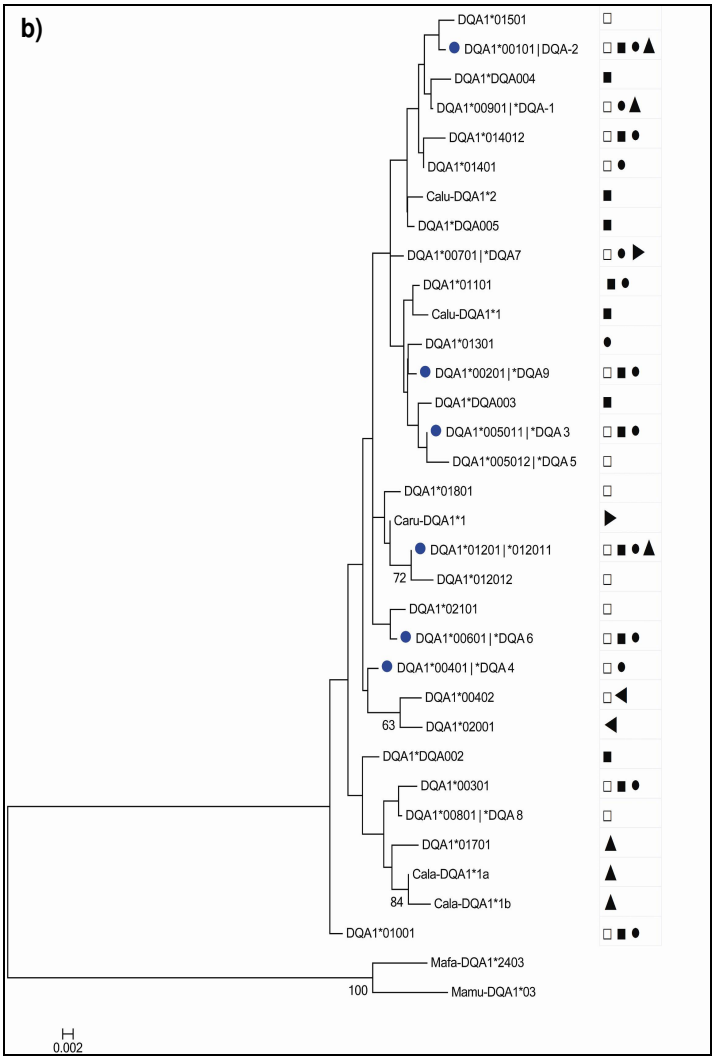
Haplo-type	Nomenclature (DRB1 / DQA1 / DQB1)	A (*n*=6)		nAp (*n*=31)		cAp (*n*=24)		sAp (*n*=13)	
		2*n*	**freq**	**2*n***	**freq**	**2*n***	**freq**	**2*n***	**freq**
1	DRB1*12801 / DQA1*005011 / DQB1*03901	4	0.33	28	0.45	18	0.38	4	0.15
2	DRB1*092013 / DQA1*005011 / DQB1*00701	7	0.58	13	0.21	11	0.23	8	0.31
3	DRB1*03601 / DQA1*012011 / DQB1*03501	1	0.08	5	0.08	8	0.17	7	0.27
4	DRB1*02001 / DQA1*00401 / DQB1*01303	0	0.00	8	0.13	1	0.02	0	0.00
5	DRB1*03202 / DQA1*00201 / DQB1*02002	0	0.00	3	0.05	2	0.04	1	0.04
6	DRB1*03701 / DQA1*005011 / DQB1*00701	0	0.00	1	0.02	0	0.00	4	0.15
7	**DRB1*01501 / DQA1*00601 / DQB1*00301**	**0**	**0.00**	**0**	**0.00**	***4***	***0.08***	**0**	**0.00**
8	**DRB1*00101 / DQA1*00101 / DQB1*00201**	**0**	**0.00**	**1**	**0.02**	**1**	**0.02**	**1**	**0.04**
9	DRB1*092011 / DQA1*00601 / DQB1*02002	0	0.00	0	0.00	1	0.02	1	0.04
10	DRB1*12801 / DQA1*005011 / DQB1*00701	0	0.00	1	0.02	1	0.02	0	0.00
11	**DRB1*01501 / DQA1*00401 / DQB1*00301**	**0**	**0.00**	**0**	**0.00**	***1***	***0.02***	**0**	**0.00**
12	DRB1*02001 / DQA1*00401 / DQB1*03901	0	0.00	*1*	*0.02*	0	0.00	0	0.00
13	DRB1*092013 / DQA1*005011 / DQB1*03501	0	0.00	*1*	*0.02*	0	0.00	0	0.00

Galaverni,_SupMat_Figures.pdf Wolf MHC variability JHered

Figure S1.



Galaverni,_SupMat_Figures.pdf Wolf MHC variability JHered



Galaverni,_SupMat_Figures.pdf Wolf MHC variability JHered

a) DRB1 (left: upper portion; right: basal portion), **b**) DQA1, **c**) DQB1 Neighbour-Joining tree. Italian wolf sequences are indicated by a spot beside the allele name. The taxa where the alleles were described to date are indicated on the right column (Dog □; European wolf ■; American wolf ●; coyote▲; red wolf ▶; Ethiopian wolf ◀). Only bootstrap values above 60 are represented on the trees.

INTRODUCTION TO PAPER II

First Evidence of Hybridization between Golden Jackal (*Canis aureus*) and Domestic Dog (*Canis familiaris*) as Revealed by Genetic Markers

In the PAPER II, we provide the first genetic evidence supporting suspect hybridization between jackals and domestic dogs in Croatia, mostly moved by the presence of anomalous phenotypes recently retrieved in south-eastern Europe. We combined 15 autosomal markers, mt DNA control region, Y STRs, K locus and three class II MHC loci haplotypes to asses putative admixed origins of three wild-living canids showing anomalous phenotypic traits. To better introduce this paper a background description of the species and its principal conservation concerns will follow.

The species: golden jackal (*Canis aureus*)

Biology and distribution

The golden jackal is considered the most typical representative medium-sized canid of the genus *Canis.* Basic coat colour is golden but varies from pale yellow to a dark tawny hue on a seasonal basis. The pelage on the back is often a mixture of black, brown, and white hairs, such that they can appear to have a dark saddle similar to the black-backed jackal (*Canis mesomelas*). The tail is bushy with a tan in the black tip. Legs are relatively long, and feet slender with small pads (Sheldon 1992). The skull of the golden jackal is more similar to that of the coyote and the gray wolf, than that of the black-backed jackal (*C. mesomelas*), side-striped jackal (*C. adustus*), and Ethiopian wolf (*C. simensis*; Clutton-Brock *et al.* 1976). Females can reach 5.8kg of body mass, while males 6.6kg (Moehlman and Hofer 1997).

These carnivores live in a wide variety of habitats, also well tolerating adverse climate varying their diet according to season depending on the availability and distribution of food resources (Macdonald 1979; Moehlman 1983, 1986, 1989; Fuller *et al.* 1989; Moehlman and Hofer 1997). Jackals range from the Sahel Desert to the evergreen forests of Myanmar and Thailand. They

occupy semi-desert, short to medium grasslands and savannahs in Africa, and forested, mangrove, agricultural, rural and semi-urban habitats in India and Bangladesh (Clutton-Brock *et al.* 1976; Poche *et al.* 1987).

As for wolves, the pack is formed by the breeding pair, and sometimes by current and previous litter members (Moehlman 1983, 1986, 1989). Social interactions are complex and also comprehend group vocalisations (Golani and Keller 1975), and a warning call different from normal howls.

Golden jackals are widespread in North Africa and north-east Africa. They have expanded their range from the Arabian Peninsula, eastwards into Turkey, Syria, Iraq, Iran, Central Asia, the entire Indian subcontinent, and then east and south to Sri Lanka, Myanmar, Thailand and parts of Indo-China, and finally into western Europe. In Europe they declined until 30–40 years ago due to human persecution and overhunting and were sometimes treated as pests and eradicated (Spassov 1989). Populations then became fragmented, particularly towards the north-west periphery of the range. Apart from Greece, where the jackals are on the decline and listed as vulnerable in the national Red List (Giannatos *et al.* 2005) the main populations occur in the Balkans (Arnold *et al.* 2012), most notably in Bulgaria, which now supports the largest jackal population in Europe (Genov and Wassilev 1989). Smaller isolated populations survived along the Adriatic coast of Albania, Montenegro and Croatia, and in the Black Sea coast of the Balkan peninsula (Arnold *et al.* 2012). Golden jackals are present in Croatia (Krystufek and Tvrtkovic 1990) in southern Dalmatia and also in the continental eastern part of Croatia (Slavonia). A permanent population was established in northern Dalmatia during the 20th century, and since the 1980s they have been spreading northwest populating Istria (Krystufek and Tvrtkovic 1990). Immigrant individuals from Bulgaria, Romania and Serbia (Banea *et al.* 2012; Selanec *et al.* 2011) are considered the main cause for recent jackal population renewal. Vagrants were also detected towards north-eastern Italy, and Switzerland (Lapini et. al 2009, 2012).

Phylogenetic and genetic findings

Mitochondrial DNA (mtDNA) is extensively used to study phylogeography across many vertebrate species (Simon C. 1991) because of its high mutation rates, small size and conserved arrangement of genes. MtDNA has been used to determine also the golden jackal phylogenetic relation to wolf-like canids. The golden jackal is described to be as the most likely sister taxon of the monophyletic group that includes the grey wolf, the coyote and the Ethiopian Wolf (Wayne et a1997; Bardeleben, Moore, and Wayne 2005). Other jackal species do not form a monophyletic group with the golden

jackal. Recent findings on mtDNA phylogenetic assignments on the cryptic *Canis lups lupaster* (Egyptian jackal) demonstrated that this jackal subspecies has more similarity to *C. lupus* than with *C.aureus*.

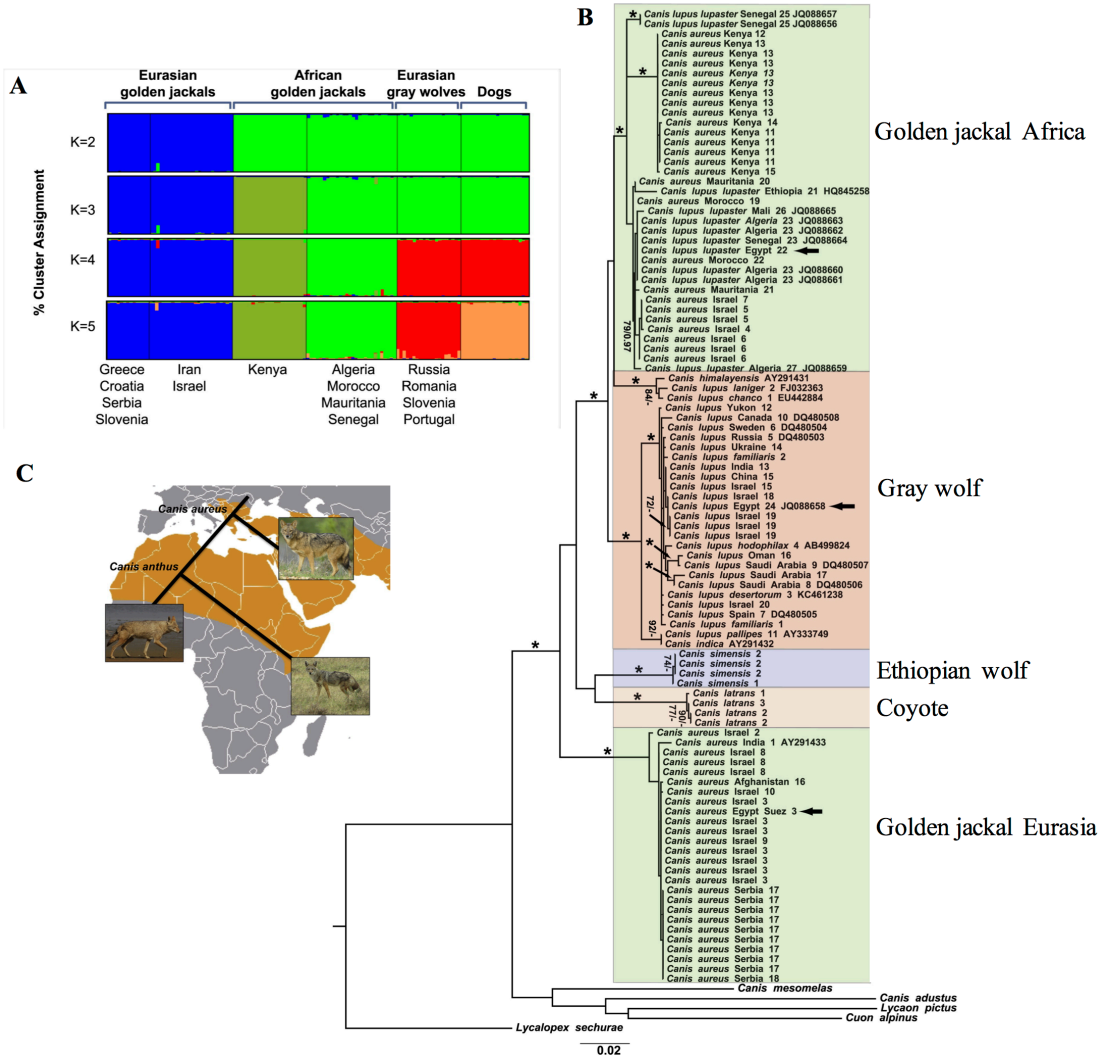


Figure 3

Divergence between golden jackals (*Canis aureus*) from Africa and Eurasia using data from the mitochondrial and nuclear genomes.. A. Maximum-likelihood phylogram of 104 cytochrome b sequences (1,140 bp). Haplotype number is shown next to taxon name and locality. Accession numbers indicate sequences downloaded from GenBank. Haplotypes without accession numbers are novel sequences. Asterisks at nodes indicate bootstrap support ≥ 80% based on maximum-likelihood analyses (500 pseudoreplicates) and ≥ 0.95 posterior probability from Bayesian inference. Canis spp. from Egypt are indicated by thick arrows. Haplotypes labeled as *Canis lupus lupaster* refer to the African wolf. Scale bar indicates the number of substitutions per site. B. Estimated population structure of 128 individuals genotyped for 38 microsatellite loci. Analysis and posterior probability assignments to each cluster assuming two (K = 2) to five (K = 5) genetic clusters were estimated using STRUCTURE. ΔK likelihood was highest for K = 3. The origin of individuals in each cluster is indicated at the bottom of the figure. C. Map of the geographic range of golden jackal based on IUCN distribution, with phylogenetic diagram showing that African and Eurasian golden jackals are genetically distinct and independent lineages, and that African golden jackals likely represent a separate species (*Canis anthus*) Images modified from Koepfli *et al.* (2015).

Therefore, the Egyptian jackal most likely represents an ancient wolf lineage that colonized Africa prior to the radiation of the Holarctic wolf and therefore should be reclassified as the African wolf (Rueness *et al.* 2011).

Golden jackals from Bulgaria, Austria, Serbia and Croatia share the same CR mtDNA haplotype (Fabbri *et al.* 2014; Zachos *et al.* 2009), while high diversity was found among Indian golden jackals populations (Yumnam *et al.* 2015). The cytochrome b (cyt b) gene is another genetic marker used to distinguish differences among mammalian species (Bradley and Baker 2001). Golden jackals from India, the Middle East and Europe formed a monophyletic clade, which was highly divergent from other canid species (cyt b genetic distance > 5%) (Rueness *et al.* 2011; Yumnam *et al.* 2015). Further haplotype networks at the cyt b showed a clear radiation of European and Middle Eastern haplotypes from India (Yumnam *et al.* 2015), suggesting that Europe was colonized a few centuries ago by small numbers of founders, which carried a limited portion of the total genetic diversity of the southern golden jackal source populations.

This pattern has been completed by the recent study by Koepfli *et al.* (2015), that definitely assigned Eurasian and African golden jackals to two different species: *C. aureus* and *C. anthus*, respectively, with the latter clustering closer to wolves (Figure 3). Wolf-like canids have colonized Africa from Eurasia at least five times throughout the Pliocene and Pleistocene (Koepfli *et al.* 2015).

Despite its large distribution, the genetic structure of golden jackal populations still remains unknown. Few studies investigated on genetic variability at microsatellites markers in Israel (Magory Cohen *et al.* 2013) and in Balkan regions (Fabbri *et al.* 2014; Zachos *et al.* 2009). In Israel the species was near-completely eradicated, but a high level of genetic diversity and no evidence of bottleneck was found. On the contrary, there are evidences in population from Serbia, Bulgaria and Croatia of low rates of variability and both historical and current gene flow, probably as consequence of isolation or even a bottleneck experienced by the Dalmatian population (Fabbri *et al.* 2014). Inter-population admixed genotypes were noted from individuals in Bulgaria, and Italy. Admixture in Bulgaria appeared connected with jackals present in Slavonia–Serbia, in agreement with recent expansion of the more abundant and widespread jackal populations in Romania and Bulgaria. Evidences of admixed animals found in Italy were the result of individuals migrating from Dalmatian and Slavonian populations towards north-eastern Italy (Fabbri *et al.* 2014).

Conservation concerns

The golden jackal is listed as a species of Least Concern by the IUCN (Jhala et. al 2008) and is included in CITES Appendix III, which permits a limited trade of pelts. In India, the golden jackal is listed in Schedule III of the Wildlife Protection Act (1972), which ensures a complete ban on hunting. In south-eastern Europe the species is not at risk of extinction, and benefits from partial legal protection. Past decline of golden jackals could have been worsened by the competition with wolves, in conjunction with the waves of wolf expansion in Europe (Genov and Wassilev 1989; Krystufek and Tvrtkovic 1990). However, there is evidence of recent expansion particularly in Bulgaria (Krystufek *et al.* 1997). The reversion of negative demographic trends and climate and habitat changes are probably favouring the expansion in areas from where the species has been absent till recent. Vagrant or reproductive individuals were recently observed in Slovenia, Austria and north-eastern Italy, probably pushed by the ongoing expansion from Bulgaria.

The golden jackal could be considered as a "species requiring no immediate protection" and no species-specific conservation efforts have been undertaken, despite populations throughout its range were likely declining and fragmented. Little quantitative information is available on jackal densities, habitat use, and ranging patterns in relation to food, and aspects of canid diseases in relation to population dynamics need to be better understood.

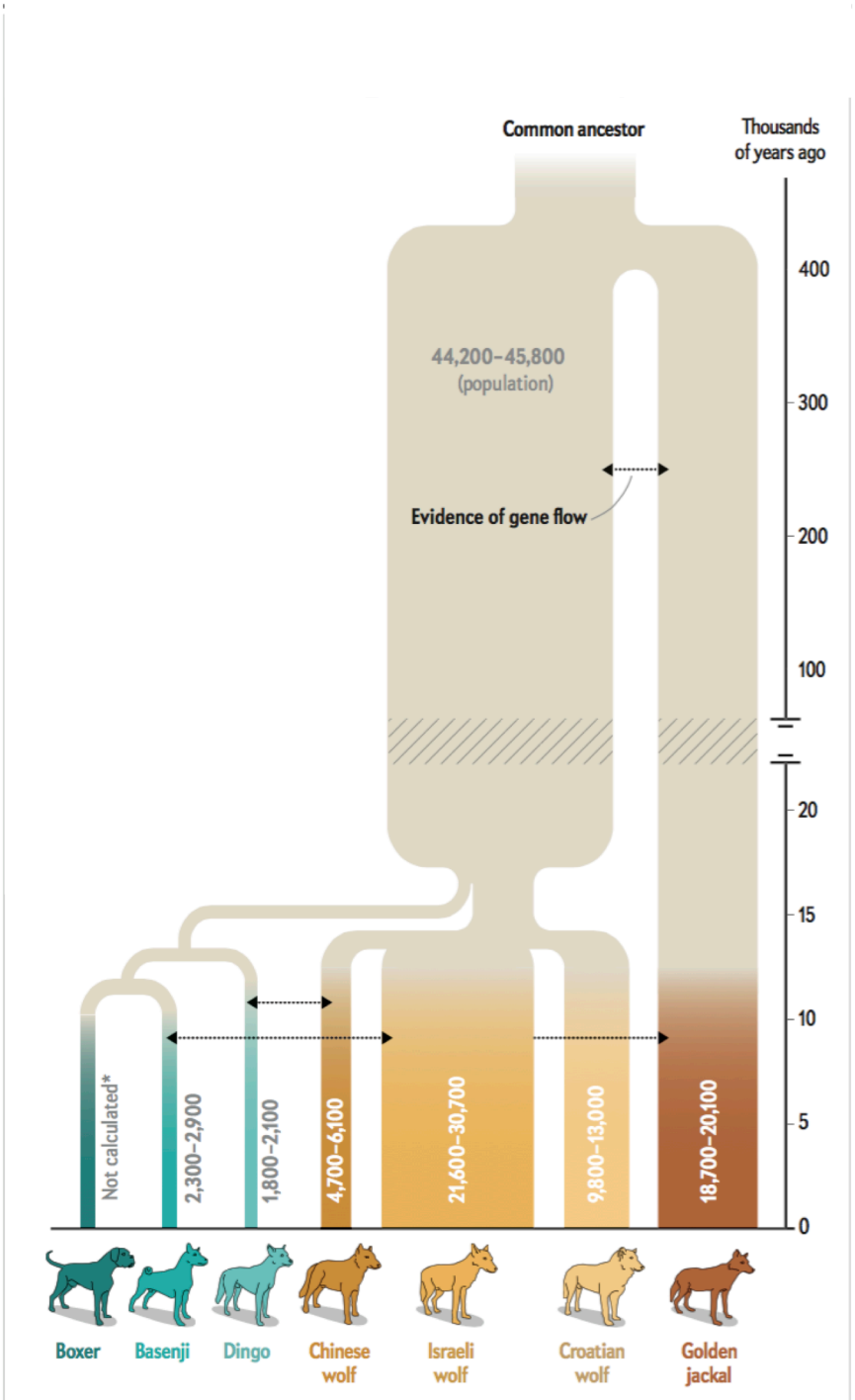


Figure 4

Demographic model of domestication modified from Freedman *et al.* (2014) in Morell *et al.* (2015) showing the population tree best supported by genome-wide sequence divergence from the Boxer reference genome, and the sequenced genomes of two basal dog breeds, three wolves, and a golden jackal. The width of each population branch is proportional to inferred population size. Black horizontal arrows show extensive gene flow between dogs (in green) and wolves (in orange).

Golden jackals' main concerns in conservation are linked to direct persecution and illegal poaching, impacts with vehicles and hybridization. The ranges of golden jackal, gray wolf and domestic dogs largely overlap in Europe and India, but also in southern Asia. Golden jackals could be inter-fertile with other canids species, thus hybridization between jackals, wolves and feral dogs may occur. Recent genome-wide SNP data also revealed evidence of ancient admixture in the histories of Eurasian golden jackals and African golden wolves. Golden jackals from Israel show signals of hybridization with grey wolves, dogs and the African golden wolf (Koepfli *et al.* 2015), evidences supported also by genome-wide findings of ancient hybridization events of Israeli wolves with jackals (Freedman *et al.* 2014).

PAPER II

Galov, Ana, Elena Fabbri, Romolo Caniglia, Haidi Arbanasić, **Silvana Lapalombella**, Tihomir Florijančić, Ivica Bošković, Marco Galaverni, and Ettore Randi. 2015. "**First Evidence of Hybridization between Golden Jackal (*Canis aureus*) and Domestic Dog (*Canis familiaris*) as Revealed by Genetic Markers**." *Royal Society Open Science* 2 (12): 150450. doi:10.1098/rsos.150450. http://rsos.royalsocietypublishing.org/content/2/12/150450.abstract.

Authors' Contributions:

AG, EF, RC, HA, ER designed the study; AG, EF, HA, RC, SL, MG carried out the molecular work and participated in data analysis; TF, IB conceived of the study and collected field data; ER coordinated the study; AG, EF, RC, HA, ER drafted the manuscript. All authors revised the manuscript and gave final approval for publication.

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research

Cite this article: Galov A, Fabbri E, Caniglia R, Arbanasić H, Lapalombella S, Florijančić T, Bošković I, Galaverni M, Randi E. 2015 First evidence of hybridization between golden jackal (*Canis aureus*) and domestic dog (*Canis familiaris*) as revealed by genetic markers. *R. Soc. open sci.* **2**: 150450. http://dx.doi.org/10.1098/rsos.150450

Received: 1 September 2015
Accepted: 3 November 2015

Subject Category:
Biology (whole organism)

Subject Areas:
ecology/evolution/genetics

Keywords:
Canis, interspecific hybridization, gene introgression, major histocompatibility complex, melanism, β-defensin CDB103

Author for correspondence:
Ana Galov
e-mail: anagalov@biol.pmf.hr

Electronic supplementary material is available at http://dx.doi.org/10.1098/rsos.150450 or via http://rsos.royalsocietypublishing.org.

THE ROYAL SOCIETY PUBLISHING

First evidence of hybridization between golden jackal (*Canis aureus*) and domestic dog (*Canis familiaris*) as revealed by genetic markers

Ana Galov[1], Elena Fabbri[2], Romolo Caniglia[2], Haidi Arbanasić[1], Silvana Lapalombella[3], Tihomir Florijančić[4], Ivica Bošković[4], Marco Galaverni[2] and Ettore Randi[2,5]

[1]Department of Biology, Faculty of Science, University of Zagreb, Rooseveltov trg 6, Zagreb 10000, Croatia
[2]Laboratorio di Genetica, Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), Ozzano dell'Emilia (BO) 40064, Italy
[3]Department of Biological, Geological and Environmental Sciences University of Bologna, Via Selmi 3, Bologna 40126, Italy
[4]Department for Hunting, Fishery and Beekeeping, Faculty of Agriculture in Osijek, Josip Juraj Strossmayer University of Osijek, Kralja Petra Svačića 1d, Osijek 31000, Croatia
[5]Department 18/Section of Environmental Engineering, Aalborg University, Sohngårdsholmsvej 57, Aalborg 9000, Denmark

AG, 0000-0002-3709-7286; ER, 0000-0002-4882-6595

Interspecific hybridization is relatively frequent in nature and numerous cases of hybridization between wild canids and domestic dogs have been recorded. However, hybrids between golden jackals (*Canis aureus*) and other canids have not been described before. In this study, we combined the use of biparental (15 autosomal microsatellites and three major histocompatibility complex (MHC) loci) and uniparental (mtDNA control region and a Y-linked *Zfy* intron) genetic markers to assess the admixed origin of three wild-living canids showing anomalous phenotypic traits. Results indicated that these canids were hybrids between golden jackals and domestic dogs. One of them was a backcross to jackal and another one was a backcross to dog, confirming that

golden jackal–domestic dog hybrids are fertile. The uniparental markers showed that the direction of hybridization, namely females of the wild species hybridizing with male domestic dogs, was common to most cases of canid hybridization. A melanistic 3bp-deletion at the *K* locus (β-defensin CDB103 gene), that was absent in reference golden jackal samples, but was found in a backcross to jackal with anomalous black coat, suggested its introgression from dogs via hybridization. Moreover, we demonstrated that MHC sequences, although rarely used as markers of hybridization, can be also suitable for the identification of hybrids, as long as haplotypes are exclusive for the parental species.

1. Introduction

Interspecific hybridization is relatively frequent in nature, occurring not only in plants but also in animals, where at least 10% of species are involved in admixture and potential introgression [1]. Interspecific hybridization in *Canis* has been described in a number of studies in North America (e.g. coyote–grey wolf [2]). Moreover, cases of anthropogenic hybridization between wild canids and widespread free-ranging domestic dogs are particularly alarming, because they may threaten the survival of endangered species (e.g. the Ethiopian wolf, *Canis simensis* [3]), or may deeply change the genetic make-up of wild populations in human-dominated landscapes (e.g. grey wolf [4–6]).

The golden jackal (*Canis aureus*) is a medium-sized species, currently distributed in northern and northeastern Africa, southeastern Europe and large parts of southern Asia [7]. European golden jackals were first reported in 1491 in the coastal region of southern Dalmatia, where they still occur. After suffering a severe decline in the first half of the twentieth century, the European population has recovered and has been expanding since the early 1980s, especially in the Balkan regions [8].

To our knowledge, there are no published cases of recent hybridization between golden jackals and any other canid species in nature, although genome-wide traces of ancestral admixture with wolves have been recently documented [9], and the assignment of the African *Canis aureus lupaster* to the wolf clade cannot rule out the hypothesis of ancient hybridization events [10,11]. The only two documented cases of present-day hybrids were questionable, since they were only based on morphological measurements: in Romania a putative jackal–wolf hybrid was shot in 1942; in Hungary, the skull of a suspected jackal–dog was discovered in 1983, but erroneously reported as the first jackal shot in the country after 41 years [12].

Molecular techniques are routinely used for the identification of closely related species and their hybrids within the first two to three generations of admixture [6,13]. The combined use of maternal and paternal markers may also reveal the hybridization direction [6,14–16]. The mitochondrial DNA control region (mtDNA CR) can identify the maternal ancestry in putative hybrids [15,17]. A single mtDNA CR haplotype has been found so far in European golden jackal populations [18,19], facilitating the identification of golden jackal–dog hybrids through the maternal line. Paternal ancestry of hybrids may be revealed by species-specific Y chromosome markers developed to discriminate between golden jackal and domestic dog males, based on an insertion found in a dog *Zfy* intron, but not in golden jackal [20]. In addition, putatively neutral biparental markers such as microsatellite loci (STR), originally developed for the domestic dog, can efficiently cross-amplify and identify golden jackal genotypes [19,21], thus providing estimates of the proportion of neutral admixture inherited from the mixing parental populations (e.g. [5,17,22]).

Functional markers could further contribute to the investigation of hybridization and introgression. The major histocompatibility complex (MHC) is a multigene family that is commonly used in adaptive variation studies [23]. MHC genes encode cell-surface glycoproteins that bind and present antigens to T cells, which trigger an appropriate immune response, thus playing a pivotal role in the vertebrate immune system. However, to our knowledge, MHC genes have been used as markers to detect hybridization in only two vertebrate species: a case of hybridization between Iberian ibex and domestic goat [24], and a study of introgression of dog MHC alleles in wild-living Italian wolves [25]. Other functional mutations that could be used in determining phenotypical variation have been discovered in recent genomic studies [26], such as a dominant three-nucleotide deletion in the β-defensin CBD103 gene (the *K* locus) correlated to black coat colour in canids, which could have been introduced from dogs into wild-living wolves in North America [27] and in Italy [16] via hybridization (hereafter referred to as ‘melanistic deletion’).

Thus, the primary objective of this study was to document the potential occurrence and directionality of golden jackal–dog hybridization in an expanding jackal population, using combined genetic analyses

of 15 microsatellite loci, mtDNA control region and a Y chromosome marker. Our secondary objectives were to assess the presence of a functional melanistic deletion at the β-defensin CBD103 gene in an individual with black coat coloration and to test the applicability of coding markers such as MHC genes to detect hybridization between closely related species, such as the dog and the golden jackal.

2. Material and methods

2.1. Putative golden jackal–dog hybrid samples

We collected muscle tissue samples from three putative golden jackal–dog hybrids legally harvested by hunters in Croatia. The putative hybrids were initially identified on the basis of their unusual morphological traits. The individual S21 (figure 1) was an adult female showing light coat colour, digital pad depigmentation and atypical long ears with rounded tip (whereas golden jackals have shorter triangle-shaped ears). However, the digital pads of the middle fingers were partially joined as commonly occurs in golden jackals, but not in dogs [12] (I. Bošković 2012, unpublished data). The individual S22 (figure 2) was a juvenile male found and shot together with female S21, probably representing a mother–son pair. It displayed a dog-like morphology, particularly similar to the Istrian shorthaired hound breed (very short hair on the head, white coat colour with sparse patches of light brown, dewclaws on hind legs). However, the animal's tail was shorter and thicker than typical for this dog breed and more similar to that of a golden jackal. Thus, we hypothesized it could have originated through a backcross between the putative hybrid female S21 and a male Istrian shorthaired hound dog. The third individual (60c) (figure 3) was a male that exhibited black coat coloration, atypical for golden jackals, and other dog characteristics such as ears with rounded tip. However, the digital pads on the middle fingers of its forelimbs were partially joined, as in golden jackals.

Tissue samples were stored in 96% ethanol at −20°C prior to analyses. DNA was extracted using a Wizard Genomic DNA Purification kit (Promega, USA).

2.2. Reference samples

To correctly identify the origin and ancestry of the putative golden jackal–dog hybrids, we used as reference source populations 50 jackal samples from Croatia [19] and 51 mixed breed dog samples from Croatia, previously genotyped at STR markers. Reference jackal samples were legally shot or road-killed in Croatia.

2.3. Mitochondrial DNA analysis

We amplified the hypervariable left domain of the mtDNA CR using primers L-Pro [28] and H-576 [29] (electronic supplementary material, table S1). The polymerase chain reactions (PCRs) were carried out in 30 µl containing 1× Qiagen HotStarTaq Master Mix (Qiagen HotStarTaq Master Mix Kit, Qiagen, USA), 0.2 µM of forward and reverse primer and 3 µl template DNA. Cycling conditions were the following: 95°C for 15 min, 35 cycles of 40 s at 94°C, 50 s at 55°C, 1 min at 72°C, and 10 min final extension at 72°C. The amplification products were purified (Wizard SV Gel and PCR Clean-Up System, Promega) and sequenced using L-Pro primer. Sequences were aligned in BioEdit [30] with the only golden jackal haplotype found so far [18,19] and 12 dog haplotypes found among mixed breed dogs from Croatia [31].

2.4. Y chromosome analysis

We analysed the two male samples S22 and 60c using a PCR-based Y chromosome marker method in which the dog DNA template produces two amplicons, whereas the golden jackal template produces only one [20] (electronic supplementary material, table S1).

2.5. Autosomal microsatellite loci analysis

We genotyped the reference populations and the putative hybrid samples at 15 unlinked autosomal canine STRs using the same procedure described in Fabbri *et al*. [19] (electronic supplementary material,

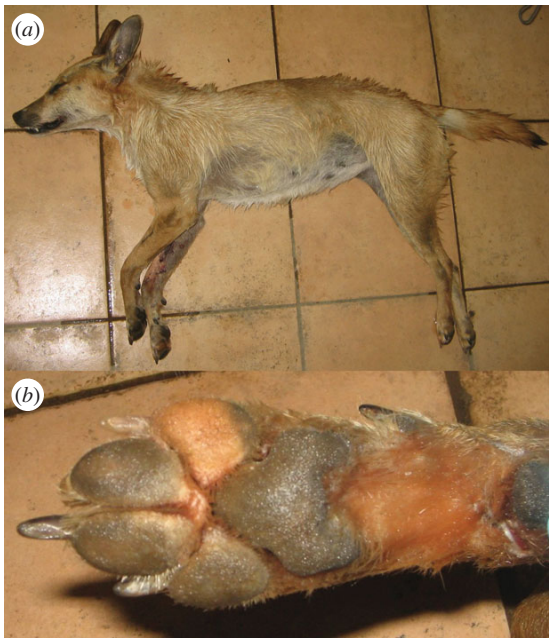


Figure 1. Female golden jackal–dog hybrid (S21) (*a*) and its forelimb with notable digital pad depigmentation (dog characteristic) and partially joined digital pads of the middle fingers (golden jackal characteristic) (*b*).

Figure 2. Male golden jackal–dog hybrid (S22) (*a*) and its hind leg with dewclaw (*b*).

table S2). The average number of alleles (N_a) and private alleles (N_p), and the observed and expected heterozygosity (H_o, H_e) were estimated using GENALEX v. 6.5 [32,33] as measures of genetic diversity. Exact tests for Hardy–Weinberg equilibrium were computed using the Guo and Thompsons' Markov chain method [34] as implemented in the software GENPOP v. 4.00 [35]. The sequential Bonferroni correction test for multiple comparisons was used to adjust significance levels [36]. The 15-STR multilocus genotypes of the reference jackals and dogs, and of the three putative hybrids were used to distinguish species and to detect putative admixed individuals and their ancestry through two

different methodologies: (i) a multivariate analysis: principal coordinates analysis (PCoA) of individual STR genotypes implemented in GENALEX [37]; (ii) a Bayesian clustering procedure implemented in STRUCTURE v. 2.3.4 [38–40], which estimates the admixture proportion of each individual genotype.

We used the Admixture model with independent allele frequencies (*I*-model) running five replicates of K from 1 to 5 using 5×10^5 iterations of MCMC following a burn-in period of 5×10^4 iterations. The optimal number of populations K was determined according to Evanno *et al*. [41] independently from any prior non-genetic information (option usepopinfo not active). For each group, we assessed the average proportion of membership (Q_i) to each different clusters, and individual assignment was consequently based on the proportions of membership (q_i) estimated for every single individual. Based on these first STRUCTURE results, admixture analyses were performed again assuming two reference groups (jackal and dog) for the assignment of the putative jackal–dog hybrids (PHy). STRUCTURE was run with $K = 2$, with the option 'usepopinfo' activated or not. In the former case, we assumed that reference jackals 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 software NEWHYBRIDS [42] was then used to compute the posterior probability for each genotype to belong to each of the six following classes: jackal (J) and dog (D) parentals, F_1 and F_2, backcrosses of F_1 with dogs (BC1D) and with jackals (BC1J). Posterior distributions were evaluated after 10^5 iterations of the Monte Carlo Markov chains, following a burn-in period of 10^4 iterations, without any individual or allele frequency prior information, with 'Jeffreys-like' or 'Uniform' priors for mixing both proportions and allele frequencies.

In addition, we used HYBRIDLAB [43] to evaluate the power of the 15 STRs to correctly detect *a priori* known parentals, hybrids and backcrosses. We used the 50 reference golden jackals and the 51 reference dogs to simulate 50 genotypes for each of the following classes: first and second generation hybrids (F_1, F_2), first and second generation backcrosses with golden jackal and dog (BC1J, BC1D, BC2J, BC2D). The simulated genotypes were analysed in STRUCTURE and NEWHYBRIDS using the run parameters described before ('Admixture' and the 'I' models, without any prior population information).

2.6. *K* locus analysis and major histocompatibility complex analyses

Since one putative hybrid showed a black coat (figure 3), we assessed the presence of the functional melanistic deletion at the β-defensin CBD103 gene (corresponding to the *K* locus), which determines black coat colours in dogs and wolves [16,27,44], following the amplification protocol described in Caniglia *et al*. [16] (electronic supplementary material, table S1).

The 50 golden jackal samples and the three putative hybrid samples were further analysed for MHC DLA-DRB1, DQA1 and DQB1 class II genes using cloning/sequencing method. The primers used to amplify exon 2 were: for DLA-DRB1, forward DRBF [45] and reverse DRB1R [46]; for DLA-DQA1, forward DQAin1 and reverse DQAIn2 [47]; for DLA-DQB1, forward DQB1BT7 [47] and reverse DQBR3, ACCTGGGTGGGGAGCCCG (primer designed in this study based on the sequence published in Wagner *et al*. [48]) (electronic supplementary material, table S1). Amplifications were carried out by PCR in a total volume of 25 µl containing 150–250 ng of genomic DNA, 1× QIAGEN HotStarTaq Master Mix (Qiagen, Hilden, Germany) (consisting of 1× PCR buffer, 200 µM of each dNTP and 2.5 units HotStartTaq DNA polymerase) and 0.2 µM of each primer. A negative control containing no DNA template was included in each amplification run to detect any contamination. All amplifications were performed using a standard touchdown PCR protocol as described in Kennedy *et al*. [49]. PCR products were visualized on 1% agarose gels stained with SYBR Safe DNA gel stain (Invitrogen) and purified by the Wizard SV Gel and PCR Purification Clean-Up System (Promega). Sequencing for typing was performed using the same primers as for PCR in reverse direction for all three loci. Confirmation of new alleles was performed by sequencing in both directions and by further DNA cloning. The PCR products were ligated into vectors and transformed into bacteria using the pGEM-T Vector System II (Promega). Plasmid DNA from 8 to 12 positive clones per individual was isolated using the Promega Wizard Plus SV Miniprep DNA Purification System, and inserts were sequenced using the PCR primers described above. Sequence processing and analysis were performed with BIOEDIT [30]. To identify alleles in heterozygous animals, we used the Applied Biosystems SeqScape® software, which is designed for analysis based on a locus-specific allele reference library, and previously described canid alleles that we obtained from the Immuno Polymorphism (IPD)—MHC Database (L. J. Kennedy 2013, personal communication), as described in Arbanasić *et al*. [50]. Three-locus haplotypes were unambiguously identified in a sequential process [51] and confirmed by ARLEQUIN v. 3.11 [52].

Figure 3. Male golden jackal–dog hybrid (60c) with black coat coloration (*a*) and ears with rounded tip (*b*) (dog characteristics), and forelimbs with partially joined digital pads of the middle fingers (golden jackal characteristics) (*c*).

3. Results

3.1. Mitochondrial DNA

We obtained fragments of 550 bp of the mtDNA CR from the three putative hybrid samples, which all carried sequences identical to the reference golden jackal mtDNA CR haplotype [18,19] (GenBank accession no. KF588364), suggesting their golden jackal maternal ancestry.

3.2. Y chromosome marker

Both male samples (S22 and 60c) produced two amplicons of the *Zfy* intron, characteristic of the domestic dog, suggesting their dog paternal ancestry.

3.3. Autosomal microsatellite loci

All the 15 STR loci were polymorphic in both reference populations, showing from three to 19 alleles in dog and from three to 10 alleles in jackal samples and an average number of alleles per locus of 8.2 (s.e. 0.98) and 4.3 (s.e. 0.50), respectively (electronic supplementary material, table S3). As expected, dogs showed an observed and expected heterozygosity higher than jackals and an average number of private alleles of 4.67 versus 1.07.

