Federation Cynologique Internationale (2008), Rafeiro of Alentejo

Unraveling the evolutionary history of livestock dog breeds

Diogo Filipe Coutinho Lima

Mestrado em Biodiversidade, Genética e Evolução
Departamento de Biologia
2018

Orientador

Raquel Godinho, Principal Researcher, CIBIO/InBio
Invited Assistant Professor, Universidade do Porto

Co-orientadores

Ignacio Doadrio, Researcher, CSIC
Museo Nacional de Ciencias Naturales, Madrid

Pedro Silva, Post-Doctoral Researcher, CIBIO/InBio

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____

“Thus man, however slow the process may have been, had, through the aid of the dog, at last firmly planted his feet on the ladder by which he was to climb to civilization”.

- R.D.S. Gwatkin

Agradecimentos

Penso que a escrita deste texto marca o final destes 5 anos de aventura. E que 5 anos foram, cheios de risos e frustrações, de trabalho e de vida boa, de persistência e de medo de falhar. Passei este ano de trabalho na tese sempre a desvalorizar este texto, a achar que não tinha nada nem ninguém a quem agradecer e mais uma vez vem a vida ensinar-me o quão errado estava. Agora no meio de lágrimas, baba e ranho é que me apercebo que não fiz este caminho sozinho, apercebo-me que não teria conseguido fazer este caminho sozinho.

Em primeiro lugar devo os maiores agradecimentos à pessoa que quis ser pai, mas que teve que ser mãe, à pessoa que me chateia todos os dias, que me põe em cima toda a pressão do mundo e mais alguma, mas acima de tudo a pessoa que nunca me deixou falhar, que sempre me apoiou quando os meus sonhos não seguiam o caminho mais fácil e que me educou para ser eu próprio. Sei que não sou a pessoa mais fácil, que posso ser negligente em relação à minha família e que não pareço grato em relação a tudo o que me deste. Sou o que sou, teimoso, orgulhoso, e uma tempestade de emoções, mas nunca ingrato. Sei bem o que custou criar um filho sozinho, sei todos os sacrifícios que fizeste por mim e sei bem o quanto te esforçaste para que a minha vida fosse melhor do que a tua. Nunca nenhuma ação minha vai ser agradecimento suficiente, por isso é que vou continuar a dar sempre o melhor de mim, vou trilhar o caminho que quero para mim por mais difícil que seja e, mais do que orgulho em ser eu próprio, tenho orgulho em ser teu filho.

Em segundo lugar quero agradecer ao resto da minha família, tanto aos que estão perto como aos que estão longe, tanto aqueles que acham que querer mais é o mesmo que ter o nariz empinado como aqueles que me ajudaram a provar que os outros estavam errados. Maria da Graça obrigado por mostrares que quem cria 8 filhos também cria 1 a mais, grande parte da pessoa que sou também se deve a ti. Maria da Glória, obrigado por me ensinares que mesmo depois de nos roubarem o que mais amamos e nos esmagarem o coração, podemos voltar a ser felizes e encarar a vida com um sorriso na cara. Tia, para ti é difícil arranjar palavras, obrigado por teres posto a tua vida em espera para criares um filho que não era teu, dizem que sou muito parecido contigo e não poderia ter mais orgulho nisso. Sempre disseram que precisava de uma mãe, nunca se aperceberam de que já tinha três.

Susana obrigado por toda a roupa lavada e por todas as refeições em cima da mesa ahahahah, falando a sério, obrigado por estares presente durante o meu crescimento e por teres participado na construção do adulto que sou hoje. Pedro e João, não podia ter irmãos mais diferentes e retardados, mas também não podia ter melhores.

Um agradecimento especial para a melhor amiga, para a minha outra metade, a pessoa que achou que matemática era uma boa ideia. Obrigado Mónica, obrigado por teres feito este caminho comigo. É estranho pensar que já passaram 5 anos de faculdade e ainda mais estranho pensar que

passaram outros 5 desde que nos conhecemos. Penso que demos uma chapada de luva branca a quem nunca acreditou em nós e a quem dizia que a faculdade separa amizades. Tenho muito orgulho no percurso que fizeste e espero que continuemos a percorrer esta estrada juntos.

Outro agradecimento especial para o Bucky para o meu Steve, o Ron para a minha April, a Diana para a minha Carolina. Muito obrigado por todas as boleias para o CIBIO, por todas as playlists de Britney e por todo conhecimento MCU. Em 2013 a Britney lançou a Work B*tch e 5 anos depois acabamos o curso, coincidência? Começamos isto juntos e vamos acabar isto juntos. Obrigado por tudo Inês “Chicken” Pinto.

Obrigado a todos os meus migos do Império do Canhão. Obrigado Afonso, Marta, Sofia, Lala, Poncito, Rui e Meia. Quem dizia que na faculdade se arranjam amigos para a vida claramente que andou na faculdade convosco. Obrigado por tudo o que me deram sem pedir nada em troca, não teria chegado aqui se não fossem vocês. Obrigado Meia por teres decidido que este mestrado era o melhor para mim numa noite de São João em a Titanium já estava a acabar e a alma da Lala já caminhava por outros mundos.

Um kiss grande para as minhas Macacas do Norte. Vocês são a prova de que as pessoas podem unir-se, ignorando as suas diferenças, para lutar por algo em que acreditam. Quando nos perguntarem por é que somos o melhor ano de sempre, respondemos apenas: somos Macacas do Norte, tornamos a praxe mais forte. Para bom entendedor meia palavra basta.

Próximo kiss para as minhas Hienas, em especial para as minhas filhotas. Vocês tomam péssimas opções de vida é verdade, mas ensinaram-me tanta coisa, mudaram tanto a pessoa que eu era, e lá no fundo até que gosto de vocês. Lutem sempre por aquilo em que acreditam. BTW shout out para a melhor madrinha do mundo, obrigado por tudo Ginny.

Um abraço para ti Ninicas, és uma parte muito grande daquilo que sou hoje. Obrigado por me teres levado ao colo quando já não tinha forças para andar.

Um obrigado gigante para ti Raquel. Ensinaste-me imensa coisa e nunca me deixaste na mão. Espero poder continuar a trabalhar contigo no futuro. Muchas gracias Nacho por tu entusiasmo por este trabajo. Obrigado Carolina por me teres praxado desde o dia 1 no laboratório e por me teres ensinado tudo o que precisei sempre com calma e sem reservas. Obrigado Pedro pela paciência e pela dedicação, prometo que defenderei sempre a linha de comandos em prol do interface gráfico. Um grande beijinho para toda a equipa da Ecogen, agradeço todo o apoio. Obrigado Susana e Patrícia do CTM, nunca me falharam e sempre foram pacientes com este rapaz que se apaixonou pelo trabalho de laboratório. Prometo que enquanto puder não vos deixo em paz.

This is for the mother that never came to be
but is always with me,

Financial Support

This work was partially supported by FCT (projects PTDC/BIA-EVF/2460/2014 and IF/00564/2012/CP0159/CT0003, both awarded to Raquel Godinho) and by Norte Portugal Regional Operational Programme (NORTE2020), under the PORTUGAL 2020 Partership Agreement, through the European Regional Development Fund (ERDF)(NORTE-01-0145-FEDER-000007 to Nuno Ferrand).

We thank the Portuguese and the Spanish Kennel Club and Joana Robalo, Paloma Garzón, Manuel Bahillo and Enrique Pérez Campos for collecting the samples. We thank the Clube Português de Canicultura for samples of all the Portuguese dog breeds and Luis Miguel Moreira for additional samples of Transmontano Mastiff.

Sumário

Os cães de guarda do gado seguem e ajudam os seres humanos desde a domesticação do gado e ovelhas, e a consequente necessidade de mover e proteger estes valiosos recursos de predadores. Apesar deste importante papel, houveram surpreendentemente poucas investigações sistemáticas sobre a história evolutiva dessas raças de cães. Na Península Ibérica, onde a prática agrícola mudou drasticamente a paisagem e onde a criação de gado assume uma parte substancial do sustento do povo ibérico, é invulgar que estes cães sejam ainda mal compreendidos.

Neste trabalho, procuramos em primeiro lugar estimar diversidade e relações entre raças de cães de guarda de gado espalhadas pela Eurásia e avaliar sua origem mais provável, e em segundo procuramos compreender a relação entre os cães de guarda e pastoreio e avaliar a sua diversidade genética. Para fazer isso, mais de 180 cães representando 24 raças de cães guardadores de gado foram genotipados usando o Illumina CanineHD BeadChip, que inclui mais de 170.000 SNPs uniformemente espaçados em todo o genoma. Além disso, 294 cães de guarda e de pastoreio foram genotipados para 43 marcadores microssatélites autossómicos e 6 do cromossoma Y.

No que respeita à análise de microssatélites, foram encontrados vinte e um haplótipos numa amostra de 150 machos analisados para o cromossoma Y, oito dos quais privados de uma determinada raça, tendo o Rafeiro do Alentejo e o Pastor Leonês apresentado o maior número. Em relação à diversidade genética, devido a intensos eventos de seleção artificial em raças modernas, esperava-se que os cães pastoreio demonstrassem menor diversidade e, embora os cães de guarda apresentassem maior diversidade, tanto para os loci Y como para os autossomas, no geral, a diferença não é considerável.

Estudos filogenéticos anteriores descobriram que os cães asiáticos são representantes dos clados basais, agrupam-se com forte suporte estatístico e podem ser separados das raças modernas, sendo possivelmente as mais antigas. A partir de uma análise filogenética de cães de guarda de gado, encontramos as mesmas evidências, sendo as raças da área de distribuição oriental deste tipo de cães as raças basais. De facto, um gradiente de perda de diversidade genética, embora tênue, pode ser encontrado das raças Orientais às Ocidentais, o que poderia indicar que os cães asiáticos são mais velhos. No entanto, as diferenças nos níveis de diversidade genética não foram significativas entre as raças orientais e ocidentais, e algumas raças ibéricas de cães de guarda, como o cão da Serra da Estrela, apresentaram elevados níveis de diversidade genética.

A maioria das raças não exibem sinais genéticos de endogamia, como encontrado para outras raças de cães em estudos anteriores; isto provavelmente deve-se ao facto dos cães de guarda, sendo mantidos com os seus rebanhos, não serem sujeitos a intensos níveis de seleção para características físicas desejadas. Além disso, a análise da estrutura genética demonstrou baixos níveis de

diferenciação e eventos de fluxo gênico foram encontrados entre as raças, apoiando a crença de um enriquecimento genético causado pelo cruzamento entre raças.

Como é de esperar, as raças ibéricas estão muito relacionadas entre si e, embora haja alguma variação entre os dois grupos baseados em funções, não foram encontradas diferenças significativas entre os cães de guarda e de pastoreio, sugerindo que as raças Portuguesas e Espanholas são derivadas do mesmo stock ancestral. Embora exista uma clara diferenciação fenotípica entre as raças de pastoreio e de guarda, essa diferenciação não foi observada no genoma.

Propomos que os cães de guarda surgiram em algum lugar no Continente Asiático a partir do stock de cães disponíveis e que as raças, como as conhecemos agora, provavelmente são fruto de diferenciação após os eventos de dispersão para seguir os rebanhos, mas também incorporam sinais genómicos de miscigenação com outras raças provavelmente resultantes de movimentos pecuários como Transumância.

PALAVRAS-CHAVE: genómica das populações, miscigenação, fluxo genético, diversidade genética, estrutura genética, SNPs, microssatélites, cromossoma Y, autossomas, cães de guarda, cães de pastoreio, Península Ibérica, Eurásia.

Abstract

Livestock dogs follow and help humans since the domestication of cattle and sheep, and the consequent need to move and protect these valuable resources from predators. Despite this important role, there have been surprisingly few systematic investigations into the evolutionary history of these dog breeds. In Iberia, where agricultural practice drastically changed the landscape and where livestock farming assumes a substantial part of the livelihood of the people, is outlandish that these dogs are still poorly understood.

In this work, we aimed to, first estimate the diversity and relationships among livestock guarding dog breeds spread over Eurasia and evaluate their most probably origin and second, to try to understand the relationship between livestock guarding and herding dogs and to assess their genetic diversity. To do so, over 180 dogs representing 24 livestock guarding dog breeds were genotyped using the Illumina CanineHD BeadChip, which includes more than 170,000 SNPs evenly spaced across the genome. Also, 294 Iberian livestock guarding, and herding dogs were genotyped for 43 autosomal and 6 Y-chromosome microsatellite markers.

Concerning the microsatellite analysis, twenty-one haplotypes were found in a sample of 150 males analysed for the Y chromosome, eight of which were private for a certain breed, with Rafeiro of Alentejo and Leonese Shepherd displaying the highest number. Regarding the genetic diversity, due to intense artificial selection events on modern breeds, herding dogs were thought to display lower diversity and although guarding livestock dogs present overall, more diversity, the difference is not considerable.

Previous phylogenetic studies found Asian dogs to be the ones representative of the basal clades, which grouped together with strong statistical support and could be separated from the modern breeds, possibly being the oldest ones. From a phylogeny analysis of livestock guarding dogs, we retrieved the same evidence, being breeds from the Eastern distribution area of this kind of dogs the basal ones. In fact, a gradient of genetic diversity loss, though tenuous, can be found from the Eastern to the Western breeds, which could indicate Asian dogs to be older. However, differences on the levels of genetic diversity were not significant between Eastern and Western breeds, and some Iberian livestock guarding dog breeds, like Estrela Mountain dog, displayed high levels of genetic diversity.

Most breeds do not exhibit genetic signals of inbreeding as found in other dog breeds by previous studies; this is probably due to livestock guarding dogs being kept with their herds and not being subject to intense levels of selection for desired physical traits. Also, analysis of genetic structure demonstrated low levels of differentiation and gene flow events were found between breeds, supporting the belief of a genetic pool enrichment caused by inter-breed mating, something not found in past studies.

As expected Iberian breeds are very related to each other, and although there is some variation between the two function-based groups, no major differences were found between either guarding and herding dogs, also suggesting that Portuguese and Spanish breeds are derived from the same ancestral stock. While there is clear phenotypical differentiation observed between herding and guarding breeds, this differentiation was not observed through in the genome.

We propose that livestock guarding dogs emerged somewhere in the Asian continent from available dog stock and that the present breeds are probably fruit of differentiation after dispersion events to follow herds but also incorporate on their genome signals of admixture with other breeds likely resulting from livestock movements like Transhumance.

KEYWORDS: population genomics, admixture, gene flow, genetic diversity, genetic structure, SNPs, microsatellites, Y chromosome, autosomes, guarding dogs, herding dogs, Iberian Peninsula, Eurasia.

Index

List of figures

List of Tables

List of Abbreviations

- AMOVA – Analysis of Molecular Variance
- FCA – Factorial Correspondence Analysis
- F_{ST} – Fixation index (Wright's F-statistics)
- nDNA – nuclear DNA
- Y-DNA – Y-linked DNA
- mtDNA – mitochondrial DNA
- SNP – Single Nucleotide Polymorphism
- STR – Short Tandem Repeat
- LGD – Livestock Guarding Dog
- LHD – Livestock Herding Dog
- PCR – Polymerase Chain Reaction

CHAPTER 1.

General Introduction

1.1. A tale of domestication

The defining separation between men and apes occurred when the former discovered fire and gained control of wood and stone. Although, another important discovery to our species, yet much disregarded, happened when humans learned how to control nature to their advantage, commencing with the dog domestication[1]. It is commonly accepted that dogs exhibit the most complete and the most useful conquest that man has made[1,2].

When mankind first emerged in the African continent, it spent his entire lifespan engaged in a fight in search of food for himself and for those who were dependent on him[3]. Having a nomadic type of life, men started moving north from their birthplace and finally arrived at the Middle East around 60 thousand years ago[4]. When they were lucky enough they would encounter some large animal too sick to offer resistance, or already dead[1]. This “feasts” helped as first contact between mankind and other carnivores that would hang around men, awaiting any scraps. Wolves and jackals were among those animals[1] and slowly, they got used to each other’s presence. Over time, both species realized the benefits that the other species could offer and wolves started changing to be less aggressive and more submissive, giving birth to the first domestic dogs. The archaeological record shows dog remains buried beside humans 14,700 years ago[4], but other, although disputed, remains occurring 35 thousand years ago[5] suggests a much older domestication time for dogs.

With the aid of dogs, the eternal hunt for food became less strenuous. With their assistance, men successfully attacked larger and more ferocious animals that supplied meat for longer periods. As it was no longer necessary for women and children to accompany hunting parties, mankind started gradually changing from nomads into building the first civilization cities, with the emergence of Jericho[6], in which is nowadays Palestinian Territories.

This led to the Neolithic revolution, which started around eleven thousand years ago[7], in Mesopotamia. The Neolithic Revolution, or First Agricultural Revolution, was a wide-scale transition, during the Neolithic period, from the nomadic lifestyle, of hunting and gathering, to the sedentary lifestyle, of agriculture and settlement[4]. The first settled communities allowed humans to observe and experiment with plants, to learn how they grew and develop, leading to the first domestication phenomenon of plants[8]. The first agriculture practices started in the region named Fertile Crescent, also known as the "cradle of civilization", with the cultivation of barley and wheat[9]. Modern-day countries within the Fertile Crescent include Iraq, Syria, Lebanon, Cyprus, Jordan, Israel, Palestine, Egypt, as well as the south eastern fringe of Turkey and the western fringes of Iran[10,11].

With humans settling in one place, and with the help of the domestic dogs, it became possible to domesticate other species, like pigs, sheep and cattle[12]. Besides being a direct source of food, certain animals could provide leather, wool, hides, and fertilizer. Animals that provided milk, such as cows and goats, offered a source of protein that was renewable and therefore quite valuable. With

the creation of the first domesticated herds and the need to protect them from wild predators, the first livestock guarding dogs emerged from the ancient domestic dog[13] (Fig. 1.1).

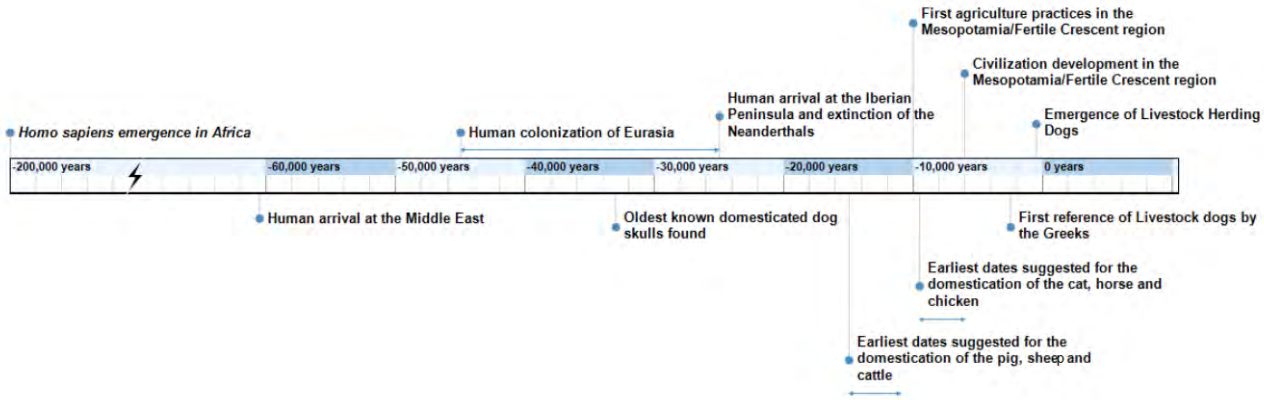


Fig. 1.1 – Timeline highlighting some of the most important events in the evolutionary history of mankind along with domestication of dog.

1.2. The origin of the domestic dog

The origin of the domestic dog is a highly-debated subject mainly due to the phenotypical and behaviour diversity found in this species[14]. However, it is consensual that wolf was the first species to be domesticated and the genomic differentiation between wolves and dogs is yet considerably low. As the dogs evolutionary history is intertwined with human history[15], knowing when, where and how the domestication process occurred is a big step forward in understanding the past of our own species[16]. The domestic dog is also ideal to study the genetic basis of morphology, susceptibility to disease and behaviour[17].

Domestication can be described as an evolutionary process where a population of animals becomes adapted to humans and to different environments through genetic changes associated with selectiveness controlled by Man[18]. Although there is a common belief that it was mankind starting the domestication progress there is other hypothesis suggesting a “Self-domestication” event started by wolves. This hypothesis proposes that some wolves moved into a symbiotic relation with humans and thereafter, these wolves found utility as barking sentinels, warning of human of other predators, and as hunters[19]. Gradually, the first primitive dogs emerged from this group[20,21].

Diverse methods are being used to determine the dog’s origin, including the search for archaeological evidence of domestication (e.g. fossils[22] and through the aid of genetic tools. Frequently these methods do not show similar results mainly due to gaps in the fossil record and the lack of transitional forms that make it difficult to establish a precise phylogenetic relationship[23]. The first theoris[23] hypothesized that the domestic dog was an independent species from the beginning and that modern dogs are a result of the admixture of an ancient *Canis familiaris,* similar to the modern form, with wolves. Later, hybridization with other members of the genus *Canis*, like coyotes was also

proposed. Recent genetic studies, however, support the hypothesis that dogs descend exclusively from the grey wolf[24], although there is no consensus of when or where the domestication started.

Dog domestication studies have shown different conclusion through the time. Equipped with fragments of mitochondrial DNA and molecular clocks, Vilà et al.[25] concluded that dogs were domesticated 135,000 years ago. Later, a separate work analysed a similar mitochondrial fragment for 654 dogs and, based on regional patterns of modern dog diversity, deduced that dogs were domesticated just once in East Asia[26]. Recent genomic studies concerning modern dogs and primitive dogs from Asia and Africa also supported the idea of and Asian origin for the domestic dog with later spreading across Eurasia[27]. Furthermore, a study of 48,000 SNPs in 912 dogs and 225 grey wolves concluded that dog breeds share a higher proportion of their genome with wolves from the Middle East, supporting a theory of a Middle Eastern origin for dogs[28]. Other work, using nuclear markers, also suggests a Middle East geographic origin for dogs[29]. Besides the theories for an Asian or Middle Eastern origin, there is also the theory for a European origin of the domestic dog[30].

The huge difference among wolf/dog domestication studies may be explained with the dependence on molecular evolutionary rates. There is still a lack of knowledge on the dog-specific mutation rate, and so the application of different estimation of the mutation rate in different studies has led to these discrepancies between results[31,32]. Besides, there are also other problems regarding the finding of the domestic dog geographical origin. Gene flow between wolves and dogs was proven to have occurred after domestication and there is the hypothesis that the wolf population from where dogs came from no longer exists[31,33]. Also, too many of these studies were built on the certainty that dogs breeds that displayed the highest genetic diversity should be the wolves, which is nowadays known to not be true[34]. Demographic history has varied enough in dogs across the world after domestication so that current levels of genetic diversity may not reflect the oldest ancestral population[35].

Another factor blurring the evolutionary history of dogs is the extreme selection events that were applied in order to create the outstanding phenotypic diversity of which modern dogs are associated to[28]. Other human events also had strong influences on their genetic diversity, pointing warfare. Across history, dogs became extinct in many regions or suffered significant bottlenecks[36,37] and most of the dog populations as we know them are the products of controlled breeding practices and have only been established in the past two centuries[28,38].

Given that, the formation of highly selected and inbred populations has created a lot of problems in the understanding of the dog origin. Even in large datasets with high numbers of individuals and breeds, methods that use genomic sequences or even thousands of SNPs, spread across the 2.4 billion DNA bases that make up the dog genome[24], are only skilled to propose a wide range of time and place for the dog domestication[39,40].

1.3. The emergence of the term “breed”

The popular and well-known term “breed” is an artificial aggregation of morphologic or behaviour characteristics, not restricted to a geographic area or a survival strategy, in which there is a strong human influence[41], occurring transmission of desired characters to the progeny. The breed is not a taxonomic category, and thus it does not have any meaning as subspecies[42].