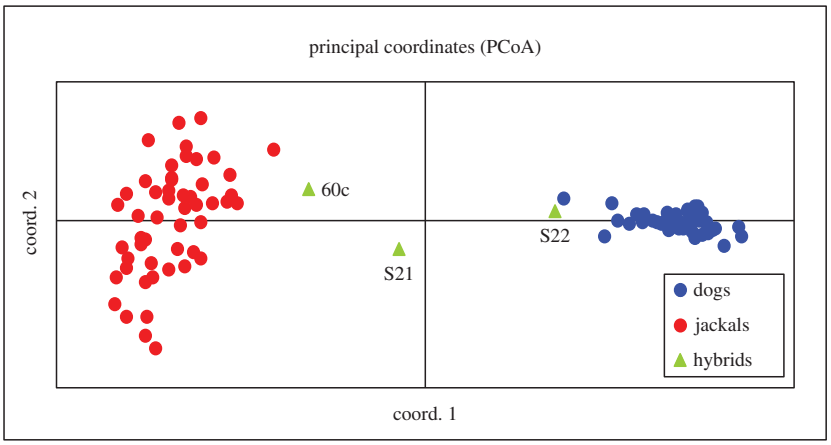


Figure 4. Principal coordinates analysis obtained by GenAlEx. The two principal axes (PC-I and PC-II) cumulatively explain 35.8% of the total genetic diversity. Blue dots represent the Croatian dog references, red dots the Croatian golden jackal references and triangles the golden jackal–dog hybrids: S21, S22 and 60c.

The PCoA results are shown in figure 4, where the individual scores were plotted onto the two principal axes (PC-I and PC-II), which cumulatively explain 35.8% of the total genetic diversity. The PCoA split jackals and dogs into two clearly separate clusters, with all dogs and jackals correctly identified by their genotypes. The only exception is three putative hybrids that showed intermediate positions between the two clusters, suggesting an admixed origin.

At $K = 2$ (the optimal number of genetic clusters), results from STRUCTURE admixture analyses showed that all dogs were assigned to a single cluster with an average membership proportion $Q_D = 0.998$ and individual proportions of admixture q_D ranging from 0.984 to 0.999. Jackals were assigned to the other cluster with $Q_j = 0.999$ and a q_J ranging from 0.994 to 0.999. The three putative hybrids S21, S22 and 60c showed $q_J = 0.588$ (90% confidence intervals CI: 0.396–0.770), 0.227 (0.063–0.411) and 0.849 (0.687–1.000), respectively (figure 5 and table 1).

When we assumed that reference jackals 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), results obtained from five STRUCTURE runs using $K = 2$, PopINFO and 'I' model were concordant with those obtained without any prior information (table 1).

Results of admixture analyses showed that 97.34% of the simulated admixed genotypes can be correctly identified as admixed at threshold $q_i = 0.980$ (after [53], we used as a threshold the minimum q_i value observed in reference populations using only real data: 0.984). All the F_1, F_2, BC1J and BC1D were correctly identified as admixed using both STRUCTURE models (Admixture and PopInfo). Only 6% of BC2J and 10% of BC2D genotypes showed a $q_i > 0.980$ (to jackal cluster and to dog cluster, respectively). Using the Admixture model with selection Flag to reference populations and PopInfo model, all simulated were correctly identified as admixed (figure 5).

In agreement with STRUCTURE results, the three putative hybrids were identified also by NEWHYBRIDS as admixed. S21 showed a posterior probability $p = 0.638$ to belong to F_1 class; S22 had $p = 0.982$ to belong to backcross with dog and 60c had $p = 0.991$ to be a backcross with jackal (table 1).

Relatedness analysis of female S21 and juvenile male S22 revealed that they shared at least one allele on each STR locus, confirming their mother–son relationship (electronic supplementary material, table S4).

3.4. *K* locus and major histocompatibility complex

The black hybrid 60c showed a heterozygote genotype at the *K* locus (K^+/K^B), whereas the deletion was absent in all reference jackals but present in 14 of the 51 reference dogs (table 2).

All 50 golden jackals and three putative golden jackal–dog hybrids successfully amplified at all three MHC loci analysed. In golden jackals, we found four DRB1, two DQA1 and two DQB1 alleles (table 3). Of these, three DLA-DRB1 (DLA-DRB1*13001, 13101 and 04503), one DLA-DQA1 (DLA-DQA1*03001) and both DLA-DQB1 alleles (DLA-DQB1*02305 and 06801) were not identified before in any canid

Table 1. Assignment of observed and simulated dog, golden jackal and admixed genotypes to reference parental populations (dogs, jackals), their first (F_1) and second (F_2) generation hybrids, their first (BC1) and second (BC2) backcrosses with jackals and dogs. The assignments of individual genotypes of samples S21, S22 and 60C are also reported. Q_j, average proportion of membership to jackal cluster; q_j, proportion of membership to jackal cluster of individual genotypes; *P*, mean posterior probability for each population to belong to each of the six classes (D, dog; J jackal; F_1, first generation hybrid; F_2, second generation hybrid; BCD, backcrosses of F_1 with dogs; BCJ, backcrosses of F_1 with jackal); *p*, individual posterior probability to belong to each of the six classes. The highest values are marked in italic.

	Structure				NewHybrids					
	only observed		observed and simulated							
	Q_j (90% CI) Admixture model	Q_j (90% CI) UsePopInfo model	Q_j (90% CI) Admixture model	Q_j (90% CI) UsePopInfo model	P_D	P_J	P_{F_1}	P_{F_2}	P_{BCD}	P_{BCJ}
dogs	0.002 (0.000–0.013)	0.000	0.020 (0.000–0.087)	0.000	*0.991*	0.000	0.000	0.000	0.009	0.000
jackals	*0.999* (0.994–1.000)	*1.000*	*0.988* (0.943–1.000)	*1.000*	0.000	*0.998*	0.000	0.000	0.000	0.002
F_1	—	—	0.511 (0.335–0.687)	0.508 (0.341–0.673)	0.000	0.000	*0.964*	0.006	0.004	0.026
F_2	—	—	0.505 (0.331–0.678)	0.502 (0.338–0.666)	0.000	0.000	0.068	*0.666*	0.126	0.139
BC1J	—	—	0.789 (0.629–0.921)	0.754 (0.600–0.884)	0.000	0.061	0.011	0.013	0.000	*0.915*
BC2J	—	—	0.917 (0.796–0.988)	0.865 (0.737–0.960)	0.000	*0.560*	0.000	0.002	0.000	*0.438*
BC1D	—	—	0.243 (0.104–0.408)	0.269 (0.132–0.426)	0.058	0.000	0.077	0.025	*0.839*	0.000
BC2D	—	—	0.090 (0.014–0.215)	0.139 (0.041–0.270)	*0.547*	0.000	0.000	0.002	*0.451*	0.000
	q_j (90% CI) Admixture model	q_j (90% CI) UsePopInfo model	q_j (90% CI) Admixture model	q_j (90% CI) UsePopInfo model	p_D	p_J	p_{F_1}	p_{F_2}	p_{BCD}	p_{BCJ}
S21	0.588 (0.396–0.770)	0.561 (0.403–0.717)	0.598 (0.416–0.772)	0.586 (0.413–0.750)	0.000	0.000	*0.638*	0.075	0.002	0.285
S22	0.227 (0.063–0.411)	0.319 (0.166–0.477)	0.255 (0.105–0.429)	0.275 (0.130–0.440)	0.000	0.000	0.000	0.018	*0.982*	0.000
60c	0.849 (0.687–1.000)	0.735 (0.584–0.875)	0.832 (0.680–0.954)	0.791 (0.646–0.912)	0.000	0.005	0.000	0.003	0.000	*0.991*

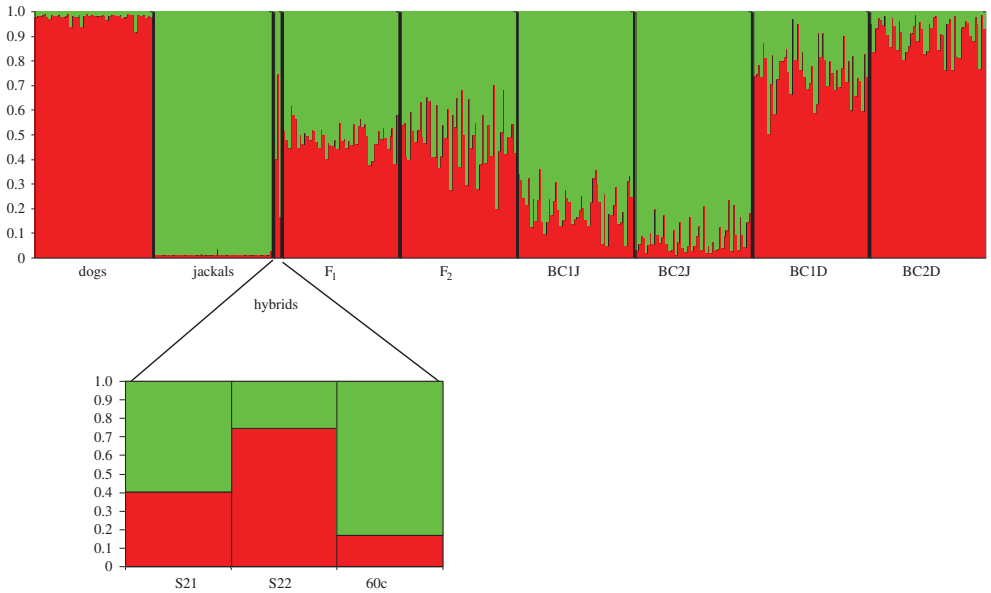


Figure 5. Bayesian analysis obtained by STRUCTURE using admixture models and $K = 2$. Each individual is represented by a vertical bar fragmented in K sections of different length, according to their membership proportion in the two inferred genetic clusters: the red represent the dog component and green the jackal component. Dogs, Croatian dog references; jackals, Croatian golden jackal references; hybrids, golden jackal–dog hybrids; simulated genotypes by HYBRIDLAB: F_1 and F_2, first and second generation hybrids, BC1J and BC2J, first and second backcrosses of F_1 with golden jackals; BC1D and BC2D, first and second backcrosses of F_1 with dogs.

Table 2. Number and frequency (in parenthesis) of genotypes at the β-defensin CBD103 gene: K^+/K^+, homozygotes wild-type (no deletion); K^+/K^B, heterozygotes for the KB melanistic deletion; K^B/K^B, homozygotes for the KB melanistic deletion. N_a, number of alleles; N_p, number of private alleles; H_o and H_e, observed and expected heterozygosity; HWE prob., probability test for Hardy–Weinberg equilibrium.

ref. pop (N)	N_a (s.e.)	N_p (s.e.)	H_o (s.e.)	H_e (s.e.)	HWE prob.	K^+/K^+	K^+/K^B	K^B/K^B
dogs (51)	8.20 (0.99)	4.67 (0.80)	0.61 (0.05)	0.68 (0.05)	0.0000	36 (0.72)	11 (0.22)	3 (0.06)
jackals (51)	4.33 (0.50)	1.07 (0.30)	0.46 (0.04)	0.54 (0.05)	0.0028	51 (1.0)	0	0

species. New alleles were assigned official names by the DLA Nomenclature Committee (GenBank accession numbers KT159767, KT182924–KT182928). Identified alleles combined to form four DLA-DRB1/DQA1/DQB1 three-locus haplotypes (table 3). In all three putative golden jackal–dog hybrids, we identified the same DLA-DRB1*00901/DQA1*00402/DQB1*02305 haplotype, which was exclusive to and the most frequent among golden jackals in our reference samples. Among the three alleles that constitute this haplotype, the allele DQB1*02305 appears to be specific for golden jackals, whereas alleles DRB1*00901 and DQA1*00402 were previously found in dogs, but not in the same haplotype (L. J. Kennedy 2013, personal communication).

The other haplotype and alleles present in hybrids were not seen in this golden jackal cohort, but are common in dogs. The haplotype DLA-DRB1*00803/DQA1*00301/DQB1*00401 found in putative hybrid S21 was previously detected in the Croatian sheepdog and border terrier (L. J. Kennedy 2013, personal communication), while haplotypes DLA-DRB1*02001/DQA1*00401/DQB1*01301 and DLA-DRB1*00101/DQA1*00101/DQB1*00201 detected in putative hybrids S22 and 60c, respectively, were found in numerous dog breeds [54] (table 3).

4. Discussion

Using genetic markers we confirmed that the three individuals with anomalous phenotypic characters were indeed interspecific hybrids, namely a first generation hybrid between golden jackal and domestic

Table 3. DLA-DRB1/DQA1/DQB1 haplotypes found in 50 golden jackals and genotypes found in three golden jackal–dog hybrids. Alleles in bold were found exclusively in golden jackal. Italicized haplotype was predominant in golden jackal population.

		haplotypes identified in 50 golden jackal individuals			haplotype frequency (%)	no. of animals with the haplotype (no. of homozygotes)
golden jackal		DRB1	DQA1	DQB1		
		00901	*00402*	***02305***	50.00	35 (15)
		13001	00402	**02305**	30.00	24 (6)
		13101	**03001**	**06801**	15.00	13 (2)
		04503	00402	**02305**	5.00	4 (1)
	individual	genotypes identified in three golden jackal–dog hybrids			haplotype determination	
golden jackal–dog hybrids	S21	*00901*	*00402*	***02305***	golden jackal	
		00803	00301	00401	Croatian sheepdog, border terrier[a]	
	S22	*00901*	*00402*	***02305***	golden jackal	
		02001	00401	01303	more than 25 dog breeds [54]	
	60c	*00901*	*00402*	***02305***	golden jackal	
		00101	00101	00201	50 dog breeds [54]	

[a]L. J. Kennedy 2013, personal communication.

dog (female S21), a backcross to dog (juvenile male S22) and a backcross to jackal (male 60c). The existence of backcrosses confirms that golden jackal–dog hybrids are fertile. Although these two species are estimated to have diverged about 1.7 Ma [2], or 0.4 Ma according to Freedman *et al*. [9], the occurrence of their hybrids and the fact that they are fertile do not come as a surprise, since reproductive isolation between pairs of geographically overlapping species evolves progressively [1] and may need hundreds to millions of generations to complete [55]. Golden jackals and dogs, together with grey wolves and coyotes, form a monophyletic clade and are more closely related than the golden jackal to two other jackal species, the side-striped jackal (*Canis adustus*) and the black-backed jackal (*Canis mesomelas*) [2]. Furthermore, Moura *et al*. [56] also reported weak evidence for current hybridization between grey wolves and golden jackals, as they identified several Bulgarian wolves exhibiting mixed ancestry with the jackal cluster.

The fact that golden jackal–dog hybrids were not recorded before in the wild might be a consequence of low research interest in golden jackals in the past, as they were not present in countries with the most active research communities. The golden jackal range has been expanding from southeastern Europe northwards and westwards in the last 30 years [8], thus a number of monitoring projects have been recently activated, increasing the possibility to record phenotypically abnormal individuals that might have always occurred in the populations. Moreover, before the advent of genetic tools, hybrids without a clear phenotypic signature could have gone undetected. On the other hand, the occurrence of golden jackal–dog hybrids might indeed be increasing because of several factors. First, higher population densities of golden jackals due to their recent expansion in the Balkan Peninsula [8] could enhance the encounters with stray and free-ranging dogs, which are quite abundant in Croatia and occur in the same areas where jackals live, hunt and gather food (I. Bošković 2012, unpublished data). It is known that the golden jackal is opportunistic in nature, primarily uses easily accessible human-derived food [57] and greatly benefits from the presence of agricultural surroundings [57,58]. Second, high mortality rates associated with jackal culling in Croatia [59], which mainly takes place from November to January and partially coincides with the jackal breeding season, can disrupt the social structure and promote hybridization with dogs, as suggested for wolf and coyote [5,60] or wolf and dog [16]. Yet, our results indicate that at least in Croatia introgressive hybridization between the two species is not a widespread phenomenon. We showed that Croatian golden jackals and dogs remain separated, forming two well-differentiated genetic entities where individuals are assigned to their respective cluster (figures 4 and 5), without significant ancestry in the other cluster.

Golden jackals exhibit lower genetic diversity measures than dogs (number of alleles, number of private alleles, observed and expected heterozygosity) (electronic supplementary material, table S3). This

is in line with previous investigations (e.g. [6,61]), which compared diversity measures between grey wolves and dogs and consistently reported higher values in dogs, probably due to the multiple events of domestication and introgression during dog history [61], or to an ascertainment bias [62]. However, this pattern could be different when taking into account the genetic variability of single dog breeds, which show much less genetic variation than what can be observed across breeds or in village dogs [63–65].

In order to evaluate the power of the STR loci in the determination of golden jackal–dog hybrids, we also used simulated genotypes for six different hybrid classes. The analysis of the simulation results revealed that these 15 STRs are variable enough to detect 100% of parentals and F_1, F_2, BC1J and BC1D hybrids using a threshold of 0.98. Only 6% and 10% of second backcrosses with jackals and dogs, respectively, were undetected using the Admixture model and no other *a priori* information in the Bayesian assignment procedure. Thus, the 15 STRs seem to be reliable and powerful enough to detect hybrids between golden jackals and dogs.

The analysis of the mtDNA CR and of the *Zfy* intron on the Y chromosome appears to be diagnostic, since these markers proved to be fixed for different haplotypes in golden jackals and dogs (this study, [19,20]). All the three anomalous individuals carried a golden jackal mtDNA CR haplotype, whereas the two males showed Y chromosome marker amplicons characteristic of dogs [20], enabling us to deduce their lines of descent (electronic supplementary material, figure S1): a female golden jackal mated with a male domestic dog to produce the hybrid F_1 female S21, which further mated with a domestic dog producing the backcross S22. The father of the male 60c was a golden jackal–dog hybrid, whose mother was a golden jackal and whose father was a dog, and he further mated with a female golden jackal. Accordingly, in both cases of F_1 hybrids (one documented—S21, and the other deduced—the father of 60c, which was not sampled), hybridization took place between female golden jackals and male dogs, which is congruent with the sexual asymmetry present in most hybridizations between domestic dogs and wild *Canis* species, e.g. grey wolf [61] and Ethiopian wolf [3], though occasionally violated [15]. The finding of a heterozygote genotype at the *K* locus (K^+/K^B) in the backcross to that of a golden jackal 60c with black coat coloration, the absence of the melanistic *K* locus mutation in all reference jackals and its presence in 14 of the dog samples (table 2) suggest that this hybrid could have received the *K* locus deletion from dogs. In this way, golden jackal could join the panel of canid species that possibly derived their melanistic *K* locus mutation through hybridization with domestic dogs, namely grey wolves and coyotes [16,27]. Further, our findings cast doubts on the hypothesis of Ambarli & Bilgin [66], according to which melanism in the golden jackal they camera-trapped was due to an independent mutation instead of introgression from the domestic dog, and suggest that this individual might be another case of golden jackal–dog hybrid.

The MHC loci further confirmed that the three anomalous individuals described here were hybrids. Three out of four DLA-DRB1 alleles, one out of two DLA-DQA1 alleles and both DLA-DQB1 alleles found in golden jackals in this research (table 3) have not been identified before in any canid species. Thus, they could represent golden jackal private alleles that can be used as species-diagnostic markers. However, as the three hybrids in our research did not possess any of those jackal private alleles on DRB1 and DQA1 loci, the species (and hence, hybridization) determination could not have been performed using either of those two loci separately. Notwithstanding, the combined use of alleles on three loci enabled us to reliably confirm golden jackal ancestry in all three individuals in our research, since all of them possessed a three-locus haplotype that is specific and exclusive for golden jackals (DRB1*00901/DQA1*00402/DQB1*02305, table 3) and has not been found in dogs (L. J. Kennedy 2013, personal communication). In other words, even when alleles on particular MHC loci are not species-specific, their two- or three-locus haplotypes might prove to be, due to high levels of linkage disequilibrium across large stretches of this genomic region, where particular combinations of alleles at neighbouring loci are maintained by selection [67]. Further, the other haplotype possessed by each of the three individuals was characteristic of dogs and none of those was found in our reference jackal samples (table 3). In addition, none of the alleles comprising those haplotypes was found in jackals, indicating that even a single locus would suffice to confirm the dog ancestry in those samples. Likewise, for the confirmation of jackal ancestry in this research, the exclusive use of DQB1 locus would suffice since both alleles found in the reference jackal population are private. However, it would come as no surprise if additional alleles were found on that locus as more jackal samples become typed, and if some of them were to be shared between golden jackals and domestic dogs. This phenomenon, known as trans-species polymorphism, is characteristic of MHC genes, mainly occurs among closely related species and is a consequence of balancing selection, which acts on MHC genes over the long term and maintains ancestral polymorphism in descendant species [68]. Indeed, when trans-species polymorphism is present, analysing three-locus haplotypes instead of individual MHC loci should add power to detect

possible introgression. The procedure for analysing MHC markers is quite straightforward, and time and cost efficient even when analysing three loci. Albeit they were applied as markers to detect hybridization in only two studies up to now [24,25], our results indicate that MHC genes can be used as suitable molecular markers for the identification of vertebrate species and for the determination of hybridization events, at least when data on reference parental populations are available and they are not closely related.

Hybridization might play an important role in the process of animal evolution, especially in rapidly changing environments. Hybridizing with dogs, golden jackals might increase their variability and thus facilitate their long-term adaptation. In addition, adaptive introgression may be facilitated for genes evolving under multi-allelic balancing selection, such as the vertebrate MHC system, where increased resistance to infectious diseases from adapted MHC variants might be transferred to closely related recipient species, as long as fertile hybrids can be formed [69,70]. This could be the case for the jackal, where standing variation to face new adaptive challenges, such as new parasites associated with human-related food sources, may be low but could be compensated by introgression from dogs, which have already adapted to human-related environments [70]. Adaptive introgression of MHC genes was recently suggested between two closely related species of newts [71] and from archaic to modern humans [72]. Sexual selection may also facilitate introgression of dog MHC alleles into the genomes of golden jackals, potentially contributing to resistance against currently prevalent parasites, as was proposed for three-spined sticklebacks (*Gasterosteus aculeatus*) [73]. It is thought that MHC-based mate choice may allow genes to cross species boundaries if parasite selection is strong enough that the benefits of hybridization overcome its costs [74].

Conversely, hybridization and introgression may also have harmful effects on the fitness of animal populations in the wild, causing loss of genetic diversity due to genetic homogenization and/or outbreeding depression in local populations [75]. In addition, possible cross-species transmission of canine diseases [76] might pose another risk for the expanding golden jackal populations.

In summary, in this paper we document the first occurrence of three cases of golden jackal–dog hybrids. However, the frequency of hybridization events, the extent of possible genetic introgression of dog genes into European golden jackal populations and the consequences on genetic diversity and population fitness (either beneficial or unfavourable) still remain to be investigated.

Ethics. Tissue samples used in this study were obtained from golden jackal individuals legally harvested by hunters or road-killed. No animal was killed for the purpose of this study.

Data accessibility. The dataset supporting this article has been uploaded as part of the electronic supplementary material.

Authors' contributions. A.G., E.F., R.C., H.A. and E.R. designed the study; A.G., E.F., H.A., R.C., S.L. and M.G. carried out the molecular work and participated in data analysis; T.F. and I.B. conceived the study and collected field data; E.R. coordinated the study; A.G., E.F., R.C., H.A. and E.R. drafted the manuscript. All authors revised the manuscript and gave final approval for publication.

Competing interests. We have no competing interests.

Funding. This research was partially funded by the Croatian Science Foundation (Project 'Molecular epidemiology of selected parasitic diseases of wildlife', code 3421).

References

1. Mallet J. 2005 Hybridization as an invasion of the genome. *Trends Ecol. Evol.* **20**, 229–237. (doi:10.1016/j.tree.2005.02.010)
2. VonHoldt BM *et al*. 2011 A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. *Genome Res.* **21**, 1294–1305. (doi:10.1101/gr.116301.110)
3. Gottelli D, Sillero-Zubiri C, Applebaum GD, Roy MS, Girman DJ, Garcia-Moreno J, Ostrander EA, Wayne RK. 1994 Molecular genetics of the most endangered canid: the Ethiopian wolf *Canis simensis*. *Mol. Ecol.* **3**, 301–312. (doi:10.1111/j.1365-294X.1994.tb00070.x)
4. 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. *Conserv. Genet.* **3**, 31–45. (doi:10.1016/j.earscirev.2006.09.002)
5. Stronen AV *et al*. 2012 Canid hybridization: contemporary evolution in human-modified landscapes. *Ecol. Evol.* **2**, 2128–2140. (doi:10.1002/ece3.335)
6. Randi E *et al*. 2014 Multilocus detection of wolf x dog hybridization in Italy, and guidelines for marker selection. *PLoS ONE* **9**, e86409. (doi:10.1371/journal.pone.0086409)
7. Jhala Y, Moehlman P. 2004 *Canis aureus* Linnaeus, 1758. In *Canids: foxes, wolves, jackals and dogs status survey and conservation action plan* (eds C Sillero-Zubiri, M Hoffman, D Macdonald), pp. 156–161. Oxford, UK: IUCN/SSC Canid Specialist Group Gland, Switzerland.
8. Arnold J, Humer A, Heltai M, Murariu D, Spassov N, Hackländer K. 2012 Current status and distribution of golden jackals *Canis aureus* in Europe. *Mamm. Rev.* **42**, 1–11. (doi:10.1111/j.1365-2907.2011.00185.x)
9. Freedman AH *et al*. 2014 Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet.* **10**, e1004016. (doi:10.1371/journal.pgen.1004016)
10. Rueness EK, Asmyhr MG, Sillero-Zubiri C, Macdonald DW, Bekele A, Atickem A, Stenseth NC. 2011 The cryptic African wolf: *Canis aureus lupaster* is not a golden jackal and is not endemic to Egypt. *PLoS ONE* **6**, e16385. (doi:10.1371/journal.pone.0016385)
11. Gaubert P, Bloch C, Benyacoub S, Abdelhamid A, Pagani P, Djagoun CAMS, Couloux A, Dufour S. 2012 Reviving the African wolf *Canis lupus lupaster* in north and west Africa: a mitochondrial lineage ranging more than 6,000 km wide. *PLoS ONE* **7**, e42740. (doi:10.1371/journal.pone.0042740)
12. Tóth T, Krecsák L, Szucs E, Heltai M, Huszár G. 2009 Records of the golden jackal (*Canis aureus* Linnaeus, 1758) in Hungary from 1800th until 2007, based on a

literature survey. *North West. J. Zool.* **5**, 386–405.

13. VonHoldt BM, Pollinger JP, Earl DA, Parker HG, Ostrander EA, Wayne RK. 2012 Identification of recent hybridization between gray wolves and domesticated dogs by SNP genotyping. *Mamm. Genome* **24**, 80–88. (doi:10.1007/s00335-012-9432-0)
14. Vilà C *et al*. 2003 Combined use of maternal, paternal and bi-parental genetic markers for the identification of wolf–dog hybrids. *Heredity* **90**, 17–24. (doi:10.1038/sj.hdy.6800175)
15. 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**, e46465. (doi:10.1371/journal.pone.0046465)
16. 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? *Eur. J. Wildl. Res.* **59**, 543–555. (doi:10.1007/s10344-013-0703-1)
17. Muñoz-Fuentes V, Vilà C, Green AJ, Negro JJ, Sorenson MD. 2007 Hybridization between white-headed ducks and introduced ruddy ducks in Spain. *Mol. Ecol.* **16**, 629–638. (doi:10.1111/j.1365-294X.2006.03170.x)
18. Zachos FE, Cirovic D, Kirschning J, Otto M, Hartl GB, Petersen B, Honnen AC. 2009 Genetic variability, differentiation, and founder effect in golden jackals (*Canis aureus*) from Serbia as revealed by mitochondrial DNA and nuclear microsatellite loci. *Biochem. Genet.* **47**, 241–250. (doi:10.1007/s10528-009-9221-y)
19. Fabbri E *et al*. 2014 Genetic structure and expansion of golden jackals (*Canis aureus*) in the north-western distribution range (Croatia and eastern Italian Alps). *Conserv. Genet.* **15**, 187–199. (doi:10.1007/s10592-013-0530-7)
20. Galov A, Sindičić M, Gomerčić T, Arbanasić H, Baburić M, Bošković I, Florijančić T. 2014 PCR-based Y chromosome marker for discriminating between golden jackal (*Canis aureus*) and domestic dog (*Canis lupus familiaris*) paternal ancestry. *Conserv. Genet. Resour.* **6**, 275–277. (doi:10.1007/s12686-013-0110-7)
21. Magory Cohen T, King R, Dolev A, Boldo A, Lichter-Peled A, Kahila Bar-Gal G. 2013 Genetic characterization of populations of the golden jackal and the red fox in Israel. *Conserv. Genet.* **14**, 55–63. (doi:10.1007/s10592-012-0423-1)
22. Pastorini J, Zaramody A, Curtis DJ, Nievergelt CM, Mundy NI. 2009 Genetic analysis of hybridization and introgression between wild mongoose and brown lemurs. *BMC Evol. Biol.* **9**, 32. (doi:10.1186/1471-2148-9-32)
23. Bernatchez L, Landry C. 2003 MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *J. Evol. Biol.* **16**, 363–377. (doi:10.1046/j.1420-9101.2003.00531.x)
24. Alasaad S, Fickel J, Rossi L, Sarasa M, Benítez CV, Granados JE, Soriguer RC. 2012 Applicability of major histocompatibility complex DRB1 alleles as markers to detect vertebrate hybridization: a case study from Iberian ibex x domestic goat in southern Spain. *Acta Vet. Scand.* **54**, 56. (doi:10.1186/1751-0147-54-56)
25. Galaverni M, Caniglia R, Fabbri E, Lapalombella S, Randi E. 2013 MHC variability in an isolated wolf population in Italy. *J. Hered.* **104**, 601–612. (doi:10.1093/jhered/est045)
26. Wayne RK, Vonholdt BM. 2012 Evolutionary genomics of dog domestication. *Mamm. Genome* **23**, 3–18. (doi:10.1007/s00335-011-9386-7)
27. Anderson TM *et al*. 2009 Molecular and evolutionary history of melanism in North American gray wolves. *Science* **323**, 1339–1343. (doi:10.1126/science.1165448)
28. Douzery E, Randi E. 1997 The mitochondrial control region of Cervidae: evolutionary patterns and phylogenetic content. *Mol. Biol. Evol.* **14**, 1154–1166. (doi:10.1093/oxfordjournals.molbev.a025725)
29. 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. *Conserv. Biol.* **14**, 464–473. (doi:10.1046/j.1523-1739.2000.98280.x)
30. Hall TA. 1999 BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98.
31. Sindičić M, Gomerčić T, Galov A, Arbanasić H, Kusak J, Slavica A, Huber đ. 2011 Mitochondrial DNA control region as a tool for species identification and distinction between wolves and dogs from Croatia. *Vet. Arh.* **81**, 249–258.
32. 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. (doi:10.1093/bioinformatics/bts460)
33. Peakall R, Smouse PE. 2006 GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **6**, 288–295. (doi:10.1111/j.1471-8286.2005.01155.x)
34. Guo SW, Thompson EA. 1992 Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* **48**, 361–372. (doi:10.2307/2532296)
35. Rousset F. 2008 GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol. Ecol. Resour.* **8**, 103–106. (doi:10.1111/j.1471-8286.2007.01931.x)
36. Rice WR. 1989 Analysing tables of statistical tests. *Evolution* **43**, 223–225. (doi:10.1016/0260-4779(86)90011-7)
37. Orloci L. 1978 *Multivariate analysis in vegetation research*. The Hague, The Netherlands: W Junk.
38. Pritchard JK, Stephens M, Donnelly P. 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.
39. Falush D, Stephens M, Pritchard J. 2003 Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587.
40. Falush D, Stephens M, Pritchard JK. 2007 Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* **7**, 574–578. (doi:10.1111/j.1471-8286.2007.01758.x)
41. Evanno G, Regnaut S, Goudet J. 2005 Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620. (doi:10.1111/j.1365-294X.2005.02553.x)
42. Anderson EC, Thompson EA. 2002 A model-based method for identifying species hybrids using multilocus data. *Genetics* **160**, 1217–1229. (doi:test statistics; hybrids)
43. Nielsen EE, Bach LA, Kotlicki P. 2006 HYBRIDLAB (version 1.0): a program for generating simulated hybrids from population samples. *Mol. Ecol. Notes* **6**, 971–973. (doi:10.1111/j.1471-8286.2006.01433.x)
44. Candille SI *et al*. 2007 A -defensin mutation causes black coat color in domestic dogs. *Science* **318**, 1418–1423. (doi:10.1126/science.1147880)
45. Kennedy LJ, Quarmby S, Fretwell N, Martin AJ, Jones PG, Jones CA, Ollier WER. 2005 High-resolution characterization of the canine DLA-DRBI locus using reference strand-mediated conformational analysis. *J. Hered.* **96**, 836–842. (doi:10.1093/jhered/esi112)
46. Wagner JL, Burnett RC, Works JD, Storb R. 1996 Molecular analysis of DLA-DRBB1 polymorphism. *Tissue Antigens* **48**, 554–561. (doi:10.1111/j.1399-0039.1996.tb02669.x)
47. Wagner JL, Burnett RC, DeRose SA, Storb R. 1996 Molecular analysis and polymorphism of the DLA-DQA gene. *Tissue Antigens* **48**, 199–204. (doi:10.1111/j.1399-0039.1996.tb02629.x)
48. Wagner JL, Hayes-Lattin B, Works JD, Storb R. 1998 Molecular analysis and polymorphism of the DLA-DQB genes. *Tissue Antigens* **52**, 242–250. (doi:10.1111/j.1399-0039.1998.tb03039.x)
49. Kennedy LJ, Angles JM, Barnes A, Carmichael LE, Radford AD, Ollier WER, Happ GM. 2007 DLA-DRB, DQAI, and DQBI alleles and haplotypes in North American gray wolves. *J. Hered.* **98**, 491–499. (doi:10.1093/jhered/esm051)
50. Arbanasić H, Huber D, Kusak J, Gomerčić T, Hrenović J, Galov A. 2013 Extensive polymorphism and evidence of selection pressure on major histocompatibility complex DLA-DRB1, DQA1 and DQB1 class II genes in Croatian grey wolves. *Tissue Antigens* **81**, 19–27. (doi:10.1111/tan.12029)
51. Kennedy LJ *et al*. 2002 Extensive interbreed, but minimal intrabreed, variation of DLA class II alleles and haplotypes in dogs. *Tissue Antigens* **59**, 194–204. (doi:10.1034/j.1399-0039.2002.590303.x)
52. Excoffier L, Laval G, Schneider S. 2005 Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **1**, 47–50.
53. Oliveira R, Godinho R, Randi E, Ferrand N, Alves PC. 2008 Molecular analysis of hybridisation between wild and domestic cats (*Felis silvestris*) in Portugal: implications for conservation. *Conserv. Genet.* **9**, 1–11. (doi:10.1007/s10592-007-9297-z)
54. Kennedy LJ *et al*. 2007 Canine DLA diversity: 1. New alleles and haplotypes. *Tissue Antigens* **69**, 272–288. (doi:10.1111/j.1399-0039.2006.00779.x)
55. Abbott R *et al*. 2013 Hybridization and speciation. *J. Evol. Biol.* **26**, 229–246. (doi:10.1111/j.1420-9101.2012.02599.x)
56. Moura AE, Tsingarska E, Dabrowski MJ, Czarnomska SD, Jedrzejewska B, Pilot M. 2014 Unregulated hunting and genetic recovery from a severe population decline: the cautionary case of Bulgarian wolves. *Conserv. Genet.* **15**, 405–417. (doi:10.1007/s10592-013-0547-y)
57. Lanszki J, Giannatos G, Dolev A, Bino G, Heltai M. 2010 Late autumn trophic flexibility of the golden jackal *Canis aureus*. *Acta Theriol. (Warsz)* **55**, 361–370. (doi:10.1007/BF03193239)

58. Borkowski J, Zalewski A, Manor R. 2011 Diet composition of golden jackals in Israel. *Ann. Zool. Fennici* **48**, 108–118. (doi:10.5735/086.048.0203)
59. Selanec I, Lauš B, Sindičić M. 2011 Golden jackal (*Canis aureus*) distribution in Croatia. In *Abstracts of VI European Congress of Mammalogy, 19–23 July 2011, Université P. et M. Curie – Muséum national d'Histoire naturelle, Paris, France* (ed. C Denys).
60. Rutledge LY, White BN, Row JR, Patterson BR. 2012 Intense harvesting of eastern wolves facilitated hybridization with coyotes. *Ecol. Evol.* **2**, 19–33. (doi:10.1002/ece3.61)
61. Godinho R *et al*. 2011 Genetic evidence for multiple events of hybridization between wolves and domestic dogs in the Iberian Peninsula. *Mol. Ecol.* **20**, 5154–5166. (doi:10.1111/j.1365-294X.2011.05345.x)
62. Cronin MA, Canovas A, Bannasch DL, Oberbauer AM, Medrano JF. 2014 Single nucleotide polymorphism (SNP) variation of wolves (*Canis lupus*) in Southeast Alaska and comparison with wolves, dogs, and coyotes in North America. *J. Hered.* **106**, 26–36. (doi:10.1093/jhered/esu075)
63. Boyko AR *et al*. 2009 Complex population structure in African village dogs and its implications for inferring dog domestication history. *Proc. Natl Acad. Sci. USA* **106**, 13 903–13 908. (doi:10.1073/pnas.0902129106)
64. 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. *Vet. J.* **189**, 155–159. (doi:10.1016/j.tvjl.2011.06.013)
65. Vaysse A *et al*. 2011 Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet.* **7**, e1002316. (doi:10.1371/journal.pgen.1002316)
66. Ambarli H, Bilgin CC. 2013 First record of a melanistic golden jackal (*Canis aureus*, Canidae) from Turkey. *Mammalia* **77**, 219–222. (doi:10.1515/mammalia-2012-0009)
67. Garrigan D, Hedrick PW. 2003 Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution* **57**, 1707–1722. (doi:10.1554/02-732)
68. Piertney SB, Oliver MK. 2006 The evolutionary ecology of the major histocompatibility complex. *Heredity* **96**, 7–21. (doi:10.1038/sj.hdy.6800724)
69. Castric V, Bechsgaard J, Schierup MH, Vekemans X. 2008 Repeated adaptive introgression at a gene under multiallelic balancing selection. *PLoS Genet.* **4**, e1000168. (doi:10.1371/journal.pgen.1000168)
70. Hedrick PW. 2013 Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol. Ecol.* **22**, 4606–4618. (doi:10.1111/mec.12415)
71. Nadachowska-Brzyska K, Zieliński P, Radwan J, Babik W. 2012 Interspecific hybridization increases MHC class II diversity in two sister species of newts. *Mol. Ecol.* **21**, 887–906. (doi:10.1111/j.1365-294X.2011.05347.x)
72. Abi-Rached L *et al*. 2011 The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* **334**, 89–94. (doi:10.1126/science.1209202)
73. Eizaguirre C, Yeates SE, Lenz TL, Kalbe M, Milinski M. 2009 MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Mol. Ecol.* **18**, 3316–3329. (doi:10.1111/j.1365-294X.2009.04243.x)
74. Wegner KM, Eizaguirre C. 2012 New(t)s and views from hybridizing MHC genes: introgression rather than trans-species polymorphism may shape allelic repertoires. *Mol. Ecol.* **21**, 779–781. (doi:10.1111/j.1365-294X.2011.05401.x)
75. Randi E. 2008 Detecting hybridization between wild species and their domesticated relatives. *Mol. Ecol.* **17**, 285–293. (doi:10.1111/j.1365-294X.2007.03417.x)
76. Hedrick PW, Lee RN, Buchanan C. 2003 Canine parvovirus enteritis, canine distemper, and major histocompatibility complex genetic variation in Mexican wolves. *J. Wildl. Dis.* **39**, 909–913. (doi:10.7589/0090-3558-39.4.909)

Supplementary materials PAPER II

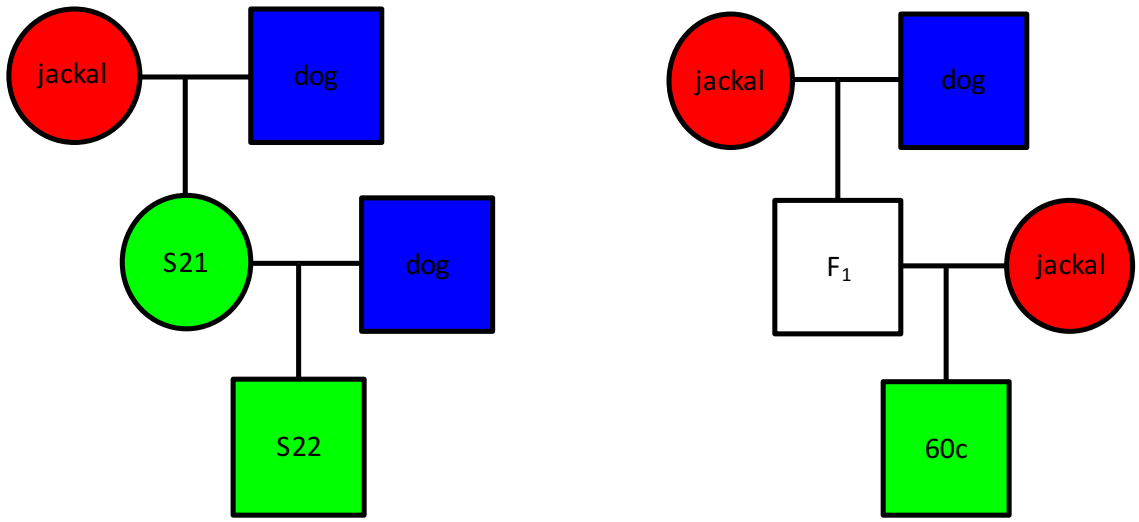


Figure S1: Pedigrees of the three hybrids (shown in green). Circles represent females, squares represent males.

Table S1. Primer sequences (forward and reverse), annealing temperatures (Ta) and product sizes (bp) for mtDNA, Y chromosome marker, K locus and DLA-DRB1, DQA1 and DQB1 genes

Marker	Product size	Ta	Forward / Reference	Reverse / Reference
mtDNA	550	55°C	CGTCAGTCTCACCATCAACCCC CAAAGC (L-Pro) / 1	TTTGACTGCATTAGGGC CGCGACGG (H576) / 2
Y chromosome marker	536 (jackal) 242 + 566 (dog)	57°C	GTCCATTGGATAATTCTTTCC (Yint2-335) / 3 GCACTGCTAAATCAACCAC (YintF2) / 4	CAAGTTCTGCTTTGGTT CT (YintR) / 4
K-locus	147-151	55°C	TGTCTTCATCCCTGTGAGGT / 5	CCAGGAGGCATTTTCAC ACT / 5
DLA-DRB1	267	62°C*	GATCCCCCCGTCCCCACAG (DRBF) / 6	TGTGTCACACACCTCAG CACCA (DRB1R) / 7
DLA-DQA1	246	54°C*	TAAGGTTCTTTTCTCCCTCT (DQAin1) / 8	GGACAGATTCAGTGAA GAGA DQAIn2 / 8
DLA-DQB1	267	73°C*	CTCACTGGCCCGGCTGTCTC (DQB1BT7) / 8	ACCTGGGTGGGGAGCC CG DQBR3 / 9

*Amplifications were performed using a touchdown PCR protocol consisting first of 95°C for 15 min, followed by 14 touchdown cycles comprising of 95°C for 30s, the annealing temperature for 1 min and 72°C for 1 min. Annealing temperatures were set initially at indicated temperatures, then reduced by 0.5°C in each cycle. This was followed by 20 cycles of 95°C for 30s; 55°C (DRB1), 47°C (DQA1) or 66°C (DQB1) for 1min and 72°C for 1 min. A final extension step was carried out at 72°C for 10 min.

[1] Douzery, E. & Randi, E. 1997 The mitochondrial control region of Cervidae: evolutionary patterns and phylogenetic content. *Mol. Biol. Evol.* **14**, 1154–1166.
[2] Randi, E., Lucchini, V., Christensen, M. F., Mucci, N., Funk, S. M., Dolf, G. & Loeschcke, V. 2000 Mitochondrial DNA variability in Italian and east European wolves: Detecting the consequences of small population size and hybridization. *Conserv. Biol.* **14**, 464–473.
[3] Shami, K. 2002 Evaluating the change in distribution of the eastern timber wolf (*Canis lycaon*) using the Y-chromosome. Master thesis, McMaster University. Cited in: Rutledge, L. Y., Garroway, C. J., Loveless, K. M., Patterson, B. R. 2010 Genetic differentiation of eastern wolves in Algonquin Park despite bridging gene flow between coyotes and grey wolves. *Heredity* **105**, 520-531.
[4] Galov, A., Sindičić, M., Gomerčić, T., Arbanasić, H., Baburić, M., Bošković, I. & Florijančić, T. 2014 PCR-based Y chromosome marker for discriminating between golden jackal (*Canis aureus*) and domestic dog (*Canis lupus familiaris*) paternal ancestry. *Conserv. Genet. Resour.* **6**, 275–277.
[5] Candille, S. I., Kaelin, C. B., Cattanach, B. M. *et al.* (2007) A *β*-defensin mutation causes black coat colour in domestic dogs. *Science* **318**:1418–1423.

[6] Kennedy, L. J., Quarmby, S., Fretwell, N., Martin, a. J., Jones, P. G., Jones, C. a. & Ollier, W. E. R. 2005 High-resolution characterization of the canine DLA-DRBI locus using reference strand-mediated conformational analysis. *J. Hered.* **96**, 836–842.
[7] Wagner, J. L., Burnett, R. C., Works, J. D. & Storb, R. 1996 Molecular analysis of DLA-DRBB1 polymorphism. *Tissue Antigens* **48**, 554–561.
[8] Wagner, J. L., Burnett, R. C., DeRose, S. A. & Storb, R. 1996 Molecular analysis and polymorphism of the DLA-DQA gene. *Tissue Antigens* **48**, 199–204.
[9] Primer designed in this study based on the sequence published in: Wagner, J. L., Hayes-Lattin, B., Works, J. D. & Storb, R. 1998 Molecular analysis and polymorphism of the DLA-DQB genes. *Tissue Antigens* **52**, 242–250.

Table S2. List of autosomal microsatellite loci used, with primer sequences (forward and reverse), annealing temperatures and product sizes (bp)

Marker	Chr	Repeat	Allele Range	Temp. Annealing	Multiplex	Forward	Reverse	Ref.
CPH6	CFA23	di	120-126	57° C	A	CATTGGCTGTTTGACTCTAGG	ACTGATGTGGGTGTCTCTGC	1
FH2004	CFA11	tetra	104-202	57° C	A	TCATTGCCTGATACAGACTGAG	TTGAGCTTAGTACTCAAAGCATAGTG	2
FH2088	CFA15	di	91-139	57° C	A	CCCTCTGCCTACATCTCTGC	TAGGGCATGCATATAACCAGC	2
FH2140	CFA5	tetra	112-168	57° C	B	GGGGAAGCCATTTTTAAAGC	TGACCCTCTGGCATCTAGGA	3
C20.253	CFA20	di	90-120	57° C	B	AATGGCAGGATTTTCTTTTGC	ATCTTTGGACGAATGGATAAGG	4
CPH8	CFA13	di	191-219	57° C	C	AGGCTCACAATCCCTCTCATA	TAGATTTGATACCTCCCTGAGTCC	1
FH2096	CFA11	tetra	80-110	57° C	C	CCGTCTAAGAGCCTCCCAG	GACAAGGTTTCCTGGTTCCA	2
C09.250	CFA9	di	117-145	57° C	D	TTAGTTAACCCAGCTCCCCCA	TCACCCTGTTAGCTGCTCAA	4
CXX.213	CFA25	di	158-162	57° C	D	AATATGGGAGAGGAGAAGAGGG	ATGCTTCCTGGTAAGCAATCA	4
CPH4	CFA15	di	130-155	57° C	E	ACTGGAGATGAAAACTGAAGATTATA	TTACAGGGGAAAGCCTCATT	1
CPH5	CFA15	di	102-124	57° C	E	TCCATAACAAGACCCCAAAC	GGAGGTAGGGGTCAAAAGTT	1
CPH12	CFA8	di	186-214	57° C	F	GGCATTACTTGGAGGGAGGAA	GATGATTCCTATGCTTCTTTGAG	1
CPH9	CFA28	di	139-154	57° C	F	CAGAGACTGCCACTTTAAACACAC	AAAGTTCTCAAATACCATTGTGTTACA	1
FH2137	CFA3	di	140-204	57° C	G	GCAGTCCCTTATTCCAACATG	CCCCAAGTTTTGCATCTGTT	2
CPH22	CFA3	di	108-120	57° C	G	TCTTTCATTTACATTTTTGGCTCA	GCCCCAAAATCCGTGTGT	5

[1] Fredholm M, Winterø AK (1995) Variation of short tandem repeats within and between species belonging to the Canidae family. Mammalian Genome 6:11-18.
[2] Francisco LV, Langston AA, Mellersh CS, Neal CL, Ostrander EA (1996) A class of highly polymorphic tetranucleotide repeats for canine genetic mapping. Mammalian Genome 7:359–62.
[3] Breen M, Jouquand S, Renier 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.
[4] Ostrander EA, Sprague GF, Rine J (1993) Identification and Characterization of dinucleotide repeat (CA)n markers for genetic mapping in dog. Genomics 16:207-213.
[5] Dolf G, Schläpfer J, Gaillard C, Randi E, Lucchini V, Breitenmoser U, Stahlberger Saitbekova N (2000) Differentiation between Italian wolf and domestic dog based on microsatellite analysis. Genetics Selection Evolution 32:533–541.

Table S3. Genetic variability at 15 autosomal microsatellites in 51 domestic dogs and 50 golden jackals sampled in Croatia

Dogs (51)	*Locus ID*	*Na*	*Np*	*Ho*	*He*	*Prob*	*Signif*
	CPH4	6	1	0,56	0,70	0,098	ns
	CPH5	5	2	0,55	0,63	0,907	ns
	CPH6	11	7	0,79	0,85	0,000	***
	CPH8	7	5	0,76	0,79	0,509	ns
	CPH9	8	4	0,65	0,64	0,863	ns
	CPH12	7	4	0,55	0,73	0,000	***
	CPH22	3	1	0,13	0,21	0,005	*
	FH2004	11	6	0,75	0,78	0,027	ns
	FH2088	6	5	0,65	0,76	0,558	ns
	FH2096	4	0	0,50	0,52	0,947	ns
	FH2137	19	13	0,88	0,92	0,023	ns
	FH2140	9	5	0,77	0,77	0,812	ns
	CXX.213	9	6	0,49	0,79	0,001	***
	C09.250	10	6	0,73	0,74	0,091	ns
	C20.253	8	5	0,37	0,41	0,388	ns
	Average	**8,20**	**4,67**	**0,61**	**0,68**		
	(SE)	**(0,99)**	**(0,80)**	**(0,05)**	**(0,05)**		
Jackals (50)	***Locus ID***	***Na***	***Np***	***Ho***	***He***	***Prob***	***Signif***
	CPH4	4	1	0,63	0,66	0,743	ns
	CPH5	4	1	0,47	0,52	0,934	ns
	CPH6	3	0	0,45	0,53	0,350	ns
	CPH8	4	2	0,28	0,28	0,001	**
	CPH9	4	0	0,51	0,53	0,525	ns
	CPH12	3	1	0,04	0,08	0,000	***
	CPH22	3	1	0,31	0,46	0,063	ns
	FH2004	7	0	0,59	0,65	0,921	ns
	FH2088	2	0	0,43	0,49	0,389	ns
	FH2096	4	0	0,49	0,64	0,020	ns
	FH2137	10	4	0,63	0,76	0,002	*
	FH2140	4	1	0,57	0,71	0,222	ns
	CXX.213	3	0	0,35	0,39	0,326	ns
	C09.250	5	1	0,61	0,69	0,653	ns
	C20.253	5	3	0,57	0,71	0,134	ns
	Average	**4,33**	**1,07**	**0,46**	**0,54**		
	(SE)	**(0,50)**	**(0,30)**	**(0,04)**	**(0,05)**		

Na = number of alleles; Np =number of private alleles; Ho and He = observed and expected heterozygosity; Prob = probability values of departures from Hardy-Weinberg proportions after Bonferroni correction: ns = not significant, * P < 0,05, ** P < 0,001, *** P < 0,0001.

Table S4. STR genotypes of female S21 and juvenile male S22

locus	CPH4		CPH5		CPH6		CPH8		CPH9		CPH12		CPH22		FH2004	
S21	137	**149**	110	**116**	**124**	**124**	**195**	**195**	138	**148**	198	**206**	108	**116**	188	188
S22	141	**149**	112	**116**	**124**	**124**	**195**	**195**	146	**148**	194	**206**	110	**116**	0	0

locus	FH2088		FH2096		FH2137		FH2140		CXX.213		C09.250		C20.253	
S21	**97**	**125**	92	**100**	**153**	162	130	**134**	**150**	158	117	**133**	98	**106**
S22	**97**	**125**	96	**100**	**153**	153	122	**134**	**150**	150	133	**133**	106	**106**

Shared genotypes are marked in bold.

INTRODUCTION TO PAPER III

Choosy Wolves ? Heterozygote Advantage But No Evidence of MHC-Based Disassortative Mating

Study of MHC loci has gained great popularity in recent years, partly due to their function in protecting vertebrates from infections and process of rejection of organ transplants, partly because of increasing interest of genes under selection rather than neutral markers. As we saw at the beginning of this part of the thesis, pathogen mediated selection and sexual selection can rule selection at MHCs. Individuals are expected to differ in their mate preferences both for assortative or more often for disassortative patterns of choice. This may lead to produce more or less pathogen resistance in progeny. Especially in species that occupy densely human inhabited areas, habitat fragmentation can result in decreased effective population size and concurrent increase in the rate of inbreeding.

In this paper we used a cohort of wolf non-invasive samples from Italian wolf population to identify three class II MHC haplotypes, and test them for random mating and allele sharing in order to detect MHC mating scheme. Combining analyses at neutral and adaptive genetic loci provides an efficient tool for evaluating the importance of various selective mechanisms, such as mate choice.

PAPER III

Galaverni, Marco, Romolo Caniglia, Pietro Milanesi, **Silvana Lapalombella**, Elena Fabbri, and Ettore Randi. 2015. "**Choosy Wolves ? Heterozygote Advantage But No Evidence of MHC-Based Disassortative Mating**" 1–9. doi:10.5061/dryad.t6n8h.

Authors' Contributions:

Authors' contributions are listed in alphabetical order. Conceived and designed the experiments: RC MG ER. Performed the experiments: RC MG SL. Analysed the data: RC MG PM SL. Contributed reagents/ materials/analysis tools: EF RC ER. Drafted the paper: RC MG SL ER. All authors revised the manuscript and gave final approval for publication.

Journal of Heredity Advance Access published November 25, 2015

American Genetic Association

Journal of Heredity, 2015, 1–9
doi:10.1093/jhered/esv090
Original Article

OXFORD

Original Article

Choosy Wolves? Heterozygote Advantage But No Evidence of MHC-Based Disassortative Mating

Marco Galaverni, Romolo Caniglia, Pietro Milanesi, Silvana Lapalombella, Elena Fabbri, and Ettore Randi

From the Laboratorio di genetica, ISPRA, via Cà Fornacetta 9, 40064 Ozzano dell'Emilia (Bologna), Italy (Galaverni, Caniglia, Milanesi, Fabbri, and Randi); Department of Biological, Geological and Environmental Sciences (BiGeA), University of Bologna, Via Selmi 3, 40126 Bologna, Italy (Lapalombella); and Department 18/Section of Environmental Engineering, Aalborg University, 9220 Aalborg, Denmark (Randi).

Address correspondence to Marco Galaverni at the address above, or e-mail: marco.galaverni@isprambiente.it.

Data deposited at Dryad: http://dx.doi.org/doi:10.5061/dryad.t6n8h

Received July 15, 2015; First decision October 26, 2015; Accepted October 27, 2015.

Corresponding Editor: Robert Wayne

Abstract

A variety of nonrandom mate choice strategies, including disassortative mating, are used by vertebrate species to avoid inbreeding, maintain heterozygosity and increase fitness. Disassortative mating may be mediated by the major histocompatibility complex (MHC), an important gene cluster controlling immune responses to pathogens. We investigated the patterns of mate choice in 26 wild-living breeding pairs of gray wolf (*Canis lupus*) that were identified through noninvasive genetic methods and genotyped at 3 MHC class II and 12 autosomal microsatellite (STR) loci. We tested for deviations from random mating and evaluated the covariance of genetic variables at functional and STR markers with fitness proxies deduced from pedigree reconstructions. Results did not show evidences of MHC-based disassortative mating. Rather we found a higher peptide similarity between mates at MHC loci as compared with random expectations. Fitness values were positively correlated with heterozygosity of the breeders at both MHC and STR loci, whereas they decreased with relatedness at STRs. These findings may indicate fitness advantages for breeders that, while avoiding highly related mates, are more similar at the MHC and have high levels of heterozygosity overall. Such a pattern of MHC-assortative mating may reflect local coadaptation of the breeders, while a reduction in genetic diversity may be balanced by heterozygote advantages.

Subject areas: Reproductive strategies and kinship analysis; Molecular adaptation and selection
Key words: *Canis lupus*, fitness, major histocompatibility complex, mate choice, microsatellites, sexual selection

Mate choice is one of the main mechanisms through which sexual selection influences genetic diversity, inbreeding avoidance and adaptability in wild populations (Tregenza and Wedell 2000). The major histocompatibility complex (MHC) is viewed as a key element in mate choice (Eizaguirre et al. 2009). MHC genes encode for glycoprotein receptors that trigger biological pathways in response to pathogens and infectious diseases. MHC proteins may also affect the composition of the bacterial flora of the host and the pleasantness of individual

odor cues. In this way, MHC peptides can be perceived via olfaction or pheromone detection (Brennan and Kendrick 2006; Spehr et al. 2006; Janeš et al. 2010; Sturm et al. 2013) and can function as signals of similarity between individuals, possibly driving active mate choice.

The effects of MHC-mediated signals in mate choice have been hypothesized in several sexual selection models. In the inbreeding avoidance hypothesis (Potts and Wakeland 1990), selection should favor unrelated parents, whose offspring will not be exposed to inbreeding depression. Parallel, heterozygote individuals can have higher fitness (heterozygote advantage) due to higher resistance to multiple parasites (Hedrick 2012). In the moving target model (Penn and Potts 1999), sexual selection could enhance the ability of the host to keep its defenses up-to-date with the continuing evolution of the "parasite weapons" thanks to a rare-allele advantage (Landry et al. 2001). In the honest signal model (Hamilton and Zuk 1982), MHC may be linked to other secondary sexual characteristics that indicate the levels of pathogen resistance and the general fitness of potential mates (Ditchkoff et al. 2001). These 3 hypotheses have been documented in a number of species (Kamiya et al. 2014), including humans (Wedekind et al. 1995; Havlicek and Roberts 2009).

However, other studies failed to detect any deviation from random mating (Huchard et al. 2010; Yang et al. 2014; Kuduk et al. 2014) and a number of works also reported cases of assortative (hereafter intended as positive-assortative) matings (Sin et al. 2015 and references therein). The theoretical models to explain assortative matings assume the reduction of outbreeding (e.g., in hybrid zones between divergent populations, where selection can act against genetic incompatibility; Tregenza and Wedell 2000), or the maintenance of allele compatibility or optimal allele combinations linked to local adaptation (Neff 2004). Preferences for partners with intermediate levels of dissimilarity can otherwise be expected to balance the costs of expressing numerous MHC alleles, which can lead to an increased chance of autoimmune diseases (Penn and Potts 1999), representing an intermediate case between disassortative and assortative mating.

Gray wolves (*Canis lupus*) offer an interesting case to study interactions between MHC variability and mate choice due to their complex social structure. The classical vision of a dominance hierarchy, once used to describe the relationships within a pack, was typically derived from observations of large groups in captivity and is mostly overcome (Packard 2003). In smaller, wild-living packs, family-like relationships can better illustrate the interactions among pack members, which are regulated by a combination of leadership skills, individual temperament, and experiences (Packard 2003). However, the mating system in wolves is strictly monogamous (Geffen et al. 1996), with a single pair that reproduces in a pack once a year (Mech and Boitani 2003). Polygamy and multiple litters per year are rare exceptions (Vonholdt et al. 2008). Breeding pairs can establish in newly colonized territories by loner or dispersing individuals, whereas in already existing packs one or both mates can be replaced by immigrant wolves, or by their own progeny (Vonholdt et al. 2008; Caniglia et al. 2014). Active inbreeding avoidance has been described within, but not across, packs (Geffen et al. 2011). Moreover, genetic diversity in wolf populations can also be maintained by heterozygote advantage or balancing selection (Bensch et al. 2006; Hagenblad et al. 2009).

The genetic organization of the canine MHC is well known (Kennedy et al. 2001; Berggren and Seddon 2005; Wayne and Ostrander 2007; Yuhki et al. 2007). MHC variability has been investigated in domestic and wild canids (Hedrick et al. 2002; Angles et al. 2005; Kennedy et al. 2007b; Fliegner et al. 2008; Galaverni et al. 2013), where a few cases of MHC-mediated resistance to pathogen outbreaks have been described (Hedrick et al. 2003; Kennedy et al. 2011). However, the role of MHC variation on mate choice in gray wolves yet has to be explored, also because obtaining data on wild-living individuals is no easy task. The extensive use of combined field and molecular monitoring tools is limited to a few cases (Vonholdt et al. 2008; Stahler et al. 2013). Nonetheless, the possibility to derive good-quality genetic data even from noninvasive samples can be successfully exploited (Yang et al. 2014).

In this study, we used wolf genotypes and pack genealogies, obtained in a long-term monitoring project based on extensive noninvasive genetic sampling, which allowed to infer the most likely reproductive parent pairs and their offspring (Caniglia et al. 2014). These wolves were genotyped at both putatively nonfunctional microsatellites (STR) and functional MHC class II loci, and fitness proxies were deduced from the reconstructed pedigrees. We aimed to identify the prevalent mate choice patterns, as well as to investigate the relations between MHC and STR variability, and individuals' fitness in wild-living wolf packs. Expecting high genetic diversity (Seddon and Ellegren 2004), inbreeding avoidance (Adams et al. 2011; Geffen et al. 2011) and heterozygote advantage or balancing selection (Bensch et al. 2006; Hagenblad et al. 2009), we assumed that the prevalent mate choice patterns can be aimed to select the maximum MHC diversity and dissimilarity between reproductive pack partners.

Materials and Methods

Sampling and Pack Reconstructions

We used DNA samples that were collected during an extensive noninvasive genetic monitoring program of the Italian wolf population carried out in the northern Apennines (Italy) from 2001 to 2011 (Caniglia et al. 2014). Biological samples (n = 5065, 99% scats) were genotyped at 12 nonfunctional and putatively neutral autosomal STRs, which allowed to identify 414 distinct wolf individual genotypes. Familial relationships were identified after Caniglia et al. (2014) using a maximum-likelihood approach implemented in the software Colony 2.0 (Wang and Santure 2009). All the genotypes that were sampled in restricted ranges (<100 km^2) at least 4 times and for periods longer than 24 months were selected. Their spatial distributions were determined by 95% kernel analysis using the adehabitathr package for R (Calenge 2006) and mapped in ArcGis 10.0 (ESRI). According to spatial overlaps, individuals were split into distinct groups that might correspond to packs, for which parentage analyses were performed. All the individuals sampled in the first year of sampling and more than 4 times in the same area were considered as candidate parents of each group. All the individuals collected within the 95% kernel spatial distribution of each pack and in a surrounding buffer area of approximately 17-km radius from the kernel were considered as candidate offspring (Caniglia et al. 2014). Colony was run with allele frequencies and PCR error rates as estimated from all the genotypes, assuming a 0.5 probability of including fathers and mothers in the candidate parental pairs. Genealogies were then compared to those obtained by an open parentage analysis, using all the males and females as candidate parents, and all the wolves sampled in the study area as candidate offspring. The best maximum-likelihood genealogies reconstructed by Colony were then compared also with those obtained by a likelihood approach in Cervus (Kalinowski et al. 2007), based on the Mendelian inheritance of the alleles, accepting only parent–offspring combinations with at most 1/24 allele incompatibilities, and father–son combinations

with no incongruities at Y-STR haplotypes. Whenever available, additional field information such as snow-tracking, wolf-howling, camera-trapping, and occasional direct observations were used to evaluate the reliability of the inferred pack structure and locations. Reliable pedigrees were identified for 26 packs, which sometimes extended over multiple years and generations (Caniglia et al. 2014). None of the alternative genealogies was supported by sampling dates, frequency, or location. Moreover, from 2/24 to 5/24 allele incompatibilities and incongruities at Y-haplotypes occurred in 93 (95%) of 98 father–son combinations of the alternative trios with significant natural log of likelihood ratio scores (for details see also Supporting Information S7 from Caniglia et al. 2014). Multiple litters per year in a pack, or extra-pair reproductions were never detected in any pack. Details on the metrics of the STR loci used to genotype the samples are reported in Supplementary Table S1 online.

For each breeding pair, we then calculated an overall fitness proxy derived from pedigree information, namely the total number of offspring of a pair (total offspring, TO), which reflects the number of pups that survived until they were detected through noninvasive genetic sampling. In addition, to better highlight different contributions to fitness, we also considered the 2 main components of TO, namely the number of years a breeding pair was sampled (hereafter: years as reproducer, YR) and the average number of identified offspring per year, or litter size per year (LSY).

We then computed the correlations between all the pairs of fitness proxies (Pearson's correlation test) and tested for the normal distribution of data with a Shapiro–Wilk normality test.

MHC Amplification and Genotyping

For each of the 63 identified breeders (32 males, 31 females; Caniglia et al. 2014), the best available DNA sample was PCR-amplified and sequenced at the second exon of 3 class II MHC genes: DRB1, DQA1, and DQB1, using the intronic and locus-specific primers DRB1F (5′ –ccg tcc cca cag cac att tc– 3′) and DRB1R (5′ –tgt gtc aca cac ctc agc acc a– 3′; Hedrick et al. 2002, after Kennedy et al. 1998); DQAin1 (5′ –taa ggt tct ttt ctc cct ct– 3′) and DQAin2 (5′ –gga cag att cag tga aga ga– 3′); DQB1B (5′ –ctc act ggc ccg gct gtc tc– 3′), and DQBR2 (5′ –cac ctc gcc gct gca acg tg– 3′; Kennedy et al. 2006, after Wagner et al. 1996). Amplifications were carried out in a 10-μL mix, including 2 μL genomic DNA solution, 1 μL bovine serum albumin (2%), and 0.2 μL of each primer (10 μM) plus 0.25 units of Taq, at conditions specific for each primer pair (Galaverni et al. 2013). PCR products were purified with Exo/SAP and sequenced in both directions using BigDye Terminator 1.1, according to the manufacturer's protocol. Sequences were analyzed in an automated sequencer ABI 3130XL, corrected and aligned with the software Seqscape 2.5, using the sequences DLA-DRB1*03101 (AF336108.1), DLA-DQA1*014012 (AJ316220.1), and DLA-DQB1*05601 (FM246843.1) as references.

MHC alleles were reconstructed in DnaSp 5.10 (Librado and Rozas 2009) using the software PHASE (Stephens and Donnelly 2003), which allows to avoid the cloning step, applying the "recombination" model (–MR0) and running 1000 iterations after 100 burn-ins (Berggren and Seddon 2008; Galaverni et al. 2013). Alleles were matched to sequences available in GenBank via Blastn (Johnson et al. 2008) and also compared with those included in the Immuno Polymorphism-MHC Database (IPD; http://www.ebi.ac.uk/ipd/mhc/dla/index.html; Robinson et al. 2010). Three-locus MHC haplotypes were then identified with a subtractive approach described in Kennedy et al. (2007a) and computationally verified in PHASE (Berggren and Seddon 2008; Galaverni et al. 2013).

The MHC genotypes of 5 individuals were independently derived from 2 different fecal samples in order to check for robustness in the sequencing and genotyping procedure. To further verify the congruent reconstruction of the alleles through trio comparisons, we also sequenced and phased MHC alleles in 15 known offspring of the analyzed breeders. In addition, 22 nonreproductive pack members unrelated to the breeding pair (hereafter "adoptees") were included for testing for preferences toward alternative mates.

Random Mating and Allele Sharing

We computed the mean value of the asymmetrical Queller and Goodnight's estimator of relatedness (r) between the members of each observed breeding pair at the 3 MHC loci [$r_{(MHC)}$] in GenAlEx (Peakall and Smouse 2006). Similarly, r was computed also at the 12 STRs [$r_{(STR)}$] to control for genome-wide effects.

We then empirically assessed the probability of deviations from random mating by comparing the breeding pairs' relatedness values, both at MHC and STR loci, to those between the members of all the potential pairs of male–female breeders. This was done by a permutation procedure implemented in PERM 1.0 (Duchesne et al. 2006), with 5000 permutations of the breeding pair members repeated for 10 iterations.

Additionally, we tested for deviations from random mating probabilities by observing the levels of allele sharing at the MHC in the breeding pairs, both at each locus (in these cases considering a Bonferroni-corrected alpha of 0.05/3, or ~0.017, to adjust for multiple comparisons) and for the 3-locus haplotypes. We hypothesized an excess of pairs where the mates shared a lower number of MHC alleles than expected by random. At each locus, there are 3 possible classes of matings: pairs where the members share no alleles (e.g., 01 01×12 12), 1 allele (01 01×01 12, or 01 16×01 12), or both (01 01×01 01). The expected frequency of a given mating (mating probability, P_M) was then calculated as $\Sigma P_{ij(f)} P_{ij(m)}$, where $P_{ij(f)}$ and $P_{ij(m)}$ are the observed frequencies of the genotype with alleles i and j in females and males, respectively (Hedrick and Black 1997). We then evaluated the difference between the observed and expected number of mating events in these 3 classes (χ^2 test) at each locus and at the 3-locus haplotypes. To avoid inflating the frequency of alleles that could be more frequent among breeders than in the whole population, genotypic allele frequencies were computed only across breeding individuals. We then carried out a power analyses to verify the robustness of the results and the possibility to detect significant differences between groups (Faul et al. 2007). The power of a statistical test is the complement of β, which denotes the Type II (or beta-error) probability of falsely retaining an incorrect null hypothesis. Thus, the power of a test ranges between 1 and 0, which indicate high or low statistical power, respectively (Faul et al. 2007).

Functional Similarity Analyses

In order to better incorporate the levels of functional similarity in our analyses (Landry et al. 2001; Miller et al. 2009), we calculated the mean number of amino acid (AA) differences between female and male alleles at each MHC locus, and in total across the 3 loci. AA differences were then compared in observed versus potential pairs of breeders with the same permutation procedure used above, expecting a higher difference within actual pairs in case of disassortative matings.

To further test for preferences toward alternative mates, we finally compared AA differences between observed mates to those between a given breeder and any unrelated wolf of the opposite sex that was eventually present in the pack in the same period (adoptee),

but who had no access to reproduction (Caniglia et al. 2014), with a Bonferroni-corrected *t*-test.

Heterozygote Advantage and Correlations to Fitness Proxies

We correlated the genetic measures of diversity and dissimilarity with the fitness proxies derived from pedigree information.

First, the observed heterozygosity of each breeder and the mean heterozygosity for each breeding pair were calculated in GenAlEx, both at MHC [$Ho_{(MHC)}$] and STR loci [$Ho_{(STR)}$]. Then we compared the average fitness values in MHC-heterozygous versus MHC-homozygous breeders and between breeders with higher- versus lower-than-average $Ho_{(STR)}$ through a Welch 2-sample *t*-test.

Finally, we related each fitness proxy to the main genetic metrics deduced from the genotypes of the breeding pair members: the relatedness estimates $r_{(STR)}$ and $r_{(MHC)}$, also re-computed according to Lynch and Ritland (1999) [$r_{(STR\text{-}LR)}$ and $r_{(MHC\text{-}LR)}$], the mean heterozygosity of the breeders $Ho_{(MHC)}$ and $Ho_{(STR)}$, the mean heterozygosity of the breeders calculated across all the polymorphic sites at the MHC loci [$Ho_{(MHC\text{-}SNPs)}$], the average number of AA differences between mates (AA_{dist}) and their normalized values ($AA_{dist\text{-}N}$). These variables were used as fixed factors in generalized linear mixed models (GLMM), while geographic locations in the study area ("Eastern," "Central," and "Western"), elevations (higher vs. lower than 800 m above sea level) and year of breeding (from 2001 to 2009, grouped in 3 classes: 2001–2003; 2004–2006; 2007–2009) were considered as random factors. We also entered quadratic terms for the genetic variables to test if they could show higher relation with each fitness proxy, as expected under an optimal allele diversity model (Thoss et al. 2011).

Model selection was based on the Akaike's information criterion corrected for sample size (AICc; Akaike 1974), where models that better fit the data produce lower AICc values. According to Anderson et al. (2001), multi-model inference was performed for models with ΔAICc < 2. All these analyses were performed in R 3.2.0 (R Development Core Team 2009; www.R-project.org).

Data Archiving

In fulfillment of data archiving guidelines (Baker 2013), we have deposited the primary data underlying these analyses as follows:

- Genetic metrics for the microsatellite used for genotyping: Supplementary Table S1 online.
- Allele frequency for the 3 MHC loci analyzed and their combined haplotypes: Supplementary Tables S2 and S3 online, respectively.
- MHC genotypes for all the breeders analyzed: Dryad.

Results

Sequencing Success and MHC Variability

We successfully sequenced and phased the alleles at all the 3 MHC loci in DNA samples from 47/63 identified wolf breeders, belonging to 19 packs and 26 breeding pairs, plus 10/15 of their offspring and 10/22 adoptees (67% of the tested individuals, overall). All the genotypes of the 5 individuals that were reconstructed from 2 independent fecal samples regularly matched one another (100%). We detected 7 alleles at DRB1, 6 at DQA1, and 7 at DQB1 (Supplementary Table S2 online), combined into 16 different MHC haplotypes (Supplementary Table S3 online). We were able to reconstruct 10 full parent-offspring trios, in all cases confirming the Mendelian inheritance of the alleles.

Random Mating and Allele Sharing

The observed relatedness between breeders at microsatellite loci [$r_{(STR)} = 0.026 \pm 0.237$] was not significantly deviating from random mating expectations [$r_{(STR)} = -0.023 \pm 0.252$; $P = 0.218 \pm 0.005$), as well as the relatedness at MHC loci [$r_{(MHC)} = 0.066 \pm 0.688$ versus -0.090 ± 0.641, respectively, $P = 0.080 \pm 0.004$].

Similarly, no signals of disassortative matings were detected from the allele sharing analysis. Actually, we found a nonsignificant excess of cases where the pair members shared one or both alleles compared to random mating expectations at each locus and MHC haplotype (Table 1). Results were associated to a high statistical power for the test at DQA1, but just to a moderate power for DRB1, DQB1, and haplotypes (Table 1), indicating that for these loci a larger sample size could have provided more significant results.

Functional Similarity

Up to 22 variable AA sites were found at DRB1, 5 at DQA1, and 22 at DQB1 (68%, 20%, and 59% of which fall within putative antigen-binding sites, respectively).

The trend toward an excess of allele sharing was stronger when looking at the mean number of pairwise AA differences between mates at the MHC genes. The number of AA differences was significantly lower between actual mates than among random combinations of breeders at DQA1 ($P = 0.030$) and DQB1 ($P = 0.026$), and in total across the 3 loci ($P = 0.039$, Table 2), as deduced from PERM permutations.

Similarly, the comparison with the 10 alternative potential mates from the same packs showed that actual mates differed on average for 12.3 ± 7.1 AA in total across loci versus 17.6 ± 10.2 AA differences for alternative mates, although not significantly (*t*-test, $P = 0.097$).

Table 1. Probabilities of allele sharing between observed and random mates

Allele combination	DRB1		DQA1		DQB1		Haplotypes	
	obs.	exp.	obs.	exp.	obs.	exp.	obs.	exp.
0 shared	8	12.8	4	5.8	10	13.3	11	15.2
1 shared	14	10.6	10	11.9	12	10.4	13	9.3
2 shared	4	2.6	12	8.3	4	2.3	2	1.5
χ^2 *P* value	0.164		0.287		0.299		0.247	
Statistical power	0.406		0.999		0.563		0.509	

The number of breeding pairs ($n = 26$) sharing 0, 1, or 2 alleles at each MHC locus and haplotype (obs.) was compared to what expected under a random chance of mating based on the breeders' allele frequencies (exp.). The *P* values were computed by a χ^2 test.

Heterozygote Advantage and Correlations to Fitness Proxies

Results of the correlations between our fitness proxies showed that the values of litter size per year (LSY) were not significantly correlated to the years as reproducers (YR; Pearson's correlation c = 0.09, P = 0.561), whereas the total offspring (TO) was significantly correlated to both YR (c = 0.76, P = 1.024e^{-09}) and LSY (c = 0.63, P = 3.652e^{-06}), as expected. When comparing MHC and genome-wide STR diversity, we found that $Ho_{(MHC)}$ and $Ho_{(STR)}$ were not significantly correlated (Pearson's correlation c = −0.08, P = 0.598).

Breeders that were heterozygote at MHC haplotypes and DQB1 had significantly higher LSY than the homozygotes, on average (respectively, 2.83 ± 1.28 versus 1.79 ± 0.95, P = 0.003; 2.92 ± 1.34 versus 1.96 ± 0.91, P = 0.004; t-test). No significant differences were found for other fitness proxies and loci (Figure 1). Breeders with higher-than-average heterozygosity at STR loci had also higher values of TO (7.2 ± 5.5 versus 4.7 ± 4.4) and LSY (2.9 ± 1.2 versus 2.4 ± 1.3), but these differences were not significant per se (t-test, P = 0.072 and P = 0.156, respectively).

By applying GLMMs, we identified a single model with ΔAICc < 2 for each fitness proxy, characterized by the same 4 explanatory genetic variables, namely the average heterozygosity $Ho_{(STR)}$ and $Ho_{(MHC)}$ of the pair members, and their relatedness $r_{(STR)}$ and $r_{(MHC)}$ (Figure 2, Table 3). The model with the lowest AICc for YR (AICc = 82.63; R^2 = 0.35) showed a significant and negative effect of $r_{(STR)}$ (β = −2.795, P = 0.0056), while LSY (AICc = 85.68; R^2 = 0.39) was mostly explained by $Ho_{(STR)}$ and $Ho_{(MHC)}$, both with positive effects (β = 1.631, P = 0.0372; and β = 7.089, P = 0.0115, respectively). Finally, we identified also for TO only one model with ΔAICc < 2 (AICc = 139.38; R^2 = 0.32) in which $r_{(STR)}$ showed the most significant, negative effect (β = −8.881, P = 0.0273), confirming its correlation to our fitness proxies.

Table 2. Average number of pairwise amino acid differences at 3 MHC loci between the members of the 26 breeding pairs compared to that expected between all the possible random pairs of breeders (after 10×5000 permutations in PERM; upper part). Average number of pairwise amino acid differences between 10 breeders and their actual mates, versus the same 10 breeders and 10 alternative mates, defined as unrelated individuals present in their same pack, but not reproducing (t-test, Bonferroni corrected; lower part). Significant values are indicated with an asterisk.

	Mean number of amino acid differences			
	DRB1	DQA1	DQB1	Total
Within observed pairs	6.6	1.5	6.7	14.8
Within random pairs	7.0	1.7	7.8	16.6
PERM P value	0.221	0.030*	0.026*	0.039*
Actual mates	5.1	1.9	5.2	12.3
Alternative mates	8.3	2.6	6.7	17.6
t-Test P	0.079	0.067	0.204	0.097

Discussion

Contrary to sexual selection models assuming advantages for disassortative matings (Kamiya et al. 2014), our study did not show

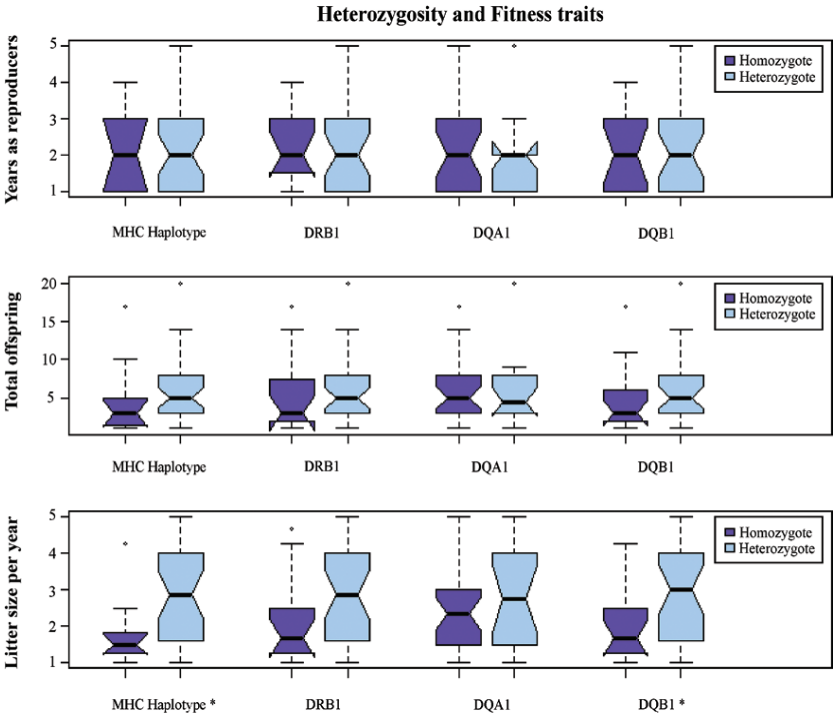


Figure 1. Comparison of the average values of fitness between homozygote and heterozygote breeding wolves at MHC haplotypes, DRB1, DQA1, and DQB1 loci. Outlier values are indicated as dots outside the boxes. Notches that do not overlap represent a 95% probability for the median values to be different. Significant differences are indicated with an asterisk next to the label.

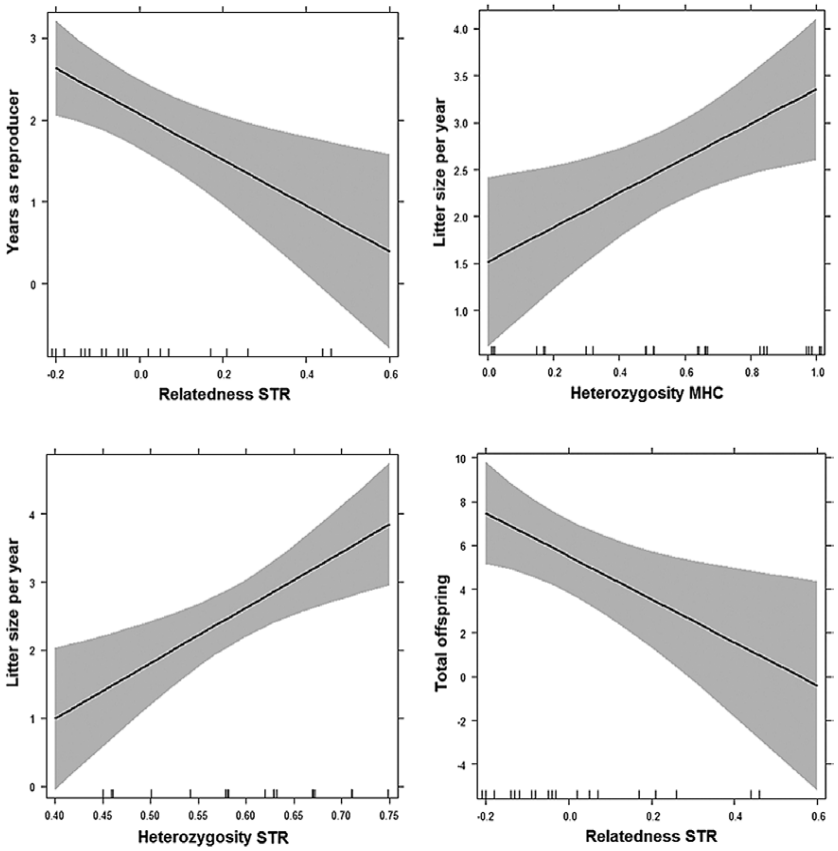


Figure 2. Effect plots of the multiple regressions between fitness proxies of breeding wolves (*Y* axis) and their significant explanatory genetic variables (*X* axis), as deduced from the best generalized linear mixed models (Table 3). Tic marks on horizontal axes indicate observations and shaded areas underlie 95% confidence intervals. The number of years as reproducers (top-left) was significantly correlated to the relatedness at the 12 microsatellites, with a negative effect. The average litter size per year (top-right and bottom-left) was positively correlated to the heterozygosity at MHC and STR loci, whereas the total offspring (bottom-right) was correlated to the relatedness at STR loci, with a negative effect.