The concept of breed in dogs is a fairly recent British invention, going back less than 200 years[43]. International dog breeds are registered in the *Federation Cynologique Internationale* (FCI), the international organization that accepted or fixed standards for each breed. Currently, and according to FCI, there are between 400-500 registered dog breeds[15]. The origin of these breeds has been debated by several authors, being common to claim that is closely linked to the dog function[44,45]. It is then possible to group breeds according to function they are used for (guardian dogs, herding dogs, hunting dogs, company dogs, etc).

Some breeds were first classified as “ancient breeds” by genetic analysis after observing their early-branching pattern in phylogenetic analysis[38]. Nevertheless, these breeds could also be together in this basal position on a phylgenetic tree because they lack recent admixture with other breeds. Currently, these lineages are referred as “basal”[39]. Basal breeds are distinctive from modern breeds because their genetic heritage mostly avoided admixture with modern breeds, while modern breeds genome has become blurred due to admixture.

Numerous dog breeds have been recognized by FCI, but many other are still not recognized by any kennel club and are facing a dark future because of a poor management and the lack of conservation measures. Recently, Dreger et al.[46] published a breed status classification based on genomic tools for pure breed dogs that includes threshold levels from four homozygosity measurements. These metrics demarcate the extremes of single-dog length of homozygosity, ten-dog shared length of homozygosity, coefficient of inbreeding, and rate of shared homozygosity decay represented by standardized pure dog breeds. Also, Talenti et al.[47] defined a fifth parameter (phylogenetic clustering on an expected clade) for which the dog population can be assigned as a breed or a breed variant (when the phylogenetic clustering is absent). Given these five genetic parameters, the population status is determined as a “breed” if the dogs follow the purebred range in at least three of the four genomic metrics and the individuals group together phylogenetically[47].

Kennel club breed standards were set to describe the perfect breed representative, which further encouraged line breeding to propagate the physical traits that are easier to spot early in the dog’s development[48]. The FCI breed classification requires that dogs present an “Ideal cynological description of a breed” and that dog populations meet certain criteria, like a high number of individuals. Since many breeds fail to meet these strict criteria, mainly due to the lack of knowledge, a breed status classification based on genomic tools could prove to be very useful.

1.4. The first livestock guarding and herding breeds

Although the time and place of origin for the first dogs is still unknown, the origin for livestock breeds is believed to be in the Neolithic Age[21,49], when men adopted a sedentary way of living, which led to higher selection pressures on the already existing dog together with the definite isolation from the wild stock, caused by the building of the first cities[44]. The first dogs were used for both hunting and guarding but over time and with an increase in the function selectiveness guarding breeds and hunting breeds diverged.

Some literature defends that the far distant ancestor of the guarding dogs was a mythical mastiff which lived on the high Chinese plateaus from pre-historic times[50]. Actually, in a recent work, Wang et al.[51] stated that the native dogs of South China are likely the most primitive form of dogs and may represent the product of the first stage of domestication, occurring more than 32 thousand years ago. Among these primitive dogs is included the Tibetan mastiff, a direct descendent from the first dogs and a predecessor of the modern-day mastiffs[50,51]. This breed is popularly known to be the ancestor of many other breeds, leading to the belief that this ancient breed may be the "mythical mastiff", although those are just conjectures and was isolated in the Himalayan area for centuries[52,53]. Besides these theories, livestock guarding dogs most likely originated in Mesopotamia and its peripheral regions, where the pastoral practices were well developed[13].

It is not known how dogs first arrived in Europe. Some authors defend that the first ancestors of guardian dogs arrived in the company of nomadic shepherds, like during the migrations between Mesopotamia and modern-day Hungary[50] and spread across Europe[54–56]. Others believe an old legend defending that European dogs evolved from dogs named molosses. These dogs were brought from India to Greece in 326 B.C. by Alexander the Great[50], with molosses being afterwards used by the Romans for circus games, for combat and to guard their properties[57]. It is also possible that dogs came from the Middle East to Europe with the Phoenician merchants[50,58]. It is possible that parts from all these theories are true, with European breeds being fruit of admixture and differentiation of different Asian breeds that arrived in Europe at different times.

It appears, from the old literature, that early shepherd dogs were used both for guarding and herding. Today, distinctions between the two types are made more clearly. Herding dogs are not 24-hour flock guardians, nor are guarding dogs good for herding livestock[13]. Guarding dogs are not used for driving cattle due to their large constitution which helps them fighting large predators but could lead to harming or scare the herds[13]. While guarding dogs were created due to the need to protect domestic stock from wild predators, herding breeds were created to help move the herds from place to place[59]. It is believed that the origin of herding dog breeds occurred recently, in the early modern period (between the late middle age [1500] and Age of Revolutions [1800]), in countries where large carnivores suffered great losses in their populations and guarding dogs became less needed.

Thomas[60] noted that the lack of wolves in England, at the beginning of "the early modern period", resulted in dogs that drove sheep, while in France or Italy, where wolves still survive, the herds follow a shepherd, and guarding dogs act as their protector. In France, herding dogs are known as a fairly recent development, supplanting LGDs (Livestock guarding dogs) when large carnivores (bear, wolf) disappeared from western Europe[61].

Livestock guarding breeds generally represent breeds with very low levels of artificial selection, as is the case of the Spanish Mastiff or most of the Persian and Kurdish dog breeds, because the only usefulness they had in the past was to protect the livestock. In contrast, livestock herding breeds, among other dogs, represent modern dogs. Modern breeds were created through backcrossing and inbreeding, to fix desired traits, resulting in a much-reduced level of heterozygosity within any breed compared to wild canids or ancient breed[24]. Until recently, the phenotype was not important to accomplish the job, but nowadays these breeds are getting more attention by the general public, who are obtaining these dogs to serve as family pets or show dogs[13].

1.5. Livestock guarding and herding dogs

Guarding dogs were used to protect livestock from wolves and other predators, often displaying aggressive behaviour, which is mostly instinctive, as the dog is bonded to the herd from an early age[62]. The puppies of these breeds display behaviour involving social games (contests, hunting, submission-domination), but not predation behaviour. Under certain conditions, LGDs were also used to move the flocks, although this does not happen commonly.

LGDs are among the biggest dogs in the world. They are often called mollossers, or mastiffs, and are usually large, powerful and muscular, although they also tend to be agile and fast. Phenotypical differences among them are usually associated with the herds they are protecting (e.g. when protecting sheep LGDs are preferred with white coat to blend in), with the conditions of their region (LGDs can support harsh climates and work for long hours) and with the native predator (the size of the predator has influence on the size of the dog).

Herding dogs, contrary to guarding dogs, do not live permanently with livestock. They are used mainly to drive or assemble the animals and show themselves to be a valuable help for man. There is a fundamental difference in the behaviour of puppies of herding and guarding dogs. Among herding dogs, the predation behaviour appears rapidly, such as fixation on an object or sibling followed by a predatory approach and sometimes by a pursuit (hunt)[63]. Thus, we get two groups of herd dog: herd dogs which encircle the livestock, such as the Border Collie; and herd dogs which pincer the livestock, such as the Appenzell Bouvier.

In contrast with guarding dogs, that do not express predation behaviour towards the herds, herding dogs were selected to display predation but instead of killing or significantly hurt the herds, they demonstrate an incomplete predation behaviour, somehow in between of Wolves and LGDs[13,64]. These dogs are usually medium size but possess great agility and speed[65]. Herding dogs are known for its intelligence and liveliness and, although they are devoted to the shepherd.

For this work, 24 livestock guarding and 7 livestock herding breeds were considered. As very few systematic works were done concerning livestock dogs, there is a lack of information regarding this subject. Most of the material here presented was retrieved from scientific journals or Kennel associations, although additional information available online had to be included. Special concern and criticism were applied concerning that information, and its usage was avoided whenever it was possible to try to maintain this work as factual as it could be.

The livestock breeds used in this work are internationally recognized by the *Fédération Cynologique Internationale* (http://www.fci.be) with the exception of Transmontano Mastiff and Terceira Cattle Dog, recognized by the *Clube Português de Canicultura* (Portuguese Kennel Club), the Basque Shepherd Dog and the Leonese Shepherd, recognized by the *Real Sociedad Canina de España* (Spanish Kennel Club), Akbash, recognized by the *United Kennel Club*, Molossus of Epirus, recognized by the *Kynologikós Omilos Elládos* (Greek Kennel Club) and Kurdish Shepherd dog, Persian Mastiff, Sage Kuchi, Turkmenistan Shepherd, Galician Palleiro and Spanish Wolf which have not been recognized yet by any Kennel Club (Table 1.1).

All the information was retrieved from the organizations where the breed is recognized. Information describing unrecognized breeds was retrieved mainly from an website called mollosserdogs.com, so may not be fully trustable.

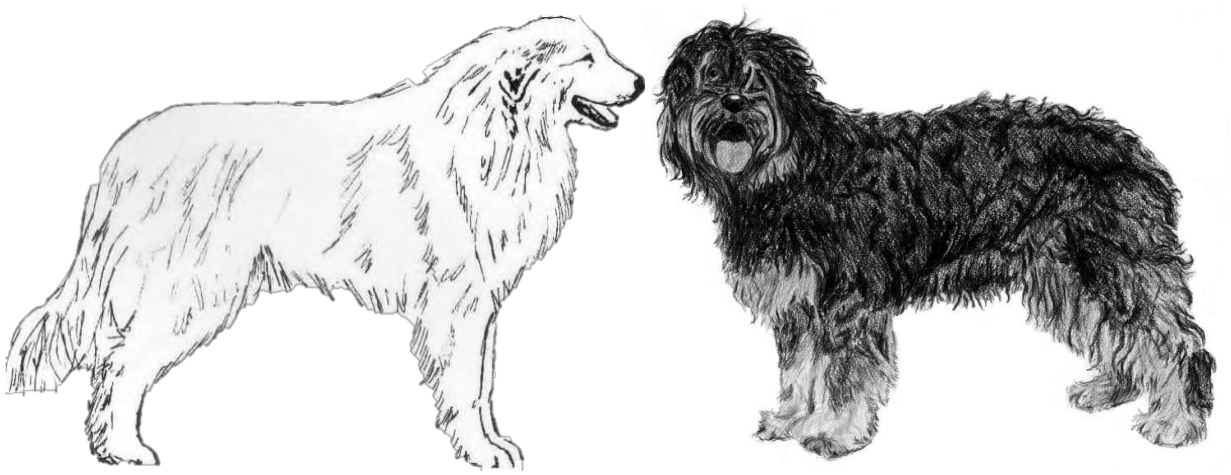


Fig. 1.2 – Illustration of classic livestock dogs. Left: Pyrenean Mountain dog[66]; Right: Portuguese Sheepdog[65]

Table 1.1 – The breed, geographical origin, type (based on physical and behavioral characteristics) and level of recognition ofthirty-one dog populations used in this study. The breeds, if not recognized internationally by the FCI (International Cynological Federation), are recognized by the KOE (Greek Kennel Club), CPC (Portuguese Kennel Club), RSCE (Spanish Kennel Club), UKC (United Kennel Club) or yet not recognized by any entity.

Breed	Abreviation	Geographical Origin	Type	Recognition
Aidi	Ai	Morocco	Livestock Guarding Dog	FCI[67]
Akbash	Ak	Turkey		UKC[68]
Anatolian Shepherd Dog	ASD	Turkey		FCI[69]
Castro Laboreiro dog	CLS	Portugal		FCI[70]
Caucasian Shepherd Dog	CSD	Caucasus Region		FCI[71]
Central Asia Shepherd dog	CAM	Central Asia Region		FCI[72]
Estrela Mountain Dog	EMD	Portugal		FCI[73]
Kangal Dog	KaD	Turkey		FCI[74]
Komondor	Kom	Hungary		FCI[75]
Kurdish Shepherd Dog	KD	Kurdistan Region		-[76]
Leonberger	Leon	Germany		FCI[77]
Maremma Sheepdog	Mare	Italy		FCI[78]
Molossus of Epirus	MoE	Greece		KOE[79]
Persian Mastiff	PeM	Iran		-[80]
Pyrenean Mastiff	PM	Spain		FCI[81]
Pyrenean Mountain Dog	PMD	France/Spain		FCI[66]
Rafeiro of Alentejo	RA	Portugal		FCI[82]
Sage Kushi	SK	Afghanistan		-[83]
Spanish Mastiff	SM	Spain		FCI[84]
St. Bernard	SB	Switzerland		FCI[85]
Tibetan Mastiff	MT	Nepal		FCI[86]
Transmontano Mastiff	TM	Portugal		CPC[87]
Turkmenistan Shepherd Dog	TSD	Turkmenistan		-[88]
Yugoslavian Shepherd dog	YGS	Former Yyugoslavia		FCI[89]
Portuguese Sheepdog	PS	Portugal	Livestock Herding Dog	FCI[65]
Basque Shepherd	BS	Spain		RSCE[90]
Galician Palleiro	GP	Spain		-[91,92]
Leonese Shepherd	LS	Spain		RSCE[93]
Saint Miguel Cattle Dog	SMC	Azores, Portugal		FCI[94]
Spanish Wolf Dog	SW	Spain		-[95]
Terceira Cattle Dog	TC	Azores, Portugal		CPC[96]

1.5.1 Other Iberian dogs

Besides Livestock working dogs there are other types of dogs, like hunting dogs, companion and toy dogs, etc. In the Iberian Peninsula, besides livestock guarding and herding breeds, there are many other autochthonous breeds like the Portuguese Podengo and Pointer and Spanish Alano and Galgo. The dominant kind of autochthonous dogs in Iberia is the hunting type, which referrers to dogs whose function is to hunt with or for their owners. There are several types of hunting dogs developed for various tasks, which include hounds, terriers, dachshunds, cur type dogs, and gun dogs. Hounds can be contrasted with gundogs, which assist hunters by identifying the location of prey and/or recovering shot quarry. The hound breeds were the first hunting dogs and appeared nearly at the same time as mastiffs. They have either a powerful sense of smell or great speed. Gun dogs were developed to assist hunters in finding and retrieving game and can be subdivided in three types: pointing, retrievers and flushing dogs.

For this study, we collected samples from both the Portuguese (Fig. 1.3) and Spanish Water Dogs, which are internationally recognized by the *Fédération Cynologique Internationale* (http://www.fci.be) and from the Villano Bulldog, which is not recognized by any association (Table 1.2).

Fig. 1.3 – Illustration of a Portuguese Water dog[97].

Table 1.2 – The breed, geographical origin, type (based on physical and behavioral characteristics) and level of recognition of three dog populations used in this study. Villano Bulldog, not recognized internationally by the FCI (International Cynological Federation), is yet not recognized by any entity.

Breed	Abreviation	Geographical Origin	Type	Recognition
Portuguese Water Dog	PWD	Portugal	Water/herding Dog	FCI[97]
Spanish Water Dog	SWD	Spain		FCI[98]
Villano Bulldog	VE	Spain	Hunting dog	-[99]

1.6. Molecular markers used in this study

Past studies on population genetics have traditionally used approaches with mitochondrial DNA (mtDNA). As mtDNA represents an only maternal lineage, additional markers targeting nuclear DNA, such as microsatellites and SNPs (single nucleotide polymorphism) need to be used for more accurate interpretation of population genetics, biodiversity, phylogeny and forensics[100].

Microsatellites are short sections of DNA where a simple motif, repeated up to about 100 times[101], highly polymorphic, abundant and evenly distributed throughout eukaryotic genomes. The popularity of these markers is due to their ease of amplification by PCR, their co-dominant nature and their typically high levels of allelic diversity at different loci. Microsatellites are a valuable tool in population genetics since they generally perform well for detecting structure at limited spatial scales[102], to measure or infer genetic diversity, gene flow, population size and populational events like bottlenecks[103–105]. However, microsatellites also have a few limitations, being the main the limited reproducibility across laboratories using different equipment or reagents[106]. This is related to allele calling because different observors may call differently the alleles (allelic binning), which is not relevant inside a study but it may be problematic when joining datasets from different studies. Different allelic binning is the cause of the inconsistences in more than 80% of studies comparison[107].

With the emergence of new genome-wide approaches, other markers have been recognized as powerful population genetic tools. Most of the mutations in the genome take the form of substitutions at a single base pair, dubbed single-nucleotide polymorphisms (SNPs)[108]. In theory, this marker ranges from bi- to tetra-allelic but is usually bi-allelic. It is a codominant marker, with the minor allele segregating at a frequency of at least 1%[109–111]. In comparison to microsatellites, SNPs present a less complex mutation pattern, as the number of possible alleles decrease from a high number in microsatellites to only two to four forms in SNPs. SNPs are a powerful and recent tools to use in bioinformatics and to analyse population genomic data, and represent the better available markers for whole-genome populational studies in the dog[28]. The major downside of using single nucleotide polymorphisms is the bias resulting from the choice of the initial panel of genotypes used to screen for polymorphisms (ascertainment bias)[112], which is considerably high. The problem underlying the use of SNPs when compared to microsatellites lies in the fact that the lower polymorphism in SNPs can only be compensated by an increase in the number of loci to achieve similar high statistical power. This can lead to an increase of two to six times the number of markers required to have similar statistical power as microsatellites and may become much more expensive[109,113].

High-density single nucleotide polymorphisms from domestic species are remarkable markers to perform population genetic analyses and have been enhancing our understanding of dog life history[28,34,39,46,47,114–116]. In this study, we used a mirroarray chip that contains more than 170,000 SNPs evenly spaced across the genome.

CHAPTER 2.

Genetic diversity and structure of Iberian livestock dogs

2.1. Introduction

Early work on the genetic diversity, divergence and relationships among dog breeds using a wide range of worldwide breeds and large panels of different molecular markers has made clear that dog breeds are grouped primarily by their function in human activities[28,38]. Very recently, a two-step process of dog breed creation was proposed, an ancient split by function and a more recent split by selection for physical attributes[117]. This genetic organization of breeds conforms well to the traditional breed categories used by the Federation Cynologique Internationale (FCI) that recognizes dog breeds based on behaviour, which are further split based on appearance, origin, or very specific usage (http://www.fci.be/en/Nomenclature/).

Dogs that are used in the livestock maintenance fall into two breed types according to their distinct function, the livestock guarding dogs (LGDs) and the livestock herding dogs (LHDs). Both are groups of breeds for which selection relied on working ability rather than on phenotypic appearance, and specially for LGDs, which were only moderately affected by the more recent split when intense breeding based on stringent physical criteria was highly encouraged[47,118]. LGDs have been selected for millennia to protect livestock from wild predators, especially wolves, and are known for their power and devotion, while LHDs are used to control the movement of the flock and are specially known by their agility and learning ability[13,44].

Iberian Peninsula is specially rich in autochnous livestock guarding dog breeds, where they have commonly been assisting shepherds. Seven molossian mountain type breeds are recognized, four in Portugal and three in Spain, six of them recognized by FCI and one recognized by the national Portuguese Kennel Club (Clube Português de Canicultura). A possible main reason is at the basis of this diversity. For many centuries the Iberian Peninsula was an important stage for a known traditional livestock practice known as transhumance, which permits the complementary exploitation of resources between highlands and lowlands[119]. This practice evolved a network of livestock movements, happening twice a year in a territory with widespread wolf presence. Sheperds protected herds and flocks using guarding dogs during these migrations[120] which may have repeatedly promoted selection for proficient guarding behaviours and thus the formation of different guarding breeds. The long history of human-wildlife conflict in Iberian Peninsula still continue, and guarding dogs have traditionally been considered an essential tool, which was robustly evidenced recently[121].

The few available genetic studies that include Iberian livestock breeds have focused on different small fractions of breeds and on a small number of genetic markers, and therefore the patterns of population diversity and structure for these breed remain poorly investigated. A first survey of nuclear genetic diversity and differentiation between the Spanish Mastiff and the Pyrenean Mastiff showed that these two LGD breeds were significantly differentiated based on protein polymorphisms[122]. Later, mitochondrial DNA diversity and structure was studied in four Iberian LGDs and two Portuguese

LHDs[123,124]. In general, LGDs exhibited more mtDNA diversity than LHDs, and both private and shared haplotypes were observed for all breeds, except for the Azores Cattle dog breed that did not share haplotypes. No geographic structure could be detected among livestock dog breeds for the mtDNA. Patterns of nuclear genetic diversity and differentiation were further assessed for five LGDs and two LHDs using microsatellites and AFLPs[125] and Y-linked SNPs and STRs[126]. Overall, the Portuguese sheepdog (LHD) and the Castro Laboreiro dog (LGD) exhibited the least diverse genome across the different compartments studied, while the Estrela mountain dog was the most diverse. Contrasting population structure among different markers hampered robust outcomes in this matter.

In this work we combine the analysis of the whole set of livestock dog breeds from Iberian Peninsula with the use of a powerful set of 47 and 6 autossomal and Y-linked microsatellites, respectively, to investigate the genetic diversity, differentiation and structure between livestock guarding and livestock herding breeds and among breeds withing this two groups. We hypothesize that past gene flow between these geographically close breeds hampered high levels of genetic differentiation within the two groups but encouragment for breeding only within own breeds promoted selection across recent times allowing structure to occur.

2.2. Material and methods

2.2.1. Sample collection and DNA extraction

We used 349 dog samples from all livestock guarding and herding Iberian dog breeds and from stray dogs (Fig. 2.1). Samples belong to the sample collections of CIBIO (Research Centre in Biodiversity and Genetic Resources, University of Porto; N=192) and CSIC (Spanish National Research Council, Museo National de Ciências Naturales, Madrid; N=157). Samples were mainly collected from working dogs and from different owners, avoiding all known direct relationships among individuals.

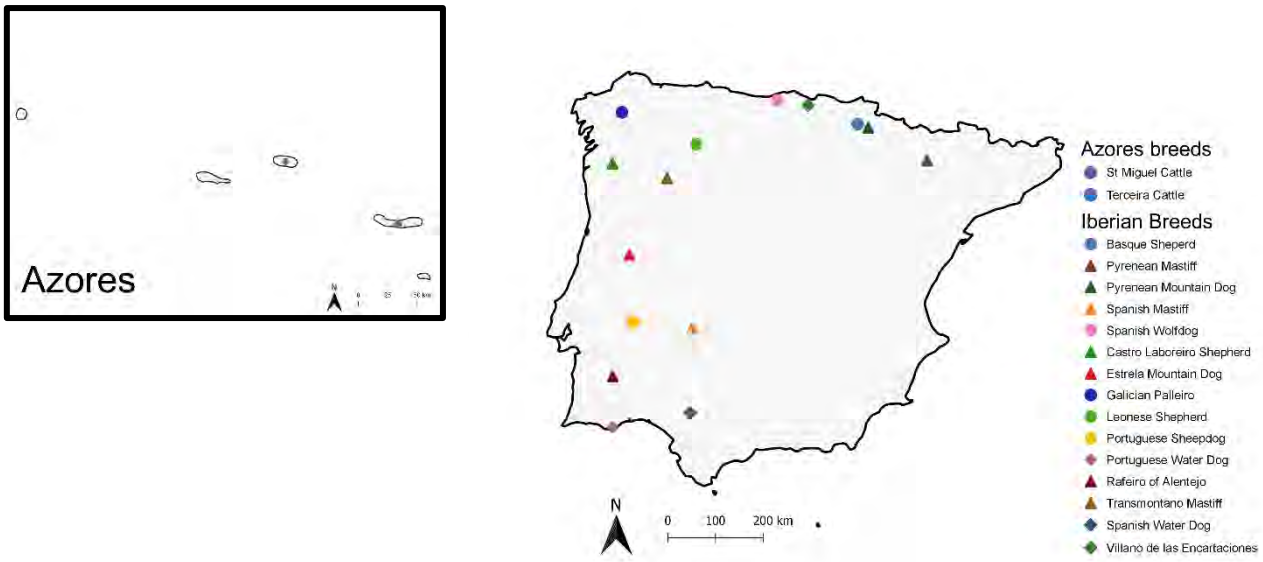


Fig. 2.1 – Region of origin for the 17 livestock guarding (triangles), herding (circles) and other (squares) dog breeds used in this study.