Table 3. Generalized linear mixed models correlating 3 fitness proxies (YR = years as reproducers, LSY = litter size per year, TO = total offspring) to relatedness (*r*) and heterozygosity (Ho) of the breeding pair members at the 12 microsatellites and 3 MHC loci

Model	$Y \sim Ho_{(MHC)} + Ho_{(STR)} + r_{(MHC)} + r_{(STR)}$						
Fitness proxy	AICc	R^2	Variable	β	SE	*t*	*P*
YR	82.63	0.35	Intercept	3.334	1.509	2.209	0.0389*
			$Ho_{(MHC)}$	0.494	0.677	0.731	0.4741
			$Ho_{(STR)}$	−2.664	2.361	−1.128	0.2724
			$r_{(MHC)}$	0.491	0.374	1.311	0.2047
			$r_{(STR)}$	−2.795	0.902	−3.099	0.0056**
LSY	85.68	0.39	Intercept	−2.583	1.629	−1.586	0.1284
			$Ho_{(MHC)}$	1.631	0.731	2.232	0.0372*
			$Ho_{(STR)}$	7.089	2.547	2.783	0.0115*
			$r_{(MHC)}$	0.633	0.404	1.568	0.1325
			$r_{(STR)}$	−1.052	0.973	−1.081	0.2928
TO	139.38	0.32	Intercept	−3.737	6.237	−0.599	0.5559
			$Ho_{(MHC)}$	4.567	2.799	1.632	0.1184
			$Ho_{(STR)}$	10.981	9.756	1.126	0.2737
			$r_{(MHC)}$	2.862	1.547	1.851	0.0591
			$r_{(STR)}$	−8.881	3.728	−2.382	0.0273*

AIC, Akaike's Information Criterion; df, degrees of freedom.
Significance "**" = 0.01 and "*" = 0.05.

evidences of MHC-based disassortative mate choice in 26 breeding pairs from 19 packs of wild-living gray wolves from the Italian population. Conversely, we found that wolves preferentially chose mates who had less amino acid differences at MHC class II loci than expected under random mating. Although not significantly, an excess of MHC similarity appeared also when comparing the number of AA differences between observed versus alternative possible mates, or when considering the levels of allele sharing at MHC loci between observed breeders. Overall, these results indicate the absence of an MHC-based disassortative pattern, but are rather compatible with signals of positive-assortative mating. Although the most popular hypotheses on the benefits of sexual selection assume a disassortative mate choice (Piertney and Oliver 2006; Ejsmond et al. 2014), a number of studies failed to document such a scheme in wild and captive populations (with exhaustive examples presented in (Huchard et al. 2010). Assortative mating was conversely observed in a wild population of house sparrow, where mates shared more MHC alleles than expected by random (Bonneaud et al. 2006). A higher reproductive success for MHC-similar mates was documented in the tiger salamander (Bos et al. 2009). Assortative mating was also recently documented in the European badger, with a smaller amino acid distance between actual breeders compared with random mates at MHC class II loci (Sin et al. 2015). Benefits of an assortative mate choice have been so far demonstrated to reduce outbreeding in hybrid zones and limit autoimmune disorders caused by genetic incompatibilities (Tregenza and Wedell 2000), maximize the adaptation to local pathogen pressures (Sin et al. 2015) and avoid the disruption of linked and co-adapted genes (Roberts 2009). Moreover, Lewis (1998) argued that in social species a preferential association with MHC-similar individuals should decrease the probability of infection with unfamiliar pathogens that could be carried by immigrant individuals who do not share the same immunological variants. Consequently, choosing an MHC-similar mate would be directly beneficial for the breeders, that will more likely avoid interaction with potential disease carriers, as long as this does not increase the general inbreeding levels (Drury 2010). Parallel, inbreeding avoidance may not be necessary in expanding populations (Jamieson et al. 2009), such as the Italian wolf, where the probability to encounter related individuals outside the natal pack is relatively low (Geffen et al. 2011). In fact, when analyzing several wolf populations at neutral loci, Geffen et al. (2011) did not find any evidence of inbreeding avoidance between mates, except within natal groups. Moreover, inbreeding in wolves can be also prevented by other commonly adopted behaviors, such as long-distance dispersals (Caniglia et al. 2014).

However, when considering the correlation of MHC and STR diversity and dissimilarity of mates to fitness proxies deduced from pedigree data, we found a significant effect of the genome-wide relatedness $r_{(STR)}$ of mates on the number of years they reproduced (YR) and on the total offspring they produced (TO): the lower the relatedness, the higher the fitness proxy (Figure 2). These findings suggest that positive assortative mating at the MHC (highlighted by the lower AA difference between mates) may be balanced by a greater fitness of mates with a higher genome-wide dissimilarity (inbreeding avoidance).

Moreover, the average litter size per year (LSY) was positively correlated to the average heterozygosity of mates, both at the MHC and STRs, as also shown by the comparisons between MHC-homozygous versus heterozygous breeders (Figure 1). These results confirm the presence of a heterozygote advantage demonstrated in a number of studies on wolves and other vertebrate species, both at neutral (García-Navas et al. 2009) and MHC loci (Setchell and Huchard 2010; Thoss et al. 2011; Knafler et al. 2012; Niskanen et al. 2014). Despite the possible presence of MHC-based assortative mating schemes, MHC polymorphism can thus be maintained via a heterozygote advantage hypothesis, by providing resistance to a wider range of diseases or pathogens and possibly resulting in a higher individual survival (Hedrick 2012).

Such a panel of potential benefits likely reflects the social structure of wolves, the pack, where a single pair of well-fit mates can reproduce each year, with few exceptions (Vonholdt et al. 2008). This mechanism allows for a constant adaptation to the environment, which includes both its resources and its pathogens. Therefore, MHC-based assortative mating could be explained by a local co-adaptation of the breeders and would allow to constantly keep up with the local environmental changes, including pathogens (Penn and Potts 1999), while reducing the interactions with possibly sick or pathogen-carrying individuals (Drury 2010). This could be especially true in expanding populations, where a significant portion of packs is founded by dispersing individuals that establish themselves in new territories (Vonholdt et al. 2008; Caniglia et al. 2014). Of course, our study would greatly benefit from a wider sampling of the studied population. However, obtaining genetic and reproductive information on elusive carnivore species such as the wolf is far from trivial, especially in populations where active management strategies or trapping of individuals are not applied. Thus, we sought to exploit a large source of non-invasive samples (Caniglia et al. 2014), although fecal DNA has been seldom used in MHC gene sequencing (Yang et al. 2014). Our study represents the first application of non-invasive genetic sampling to investigate the role of MHC on sexual selection in canids. Our procedure showed a good MHC genotyping success (comparable to values reported in other noninvasive genetic studies on MHC; Yang et al. 2014) and reliability (as confirmed by the double-sequencing controls), and promises to be extended to future large-scale monitoring projects.

In conclusion, although contrasting with the most common research results reporting dissimilar mating schemes, our models were able to explain part of the variance of multiple fitness proxies in an expanding wolf population and to elucidate the genetic bases driving the observed mating patterns.

This type of information could be also useful to increase the chance of success of active management actions such as reintroduction projects or the release of rescued individuals.

Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/

Funding

Italian Ministry of Environment; the Emilia-Romagna Region (thanks to Dr. M. L. Bargossi and Dr. M. L. Zanni) sponsored the genotyping work at the Laboratorio di Genetica, Istituto Superiore per la Protezione e la Ricerca Ambientale; and University of Bologna (to M.G. and S.L.).

Acknowledgments

We kindly thank Phil Hedrick (University of Arizona), Edoardo Velli (University of Roma 3), and Aritz Ruiz-Gonzales (University of the Basque Country) for their useful comments, and many collaborators that contributed to collect the wolf samples in Italy. We also thank the Associated Editor and

2 anonymous reviewers for their useful comments that helped to considerably improve the manuscript.

References

Adams JR, Vucetich LM, Hedrick PW, Peterson RO, Vucetich JA. 2011. Genomic sweep and potential genetic rescue during limiting environmental conditions in an isolated wolf population. *Proc Biol Sci*. 278:3336–3344.

Anderson DR, Link WA, Johnson DH, Burnham KP. 2001. Suggestions for presenting the results of data analyses. *J Wildl Manag*. 65:373–378.

Angles JM, Kennedy LJ, Pedersen NC. 2005. Frequency and distribution of alleles of canine MHC-II DLA-DQB1, DLA-DQA1 and DLA-DRB1 in 25 representative American Kennel Club breeds. *Tissue Antigens*. 66:173–184.

Baker CS. 2013. Journal of heredity adopts joint data archiving policy. *J Hered*. 104:1.

Bensch S, Andrén H, Hansson B, Pedersen HC, Sand H, Sejberg D, Wabakken P, Akesson M, Liberg O. 2006. Selection for heterozygosity gives hope to a wild population of inbred wolves. *PLoS One*. 1:e72.

Berggren KT, Seddon JM. 2005. MHC promoter polymorphism in grey wolves and domestic dogs. *Immunogenetics*. 57:267–272.

Berggren KT, Seddon JM. 2008. Allelic combinations of promoter and exon 2 in DQB1 in dogs and wolves. *J Mol Evol*. 67:76–84.

Bonneaud C, Chastel O, Federici P, Westerdahl H, Sorci G. 2006. Complex Mhc-based mate choice in a wild passerine. *Proc Biol Sci*. 273:1111–1116.

Bos DH, Williams RN, Gopurenko D, Bulut Z, DeWoody JA. 2009. Condition-dependent mate choice and a reproductive disadvantage for MHC-divergent male tiger salamanders. *Mol Ecol* 18:3307–3315.

Brennan PA, Kendrick KM. 2006. Mammalian social odours: attraction and individual recognition. *R Soc Philos Trans Biol Sci*. 361:2061–2078.

Calenge C. 2006. The package AdeHabitat for the R software: a tool for the analysis of space and habitat use by animals. *Ecol Model*. 197:516–519.

Caniglia R, Fabbri E, Galaverni M, Milanesi P, Randi E. 2014. Noninvasive sampling and genetic variability, pack structure, and dynamics in an expanding wolf population. *J Mammal*. 95:41–59.

Ditchkoff SS, Lochmiller RL, Masters RE, Hoofer SR, Van Den Bussche RA. 2001. Major-histocompatibility-complex-associated variation in secondary sexual traits of white-tailed deer (*Odocoileus virginianus*): evidence for good-genes advertisement. *Evolution*. 55:616–625.

Drury JP. 2010. Immunity and mate choice: a new outlook. *Anim Behav*. 79:539–545.

Duchesne P, Étienne C, Bernatchez L. 2006. PERM: a computer program to detect structuring factors in social units. *Mol Ecol Notes*. 6:965–967.

Ejsmond MJ, Radwan J, Wilson AB. 2014. Sexual selection and the evolutionary dynamics of the major histocompatibility complex. *Proc R Soc B Biol Sci*. 281:1662.

Eizaguirre C, Yeates SE, Lenz TL, Kalbe M, Milinski M. 2009. MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Mol Ecol*. 18:3316–3329.

Faul F, Erdfelder E, Lang AG, Buchner A. 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 39:175–191.

Fliegner RA, Holloway SA, Lester S, McLure CA, Dawkins RL. 2008. Evaluation of the class II region of the major histocompatibility complex of the greyhound with the genomic matching technique and sequence-based typing. *Tissue Antigens*. 72:131–136.

Galaverni M, Caniglia R, Fabbri E, Lapalombella S, Randi E. 2013. MHC variability in an isolated wolf population in Italy. *J Hered*. 104:601–612.

García-Navas V, Ortego J, Sanz JJ. 2009. Heterozygosity-based assortative mating in blue tits (*Cyanistes caeruleus*): implications for the evolution of mate choice. *Proc Biol Sci*. 276:2931–2940.

Geffen E, Gompper ME, Gittleman JL, Luh H-K, MacDonald DW, Wayne RK. 1996. Size, life-history traits, and social organization in the Canidae: a reevaluation. *Am Nat*. 147:140–160.

Geffen E, Kam M, Hefner R, Hersteinsson P, Angerbjörn A, Dalèn L, Fuglei E, Norèn K, Adams JR, Vucetich J, *et al.* 2011. Kin encounter rate and inbreeding avoidance in canids. *Mol Ecol*. 20:5348–5358.

Hagenblad J, Olsson M, Parker HG, Ostrander EA, Ellegren H. 2009. Population genomics of the inbred Scandinavian wolf. *Mol Ecol*. 18:1341–1351.

Hamilton WD, Zuk M. 1982. Heritable true fitness and bright birds: a role for parasites? *Science*. 218:384–387.

Havlicek J, Roberts SC. 2009. MHC-correlated mate choice in humans: a review. *Psychoneuroendocrinology*. 34:497–512.

Hedrick P, Black F. 1997. HLA and mate selection: no evidence in South Amerindians. *Am J Hum Genet*. 61:505–511.

Hedrick PW. 2012. What is the evidence for heterozygote advantage selection? *Trends Ecol Evol*. 27:698–704.

Hedrick PW, Lee RN, Buchanan C. 2003. Canine parvovirus enteritis, canine distemper, and major histocompatibility complex genetic variation in Mexican wolves. *J Wildl Dis*. 39:909–913.

Hedrick PW, Lee RN, Garrigan D. 2002. Major histocompatibility complex variation in red wolves: evidence for common ancestry with coyotes and balancing selection. *Mol Ecol*. 11:1905–1913.

Huchard E, Knapp LA, Wang J, Raymond M, Cowlishaw G. 2010. MHC, mate choice and heterozygote advantage in a wild social primate. *Mol Ecol*. 19:2545–2561.

Jamieson IG, Taylor SS, Tracy LN, Kokko H, Armstrong DP. 2009. Why some species of birds do not avoid inbreeding: insights from New Zealand robins and saddlebacks. *Behav Ecol*. 20:575–584.

Janeš D, Klun I, Vidan-Jeras B, Jeras M, Kreft S. 2010. Influence of MHC on odour perception of 43 chemicals and body odour. *Cent Eur J Biol*. 5:324–330.

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden T. 2008. NCBI BLAST: a better web interface. *Nucleic Acids Res*. 36:W5–W9.

Kalinowski ST, Taper ML, Marshall TC. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol*. 16:1099–1106.

Kamiya T, O'Dwyer K, Westerdahl H, Senior A, Nakagawa S. 2014. A quantitative review of MHC-based mating preference: the role of diversity and dissimilarity. *Mol Ecol*. 23:5151–5163.

Kennedy LJ, Angles JM, Barnes A, Carter SD, Francino O, Gerlach JA, Happ GM, Ollier WE, Thomson W, Wagner JL. 2001. Nomenclature for factors of the dog major histocompatibility system (DLA), 2000: second report of the ISAG DLA Nomenclature Committee. *Anim Genet*. 32:193–199.

Kennedy LJ, Angles JM, Barnes A, Carmichael LE, Radford AD, Ollier WER, Happ GM. 2007a. DLA-DRBI, DQAI, and DQBI alleles and haplotypes in North American gray wolves. *J Hered*. 98:491–499.

Kennedy LJ, Barnes A, Short A, Brown JJ, Lester S, Seddon J, Fleeman L, Francino O, Brkljacic M, Knyazev S, *et al.* 2007b. Canine DLA diversity : 1. New alleles and haplotypes. *Tissue Antigens*. 42:272–288.

Kennedy LJ, Carter SD, Barnes A, Bell S, Bennett D, Ollier WE, Thomson W. 1998. Nine new dog DLA-DRB1 alleles identified by sequence-based typing. *Immunogenetics*. 48:296–301.

Kennedy LJ, Huson HJ, Leonard J, Angles JM, Fox LE, Wojciechowski JW, Yuncker C, Happ GM. 2006. Association of hypothyroid disease in Doberman Pinscher dogs with a rare major histocompatibility complex DLA class II haplotype. *Tissue Antigens*. 67:53–56.

Kennedy LJ, Randall DA, Knobel D, Brown JJ, Fooks AR, Argaw K, Shiferaw F, Ollier WER, Sillero-Zubiri C, Macdonald DW, *et al.* 2011. Major histocompatibility complex diversity in the endangered Ethiopian wolf (*Canis simensis*). *Tissue Antigens*. 77:118–125.

Knafler GJ, Clark JA, Boersma PD, Bouzat JL. 2012. MHC diversity and mate choice in the magellanic penguin, *Spheniscus magellanicus*. *J Hered*. 103:759–768.

Kuduk K, Babik W, Bellemain E, Valentini A, Zedrosser A, Taberlet P, Kindberg J, Swenson JE, Radwan J. 2014. No evidence for the effect of MHC on male mating success in the brown bear. *PLoS One*. 9:e113414.

Landry C, Garant D, Duchesne P, Bernatchez L. 2001. "Good genes as heterozygosity": the major histocompatibility complex and mate choice in Atlantic salmon (*Salmo salar*). *Proc Biol Sci*. 268:1279–1285.

Lewis K. 1998. Pathogen resistance as the origin of kin altruism. *J Theor Biol*. 193:359–363.

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

Lynch M, Ritland K. 1999. Estimation of pairwise relatedness with molecular markers. *Genetics*. 152:1753–1766.

Mech LD, Boitani L. 2003. Wolf social ecology. In: Mech LD, Boitani L, editors. *Wolves: behavior, ecology and conservation*. Chicago (IL): The University of Chicago Press. p. 1–34.

Miller HC, Moore JA, Nelson NJ, Daugherty CH. 2009. Influence of major histocompatibility complex genotype on mating success in a free-ranging reptile population. *Proc Biol Sci*. 276:1695–1704.

Neff BD. 2004. Stabilizing selection on genomic divergence in a wild fish population. *Proc Natl Acad Sci U S A*. 101:2381–2385.

Niskanen AK, Kennedy LJ, Ruokonen M, Kojola I, Lohi H, Isomursu M, Jansson E, Pyhäjärvi T, Aspi J. 2014. Balancing selection and heterozygote advantage in MHC loci of the bottlenecked Finnish wolf population. *Mol Ecol*. 23:875–889.

Packard JM. 2003. Wolf behavior: reproductive, social and intelligent. In: Mech LD, Boitani L, editors. *Wolves: behavior, ecology and conservation*. Chicago (IL): The University of Chicago Press. p. 35–65.

Peakall ROD, Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes*. 6:288–295.

Penn DJ, Potts WK. 1999. The evolution of mating preferences and major histocompatibility complex genes. *Evolution*. 153:145–164.

Piertney SB, Oliver MK. 2006. The evolutionary ecology of the major histocompatibility complex. *Heredity*. 96:7–21.

Potts WK, Wakeland EK. 1990. Evolution of diversity at the major histocompatibility complex. *Trends Ecol Evol*. 5:181–187.

Roberts SC. 2009. Complexity and context of MHC-correlated mating preferences in wild populations. *Mol Ecol*. 18:3121–3123.

Robinson J, Mistry K, McWilliam H, Lopez R, Marsh SGE. 2010. IPD—the Immuno Polymorphism Database. *Nucleic Acids Res*. 38:D863–D869.

Seddon JM, Ellegren H. 2004. A temporal analysis shows major histocompatibility complex loci in the Scandinavian wolf population are consistent with neutral evolution. *Proc Biol Sci*. 271:2283–2291.

Setchell JM, Huchard E. 2010. The hidden benefits of sex: evidence for MHC-associated mate choice in primate societies. *Bioessays*. 32:940–948.

Sin YW, Annavi G, Newman C, Buesching C, Burke T, Macdonald DW, Dugdale HL. 2015. MHC class II assortative mate choice in European badgers (*Meles meles*). *Mol Ecol*. 24:3138–3150.

Spehr M, Kelliher KR, Li X-H, Boehm T, Leinders-Zufall T, Zufall F. 2006. Essential role of the main olfactory system in social recognition of major histocompatibility complex peptide ligands. *J Neurosci*. 26:1961–1970.

Stahler DR, Macnulty DR, Wayne RK, VonHoldt B, Smith DW. 2013. The adaptive value of morphological, behavioural and life-history traits in reproductive female wolves. *J Anim Ecol*. 82:222–234.

Stephens M, Donnelly P. 2003. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet*. 73:1162–1169.

Sturm T, Leinders-Zufall T, Maček B, Walzer M, Jung S, Pömmerl B, Stevanović S, Zufall F, Overath P, Rammensee H-G. 2013. Mouse urinary peptides provide a molecular basis for genotype discrimination by nasal sensory neurons. *Nat Commun*. 4:1616.

Thoss M, Ilmonen P, Musolf K, Penn DJ. 2011. Major histocompatibility complex heterozygosity enhances reproductive success. *Mol Ecol*. 20:1546–1557.

Tregenza T, Wedell N. 2000. Genetic compatibility, mate choice and patterns of parentage: invited review. *Mol Ecol*. 9:1013–1027.

Vonholdt BM, Stahler DR, Smith DW, Earl DA, Pollinger JP, Wayne RK. 2008. The genealogy and genetic viability of reintroduced Yellowstone grey wolves. *Mol Ecol*. 17:252–274.

Wagner JL, Burnett RC, Works JD, Storb R. 1996. Molecular analysis of DLA-DRBB1 polymorphism. *Tissue Antigens*. 48:554–561.

Wang J, Santure AW. 2009. Parentage and sibship inference from multilocus genotype data under polygamy. *Genetics*. 181:1579–1594.

Wayne RK, Ostrander EA. 2007. Lessons learned from the dog genome. *Trends Genet*. 23:557–567.

Wedekind C, Seebeck T, Bettens F, Paepke AJ. 1995. MHC-dependent mate preferences in humans. *Proc Biol Sci*. 260:245–249.

Yang B, Ren B, Xiang Z, Yang J, Yao H, Garber PA, Li M. 2014. Major histocompatibility complex and mate choice in the polygynous primate: the Sichuan snub-nosed monkey (*Rhinopithecus roxellana*). *Integr Zool*. 9:598–612.

Yuhki N, Beck T, Stephens R, Neelam B, O'Brien SJ. 2007. Comparative genomic structure of human, dog, and cat MHC: HLA, DLA, and FLA. *J Hered*. 98:390–399.

Supplementary materials PAPER III

Table S1 Summary statistics of the 12 microsatellite loci (STR) used to reconstruct the individual genotypes for the 47 breeding individuals from the 26 tested pairs. Means and standard errors (SE) are indicated.

STR Locus	**Na**	**Ne**	**Ho**	**He**	**F**
2004	5	2.4	0.681	0.587	-0.159
2079	4	2.9	0.739	0.655	-0.129
2088	5	3.0	0.638	0.668	0.044
2096	3	2.8	0.596	0.642	0.072
2137	10	5.6	0.787	0.823	0.043
cph2	6	2.7	0.617	0.623	0.010
cph4	3	1.8	0.447	0.438	-0.020
cph5	3	2.3	0.574	0.561	-0.023
cph8	5	3.6	0.723	0.720	-0.005
cph12	4	1.9	0.543	0.461	-0.179
u250	5	2.8	0.702	0.639	-0.098
u253	1	1.0	0.000	0.000	N/A
Mean	4.500	2.723	0.587	0.568	-0.040
SE	0.634	0.329	0.060	0.060	0.025

Na: number of alleles; Ne: effective number of alleles; Ho: observed heterozygosity; He: expected heterozygosity; F: Fixation Index.

Table S2

Names and frequencies of the alleles found at three MHC class II loci in the wolf breeders from the northern Apennine, Italy, analysed in this study (n = 47). The corresponding GenBank names and accession numbers (AN) are indicated. Those previously undescribed in the study area (Galaverni *et al.* 2013) are shown in bold.

Locus and allele	Frequency	GenBank	AN
DRB1			
DLA-DRB1*12801	0.457	12801	JX206798
DLA-DRB1*092013	0.266	092013	JX206799
DLA-DRB1*02001	0.117	D20	U58684.1
DLA-DRB1*03601	0.064	03601	AF336110.1
DLA-DRB1*09201	0.043	09201	AM408904.1
DLA-DRB1*01501	0.043	DRB1-W	DQ056281.1
DLA-DRB1*03901	0.011	03901	AF343740.1
DQA1			
DLA-DQA1*005011	0.723	DQA3	U44787.1
DLA-DQA1*00401	0.096	DQA4	U44788.1
DLA-DQA1*01201	0.064	01201	AF343734.1
DLA-DQA1*00601	0.064	DQA6	U44790.1
DLA-DQA1*00901	**0.043**	**DQA1**	**U44785.1**
DLA-DQA1*00201	0.011	DQA9	U75455.1
DQB1			
DLA-DQB1*03901	0.415	03901	AY126651.1
DLA-DQB1*00701	0.319	DQB4	AF043149.1
DLA-DQB1*01303	0.085	DQB7	AF043152.1
DLA-DQB1*03501	0.074	03501	AJ311107.1
DLA-DQB1*02002	0.053	DQB19	AF043164.1
DLA-DQB1*00101	**0.043**	**DQB19**	**AF043164.1**
DLA-DQB1*02901	0.011	02901	AY126648.1

Table S3

Number and frequency of the three-locus MHC haplotypes inferred in the 47 breeders successfully analysed in this study. Haplotypes previously undescribed in the study area (Galaverni *et al.* 2013) are highlighted in bold.

Haplotype	Nomenclature (DRB1 / DQA1 / DQB1)	Frequency
1	DRB1*12801 / DQA1*005011 / DQB1*03901	0.362
2	DRB1*092013 / DQA1*005011 / DQB1*00701	0.298
3	DRB1*03601 / DQA1*01201 / DQB1*03501	0.064
4	DRB1*02001 / DQA1*00401 / DQB1*01303	0.085
5	DRB1*09201 / DQA1*00601 / DQB1*02002	0.043
6	DRB1*12801/ DQA1*005011 / DQB1*00701	0.053
7	**DRB1*01501 / DQA1*00901 / DQB1*00101**	**0.043**
8	**DRB1*03901 / DQA1*00201 / DQB1*02901**	**0.011**
9	DRB1*02001 / DQA1*00401 / DQB1*03901	0.011
10	**DRB1*12801 / DQA1*00601 / DQB1*03901**	**0.011**
11	**DRB1*02001 / DQA1*005011 / DQB1*01303**	**0.011**
12	**DRB1*02001 / DQA1*00601 / DQB1*02002**	**0.011**

Four additional haplotypes (DRB1*03901 / DQA1*00201 / DQB1*02002; DRB1*092013 / DQA1*005011 / DQB1*03501; **DRB1*12801/ DQA1*005011 / DQB1*03501; DRB1*092013 / DQA1*00401 / DQB1*01303**) have been found a single time among the offspring and the unrelated non-reproductive wolves analysed for comparative purposes.

FINAL COMMENTS

In conservation genetics variation within and among threatened populations is often identified using neutral genetic markers which are commonly known as not being directly targeted by selection, hence by genetic drift or demographic events. These markers are excellent for estimating effective population size, migration rates, and other population genetic processes. However, among the genome selective forces can shape particular traits, which may retain a significant role in survival. Those traits generally include encoding genes with adaptive functions. The study of those genes is necessary in conservation biology to comprehend the minimum ecologically meaningful genetic variation required to maintain resilience and resistance of natural populations to habitat and environmental changes.

The MHC genes belong to one of the most polymorphic multi-gene family that constitutes important adaptive traits in vertebrates. MHCs became the best candidates for the study of adaptive genetic diversity thanks to their homologous presence in several mammals, their extraordinary variability and for their obvious ecological relevance. The sequencing of MHC class II genes in conjunction with neutral markers has then been extensively used in conservation genetics.

Previously, higher differentiation in neutral than in MHC loci has been found in several wildlife species (Aguilar *et al.* 2004; van Oosterhout, Joyce, and Cummings 2006). However, also opposing results possibly caused by diversifying selection in MHC have been observed (Ekblom *et al.* 2007; Clare D. Marsden *et al.* 2012). In some other cases, populations were equally differentiated with MHC and neutral markers (P W Hedrick *et al.* 2000), implying that in small populations the effect of drift or bottleneck can be stronger than selection.

Genetic variation is a prerequisite for any population's ability to adapt to a changing environment. When populations are small and isolated, they are generally characterized by the presence of low levels of genetic variation. Consequently it is believed that these populations are less able to adapt if rapid changes occurs. In the case of the MHC loci low diversity it is expected to cause more susceptibility to: i) rapid environment changings (Weber *et al.* 2013), ii) wider range of diseases (Sommer 2005), iii) or specific genetic vulnerability to a pathogen (Quinnell *et al.* 2003). However there are also cases in which the species have not shown negative effects from poor MHC variability (Ellegren *et al.* 1993; Sommer 2005).

The Italian wolf population have experienced a long-term isolation and extreme demographic reduction followed by a recent re-expansion. Therefore, in order to monitor its genetic variability

and ability to cope with future adverse changes, we used a combination of several neutral markers and three MHC class II genes. We provided evidences of absence of loss of genetic variation due to the past bottleneck, and traces of historical selection were found, but we could not detect clear signals of on-going selection. Difference in heterozygosity between MHC and STRs loci were not significant but more marked in the putatively admixed wolf x dog individuals, suggesting higher genetic differentiation at the STRs than MHC among parental populations. We found private MHC multi locus haplotypes in admixed individuals and these haplotypes were mostly constituted by likely dog-derived alleles in accordance to the putative hybrid origin of the individual. This result raised interest in the use of MHC class II loci typing in admixed individuals, thus we also decided to test the utility of using the MHC loci in the identification of admixed patterns of variability in a already developed multi-locus method (Randi *et al.* 2014), applied to several Italian and west European canid populations (Lapalombella et. al in preparation).

Fragmentation and isolation experimented by the Italian wolf population may have extremely detrimental effects on the fitness, thus we investigated mating preferences in three MHC class II genes. We found assortative mating scheme rather than disassortative, which means that breeders tent to mate individuals that share MHC alleles with less AA changes at each gene. Choosing similar mate could directly benefit breeders that will not interact with potential disease carriers. Combining MHC and STRs and fitness proxies deduced from pedigree data, we found that assortative mating scheme was balanced by a clear signal of inbreeding avoidance, and by heterozygote advantage.

Wolf-like canids (genus *Canis*) evolved during the last 2–4 million years (vonHoldt *et al.* 2011), and separated recently retaining the potential to hybridize in nature (Stronen *et al.* 2012). Wild canids conservation is thus complicated by hybridization. Particular concern is created by the overlap of the diffusion of wild canids with free dogs, which usually outnumbers the natural populations. Due to this imbalance, the main threaten to population genetic conservation is the possible undirectional gene flow of domestic genes facilitated by the presence of fragmented and isolated breeders. The spread of domestic characters into natural populations may disrupt local adaptation and or increase genetic homogenization, eventually leading to the extinction of species through introgressive hybridization (Allendorf *et al.* 2001). The perceived problem by media is the recovering of wild canids, while the real problem for conservation management is the presence of free domestic dogs. Several national and local campaigns encourage more responsible ownership and sterilizations to reduce feral dog populations. However we are far from educating all dog owners to declare, register, micro-chipping and prevent to free their dogs.

We described in this thesis different examples of hybridization among canids, and in particular we

examined the genetic evidences found in three canids carrying anomalous phenotypes, supposed to belong to jackal population in Croatia. The hybrid nature of the three samples was confirmed by principal component analysis at STRs, and Bayesian cluster analysis. After computing posterior probabilities to determine individual identity to different hybrid classes, one individual was assigned to first generation of hybrid, and two to backcross, one with dog and one with jackal. Hybridization took place between female golden jackals and male dogs since all the three anomalous individuals carried a golden jackal mtDNA CR haplotype, whereas the two males showed Y chromosome marker haplotypes characteristic of dogs. None within Croatian jackal samples retained K locus deletion, while hybrids did, suggesting hybrids could have received the K locus deletion from dogs. Three out of four DLA-DRB1 alleles, one out of two DLA-DQA1 allele and both DLA-DQB1 alleles found in golden jackals analysed have not been identified in any canine species before. Thus, they could represent golden jackal private alleles that can be used as species-diagnostic markers. No dog's derived MHC multilocus haplotypes were described in jackal population, while hybrid individuals retained both dog and jackal's derived multilocus haplotypes, further confirming their admixed origin.

Nevertheless, it is also important to understand more about the reasons, which caused local or specific admixed patterns of variability. Management should take actions to avoid further fragmentation of wild canids populations, and improve the maintaining of a network of ecological corridors between protected areas, potentially allowing for natural dispersion and recovery. Further, giving the actual distribution of wild canids in Europe, these actions should include also areas across national borders.

Besides hybridization, poaching and accidental killing are the main cause of concern which cause not only genetic but also economic losses. People still perceive the wolf as a real treat for humans. Poaching and illegal hunting gains social acceptance in rural areas and management actions in favour of the species are often used to increase political divisions too.

In Italy has been esteemed almost 321 wolf packs corresponding to 1269-1800 wolves, with 1212–1711 distributed along the Apennines and 57–89 in the Alps (Galaverni *et al.* 2015). Incredibly high part (15-20%) of the population is illegally killed each year. The reasons for resentment and hatred to wolves' depredation on livestock are already addressed through practical support and education. Clearly, we are facing times in which the species is transitioning from a threatened and strictly protected to locally abundant state which calls for updated conservation and management approaches (Kaczensky *et al.* 2013).

In conclusion, as we showed in this thesis the study of MHCs variability can be differently applied in conservation of wild canids. Of course conservation genetics focus on the preservation of genetic

variation and on the identification of the crucial mechanisms involved in preserving natural populations of threatened species. However, preservation of genetic diversity should be done setting genetic goals in the protection plans (Frankham 2010), since this line of reasoning may often lead to the conclusion that all that matters in conservation is to preserve genetic variation (Höglund 2009). It must be said that the majority of the MHC studies focus on describing variability and possible selection pressure and consequences in viability, but there are few studies that directly demonstrate link between pathogen mediated population decline and low MHC variation. These studies should be performed to answer to this lack of data. Since diversity should be managed with caution, captive management and reintroduction programs could benefit from these studies and from genetic data coming from MHC typing, to avoid inbreeding and to prevent pathogen invasion suggesting with foresight and knowledge the introduction of different MHC haplotype alleles carriers among breeders.

REFERENCES

Aguilar A, Roemer G, Debenham S *et al.* (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 3490–4.

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology & Evolution*, **16**, 613–622.

Angles JM, Kennedy LJ, Pedersen NC (2005) Frequency and distribution of alleles of canine MHC-II DLA-DQB1, DLA-DQA1 and DLA-DRB1 in 25 representative American Kennel Club breeds. *Tissue antigens*, **66**, 173–84.

Anonymous (1972) The Indian Wildlife (Protection) Act, (1972). Ministry of Environment and Forests, Government of India, Nataraj Publishers, Dehradun, India.

Arbanasić H, Huber Đ, Kusak J *et al.* (2013) Extensive polymorphism and evidence of selection pressure on major histocompatibility complex DLA-DRB1, DQA1 and DQB1 class II genes in Croatian grey wolves. *Tissue Antigens*, **81**, 19–27.

Arnold J, Humer A, Heltai M *et al.* (2012) Current status and distribution of golden jackals Canis aureus in Europe. *Mammal Review*, **42**, 1–11.

Banea OC, Krofel M, C˘ ervinka J, Gargarea P, Szabo´ L (2012) New records, first estimates of densities and questions of applied ecology for jackals in Danube Delta Biosphere Reserve and hunting terrains from Romania. Acta Zool Bulg **64**, 353–366

Bardeleben C, Moore RL, Wayne RK (2005) A molecular phylogeny of the Canidae based on six nuclear loci. *Molecular phylogenetics and evolution*, **37**, 815–31.

Beck S, Trowsdale J (2000) The Human Major Histocompatibility Complex: Lessons from the DNA Sequence. *Annual Review of Genomics and Human Genetics*, **1**, 117–137.

Berggren KT, Seddon JM (2005) MHC promoter polymorphism in grey wolves and domestic dogs. *Immunogenetics*, **57**, 267–72.

Berggren KT, Seddon JM (2008) Allelic combinations of promoter and exon 2 in DQB1 in dogs and wolves. *Journal of molecular evolution*, **67**, 76–84.

Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of evolutionary biology*, **16**, 363–77.

Billingham RE, Brent L, Medawar PB (1953) "Actively Acquired Tolerance" of Foreign Cells. *Nature*, **172**, 603–606.

Billingham RE, Medawar PB (1951) The Technique of Free Skin Grafting in Mammals. *Journal of Experimental Biology*, **28**, 385–402.

Bollmer JL, Vargas FH, Parker PG (2007) Low MHC variation in the endangered Galápagos penguin (Spheniscus mendiculus). *Immunogenetics*, **59**, 593–602.

Bradley RD, Baker RJ (2001) A Test Of The Genetic Species Concept: Cytochrome- B Sequences And Mammals. *Journal of Mammalogy*, **82**, 960–973.

Brennan P A, Kendrick KM (2006) Mammalian social odours: attraction and individual recognition. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **361**, 2061–2078.

Brennan P A, Zufall F (2006) Pheromonal communication in vertebrates. *Nature*, **444**, 308–315.

Chaix R, Cao C, Donnelly P (2008) Is Mate Choice in Humans MHC-Dependent? *PLoS Genetics*, **4**, e1000184.

Clarke B, Kirby DRS (1966) Maintenance of Histocompatibility Polymorphisms. *Nature*, **211**, 999–1000.

Clutton-Brock, J., Corbet, G.B. and Hills, M. 1976. A review of the family Canidae, with a classification by numerical methods. Bulletin of the British Museum (Natural

History), Zoology **29**,119–199.
Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biology*, **3**.
Doherty PC, Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature*, **256**, 50–52.
Egid, K., and J. L. Brown. 1989. The major histocompatibility com- plex and female mating preference in mice. Anim. Behav. **38**, 548–560.
Eizaguirre C, Yeates SE, Lenz TL, Kalbe M, Milinski M (2009) MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Molecular ecology*, **18**, 3316–29.
Ekblom R, Stein Are Sæther, Jacobsson P, Fisker P, Sahlman T *et al.* (2007) Spatial pattern of MHC class II variation in the great snipe (Gallinago media). *Molecular Ecology*, **16**, 1439–1451.
Ellegren H, Hartman G, Johansson M, Andersson L (1993) Major histocompatibility complex monomorphism and low levels of DNA fingerprinting variability in a reintroduced and rapidly expanding population of beavers. *Proceedings of the National Academy of Sciences*, **90**, 8150–8153.
Fabbri E, Caniglia R, Galov A *et al.* (2014) Genetic structure and expansion of golden jackals (Canis aureus) in the north-western distribution range (Croatia and eastern Italian Alps). *Conservation Genetics*, **15**, 187–199.
Flajnik MF, Kasahara M (2010) Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nature Reviews Genetics*, **11**, 47–59.
Freedman AH, Gronau I, Schweizer RM *et al.* (2014) Genome sequencing highlights the dynamic early history of dogs. *PLoS genetics*, **10**, e1004016.
Fuller, T.K. 1989. Population dynamics of wolves in north-central Minnesota. Wildlife Monographs **105**, 1– 41.
Galaverni M, Caniglia R, Fabbri E, Lapalombella S, Randi E (2013) MHC variability in an isolated wolf population in Italy. *Journal of Heredity*, **104**, 601–612.
Galaverni M, Caniglia R, Fabbri E, Milanesi P, Randi E (2015) One, no one, or one hundred thousand: how many wolves are there currently in Italy? *Mammal Research*, **61**, 13-24.
Garrigan D, Hedrick PW (2003) Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution; international journal of organic evolution*, **57**, 1707–22.
Genov, P. and Wassiley, S. 1989. Der schakal (Canis aureus L.) in Bulgarian. Ein beitrag zu seiner verbreitung und biologie. Zeitschrift fur Jagdwissenschaft **35,** 145–150.
Giannatos G, Marinos Y, Maragou P, Catsadorakis G (2005) The status of the Golden Jackal (*Canis aureus* L.) in Greece. *Belgian Journal of Zoology*, **135**, 145–149.
Golani, I. and Keller, A. (1975). A longitudinal field study of the behavior of a pair of golden jackals. Pp. 303–335 in M.W. Fox, ed. The wild canids. Van Nostrand Reinhold Company, New York, USA.
Greer K a, Wong a K, Liu H *et al.* (2010) Necrotizing meningoencephalitis of Pug dogs associates with dog leukocyte antigen class II and resembles acute variant forms of multiple sclerosis. *Tissue antigens*, **76**, 110–8.
Griggio M, Biard C, Penn DJ, Hoi H (2011) Female house sparrows "count on" male genes: experimental evidence for MHC-dependent mate preference in birds. *BMC evolutionary biology*, **11**, 44.
Hedrick PW (1986) Genetic Polymorphism in Heterogeneous Environments: A Decade Later. *Annual Review of Ecology and Systematics*, **17**, 535–566.
Hedrick PW, Lee RN, Garrigan D (2002) Major histocompatibility complex variation in red wolves: evidence for common ancestry with coyotes and balancing selection. *Molecular ecology*, **11**, 1905–13.
Hedrick PW, Lee RN, Parker KM (2000a) Major histocompatibility complex (MHC) variation in the endangered Mexican wolf and related canids. *Society*, **85**, 617–624.