We extracted DNA from blood and tissue samples using the Qiagen DNeasy® Blood & Tissue kit, and from saliva samples using the Qiagen QIAamp® DNA Blood Mini Kit. The salting out DNA extraction procedure from Miller et al.[127] with some modifications (see the complete protocol used in the Supplementary metodology) was additionally perfomed for some samples to enhance the quantity and/or quality of extracted DNA. DNA was tested by electrophoresis using 0.8% agarose gels with GelRed fluorescent dye incorporated and evaluated in a UV transilluminator (BIO-RAD). Negative controls were included throughout the entire process to monitor for potential DNA contaminations. DNA concentration was quantified using the Qubit® dsDNA HS assay from Invitrogen™, which uses fluorometric quantitation to test the presence of double stranded DNA.

2.2.2. Y-linked and autosomal microsatellite genotyping

Six Y-linked microsatellite markers (Bannasch et al.[128]) were genotyped for 150 male dogs using a single multiplex reaction. On average, 10 males were genotyped per breed (a minimum of four individuals were used). For autosomal markers, a total of 47 loci[129,130,139–141,131–138] (Table 2.1) were genotyped for 300 dogs using five multiplex reactions. On average 20 individuals were analysed per breed (a minimum of 10 individuals per breed were used). Multiplex PCR reactions were performed using the QIAGEN Multiplex PCR Kit following the manufacturer's instructions (see Supplementary Table 1 to 5). Forward primers were M13-tailed to follow a fluorescent labelling protocol[142]. Reactions were prepared for a final volume of 10µL using 1µL of genomic DNA. The PCR reactions were performed in T100™ BIO-RAD Thermal Cyclers, always using negative controls to monitor possible contaminants (see Supplemneary Tables 6 to 11 for thermocycling conditions).

PCR products were tested by electrophoresis in 2% agarose gels with GelRed fluorescent dye) and the result was visualized in a UV transilluminator (BIO-RAD). Successful PCR products were diluted to a required concentration and then separated by capillary electrophoresis in an 3130xl Applied Biosystems® automated sequencer with subsequent fragment sizes scored in the GeneMapper software (Applied Biosystems) and checked manually. Individuals with more than 25% missing data were removed from the analysis.

2.2.3. Data analysis

Population diversity and differentiation, and the performing of population comparisons were calculated using ARLEQUIN vs. 3.5.2.2[143] software package, which included the estimations of molecular diversity indices for each breed. Haplotype diversity was calculated for the Y-linked dataset and observed/expected heterozigosity and number of alleles were estimated for the autosomal dataset. The number of polymorphich loci was estimated for both analysis. Allelic richness was calculated using the FSTAT vs 1.02[144] software package standardized for 294 samples. In this analysis, 5 loci were removed from the dataset since there was excessive missing data for these loci in some breeds, which decisively influences the result of the analysis.

ARLEQUIN was also used for testing deviation from the Hardy–Weinberg equilibrium and linkage disequilibrium in the autosomal dataset, with the level of statistical significance adjusted using strict Bonferroni corrections.

Table 2.1 – Autosomal and Y-linked loci analyzed in this study, along with repeat motif, allelic range, the multiplex that included each marker and references.

Name	Repeat	Allelic Range	Multiplex	Reference
AHT121	Di	74-118	Kit3	Holmes et al. 1995
AHT103	Di	71-89	Kit4	Holmes et al. 1995[130]
AHT111	Di	72-92	Kit 2	Holmes et al. 1993[129]
AHT132	Di	170-182	AP2	by N. Holmes
AHT137	Di	128-160	Kit3	Holmes et al. 1995
AHTh171	Di	216-240	Kit3	Breen et al. 2001
AHTh260	Compond	229-265	Kit3	Breen et al. 2001
AHTk211	Di	82-98	Kit3	Lingaas et al. 1997
AHTk253	Di	280-300	Kit3	Lingaas et al. 1997
C04.140	Di	132-160	Kit2	Ostrander et al.1993
C08.410	Di	95-125	Kit4	Ostrander et al.1995
C08.618	Di	188-208	Kit4	Ostrander et al. 1995
C09.173	Di	100-118	Kit2	Ostrander et al. 1993
C09.474	Di	111-133	Kit4	Ostrander et al. 1995
C13.758	Di	220-244	Kit2	Mellersh et al. 1997
C14.866	Di	221-257	Kit2	Mellersh et al. 1997
C20.253	Di	95-125	Kit2	Ostrander et al. 1993
C20.446	Di	173-201	Kit4	Ostrander et al. 1995
C22.279	Di	108-132	Kit3	Ostrander et al. 1993
C22.763	Di	192-216	Kit4	Breen et al. 2001
C27.442	Di	158-172	AP2	Ostrander et al. 1995
CPH02	Di	87-113	Kit4	Fredholm & Wintero 1995
CPH05	Di	95-131	Kit4	Fredholm & Wintero 1995
CPH09	Di	133-163	Kit4	Fredholm & Wintero 1995
CPH14	Di	185-205	Kit2	Fredholm & Wintero 1995
CXX.459	Di	141-167	Kit4	Ostrander et al. 1995
FH2001	Tetra	123-155	Kit3	Francisco et al. 1996
FH2010	Tetra	216-240	AP2	Francisco et al. 1996
FH2054	Tetra	121-181	Kit3	Francisco et al. 1996
FH2079	Tetra	246-292	AP1	Francisco et al. 1996
FH2161	Tetra	228-260	Kit4	Francisco et al. 1996
FH2848	Di	224-244	Kit3	Breen et al. 2001
INRA21	Di	87-103	Kit3	Mariat et al. 1996
INU005	Di	104-136	Kit3	Finnzymes, Inc
INU030	Di	136-156	Kit3	Finnzymes, Inc
INU055	Di	196-210	Kit3	Finnzymes, Inc
PEZ1	Tetra	99-131	AP1	Neff et al. 1999
PEZ3	Tri	106-147	AP2	Neff et al. 1999
PEZ5	Tetra	95-119	AP1	Neff et al. 1999
PEZ8	Compond	216-252	AP1	Neff et al. 1999
REN162C04	Di	189-215	Kit3	Guyon et al. 2003
REN169D01	Di	192-220	Kit3	Guyon et al. 2003
REN169O18	Di	145-171	Kit3	Guyon et al. 2003
REN247M23	Di	263-283	Kit3	Guyon et al. 2003
REN54P11	Di	223-245	Kit3	Guyon et al. 2003
REN64E19	Di	132-181	Kit4	Breen et al. 2001
VWF	Hexa	138-192	Kit2	Shibuya et al. 1994
990.35	Compond	125-134	CrY	Bannasch et al. 2005
650.79.2	Di	128-136	CrY	Bannasch et al. 2005
650.79.3	Di	124-134	CrY	Bannasch et al. 2005
MS41	Compond	212-230	CrY	Bannasch et al. 2005
MS34A	Di	318-326	CrY	Bannasch et al. 2005
MS34B	Di	316-330	CrY	Bannasch et al. 2005

Autosomal loci that were found to be in linkage disequilibrium were eliminated, leaving the final dataset with 43 autosomal markers. Population differentiation was estimated by pairwise mean F_{ST}[145] and by an analysis of molecular variance (AMOVA) also using ARLEQUIN. For the AMOVA, the genetic variation was considered for each dog function (guarding and herding) in a three-level structure, i.e., among functions, among breeds within each function, among individuals within breeds. Statistical significance of results was tested for $\alpha = 0.05$ using 10,000 permutations.

An intraspecific gene genealogy was inferred for Y-linked haplotypes using first the reduced median algorithm[146,147], whose result was then used as input for the median-joining network algorithm[148] in the software NETWORK vs. 5[147]. Gaps were treated as evolutionary events, and data were analysed with all characters weighting equally. For graphical output, haplotype frequencies were converted into proportional areas.

Bayesian clustering procedure implemented in STRUCTURE vs. 2.3.4[149] software was used to assign individuals to populations according to their genetic similarities. First, we carried out an analysis including all individuals and using model with admixture and correlated allele frequencies and the Usepopinfo tool disabled (no a priori information of the source population for each individual). We performed 10 independent runs for each number of populations (K) between 1 and 15, using a burn-in period of $1x10^5$ followed by $5x10^6$ MCMC iterations to guarantee the achievement of similar posterior probabilities of the data in each run. On a second analysis using the same procedure, we split guarding and herding dogs in two sub-datasets. We conducted 10 independent runs for each value of the putative number of populations between 1 and n, where n is the number of breeds in each sub-dataset. CLUMPAK[150] was used to truncate the different runs of each K and to display the graphical output for the best K for each analysis.

Finally, the genetic variance between individuals was structure in the dataset a Factorial Correspondence Analysis (FCA) was performed using the software GENETIX 4.2[151].

2.3. Results

To understand the relationships among Iberian livestock guarding and herding dog breeds, we have collected genetic data from 284 dogs representing 14 breeds. 10 Portuguese Water dogs were added to the dataset to serve as a comparison. The data set included all breeds recognised by the Fédération Cynologique Internationale, 3 breeds recognised by the two national kennel clubs, the Portuguese Kennel Club (Clube Português de Canicultura) and the Spanish Dog Royal Society (Real Sociedad Canina de España) and 2 additional breeds not yet recognized by any association.

2.3.1. Y-chromosome diversity and differentiation

The 14 breeds exhibited a total of 21 Y-linked haplotypes. Concerning the Y-linked loci for livestock dogs, almost the breeds are polymorphic, with LGDs displaying higher number of polymorphic loci (Table 2.2). Also, LGDs exhibited higher diversity, as measured for the number of haplotypes (Nh_{LGDs}=4.57; Nh_{LHDs}=2.67), and for the mean haplotype diversity (Hd_{LGDs}=0.40; Hd_{LHDs}=0.34). Despite this, within each group, the maximum and minimum values for haplotype diversity of breeds were not very different. Within LGDs, Rafeiro of Alentejo emcompassed the highest haplotype diversity (Hd=0.53) while Transmontano Mastiff presented the lowest (Hd=0.25). For LHDs, the highest diversity for the Y chromosome was observed for one Azorean breed, Terceira Cattle dog (Hd=0.51) and the lowest for the second Azorean breed, Saint Miguel Cattle dog and for the Portuguese sheepdog (Hd=0.22). Several breeds exhibited one or two private haplotypes, with Rafeiro of Alentejo and Leonese Shepherd having two (Table 2.2). Galician Palleiro was excluded from this analysis because genotyping was not achieved for a minimum of four samples.

Y chromosome differences among breeds accounted for 40% of genetic variation within the two groups considered (guarding and herding dogs) and 60% among individuals within breeds (Table 2.3).

Genetic differentiation for the Y chromosome between LGDs and LHDs (Fst=0.382; $p<0.05$), and among breeds within each dog function (F_{STLGDs}= 0.206; F_{STLHDs}= 0.348; Table 2.3) was generally very high. The most differentiated breeds withing LGDs were the Pyrenean Mastiff and the Castro Laboreiro dog (F_{ST} = 0.670), while the most differentiated pair of breeds within LHDs were Terceira cattle dog and the Leonese shepherd (F_{ST} = 0.780).

Table 2.2 – Comparison of Y chromosome gene diversity across Iberian breeds and for livestock guarding and herdings dog groups using six Y-linked microsatellites. Breeds, number of males typed, number of haplotypes, number of polymorphic loci, the haplotype diversity and its associated standard-deviation (sd) and number of private haplotypes are listed.

	N males	N haplotypes	N polymorphic loci	Haplotype diversity (sd)	No. private haplotypes
Livestock guarding dogs	12.43	4.57	4.00	0.40 (0.13)	0.71
Estrela Mountain Dog	11	4	4	0.39 (0.27)	1
Castro Laboreiro dog	17	4	5	0.28 (0.22)	0
Transmontano Mastiff	16	6	5	0.25 (0.14)	0
Rafeiro of Alentejo	14	7	5	0.53 (0.17)	2
Pyrenean Mastiff	20	6	3	0.44 (0.08)	1
Spanish Mastiff	4	3	2	0.50 (0.00)	1
Pyrenean Mountain Dog	5	2	4	0.40 (0.00)	0
Livestock herding dogs	9.17	2.67	2.83	0.34 (0.08)	0.83
Portuguese Sheepdog	9	2	4	0.22 (0.00)	0
Saint Miguel Cattle Dog	9	2	2	0.22 (0.00)	1
Terceira Cattle Dog	9	3	3	0.51 (0.15)	1
Basque Shepherd	9	3	3	0.44 (0.09)	1
Leonese Shepherd	12	3	3	0.26 (0.08)	2
Spanish Wolfdog	7	3	2	0.41 (0.17)	0
Portuguese Water Dog	11	4	3	0.36 (0.16)	0

Table 2.3 – Analysis of Molecular Variance for the Livestock Guarding and Herding dog groups for the Y-chromosome.

Source of variation	d.f.	Sum of squares	Variance components	% of variation
Between LGDs and LHDs	1	6.188	0.00166 Va	0.14
Among breeds within groups	11	63.771	0.46353 Vb	40.03
Among individuals within breeds	133	92.136	0.69275 Vc	59.83
Total	141	162.095	1.15794	

The neighbor joining network exhibits a central group of haplotypes (HY4, HY9-12 and HY14-15) that are more frequent and more shared among breeds (Fig. 2.2). The displacement of haplotypes in the network has no apparent structure in relation to LGD and LHD breeds. None haplotype was shared by all breeds, but HY4, 16 and 18 were all shared by the guarding breeds and HY11 and 12 were shered by the herding breeds. The most central haplotype in the network torso (HY14) is shared by 10 of the 14 breeds, both LGDs and LHDs. Private haplotypes are often the most distant from the network central torso. LGDs and LHDs presented eight and seven private haplotypes, respectively (Supplementary Table 12).

Table 2.4 – Genetic differentiation (FST) values between LGDs and LHDs Iberian dog breeds for the Y chromosome.

	Estrela Mountain dog	Castro Loboreiro dog	Transmontano Mastiff	Rafeiro of Alentejo	Pyrenean Mastiff	Spanish Mastiff	Pyrenean Mountain dog
Estrela Mountain dog	0	+	-	+	+	+	-
Castro Loboreiro dog	0.26365	0	+	+	+	+	-
Transmontano Mastiff	0.02106	0.19864	0	+	+	+	-
Rafeiro of Alentejo	0.22183	0.3774	0.24483	0	+	+	-
Pyrenean Mastiff	0.54912	0.67202	0.52311	0.30767	0	+	+
Spanish Mastiff	0.12936	0.36786	0.08829	0.25099	0.54228	0	-
Pyrenean Mountain dog	0.01212	0.30499	-0.07059	0.16682	0.6449	-0.05469	0

	Saint Miguel Cattle dog	Portuguese Sheepdog	Terceira cattle dog	Basque Shepherd	Leonese Shepherd	Spanish Wolfdog
Saint Miguel Cattle dog	0	+	+	+	+	+
Portuguese Sheepdog	0.44096	0	-	+	+	-
Terceira cattle dog	0.72839	0.66043	0	+	+	+
Basque Shepherd	0.21827	0.0608	0.67964	0	+	+
Leonese Shepherd	0.34966	0.56537	0.77872	0.44862	0	+
Spanish Wolfdog	0.40308	0.34446	0.76713	0.28684	0.57107	0

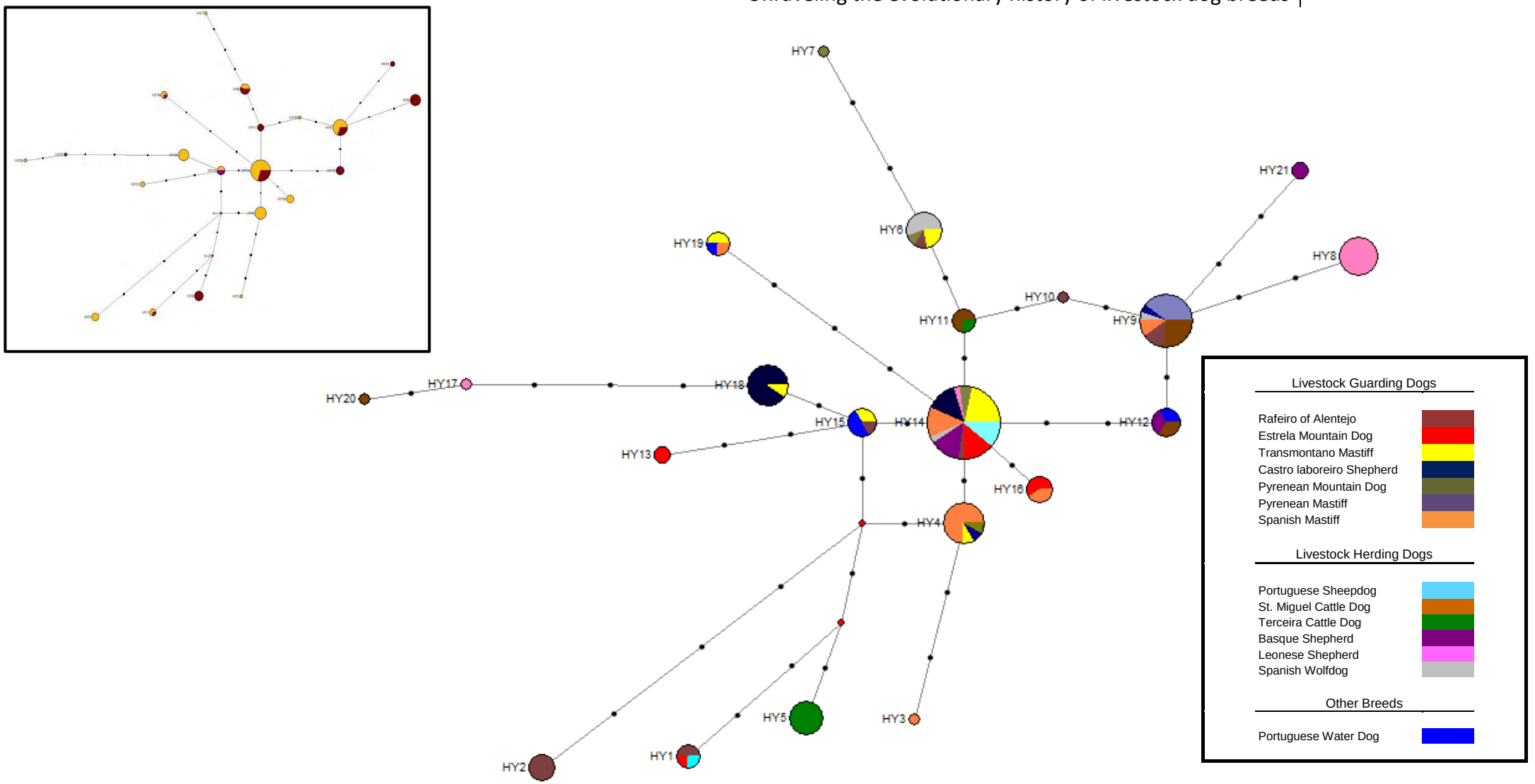


Fig. 2.2 – Median-joining network depicting relationships among Iberian livestock dog breeds. Circle areas are proportional to the corresponding haplotype frequency in each bread. Median vectors illustrated by red dots represent not sampled haplotypes. Mutational steps between haplotypes are represented by black dots. Inset: distribution of LGDs (yellow) and LHDs (brown) in the same network.

2.3.2. Autosomal diversity and differentiation

A total of 43 loci were analysed in 294 individuals (194 females and 148 males) of 15 different Iberian breeds. Between 13 (Spanish Wolfdogs) and 30 (Estrela Mountain Dog) individuals were analysed within each breed. LGDs showed higher genetic diversity values for all parameters. Within LGDs, Rafeiro of Alentejo (He=0.74; AR=4.26) and Transmontano Mastiff (He=0.73; AR=4.39) were the most diverse breeds while the Pyrenean Mastiff was the least diverse (He=0.59; AR=3.12). Among LHDs, genetic diversity ranged between the Portuguese Sheepdog (He=0.63; AR=3.58) and the Leonese Shepherd (He=0.69; AR=3.89) (Table 2.5).

LGDs and LHDs exhibited 10% of variation among breeds. However, most of the variation for all comparisons was found within individuals (89%) for both groups of dogs. Almost no variation was found between LGD and LHD groups (Table 2.7). The total genetic variation over all loci was moderate to highly partitioned between LGDs and LHDs (F_{ST} = 0.106; P < 0.001). Within each group, the global F_{ST} among breeds was 0.075 for LGDs and 0.092 for LHDs, with values ranging from 0.035 to 0.247 between the Basque Shepherd and Terceira Cattle dog and Saint Miguel Cattle dog and Portuguese Sheepdog comparisons, respectively. All pairwise F_{ST} comparisons among breeds within LGD and LHDs were statistically supported (P < 0.05; Table 2.6).

Table 2.5 – Genetic diversity indices for Livestock guarding and herding Iberian dog breeds using 43 autosomal microsatellites. Breeds, number of individuals typed, number of polymorphic loci observed heterozygosity, expected heterozygosity, the number of alleles and the allelic richness are listed.

Breeds	N Individuals	N polymorphic loci	Ho	He	N alleles	Allelic Richness (AR)
Livestock guarding dogs	22.14	31.14	0.66	0.69	5.47	3.91
Estrela Mountain dog	30	34	0.72	0.72	6.14	4.12
Castro Laboreiro dog	18	21	0.68	0.7	5.42	3.99
Transmontano Mastiff	24	39	0.7	0.73	6.51	4.39
Rafeiro of Alentejo	29	35	0.7	0.74	6.4	4.26
Pyrenean Mastiff	19	20	0.6	0.59	5.51	3.12
Spanish Mastiff	21	35	0.65	0.69	3.93	3.93
Pyrenean Mountain dog	14	34	0.6	0.63	4.4	3.55
Livestock herding dogs	18.43	25.14	0.63	0.66	4.99	3.74
Portuguese Sheepdog	18	20	0.58	0.63	4.54	3.58
Saint Miguel Cattle dog	17	12	0.7	0.66	5.02	3.72
Terceira Cattle dog	20	28	0.63	0.65	5	3.74
Basque Shepherd	18	27	0.61	0.67	5.12	3.90
Leonese Shepherd	21	29	0.66	0.69	5.49	3.89
Spanish Wolfdog	13	28	0.63	0.67	4.65	3.75
Galician Palleiro	22	32	0.58	0.64	5.12	3.61
Portuguese Water dog	10	27	0.65	0.66	4.3	3.64

Table 2.6 – Genetic differentiation (FST) values between LGDs and LHDs Iberian dog breeds using 43 autosomal microsatellites.

Livestock guarding dogs	Transmontano Mastiff	Castro Laboreiro dog	Pyrenean Mastiff	Spanish Mastiff	Pyrenean Mountain dog	Rafeiro of Alentejo	Estrela Mountain dog
Transmontano Mastiff	0	+	+	+	+	+	+
Castro Laboreiro dog	0.10276	0	+	+	+	+	+
Pyrenean Mastiff	0.10596	0.21105	0	+	+	+	+
Spanish Mastiff	0.04292	0.17198	0.10466	0	+	+	+
Pyrenean Mountain dog	0.07075	0.11199	0.16168	0.05873	0	+	+
Rafeiro of Alentejo	0.05085	0.04765	0.15299	0.09078	0.06722	0	+
Estrela Mountain dog	0.03952	0.1127	0.16527	0.08808	0.09305	0.0606	0
Livestock herding dogs	**Portuguese Sheepdog**	**Galician Palleiro**	**Saint Miguel Cattle dog**	**Terceira Cattle dog**	**Leonese Shepherd**	**Basque Shepherd**	**Spanish Wolfdog**
Portuguese Sheepdog	0	+	+	+	+	+	+
Galician Palleiro	0.11184	0	+	+	+	+	+
Saint Miguel Cattle	0.24722	0.13676	0	+	+	+	+
Terceira Cattle dog	0.08379	0.0996	0.16206	0	+	+	+
Leonese Shepherd	0.21318	0.15611	0.16423	0.09388	0	+	+
Basque Shepherd	0.09335	0.06125	0.13288	0.03464	0.13309	0	+
Spanish Wolfdog	0.07232	0.07367	0.19308	0.06778	0.14936	0.08967	0

Table 2.7 – Analysis of Molecular Variance for the Livestock Guarding and Herding dog groups using 47 autosomal microsatellites.

Source of variation	d.f.	Sum of squares	Variance components	% of variation
Between LGDs and LHDs	1	75.413	0.06820 Va	0.62
Among breeds within groups	12	653.009	1.10591 Vb	10.01
Among individuals within breeds	554	5469.050	9.87193 Vc	89.37
Total	567	6197.472	11.04603	

To determine if each breed form a breed-specific cluster we investigated the patterns of genetic structure using Structure software (Fig. 2.3). At the first possible partition (K=2) breeds were mostly clustered following their function, with LGDs and LHDs spliting apart. However, St Miguel Cattle Dog assigned on this first split with the LGDs and keeps its association to Spanish Mastiff and Rafeiro do Alentejo until K=6 where it became an different cluster. Fourteen out of the 15 breeds were correctly assigned when considering 15 different clusters (Fig. 2.5). K=13 was identified as the most likely number of populations in the datset using the delta K method[152] (Fig. 2.4). This number was also investigated using the mean ln(Pr(X|K) values, for which K=13 seems to be the best number of populations. Transmontano Mastiff is not assigned to its own cluster, instead, it shares with Spanish Mastiff most of its ancestry, but also with other different breeds possible indicating several origins. Galician Palleiro and Spanish Mastiff are assigned to their own clusters as soon as K=2. Furthermore, for the Leonese Shepherd it seems that part of the population investigated is fully assigned to its own private cluster, the other part is exhibiting admixture with other breeds.

LGDs and LHDs where then run apart (Fig. 2.6) and for LGDs the best K found was K=4 (Supplementary Fig. 1), assigning Spanish Mastiff, Rafeiro of Alentejo and Transmontano Mastiff to the same cluster and the remaining breeds to private clusters. For LGDs, the clustering analysis showed that the Transmontano Mastiff did not form a single cluster but instead shows the involvement of more than one founding breed, was the one not assigned to its own cluster and apparently sharing with Rafeiro of Alentejo a large fraction of its genome. The first split (K=2) occurs between the Pyrenean breeds and all the others. For LHDs the best K was achieved for 5 clusters (Supplementary Fig. 3), assigning Portuguese sheepdog, Terceira Cattle, Galician Palleiro and half of the Leonese Shepherds to their own cluster and assigning all the remaining breeds to their own clusters. In this analysis, we observed for the first split (K=2) a almost perfectly match for Portuguese and Spanish breeds. Like in the overall analysis, some of the Leonese Shepherds can not be clustered with the other individuals of that breed, instead being assigned to the same cluster as the Spanish Wolfdog.

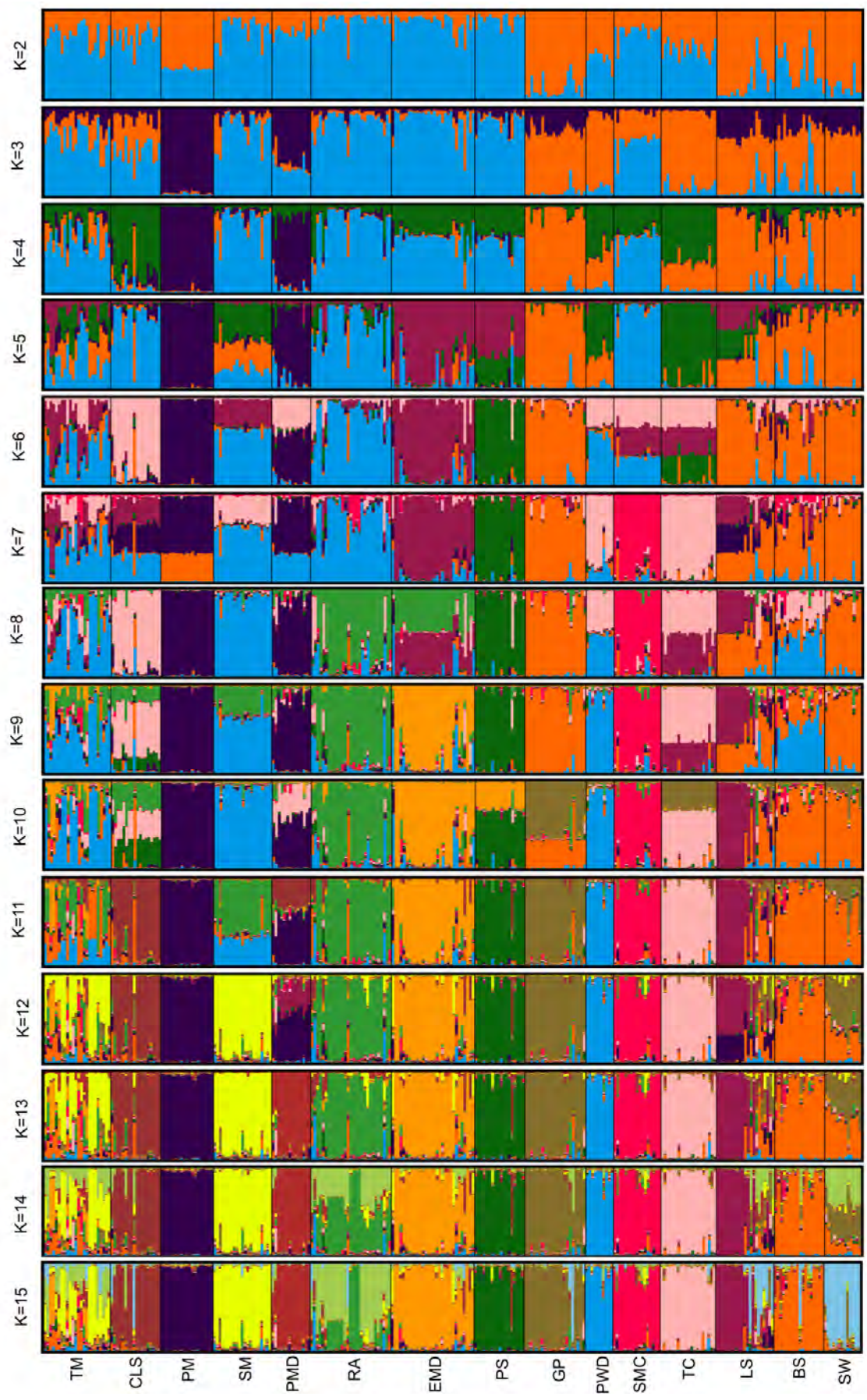


Fig. 2.3 – Proportion of membership of 324 individuals from 15 dog breeds for K=2 to 16, as calculated by Structure software. Multiple runs for each K were concatenated using ClumpaK, and distruct was used to generate images. SD = Stray Dogs; TM = Transmontano Mastiff; CLS = Castro Laboreiro dog; PM = Pyrenean Mastiff; SM = Spanish Mastiff; PMD = Pyrenean Mountain Dog; RA = Rafeiro of Alentejo; EMD = Estrela Mountain Dog; SMC= St Miguel Cattle Dog; PS = Portuguese Sheepdog; PWD= Portuguese Water Dog; TC = Terceira Cattle Dog; LS = Leonese Shepherd; BS = Basque Shepherd; SW = Spanish Wolfdog; GP = Galician Palleiro.

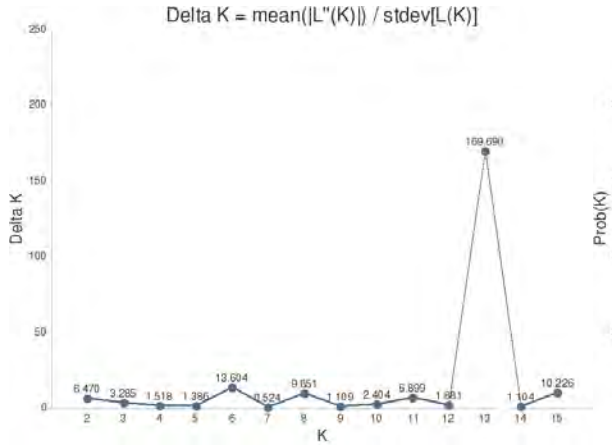


Fig. 2.4 – DeltaK graph for the whole dataset analysis using STRUCTURE. The highest probability by Evanno method is achieved for K=13.

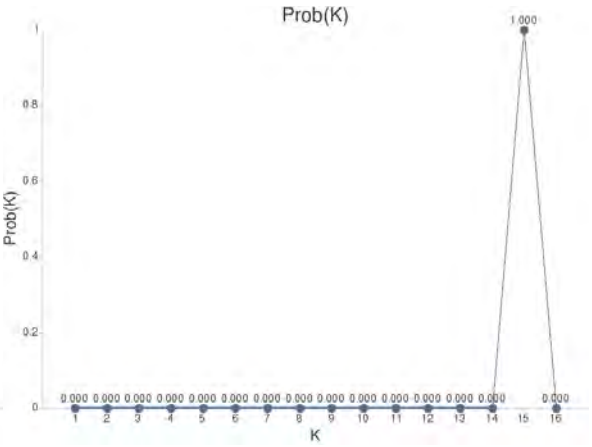


Fig. 2.5 – Log likelihood of the data for each K, for the whole dataset analysis using STRUCTURE. The highest probability of the data is K=15.

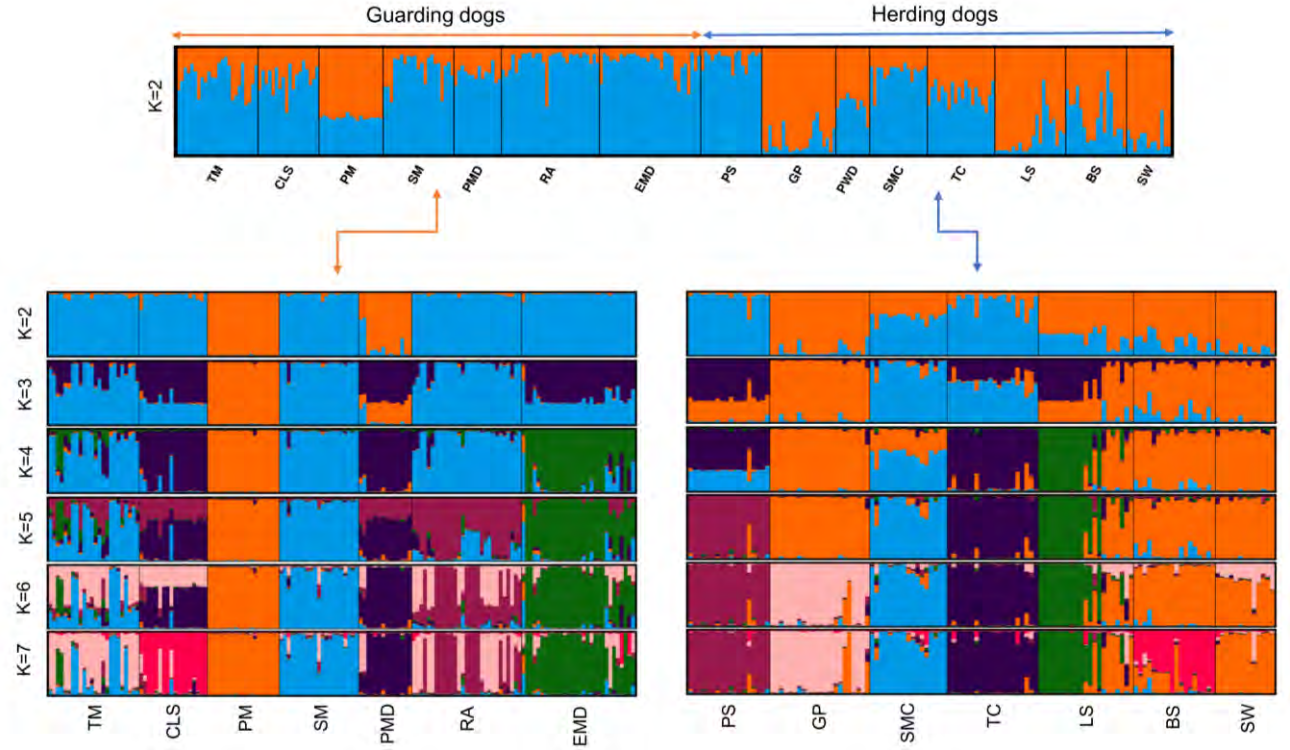


Fig. 2. 6 – Sub-dataset analysis, in which the two separate sub-datasets were considered (following the separation observed in K=2 from the overall analysis, middle top), of 324 individuals from 15 dog breeds, as calculated by STRUCTURE software. Multiple runs for each K were concatenated using ClumpaK, and distruct was used to generate images. In right bottom, sub-dataset with the LGDs, from K=2 to 7; In left bottom, sub-dataset with the LHDs, from K=2 to 7. TM = Transmontano Mastiff; CLS = Castro Laboreiro dog; PM = Pyrenean Mastiff; SM = Spanish Mastiff; PMD = Pyrenean Mountain Dog; RA = Rafeiro of Alentejo; EMD = Estrela Mountain Dog; PS = Portuguese Sheepdog; GP = Galician Palleiro; SMC= St Miguel Cattle Dog; TC = Terceira Cattle Dog; LS = Leonese Shepherd; BS = Basque Shepherd; SW = Spanish Wolfdog; PWD= Portuguese Water Dog.

The two-dimensional plot defined by the first two principal axes of a FCA explain 2.50% and 2.37% for axes 1 and 2 (Fig. 2.7). A great overlap among dog breeds can be observed (for both LGDs and LHDs), reinforcing the lack of a strong genetic differentiation between guarding and herding breeds.

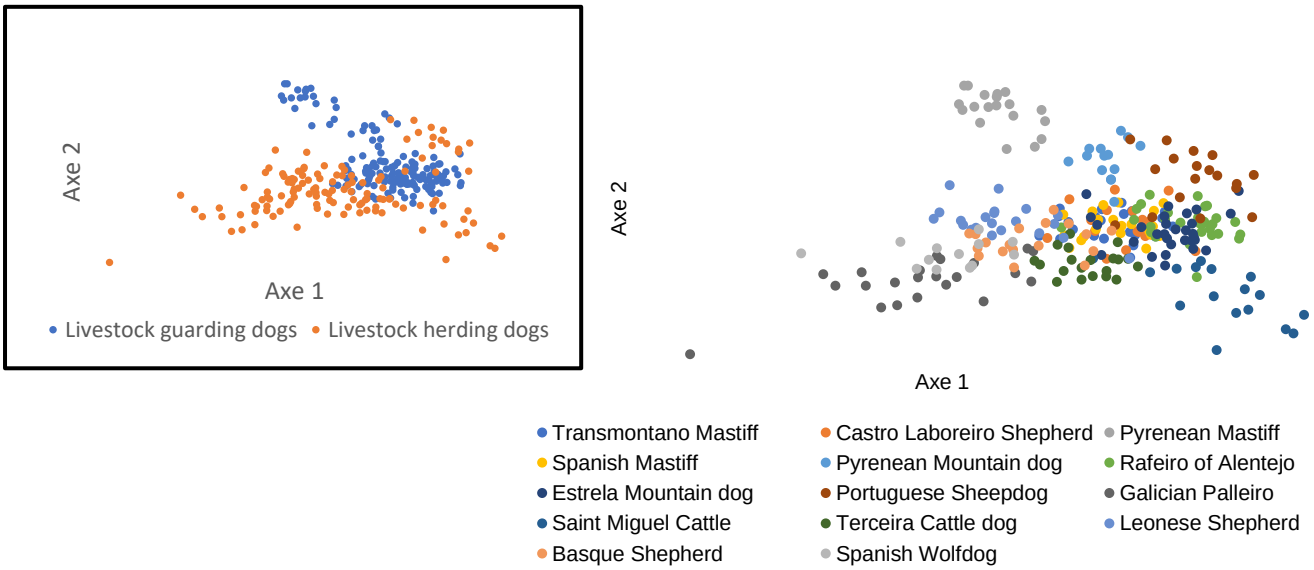


Fig. 2.7 - Scatter plot of the Factorial Correspondence Analysis for the autosomal analysis. Each color represents a dog breed. the first two principal axes of a FCA explain 2.50% and 2.37% for axes 1 and 2. Inset: distribution of LGDs and LHDs in the same FCA.

2.4. Discussion

Livestock guarding and herding dogs have always played a crucial role in human-wildlife co-existence. LGDs develop a life bond with the herd and are integrated as group members, protecting it by facing predators[153]. LHDs round up and guide the herd maintaining the integrity of the unit and hindering the access of predators[153]. These dogs are thus preventing livestock losses while also mitigating potential conflicts generated by each attack, and thus playing a proeminent role in carnivores conservation[121]. However, factors such as abandonment of traditional agriculture practices and local decline of predators' populations may jeopardize the interest of preserving working lineages of these breeds. In Iberian Peninsula, recent periods of critically low population size for some breeds[154,155] may have an impact in the genetic diversity of these dogs. Our investigation adresses this question using two different genomic compartments, the Y chromosme and the autosomes and providing a comparison basis among different breeds.

2.4.1. Patterns of genetic diversity

Overall consistency was observed among different markers and different analysis for the higher genetic diversity exhibited by LGDs in relation to LHDs. This is in agreement with expectations because herding dogs may suffer higher selection pressure for conformation and phenotype[156], which may be achieved through inbreeding. It also agrees with historical records[153,157] and previous genetic studies for mitochondrial DNA[123,124]. However, the LGD Pyrenean Mastiff exhibited the lowest genetic autosomal diversity among all studied breeds, while Castro Laboreiro dog and Transmontano Mastiff showed depauperated diversity for the Y chromosme and no private haplotypes, but a high autosomal diversity. In general, the haplotypic diversity observed for the groups of breeds or by the breeds themselves is lower than values reported by Pires et al.[126] for Iberian dog breeds (0.59 ± 0.11) however this values was achieved for a set of breeds from other groups than LGDs and LHDs. More general comparisons reveal that Iberian LGDs and LHDs have slightly higher Y diversity than the average worldwide breed dogs (0.38 ± 0.03[128]; 0.1007 ± 0.088[158]).

The low diversity values exhibited by the Pyrenean Mastiff support thedescribed history of the breed[81,159] that suffered a sharp population decline associated with wolf extinction in the Pyrenees and the consequent loss of its usefulness. It is also referred to that after the Spanish Civil War, in the mid 20th century, the country suffered a serious economic crisis and the costs to preserve the breed were too high, leading to the mentioned decline. In the seventies, protection measures and breed programs were implemented as an attempt to preserve the breed, having as an outcome a "new type" of Pyrenean Mastiff with typical features of the ancient mastiff[159]. Historical records of Castro

Laboreiro reveal that the breed was not used in the transhumance movements being isolated for long periods[160], which would be in accordance with observations by Pires et al.[124,125] and Amaro[161] that describe It as one of the least diverse breeds in the Iberian Peninsula. Our data, however, suggest otherwise since the nuclear diversity observed was high (He=0.70) and similar to other values LGDs in Iberia. Similar results were observed for the Azores Cattle and Saint Miguel Cattle dogs, isolated in the Azores islands, showing low values of Y haplotype diversity but high levels of nuclear heterozygosity.

Rafeiro of Alentejo, which is believed to be among the most ancient breeds in the peninsula[82,162], shares haplotypes with almost every Iberian breed studied and is a very diversed dog from both the Y-chromosome and autosomal point of view. Surprisingly it only shares one haplotype with Estrela Mountain Dog, which is considered to be its ancestor[162]. Thus, Y chromosome is not showing the patterns of common ancestry found previously for these two breeds using other compartments of the genome, namely the mitochondrial DNA[124] and the nuclear DNA[125].

AMOVA results suggest that the percentage of variation explained by dividing the two types of dogs, herding and guarding, is negligible for both the Y chromosome and the autosomes. Nonetheless, the diversity between breeds explain 40% of variation to the Y-chromosome, while only 10% is explained for the autosomes. Additionally, 60% and 88% of variation is found within individuals for the Y chromosme and the autosomes, respectively. Previous works[38,163,164] detailed values ranging from 18% to nearly 30% of diversity explained by breeds. Most of the different purebred dogs are morphologically distinct and also differ in behavior and physical properties, which was expected to indicate a higher level of differentiation among Iberian livestock dogs, but the highest differentiation levels were only found between individuals with the breeds.

2.4.2. Patterns of genetic structure

LGDs and LHDs exhibit a clear and well defined genetic substructure. We inferred high and significant F_{ST} values among the two groups of breeds ($Fst_{Y\text{-}Chr}$ = 38.2%; $Fst_{Autosomes}$ = 10.6%). LGDs were less differentiated than LHDs for both genomic compartments (Fst_{LGDs} = 20.6% and 7.5%, and Fst_{LHDs} = 34.8% and 9.2%, respectively for Y and autosomes). This was an expected results because of the recent more intense selection for herding dogs[165]. In this situations, it is likely that some animals are more intensively used in breeding favoring an inbreeding process and diversity loss.

The two herding dogs breeds originally from the Azores islands generally presented high levels of differentiation. Saint Miguel Cattle Dog and Terceira Cattle Dog exhibited F_{ST} > 10% with almost all breeds studied. This result could be either a consequence of isolation or of possible breeding events with non-Iberian breeds. Dogs brought by foreign ships in remote times when Azores was the

safe harbor in the middle of the Atlantic may have had an important role in the evolution of these island breeds[94,96,166].

The values of differentiation for the Y chromosome were, however, not evident in the relationships between Y haplotypes depicted in the median-joining network, in which haplotypes of the same group or the same breed did not group together (Fig. 2.2). This indicate that breeds shared the same Y chromosome information likely retained from ancestral periods or resulting from breed admixtures. Also, a recent study by Pires et al.[126] revelead that Iberian dogs do not share the paternal lineage with the Iberian wolf. Instead, the genome of Iberian dogs is composed by a panopoly of Y chromosome lineages that came from outside the Peninsula, probably at different points in time. Similar results were obtained using mtDNA haplotypes[26,124], apparently reflecting similar events.

Furthermore, to test whether these closely related breeds were genetically structured for the diploid genome, we applied the bayesian algorithm of STRUCTURE and could observe that in all but one case, the clusters separated individuals into individual breeds. The result provides evidence for the current breeding practice that maintain breeds as almost closed genetic pools, remarkably differentiated for allele frequencies. The exceptions was the Transmontano Mastiff, the newerly recognized breed by the Portuguese Kannel Club that likely represents a breed under development. The genomic information of Transmontano Mastiff is assigned to several clusters with a predominance of Spanish Mastiff (59.81%), Estrela Mountain dog (6.63%) and Rafeiro of Alentejo (26.58%). In fact, some authors have said that the Transmontano Mastiff is a variant of the result of a Rafeiro of Alentejo that was admixted with short-haired Estrela Mountain dogs[167,168]. Rafeiro of Alentejo used to move northwards from its original region in Alentejo (south of Portugal) during transhumance, and along the centuries, some dogs may have remained in the high regions of “Trás-os-Montes” giving birth to a new ecotype which is nowadays acepted as a new breed. This new breed is believed to have been bred with the Spanish Mastiff and with the Estrela Mountain dog[167]. In contrast with Transmontano Mastiff, other breeds that are not officially recognized by FCI or regional official entities, like the Galician Palleiro or the Spanish Wolfdog, were well assigned to their own cluster in the structure analysis. The Galician Palleiro was assigned to an independent cluster from K=10 onwards.