Hedrick PW, Parker KM, Gutiérrez-Espeleta G a, Rattink A, Lievers K (2000b) Major Histocompatibility Complex Variation In The Arabian Oryx. *Evolution*, **54**, 2145–2151.

Hill AS (1991) HLA Associations with Malaria in Africa: Some Implications for MHC Evolution. In: *Molecular Evolution of the Major Histocompatibility Complex SE - 33* NATO ASI Series. (eds Klein J, Klein D), pp. 403–420. Springer Berlin Heidelberg.

Horton R, Wilming L, Rand V *et al.* (2004) Gene map of the extended human MHC. *Nature reviews. Genetics*, **5**, 889–899.

Hughes a. L, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: vidence for overdominant selection. *Proceedings of the National Academy of Sciences of the United States of merica*, **86**, 958–62.

Ilmonen P, Stundner G, Thoss M, Penn DJ (2009) Females prefer the scent of outbred males: good-genes-as-heterozygosity? *BMC evolutionary biology*, **9**, 104.

Jacob S, McClintock MK, Zelano B, Ober C (2002) Paternally inherited HLA alleles are associated with women's choice of male odor. *Nature genetics*, **30**, 175–179.

Jamieson IG, Taylor SS, Tracy LN, Kokko H, Armstrong DP (2009) Why some species of birds do not avoid inbreeding: Insights from New Zealand robins and saddlebacks. *Behavioral Ecology*, **20**, 575–584.

Janeš D, Klun I, Vidan-Jeras B, Jeras M, Kreft S (2010) Infuence of MHC on odour perception of 43 chemicals and body odour. *Central European Journal of Biology*, **5**, 324–330.

Jhala, Y. & Moehlman, P.D. (2008). Canis aureus. The IUCN Red List of Threatened Species 2008: e.T3744A10054631.

Jokinen P, Rusanen EM, Kennedy LJ, Lohi H (2011) MHC class II risk haplotype associated with canine chronic superficial keratitis in German Shepherd dogs. *Veterinary immunology and immunopathology*, **140**, 37–41.

Kaufman J, Milne S, Göbel TW *et al.* (1999) The chicken B locus is a minimal essential major histocompatibility complex. *Nature*, **401**, 923–925.

Kelley J, Walter L, Trowsdale J (2005) Comparative genomics of major histocompatibility complexes. *Immunogenetics*, **56**, 683–695.

Kennedy LJ, Angles JM, Barnes a *et al.* (2001) Nomenclature for factors of the dog major histocompatibility system (DLA), 2000: second report of the ISAG DLA Nomenclature Committee. *Animal genetics*, **32**, 193–9.

Kennedy LJ, Barnes a, Happ GM *et al.* (2002) Extensive interbreed, but minimal intrabreed, variation of DLA class II alleles and haplotypes in dogs. *Tissue antigens*, **59**, 194–204.

Kennedy LJ, Carter SD, Barnes A *et al.* (1998) Nine new dog DLA-DRB1 alleles identified by sequence-based typing. *Transplantation*, 296–301.

Kennedy LJ, Huson HJ, Leonard J *et al.* (2006) Association of hypothyroid disease in Doberman Pinscher dogs with a rare major histocompatibility complex DLA class II haplotype. *Tissue antigens*, **67**, 53–6.

Kennedy LJ, Randall D a, Knobel D *et al.* (2011) Major histocompatibility complex diversity in the endangered Ethiopian wolf (Canis simensis). *Tissue Antigens*, **77**, 118–125.

Klein J (1987) Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. *Human immunology*, **19**, 155–162.

Knapp LA, Robson J, Waterhouse JS (2006) Olfactory signals and the MHC: a review and a case study inLemur catta. *American Journal of Primatology*, **68**, 568–584.

Koepfli K-P, Pollinger J, Godinho R *et al.* (2015) Genome-wide Evidence Reveals that African and Eurasian Golden Jackals Are Distinct Species. *Current Biology*, **25**, 2158–2165.

Krystufek Boris, Tvrtković, Nikola 1990 "Range Expansion by Dalmatian Jackal Population in the 20th-Century (Canis Aureus Linnaeus, 1758)." Folia Zoologica **39**, 291–296

Krystufek B,Murariu D, Kurtonur C (1997) Present distribution of the Golden Jackal Canis aureus in the Balkans and adjacent regions. Mamm Review **27**, 109–114

Lapini L (2012) Lo sciacallo dorato Canis aureus moreoticus (I.GEOFFROY SAINT HILAIRE,

1835) nell'Italia nord-orientale (Carnivora: Canidae). , 1–118.
Lapini, L., Molinari, P., Dorigo, L., Are, G., & Beraldo, P. (2009). Reproduction of the golden jackal (Canis aureus moreoticus I. Geoffroy Saint Hilaire, 1835) in Julian pre-Alps, with new data on its range-expansion in the high-adriatic Hinterland (Mammalia, Carnivora, Canidae). Boll Mus Civ St Nat Venezia, **60**, 169-186.
Lindblad-Toh K, Wade CM, Mikkelsen TS *et al.* (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, **438**, 803–819.
Macdonald, D.W. 1979. The flexible social system of the golden jackal, Canis aureus. Behavioural Ecology and Sociobiology **5**, 17–38.
Magory Cohen T, King R, Dolev A *et al.* (2013) Genetic characterization of populations of the golden jackal and the red fox in Israel. *Conservation Genetics*, **14**, 55–63.
Marsden CD, Mable BK, Woodroffe R *et al.* (2009) Highly Endangered African Wild Dogs (Lycaon pictus) Lack Variation at the Major Histocompatibility Complex. *Journal of Heredity*, **100**, S54–S65.
Marsden, C. D., Woodroffe, R., Mills, M. G., McNutt, J. W., Creel, S., Groom, R., ... & Ginsberg, J. (2012). Spatial and temporal patterns of neutral and adaptive genetic variation in the endangered African wild dog (Lycaon pictus). Molecular Ecology, 21(6), 1379-1393.
Martínez-Borra J, López-Larrea C (2012) The emergence of the Major Histocompatilibility Complex. *Advances in Experimental Medicine and Biology*, **738**, 277–289.
Milinski M, Griffiths S, Wegner KM *et al.* (2005) Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 4414–4418.
Moehlman, P.D. (1983). Socioecology of silverbacked and golden jackals (Canis mesomelas and Canis aureus). Pp. 423–453 in J.F. Eisenberg and D.G. Kleiman, eds. Recent advances in the study of mammalian behavior. American Society of Mammologists Spec. Publ. No. 7, Pittsburgh Pennsylvania, USA.
Moehlman, P.D. (1986). Ecology of cooperation in canids. Pp. 64–86 in D.I. Rubenstein and R.W. Wrangam, eds. Ecological aspects of social evolution. Princeton University Press, Princeton, NJ, USA.
Moehlman, P.D. (1989). Intraspecific variation in canid social systems. Pp. 143–163 in J.L. Gittleman, ed., Carnivore behavior, ecology and evolution. Cornell University Press, Ithaca, NY, USA.
Moehlman, P.D. and Hofer, H. (1997). Cooperative breeding, reproductive suppression, and body mass in canids. Pp. 76–128 in N.G. Solomon and J.A. French, eds. Cooperative breeding in mammals. Cambridge University Press, Cambridge, UK.
Morell V (2015) From Wolf to Dog. Scientific American, 313, 60–67.
Murgia C, Pritchard JK, Kim SY, Fassati A, Weiss RA (2006) Clonal Origin and Evolution of a Transmissible Cancer. *Cell*, 477–487.
Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 7799–7806.
Ninomiya K, Brown RE (1995) Removal of the preputial glands alters the individual odors of male MHC-congenic mice and the preferences of females for these odors. *Physiology & behavior*, **58**, 191–4.
Niskanen AK, Hagström E, Lohi H *et al.* (2013) MHC variability supports dog domestication from a large number of wolves: high diversity in Asia. *Heredity*, **110**, 80–85.
Niskanen AK, Kennedy LJ, Ruokonen M *et al.* (2014) Balancing selection and heterozygote advantage in major histocompatibility complex loci of the bottlenecked Finnish wolf population. *Molecular Ecology*, **23**, 875–889.
Ohno S (1970) Evolution by Gene Duplication. *Springer-Verlag*.
van Oosterhout C, Joyce D a, Cummings SM (2006) Evolution of MHC class IIB in the genome of

wild and ornamental guppies, Poecilia reticulata. *Heredity*, **97**, 111–118.
Penn DJ, Potts WK (1999) The Evolution of Mating Preferences and Major Histocompatibility Complex Genes. *Evolution*, **153**, 145–164.
Pierson RN, Winn HJ, Russell PS, Auchincloss H (1989) Xenogeneic skin graft rejection is especially dependent on CD4+ T cells. *The Journal of Experimental Medicine* , **170** , 991–996.
Poche, R.M., Evans, S.J., Sultana, P., Haque, M.E, Sterner, R. and Siddique, M.A. 1987. Notes on the golden jackal (Canis aureus) in Bangladesh. Mammalia **51**, 259–270.
Potts WK, Wakeland EK (1990) Evolution of diversity at the major histocompatibility complex. *Trends in ecology & evolution*, **5**, 181–187.
Quinnell RJ, Kennedy LJ, Barnes A *et al.* (2003) Susceptibility to visceral leishmaniasis in the domestic dog is associated with MHC class II polymorphism. *Immunogenetics*, **55**, 23–8.
Randi E, Hulva P, Fabbri E *et al.* (2014) Multilocus Detection of Wolf x Dog Hybridization in Italy, and Guidelines for Marker Selection (G Larson, Ed,). *PLoS ONE*, **9**, e86409.
Reusch TBH, Häberli MA, Aeschlimann PB, Milinski M (2001) Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. *Nature*, **414**, 300–302.
Rosenberg a S, Singer a (1992) Cellular basis of skin allograft rejection: an in vivo model of immune-mediated tissue destruction. *Annual review of immunology*, **10**, 333–358.
Rueness EK, Asmyhr MG, Sillero-Zubiri C *et al.* (2011) The cryptic African wolf: Canis aureus lupaster is not a golden jackal and is not endemic to Egypt. *PloS one*, **6**, e16385.
Runstadler JA, Angles JM, Pedersen NC (2006) Dog leucocyte antigen class II diversity and relationships among indigenous dogs of the island nations of Indonesia (Bali), Australia and New Guinea. *Tissue antigens*, **68**, 418–26.
Seddon JM, Ellegren H (2002) MHC class II genes in European wolves: a comparison with dogs. *Immunogenetics*, **54**, 490–500.
Seddon JM, Ellegren H (2004) A temporal analysis shows major histocompatibility complex loci in the Scandinavian wolf population are consistent with neutral evolution A temporal analysis shows major histocompatibility complex loci in the Scandinavian wolf population are consistent wi. *B, Proc R Soc Lond Society*.
Selanec J, Lausˇ B, Sindicˇic´ M (2011) Golden jackal (Canis aureus) distribution in Croatia. In: Abstract of VI European Congress of mammalogy, Paris, France, p 60
Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in zoology*, **2**, 16.
Spassov, N. 1989. Signalling - communicative marking in the colouration of the genus Canis L. In Proceedings of the Fifth International Theriological Congress, Rome, 22-29 August 1989. Vol. 2. NAGE Snc, Rome. pp. 799-800.
Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 979–988.
Stronen A V, Tessier N, Jolicoeur H *et al.* (2012) Canid hybridization: contemporary evolution in human-modified landscapes. *Ecology and evolution*, **2**, 2128–40.
Tregenza T, Wedell N (2000) Genetic compatibility, mate choice and patterns of parentage: invited review. *Molecular ecology*, **9**, 1013–27.
Väli Ü, Einarsson A, Waits L, Ellegren H (2008) To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology*, **17**, 3808–3817.
van Oosterhout, C, D a Joyce, and S M Cummings. (2006). "Evolution of MHC Class IIB in the Genome of Wild and Ornamental Guppies, Poecilia Reticulata." Heredity **97**, 111–118.
Vilà C, Seddon J, Ellegren H (2005) Genes of domestic mammals augmented by backcrossing with wild ancestors. *Trends in genetics : TIG*, **21**, 214–8.
vonHoldt BM, Pollinger JP, Earl D a *et al.* (2011) A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. *Genome Research*, **21**, 1294–1305.
vonHoldt BM, Pollinger JP, Lohmueller KE *et al.* (2010) Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature*, **464**, 898–902.

Wang, X., Tedford, R. H., Van Valkenburgh, B., & Wayne, R. K. (2004). Evolutionary history, molecular systematics, and evolutionary ecology of Canidae. Biology and conservation of wild canids (DW Macdonald and C. Sillero-Zubiri, eds.). Oxford University Press, Oxford, United Kingdom, 39-54.

Wagner JL, Burnett RC, Works JD, Molecular RS (1996) Molecular analysis of DLA-DRBB 1 polymorphism. Cancer Research, 554–561.

Wagner JL, Storb R, Storer B, Mignot E (2000) DLA-DQB1 alleles and bone marrow transplantation experiments in narcoleptic dogs. *Tissue antigens*, **56**, 223–31.

Wayne RK, Bridgett M (2012) Evolutionary genomics of dog domestication. , 3–18.

Wayne RK, Geffen ELI, Girman DJ *et al.* (1997) MOLECULAR SYSTEMATICS OF THE CANIDAE jackal- , and foxlike carnivores that includes gentina . The first appearance of South American canids followed the immigration of North American mammals into South America during the early Pleistocene after the geolog. , **46**, 622–653.

Weber DS, Van Coeverden De Groot PJ, Peacock E *et al.* (2013) Low MHC variation in the polar bear: implications in the face of Arctic warming? *Animal Conservation*, **16**, 671–683.

Wedekind C, Seebeck T, Bettens F, Paepke a J (1995) MHC-dependent mate preferences in humans. *Proceedings. Biological sciences / The Royal Society*, **260**, 245–249.

Yuhki N, Beck T, Stephens R, Neelam B, O'Brien SJ (2007) Comparative genomic structure of human, dog, and cat MHC: HLA, DLA, and FLA. *The Journal of heredity*, **98**, 390–9.

Yumnam B, Negi T, Maldonado JE, Fleischer RC, Jhala Y V. (2015) Phylogeography of the Golden Jackal (Canis aureus) in India (A Janke, Ed,). *PLOS ONE*, **10**, e0138497.

Zachos FE, Cirovic D, Kirschning J *et al.* (2009) Genetic variability, differentiation, and founder effect in golden jackals (Canis aureus) from Serbia as revealed by mitochondrial DNA and nuclear microsatellite loci. *Biochemical genetics*, **47**, 241–50.

PART II – The OR typing project

PART II – The OR genes typing project

Introduction to the olfactory system

In this section of the thesis, I present a general background on the olfactory system in order to introduce the results obtained by a comparative genetic analysis of the olfactory system that we performed between domestic dogs and wolves.

Mammals developed very sophisticated systems to communicate. Signals and information are the basis of communication. Signals can be modulated individually to contain a variety of qualitative and quantitative data from a sender to every possible receiver. The extraordinary fact is that in the natural world there are some signals that are comprehensible by a great variety of organisms, independently from sender and receiver species identity. This is the case of odour signals, where a molecule can convey a message to a wide number of receivers, but may reach greater complexity if the sender and the receiver belong to the same species. High complex messages indeed require that the receiver possess the ability to comprehend the additive meta data enclosed in them. This means primarily the ability to recognize additional information from the chemical clues, and also the ability to associate olfactory signals to other type of signals, such as visual and auditory ones.

The study of olfaction has traditionally focused on rats and mice and recently on canids, in particular on dogs, where peculiar smelling abilities have been observed especially in working breeds. Comparably, direct observations on their wild counterparts, the wolves, proved that they often use olfactory signals as a crucial complement of social communication, and although olfaction is considered the wolf primary sensory modality, few academic papers have been released to prove it.

The anatomy of olfaction

In mammals, olfaction is mediated by two main systems: the main olfactory system (MOS) and the accessory olfactory system (AOS). The MOS comprises the olfactory mucosa from which axons of the olfactory sensory neurons (OSNs) convey signals into the olfactory nerve to reach the main olfactory bulb (MOB). The nasal mucosa has been especially studied in mice, where it generally consists of four areas: the main olfactory epithelium (MOE), the septal organ (SO), the ganglion of Grüneberg (GG), and the vomeronasal sensory epithelium (VNE). Special epithelial layers mixed

with clusters of neurones characterize all these structures. The GG and SO are spatially separated chemosensory clusters of neurons. Originally discovered by Grüneberg in 1973, the GG is placed on the roof of nasal cavity. It was considered a non-sensory structure afferent to the accessory olfactory bulb (AOB) because of cilia and dendrites lack. The SO, also known as Masera organ (Rodolfo Masera, 1943), is a small spot of olfactory epithelium on the nasal septum above the vomero nasal organ (VNO). In contrast with what has been observed in other parts of the MOB, the SO is characterized by a fraction of MOB receptors and by one type of receptor expressed per cell (Tian & Ma 2004). While the GG function is related to alarm pheromones (Brechbühl *et al.* 2008) and cold temperature sensing (Mamasuew *et al.* 2008; Schmid *et al.* 2010), the Masera organ still remains poorly known (Kaluza *et al.* 2004; Tian & Ma 2004) although a connection with the regulation of respiration phases was recently proposed (Mori *et al.* 2014).

Nonetheless, the murine olfactory substructure is not always comparable with all macrosmatic mammals (Salazar & Quinteiro 2009). Post-mortem dissection of German shepherd dog MOS showed no GG nor SO (Barrios *et al.* 2014) suggesting indeed that these organs regressed during MOS evolution or ontogeny in dogs.

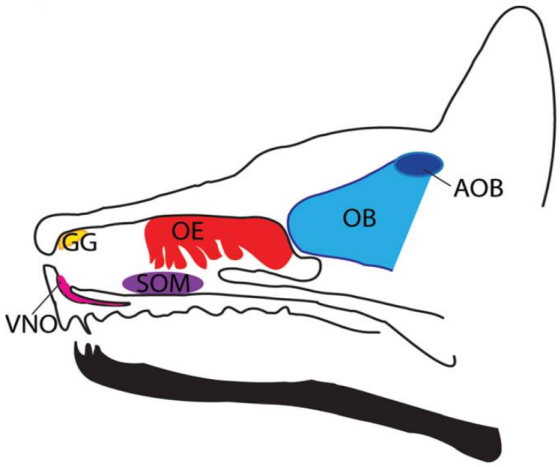


Figure 1

A diagrammatic representation of the olfactory cavity of a mammal. The location of the four main odour perception tissues is shown: olfactory epithelium (OE), vomeronasal organ (VNO), the Grueneberg ganglion (GG), and the septal organ of Masera (SOM). Other abbreviations used: olfactory bulb (OB) and accessory olfactory bulb (AOB). (from: Hayden and Teeling 2014)

Odorant binding

Two types of molecules, pheromones and odorants, can stimulate a neural cascade from OR receptors to the brain. Olfactive cues can be detected diffusing at very low concentrations into the mucus that covers the MOE, and thus be transported to ORs on OSNs cilia by odour-binding proteins. Despite there are ~1,000 ORs genes in mammals, each OSN expresses only one allele from one gene (Monahan & Lomvardas 2015). This 'one neuron–one receptor rule' is thought to be important for olfactory coding, such that only a given population of olfactory neurons responds to a restricted number of odour molecules.

Many studies have been performed to identify the binding sites for odorous ligands in the receptor structure. It is thought that the basis for the wide spectrum of odorous ligands recognition is in the sequence variability of TM3, 5 and 6 (Fleischer *et al.* 2006).

The ligand-binding phenomenology of ORs is complex. The binding pocket created by alpha chains is where odorant docks and changes receptor protein shape, leading to the G protein activation. The binding step can be performed by a finite number of receptors with an infinite number of olfactory cues combinations (Malnic *et al.* 1999). The olfactory code is thus similar to language complexity with odorant molecules performing like letters composing words. Each OR can recognize several chemical cues, and specific odorants may bind to several ORs (Malnic *et al.* 1999) likely with different response amplitudes (Young and Trask 2002; Figure 2. A - the third receptor reacts strongly than the last receptor).

Each time an OR is activated, a signal triggers the G-protein to which the receptor is coupled. The G-protein passes from an inactive state, binding GDP (guanosine diphosphate) to an active state, binding GTP (guanosine triphosphate), which activates the ATP conversion into cAMP (cyclic adenosine monophosphate) by adenylyl cyclase. The cAMP acts as a messenger and activates ion channels in the cell, inducing membrane depolarization and allowing the transduction of the signal trough OSN axons to the olfactory glomeruli in the olfactory bulb. Each OSN in MOE sends its projections to glomeruli, while in VNO OSNs are clustered into apical and basal zones (Dulac & Wagner 2006). Secondary neurons, the mitral cells, receive the signal inside glomeruli and covey it to the upper part of the brain, including the piriform cortex (MOE), hippocampus and amigdala, in turn connected to higher cortical brain centres for further processing and odorant identification (Firestein 2001; Dulac & Wagner 2006).

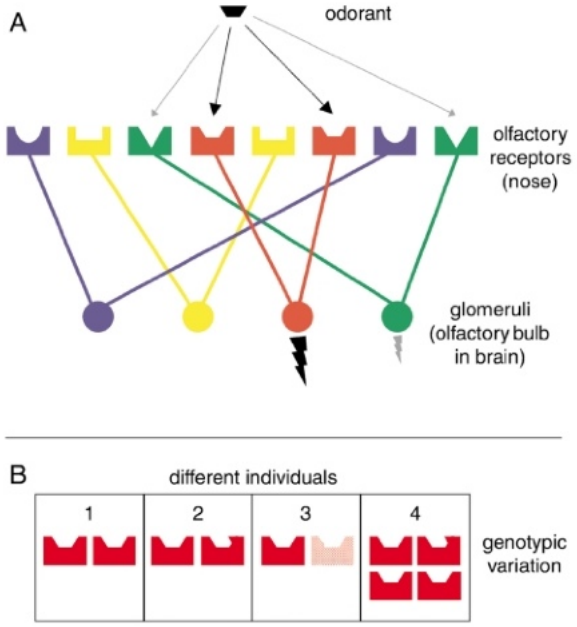


Figure 2

(A) Example of combinatorial code of olfaction whit a same odorant being recognized from several receptors. (B) Sources of phenotypic variation in olfaction. (Young 2002).

Thousands of neurons expressing a given olfactory receptor are spread throughout one zone of the olfactory epithelium, but their axons converge on one or two glomeruli in the olfactory bulb. As shown in Figure 1 B, individuals with different genotypes may (1) be homozygous for a given olfactory receptor, (2) be heterozygous and express sequence variants with slightly different odorant-binding capabilities, (3) possess non-functional variants (hatched receptor) and/or (4) have duplicate gene copies, perhaps changing the relative numbers of responsive neurons in the olfactory epithelium.

Olfactory repertoire size and structure

The detection of environmental chemicals is mediated by different transmembrane receptors. There are five receptor categories: olfactory receptors (ORs), the molecular receptors that recognize odorant molecules, expressed in MOE; the vomeronasal receptors (V1Rs and V2Rs) (Dulac & Axel 1995; Herrada & Dulac 1997; Matsunami & Buck 1997; Ryba & Tirindelli 1997) expressed

respectively in the basal and apical zones of the VNO (Dulac & Wagner 2006); the trace amine-associated receptors (TAARs), expressed primarily in MOE (Liberles & Buck 2006); the formyl peptide receptors (FPRs) expressed in VNO (Liberles *et al.* 2009; Rivière *et al.* 2009); and the guanylyl cyclase GC-D receptors (Yu *et al.* 1997) expressed in septal organ (SO) and MOE. ORs, TAARs, VRs and FPRs belong to the G protein-coupled receptor family (GPCR) and share seven hydrophobic transmembrane domains (TM).

On average, ORs are 310 amino acid long (Niimura et al. 2012) and are encoded by intronless genes present in almost all chromosomes. ORs have been grouped into families (sequence similarity > 40%) and subfamilies (similarity > 60%). Due to the level of receptor diversification, there are large numbers of subfamilies. These genes tend to form also genomic clusters with different sizes, which can contain members of several subfamilies or even families. ORs in mammals can be divided into Class I and Class II genes. These have been subdivided into several groups among vertebrates, but only Class I α group and Class II γ group receptors are present in mammals (Niimura 2012: Figure 3).

Interestingly in human, mouse and dog genomes, all Class I OR genes form a single cluster, contrary to Class II. Olfaction genetic repertoire could vary in sizes and pseudogenes rates. Humans have the smallest olfactory subgenome with almost 1000 genes spread in all chromosomes except 20 and Y (Glusman *et al.* 2001). Our species possess another olfactory record, since we have the highest pseudogenes rate (65%) and more than 100 subfamilies composed by pseudogenes (Rouquier & Giorgi 2007).

Conversely, mouse and rat have the largest OR repertoire, consisting of compact gene clusters unevenly distributed among chromosomes. In mice 1200 and 1400 genes with 20% pseudogenes (Rouquier & Giorgi 2007) were found, and about 1600 genes divided in 282 families in rat, with 13% pseudogenes (Quignon *et al.* 2005; Zhang *et al.* 2007).

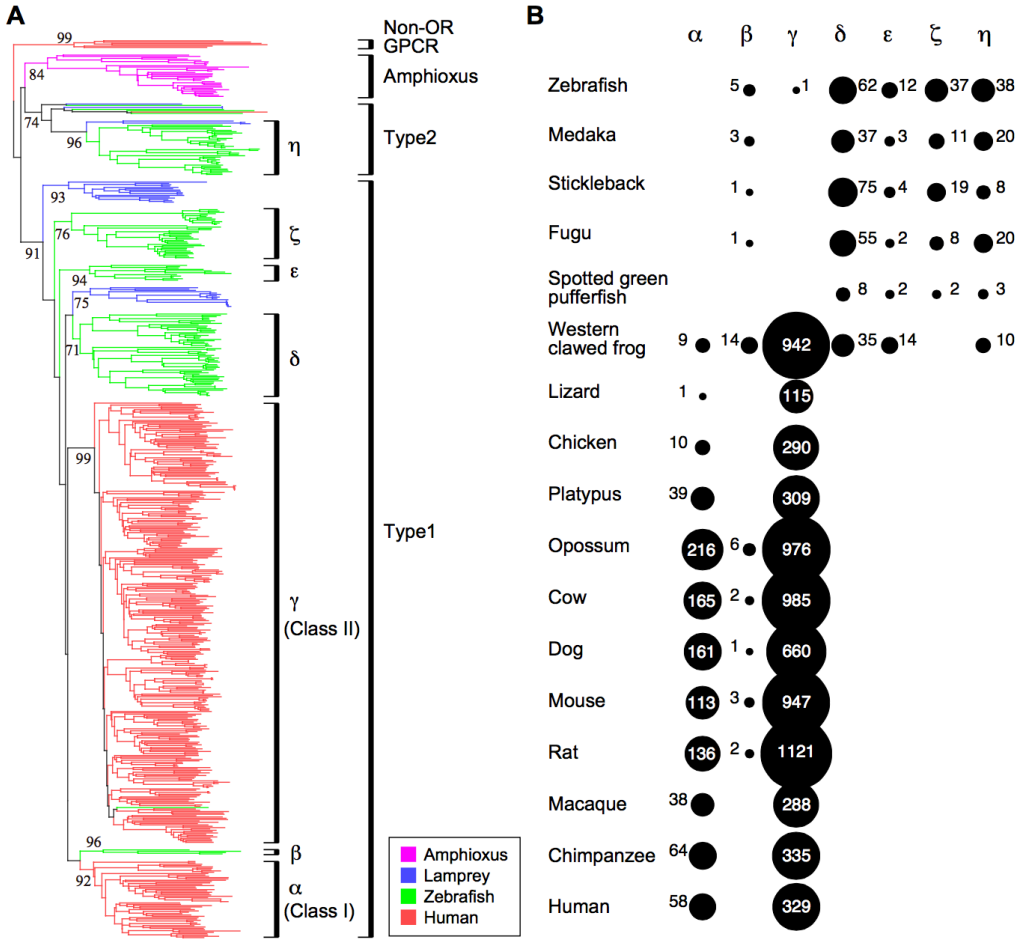


Figure 3

(A) Phylogenetic tree constructed using all intact OR genes from amphioxus, lamprey, zebrafish, and human. Several non-OR GPCR genes were used as outgroups. (B) Number of functional genes (the sum of intact genes and truncated genes) belonging to each group for each species. (Niimura et al. 2012)

Canine OR repertoire

Dog, mouse, and rat repertoire have 2.5–3.5 times more genes than human, consistent with the hypothesis that the repertoire size is related to the olfaction abilities of a species (Quignon *et al.* 2012). Notably, dogs, which are supposed to have a good sense of smell, do not have the largest number of functional OR genes (Niimura *et al.* 2014).

The canine olfactory subgenome was firstly identified via genomic DNA cloning of olfactory sequence tags (Olender *et al.* 2004) and then searching for five amino acidic patterns through dog whole genome shotgun sequence (Quignon *et al.* 2005). Dispersed into 24 chromosomes, the OR canine subgenome forms 49 clusters and 300 subfamilies (Quignon *et al.* 2005, 2012). The majority of ORs canine genes are class II, and about 200 belong to class I in chromosome 21 (CFA21),

between 29 and 31 Mb (Quignon et al. 2012). Five class I and 18 class II families were described (Robin *et al.* 2009). The two largest clusters on canine chromosomes 18 and 21 harbour only 4% and 10% pseudogenes, respectively, and at the family level a rather uneven distribution of pseudogenes was noted (Quignon *et al.* 2003). Data are in constant evolution, since now 1094 genes have been identified, with 20% pseudogenes (Quignon et al. 2012).

OR diversity in dogs greatly varies among breeds and individuals (Robin et al. 2009). Some breeds showed absence or weak polymorphism (German Shepherd) and other extreme gene cluster expansions, resulting in a breed-specific haplotype signature. Moreover, the level of OR polymorphism in dogs tends to be related to cluster organization, with the least polymorphic genes mostly localized in small clusters and the highest in large clusters. Due to the high values of SNP density found, Robin et al. (2009) suggested that the evolutionary model based only on duplication events could not explain such high variability, thus a gene conversion mechanism should be responsible for the large proportion of the mutations not being counter selected.

ORs evolution: neutral, evolutionary and adaptive theories

Until 1990 most multi-gene families were thought to be subject to a concerted evolution model, which proposes that genes belonging to the same family evolve as a unit in concert, by randomly repeated or unequal crossing-over and gene conversion (Nei & Rooney 2005). The concerted evolution theory thus proposes homogenization of the duplicated entities. The organization of OR repertoires among mammals is similar despite the number of genes. Synteny in clusters has been maintained and orthologous pairs can be found. Surprisingly, phylogenetically distantly-related OR genes were often observed in orthologous clusters (Glusman *et al.* 2000), a fact that was incompatible with the concerted evolution theory. Conversely, tandem duplications could have been the primary mechanism for OR family expansions and gene clustering (Young 2002; Niimura & Nei 2005). This mechanisms is well known as the birth-and-death model (Nei 1969; Nei *et al.* 1997). Repeated gene duplication events can give birth to new genes and pseudogenes by chance. Some of newly formed genes can persist in a genome for long time and then diverge, while others can be deleted or subjected to frameshift deleterious mutations.

In other studies it was suggested that gene duplication should be accompanied by gene conversion too. This was the case of dog's olfaction repertoire, where high values of SNPs density were found, and authors suggested that the only duplication evolutionary model could not explain such high variability, thus, gene conversion mechanism should be responsible for the large proportion of the

mutations not being counter selected.

The neutralist view of olfactory subgenome evolution (Niimura & Nei 2007) suggested that differentiation among species is almost due to random genes' duplication and inactivation, as the only evolutionary explanation for ORs different expression among mammals. Nevertheless, adaptive evolution too might have shaped OR plasticity (Hayden *et al.* 2010), thus different selective pressure may act during evolution driving olfaction abilities to be specialized or lost. During tetrapod evolution, the transition from water to land likely increased the importance of olfaction and consequently the number of OR genes encoding for receptors in the tetrapod lineage. Class I ORs were originally identified in fish (Ngai et al. 1993), thus classified as fish-like receptors, and class II in mammals, but subsequently both classes were found also in the amphibian species *Xenopus laevis* (Freitag et al. 1995). Class I ORs were supposed to be more sensitive to relatively hydrophilic compounds, whereas Class II ORs might favour more hydrophobic compounds (Freitag et al. 1995; Xinmin Zhang and Firestein 2002; Saito et al. 2009).

Recently Niimura et al. (2014) investigated the possible differences between class I and II receptors. In particular authors suggested that the positive correlation that exists between the number of OR genes and the number of ligands, and the generally lower expansion observed in class I than class II genes, could be due to specific functionalities retained by these two gene lineages. In particular class I genes were considered "the specialist" class, whose functionality could have been maintained for long periods during evolution in mammals by a combination of limited gene gains and losses, and directional selection.

Positive selection on OR genes appeared to promote OR repertoire evolution through changes in aminoacid sequence in order to increase discriminatory powers. This have been described in human (Gilad *et al.* 2003), mouse (Zhang *et al.* 2004) and dog (Robin *et al.* 2009). However cases of purifying selection too have been described in 13 placental mammals (Niimura et al. 2014), and in chimpanzees in comparison with humans, together with a higher proportion of intact genes (Gilad *et al.* 2003). In humans the massive presence of pseudogenes has been explained by the vision priority hypothesis, which correlates the increasing OR pseudogenization rate with the gain of the trichromatic vision in primates (Gilad *et al.* 2004).

The behavioural aspect of smell

In animals, chemical recognition is of fundamental importance and could drive to survival or death For humans, although our survival is less deeply dependent from smell capabilities, olfaction

is still centrally involved in the quality of life. It is intimately related with taste and can activate the smell-related memory, which is linked both to positive or negative feelings. All interactions among individuals of the same (or even different) species are dependent on the recognition of sex- and species-specific traces, which could be dispersed by urine, tears and saliva. These chemo-signals can be decoded both by the main olfactory and the vomeronasal system (Kelliher 2007). Individual chemo-signals are usually odorants or pheromones, but there is evidence that also MHC peptides can be recognized by V2R receptors in VNs (Brennan & Kendrick 2006; Ziegler *et al.* 2010; Sturm *et al.* 2013). The understanding of vertebrate pheromone communication will continue to be driven by molecular genomic approaches. However, an important limitation is that most works are carried out in rodents, particularly mice, whereas pheromones are, by definition, species-specific signals. Therefore, caution must be exercised in extrapolating between species (and between sexes), that have different reproductive strategies and behavioural priorities.

Canids, like other mammals, most rely on olfaction to communicate, but their characteristic social system made them an interesting complex model to investigate, especially in behavioural ecology and neuroscience. Often the only way to hypothesize wild canids olfactory abilities is by comparison with demonstrated dog olfactory abilities, assuming that also other canids may retain the same sensitivity. Odours in canids are supposed to contain a panel of individual information such as its gender, breeding condition, social status, age and even diet. Scent marks can result in a distinctive odour fingerprint as a result of typical individual microflora and diet, thus it is likely that also wolves are capable to identify such information (Mech & Boitani 2003).

Olfactory communication can convey signals both in presence and in absence of the sender.

In marking their territory wolves use both vocal and scent signals. Scent marks are often used as indirect deterrent for the defence of territory, thereby reducing fights with conspecifics (indirect territorial defence hypothesis). Scent marking by urination is a common practice in the *Canidae* family (Cafazzo *et al*. 2012). Several studies demonstrated that marks are more often left by animals with high social status (dominance/treat hypothesis (Peters & Mech 1975; Mech 1999; Cafazzo *et al.* 2012; Jordan *et al.* 2013). Marks are usually deposited along territory periphery (Cafazzo *et al.* 2012).

Olfactory communication guides many different social behaviours such as kin recognition, mating, territorial aggression, defensive responses (Brennan & Kendrick 2006; Isogai *et al.* 2011). However, olfaction in canids is often complemented by other types of communication. For example during mating periods olfactory, visual and vocal communication are used in combination. Breeders usually leave marks full of pheromones, and the female in particular changes its body posture and uses vocal signals to show to the male that the moment is favourable to mate.

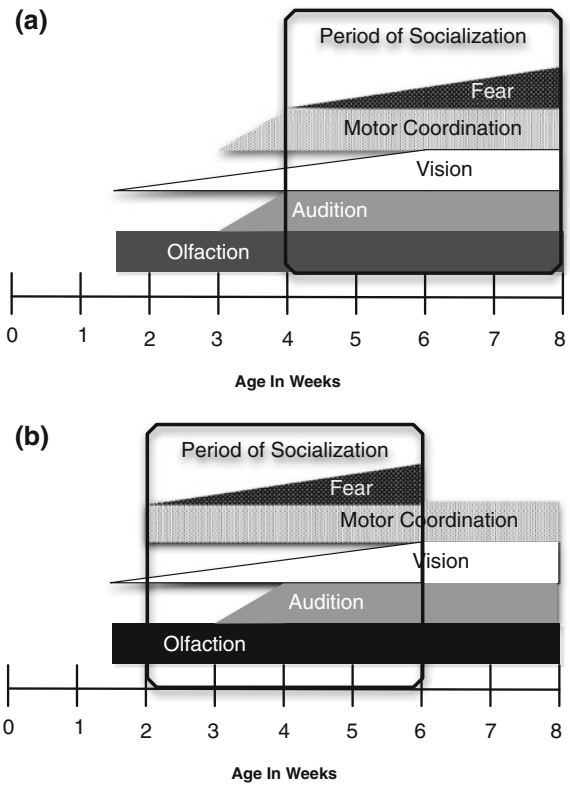


Figure 4

a) Classification of early development in dogs based on findings from previous literature (see discussion for sources) The beginning of a bar represents the first appearance; the maximum width of a bar repre- sents maturity. (b) Classification of early development in wolves given current data. Wolves and dogs develop senses at the same time. The difference between dogs and wolves is in the time they start socialization, which deeply impacts on the adult behaviour, thus dogs evolved a reduced flight distance and higher ability to form interspecies social bounds. (From Lord 2013)

Comparing dog and wolf behaviours correlated to olfaction, Lord (2013) tried to assess the differential importance of the smell in these two canids. During pups weaning, olfaction is important to create a bound between the mother and the litter and also with siblings and other relatives, but parental cares are very different in wolves and dogs.

Dogs and wolves differ in the timing they usually start to socialize. Although wolves and dogs develop their senses at the same time, wolves usually start to explore at two weeks of age, when they are able to smell but still partially blind and deaf but, whereas dogs begin at four weeks, when their senses of sight, hearing, and smell are already developed. This difference means that while dog pups can use all their senses in familiarizing with the world around them, wolves prevalently rely on smell, which will likely cause an elusive and more vigilant behaviour than dogs in the adulthood (Figure 4; Lord 2013).

INTRODUCTION TO PAPER IV

Silvana Lapalombella, Marco Galaverni, Romolo Caniglia, Sara De Fanti, Douglas Scofield, Elena Fabbri, Robert Ekblom, Ettore Randi. "**Canid olfactomics: olfactory receptors in wolves and dogs**" (in preparation)

According to their sense of smell, animals can recognize clues related to the perception of the external environment such as presence of food, preys and predators, mates and kin or even intruders. Depending on the species, olfaction abilities can play marginal or fundamental role in social behaviour and survival. This sense is particularly complex from many aspects, and still poorly known. A combinatorial interaction of possibly unlimited number of ligands with receptors drove biological systems to specialize in anatomy with peculiar structures, and increasing the number of genes. Nevertheless, the expanding availability of genomic and bioinformatic tools is opening the way to study the complex genomics of olfaction. Here we present results from a comparative study on olfactory receptor genes sequenced using next-generation technologies in dog and wolf samples in the PAPER IV.

PAPER IV

Silvana Lapalombella, Marco Galaverni, Romolo Caniglia, Sara De Fanti, Douglas Scofield, Elena Fabbri, Robert Ekblom, Ettore Randi. "**Canid olfactomics: olfactory receptors in wolves and dogs**" (in preparation)

Author Contributions:

Authors' contributions are listed in alphabetical order. Conceived and designed the experiments: MG SL ER. Performed the experiments: SDF SL. Analysed the data: MG SL. Contributed reagents/ materials/analysis tools: RE SDF EF RC DGS ER. Wrote the first draft of the paper: MG SL. Critically revised the manuscript RC DGS ER RE. All authors approved the final version.