The first possible split (K=2), in which a moderate split between LGDs and LHDs can be observed, the Portuguese Sheepdog exhibits a genomic information shift towards LGDs.The available information about the breed states that it was created from the Briards, a French herding dog imported into Portugal in the early 1900s, that was by then crossed with the Pyrenean Sheepdog[65]. This might be a possibel explanation for the observed result. Remarkably different results were found in a previous work, where the Portuguese Sheepdog and the Spanish Mastiff did not group with the LGDs but grouped with other Iberian dogs, with different functions[125].

2.5. Conclusion

Our results from both genomic compartments were coincident in showing that the variation present in Iberian livestock dogs happens among individuals or among breeds but not between function determined groups. Centuries of artificial selection created clear different phenotypes and clear clusters of individuals that share the same basal genomic information and can be grouped in breeds, but Iberian livestock dogs still retain a lot of their ancestral stock and are not differentiated enough to clearly separate the breeds according to their function. Similarly, no pattern of genetic differentiation was consistently found between Portuguese and Spanish dogs, allowing us to consider that dog breeds within each country are derived from the same ancestral stock and maintained gene flow across the time, as expected in geographically closed regions.

Dog breeds should be viewed as dynamic populations, in which historic phenomena such as admixture, introgression, genetic isolation, and drift have occurred[169]. Our work contributed to additional clarifications on the history of Iberian livestock dogs that should encourage efforts by official entities in in Portugal and Spain to protect them from further periods of low population size.

CHAPTER 3.

The origin of livestock guarding dogs

3.1. Introduction

Although many things have been written about the origins of LGDs there have been surprisingly few systematic investigations into their origin. One thing we know is that the co-occurrence of livestock and dogs is necessary, although not sufficient for the emergence of LGDs from the primitive dogs. There are some historical preconditions for the presence of LGDs. The first one is the practice of large-scale extensive sheep-farming[56,170]. The second precondition is, of course, the necessity to protect the flocks from predators, which is confirmed by the absence of LGDs in places with no large carnivores as well as their disappearance from places where large carnivores became extinct[56,170]. The last condition for the presence of LGDs is the capability for livestock owners to feed them since large dogs need big quantities of meat to survive.

It is probable that early LGDs were primitive dogs that were raised with and bonded to livestock and simply adapted to the task they were educated to do. These primitive livestock guardians then most likely spread from their native regions with the still existing nomadic tribes along different migration routes or with merchants[64].

The origin of transhumance and of livestock and shepherd movements partially explains cultural phenomena and shaped the landscape across Eurasia[120]. These movements are likely to have facilitated gene flow among livestock guarding dog populations when they were brought into contact during these seasonal migrations[171]. The difference between nomadism and transhumance is that in nomadism the entire population moves with the herd, taking their homes with them[44]. In Transhumance, only the shepherds accompany the flocks while the rest of the population remains stationary at home, usually engaging in agriculture[56].

Over time, shepherds started to favour dogs that had morphological or behavioural characteristics that enabled them to outperform other dogs in pastoral tasks, giving birth to different populations of livestock dogs that became adapted to specific landscapes and livestock types[64]. These animals were not sexually isolated from the greater dog population until recent times, and even then, sexual isolation occurs mostly in the West as performed by kennel associations[44].

The main interest in studying livestock dog breeds in favour of other dog breeds resides in the fact that these animals could be the key to start to explain the first process of domestication mankind ever performed and also, they represent an indirect way of studying nomadic movements and the diffusion of culture and livestock practices across Eurasia, since they followed human movements and only recently were subject to intensive selection[56].

Also, the usage of livestock dogs is being publicized in many countries as being the "green solution" to predation control. LGDs have been used for millennia in Europe and Asia for flock protection, and today they are used in other parts of the world as well. Unfortunately, with the past decline in predator species worldwide, the use of livestock guardian dogs has decreased in many

regions. With the recent recover of large predators, the use of LGDs in Europe has resurged, for example, more than 1000 LGDs are now working in the Alps. Since predators usually flee at the sight of these dogs, their presence in flocks tends to ward off predators without seriously affecting endangered wildlife populations[153].

As previously pointed few systematic investigations were done for livestock dog breeds. There is not a lot of information available concerning these breeds, especially for the Asian and Middle Eastern dogs. Some of these breeds are in war zones where it is very difficult to reach, and their history remains in documents written in native languages hard to access. More research is needed to assess the presence of LGDs, either in archaeological remains or in ancient texts from other places.

The CanineHD BeadChip (created by Illumina) array enables the interrogation of genetic variation in any domestic dog breed, providing ample SNP[172]. Some work has been done using this and others genetic tools although they generally mix livestock guarding dogs with modern breeds[27,47,48,116,173] and there is no study published only focusing on LGDs.

The main goal of this project was to use genomic data to estimate diversity and relationships among livestock guarding dog breeds spread over the Mediterranean and Western Asia regions, evaluate their most probably origin and track their colonization routes, and ultimately, of humans, using a genome-wide coverage of 170.000 evenly spaced SNPs present on the BeadChip of Illumina and annotated on the dog reference genome sequence. We thus aim to uncover the Mediterranean livestock guarding dog's evolutionary history. We hypothesize that LGDs have originated somewhere near the Fertile Crescent and have evolved and differentiated in other breeds through multiple processes of isolation and gene flow.

3.2. Material and Methods

3.2.1. Sample collection and DNA extraction

Samples from 203 livestock guarding dogs spread across Eurasia (Fig. 3.1) were obtained. The samples belong to the sample collections of CIBIO (Research Centre in Biodiversity and Genetic Resources, University of Porto; N=70) and CSIC (Spanish National Research Council, Museo National de Ciências Naturales, Madrid; N=133). Samples were mainly collected from working dogs and from different owners, avoiding all known direct relationships among individuals.

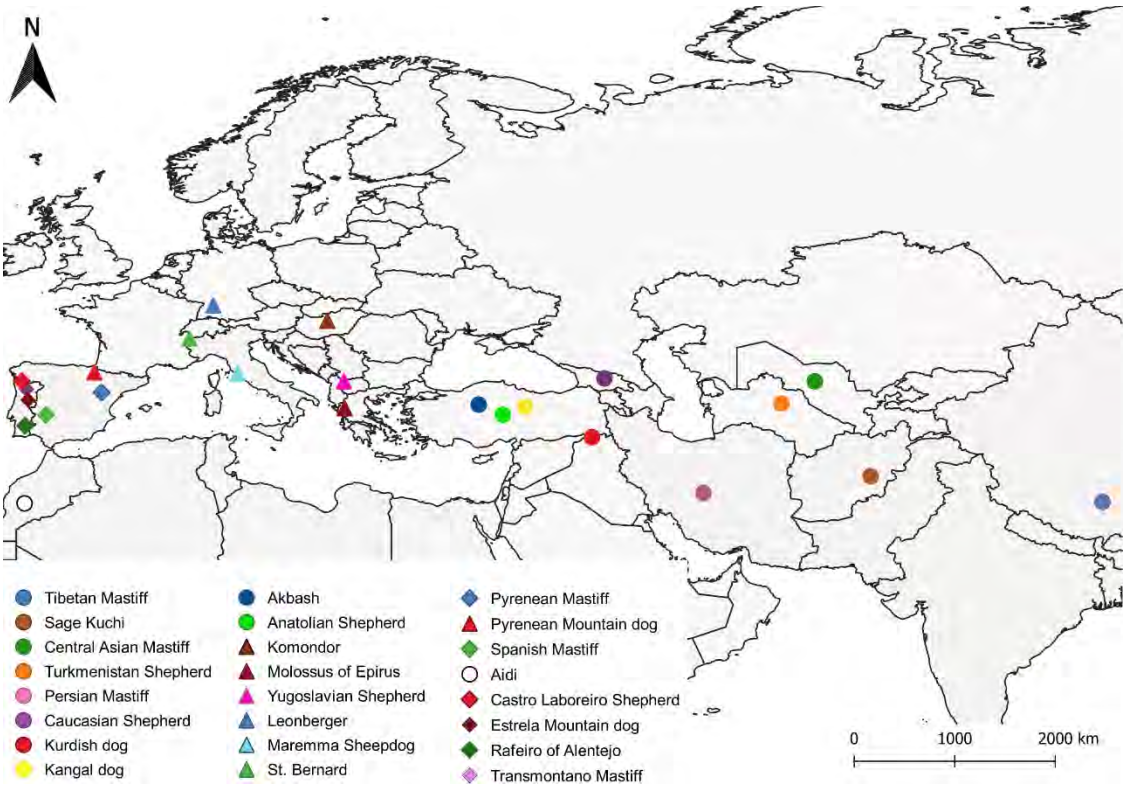


Fig. 3.1 – Region of origin for the 24 livestock guarding dog breeds used in this study.

Methods used to isolate the genomic DNA were chosen according to sample type. The Qiagen DNeasy® 96 Blood & Tissue kit was used to obtain high-quality DNA from both tissue and blood samples and the Qiagen QIAamp® DNA Blood Mini Kit was used to isolate genomic DNA from saliva samples. Also, a high salt method (see attachments) was implemented for the extraction of samples when the Qiagen kits proved to be inefficient.

The presence of isolated DNA was tested through electrophoresis (agarose 0.8% with GelRed fluorescent dye; 3µL bromophenol blue; 2µL genomic DNA), evaluating the output in a UV

transilluminator (BIO-RAD). Negative controls were included throughout the entire process to monitor for potential DNA contaminations. To assess the concentration of DNA, samples were quantified using the Qubit® dsDNA HS assay from Invitrogen™, which uses fluorometric quantitation to test the presence of double stranded DNA.

3.2.2. SNP genotyping and data cleaning

The Illumina canine high-density bead chip containing the 173,662 SNPs was used to genotype the dog samples. 10 grey wolves (C*anis lupus*) and 10 stray dogs were also genotyped using the bead chip. Although the canine bead chip was conceived using the dog genome, it is possible to apply it to wolves because the domestic dog is an extremely close relative of the grey wolf, from which it differs by only ~0.04% in nuclear coding-DNA sequence[174].

The Illumina Genome Studio software was used, applying the Infinium® Genotyping Data Analysis assay, to assure quality control of both sample and marker. A sample call rate threshold of 90% was first applied to the dataset and samples under the threshold were removed, although samples slightly under were still considered and carefully analysed (example, Fig. 3.2).

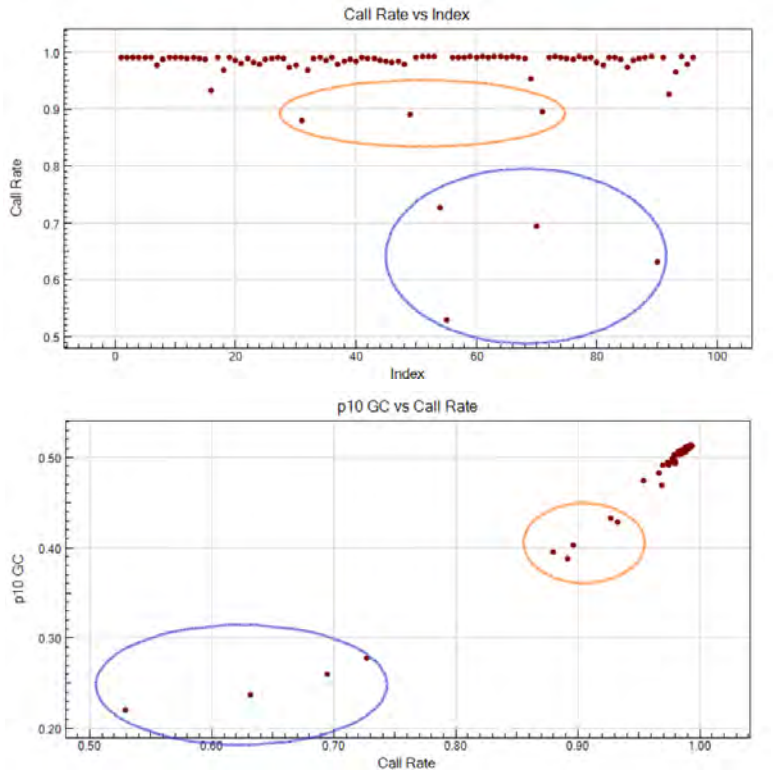


Fig. 3.2 – Scatter plot of call rates across a set of samples (Index) (above); Scatter plot of 10% GC scores compared to call rates of a sample set (below). Samples with abnormally call rates are easy identifiable and may: need to be excluded (blue circle) or manually edited (orange circle).

Initial marker quality criteria were then applied. First, the markers were evaluated according to their cluster separation, i.e. if sample genotypes form three non-overlapping clusters (example in Fig 3.3). Second, samples with low or confounded signal (ABR mean < 0.3) and low reproducibility (GTS < 0.6) were discarded. The markers were also evaluated for heterozygosity excess or deficiency and for false homozygotes and manually edited or erased, according to the need. After this, all the individuals from the Kangal dog and Pyrenean Mastiff breeds were eliminated, leaving the dataset with 22 breeds and 139 LGDs.

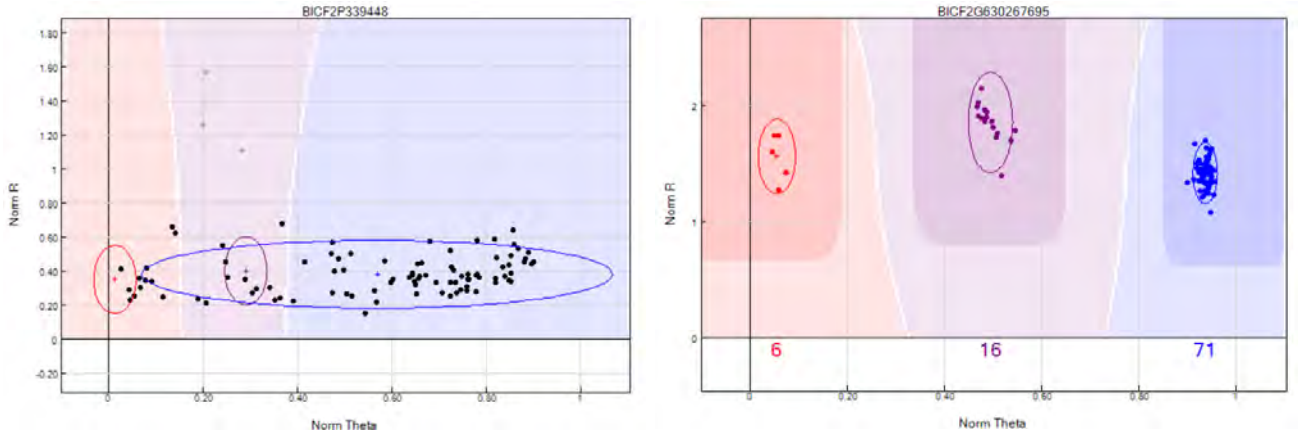


Fig. 3.3 – SNP NORM-R vs. NORM-THETA Cluster Profiles. Left: A poor-performing SNP. No genotypes can be obtained from this intensity plot, as no distinct clusters exist. Right: A valid SNP with three distinct clusters representing AA, AB, and BB genotypes, coloured red, purple, and blue.

The dataset was exported from Genome Studio and PLINK 1.9[175] was used for further quality control on the markers. SNPs on the allosomes were discarded since there would be a lack of information for those SNPs depending on the sex of the individuals. Other filters were applied to analyse the dataset, including filters of minor allele frequency (MAF=0.01), call frequency (GENO=0.1), and maximum individual missing rate (MIND=0.1). Finally, the dataset was filtered to find loci in linkage with the "—indep-pairwise" function of PLINK1.9 (--indep-pairwise 50 10 0.8).

The "--genome" function of PLINK 1.9 was used to calculate the relatedness between each pair of individuals. Levels of relatedness were found to be generally low between the majority of the individuals and none was alertly higher than the values commonly accepted for dog breeds in previous publications, but individuals that had more than 25% of relatedness were eliminated from the dataset since some of the Software's that were used assume that the individuals are unrelated.

Additionally, due to the disparity on the number of samples for each breed, individuals from breeds with a sample size bigger than 10 were randomly chosen and removed, in order to reduce a possible size bias in the analysis. Finally, the dataset was left with 106,072 SNPs for 98 LGDs, 10 Stray Dogs and 10 Wolves for further analysis.

3.2.3. Data analysis

3.2.3.1. Diversity analysis

Population diversity and differentiation were estimated using the ARLEQUIN v. 3.5.2.2[143] software package which was also used to perform population comparisons. The genetic differentiation between breeds was calculated through the aid of pairwise mean F_{ST}[145]. Additionally, an analysis of molecular variance (AMOVA) was performed using ARLEQUIN, in which dog breeds were assembled to three different groups (Asian, European and Iberian) in an attempt to try to explain their genetic variation. The significance for the F_{ST} and the AMOVA was tested using the typical criteria of Arlequin (10,000 permutations, with significance set at $\alpha = 0.05$). Also, Plink 1.9[175] function '--het' was used to calculate the inbreeding coefficient (F) for each population. F is calculated through: $\frac{[\text{observed hom.count}] - [\text{expected count}]}{[\text{total observations}] - [\text{expected count}]}$.

3.2.3.2. Population structure analysis

After data cleaning, the database was summited to various analysis of population genetics. ADMIXTURE[176] is a program for estimating ancestry in a model-based manner from large autosomal SNP genotype datasets, assuming the individuals are unrelated. Population divisions (K) ranged from two to the number of breeds (22), with each K being assessed in three different runs, in order to estimate the most likely number of partitions. ADMIXTURE includes a cross-validation (cv error) procedure that we used to identify the value of K for which the model has best predictive accuracy.

To ensure fidelity in the results three different preliminary analysis were done. The first included grey wolves as an outgroup in the dataset, the second included stray dogs as an outgroup in the dataset, and the third only had LGDs and no outgroup. The cross-validation errors for each analysis were compared (see Fig. 3.5) and the only significant difference found was that the optimal K differed when using wolves as an outgroup, which was expected. Since the usage of an outgroup in this particular analysis seemed to be almost irrelevant for K values higher than 2, further analysis continued without inclusion of wolves or stray dogs.

After K=2 was found to be the best partition (lowest CV error) model to divide the breeds, 3 different groups were defined, i.e. the breeds in the extremities that displayed different genetic signals and the middle breeds that displayed both signals, for subsequent analysis:

- The Asian group, comprising all breeds that have originated east of Turkey, including the Turkish ones;

- The European group, which comprises all breeds originating from west of Turkey to the Pyrenean Mountains, including Aidi, the Moroccan Breed;
- The Iberian group, comprising all breeds originating in the Iberian Peninsula.

On a second analysis, breeds from the previous defined groups were split in three different sub-datasets, using the same analysis procedure from K=1 to K=N (where N is the number of breeds in each sub-dataset). This analysis is particularly important to assess nested structure or sub-structuring that the general analysis of ADMIXTURE is not able to identify. CLUMPAK[150] was used to truncate the different runs of each K and display the graphical output.

3.2.3.3. Phylogenetic analysis and migration events estimation

Genetic distances between individuals were estimated using the '--distance' function of Plink 1.9[175] and the "1-ibs," "square," and "flat-missing" modifiers. The dataset was bootstrapped 1000 times to ensure statistical power, using a custom script that subsamples the original dataset, since Plink has not the ability to calculate bootstraps on its own. A neighbour-joining phylogeny was created using the PHYLIP software package[177]. Neighbour-joining is a distance matrix method that produce an unrooted tree without the assumption of a clock and is capable of handling large datasets. This method builds trees by successive clustering lineages, setting branch lengths as the lineages join.

When using bootstrap samples ties in the distance matrix are probable of appearing, with Neighbour resolving this problem in a way dependent of the order of the input file. When having many bootstrap samples from a data set, the apparent strong support for one resolution of a multifurcation may be the result of an artefact of the order of species in the input file hence, the "J" option was implemented to randomize the input order of individuals in the dataset.

The consensus package of PHYLIP was used to generate a consensus neighbour-joining tree. This tree was later depicted in FigTree v1.4.2[178] and different colours were used to differentiate the distinct dog breeds.

An extended analysis of the relationships among dog breeds was performed using Treemix software[179], a software which method does not consider the populations' history only as a phylogenetic tree, but also considers events of gene flow between populations. First, allele frequencies and missing data for each marker were calculated in Plink 1.9, using the "--freq" and "--missing" options and the individuals were grouped according to their breed using the "--within" function. Second, the dataset was converted to the input format of Treemix using the "plink2treemix" script. After, Treemix was used, in which the number of admixing events ranged from 0 (no migratory events) to N, where N is the number of breeds.

When TreeMix[179] calculates the covariance matrix among populations, it includes a correction for sample size, since this program considers populations as an entity and not the individuals. In some

cases (e.g., with single individuals in a population), this can lead to overcorrection. So, in order to avoid this overcorrection and improve the statistical power of the analysis, all the populations with only one individual were withdrawn from the dataset. After, the analysis was redone, first with sample size correction on and second with it off ('-noss'), in order to compare the results.

Due to implausible migration events inferred in the first analyses, a "stepping-stone" model was created in which, instead of inferring migratory events across the total number of breeds, the analysis was done considering migratory events within the previously defined groups (Asian, European and Iberian). Since this could also create problems because migratory events between populations on different groups wouldn't be considered, two "transitional" groups were created:

- The Eastern groups, comprising all the breeds from the Asian and European groups;
- The Western groups, comprising all the breeds from European and Iberian groups;

After this, the same analysis that was previously done was applied to the five sub-datasets, and the results were compared (see Fig. 3.10). The sample size correction function did not seem to interfere in the analysis although, it was decided to turn it off through the rest of the experimentation.

To increase the confidence in the results of Treemix[179], considering past studies and the previous concerns regarding the statistical power of the dataset, a number of conditions were created. To accept a given number of migratory events for a certain tree, two conditions must be met:

1) The maximum likelihood values must have stabilized when increasing the number of migrations;
2) The residuals must not significantly change after the moment when the maximum likelihood stabilizes;

A consensus tree was then constructed, with only the migratory events that had significant p-values, using the migrations inferred in the stepping stone model and compared to the tree obtained in the whole dataset analysis.

The "RColorBrewer" package of the R software[180,181] was used to plot the trees and residual errors for each number of admixing events. FigTree v1.4.2[138] was used to depict the maximum likelihood trees generated by Treemix[179].

To confirm significant admixture among the dog breeds a test included in the Treemix software, the f3 statistical analysis[182], was performed. This test was implemented using the treepop command line and is of the form f3(A; B, C), where a significantly negative value of the statistic implies that population A is admixed.

3.3. Results

3.3.1. Diversity analysis

Diversity estimations were calculated using the 106,072 SNPs for 98 livestock dog breeds, 10 grey wolves and 10 stray dogs (Table 3.1). Concerning the observed and expected heterozygosity, the values were, in general very similar to each other, ranging from 0.30711 to 0.38164, in the Maremma Sheepdog and the Kurdish Shepherd, respectively. The coefficient of inbreeding is used to classify pure breed dogs in many kennel clubs, with values higher than 0.13 being in the normal range for pure-breed dogs. Here, western breeds displayed higher values of F when comparing to the Asian breeds, with Maremma Sheepdog having the highest levels of inbreeding in the whole dataset. Also, western breeds tend to have a higher number of polymorphic loci. Transmontano Mastiff and Estrela Mountain dog, opposing to the other western breeds, displayed low levels of inbreeding but, have a high number of polymorphic loci.

Table 3.1 – Comparison of autosomal SNP diversity from livestock guardings dogs. Breeds, number of individuals typed, number of polymorphic loci, observed heterozygosity (Ho), expected heterozygosity (Ho) and F-coefficient of inbreeding (F) are listed. Values for some breeds are missing due to their low sample size.

Breed	N individuals	Polymorphic loci	Ho	He	F
Aidi	3	-	-	-	-
Akbash	7	88712	0.37803	0.38834	0.0564857
Anatolian Shepherd dog	1	-	-	-	-
Castro Laboreiro dog	8	89333	0.36379	0.37753	0.1017925
Caucasian Shepherd dog	1	-	-	-	-
Central Asian Shepherd dog	1	-	-	-	-
Estrela Mountain dog	10	91400	0.37052	0.36887	0.071238
Komondor	1	-	-	-	-
Kurdish Shepherd dog	8	89646	0.38164	0.38078	0.06382
Leonberger	2	-	-	-	-
Maremma Sheepdog	8	93133	0.30711	0.38359	0.2292463
Molossus of Epirus	3	-	-	-	-
Persian Mastiff	1	-	-	-	-
Pyrenean Mountain dog	2	-	-	-	-
Rafeiro of Alentejo	9	93600	0.33766	0.37635	0.1491014
Sage Kushi	2	-	-	-	-
Spanish Mastiff	10	92134	0.36013	0.36959	0.110226
St. Bernard	1	-	-	-	-
Tibetan Mastiff	2	-	-	-	-
Transmontano Mastiff	8	93896	0.36053	0.3826	0.0821003
Turkmenistan Shepherd dog	3	-	-	-	-
Yugoslavian shepherd dog	7	88712	0.37803	0.38834	0.0634457
Stray Dogs	10	98002	0.37003	0.37389	0.0273333
Grey Wolves	10	66879	0.28214	0.34632	0.142642

Almost no variation was found between Iberian, European and Asian groups in the AMOVA. The groups exhibited around 5% of variation among the breeds within the groups. Thus, most of the variation for all comparisons was found among the individuals (93.61%) for the groups of dogs (Table 3.2).

Table 3.2 – Analysis of Molecular Variance for the Asian, European and Iberian dog groups for the SNPs.

Source of variation	d.f.	Sum of squares	Variance components	% of variation
Between groups	2	55758.987	142.68647 Va	1.20
Among breeds within groups	19	309266.091	619.43153 Vb	5.20
Among individuals within breeds	174	1941216.595	11156.41721 Vc	93.61
Total	195	2306241.673	11918.53521	

Regarding the F_{ST} (Table 3.3), breeds appear to be generally very low differentiated, with the mean value among livestock guarding breeds being quite low (F_{ST}=0.0476). F_{ST} values ranged from 0.01588 between the Transmontano Mastiff and Stray dogs to 0.28327 between the Castro Laboreiro dog and the Grey Wolf. Non-significant values appeared for the Kurdish dog and the Akbash, with each other and with the Yugoslavian Shepherd and the Maremma Sheepdog.

The highest value of F_{ST} found among the livestock guarding dogs is between the Spanish Mastiff and the Kurdish dog (F_{ST}=0.07741).

Table 3.3– Comparisons of F_{ST} values between populations. TM = Transmontano Mastiff; CLS = Castro Laboreiro dog; RA = Rafeiro of Alentejo; EMD = Estrela Mountain Dog; SM = Spanish Mastiff; = Maremma Sheepdog; YGS = Yugoslavian Shepherd; KD = Kurdish dog; Ak = Akbash.

	TM	CLS	RA	EMD	SM	Mare	YGS	KD	Ak	SD	Wolf
TM	0	+	+	+	+	+	+	+	+	+	+
CLS	0.03848	0	+	+	+	+	+	+	+	+	+
RA	0.02009	0.05202	0	+	+	+	+	+	+	+	+
EMD	0.03179	0.06358	0.04197	0	+	+	+	+	+	+	+
SM	0.03191	0.07158	0.05018	0.06552	0	+	+	+	+	+	+
Mare	0.02417	0.05032	0.0357	0.04768	0.05728	0	+	-	-	+	+
YGS	0.02676	0.05497	0.03901	0.05176	0.05924	0.03427	0	-	-	+	+
KD	0.04585	0.07521	0.05876	0.0696	0.07741	0.05241	0.03668	0	-	+	+
Ak	0.03396	0.06149	0.04421	0.05613	0.06467	0.03982	0.0227	0.0253	0	+	+
SD	0.01588	0.03794	0.03067	0.04039	0.05062	0.02508	0.02978	0.0498	0.03735	0	+
Wolf	0.26075	0.28327	0.26946	0.27835	0.28195	0.26753	0.25484	0.25549	0.24095	0.25241	0

3.3.2. Population structure analysis

3.3.2.1. Results for the preliminary analysis

Regarding the preliminary analysis for the structure population analysis using the ADMIXTURE software, it was found the use of an outgroup, such as Stray dogs or Grey Wolves, to be nearly irrelevant. The results, when compared to the analysis without an outgroup, are identical (Fig. 3.4). It seems that the usage of Stray Dogs is irrelevant and although there is a difference in the analysis using Grey Wolves it doesn't seem necessary to use them to obtain good and strong results. The only interesting result is that, when using the Grey Wolves as an outgroup, signals of wolf admixture can be found in some dog breeds, mainly in the Middle Eastern and Asian breeds. Besides, some admixture can be also found in western breeds, pointing the Leonberger as an example.

About the cross-validation error for the three analysis (Fig. 3.5) it can be perceived that again, there is no significant difference between the analysis and, for K higher than 2, results for the 3 different analysis are closely identical.

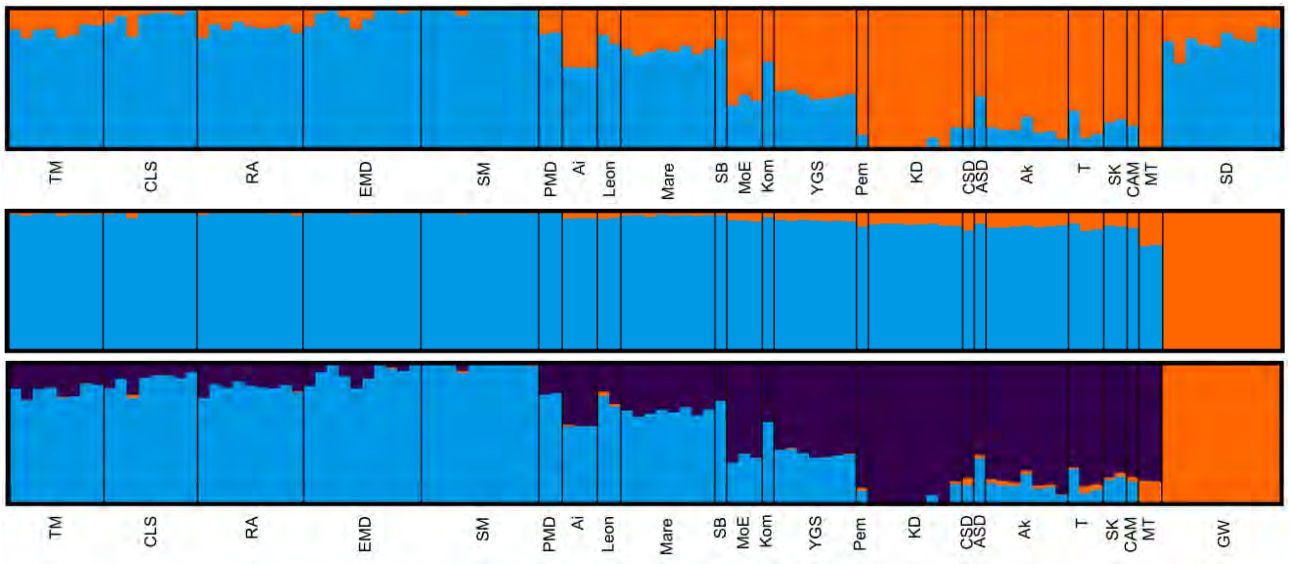


Fig. 3.4 – Proportion of membership of 98 individuals from 22 dog breeds for K=2, using the Stray Dogs as outgroup (above) and for K=2 and 3, using the Grey Wolves as an outgroup (down), as calculated by ADMIXTURE software. Multiple runs for each K were concatenated using ClumpaK, and distruct was used to generate images. TM = Transmontano Mastiff; CLS = Castro Laboreiro dog; RA = Rafeiro of Alentejo; EMD = Estrela Mountain Dog; SM = Spanish Mastiff; PMD = Pyrenean Mountain Dog; Ai = Aidi; Leon = Leonberger; Mare = Maremma Sheepdog; SB = St. Bernard; MoE = Molossus of Epirus; Kom = Komondor; YGS = Yugoslavian Shepherd; PeM = Persian Mastiff; KD = Kurdish dog; CSD = Caucasian Shepherd; ASD = Anatolian Shepherd; Ak = Akbash; T = Turkmenistan Shepherd; SK = Sage Kuchi; CAM = Central Asian Shepherd; MT = Tibetan Mastiff; SD = Stray Dogs; GW = Grey Wolf.

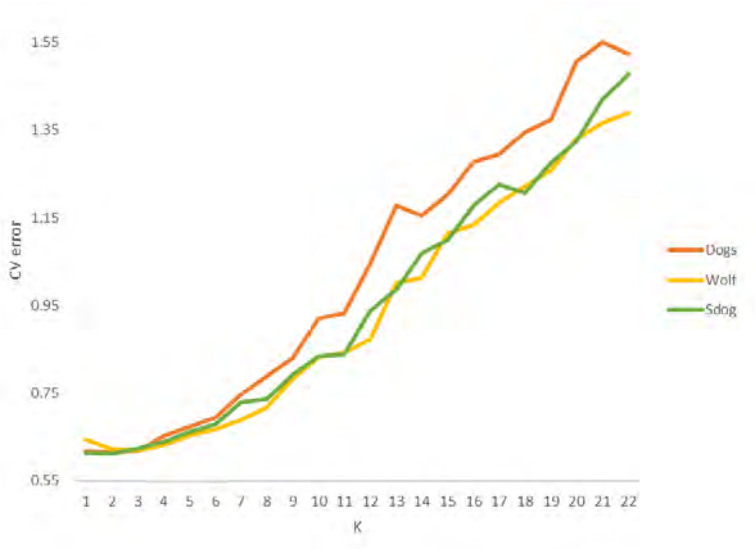


Fig. 3.5 – Cross-validation plot for the three preliminary analysis with Admixture. Yellow – grey wolves as an outgroup in the dataset; Green – stray dogs as an outgroup in the dataset; Orange – only LGDs and no outgroup in the dataset.

3.3.2.2. Results for the final analysis

In the ADMIXTURE analysis (Fig. 3.6), for K=2, which was the best partition found in the analysis (Supplementary Fig. 5), an isolation by distance phenomenon seems to be occurring. In this, the genetic signal that occurs almost completely in the Iberian breeds seems to be fading in the European breeds, almost completely lost the Asian breeds.

In this analysis 22 dog breeds were considered but, because in six of those breeds there is only one individual in the dataset and these types of software's do not create cluster with only one individual, 16 would be the maximum K for which ADMIXTURE could be capable of allocating each breed (with 2 or more individuals) in a private cluster. Although, for K=16, the Asian breeds can not be separated, forming instead a cohesive cluster throughout the analysis. The only exceptions to this are the Kurdish dog and the Tibetan Mastiff that are allocated to their own cluster as soon as K=5 and K=14, respectively. Throughout the analysis, the sofware seems to more easily forms clusters with individuals within certain breeds, pointing the Castro Laboreiro dog, Rafeiro of Alentejo or the Maremma Sheepdog, than allocate the Asian breeds to private clusters.

The Iberian breeds are the first one to be allocated to their own cluster, starting with the Spanish Mastiff in K=3, followed by the Castro Laboreiro dog, Estrela Mountain dog and the Rafeiro of Alentejo. In the Iberian group, the only breed that is never assigned to its own cluster is the Transmontano Mastiff, which seems to be a mixture of other Iberian breeds, like the Spanish Mastiff and Estrela Mountain dog.

Other breeds that seem to be a mixture of other breeds include the Aidi, the Maremma Sheepdog and the Molossus of Epirus. Also, the individuals from both the Yugoslavian Shepherd and Maremma Sheepdog breeds are being assigned to different clusters throughout the analysis. Actually, except for the Kurdish dog, the breeds seem they are being allocated to their own cluster in a gradient that starts with Iberian breeds towards the eastern breeds.

In the sub-dataset analysis with ADMIXTURE (Fig. 3.7), the whole dataset was divided in three different sub-datasets, according to their separation in K=2, as previously pointed. For all the sub-datasets, regarding the cross-validation errors, the best K found was 1 (Supplementary Fig. 5).

In the Iberian group, the Spanish Mastiff was again the first breeds to be allocated to its own cluster, followed by the Castro Laboreiro and Estrela Mountain dog and last the Rafeiro of Alentejo. Again, the Transmontano Mastiff could not be allocated to a private cluster, showing admixture signs with all the other breeds in the Iberian Peninsula. Interestingly, certain individuals of the Rafeiro of Alentejo breeds also seem to have admixture with the Estrela Mountain dog, Spanish Mastiff and Castro Laboreiro breeds.

Regarding the European group, the breeds seem to be divided in two clusters, one comprising the breeds from Central Europe and the other with the Eastern Europe breeds. Once again, some individuals of the Maremma Sheepdog are assigned to their own cluster, with the remaining being a mixture of other breeds. A similar phenomenon happens with the Yugoslavian Shepherd with a third of the individuals of this breed sharing a cluster with the Komondor breeds.

About the Asian group, the Kurdish dog was the first breed to be assigned to a private cluster, although two individuals share some admixture with other Asian breeds throughout the whole analysis. Although ADMIXTURE continues to be unable to assign all the breeds to a private cluster of their own, new signs of structure were found in this analysis. One Turkmenistan Shepherd shares a cluster with the Caucasian Shepherd breed, the Sage Kuchi individuals show some admixture with other breeds, including the Central Asian Shepherd and the Anatolian Shepherd and the later dogs share the same cluster. The Tibetan Mastiff appears to be the only breed that has a private cluster well defined for all the individuals with no signs of admixture with any other breed.

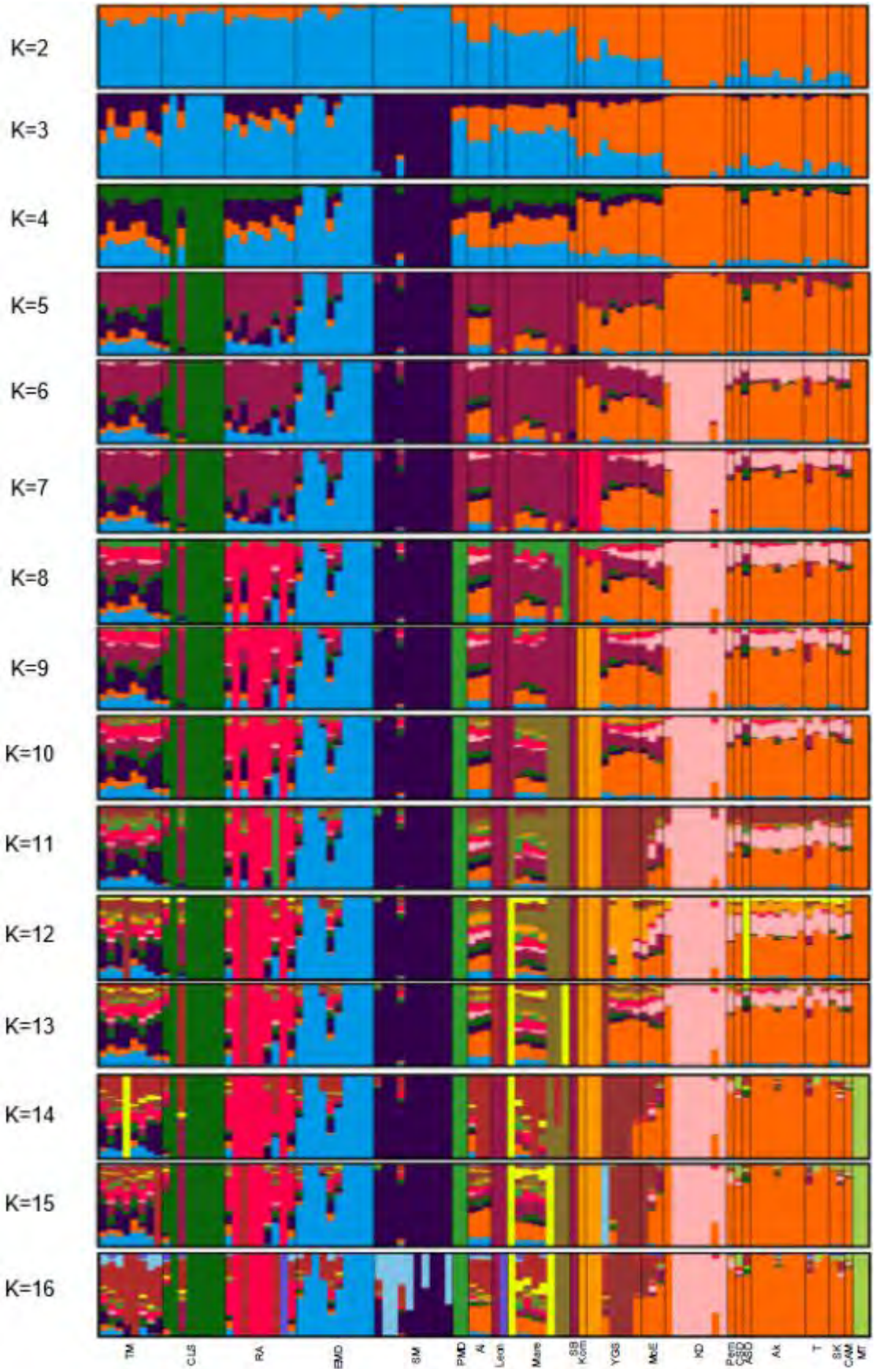


Fig. 3. 6 – Proportion of membership of 98 individuals from 22 dog breeds for K=2 to 16, as calculated by ADMIXTURE software. Multiple runs for each K were concatenated using ClumpaK, and distruct was used to generate images. TM = Transmontano Mastiff; CLS = Castro Laboreiro dog; RA = Rafeiro of Alentejo; EMD = Estrela Mountain Dog; SM = Spanish Mastiff; PMD = Pyrenean Mountain Dog; Ai = Aidi; Leon = Leonberger; Mare = Maremma Sheepdog; SB = St. Bernard; Kom = Komondor; YGS = Yugoslavian Shepherd; MoE = Molossus of Epirus; KD = Kurdish dog; PeM = Persian Mastiff; CSD = Caucasian Shepherd; ASD = Anatolian Shepherd; Ak = Akbash; T = Turkmenistan Shepherd; SK = Sage Kuchi; CAM = Central Asian Mastiff; MT = Tibetan Mastiff.

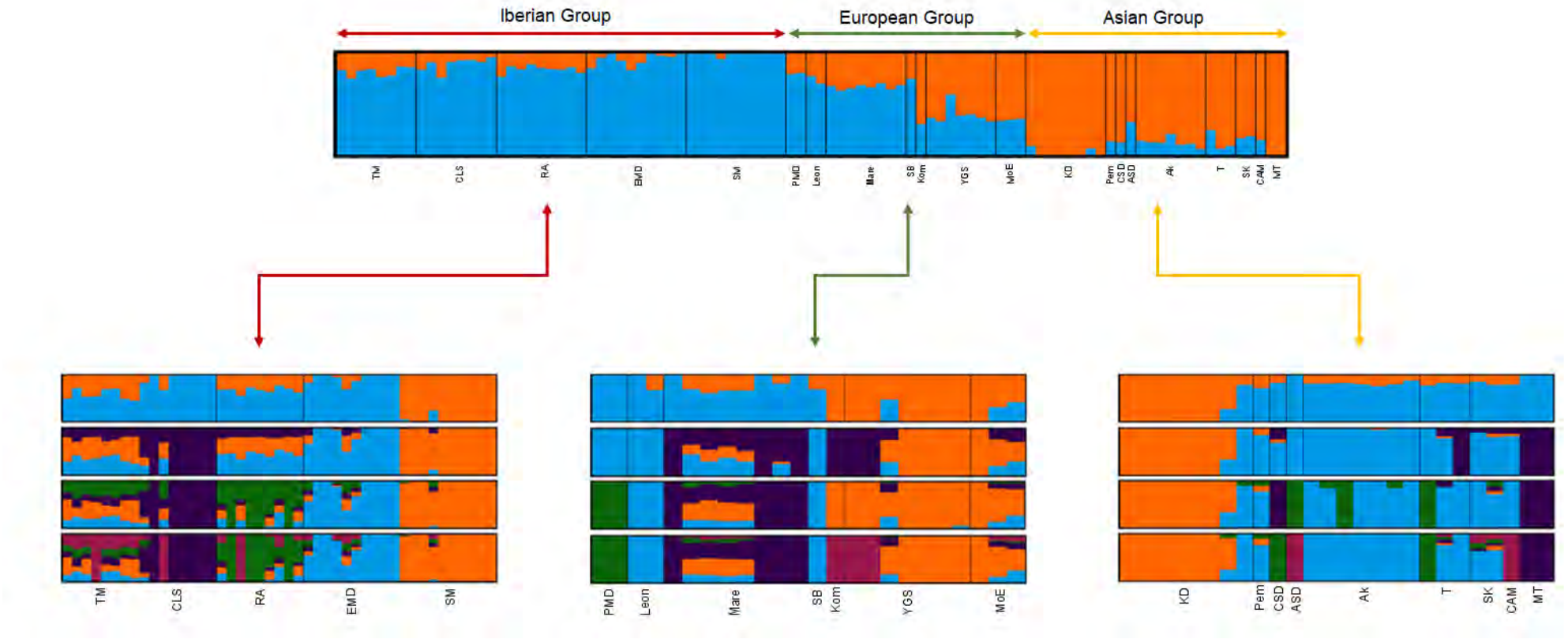


Fig. 3.7 – Sub-dataset analysis, in which the three separate sub-datasets were considered (following the separation observed in K=2 form the overall analysis) from K=2 to K=5 (top to bottom), of 93 individuals from 21 dog breeds, as calculated by ADMIXTURE software. Multiple runs for each K were concatenated using ClumpaK, and distruct was used to generate images. TM = Transmontano Mastiff; CLS = Castro Laboreiro dog; RA = Rafeiro of Alentejo; EMD = Estrela Mountain Dog; SM = Spanish Mastiff; PMD = Pyrenean Mountain Dog; Leon = Leonberger; Mare = Maremma Sheepdog; SB = St. Bernard; Kom = Komondor;YGS = Yugoslavian Shepherd; MoE = Molossus of Epirus; KD = Kurdish dog; PeM = Persian Mastiff; CSD = Caucasian Shepherd; ASD = Anatolian Shepherd; Ak = Akbash; T = Turkmenistan Shepherd; SK = Sage Kuchi; CAM = Central Asian Shepherd; MT = Tibetan Mastiff.

3.3.3. Phylogenetic analysis

The final consensus neighbour-joining tree of 1000 bootstraps, with the Grey Wolf as root, included almost every individual grouped within their breed and positioned were it would be expected (Fig. 3.8). However, the phylogenetic placement of the Iberian breeds showed some surprising results. Individuals from the Rafeiro of Alentejo, Transmontano Mastiff and Castro Laboreiro breeds are not all assigned within their breeds. The Transmontano Mastiffs does not form a single clade, instead, it appears both in the clade of the Spanish Mastiff and in the clade of the Castro Laboreiro dog. Most of Rafeiro of Alentejo individuals appear in a single clade next to the Estrela Mountain dog, but some individuals are clustered together with Transmontano Mastiffs in the Spanish Mastiff clade. Castro Laboreiro yielded another surprising result, appearing in the bottom of the Iberian clade and having one individual assigned in the European clade, with the Leonbergers, along with the St. Bernard.