PAPER IV

Canid olfactomics: olfactory receptor genes in wolves and dogs

RUNNING TITLE: Detecting selection and variability of olfactory genes in dogs and wolves

Silvana Lapalombella (SL), [*] Marco Galaverni (MG), [¶] Romolo Caniglia (RC), [¶] Sara De Fanti (SDF), [**] Douglas G. Scofield (DGS),[†§], Elena Fabbri (EF), [¶] Robert Ekblom (RE) [†] and Ettore Randi (ER) [¶††]

[*] Department of Biological, Geological and Environmental Sciences (BiGeA), University of Bologna, via Selmi 3, 40126, Bologna, Italy

[¶] Laboratory of Conservation Genetics, ISPRA, via Cà Fornacetta 9, 40064 Ozzano dell'Emilia (Bologna), Italy

[**] Laboratory of Molecular Anthropology, Centre for Genome Biology Department of Biological, Geological and Environmental Sciences (BiGeA), University of Bologna, via Selmi 3, 40126, Bologna, Italy

[†] Department of Ecology and Genetics: Evolutionary Biology, Uppsala University, Evolutionary Biology Centre, 75236 Uppsala, Sweden

[§] Uppsala Multidisciplinary Center for Advanced Computational Science, Uppsala University, Uppsala, Sweden

[††] Department 18/Section of Environmental Engineering, Aalborg University, 9220, Aalborg, Denmark

Corresponding author:

Silvana Lapalombella (silvapaloma@gmail.com)

Department of Biological, Geological and Environmental Sciences (BiGeA), University of Bologna, via Selmi 3, 40126, Bologna, Italy

Key words: olfactory receptors genes; OR; degenerate primers, NGS; Ion Torrent PGM; wolf; dog; selection

Abstract

Olfaction plays a crucial role in social behaviour and prey detection in large carnivores. These abilities are mediated by the integrated actions of different olfactory receptor (OR) genes. More than one thousand OR genes have been described in the dog genome, but the variability of OR genes is still poorly known in dog closest relative, the wolf. Here, we present comparative analyses of dog and wolf OR genes sequenced through the Ion Torrent PGM platform using canid-specific degenerate primer pairs, to amplify an estimated 59% of the known canine OR sub-genome. We sequenced samples from 16 dogs in Italian shelters and 16 wolves from wild-living packs in Italy. Sequenced reads were mapped to the dog genome assembly, and after a stringent filtering pipeline, 775 high coverage polymorphisms (733 SNPs and 42 InDels) were discovered in a total of 250 OR genes (50 class I, 196 class II, 4 unclassified). A significantly higher number of variants was detected in dogs class I vs. class II genes, together with a significantly higher number of haplotypes and gene diversity. We found no difference in the mean number of variant sites per locus between dogs and wolves. We observed significant skews from neutrality in both groups, averaging more positive values of Tajima's *D* in wolves, likely a sign of the recent bottleneck experienced by the Italian wolf population (Fabbri *et al.* 2007). Ka/Ks values suggested overall stabilizing selection acting on both dogs and wolves, but with differential pressures on some specific ORs possibly responsible for recent adaptations.

Introduction

Olfaction is a complex sense that deeply impacts animal survival by allowing the detection of environmental cues useful to discriminate the presence of prey and predators, food, kin and mates. The Olfactory Receptor (OR) genes, firstly described in *Rattus norvegicus* in 1991 (Buck & Axel 1991), play essential functional roles in vertebrates, and are the largest gene family yet identified, corresponding to more than 2% of the mammalian genes (Firestein 2001). The OR genes reside in clusters mapping to several chromosomes, likely derived from expansion or dispersion of ancient groups of genes present in a common ancestor (Aloni *et al.* 2006). ORs belong to the G protein-coupled receptor (GPCR) group, the structure of which contains seven hydrophobic transmembrane domains (TM). Within ORs, families and subfamilies are identified by sequence homology. Based on their phylogenetic relationships (Glusman *et al.* 2000), ORs are classified into two classes: class I, defined as the fish-like receptors (Freitag et al. 1995; 1998), but present also in terrestrial vertebrate genomes (Niimura 2010); and class II, the tetrapodan receptors. The different expression of the two classes of genes in vertebrates and the preferential interaction of the receptors to hydrophilic and hydrophobic agonists, likely mean that first class receptors evolved to bind preferentially water-born molecules and the second class volatile odorants (Eisthen 1997; Saito *et al.* 2009; Niimura 2010).

Among mammals, dogs have an extremely sensitive olfactory system, and they are often trained to use this system to assist humans in finding hidden substances like explosives and drugs, biological clues such as volatile organic compounds that are linked to cancers (Boedeker *et al.* 2012), human traces in rescue operations, and scats from wild animals in wildlife management (Long *et al.* 2007; Arandjelovic *et al.* 2015). Canine olfactory abilities are the result of a combination of anatomical and genomic characteristics. The main olfactory epithelium in the dog nose is almost 20 times larger than in humans (Issel-tamer *et al.* 1997)

and is accompanied by a peculiar airflow generated by the presence of the dorsal meatus, which bypasses the respiratory airways and prevents the purging of odorants from receptors during expiration, enhancing odor discrimination (Craven *et al.* 2010).

Several studies have examined OR genes in dogs and have found high levels of diversity, likely due to a combination of low selective constraints and on-going gene conversion (Quignon et al. 2012). However, only two papers (Zhang *et al.* 2011; Chen *et al.* 2012) thus far have examined OR genes in wolves, the wild sister group to dogs, and these have compared the genetic variability of a limited number of pseudogenes observing not significant difference in the proportion of pseudogenes between dogs and wolves but presence of strong purifying selection in dogs breeds.

The patterns of OR variation within the genomes of dogs and wolves could have been shaped by differing natural or artificial selection pressures, as well as by distinct demographic histories. Wolves are widespread in the Old and New World, owing to their adaptability to a variety of climatic and ecological conditions (Geffen *et al.* 2004; PILOT *et al.* 2006; Musiani *et al.* 2007). Nevertheless, European wolf populations have declined and have disappeared from parts of their historical range due to habitat changes and direct human persecution (Boitani 2000). Although some European wolf populations have started to recover in the last few decades, wolves are still locally threatened by illegal and incidental killing and hybridization with free-ranging dogs (Randi 2011; Godinho *et al.* 2011; Randi *et al.* 2014). Domestic dogs diverged from their wild counterpart between 11,000 and 36,000 years ago, possibly through multiple events of domestication and backcrosses involving various possibly now-extinct wolf populations (Lindblad-Toh *et al.* 2005; Freedman *et al.* 2014; Wang *et al.* 2016; Fan *et al.* 2016).

Domestication and associated selection of modern breeds have deeply modified a number of phenotypic and genetic traits in dogs (Vonholdt *et al.* 2010; Vaysse *et al.* 2011; Axelsson *et al.*

2013), and have also left a peculiar genetic signature in dog ORs (Robin et al. 2009). In the past, given the large number of genes and high level of polymorphism, sequencing and characterizing a significant portion of OR genes was difficult or exceedingly expensive, even when using degenerate primers (Buck & Axel 1991; Young *et al.* 2002; Linhart & Shamir 2002; Gilad *et al.* 2004). However, with the advent of new molecular methods via high-throughput sequencing, it is now possible to sequence thousands of genes in a few hours (Mainland *et al.* 2013) at reasonable cost. To take best advantage of this technology, we did not use human degenerate primers for ORs as has been done in past studies (Gilad *et al.* 2003, 2004; Hughes *et al.* 2013; Hayden *et al.* 2014); instead we designed new dog-specific degenerate primers. We also chose to sequence with Ion Torrent because of its scalability in library preparation and its long read length, which is helpful when mapping highly similar sequences.

Our main objectives were: 1) develop a reference OR database containing all sequences that could be reliably recognized as ORs; 2) design dog-specific degenerate primers and test their amplification efficiency and accuracy via a novel bioinformatic pipeline able to cope with degenerate pairing; 3) identify OR variants and their differences between dogs and wolves; and 4) determine OR genes under selection in dogs, wolves, or both.

Materials and Methods

Olfactory subgenome reference database reconstruction

In order to construct a reference database of canine OR sequences, we collected the 1121 OR sequences described in Robin et al. (2009) and all transcripts from the NCBI Seq-Gene database annotated as ORs (ftp://ftp.ncbi.nih.gov/genomes/Canis_lupus_familiaris/mapview/). We used these sequences as queries to search the latest dog genome assembly (canFam3.1) using the Basic Local Alignment Search Tool (Blast) implemented in Geneious v.7 (Biomatters, Auckland New

Zealand, http://www.geneious.com/). For each sequence and transcript, only hits with a minimum threshold of 98% identity were retained, and if two or more hits mapped to the same region, only the longest was kept. If one sequence mapped with high identity to two different genomic locations, only the hit with the highest identity was retained, whereas hits with equally high identity scores were all discarded. The locations of the hits retained from Blast were used to extract the corresponding sequences from the canFam3.1 reference genome using BEDTools package v. 2.21.0 (Quinlan & Hall 2010). These sequences were further checked to be reliable OR genes using the HMM-based Olfactory Receptor Family Assigner in the BioPerl module Bio::ORA (Hayden *et al.* 2010). We thus reconstructed a reference OR dataset using only genomic regions that satisfied both these criteria.

OR protein classification

For the nomenclature and classification of OR sequences we followed Robin et al. (2009). We assigned the few unclassified sequences to the respective class I or II genes using the following pipeline. Sequences were grouped into smaller sets and translated and aligned with MACSE (Ranwez *et al.* 2011) using default settings. All sequences showing frameshift or disrupting codons were manually discarded and only those with a complete open reading frame (ORF) were selected and aligned with Muscle (Edgar 2004). From this alignment, a phylogenetic neighbour-joining (NJ) tree was constructed in MAFFT v.7 (Katoh & Standley 2013) under the Jones-Thorton-Taylor (JTT) model for amino acid sequences (Jones *et al.* 1992). OR sequences were then assigned to class I or II according to their clade in the tree. Pseudogenes were identified based on previous classification (Robin et al. 2009). Polymorphic pseudogenes were further checked by comparison to recently released NCBI RefSeq Genes 103 using Golden Helix Genome Browser 2.0, and retained as pseudogenes only if falling into RefSeq-classified noncoding genes, or if containing a stop codon such that the open reading frame (ORF) was shorter than 290 amino acids (Olender *et al.* 2004).

Degenerate primer design

In previous studies, human degenerate primers (Hughes et al. 2013; H. Zhang et al. 2011; Olender et al. 2004) were used to try to amplify the majority of OR genes in different species. Here we developed two new canid-specific degenerate primer pairs designed using the program HYDEN (Linhart & Shamir 2002). In order to maximize the chance of amplifying both class I and II genes, primers were designed for each class separately. After removing sequences with ambiguous positions from the two input alignments, for each targeted class, the best pair of forward and reverse degenerate primers was retained, according to the highest number of matching sequences, with a degeneracy threshold of 10^4. The obtained degenerate primer sequences were: class I 5' (cI_DP_F) CTNCAYVARCCYATGTAYYWYTTBYT, class I 3' (cI_DP_R) GTYYTNACHCYRTAVAYRATRGGRTT, class II 5' (cII_DP_F) CTNCANWCNCCHATGTAYTTYYTBCT, class II 3' (cII_DP_R) TTYCTBARRSTRTARATNADRGGRTT. The expected amplification success was defined by HYDEN as the proportion of OR genes amplified with up to 2 bp mismatch in both directions.

Choice of samples, DNA extraction and quantification

Muscular tissue samples were taken from 16 wolf individuals (killed in road accidents or by poaching) collected throughout the distribution range in Italy. Blood samples from 16 dogs not belonging to any specified breed were collected in Italian kennels by authorized staff on behalf of ISPRA. No animal was hurt nor sacrificed for the purpose of this study. DNA was extracted using the DNAeasy Blood & Tissue kit (Qiagen, Valencia, California, USA) according to the manufacturer's instructions. Genomic DNA was visualized on 1.2% agarose SYBR Safe DNA e-gel (Life Technologies, Carlsbad, California, USA) and quantified with a Qubit fluorimeter v.2.0 using the dsDNA BR assay kit (Life Technologies).

PCR conditions

PCRs were performed using a Veriti® 96-Well Thermal Cycler (Applied Biosystems) with a temperature gradient specific for each degenerate primer pair to maximize the number of potentially amplified genes (Gilad *et al.* 2004). The reactions were carried out in a total volume of 20 µL containing: 100 µmol/L deoxynucleotides, 0.5 µmol/L of each primer, 1x PCR buffer, 0.25 unit of Taq (5Prime), and 25 ng DNA. PCR conditions were modified from (Hayden *et al.* 2010): a first step of denaturation was run at 94°C for 2 min, followed by 45 cycles of denaturation at 94°C for 15 s, with annealing at a gradient temperature of 38°C to 48°C for 40 s for cI_DP, and 38°C to 58°C for cII_DP (both with a ΔT of 2°C), and extension at 72°C for 1 min. The final step was an extension at 72°C for 10 min. PCR products were purified using Exo-SAP IT (Affymetrix Inc), visualized on 2% E-gel with E-gel low-range quantitative DNA ladder (Life Technologies), and quantified with a Qubit fluorometer (Life Technologies). The best range of amplification temperatures was then chosen based on the presence of good-quality products in both dog and wolf samples. PCR products amplified at 42°C and 44°C were thus gathered for cI_DP pairs, and 48°C and 50°C for cII_DP pairs. For each individual, the amplification products at each selected temperature were quantified, equimolarized and joined into one mix.

Library preparation and sequencing

For each sample, 100 ng of PCR product were enzymatically fragmented using the Ion Xpress Plus Fragment Library Kit protocol. Amplicons sizes were selected on Size Selection e-gels (Invitrogen) to reach a target read length of about 400 bp. Each sample was labelled using IonXpress Barcode Adapters and quantified by qPCR Real-Time. The labelled fragments were clonally amplified using the Ion PGM Template OT2 400 Kit with the Ion OneTouch 2 Instrument. A final step of enrichment was performed with the Ion OneTouch ES Instrument. The quality of the libraries was checked with a Qubit fluorometer to quantify the presence of

polyclonal particles. Two fully amplified and purified template pools, one for each degenerate primer pairs' product, were injected on two Ion Torrent PGM 316 sequencing chips v.2 (Life Technologies), respectively named chip I and chip II.

Sequencing validation and mapping

Sequencing data were imported in the Ion Torrent Suite v.4.0 Software. Reads were trimmed as default for low signal quality and/or for the presence of mixed DNA templates in a given well. The 3' ends of reads were scanned for matches to the adapter sequences and for trailing regions of low quality. Short reads (< 4bp) or with off-scale ionograms, based on default values, were filtered and removed, along with reads likely to be polyclonal or with adapter dimers. Only high-quality portions of reads, with Phred per-base quality scores > 15, were written into an unmapped BAM file, and then mapped to the reference canFam3.1 genome with the Torrent Mapping Alignment program (TMAP), which integrates different alignment algorithms: BWA-short (Li & Durbin 2009), BWA-long (Li & Durbin 2010), SSAHA (Ning *et al.* 2001), and Super-maximal Exact Matching (Li 2012).

SAMtools mpileup (Li *et al.* 2009) was used to extract coverage information from BAM files (chip I and chip II data), which were then processed with the intervalBed script (https://github.com/douglasgscofield/bioinfo/blob/master/scripts/intervalBed) to obtain a BED file describing the genomic intervals with a read coverage of at least 15x. Length of intervals and distances of separation were calculated. Intervals that ranged between 650 and 750 bp, compatible with the expected amplicon sizes (Figure 1), were extracted from both chips as putative targets. Finally, sequence coverage per locus and per sample metrics were assessed with the Genome Analysis ToolKit (GATK) v.3.3.0 (McKenna et al. 2010) DepthOfCoverage tool.

OR identification

The locations of the intervals passing all the filtering steps were compared with our OR reference database using BEDTools, then classified into amplified loci falling inside our reconstructed OR genomic locations (INREF) or other amplified regions not falling into our known OR intervals (OUTREF). In order to verify their possible match with ORs not included in the reference database, the corresponding OUTREF sequences were extracted from canFam3.1 and checked by both NCBI blast search, avoiding the retention of any sequence with 100% identity to two or more different locations, and with Bio::ORA. Regions recognized as putative ORs from both methods were manually annotated following the Bio::ORA starting frame prediction. As above, annotation was then visually checked against the NCBI RefSeq Genes 103 genomic track in GenomeBrowse whenever a corresponding hit was available.

Sequencing success

From all validated loci, we calculated the observed amplification success (AS) in terms of total AS, the percentage of amplified loci on the total number of OR genes identified, and class-specific AS, the percentage of amplified class I and II loci over the total number of class I and II reference genes, respectively. We then calculated the primer-specific AS as the proportion of amplified base pairs on the total gene length. Per-family and per-cluster mean rates of amplification were calculated as the mean proportion of amplified loci per family and per cluster, over the total number of loci described in the OR reference dataset. Amplification bias was checked using a Fisher's exact test (FET), comparing the number of amplified genes over the number of reference genes, counted per family and per cluster. Among families and clusters showing a significant deviation from expectations, we only considered as significant those with at least 3 genes expected to be amplified and with more than 50% deviation from expectation.

Trimming degenerate primer sequences from sequenced data

In order to trim the degenerate primer sequences from the sequenced data and avoid the possible inclusion of false variants, we applied the DISPR tool (https://github.com/douglasgscofield/dispr), which implements degenerate in-silico PCR by scanning a reference sequence using pairs of degenerate primer sequences. DISPR was used to find all sequences in the dog genome (canFam3.1) matching any of the $\sim 10^4$ unambiguous 26-bp sequences of the degenerate primers, allowing up to 5 mismatches in the first 10 bp from each 5' degenerate primer start and up to 2 additional mismatches in the remaining 16 bp of each degenerate primer. Once the primer pairings were retrieved along the genome, DISPR identified in-silico amplicons produced by forward and reverse primers and produced a BED file of internal amplicon portions with primer sequences subtracted. The BAM files of read alignments were then subset using this BED file to include just those reads which overlapped the internal portions of amplicons.

Identification of variants

To identify single nucleotide polymorphisms (SNPs) and insertions or deletions (InDels), we used GATK. Starting from subset BAM files covering internal portions of amplicons, we performed a local realignment around InDels (DePristo et al. 2011) and separated the resulting realigned multisample bam files into single-sample files. We produced genotypes using the HaplotypeCaller (HC) method, with the options ERC (emit reference call), GVCF (genotype variant calling format) and DISCOVERY mode, setting the emission confidence threshold to 10, and the minimum calling confidence to 30, filtering out reads with Mapping Quality (MAPQ) < 20. Calls were thus limited to the internal portions of amplicons as identified by DISPR. We then joined all emitted genotypes (GVCF) for each sample with GenotypeGVCF tool, to perform multisample variant calling of all sites evaluated to be potentially variable by the HC method. Finally, standard hard filtering parameters was

applied, according to GATK Best Practices recommendations (Van der Auwera *et al.* 2013), followed by an additional set of user-defined filters (Supplemental Table 3).

Variant analysis

The SNP & Variant Suite (SVS) software v.8.0.1 was used for the analysis of variant sites. We filtered the VCFs obtained from chip I and chip II, only retaining sites with genotype quality (GQ) > 20, which should reflect a 99% accuracy (Carson *et al.* 2014). VCFs from the two chips were then merged into a single file. When only one genotype was present, it was directly retained, whereas in case of discordant genotypes only the one with highest GQ was considered. We then retained only variant sites that were reliably called in at least two individuals per group. Genotypes from sites that were called in both sequencing chips (overlapping sites) were compared, and a discrepancy rate was calculated as the number of discordant genotypes over the total number of overlapping genotypes. We also determined the frequency of false alleles (FA) and allelic dropout (ADO), considering the genotype with the highest GQ as the correct one.

Bias toward a particular family or cluster in the identification of variants was evaluated by a FET, comparing the number of polymorphic loci identified vs. the number of amplified loci per family and per cluster. Families and clusters showing at least 3 expected polymorphic loci and more than 50% of deviation from expectation were considered significantly deviating.

To test the hypothesis that larger clusters host more polymorphic loci (Robin et al. 2009), Spearman's correlation was checked between cluster size (the number of OR genes described in a cluster in the reference database) and the mean number of variable sites per locus. Genes were not even distributed among clusters and thus cluster sizes were classified as small (including from 1 to 60 ORs) and large (170 to 222 ORs). The distribution of the mean number of SNPs per locus was checked over small and large clusters in both groups. We also

compared the mean number of genes that had a smaller number of SNPs (from 0 to 5) and a larger number of SNPs (more than 6).

Finally, the SVS Variant Classification function was applied to identify and summarize the different types of mutations (synonymous, non-synonymous, frameshift, or stop).

F_{ST} was calculated at each polymorphic site, and a genotype principal component analysis (PCA) was carried out over all samples. An estimate of genotype uniqueness was reported as the percentage of samples from each group that contained a unique genotype or a genotype that was not present in the other group of samples. Finally, SNP density was calculated as the frequency of SNPs across the total sequenced base pairs.

Haplotype reconstruction

Haplotypes from loci found to be polymorphic were reconstructed using all sites that had no missing alleles in all sequenced samples, in order not to add any possible bias. Complete gene sequences were obtained with the GATK tool "Fasta Alternate ReferenceMaker", which replaces the reference CanFam3.1 bases at variant sites with the bases provided by the multisample VCF file. The PHASE algorithm implemented in DnaSP software (Librado & Rozas 2009) was then used with 3000 iterations and 100 burn-ins to reconstruct phased haplotypes at each locus, with the exception of loci that included InDels.

Neutrality and selection tests

The number of haplotypes and segregating sites, nucleotide diversity, and values of Tajima's *D* and Fu-Li's *D* with associated statistical significance were obtained with DnaSP in batch mode. Ka/Ks values were calculated with KaKs-Calculator (Zhang *et al.* 2006) using the NG (Nei & Gojobori 1986) method on reconstructed haplotypes in functional loci. Significance was tested with Fisher exact tests.

A rank-based meta-analysis was performed to find loci showing different selective pressures between dogs and wolves. For this purpose we used differences in measures of Tajima's *D* (ΔD) and nucleotide diversity ($\Delta\pi$) (W. Zhang et al. 2014; Freedman et al. 2015 *under review*) between the groups. Genes were ranked considering the first 5% of most positive and most negative values, and their absolute differences. Top high-ranking loci were then identified, only retaining those loci that showed highest 5% ranks on both ΔD and $\Delta\pi$ tests.

Results

Olfactory reference database

We identified 1012 OR genes mapping on 26 chromosomes (including chromosome X) and on 22 unknown chromosomal contigs (Supplemental Table 1). The OR reference dataset included 781 class II, 174 class I and 57 un-classified (n.a.) genes (Supplemental Table 1), assigned to 23 families and 296 subfamilies. Forty-eight genomic clusters were identified and renamed following the new mapping locations on canFam3.1. All OR loci were included in clusters already described in the previous assembly, except for locus CfOR16F03/ST_XM_538213.3, that we assigned to a new pseudo-cluster, CFA31@1.

Raw sequencing results and amplification success

We obtained successful loading of 88% and 69% of the wells from chip I and chip II, respectively, corresponding to 3,418,815 and 2,980,317 total reads, with a mean length of 300 and 308 bp (1,000 Mb and 915 Mb of total sequences). The percentage of polyclonal reads was 33% and 24% (S. Table 2 and S. Figure 1). About 99% of reads from both runs mapped to the dog genome assembly canFam3.1.

The majority of amplicons (97% in both chips) were mapped within olfactory genomic regions described in our reference dataset, corresponding to portions of 375 and 443 OR

337 genes in chip I and II, respectively (Table 1). From reads mapping outside our reference OR
338 dataset, 22 additional regions (13 at chip I and 15 at chip II) were identified as putative ORs
339 and included in the dataset. The total amplification success (AS) was 37% at chip I and 43%
340 at chip II. At chip I, both class I and II loci were amplified (class I 28%, class II 67%, n.a. 3%),
341 whereas at chip II the large majority of loci were from class II (class I 5%, class II 95%, n.a.
342 3%). Class-specific AS was 60% and 54% at class I and II respectively, compatible with our
343 expectation of 51% of class I ORs amplified with DP_cI primers and 61% of class II ORs with
344 DP_cII primers. On average, we amplified 74% of the full gene lengths (the primer-specific AS)
345 with DP_cI (703 ± 16 bp) and 75% with DP_cII (704 ± 12 bp).
346 After joining data from both chips, we obtained sequences for 597 different genes, with 221 of
347 them amplified in both chips. Amplified genes were distributed across 19 families, 124
348 subfamilies, and 42 clusters (Supplementary Tables 8 and 9). Overall, we found no bias in
349 amplification across families (FET = 393, 2-tailed P = 1), but a skew across clusters was
350 detected (FET = 908, 2-tailed $P < 0.01$), with a significant excess of amplified loci in cluster
351 11@52 and a deficit in another five (3@31, 10@16, 27@1, 33@5 and X@102).
352 Coverage was variable among individuals and genes. Across the two sequencing runs, 395
353 genes (252 and 294 respectively for chip I and chip II) were sequenced at the fixed minimum
354 coverage and quality (15x, MAPQ ≥ 40) to be considered for the calling of variants
355 (Supplementary Figure 2), with 45 OR loci from chip I and 64 from chip II sequenced in all
356 samples.

357 **Genotype calling and variant identification**

358 After applying stringent hard filters (Supplementary Table 3), we identified a total of 775
359 polymorphic sites. The cross-validation of genotype calls between chips was performed on
360 5152 overlapping genotypes, in which we found only 52 discrepancies (corresponding to a
361 0.99 concordance rate), consisting in 14 false alleles (FA frequency 0.0027) and 38 cases of

allelic dropout (ADO frequency 0.0073). We could reliably call a total of 10,467 genotypes at polymorphic sites across dog samples and 9,570 across wolf samples. The average number of variant genotypes called per sample was significantly higher in dogs (654 ± 64) than in wolves (598 ± 78; 2-tailed t_{30} = 2.2, P = 0.04), as well as the mean call rate per sample (0.84 ± 0.08 in dogs and 0.77 ± 0.10 in wolves; 2-tailed t_{30} = 2.2, P = 0.04). We found no difference in the genotype quality distribution, measured as the number of sequenced bases at growing Genotype Quality between the two groups of samples in both chips (Mann Whitney U test: Chip I, U = 30, 2-tailed P = 0.14; Chip II, U = 39, 2-tailed P = 0.44; Supplementary Figure 3).

The total number of variants per group was not significantly higher in dogs than in wolves, with 624 dog variants (584 SNPs, DEL 37, INS 3) vs. 577 wolf variants (538 SNPs, DEL 36, INS 3; chi-square test P = 0.18). The observed SNP density was 4.4/kbp overall, with 3.5/kbp in dogs and 3.2/kbp in wolves. The mean rate of transitions over transversions was not significantly different between the two groups (Ti/Tv 3.57 ± 0.37 vs. 3.45 ± 0.35; Mann Whitney U = 99, 2-tailed P = 0.29).

Variants were identified in a total of 250 loci (50 Class I, 196 Class II, 4 n.a.) falling respectively into 232 genes in dogs and 221 in wolves, distributed across 15 families (3 Class I, 12 Class II; 15 dog and 14 wolf) and 30 clusters (29 dog, 30 wolf; Supplementary Tables 8 and 9). Differences in the occurrence of variants were observed among families and clusters in both groups (Supplemental Table 5). Most of the polymorphic loci (90%) contained from 1 to 5 variable sites, with only 10% having more than 6 variable sites (Figure 5). Notably, the genes with the highest number of SNPs in class I and class II were the pseudogenes CfOR2195 (17 SNPs in dog, 13 in wolf) and CfOR10Fo7 (14 SNPs in dog, 15 in wolf).

No significant differences were found in the mean number of polymorphic loci per family and per cluster between dog and wolf (Supplementary Table 6). The mean number of variants per locus was not different between dogs and wolves (2.50 ± 0.14 in dogs vs. 2.31±0.14 in

wolves), as well as per family and per cluster (Supplementary Table 6). In dogs, the mean number of variants per locus was significantly higher in class I (3.20 ± 0.40) than in class II (2.31 ± 0.15), but this difference was not significant in wolves (2.58 ± 0.34 class I vs. 2.25 ± 0162 in class II; Supplementary Table 6). Moreover, differences between dogs and wolves within classes were not significant, neither at class I nor at class II.

We found a positive correlation between cluster size and the mean number of polymorphic sites in both groups (dogs, Spearman's ρ = 0.5; wolves, Spearman's ρ = 0.7; both $P < 0.01$). More specifically, the mean number of SNPs per locus was significantly higher in large than in small clusters in dogs (small 1.98 ± 0.15 vs. large 2.86 ± 0.27), but not in wolves (small 2.04 ± 0.18 vs. large 2.32 ± 0.23; Table 2a, Figure 6a). Moreover, we found that loci with a small number of SNPs (0-5) belonged to clusters with a mean size smaller than loci with a large number of SNPs (6-17). Such difference was significant in dogs, but not in wolves (Table 2b, Figure 6b).

Heterozygosity was not significantly higher in wolves than in dogs (0.28 ± 0.04 vs. 0.31 ± 0.07, respectively; 2-tailed t_{30} = 1.6, P = 0.11), although its variance was significantly larger in wolves (Levene's test 4.5, P = 0.042).

Some variants were polymorphic within one group, but not in the other; 151 sites were fixed in dogs, but variable in wolves, and 198 were fixed in wolves, but variable in dogs. None of the sites were found to be fixed for different alleles in the two groups.

Genetic distance between dogs and wolves

The PCA showed a clear subdivision of the two groups, well differentiating dogs and wolves (Figure 2). The overall F_{ST} between dogs and wolves was 0.267. Most of the markers (436; 56%) had a small F_{ST} (<0.15), but a few sites (15; 2%) showed F_{ST} values above 0.85 (Figure 3): 6 SNPs relied on 3 class I genes (family 51) and 9 SNPs fell into 4 class II loci (in families 5, 4 and 14) (Figure 4; Supplementary Table 4). Of these 15 SNPs, 12 were variable in wolves

412 but homozygous in all dogs, another two were heterozygous in both groups and only one was
413 homozygous in all wolves, but variable in dogs. These 15 SNPs had also a percentage of
414 uniqueness in both groups > 80% (Supplemetary Table 4).

415 **Mutation types**

416 Among the polymorphic functional ORs, we identified 263 synonymous, 223 nonsynonymous
417 and 33 frameshift mutations in dogs, and 226 synonymous, 224 nonsynonymous and 34
418 frameshift in wolves. Moreover, 3 stop mutations caused by 2 SNPs and one deletion were
419 identified in 3 class II loci (all 3 observed in dogs, 2 in wolves). However, frameshift and stop
420 mutations were always observed only in some of the samples, therefore we did not consider
421 the genes in which they were found as complete pseudogenes. The distribution of each
422 mutation type per gene was not significantly different between dogs and wolves (Table 3a). A
423 significant difference in the number of synonymous and frameshift mutations per gene
424 between class I and class II was found within dogs, but not within wolves (Table 3b). No
425 significant differences were found between dogs and wolves in the classes considered
426 individually (Table 3b).

427 **Pseudogenes**

428 Among the 250 polymorphic ORs, we identified 34 pseudogenes (6 class I, 25 class II, 3 n.a.)
429 belonging to 11 families and 14 clusters (Supplementary Table 8 and 9). The families with
430 more pseudogenes were families 51 and 52 for class I (3) and family 7 for class II (9), whereas
431 the clusters with the highest number of pseudogenes were CFA21@26-29 (7 class I genes)
432 and CFA20@46-47 (7 class II genes). Of these 34 pseudogenes, 31 were polymorphic in dogs
433 and 30 in wolves, with 102 variable sites (36 class I; 58 class II; 8 n.a.) identified in dogs and
434 92 in wolves (31 class I; 55 class II; 6 n.a).

No significant differences were found comparing the mean number of variant sites per group, per family, and per cluster (Supplementary Table 7a). The mean number of variants per pseudogene was significantly higher in class I than in class II in D (class I 6.00 ± 2.37 vs. class II 2.32 ± 0.53), but not in W (class I 5.17 ± 2.01 vs. class II 2.20 ± 0.57; Table 4). No significant differences were found testing D vs. W in each class, as well (Table 4). Finally, three SNPs with $F_{ST} > 0.85$ in pseudogene CfOR1294 were found to be variable only within wolves, but not in dogs.

Haplotypes in functional and pseudogenes

Haplotypes were reconstructed in 96 dog and 97 wolf functional genes. No significant differences were found between dogs and wolves in haplotype diversity, nucleotide diversity π and number of segregating sites (Table 5a). However, the mean number of haplotypes per gene was significantly higher in dogs than in wolves (2.6 ± 0.11 vs. 2.2 ± 0.09, respectively; Table 5a).

Comparing haplotypes between classes within each group, the mean haplotype number (class I 3.0 ± 0.2 vs class II 2.4 ± 0.1) and the haplotype gene diversity (class I 0.5 ± 0.04 vs. class II 0.3 ± 0.02) were found to be significantly higher in class I than in class II loci only in dogs (Table 5b).

Haplotypes were also identified for 26 dog and 25 wolf pseudo-alleles belonging to 10 pseudogenes (1 in class I and 9 in class II). No significant differences were detected between dogs and wolves in haplotype diversity, nucleotide diversity, number of segregating sites nor in haplotype number (Supplementary Table 7b).

Neutrality tests

Neutrality tests via Tajima's *D* and Fu-Li's *D* were performed in all those loci having more than one haplotype in dogs (87) and wolves (80) (Supplemental Table 10). Tajima's *D* and Fu-Li's *D*

459 median values were significantly different from zero (Wilcoxon test, $P < 0.001$). We found
460 positive mean values of Tajima's *D* (3.00 ± 0.93 in dogs, 5.40 ± 1.07 in wolves) and Fu-Li's *D*
461 (0.03 ± 0.63 in dogs, 0.3 ± 0.74 in wolves). On average, wolves showed significantly more
462 positive values of Tajima's *D* than dogs, while no significant difference was found at Fu-Li's *D*
463 (Table 5a). Among analysed genes, only a few showed significant deviations from neutrality
464 (FET, $P < 0.05$; Supplementary Table 10). In dogs, a positive deviation was found for two
465 genes in class II (CfOR0393/ST_XM_844824.1 and CfOR0219), and one in class I
466 (CfOR0100/ST_XM_003433014.1), whereas a significantly negative value with both Tajima's
467 *D* and Fu-Li's *D* was found for the class II gene CfOR01C09/ST_XM_542000.3. In wolves,
468 values of Tajima's *D* were significantly positive in class I gene CfOR0104 ($P < 0.05$), in class II
469 CfOR10F07 ($P < 0.01$) and CfOR16D01/ST_XM_003433509.1 ($P < 0.001$), which also had a
470 significantly positive value of Fu-Li's *D*.
471 Ranking the genes with the highest differences in the values of Tajima's *D* (ΔD) and nucleotide
472 diversity ($\Delta\pi$) between dogs and wolves, two class II genes showed to be in the top 5% rank
473 in both categories: CfOR16D01 and CfOR10F07 (Supplemental Table 11).

474 **Selection tests**

475 The Ka/Ks ratio was calculated in 64 functional genes in dogs and 55 in wolves (Supplemental
476 Table 8). No differences were detected in the mean values of Ka/Ks when comparing dogs vs.
477 wolves, or class I vs. class II genes in both groups (Table 6). Only one gene
478 CfOR08D04/ST_XM_542002.2 in wolves had a Ka/Ks ratio higher than 1, although not
479 significant after Fisher's exact test. Few genes were significantly divergent from neutral
480 expectations (at $P < 0.05$), all with Ka/Ks < 1: CfOR0104 (both dogs and wolves) and
481 CfOR2238/ST_XM_542382.3 (dogs) were found at class I, while CfOR5812/ST_XM_540679.4
482 (dogs and wolves), ST_XM_540680.4/CfOR2248 and CfOR08A02/ST_XM_003432842.1 (dogs),
483 CfOR0085/ST_XM_539687.3, CfOR0457/ST_XM_540562.1, CfOR0423/ST_XM_003432364.1

484 and CfOR4598 (wolves) were found at class II. Only one gene was highly significant (P <
485 0.005), namely CfOR16D01/ST_XM_003433509.1 (wolves; Supplemental Table 11).

486 Discussion

487 The advent of high throughput sequencing technologies made sequencing experiments on
488 wide sets of genes more affordable. However, amplifying most of the olfactory subgenome in a
489 fast and cost-effective way is still a challenging task. This project was the first attempt to
490 sequence the two main classes of OR genes in a panel of dog and wolf samples, by applying
491 two newly designed canid-specific degenerate primer pairs on the Ion Torrent PGM
492 sequencing platform.

493 More than one thousand genes characterize the canine olfactory subgenome. We located 1012
494 OR genes in the latest dog genome assembly CanFam3.1, similarly to what described by other
495 studies, based on previous assemblies, in terms of number of genes, clusters and families
496 (Quignon et al. 2005; Robin et al. 2009; Quignon et al. 2012).

497 Sequencing results showed a total output per chip of almost 1 Gb, close to the highest possible
498 output for a 316 chip, which should range between 600 Mb and 1 Gb. We found a significantly
499 higher call rate in dogs than in wolves, which could be caused by different types of samples
500 used for DNA extraction (blood vs. muscular tissue) and/or could be caused by various
501 sources of reference bias such as better pairing of primers that were designed based on the
502 dog genome sequence, and better mapping of dog reads to the dog reference genome.
503 However, read mapping rates were uniformly high for both groups and genotype qualities did
504 not significantly differ between dogs and wolves.

505 Another major challenge in high throughput sequencing is the correct identification of
506 insertions and deletions, especially in non-model species (Ekblom & Galindo 2011). Ion
507 Torrent PGM has been usually described to be less accurate than light-based technologies

508 such as Illumina, but its accuracy becomes comparable with sufficient coverage (Quail *et al.*
509 2012; Loman *et al.* 2012). Our application of rigid filters to avoid high rates of false positives
510 (Quail et al. 2012) was successful, as shown by less than 1% discordance rate on genotypes
511 called independently from the two runs.
512 The use of degenerate primers offers the advantage of enhancing the number of amplified
513 targets, but also emphasizes the difficulty of retrieving the complete gene sequence, since
514 degenerate primers are usually designed to match the most conserved domains within genes.
515 Overall, we could amplify 597 genes, corresponding to 59% of canine ORs, and for these we
516 could sequence more than 70% of the full length, as expected from our primer design. The
517 new degenerate primer pairs used in this project (DP_cI and DP_cII) were designed to
518 preferentially amplify canid class I and II genes, respectively, and successfully amplified more
519 than half of the genes targeted by each (60% and 54%). However, DP_cI amplified ORs from
520 both classes, possibly due to bias in the high levels of sequence identity that exist between OR
521 classes, and also due to the larger number of class II than class I genes. Conversely DP_cII was
522 more specific, since the majority of the amplified genes belonged to class II.
523 OR genes can be classified into families and subfamilies and are unevenly distributed across
524 the genome. Following multiple duplication events in their evolutionary history, genes from
525 the same family are present in different genomic clusters and chromosomes (Glusman *et al.*
526 2000), with the exception being class I genes, which all rely on the same chromosome and
527 cluster. Despite this complexity, we succeeded in amplifying the majority of families and
528 clusters.
529 Based on our samples, polymorphisms were identified in 250 genes, corresponding to 42% of
530 the amplified loci and to 775 variants. The missed identification of variants in the remaining
531 genes is unlikely to be due to a real absence of polymorphisms in the studied samples, at least
532 to the extent that it could have been caused by a high variation in quality metrics across

samples and genes. Application of our stringent filtering criteria could have resulted in a number of positions not reaching the minimum thresholds imposed for the reliable calling of variants. If stringent filtering has removed variants, it has done so with minimal impact on our comparisons; while we did detect some differences in the number of variants per gene identified across families and clusters, we did not find a difference between dogs and wolves in the mean number of polymorphic loci observed per family and per cluster. Overall, the observed SNP density in ORs in dogs (3.5/kbp) and wolves (3.2/kbp) was in line with the very high values found in previous studies on dogs, and even higher than typically observed within non-coding regions (Tacher et al.2005; Robin et al. 2009).

The total number of variant sites as well as the mean number of variants per locus was not significantly different between dogs and wolves, nor was the distribution of polymorphic sites per family and per cluster. As Robin et al. (2009) found on different dog breeds, we also found that larger clusters were more polymorphic than smaller ones in dogs, but this difference was not significant in wolves. Perhaps this could be interpreted as another molecular signature of domestication.

There were no differences in mean values of heterozygosity per sample. Several polymorphisms were fixed in one group and variable in the other, but interestingly we found no cases of variant sites with fixed differences between dogs and wolves. Indeed, the occurrence of such sites has been described to be very rare (0.5% of variant sites genome-wide) (Freedman *et al.* 2014), which is unsurprising given the relatively recent divergence between dogs and wolves. Similarly, although direct conclusions on the rate of pseudogenization could not directly be derived from our study because of the impossibility of obtaining full-length gene sequences, we did not find any cases of genes being pseudogenized exclusively in one group.

The principal component analysis clearly distinguished dogs from wolves based on OR genotypes. We found that the majority of markers had small F_{ST} values, and only 2% of variant sites had F_{ST} values above 0.85, indicating that these sites were likely the top markers differentiating groups. Most were fixed in dogs and heterozygous in wolves, and most were found in few genes (i.e., genes with high F_{ST} values) and could be applied as diagnostic markers in future studies of wolf-dog hybridization. We found a significantly higher mean number of variant sites per locus in class I than class II ORs in dogs, mostly due to a significantly higher number of synonymous mutations, but also to more nonsynonymous mutations. Conversely, a significantly higher number of frameshift mutations was found in class II than class I in dogs. In addition to their utility as markers, these patterns may indicate a relatively more dynamic recent history of OR genes of both classes within genomes of dogs.

The most polymorphic loci in both dogs and wolves were pseudogenes, with significantly more variants observed in dog class I than class II ORs. These differences were paralleled by a higher mean number of haplotypes per gene in dogs than in wolves, similarly explained by a significantly higher haplotype diversity in class I than class II genes.

Neutrality tests allow making inferences about the strength of selection, but can be strongly influenced by the effects of demography, such as recent population expansions or contractions. On average, we observed more positive values of Tajima's *D* in wolves than in dogs, which indicates an excess of intermediate frequency mutations caused either by balancing selection or by a bottleneck effect. Dogs have been strongly selected over time and breeds are affected by two bottlenecks, one during domestication and a later one during breed creation (Lindblad-Toh *et al.* 2005). The italian wolf population was severely reduced by direct persecution in the past century, but also during the previous thousands of years (Fan et al. 2015; Silva et al. 2016 in prep). It may thus be possible that the strong bottleneck experienced by Italian wolves has caused a loss of variability even higher than that

582 experienced by dogs. Another possible source of bias that could create skew toward positive
583 values of Tajima's *D* in both groups is the multisample genotyping pipeline, which possibly
584 underestimates low frequency mutations (Han et al. 2014).
585 The Ka/Ks ratio can give an idea of selective pressures acting on functional genes. We found
586 an average Ka/Ks ratio << 1 in both dogs (0.09) and wolves (0.12), which indicates the
587 presence of purifying or stabilizing selection acting on ORs. A recent comparative study on OR
588 orthologs from 13 placental mammals (Niimura *et al.* 2014) showed evidences of purifying
589 selection, with significantly smaller Ka/Ks values detected in class I than in class II genes.
590 Although not strictly significant, we found similar results, which could indicate stabilizing
591 selection acting to preserve OR functionality, especially on class I genes.
592 Only a few genes showed significant and specific signs of selection, deviating from neutrality
593 either at Tajima's *D* or at Ka/Ks. Interestingly, only genes CfOR0104 (class II) and CfOR5812
594 (class I) were shown to be under selection in both dogs and wolves, likely representing key
595 olfactory genes for both canid groups. Genes CfOR2238 (class I) and CfOR2248 and
596 CfOR08A02 (class II) deviated significantly from neutrality only in dogs, whereas CfOR0085,
597 CfOR0457, CfOR0423 and CfOR4598 (class II) did so only in wolves. These genes could have
598 been exposed to different and more recent selective pressures in the two groups as confirmed
599 by the top F_{ST} values found in CfOR2238 and CfOR4598 genes, similarly to the few genes that
600 were outliers for ΔD and Δπ, namely CfOR0473 and CfOR16D01. Although they were not
601 previously described in literature, such genes are interesting targets for further investigation
602 in follow-up studies on larger sets of samples, since they could be responsible for adaptation
603 to the strongly different olfactory conditions to which dogs and wolves are exposed.
604 In conclusions, this study is the first genome-wide view of the divergence in olfactory
605 repertoire between dogs and wolves. Our comparative analysis confirms the high levels of
606 polymorphism that characterize canine ORs, as well as targeted adaptations as highlighted by

607 significant signs of selection acting on a small set of genes. However, the observed levels of
608 polymorphism did not significantly differ between dogs and wolves, indicating the general
609 action of stabilizing selection in both canid groups, although the demographic history of
610 Italian wolves likely also contributed to a strong reduction of OR variability due to their
611 extensive bottleneck. Notably, higher diversity at class I than class II loci was found in dogs
612 and deserves further investigation.
613 Of course, more conclusive hypotheses on the evolution of canine olfaction might be proposed
614 when many more complete canid genomes are available. Until then, our cost-effective pipeline
615 can easily be applied to characterize a wide range of ORs in dog breeds and wolf populations,
616 as well as in other groups for which a suitable reference genome is available. We need not
617 wait to deepen our knowledge of OR variability and evolution.
618

References

Aloni R, Olender T, Lancet D (2006) Ancient genomic architecture for mammalian olfactory receptor clusters. *Genome Biology*, **7**, 1–16.

Arandjelovic M, Bergl RA, Ikfuingei R *et al.* (2015) Detection dog efficacy for collecting faecal samples from the critically endangered Cross River gorilla (Gorilla gorilla diehli) for genetic censusing. *Royal Society Open Science*, **2**, 140423–140423.

Van der Auwera G a., Carneiro MO, Hartl C *et al.* (2013) From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. In: *Current Protocols in Bioinformatics*, pp. 11.10.1–11.10.33. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Axelsson E, Ratnakumar A, Arendt M-L *et al.* (2013) The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature*, **495**, 360–4.

Boedeker E, Friedel G, Walles T (2012) Sniffer dogs as part of a bimodal bionic research approach to develop a lung cancer screening. *Interactive Cardiovascular and Thoracic Surgery*, **14**, 511–515.

Boitani L (2000) *Action Plan for the conservation of the wolves in Europe (Canis lupus)* (No. 18-113). Council of Europe.

Buck L, Axel R (1991) A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell*, **65**, 175–187.

Carson AR, Smith EN, Matsui H *et al.* (2014) Effective filtering strategies to improve data quality from population-based whole exome sequencing studies. *BMC bioinformatics*, **15**, 125.

Chen R, Irwin DM, Zhang Y-P (2012) Differences in Selection Drive Olfactory Receptor Genes in Different Directions in Dogs and Wolf. *Molecular Biology and Evolution*, **29**, 3475–3484.

Craven B a, Paterson EG, Settles GS (2010) The fluid dynamics of canine olfaction: unique nasal airflow patterns as an explanation of macrosmia. *Journal of the Royal Society, Interface / the Royal Society*, **7**, 933–943.

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics*, **5**, 113.

Eisthen HL (1997) Evolution of Vertebrate Olfactory Systems. *Brain, Behavior and Evolution*, **50**, 222–233.

Fabbri E, Miquel C, Lucchini V *et al.* (2007) From the Apennines to the Alps: colonization

651 genetics of the naturally expanding Italian wolf (Canis lupus) population. *Molecular*
652 *ecology*, **16**, 1661–71.
653 Fan Z, Silva P, Gronau I *et al.* (2016) Worldwide patterns of genomic variation and admixture
654 in gray wolves. *Genome Research*, **26**, 163–173.
655 Firestein S (2001) How the olfactory system makes sense of scents. *Nature*, **413**, 211–218.
656 Freedman AH, Gronau I, Schweizer RM *et al.* (2014) Genome Sequencing Highlights the
657 Dynamic Early History of Dogs (L Andersson, Ed,). *PLoS Genetics*, **10**, e1004016.
658 Freitag J, Krieger J, Strotmann J, Breer H (1995) Two classes of olfactory receptors in Xenopus
659 laevis. *Neuron*, **15**, 1383–1392.
660 Freitag J, Ludwig G, Andreini I, Rössler P, Breer H (1998) Olfactory receptors in aquatic and
661 terrestrial vertebrates. *Journal of Comparative Physiology - A Sensory, Neural, and*
662 *Behavioral Physiology*, **183**, 635–650.
663 Geffen E, Anderson MJ, Wayne RK (2004) Climate and habitat barriers to dispersal in the
664 highly mobile grey wolf. *Molecular Ecology*, **13**, 2481–2490.
665 Gilad Y, Man O, Pääbo S, Lancet D (2003) Human specific loss of olfactory receptor genes.
666 *Proceedings of the National Academy of Sciences of the United States of America*, **100**,
667 3324–7.
668 Gilad Y, Wiebe V, Przeworski M, Lancet D, Pääbo S (2004) Loss of Olfactory Receptor Genes
669 Coincides with the Acquisition of Full Trichromatic Vision in Primates (David Hillis, Ed,).
670 *PLoS Biology*, **2**, e5.
671 Glusman G, Bahar A, Sharon D *et al.* (2000) The olfactory receptor gene superfamily: data
672 mining, classification, and nomenclature. *Mammalian Genome*, **11**, 1016–1023.
673 Godinho R, Llaneza L, Blanco JC *et al.* (2011) Genetic evidence for multiple events of
674 hybridization between wolves and domestic dogs in the Iberian Peninsula. *Molecular*
675 *ecology*, **20**, 5154–66.
676 Han E, Sinsheimer JS, Novembre J (2014) Characterizing Bias in Population Genetic Inferences
677 from Low-Coverage Sequencing Data. *Molecular Biology and Evolution*, **31**, 723–735.
678 Hayden S, Bekaert M, Crider T a. *et al.* (2010) Ecological adaptation determines functional
679 mammalian olfactory subgenomes. *Genome Research*, **20**, 1–9.
680 Hayden S, Bekaert M, Goodbla A *et al.* (2014) A cluster of olfactory receptor genes linked to
681 frugivory in bats. *Molecular biology and evolution*, **31**, 917–27.
682 Hughes GM, Gang L, Murphy WJ, Higgins DG, Teeling EC (2013) Using Illumina next generation
683 sequencing technologies to sequence multigene families in de novo species. *Molecular*
684 *Ecology Resources*, **13**, 510–521.

685 Issel-tamer L, Rine J, Issel-Tarver L, Rine J (1997) The Evolution of Mammalian Olfactory
686 Receptor Genes. *Genetics*, **145**, 185–195.
687 Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices
688 from protein sequences. *Computer applications in the biosciences : CABIOS*, **8**, 275–282.
689 Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7:
690 Improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772–
691 780.
692 Li H (2012) Exploring single-sample SNP and INDEL calling with whole-genome de novo
693 assembly. *Bioinformatics* , **28** , 1838–1844.
694 Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler
695 transform. *Bioinformatics*, **25**, 1754–1760.
696 Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler
697 transform. *Bioinformatics*, **26**, 589–595.
698 Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and
699 SAMtools. *Bioinformatics*, **25**, 2078–2079.
700 Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA
701 polymorphism data. *Bioinformatics (Oxford, England)*, **25**, 1451–2.
702 Lindblad-Toh K, Wade CM, Mikkelsen TS *et al.* (2005) Genome sequence, comparative analysis
703 and haplotype structure of the domestic dog. *Nature*, **438**, 803–819.
704 Linhart C, Shamir R (2002) The degenerate primer design problem. *Bioinformatics* , **18** , S172–
705 S181.
706 Loman NJ, Misra R V, Dallman TJ *et al.* (2012) Performance comparison of benchtop high-
707 throughput sequencing platforms. *Nature biotechnology*, **30**, 434–9.
708 Long R a, Donovan TM, Mackay P, Zielinski WJ, Buzas JS (2007) Effectiveness of scat detection
709 dogs for detecting forest carnivores. *Journal Of Wildlife Management*, **71**, 2007–2017.
710 Mainland JD, Willer JR, Matsunami H, Katsanis N (2013) Olfactory Receptors: Methods and
711 Protocols. In: (ed Crasto JC), pp. 133–147. Humana Press, Totowa, NJ.
712 McKenna A, Hanna M, Banks E *et al.* (2010) The genome analysis toolkit: A MapReduce
713 framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**,
714 1297–1303.
715 Musiani M, Leonard J a, Cluff HD *et al.* (2007) Differentiation of tundra/taiga and boreal
716 coniferous forest wolves: genetics, coat colour and association with migratory caribou.
717 *Molecular ecology*, **16**, 4149–70.
718 Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and

719 nonsynonymous nucleotide substitutions. *Molecular biology and evolution*, **3**, 418–26.
720 Niimura Y (2010) On the Origin and Evolution of Vertebrate Olfactory Receptor Genes:
721 Comparative Genome Analysis Among 23 Chordate Species. *Genome Biology and*
722 *Evolution*, **1**, 34–44.
723 Niimura Y, Matsui A, Touhara K (2014) Extreme expansion of the olfactory receptor gene
724 repertoire in African elephants and evolutionary dynamics of orthologous gene groups in
725 13 placental mammals. *Genome research*, **24**, 1485–96.
726 Ning Z, Cox AJ, Mullikin JC (2001) SSAHA: A fast search method for large DNA databases.
727 *Genome Research*, **11**, 1725–1729.
728 Olender T, Fuchs T, Linhart C *et al.* (2004) The canine olfactory subgenome. *Genomics*, **83**,
729 361–372.
730 Pilot M, Jedrzejewski W, Branicki W *Et Al.* (2006) Ecological factors influence population
731 genetic structure of European grey wolves. *Molecular Ecology*, **15**, 4533–4553.
732 Quail M a, Smith M, Coupland P *et al.* (2012) A tale of three next generation sequencing
733 platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers.
734 *BMC genomics*, **13**, 341.
735 Quignon P, Giraud M, Rimbault M *et al.* (2005) The dog and rat olfactory receptor repertoires.
736 *Genome biology*, **6**, R83.
737 Quignon P, Rimbault M, Robin S, Galibert F (2012) Genetics of canine olfaction and receptor
738 diversity. *Mammalian Genome*, **23**, 132–143.
739 Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic
740 features. *Bioinformatics*, **26**, 841–842.
741 Randi E (2011) Genetics and conservation of wolves Canis lupus in Europe. *Mammal Review*,
742 **41**, 99–111.
743 Randi E, Hulva P, Fabbri E *et al.* (2014) Multilocus Detection of Wolf x Dog Hybridization in
744 Italy, and Guidelines for Marker Selection (G Larson, Ed,). *PLoS ONE*, **9**, e86409.
745 Ranwez V, Harispe S, Delsuc F, Douzery EJP (2011) MACSE: Multiple Alignment of Coding
746 SEquences Accounting for Frameshifts and Stop Codons (WJ Murphy, Ed,). *PLoS ONE*, **6**,
747 e22594.
748 Saito H, Chi Q, Zhuang H, Matsunami H, Mainland JD (2009) Odor Coding by a Mammalian
749 Receptor Repertoire. *Science Signaling*, **2**, ra9–ra9.
750 Vaysse A, Ratnakumar A, Derrien T *et al.* (2011) Identification of Genomic Regions Associated
751 with Phenotypic Variation between Dog Breeds using Selection Mapping (JM Akey, Ed,).
752 *PLoS Genetics*, **7**, e1002316.

753 Vonholdt BM, Pollinger JP, Lohmueller KE *et al.* (2010) Genome-wide SNP and haplotype
754 analyses reveal a rich history underlying dog domestication. *Nature*, **464**, 898–902.
755 Wang G-D, Zhai W, Yang H-C *et al.* (2016) Out of southern East Asia: the natural history of
756 domestic dogs across the world. *Cell Research*, **26**, 21–33.
757 Young JM, Friedman C, Williams EM *et al.* (2002) Different evolutionary processes shaped the
758 mouse and human olfactory receptor gene families. *Human molecular genetics*, **11**, 535–
759 546.
760 Zhang W, Fan Z, Han E *et al.* (2014) Hypoxia Adaptations in the Grey Wolf (Canis lupus
761 chanco) from Qinghai-Tibet Plateau (JM Akey, Ed,). *PLoS Genetics*, **10**, e1004466.
762 Zhang Z, Li J, Zhao X-Q *et al.* (2006) KaKs_Calculator: calculating Ka and Ks through model
763 selection and model averaging. *Genomics, proteomics & bioinformatics*, **4**, 259–63.
764 Zhang H, Wei Q, Zhang H, Chen L (2011) Comparison of the fraction of olfactory receptor
765 pseudogenes in wolf (Canis lupus) with domestic dog (Canis familiaris). *Journal of*
766 *Forestry Research*, **22**, 275–280.
767

Tables and figures

Table 1: Number of genes amplified and sequenced in chip I and chip II, overlapping between the two, and totals across both runs

Counted in	Class I	Class II	n.a.	Total amplified
Chip I	105	253	17	375
Chip II	2	422	19	443
Overlap	2	212	7	221
Total	105	463	29	597

Table 2. SNPs identified in OR genes in dogs and wolves. Results compared with two-sided Mann Whitney *U* tests. Significant differences are highlighted in bold. **(a)** Mean number of SNPs per locus detected in small (1 to 60 OR genes) and large (170 to 222 ORs) clusters, and test results for differences between cluster sizes. No significant differences were found comparing dogs vs. wolves in small clusters ($U = 10867$, $P = 0.75$) nor in large clusters ($U = 4425$, $P = 0.09$). **(b)** Mean number of genes with a small or large number of SNPs (0-5 and 6-17, respectively) and test results for differences between cluster containing more or less polymorphic genes. No differences were found comparing dogs vs. wolves in the 0-5 SNPs group ($U = 26272$, $P = 0.72$) nor in the group of genes with 6-17 SNPs ($U = 152$, $P = 0.63$).

a)

Group	Clusters	Mean	Variance	Mann Whitney *U*
Dog	small	1.98	0.15	$U = 6146$, $\mathbf{P = 0.01}$
	large	2.86	0.27	
Wolf	small	2.04	0.18	$U = 7083$, $P = 0.42$
	large	2.32	0.23	

b)

Group	SNPs	Mean	Variance	Mann Whitney *U*
Dog	0-5	96.02	5.47	$U = 1553$, $\mathbf{P = 0.01}$
	6-17	158.76	17.32	
Wolf	0-5	98.94	5.43	$U = 1557$, $P = 0.26$
	6-17	135.75	23.46	

Table 3. Mutations per group. Results compared with two-sided Mann Whitney *U* tests. Significant differences are in bold. **(a)** Total number of synonymous, non-synonymous and frameshift mutations, and mean number per gene in dog and wolf samples. **(b)** Mean number of mutations per gene at class I and II genes in dog and wolf samples.

a)

Mutation type	Group	Number of mutations	Mean per gene	Variance	Mann Whitney *U*
Synonymous	Dog	263	1.05	0.08	*U* = 28420, *P* = 0.06
	Wolf	226	0.90	0.08	
Non-synonymous	Dog	223	0.90	0.08	*U* = 30977, P = 0.86
	Wolf	224	0.90	0.07	
Frameshift	Dog	33	0.13	0.02	*U* = 31250, *P* = 1.00
	Wolf	34	0.13	0.02	

b)

		Dog			Wolf			Dog vs. Wolf
Mutation type	Class	Mean per gene	Variance	Mann Whitney *U*	Mean per gene	Variance	Mann Whitney *U*	Mann Whitney *U*
Synonymous	I	1.40	0.19	*U* = 3917, ***P* = 0.02**	1.04	0.16	*U* = 4409, *P* = 0.24	*U* = 1044, *P* = 0.14
	II	0.97	0.08		0.88	0.09		*U* = 17840, *P* = 0.19
Non-synonymous	I	1.04	0.18	*U* = 4588, *P* = 0.45	0.86	0.15	*U* = 4896, *P* = 0.99	*U* = 1193, *P* = 0.67
	II	0.87	0.08		0.91	0.09		*U* = 18780, *P* = 0.68
Frameshift	I	0.04	0.03	*U* =4321, ***P* = 0.03**	0.06	0.03	*U* = 4444, *P* = 0.10	*U* = 1225, *P* = 1.00
	II	0.16	0.03		0.15	0.03		*U* = 19098, *P* = 1.00

797 **Table 4.** Mean number of variants found per pseudogene per OR class in dogs and wolves, and
798 comparison results via two-sided Mann Whitney *U* tests (used for not normal
799 distributed values) or two-sided *t*-test (used for normal distributed values). Significant
800 differences are highlighted in bold.

801

Class	Group	Mean per p-gene	Variance	Mann Whitney *U*	Class	Dog vs. Wolf
I	Dog	6.00	2.37	$U = 35$, $\mathbf{P = 0.04}$	I	$t_{10} = 0.27$, $P = 0.80$
II	Dog	2.32	0.53			
I	Wolf	5.17	2.01	$U = 50$, $P = 0.20$	II	$U = 287$, $P = 0.63$
II	Wolf	2.20	0.57			

802
803

Table 5. Haplotype variability and neutrality tests in functional OR genes in dog and wolf samples. Results compared with two-sided Mann Whitney *U* tests. Significant differences are highlighted in bold. **(a)** All functional OR genes **(b)** Functional OR genes separated into class I and II.

a)

Metric	Group	Mean	Variance	Mann Whitney *U*
Number of segregating sites	Dog	1.9	0.16	$U = 4168$, $P = 0.19$
	Wolf	1.7	0.17	
Haplotype number	Dog	2.6	0.11	$U = 3829$, $\mathbf{P = 0.02}$
	Wolf	2.2	0.09	
Haplotype gene diversity	Dog	0.4	0.02	$U = 4555$, $P = 0.79$
	Wolf	0.3	0.02	
Nucleotide diversity	Dog	5.9×10^{-4}	5.0×10^{5}	$U = 4502$, $P = 0.69$
	Wolf	6.3×10^{4}	7.8×10^{5}	
Tajima's *D*	Dog	3	0.93	$U = 2842$, $\mathbf{P = 0.04}$
	Wolf	5.4	1.07	
Fu-Li's *D*	Dog	0.03	0.63	$U = 3012$, $P = 0.13$
	Wolf	0.3	0.74	

b)

		Dog			Wolf			Dog vs. Wolf
Metric	Class	Mean	Variance	Mann Whitney *U*	Mean	Variance	Mann Whitney *U*	Mann Whitney *U*
Number of segregating sites	I	2.4	0.3	*U* = 627, *P* = 0.07	1.8	0.3	*U* = 781, *P* = 0.6	*U* = 215, *P* = 0.263
	II	1.8	0.2		1.7	0.2		*U* = 2429, *P* = 0.409
Haplotype number	I	3	0.2	*U* = 578, *P* = 0.02	2.4	0.2	*U* = 768, *P* = 0.51	*U* = 185, *P* = 0.073
	II	2.4	0.1		2.2	0.1		*U* = 2276, *P* = 0.127
Haplotype gene diversity	I	0.5	0.04	*U* = 565, *P* = 0.02	0.4	0.1	*U* = 739, *P* = 0.39	*U* = 232, *P* = 0.481
	II	0.3	0.02		0.3	0.03		*U* = 2586, *P* = 0.869
Nucleotide diversity	I	8.0×10^{4}	1.4×10^{4}	*U* = 670, *P* = 0.17	7.0×10^{4}	1.0×10^{4}	*U* = 777, *P* = 0.59	*U* = 242, *P* = 0.628
	II	5.0×10^{4}	4.90E-05		6.0×10^{4}	9×10^{5}		*U* = 2593, *P* = 0.892
Tajima's *D*	I	3.7	1.9	*U* = 644, *P* = 0.56	9.7	1.8	*U* = 421, *P* = 0.13	*U* = 47, *P* = 0.169
	II	2.8	1.1		4.2	1.3		*U* = 1648, *P* = 0.134
Fu-Li's *D*	I	0.9	1	*U* = 631, *P* = 0.45	1.4	1.4	*U* = 507, *P* = 0.61	*U* = 184, *P* = 0.709
	II	-0.4	0.8		-0.1	0.9		*U* = 1586, *P* = 0.067

812 **Table 6.** Mean Ka/Ks values detected in functional OR genes, compared between dogs and
813 wolves using two-tailed Mann Whitney *U* tests for all genes, and for class I and class II
814 genes separately within each group. No significant differences were also detected
815 comparing dogs vs. wolves for class I (Mann-Whitney $U = 177$, *2-tailed* $P = 1$) nor
816 class II ($U = 754$, *2-tailed* $P = 0.8$).

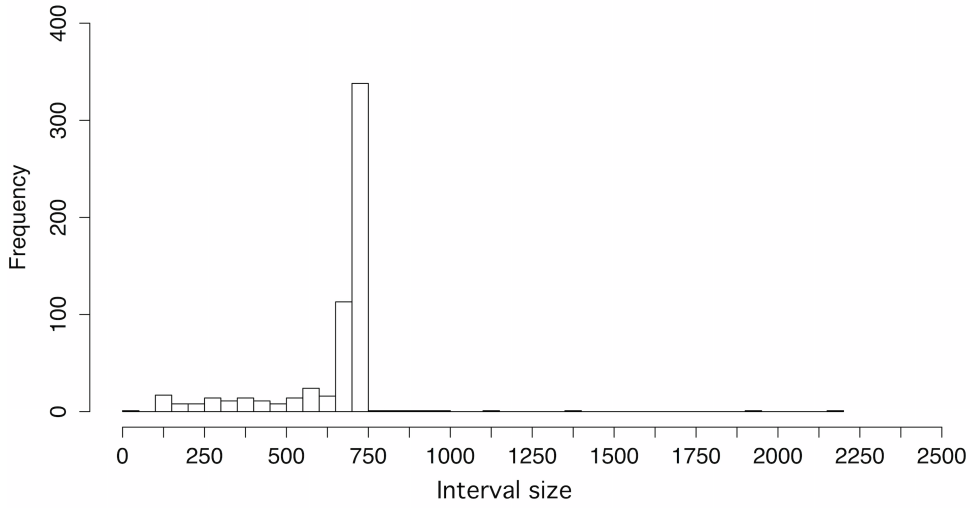
817

	Group	Mean Ka/Ks	Variance	Mann Whitney *U* test
All genes	Dog	0.09	0.02	$U = 1651$, $P = 0.8$
	Wolf	0.12	0.03	
Dog	Class I	0.07	0.02	$U = 387$, $P = 0.4$
	Class II	0.11	0.02	
Wolf	Class I	0.08	0.03	$U = 280$, $P = 0.5$
	Class II	0.14	0.04	

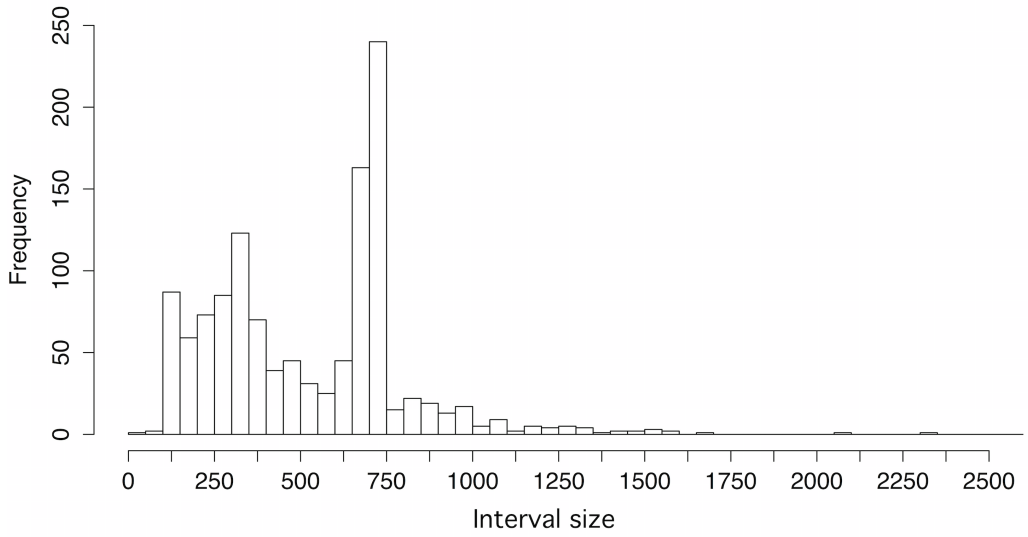
818

819 **Figure 1.** Frequency histogram of detected OR interval sizes sequenced with coverage of at least
820 15× in **(a)** Chip I and **(b)** Chip II.

821 **a)**

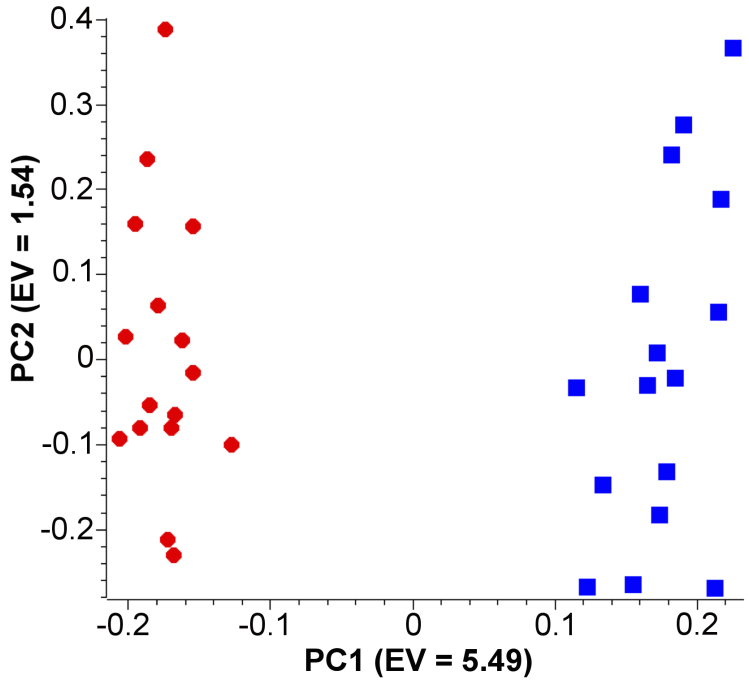


822
823 **b)**



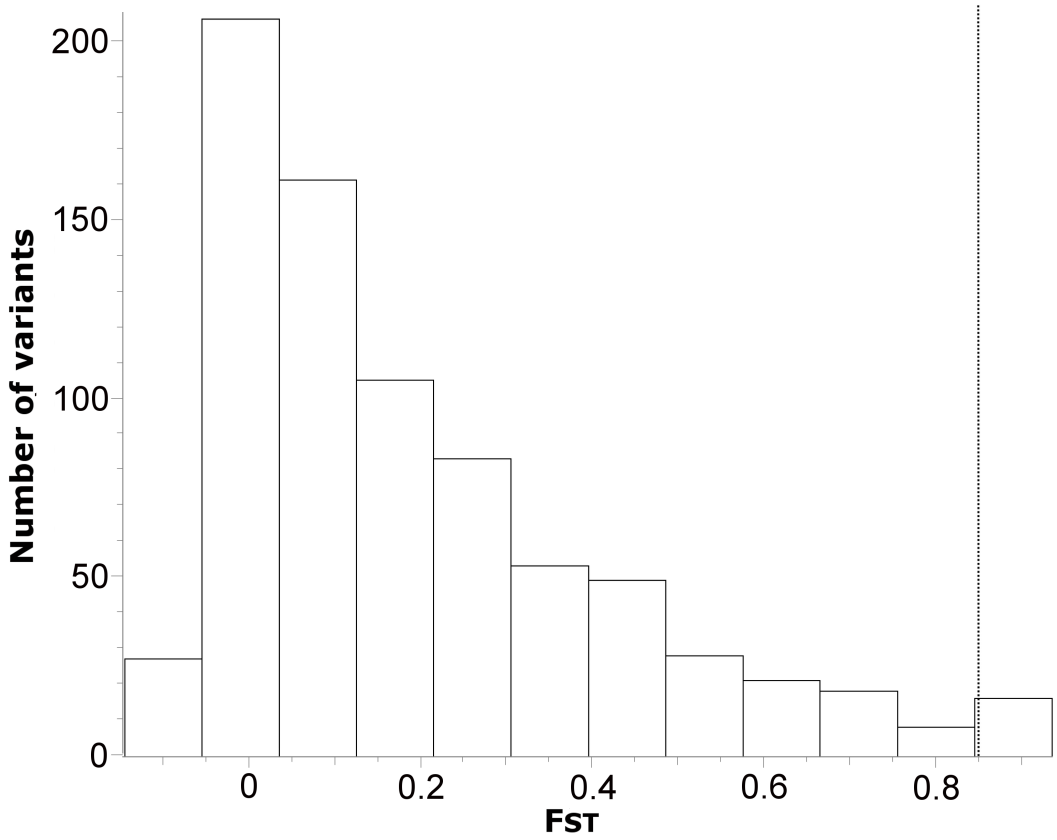
824
825

826 **Figure 2.** Results from principal component analysis on 775 variant sites within OR genes found
827 in dogs (red dots) and wolves (blue squares).



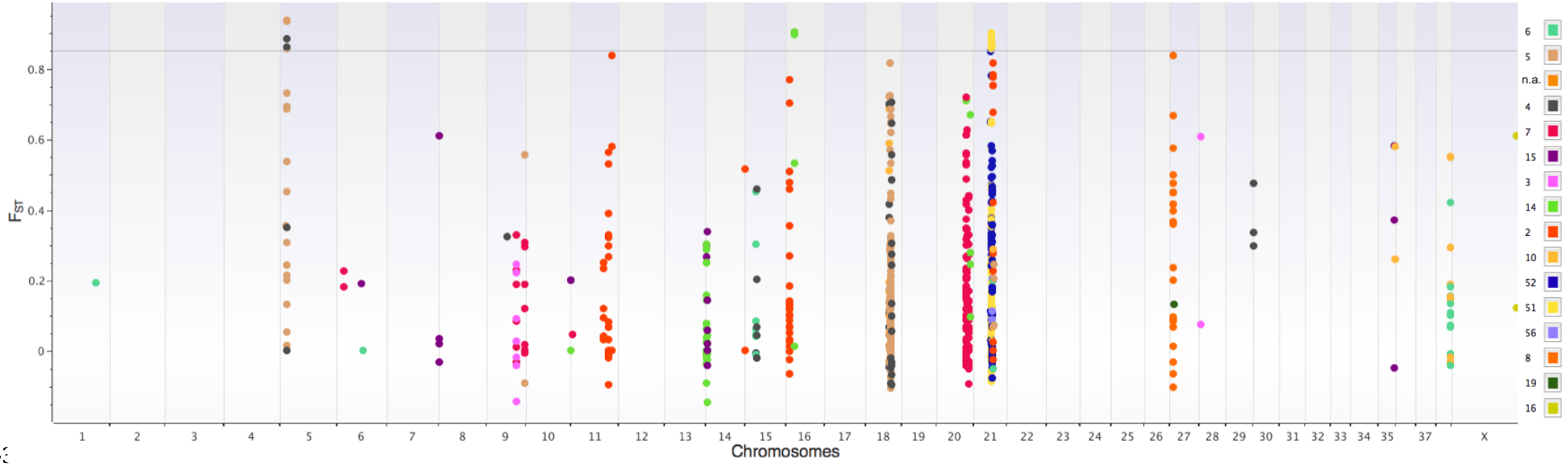
828
829
830
831

Figure 3. Distribution of the fixation index (F_{ST}) values between dogs and wolves among 775 variant sites identified in 250 polymorphic OR genes. The vertical line corresponds to the cut off used to define highly divergent variants (F_{ST} = 0.85).



838 **Figure 4.** Manhattan plot of F_{ST} values. Each dot is a variant site, vertical bands indicate the chromosomes and the colour of each variant site
839 describes the corresponding gene family. The horizontal line marks the F_{ST} cut off used (0.85) to define the most highly divergent
840 variants in wolves and dogs.

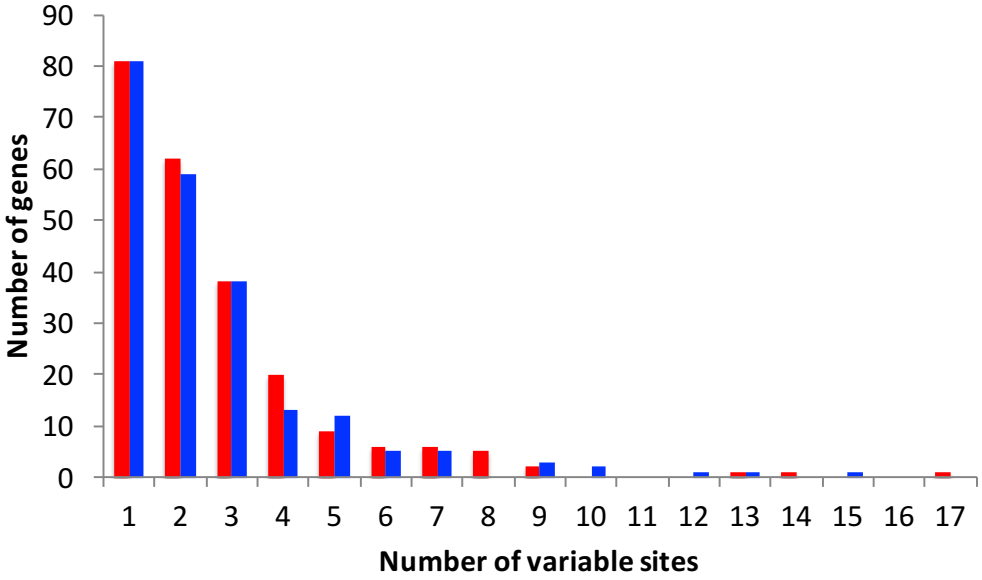
841
842



843

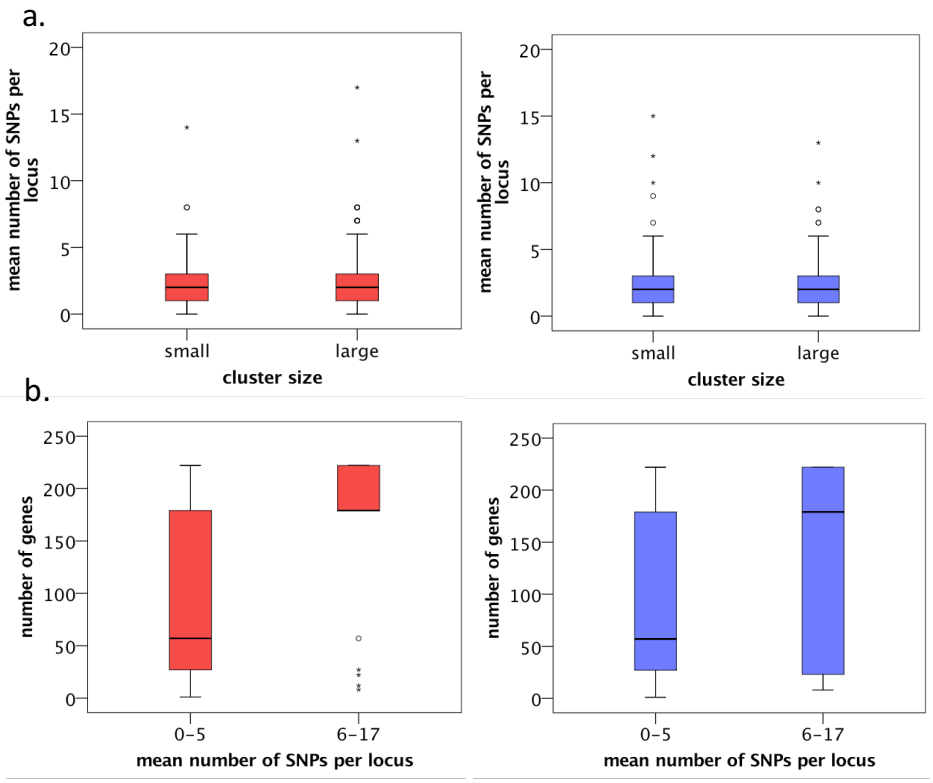
844 **Figure 5.** Distribution of the number of variable sites per locus in dogs (red bars) and wolves
845 (blue bars).

846
847



848
849
850

Figure 6. Identified SNPs per locus in dogs (red) and wolves (blue grey). **(a)** Cluster size (number of genes per cluster) in polymorphic genes hosting a small (0-5) vs. a large (6-17) number of SNPs **(b)** Mean number of SNPs per locus in small vs. large clusters.



Supplementary tables

Supplementary Table 1. (Electronic file) Reference ORs dataset containing 1012 validated genes with their genomic start and stop positions mapped on CanFam3.1, OR class, and family identification.

Supplementary Table 2. Raw Ion Torrent sequencing results for chip I and chip II.

	Loaded wells	Total reads	Mean read length (bp)	Polyclonal beads	Total sequences (Mbp)
Chip I	88%	3,418,815	300 ± 9	33%	1,000
Chip II	69%	2,980,317	308 ± 8	24%	915

865 **Supplemental Table 3.** Criteria for hard filtering of variant sites (SNPs and InDels).

	Parameters	Description	Threshold value for SNPs	Threshold value for INDELs	Motivation
HARD FILTERS from Best Practice GATK pipeline	(QD) Quality by depth	Variant confidence divided by the unfiltered depth of non-reference samples	>2.0	>2.0	The minimum quality score is normalized for the amount of coverage available.
	(FS) Fisher Strand	Phred-scaled p-value using Fisher's Exact Test to detect strand bias (the variation being seen only on the forward or only on the reverse strand)	<60	<200	More bias is indicative of false positive calls.
	(MQ) RMS Mapping Quality	Root Mean Square of the mapping quality of the reads across all samples	>40.0	>40.0	For major accuracy based on raw variants distribution.
	(MQRankSum) Mapping Quality Rank Sum Test	The u-based z-approximation from the Mann Whitney Rank Sum Test for mapping qualities (reads with reference bases vs. those with the alternate allele)	>-12.5	-	≅ 0: little to no difference; < 0: reads supporting the alternate allele have lower mapping quality scores than those supporting the reference allele, > 0: reads supporting the alternate allele have higher mapping quality scores than those supporting the reference allele.
	(ReadPosRankSum) ReadPosRankSumTest	The u-based z-approximation from the Mann Whitney Rank Sum Test for the distance from the end of the read for reads with the alternate allele	>-8.0	>-20	If the alternate allele is only seen near the ends of reads, this is indicative of error.
ADDITIONAL FILTERS	ALLELE type	Number of alternative alleles to be counted	biallelic	biallelic	Triallelic variants can be more likely due to sequencing errors.
	(AC) Allele count	Number of times each alternative allele is represented	>2	>3	5% MAF over all samples.
	(AN) Allele number	Total number of alleles in called genotypes	>32	>58	Major accuracy based on more than 50% and 90%, respectively, of the experimental population typed genotypes.
	(QD) Quality by depth	The variant confidence value divided by the unfiltered depth of non-reference samples	>5	>5	For major accuracy based on raw variants distribution.
	(MQ) RMS Mapping Quality	The Root Mean Square of the mapping quality of the reads across all samples	-	>80.0	For major accuracy based on raw variants distribution.
	Distance from nearest InDel	Nearby InDels can cause mapping quality drops and alignments errors	>5 bp	>1000 bp	We don't expect more than one InDel per locus.
	(GQ) Genotype Quality	Confidence in genotype assignment to a particular sample	>20	>20	Phred quality score corresponding to 99% accuracy.

866
867

868 **Supplemental Table 4.** (Electronic File) All polymorphic sites (corresponding to 733 SNPs, 38
869 deletions and 4 insertions, distributed across 250 genes) found in dogs and wolves with
870 their respective gene name and genomic positions. Variants identified in each run are
871 also specified for a total of 588 variants identified in chip I and 549 in chip II, with 362
872 sites found in both runs. Compared to the reference genome, all InDels caused a one-
873 base loss or gain, except one 2-bp loss (CFA20:47749184; locus CfOR03H07) and one
874 13-bp gain (CFA20:51576879; locus CfOR28H08/ST_XM_003432860.1). Genetic
875 distances are also reported with values of F_{ST} and Uniqueness. Independently from F_{ST}
876 values, 30 SNPs were found in wolves at a threshold of uniqueness > 90%, 11
877 belonging to class I and 19 to class II genes, with one of them being fixed for a single
878 allele and 14 of them with F_{ST} above 0.85. Similarly, 32 highly unique SNPs were
879 identified in dogs, 12 class I and 20 class II with 21 of them being monomorphic and 14
880 having F_{ST} above 0.85. Twenty-six of these SNPs were shared between dogs and
881 wolves, 12 of them with $F_{ST} > 0.85$.