The European clade is overall as expected, although the Maremma sheepdogs appeared separated in two different clades with Pyrenean Mountain dogs in the middle. Besides, the African Aidi breed is also assigned within the European clade.

The Asian group is rooted with the Tibetan Mastiff that follows the Grey Wolves in the bottom of the tree, followed by the Caucasian Shepherd. Two of the three Turkmenistan Shepherd are assigned together, with the third individual of this breed assigned with one Yugoslavian Shepherd, although with low support. The Central Asian Shepherd and the Persian Mastiff are assigned with the Sage Kuchi and the Akbash, respectively. This could be the fault of these breeds only having one individual, although the Anatolian Shepherd is assigned in a single clade.

Concerning the maximum likelihood tree obtained with Treemix (Fig. 3.9), it appears to be similar to the neighbour-joining tree. Almost every breed is assigned to the group where it was supposed to be, with the exception of Castro Laboreiro dog. This breed, instead of being assigned with the rest of the Iberian breeds is assigned within the European cluster, with the Leonberger and St. Bernard. Curiously it is positioned like one individual of the same breed in the previous tree. Also, in contrast with the neighbour-joining tree, the Pyrenean Mountain dog is not clustered with the Maremma Sheepdog, but instead it seems to be the basal breed of the Iberian cluster.

At the Asian group, Tibetan Mastiff is still the first breed to appear followed by the Caucasian Shepherd and the Central Asian Shepherd. Interestingly the Central Asian do not share the same clade as Sage Kuchi, which happened at the previous tree, but the Persian Mastiff and the Akbash dog still remain in the same clade.

In both trees, almost every single breed seems to be assigned according to their geographic origin, starting always with the Tibetan Mastiff in Central Asia and ending with the Spanish Mastiff, in the Iberian Peninsula.

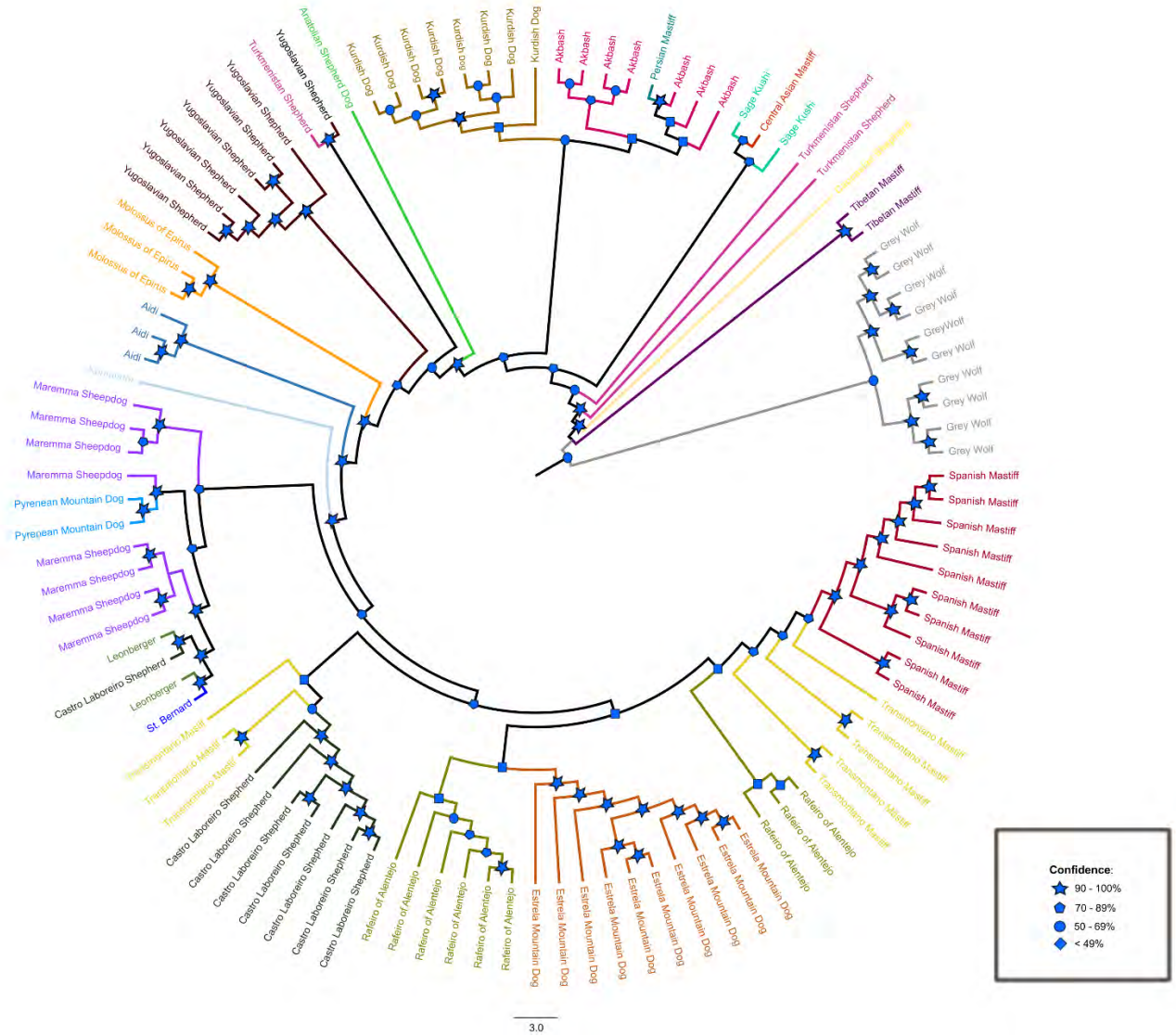


Fig. 3.8 – Neighbour-joining tree of Livestock guarding dogs compared to the grey wolf, as root. Different colours were used to differentiate among dog breeds.

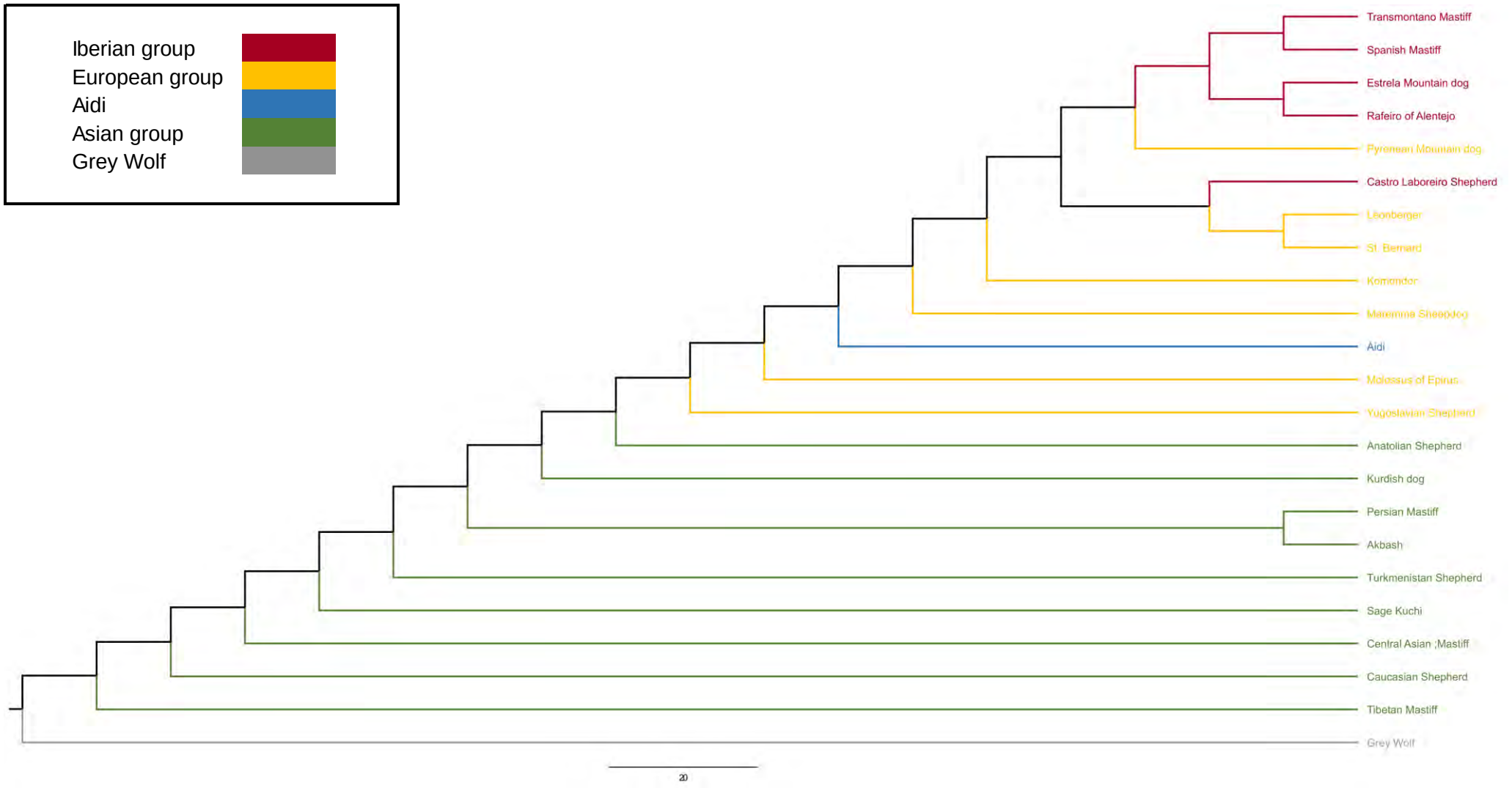


Fig. 3.9 – Maximum likelihood cladogram obtained by Treemix, with 0 migration events. Livestock dog breeds are differentiated into three different clades, with the exception of Aidi and the Grey Wolf, which was used to root the tree. Different colours were used to differentiate the different clades.

3.3.4. Migration events estimation

3.3.4.1. Results for the preliminary analysis

Analysing the outputs obtained by Treemix it was found out that the outputs were biologically and geographically incoherent because certain migratory events that were inferred, were very unlikely to have happened (eg. Migration event from the Leonberger to the Grey Wolf). Also, although the migration model with the lowest absolute residual error is generally the best, it could not be found a number of migrations for which the residuals were consistent. Thus, the stepping stone model was created, and the dataset was divided into 5 sub-datasets to try to overcome this problem.

As previous pointed, the "-noss" function was implemented to try to improve the quality of the analysis but the difference with this function seemed to be not significant (Fig. 3.10) for each of the five sub-datasets that were created.

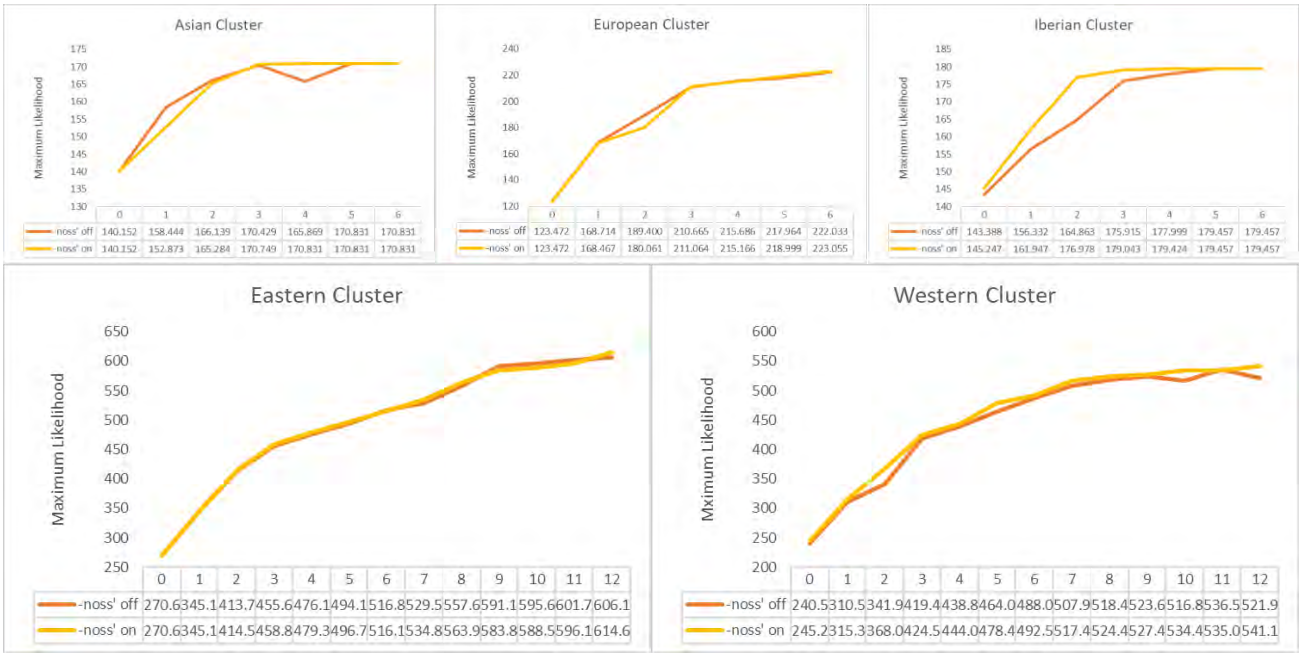


Fig. 3.10 – Maximum likelihood plot of Treemix analysis for migratory events from 0 to 6 (above) and 0 to 12 (down). Comparison of using the '-noss' function in the stepping-stone analysis.

3.3.4.2. Results for the final analysis

Several migration events were observed in the TreeMix results concerning the stepping-stone model (Fig. 3.11 and 3.12). The grey wolf was set as the root of the ancestry model. Graphics of residuals were inferred to all the groups (Supplementary Fig. 6 to 10). Residuals above zero represent populations that are more closely related to each other in the data than in the best-fit tree and are

candidates for admixture. Negative residuals indicate that a pair of populations are less closely related, based on the data, than represented in the best-fit tree.

Regarding the tree for the Asian group (Fig. 3.11A) three migrations were found. Considering the residuals (Supplementary Fig. 6), a migration between Akbash and Tibetan Mastiff is likely to be missing in this model. About the European tree (Fig. 3.11B), 4 migrations were identified, all of them with significant p-values. There seemed to have been two migrations from the Grey wolf, one to the Aidi and another to the Leonberger, a migration from the Molossus of Epirus to the ancestral of Pyrenean Mountain and Maremma sheepdog and a migration, with a high migration weight, from the Maremma Sheepdog to the Yugoslavian Shepherd. The residuals show that other migration events may have happened (Supplementary Fig. 8). The Iberian tree (Fig. 3.11C) depicted three migration events where only, from Rafeiro of Alentejo to Spanish Mastiff, had a significant p-value. Migratory events that might be missing, according to the residuals (Supplementary Fig. 10), are between the Estrela Mountain dog and the Grey Wolf and between Rafeiro of Alentejo and Castro Laboreiro.

Considering the East group tree (Fig. 3.12A), 9 migratory events were identified. Of those 7, 5 had significant p-values and one migration, from the Grey Wolf to the Leonberger, was previously seen in the European tree. The other migrations include 3, with high migration weight, from the Akbash dog and the ancestral of the Pyrenean Mountain and Maremma Sheepdog to the Sage Kuchi and from Pyrenean Mountain to The Molossus of Epirus. Interestingly, the European tree showed a reverse migration to this last one. The residuals (Supplementary Fig. 7) identified migrations between the Pyrenean Mountain/Aidi and Molossus of Epirus/Tibetan Mastiff to be probably missing.

In the West group tree (Fig. 3.12B), 7 migratory events were found, all with significant p-values. Curiously, the migration found between Leonberger and Grey Wolf is seen again, but this time the direction of the migration was inverted, going from the Leonberger to the Grey Wolf. The migrations from the Grey Wolf and Rafeiro of Alentejo to the Spanish Mastiff are depicted again, as in the Iberian tree, and the migration from Transmontano Mastiff to the Castro Laboreiro dog that was seen in the Iberian tree happens again in this tree. There is also two never seen migrations, both with high migration weights, goinng from the Maremma Sheepdog to the Pyrenean Mountain dog and from the Grey Wolf to Molossus of Epirus. A strong migration between Pyrenean Mountain/Molossus of Epirus is likely missing, regarding the residuals (Supplementary Fig. 9) for this model.

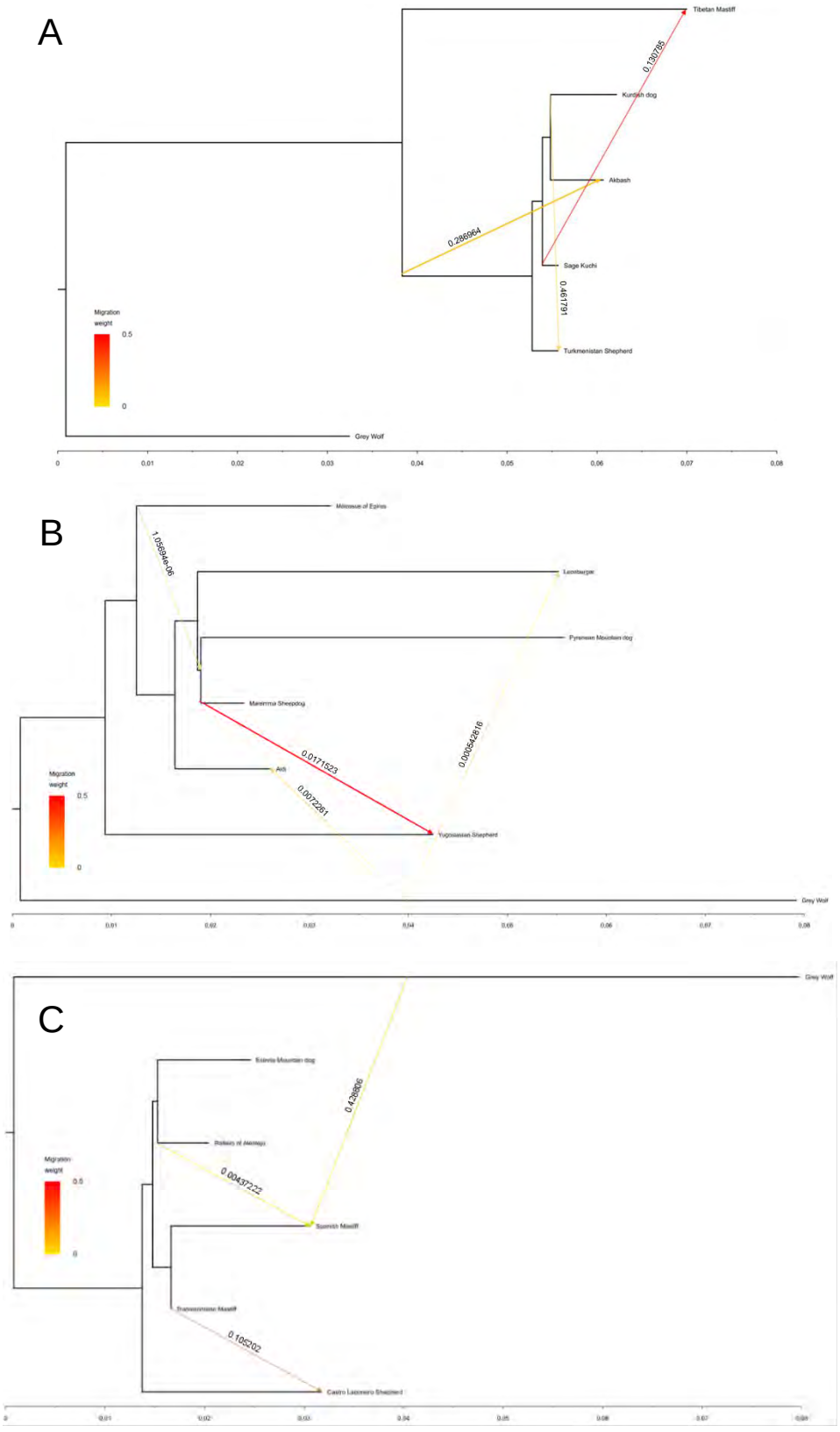


Fig. 3.11 – Maximum likelihood tree with migration events, calculated for the stepping-stone model, using the Grey wolf as root for all trees. From top to bottom: A - Asian group with 3 migratory events; B - European group with 4 migratory events; C - Iberian group with 4 migratory events. Admixture boundaries are denoted with arrows in the direction from the migrant's origin to the recipient breed and heat coloured according to the mixture percentage. P-value for each migratory event next to the arrow. Horizontal branch lengths are proportional to the amount of genetic drift that has occurred along that branch.

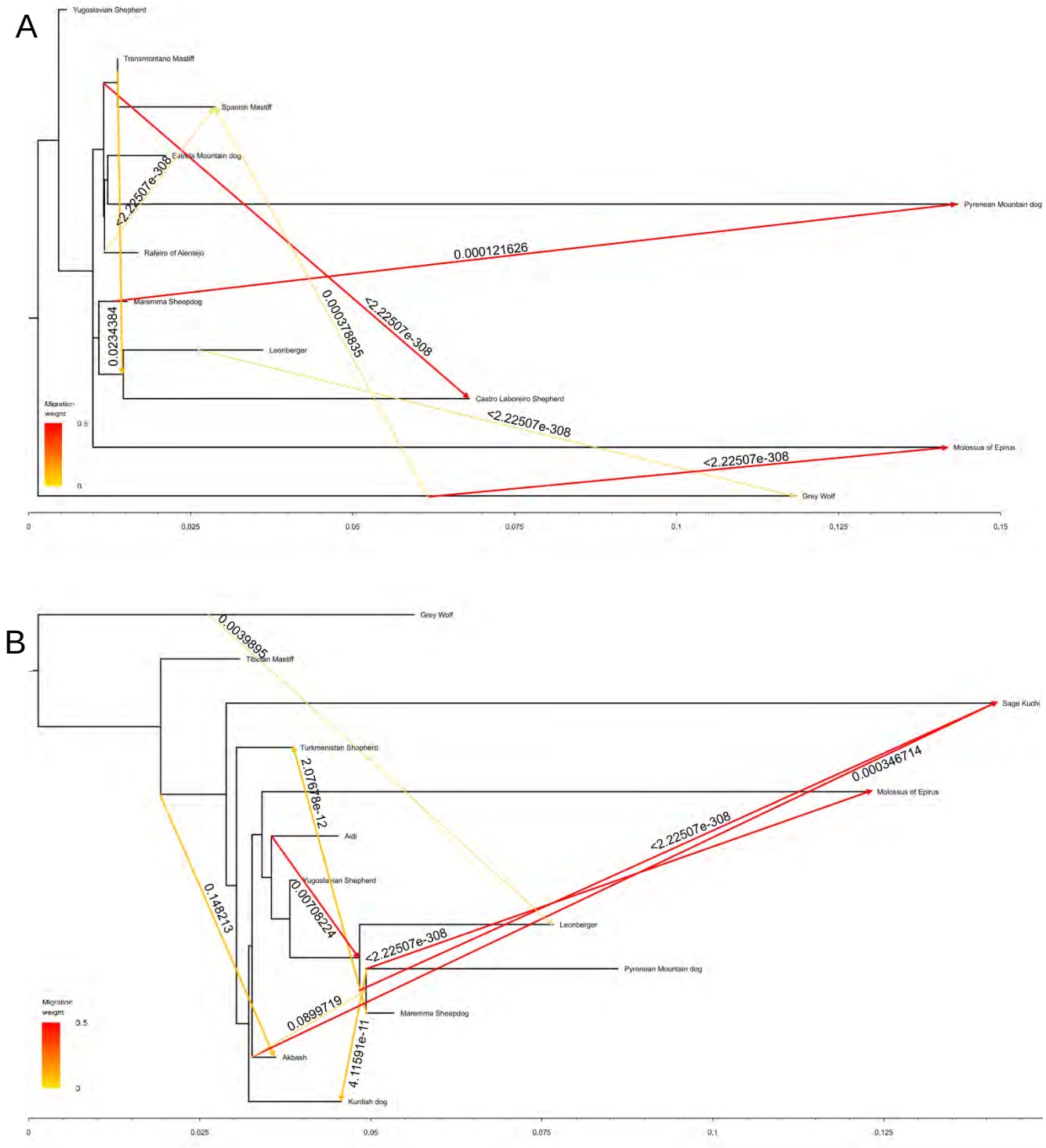


Fig. 3.12 – Maximum likelihood tree with migration events, using the Grey wolf as root for all trees. From top to bottom: A - Western group with 7 migratory events; B - Eastern group with 9 migratory events. Admixture boundaries are denoted with arrows in the direction from the migrant's origin to the recipient breed and heat coloured according to the mixture percentage. P-value for each migratory event next to the arrow. Horizontal branch lengths are proportional to the amount of genetic drift that has occurred along that branch.