882
883

884 **Supplemental Table 5.** Evaluation of possible bias toward a particular family or cluster in the
885 identification of variants. A two-sided Fisher exact test between the numbers of
886 amplified loci vs. the number of polymorphic loci identified per family and per cluster
887 was used. A bias in the identification of variants per gene was observed among families,
888 in particular with an excess in family 51 (class I) and family 15 (class II) in dogs, an
889 excess in family 7 (class II) in wolves, and a lack in family 56 (class I) in both dogs and
890 wolves. A significant skew was also found among clusters with a deficit in clusters
891 CFA20@46-47 (dogs and wolves) and CFA21@30-31 (wolves), and excess in
892 CFA30@1 (dogs and wolves).

893

	Group	Fisher exact test value	*P*
Per family	Dog	221	< 0.001
	Wolf	224	0.002
Per cluster	Dog	338	< 0.001
	Wolf	354	< 0.001

894
895

Supplemental Table 6. Comparisons of mean number of polymorphic OR genes and variable sites per locus (SNPs and InDels) in the sequenced dog and wolf samples, following their family, cluster and class distribution. Comparisons are via two-tailed Mann Whitney *U* tests, or two-tailed *t*-tests. Significant differences are highlighted in bold. Differences between dogs and wolves within classes were not significant, neither at class I (U = 1050, P = 0.162) nor at class II (U = 1829, P = 0.532).

	Comparison	Group	Mean	Variance	Test Result
Mean number of polymorphic genes	per family	Dog	15.07	3.46	t_{28} = 0.15, P = 0.882
		Wolf	14.33	3.44	
	per cluster	Dog	7.73	2.30	U = 443, P = 0.913
		Wolf	7.37	2.17	
Mean number of variable sites	per group	Dog	2.50	0.14	U= 29274, P = 0.209
		Wolf	2.31	0.14	
	per family	Dog	2.35	0.24	t_{28} = 0.26, P = 0.800
		Wolf	2.25	0.32	
	per cluster	Dog	2.14	0.21	U = 448, P = 0.979
		Wolf	2.16	0.20	
	class I	Dog	3.20	0.40	U = 3744, ***P* = 0.008**
	class II		2.31	0.15	
	class I	Wolf	2.58	0.34	U = 4423, P = 0.277
	class II		2.25	0.16	

Supplemental Table 7. Psuedogene variation. Comparisons are via two-tailed Mann Whitney *U* tests, or two-tailed *t*-tests. **(a)** Mean number of variants per pseudogene between groups, per-family and per-cluster. **(b)** Comparisons of pseudogene variability between dog and wolf haplotypes

a)

	Comparison	Group	Mean	Variance	Test Result
Mean number of variants per pseudogene	All	Dog	3.00	0.60	$U = 511$, $P = 0.40$
		Wolf	2.71	0.57	
	Per family	Dog	2.78	0.59	$t_{20} = 0.545$, P = 0.59
		Wolf	2.35	0.53	
	Per cluster	Dog	3.04	0.91	$U = 97$, $P = 0.55$
		Wolf	2.67	0.86	

b)

	Group	Mean	Variance	Test Result
Segregating sites	Dog	2.9	1.3	$U = 43$, $P = 0.6$
	Wolf	2.7	1.4	
Haplotype number	Dog	2.6	0.3	$U = 41$, $P = 0.5$
	Wolf	2.5	0.5	
Haplotype gene diversity	Dog	0.4	0.1	$t_{18} = 0.038$, $P = 1$
	Wolf	0.4	0.1	
Nucleotide diversity	Dog	7.8×10^{4}	2.5×10^{4}	$U = 49$, $P = 0.9$
	Wolf	1.2×10^{3}	7.1×10^{4}	

912

913 **Supplementary Table 8**. Number of genes per family respectively to OR reference database, together with identified number of variant genes and
914 pseudogenes, mean number of variant sites and SNPs as seen in dog and wolf samples in this study.

Family	Class	Reference genes	Number amplified genes	Number poly-morphic genes	Dog number poly-morphic genes	Wolf number poly-morphic genes	Dog mean number variant sites per locus	Wolf mean number variant sites per locus	Dog pseudo-genes	Wolf pseudo-genes	Dog mean variant sites per pseudo-gene	Wolf mean variant sites per pseudo-gene
5	II	179	120	40	39	34	3	3	3	2	2	2
4	II	140	60	20	18	16	2	2	1	1	2	2
6	II	119	34	11	10	9	2	2	1	1	3	2
7	II	87	80	49	44	47	3	3	8	9	4	4
2	II	57	50	23	22	20	2	3	1	1	1	1
14	II	47	37	21	19	20	2	2	3	4	3	3
10	II	43	25	10	10	10	3	2				
15	II	23	13	9	8	7	2	2	3	2	2	2
8	II	18	13	6	6	5	2	5	1		1	
20	II	10										
3	II	9	5	2	2	2	4	3				
12	II	5										
16	II	5	3	2	1	1	1	1	1	1	1	1
19	II	4	3	1	1		1					
21	II	3	2									
9	II	3	1									
17	II	2	2									
18	II	1	1									
52	I	87	52	22	19	21	4	3	3	3	8	6

51	I	61	39	26	25	21	3	3	3	3	4	4
56	I	20	11	2	2	2	2	1				
55	I	2										
57	I	1										

915

916 **Supplementary Table 9.** Number of genes per cluster respectively to OR reference database, together with identified number of variant genes and
917 pseudogenes, mean number of variant sites and SNPs as seen in dog and wolf samples in this study..

Cluster	Class	Reference genes	Number families	Number amplified genes	Number amplified families	Number poly-morphic genes	Dog number poly-morphic genes	Wolf number poly-morphic genes	Dog mean number variant sites per locus	Wolf mean number variant sites per locus	Dog mean number SNPs per locus	Wolf mean number SNPs per locus	Dog pseudo-genes	Wolf pseudo-genes	Dog mean variant sites per pseudo-gene	Wolf mean variant sites per pseudo-gene
CFA01@101	II	6	4	4	4	1	1	1	1	1	1	1				
CFA02@57	II	4	2													
CFA03@31	II	13	1	**3**	1											
CFA05@1	II	4	2	1	1											
CFA05@9-10	II	49	3	31	3	10	9	9	2	2	2	2	1	1	2	2
CFA06@9-12	II	3	3	1	1	1		1		2		2				
CFA06@37-38	II	2	2	2	2	1	1	1	1	1	1	1				
CFA06@40-41	II	4	1	1	1	1	1	1	1	1						
CFA08@1	II	14	4	5	2	3	3	1	1	1	1		2	1	2	1
CFA08@3	II	1	1	1	1											
CFA09@32	II	3	1	3	1	1	1	1	1	1	1	1				
CFA09@47	II	14	3	9	2	3	3	3	4	3	3	2	1	1	2	1
CFA09@59-60	II	8	2	7	2	2	2	2	5	5	5	5				
CFA10@1	II	2	1	1	1											
CFA10@16	II	16	3	**3**	2											
CFA11@1-2	II	5	3	5	3	3	2	3	1	1		1				
CFA11@52	II	8	2	**8**	2	4	3	4	2	2	1	2				
CFA11@60	II	12	1	10	1	5	5	5	3	3	3	3				
CFA11@66	II	1	1	1	1	1	1	1	3	1	2					

Cluster	Class	Reference genes	Number families	Number amplified genes	Number amplified families	Number poly-morphic genes	Dog number poly-morphic genes	Wolf number poly-morphic genes	Dog mean number variant sites per locus	Wolf mean number variant sites per locus	Dog mean number SNPs per locus	Wolf mean number SNPs per locus	Dog pseudo-genes	Wolf pseudo-genes	Dog mean variant sites per pseudo-gene	Wolf mean variant sites per pseudo-gene
CFA14@1-3	II	51	5	36	5	19	18	18	2	2	2	1	3	3	3	2
CFA15@1	II	4	1	4	1	2	2	1	1	1	1					
CFA15@17-18	II	26	2	10	2	6	5	5	2	2	2	2				
CFA16@5-7	II	27	4	18	2	8	8	6	3	3	3	3	1	1	1	1
CFA16@13-14	II	3	2	2	2	2	1	2	1	3		2		1	0	3
CFA17@57	II	2	1	2	1											
CFA18@25	II	1	1													
CFA18@37-41	II	222	6	137	5	50	48	42	3	3	3	3	3	2	2	2
CFA20@46-47	II	57	3	43	2	36	32	34	2	2	2	2	6	7	2	2
CFA20@51	II	22	1	19	1	10	10	10	4	4	4	4	1	1	14	15
CFA20@53	II	1	1	1	1	1	1	1	4	2	4	2				
CFA21@23	II	4	1	2	1											
CFA21@26-29	I	179	5	108	3	51	47	45	3	3	3	3	7	7	5	5
CFA21@30-31	II	36	5	28	5	7	7	5	2	3	2	3				
CFA25@50	II	17	2	14	1											
CFA27@1	II	20	3	**5**	3											
CFA27@5-6	II	23	2	17	2	7	7	5	2	5	2	5	1		1	0
CFA27@15	II	1	1													
CFA28@2	II	5	2	2	2	1	1	1	2	2	2	2				
CFA28@40-41	II	4	2													
CFA29@41	II	4	2													
CFA30@1	II	28	1	9	1	1	1	1	2	2	2	2	1	1	2	2
CFA31@1	II	1	1	1	1											
CFA33@5	II	14	1	**2**	1								2	2	4	3

Cluster	Class	Reference genes	Number families	Number amplified genes	Number amplified families	Number poly-morphic genes	Dog number poly-morphic genes	Wolf number poly-morphic genes	Dog mean number variant sites per locus	Wolf mean number variant sites per locus	Dog mean number SNPs per locus	Wolf mean number SNPs per locus	Dog pseudo-genes	Wolf pseudo-genes	Dog mean variant sites per pseudo-gene	Wolf mean variant sites per pseudo-gene
CFA35@25-26	II	16	3	6	2	3	3	3	3	2	2	1				
CFA38@22-23	II	30	2	17	2	8	8	8	2	2	2	2	1	1	3	2
ChrX@49	II	1	1													
ChrX@102	II	9	4	2	1	2	1	1	1	1	1	1	1	1	1	1
ChrX@122	II	4	3	**1**	1											

Supplementary Table 10. (Electronic File) Results of neutrality and selection tests for each polymorphic gene with a complete haplotype reconstruction across all 250 polymorphic genes identified.

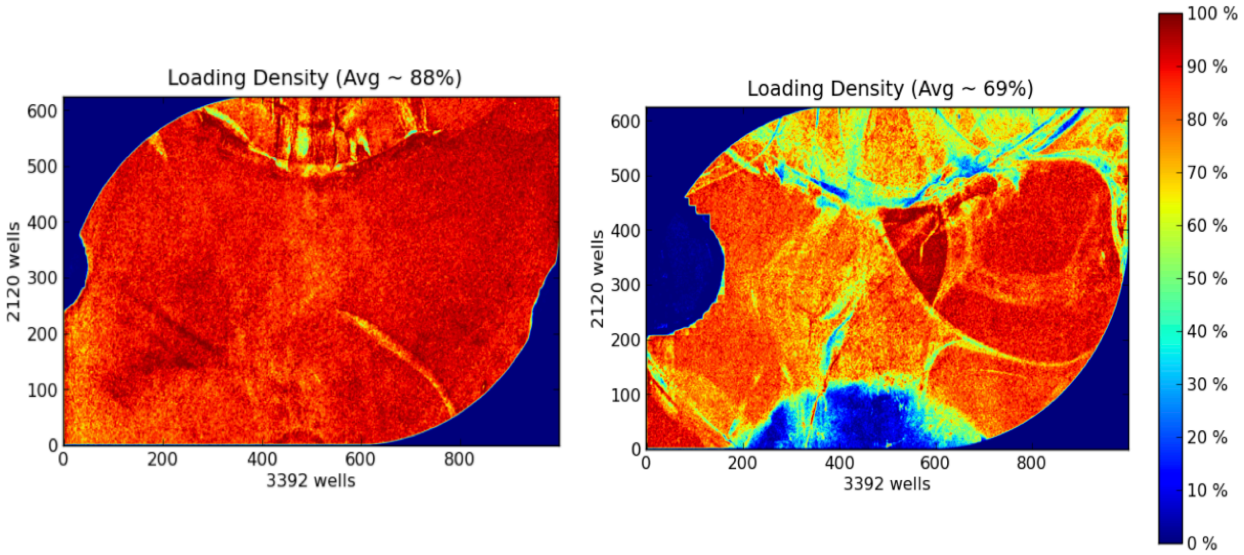
Supplementary Table 11. List of genes showing significant traces of selection at Tajima's *D* and Ka/Ks ratio. Values with significant meaning are indicated by one star for P < 0.005 and by two stars for P < 0.001. Nucleotide diversity (π) and Δ π values are also reported for those loci relying in high-ranking positions (specified in brackets) for Δ π and Δ Tajima's *D* absolute values.

Gene name	Functional	Class	Family / Subfamily	Cluster name	Dog π	Wolf π	Δ π absolute values	Dog Tajima D	Wolf Tajima D	Δ Tajima's D absolute values	Dog Fu-Li D	Wolf Fu-Li D	Dog Ka/Ks	Wolf Ka/Ks	Number sites FST > 0.85
CfOR4598	yes	II	5AP	CFA05@9-10	0.0008	0.0004	0.00039	0.98	-15.92	16.90	0.80	-1.00	0.00	0.16 *	3
CfOR0393/ST_XM_844824.1	yes	II	15B	CFA14@1-3	0.0011	0.0011	0.00002	20.86 *	20.21	0.65	0.80	0.80	NA	NA	-
CfOR0085/ST_XM_539687.3	yes	II	4P	CFA15@17-18	0.0010	0.0017	0.00073	-0.68	0.80	1.49	11.38	-0.64	0.08	0.08 *	-
CfOR0457/ST_XM_540562.1	yes	II	10G	CFA18@37-41	0.0012	0.0012	0.00002	-0.99	0.45	1.43	-0.15	10.51	0.40	0.09 *	-
CfOR0423/ST_XM_003432364.1	yes	II	5N	CFA18@37-41	0.0005	0.0013	0.00074	16.34	0.57	15.78	0.59	10.51	0.00	0.08 *	-
CfOR5812/ST_XM_540679.4	yes	II	5I	CFA18@37-41	0.0021	0.0017	0.00039	0.42	0.82	0.41	-0.15	11.38	0.16 *	0.29 *	-
ST_XM_540680.4/CfOR2248	yes	II	5I	CFA18@37-41	0.0010	0.0000	0.00095	-13.68	NA.	NA.	-11.50	n.a.	0.37 *	NA	-
CfOR0219	yes	II	4H	CFA18@37-41	0.0011	0.0000	0.00111	21.67 *	NA	NA	0.80	NA	0.33	NA	-
CfOR01C09/ST_XM_542000.3	yes	II	7A	CFA20@46-47	0.0003	0.0009	0.00064	-18.89 *	-0.39	18.50	-30.49	10.51	0.32	0.32	-
CfOR08A02/ST_XM_003432842.1	yes	II	7A	CFA20@46-47	0.0008	0.0008	0.00002	-0.61	11.23	11.85	10.51	0.80	0.09 *	0.31	-
CfOR10F07	pseudo	II	7B	CFA20@51	0.0029	0.0075	0..004658 (2)	-0.74	29.8 **	30.53 (2)	15.24 *	15.47 **	NA	0.08*	-
CfOR0100/ST_XM_003433014.1	yes	I	51L	CFA21@26-29	0.0027	0.0017	0.00096	20.83 *	17.18	3.65	12.09	10.51	0.27	0.34	-
CfOR2238/ST_XM_542382.3	yes	I	51J	CFA21@26-29	0.0017	0.0011	0.00054	0.75	10.35	9.60	11.38	-0.28	0.18 *	0.00	1
CfOR0104	yes	I	52Q	CFA21@26-29	0.0024	0.0027	0.00029	15.66	20.83 *	5.17	0.42	12.09	0.16 *	0.12 *	-
CfOR16D01/ST_XM_003433509.1	yes	II	8A	CFA27@5-6	0.0002	0.0060	0.00586 (1)	-0.45	33.84 ***	34.29 (1)	0.59	14.4 *	NA	0.08 **	-

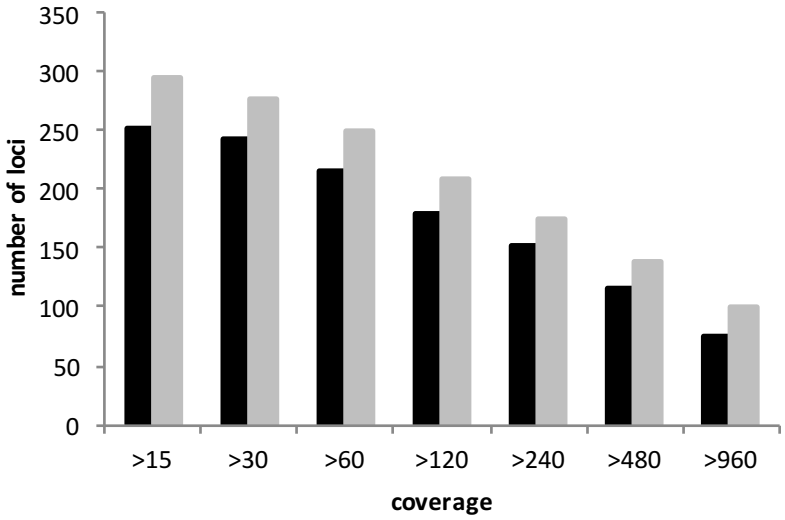
Gene name	Functional	Class	Family / Subfamily	Cluster name	Dog π	Wolf π	Δ π absolute values	Dog Tajima D	Wolf Tajima D	Δ Tajima's D absolute values	Dog Fu-Li D	Wolf Fu-Li D	Dog Ka/Ks	Wolf Ka/Ks	Number sites FST > 0.85
CfOR08D07/ST_XM_542007.3 CfOR08D07/ST_XM_542007.3	yes	II	7A	CFA20@46-47	0.0001	0.0005	0.00045	-11.42	15.97	27.39 (3)	-17.03	0.59	0.00	0.00	-
CfOR0501/ST_XM_543696.2	yes	II	8A	CFA27@5-6	0.0005	0.0001	0.00048	15.97	-11.42	27.39 (4)	0.59	-17.03	NA	NA	-
CfOR08D04/ST_XM_542002.2	yes	II	7A	CFA20@46-47	0.0002	0.0017	0.00156 (3)	-12.67	0.99	13.66	-0.76	11.38	0-15	1.08	-
CfOR1502/ST_NM_001083632	yes	I	51B	CFA21@26-29	0.0000	0.0014	0.00136 (4)	NA	16.42	NA	NA	0.94	NA	NA	-
CfOR0020/ST_XM_848281.2	yes	II	5I	CFA18@37-41	0.0015	0.0002	0.00124 (5)	20.05	-0.14	20.19	0.94	0.59	0.09	0.00	-

Supplementary figures

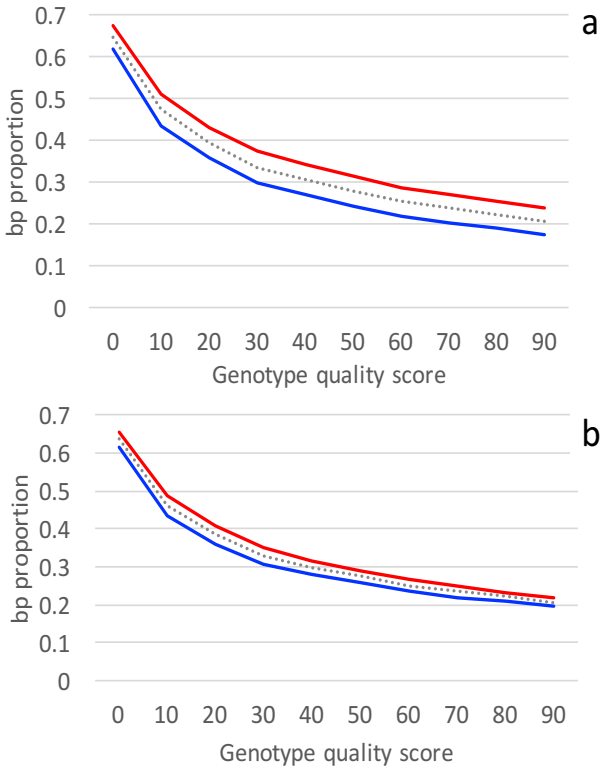
Supplementary Figure 1. Beads density pseudo-color image of the Ion Chip™ Plate showing percentage loading across the physical surface in chip I (left) and chip II (right).



Supplementary Figure 2. Number of genes sequenced at increasing coverage values (minimum value: 15×) at chip I (black) and chip II (grey).



Supplemental Figure 3. Genotype quality distribution in **(a)** Chip I and **(b)** Chip II, expressed as the fraction of base pairs (bp) sequenced at increasing values of genotype quality in dogs (red), wolves (blue) and in total (dotted line). Mann Whitney *U* test: Chip I, $U = 30$, 2-tailed $P = 0.14$; Chip II, $U = 39$, 2-tailed $P = 0.44$.



Final comments

To reach a comprehensive understanding of olfaction a multidisciplinary approach is necessary, including anatomical, physiological, genetic, and behavioural comparative analysis.

During the evolutionary history of vertebrates, olfaction raised in importance for survival, causing a progressive increase in size and complexity of brain. In primitive vertebrates, cerebral hemispheres developed from small olfactory lobes (Stoddart 1980). Olfactory bulb size has been positively associated to different allometric scales, with correlation of equal magnitude in different mammal families, but also to ecological adaptations (Gittleman 1991).The olfactory bulb and its associated neurons have undergone different changes, from the differentiation of cells carrying impulses, to the secondary connections and stratification reached in higher vertebrates. These changes were also accompanied by the evolution of other anatomical structures such as the second palate, new dental and cranial structures, the formation of diaphragm for active respiration and secretory glands (Aboitiz & Montiel 2015). The major advantages in olfaction, compared to the visual and acoustic senses, is that a scent mark can continue to broadcast its message to others for longer times. Smaller olfactory bulbs have thus been observed in aquatic species for the scarcity of chemosensory communication in aquatic environments (Gittleman 1991).

Early mammals mostly adapted to nocturnal life, which increased the necessity to exploit all senses dependent on orientation, which is mostly based on olfactory, tactile and proprioceptive cues. In the past it was commonly believed that all vertebrates retained a well-developed sense of smell except birds, which conquering a narrow ecological niche may have lost their olfactory functionality in favour of visual and auditory ones (Hagelin & Jones 2007). Recent findings show that the scale of avian repertoire may have been underestimated (Steiger *et al*. 2008, 2010). On the contrary bats, which partially exploit bird ecological niches, have developed fine-tuned olfactory abilities (Hayden *et al*. 2010, 2014; Young *et al*. 2010). Finally, macrosmatic mammals reached the sharpest sense of smell to better survive and specialize to different ecological niches (Hayden *et al*. 2010). Wild canids are supposed to be comprised in this ecological tag, because of their close evolutionary relation to dogs, which are currently more studied. The same comparison is done for all those wild mammals that have been domesticated. In particular between domestic pig and their wild relatives a brain reduction has been observed (Stoddart 1980), parallel to an olfactory structure reduction. However it seems that those structures can be re-gained when domesticated species return for long time to the wild, as observed in Sardinian free-living pigs (Maselli *et al*. 2013).

Direct observation and different scent marking studies have been performed on canids, showing that

dogs and wolves do not differ in the time when sensory systems start to work. Nevertheless, they likely have a different behaviour development, such as for the socializing period in pups, which is considered the cause of important adaptations in early-domesticated dogs to start scavenging on human trash dumps instead of hunting live prey (Udell 2014). In particular, wolves start socializing with only olfaction fully functional, while dogs have their senses mostly developed during this phase. Therefore dogs have an easier attitude to be socialized to objects at lower distances, a favourable adaptation which probably allowed dogs to forage close to human settlements, with higher tolerance of human proximity than wolves.

Canids live (as most of mammals do) in an olfactory world, full of invisible or at least unconsciously perceivable information to us. We are just starting to understand the possible specific connection existing between chemo-signals, genetic variability and olfactory abilities. To evaluate this relation, complex multi disciplinary studies are needed, and we are just starting to unravel questions about different olfaction abilities by comparing dog and wolf olfactory genomic repertoire.

Our analyses focused on the two main OR classes of genes and on the extremely high level of polymorphism which characterizes them in canids. As seen for other species, differences in the number of ORs are results of evolutionary forces driving organism to ecological and behavioural adaptations, and could reflect differences in potential odorants recognition. With our experimental design, we could not retrieve an unbiased representation of the OR repertoire, nor estimate the number of pseudogenes. Nonetheless we amplified 59% and sequenced polymorphisms in 25% of full canine repertoire. We found traces of stabilizing selection acting on both dog and wolf OR genes, with significant differences in selective pressures detectable only in few genes, which deserves further investigations.

On average, the observed levels of polymorphisms were not significantly different between dogs and wolves. The possibility that both canids share the same molecular structures especially in rapidly evolving genes such as olfactory was previously suggested by extended genomic comparisons among canids (Ostrander & Wayne 2005), supported by the narrow evolutionary period that exists between the divergence of domestic dogs from wolves. Our results point also toward a possible olfactory reduction of variability in Italian wolves, caused by past bottleneck, which could be definitively verified when a number of complete wolf genomes will be sequenced.

REFERENCES

Aboitiz F, Montiel JF (2015) Olfaction, navigation, and the origin of isocortex. *Frontiers in Neuroscience*, **9**, 1–12.

Barrios AW, SÃ¡nchez-Quinteiro P, Salazar I (2014) Dog and mouse: toward a balanced view of the mammalian olfactory system. *Frontiers in Neuroanatomy*, **8**, 1–7.

Brechbühl J, Klaey M, Broillet M-C (2008) Grueneberg ganglion cells mediate alarm pheromone detection in mice. *Science (New York, N.Y.)*, **321**, 1092–1095.

Brennan P a, Kendrick KM (2006) Mammalian social odours: attraction and individual recognition. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **361**, 2061–2078.

Cafazzo S, Natoli E, Valsecchi P (2012) Scent-Marking Behaviour in a Pack of Free-Ranging Domestic Dogs. *Ethology*, **118**, 955–966.

D. Michael Stoddart (1980) *The Ecology of Vertebrate Olfaction*.

Dulac C, Axel R (1995) A novel family of genes encoding putative pheromone receptors in mammals. *Cell*, **83**, 195–206.

Dulac C, Wagner S (2006) Genetic Analysis of Brain Circuits Underlying Pheromone Signaling. *Annual Review of Genetics*, **40**, 449–467.

Firestein S (2001) How the olfactory system makes sense of scents. *Nature*, **413**, 211–218.

Fleischer J, Schwarzenbacher K, Besser S, Hass N, Breer H (2006) Olfactory receptors and signalling elements in the Grueneberg ganglion. *Journal of Neurochemistry*, **98**, 543–554.

Freitag, J., G. Ludwig, I. Andreini, P. Rössler, and H. Breer. (1998) Olfactory Receptors in Aquatic and Terrestrial Vertebrates. Journal of Comparative Physiology - A Sensory, Neural, and Behavioral Physiology 183 (5): 635–650.

Freitag, Joachim, Jürgen Krieger, Jörg Strotmann, and Heinz Breer. (1995) Two Classes of Olfactory Receptors in Xenopus Laevis. Neuron 15 (6): 1383–1392.

Gese, E. M. & Ruff, R. L. 1997: Scent marking by coyotes, Canis latrans: the influence of social and ecological factors. Anim. Behav. **54**, 1155—1166.

Gilad Y, Bustamante CD, Lancet D, Paabo S (2003) Natural selection on the olfactory receptor gene family in humans and chimpanzees. *American Journal of Human Genetics*, **73**, 489–501.

Gilad Y, Wiebe V, Przeworski M, Lancet D, Pääbo S (2004) Loss of Olfactory Receptor Genes Coincides with the Acquisition of Full Trichromatic Vision in Primates (David Hillis, Ed,). *PLoS Biology*, **2**, e5.

Gittleman JL (1991) Carnivore olfactory bulb size: allometry, phylogeny and ecology. *Journal of Zoology*, **225**, 253–272.

Glusman G, Bahar A, Sharon D *et al.* (2000) The olfactory receptor gene superfamily: data mining, classification, and nomenclature. *Mammalian Genome*, **11**, 1016–1023.

Glusman G, Yanai I, Rubin I, Lancet D (2001) The complete human olfactory subgenome. *Genome research*, **11**, 685–702.

Grüneberg H. (1973) A ganglion probably belonging to the N. terminalis system in the nasal mucosa of the mouse. Z. Anat. Entwickl.-Gesch. **140**, 39–52.

Hagelin J, Jones I (2007) Mechanism or overlooked mode of intraspecific. , 124, 741–761.

Hayden S, Bekaert M, Crider T a. et al. (2010) Ecological adaptation determines functional mammalian olfactory subgenomes. *Genome Research*, **20**, 1–9.

Hayden S, Bekaert M, Goodbla A *et al.* (2014) A cluster of olfactory receptor genes linked to frugivory in bats. *Molecular biology and evolution*, **31**, 917–27.

Hayden S, Teeling EC (2014) The molecular biology of vertebrate olfaction. *Anatomical record (Hoboken, N.J. : 2007)*, **297**, 2216–26.

Herrada G, Dulac C (1997) A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. *Cell*, **90**, 763–773.

Horowitz A (2014) *Domestic Dog Cognition and Behavior*.
Isogai Y, Si S, Pont-Lezica L *et al.* (2011) Molecular organization of vomeronasal chemoreception. *Nature*, **478**, 241–5.
Jordan NR, Golabek KA, Apps PJ, Gilfillan GD, Mcnutt JW (2013) Scent-mark identification and scent-marking behaviour in African wild dogs (Lycaon pictus). *Ethology*, **119**, 644–652.
Kaluza JF, Gussing F, Bohm S, Breer H, Strotmann J (2004) Olfactory Receptors in the Mouse Septal Organ. *Journal of Neuroscience Research*, **76**, 442–452.
Kelliher KR (2007) The combined role of the main olfactory and vomeronasal systems in social communication in mammals. *Hormones and Behavior*, **52**, 561–570.
Liberles SD, Buck LB (2006) A second class of chemosensory receptors in the olfactory epithelium. *Nature*, **442**, 645–50.
Liberles SD, Horowitz LF, Kuang D *et al.* (2009) Formyl peptide receptors are candidate chemosensory receptors in the vomeronasal organ. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 9842–9847.
Lord K (2013) A Comparison of the Sensory Development of Wolves (*Canis lupus lupus*) and Dogs (*Canis lupus familiaris*). *Ethology*, **119**, 110–120.
Malnic B, Hirono J, Sato T, Buck LB (1999) Combinatorial receptor codes for odors. *Cell*, **96**, 713–723.
Mamasuew K, Breer H, Fleischer J (2008) Grueneberg ganglion neurons respond to cool ambient temperatures. *European Journal of Neuroscience*, **28**, 1775–1785.
Maselli V, Polese G, Larson G *et al.* (2013) A Dysfunctional Sense of Smell: The Irreversibility of Olfactory Evolution in Free-Living Pigs. *Evolutionary Biology*.
Matsunami H, Buck LB (1997) A multigene family encoding a diverse array of putative pheromone receptors in mammals. *Cell*, **90**, 775–784.
Mech D (1999) Alpha status , dominance , and division of labor in wolf packs. , **1203**, 1196–1203.
Mech, L. D., and L. Boitani (Eds.). 2003. Wolves: Behaviour, Ecology and Conservation. The University of Chicago Press.
Monahan K, Lomvardas S (2015) Monoallelic Expression of Olfactory Receptors. *Annual Review of Cell and Developmental Biology*, **31**, annurev–cellbio–100814–125308.
Mori K, Manabe H, Narikiyo K (2014) Possible functional role of olfactory subsystems in monitoring inhalation and exhalation. *Frontiers in Neuroanatomy*, **8**.
Nei M (1969) Gene duplication and nucleotide substitution in evolution. *Nature*, **221**, 40–42.
Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 7799–7806.
Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annual review of genetics*, **39**, 121–152.
Ngai J, Dowling MM, Buck L, Axel R, Chess A (1993) The family of genes encoding odorant receptors in the channel catfish. Cell 72 (5):657–666
Niimura Y (2012) Olfactory Receptor Multigene Family in Vertebrates: From the Viewpoint of Evolutionary Genomics. *Current Genomics*, **13**, 103–114.
Niimura Y, Matsui A, Touhara K (2014) Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. *Genome research*, **24**, 1485–96.
Niimura Y, Nei M (2005) Comparative evolutionary analysis of olfactory receptor gene clusters between humans and mice. *Gene*, **346**, 13–21.
Niimura Y, Nei M (2007) Extensive Gains and Losses of Olfactory Receptor Genes in Mammalian Evolution (M Hahn, Ed,). *PLoS ONE*, **2**, e708.
Olender T, Fuchs T, Linhart C *et al.* (2004) The canine olfactory subgenome. *Genomics*, **83**, 361–372.
Ostrander E a, Wayne RK (2005) The canine genome. *Genome research*, **15**, 1706–16.

Peters RP, Mech LD (1975) Scent-marking in wolves. *American Scientist*, **63**, 628–637.
Quignon P, Giraud M, Rimbault M *et al.* (2005) The dog and rat olfactory receptor repertoires. *Genome biology*, **6**, R83.
Quignon P, Kirkness E, Cadieu E *et al.* (2003) Comparison of the canine and human olfactory receptor gene repertoires. *Genome biology*, **4**, R80.
Quignon P, Rimbault M, Robin S, Galibert F (2012) Genetics of canine olfaction and receptor diversity. *Mammalian Genome*, **23**, 132–143.
Rivière S, Challet L, Fluegge D, Spehr M, Rodriguez I (2009) Formyl peptide receptor-like proteins are a novel family of vomeronasal chemosensors. *Nature*, **459**, 574–577.
Robin S, Tacher S, Rimbault M *et al.* (2009) Genetic diversity of canine olfactory receptors. *BMC Genomics*, **10**, 21.
Rodolfo-Masera T. (1943) Su l'estizenza di un particulare organo olfacttivo nel setto nasale della cavia e di altri roditori. Arch Ital Anat Embryol **48**,157–212.
Rouquier S, Giorgi D (2007) Olfactory receptor gene repertoires in mammals. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, **616**, 95–102.
Rothman, R. J. & Mech, L. D. 1979: Scent-marking in lone wolves and newly formed pairs. Anim. Behav. **27**, 750—760.
Ryba NJP, Tirindelli R (1997) A new multigene family of putative pheromone receptors. *Neuron*, **19**, 371–379.
Saito H, Chi Q, Zhuang H, Matsunami H, Mainland JD (2009) Odor Coding by a Mammalian Receptor Repertoire. *Science Signaling*, **2**, ra9–ra9.
Salazar I, Quinteiro PS (2009) The risk of extrapolation in neuroanatomy: the case of the Mammalian vomeronasal system. *Frontiers in neuroanatomy*, **3**, 22.
Schmid A, Pyrski M, Biel M, Leinders-Zufall T, Zufall F (2010) Grueneberg ganglion neurons are finely tuned cold sensors. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **30**, 7563–7568.
Steiger SS, Fidler AE, Mueller JC, Kempenaers B (2010) Evidence for adaptive evolution of olfactory receptor genes in 9 bird species. *The Journal of heredity*, **101**, 325–33.
Steiger SS, Fidler AE, Valcu M, Kempenaers B (2008) Avian olfactory receptor gene repertoires: evidence for a well-developed sense of smell in birds? *Proceedings. Biological sciences / The Royal Society*, **275**, 2309–17.
Sturm T, Leinders-Zufall T, Maček B *et al.* (2013) Mouse urinary peptides provide a molecular basis for genotype discrimination by nasal sensory neurons. *Nature communications*, **4**, 1616.
Tian H, Ma M (2004) Molecular organization of the olfactory septal organ. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **24**, 8383–8390.
Young JM (2002) The sense of smell: genomics of vertebrate odorant receptors. *Human Molecular Genetics*, **11**, 1153–1160.
Young JM, Massa HF, Hsu L, Trask BJ (2010) Extreme variability among mammalian V1R gene families. *Genome research*, **20**, 10–8.
Yu S, Avery L, Baude E, Garbers DL (1997) Guanylyl cyclase expression in specific sensory neurons: a new family of chemosensory receptors. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 3384–3387.
Zhang X, Firestein S (2002) The olfactory receptor gene superfamily of the mouse. *Nature neuroscience*, **5**, 124–33.
Zhang X, Rodriguez I, Mombaerts P, Firestein S (2004) Odorant and vomeronasal receptor genes in two mouse genome assemblies. *Genomics*, **83**, 802–811.
Zhang X, Zhang X, Firestein S (2007) Comparative genomics of odorant and pheromone receptor genes in rodents. *Genomics*, **89**, 441–450.
Ziegler A, Santos PSC, Kellermann T, Uchanska-Ziegler B (2010) Self/nonself perception, reproduction and the extended MHC. *Self/nonself*, **1**, 176–191.

ACKNOWLEDGMENTS

In questi anni ho scoperto un profondo amore per i cani, nonostante vent'anni di gattofilia, quello che un cane sa dare è incomparabile, l'empatia e la straordinaria comprensione del linguaggio che questi animali possono talvolta avere è sconvolgente. E' cresciuta quindi la voglia di capire e di studiare questi splendidi animali. Quando mi è stata proposta la possibilità di lavorare sulla genetica dei lupi loro nobili ancestori, cinque anni fa, per me è stata una grande gioia, ma non pensavo di arrivare fin qui.
La passione, la pazienza mi hanno spinto in questi anni a completare questo dottorato nella speranza che tutto l'impegno profuso in questo lavoro non fosse solo fine a se stesso, ma potesse dare un apporto in qualche modo alla conoscenza e protezione del lupo in Italia. La mia grande fortuna è stata avere la possibilità di conoscere Ettore Randi all'Università di Bologna, e dal suo lavoro e dall'eccezionale gruppo di ricerca fondato da lui, tutto è cominciato.

Procedo quindi con i ringraziamenti.
Prima di tutti, devo e dovrò sempre un grazie profondo ai miei genitori, che hanno sempre creduto in me e non mi hanno mai mollato nemmeno nei momenti più avversi, mi hanno sempre spinto a continuare, a non lasciarmi sopraffare dalle emozioni e ad affrontare la vita qualunque ostacolo ci riservi. Siete meravigliosi, non so cosa farei senza di voi. Quello che potrò fare io per voi non sarà mai abbastanza, siete parte di me e sempre lo sarete. Il vostro coraggio e forza nell'affrontare i dolori del corpo e dell'anima mi hanno insegnato e continuano a insegnarmi tanto.

Grazie al mio mostro preferito nonchè davvero 'Angelo' del focolare, che mi ha sopportato e supportato con estrema pazienza e dolcezza, mi sento viziata, grazie. Sei il mio migliore amico e compagno di viaggio perfetto. Siamo imperfettamente perfetti insieme. Dopo questo traguardo penso che ci vorrebbe veramente un uragano per separarci. Comunque sia e comunque andrà sono stati anni bellissimi, difficili ma anche pieni di gioia e amore incondizionato. Ora basta altrimenti ti monti la testa.

Grazie a Ettore ancora una volta per avermi dato la possibilità di fare questa esperienza professionale importantissima e avermi dato la possibilità di poter "rischiare" e seguire un nuovo progetto nonostante tutte le difficoltà del caso.

E ovviamente grazie al mio branco Elena, Romolo e Marco! Un team di ricerca che insieme a Ettore

mi ha permesso di imparare tantissime cose, e che ha dato con il proprio lavoro tantissimo alla comunità scientifica del lupo Italiano. E' stato un onore lavorare con voi, e imparare da voi. Davvero un ottimo gruppo unito e forte. La cosa straordinaria è stata che in questi anni è successo di tutto e le nostre vite sono cambiate tantissimo, siamo davvero cambiati e cresciuti assieme e non rimpiango nulla.

Grazie ai mitici Pietro e Aritz, la vostra esperienza è davvero grande, siete dei ricercatori brillanti da cui c'è molto da imparare, per non parlare del fatto che siete dei personaggi fantastici … senza nulla togliere al mio branco che comunque non scherza eh!! Grandiosi!

Grazie a Vale e Edo per i taralli della coop :-) e le ore passate assieme davanti ai monitor con facce spaurite, ma anche piene di risate e passione. In particolare grazie a Vale per una rara capacità di comprensione che serberò come un prezioso tesoro, grazie amica mia, adesso ci dobbiamo rifare di tutto il tempo perso.

Grazie a Rossy e Andre, per il grande amore che mi volete, e che è ricambiato. Il vostro punto di vista è sempre spunto di riflessione. In particolare vorrei ringraziare Andrea per le lunghe chiacchierate, sei il fratello che non ho mai avuto.

Grazie a Vera per le serate di follia cerebrale, mi piace circondarmi di personaggi e tu sei una personaggiona fantastica. Bella gajarda! Ci sarò sempre, promesso.

Finally, I would thank Robert and Douglas for their patience and competence that gave me the strength needed to keep on working regardless the tough period I was passing through.

I would also thank Maja and David that shared their cosy super organised home with me, but overall thanks to give me true love and friendship. I was very lucky to meet you guys, and I am waiting for you here in Bologna.

E infine un grazie va Trudy, la mia piccola compagna di vita, anche se non sai leggere voglio lasciare nero su bianco su questo pezzo della mia vita, l'emozione che il tuo legame ha significato per me …

E infine grazie a tutti gli enti e lavoratori del settore protezione fauna selvatica che ogni giorno si

impegnano nel loro lavoro e hanno reso possibili questi risultati.

Viva il lupo, lunga vita al lupo!

END