About the consensus tree (Fig. 3.14) generated using the migratory events inferred in the stepping-stone model, 15 migrations were considered, 4 of which are going from the Grey Wolf to Molossus of Epirus, Spanish Mastiff, Leonberger and Aidi. Of the considered migrations events, 8 of them had a high migration weight. Also, it is important to notice that only migrations with significant p-values were considered when constructing this tree.

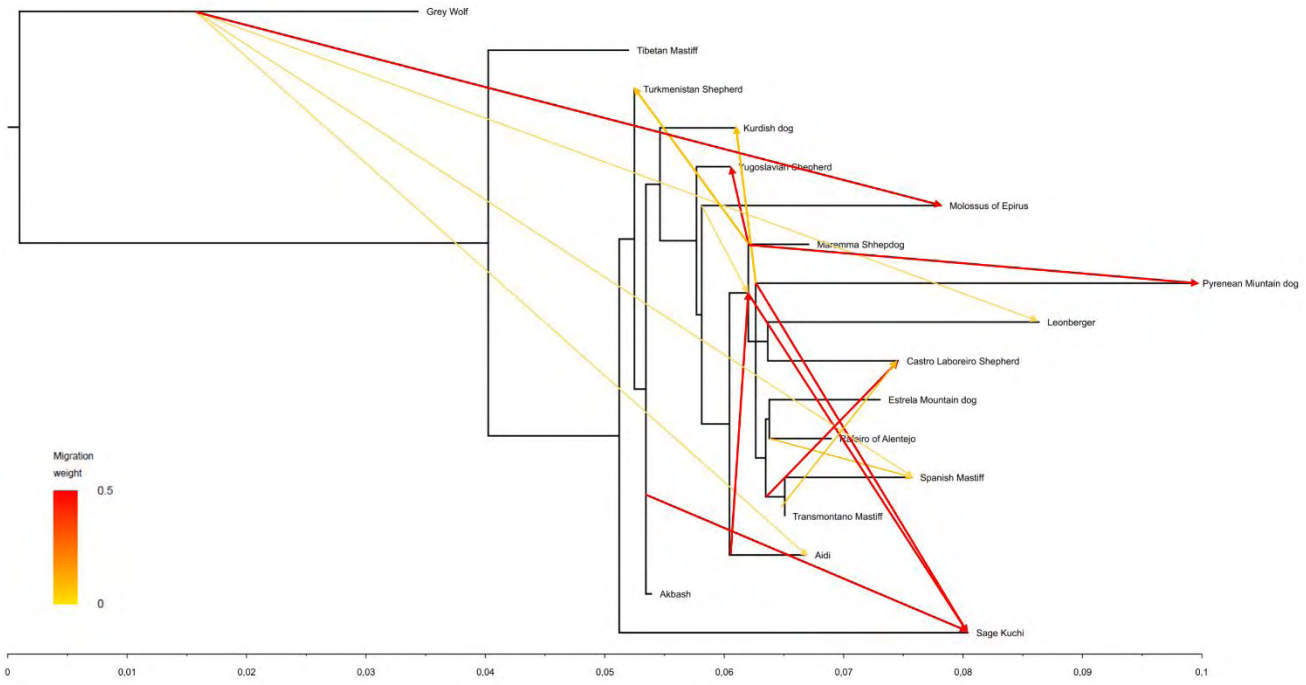


Fig. 3.13 – Maximum likelihood tree with 15 migration events, using the Grey wolf as root. Consensus migratory events inferred in each tree from the stepping-stone model. Admixture boundaries are denoted with arrows in the direction from the migrant's origin to the recipient breed and heat coloured according to the mixture percentage. Horizontal branch lengths are proportional to the amount of genetic drift that has occurred along that branch.

Regarding the tree obtained with Treemix, using the whole dataset (Fig. 3.15), 11 migration events were obtained, 2 of which had a extremely high migration weight, and 4 others having almost as high weight also. Among Iberian breeds only one migratory event was found, from the Spanish Mastiff to Rafeiro of Alentejo, though it did not have a significant p-value. The migration found in all the trees before, between the Grey Wolf and Leonberger, was also depicted here and also with a low weight. About the Moroccan breed Aidi, a migration was found between this breed and the closest ancestor of the Kurdish dog. A migration, with high weight, was also found coming from the Turkmenistan Shepherd to the Yugoslavian Shepherd.

Regarding the residuals for this model (Fig. 3.16), some migrations might be missing, pointing migratory events between the Grey Wolf and the Spanish Mastiff or between Molossus of Epirus and Kurdish dog.

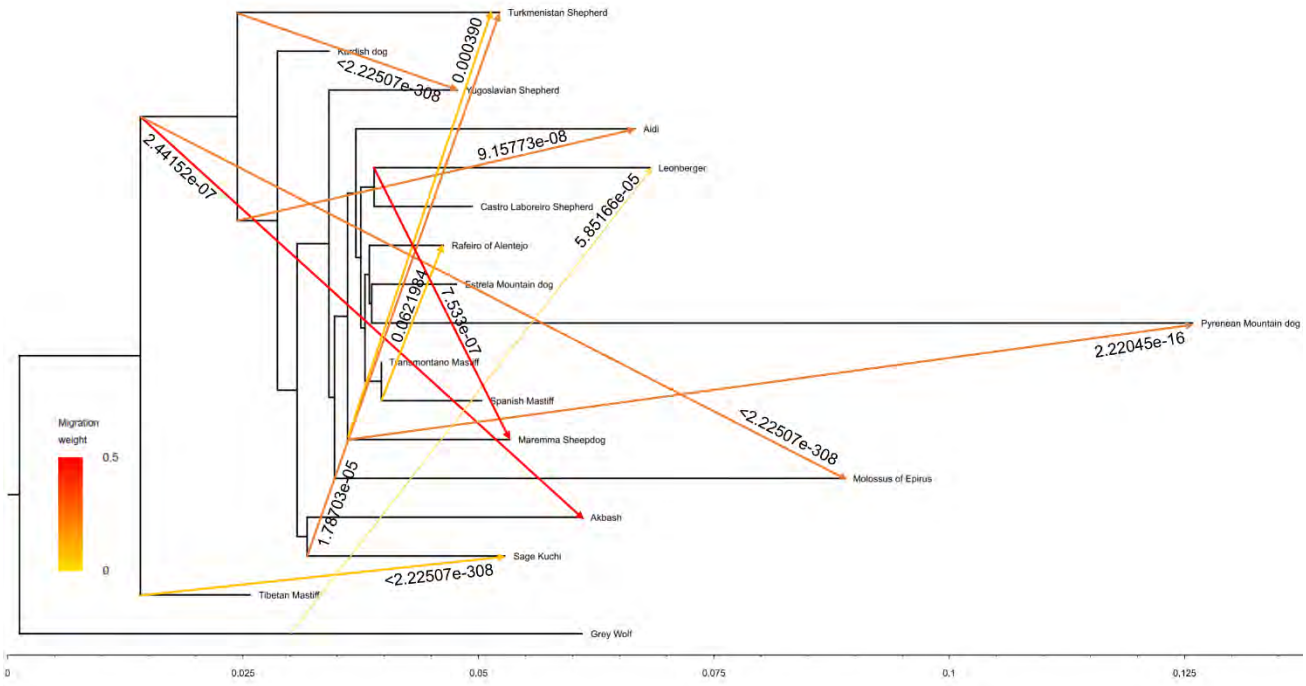


Fig. 3.14 – Maximum likelihood tree with migration events, using the Grey wolf as root for all trees, considering the whole dataset. Admixture boundaries are denoted with arrows in the direction from the migrant's origin to the recipient breed and heat coloured according to the mixture percentage. P-value for each migratory event next to the arrow. Horizontal branch lengths are proportional to the amount of genetic drift that has occurred along that branch.

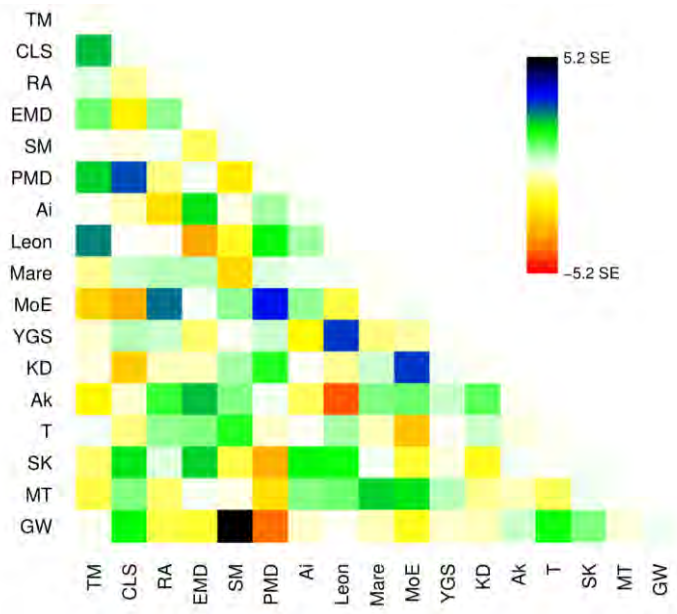


Fig. 3.15 – Residual fit of the observed versus predicted squared allele frequency difference, expressed as the number of s.e. of the deviation. Colours are in the palette on the right. TM = Transmontano Mastiff; CLS = Castro Laboreiro dog; RA = Rafeiro of Alentejo; EMD = Estrela Mountain Dog; SM = Spanish Mastiff; PMD = Pyrenean Mountain Dog; Ai = Aidi; Leon = Leonberger; Mare = Maremma Sheepdog; SB = St. Bernard; YGS = Yugoslavian Shepherd; MoE = Molossus of Epirus; KD = Kurdish dog; PeM = Persian Mastiff; CSD = Caucasian Shepherd; ASD = Anatolian Shepherd; Ak = Akbash; T = Turkmenistan Shepherd; SK = Sage Kuchi; CAM = Central Asian Shepherd; MT = Tibetan Mastiff; GW = Grey Wolf.

The f3 statistics were generated to trace the possible ancestry mixtures in Livestock Dog breeds and the gray wolf. A set of the significant f3 statistics (standardized Z score <-2) is shown in Table 3.4. Akbash, Sage Kuchi, Transmontano Mastiff and Turkmenistan Shepherd were the only breeds that appeared to have had significant admixture events with other breeds.

Table 3.4 – The most significant f3 statistics (Z<-2) shown the possible ancestor mixture of Livestock guarding dogs and Grey wolves.

Population A	Population B	Population C	*F3* statistic	Standard error	Z-score
Akbash	Pyrenean Mountain dog	Tibetan Mastiff	-0.00060	0.00028	-2.17905
Akbash	Grey Wolf	Maremma Sheepdog	-0.00057	0.00020	-2.87561
Akbash	Grey Wolf	Rafeiro of Alentejo	-0.00072	0.00020	-3.56553
Akbash	Grey Wolf	Pyrenean Mountain dog	-0.00089	0.00032	-2.80193
Akbash	Grey Wolf	Estrela Mountain dog	-0.00082	0.00021	-3.92291
Akbash	Tibetan Mastiff	Rafeiro of Alentejo	-0.00060	0.00018	-3.36119
Akbash	Tibetan Mastiff	Spanish Mastiff	-0.00045	0.00019	-2.36635
Akbash	Estrela Mountain dog	Tibetan Mastiff	-0.00057	0.00018	-3.11949
Sage Kuchi	Tibetan Mastiff	Spanish Mastiff	-0.00066	0.00032	-2.06471
Sage Kuchi	Tibetan Mastiff	Transmontano Mastiff	-0.00082	0.00030	-2.70936
Sage Kuchi	Tibetan Mastiff	Rafeiro of Alentejo	-0.00092	0.00031	-2.98615
Sage Kuchi	Estrela Mountain dog	Tibetan Mastiff	-0.00100	0.00031	-3.21901
Transmontano Mastiff	Spanish Mastiff	Maremma Sheepdog	-0.00044	0.00010	-4.20308
Transmontano Mastiff	Aidi	Spanish Mastiff	-0.00027	0.00013	-2.06011
Transmontano Mastiff	Leonberger	Spanish Mastiff	-0.00069	0.00016	-4.36248
Transmontano Mastiff	Pyrenean Mountain dog	Spanish Mastiff	-0.00069	0.00017	-4.13307
Transmontano Mastiff	Estrela Mountain dog	Spanish Mastiff	-0.00048	0.00010	-4.63463
Transmontano Mastiff	Yugoslavian Shepherd	Spanish Mastiff	-0.00022	0.00011	-2.00458
Transmontano Mastiff	Spanish Mastiff	Castro Laboreiro dog	-0.00050	0.00011	-4.55407
Turkmnenistan Shepherd	Tibetan Mastiff	Transmontano Mastiff	-0.00054	0.00024	-2.25052
Turkmnenistan Shepherd	Tibetan Mastiff	Rafeiro of Alentejo	-0.00059	0.00025	-2.3665
Turkmnenistan Shepherd	Tibetan Mastiff	Spanish Mastiff	-0.00057	0.00026	-2.22018
Turkmnenistan Shepherd	Tibetan Mastiff	Maremma Sheepdog	-0.00053	0.00024	-2.16302
Turkmnenistan Shepherd	Grey Wolf	Estrela Mountain dog	-0.00092	0.00028	-3.34825
Turkmnenistan Shepherd	Grey Wolf	Pyrenean Mountain dog	-0.00122	0.00037	-3.26666
Turkmnenistan Shepherd	Grey Wolf	Transmontano Mastiff	-0.00071	0.00026	-2.70100
Turkmnenistan Shepherd	Grey Wolf	Maremma Sheepdog	-0.00110	0.00027	-4.12255
Turkmnenistan Shepherd	Grey Wolf	Rafeiro of Alentejo	-0.00091	0.00027	-3.38772
Turkmnenistan Shepherd	Pyrenean Mountain dog	Tibetan Mastiff	-0.00073	0.00034	-2.16315

3.5. Discussion

3.5.1. Patterns of genetic diversity

Diversity levels described for domestic dogs, using SNPs, are typically around 0.30 with Asian dogs being the ones that display higher levels[51,114]. The observed heterozygosity here presented ranged from 0.307 to 0.382 in the Maremma Sheepdog and in the Kurdish dog, respectively, which are slightly higher than what was previously described for the domestic dog[28,38,48,173,183]. The Asian dogs, like the Akbash and Kurdish dog, displayed the highest observed heterozygosity although, the differences when compared to western breeds are not significant, with Iberian dogs, like Castro Laboreiro and Estrela Mountain dog, displaying levels of diversity almost as high as Asian dogs. While these values are in concordance with published works, those studies often consider only modern breeds that after years of selectiveness and consequent bottlenecks to produce specific phenotypes are now displaying low levels of heterozygosity[51]. Thus, in livestock guarding dogs, which are not exposed to high selection pressures, the low levels of diversity seen in the Maremma Sheepdog may be caused by a recent increasing of inbreeding.

The inbreeding coefficient (F) is a measure used by dog associations and is frequently associated with breed pureness, with values of F higher than 0.10 found in pure-breed dog populations[46,47,115]. Only two of the breeds here considered had an inbreeding coefficient expected to be of a pure breed dog, although in Rafeiro of Alentejo the coefficient is only slightly above the common pure-breed value. Considering both the heterozygosity and the inbreeding coefficient, it is possible that, as a result of the breed standard that enforces breed pureness, these breeds are already displaying losses in their genetic pool.

Most of the other breeds revealed extremely low levels of inbreeding and are in concordance with the literature[13,50,153,170] which defends that LGDs kept with their herds are not exposed to selectiveness to physical traits.

3.5.2. Patterns of genetic structure

When the AMOVA was performed to identify variation between the three considered groups it was perceived that the highest variation source resides among the individuals within the breeds with only 1% of differentiation explained by the grouping of the breeds. This was expected since dog populations are not wild and are not confined in one place by physical barriers. Although between central Europe and Iberia exists a mountain range, the Pyrenees, the Pyrenean Mountain dog can be found in both sides of the mountains, for example. Also, dogs travel with humans so the geographical

limitations that they face are almost none. Alexander, the Great brought dogs from the middle east to Europe, the Phoenicians traded dogs across the Mediterranean, reaching all its coast and the Iberian Peninsula, and the Romans helped diffusing the livestock dogs across Eurasia during the conquests[13]. Also, the differentiation calculated using F_{ST} presented very low levels, reinforcing the idea that these breeds should be considered as a whole and are not very differentiated in a genomic point of view.

The results from ADMIXTURE identified 2 as the best partition model for the LGDs showing a change in the dominant genetic signal present in the Asian breeds as dogs move west until it reaches the Iberian Peninsula where this signal is almost lost. This progressive “loss” in diversity, although tenuous considering the observed heterozygosity, may be cause of an “isolation by distance” phenomenon where, as the breeds accompanied human migrations while the latter species conquered Europe[13], they got more and more differentiated from the original stock.

Considering the Asian breeds, in the ADMIXTURE analysis, only one of them can be allocated in private clusters. The Tibetan Mastiff is widely considered as the oldest extant LGD, with the approximate divergence time between Tibetan Mastiff and grey wolf nearly 20 thousand years before the divergence between other domestic dogs from the wolves[184–186]. In the analysis, it is the only breed that fully displays the genetic signal that seems to be characteristic of the Asian breeds although at K=14 it is assigned to its own cluster. According to studies using the mitogenome, Tibetan Mastiff diverged from the Grey Wolf at 58,000 years ago and it was introduced to the west during the Mongolian conquests where it played an important role in the breed formation of other dogs[53]. There is a genetic relationship between this breed and others like the Pyrenean Mountain, Saint Bernard and the Leonberger and that these large breed dogs are probably partially descended from the Tibetan Mastiff[184,186]. Since Tibet closed its doors for centuries to Westerners, this breed developed for centuries in relative isolation[186], which could explain why there is enough differentiation to segregate the Tibetan Mastiff from the other Asian breeds.

The Kurdish dog descended from the Molossers of Central Asia, and while some say the Kurdish Shepherd Dog is a result of crossing original Kangals and Yoruk Sheepdogs with the Kurdish Greyhound centuries ago, others believe that the pictures of inscriptions made from the Babylonian and Assyrian civilizations are not from Kangal dogs but from Kurdish making this breed at least 5000 years old[76]. Whether these theories are true or false the Kurdish dog was assigned to a private cluster since the beginning of the analysis, appearing to be the most differentiated Asian breed.

About the European breeds, it can be stated that there is a certain separation between the breeds from Eastern Europe (Balkans and old Yugoslavia) and Central Europe since the former breeds show high levels of admixture with Asian breeds. This would be expected because there is no geographical barrier between Eastern Europe and Turkey and there are centuries of commerce between these regions with the possibility of dogs being traded as well.

The Pyrenean Mountain dogs appears to be the most differentiated breed from this group, being assigned to its own cluster early in the analysis. Previous works found this breeds' ancestry in the dogs of Asia, but others also believe that this breeds' primary ancestor is the Maremma Sheepdog, descended from the ancient Molossers of the Balkans, Caucasus and Central Asia[50]. Due to the relative isolation of the Pyrenees, it is believed that the Pyrenean Mountain dog has been isolated for at least 1500 years, leading to the differentiation found.

The Maremma Sheepdog was until recently (1958) two different breeds: the Pastore Maremmano and the Pastore Abruzzese, but the breeds were unified by the national dog association of Italy (ENCI). This was justified by their phenotypical similarities and "natural fusion" of the two types as a result of transhumance, particularly after the unification of Italy[187]. The results from ADMIXTURE show that there are two types of Maremma Sheepdog, which it is not surprising since the recent breed fusion does not override the genetic patterns originated by breed history.

Concerning the Iberian breeds, the Spanish Mastiff was the first breed assigned to a single cluster being separated from the four Portuguese breeds at K=2 in the Iberian analysis. This could be the result of isolation caused by country frontiers, although the Transhumance migrations, where both Portuguese and Spanish breeds came together, only stopped recently[44,188] and the results from Chapter 2 suport that there is no strong differentiations between Portuguese and Spanish dogs.

Rafeiro of Alentejo shows traces of admixture with other breeds, particularly the Estrela Mountain dog and Spanish Mastiff, due either to the Transhumance movements or the popular belief that these breeds originated from crossings between the former breeds. As mentioned in Chapter 2, Transmontano Mastiff is a variant of a the result of a Rafeiro of Alentejo that was admixted with short-haired Estrela Mountain dogs[167,168] and its registry was only established in 2004 based on 170 founders[125]. Results here presented also support that Transmontano Mastiff is a mixture of other Iberian dogs and has been been bred with the Spanish Mastiff and with the Estrela Mountain dog[167].

3.5.3. Phylogeny and migration events

3.5.3.1. Treemix limitations

The TreeMix model makes several assumptions about the process of population splits and gene flow. When these assumptions are violated, the reliability of the results created by Treemix may be reduced and the output may be even unrealistic. The first assumption that TreeMix does is that the migration occurred at a single point in time, thus events of on-going long-term gene flow may result in a poorly supported graph[189]. For example, transhumance movements that occurred in the Iberian Peninsula until recently could represent a fixed low level of migration between the autochthonous dog breeds and influence the topology of the maximum likelihood tree. Second, TreeMix also assumes that the history of the populations is basically tree like, but in cases of complex structure, many graphs representing different histories may be equally supported by identical covariance matrices. As previously pointed, some dog breeds may have been created through both processes of differentiation and through mixture with other breeds.

As pointed in the analysis methods, to supress the limitations of the software some preliminary analysis was done (see Preliminary results) with different datasets and using different functions of Treemix. Although dividing the whole dataset in five sub-datasets made an improvement in the quality of the results, some migratory events inferred by Treemix still were highly doubtful, mainly in the direction that the migratory event was taking. When describing the software's method for the first time, Pickerel and Pritchard[179] noted that its accuracy dropped considerably when the migratory events were happening between closely related populations. Also, the main forms of errors produced were the incorrect direction of migration arrows and inference of admixture in populations related to the truly admixed population.

Other problem faced was that many branches in the maximum likelihood tree had length of zero. According to the manual, this can happen in some cases, especially with single individuals in a population. These populations were removed, and it was also decided to turn off the sample size correction (adding the function "-noss") since, when TreeMix calculates the covariance matrix among populations, it includes a correction for sample size by default, which may lead to overcorrection. However, this correction seemed to not influence the output.

Both the results from the stepping stone model and for the whole dataset analysis were presented for comparison and to find the best outcome possible. These limitations of Treemix should always be of main concern and the results should always be evaluated considering the model restrictions and what makes or not biological sense.

3.4.5.2. Discussion

Dogs have been used to work on different purposes for ages and acquired large phenotypical and behaviour differences between breeds, although their purposes remain predictable across cultures. Attributes of sense acuity, guardianship, predatory nature and innate companionability with humans have been exploited for thousands of years[116,190–195] however, the pursuit of a distinct function ultimately relied on selection of breeding animals from a small source pool[196,197]. As such, dog breeds which perform analogous tasks are frequently closely related when compared to breeds with different occupations, being allocated in the same phylogenetic breed clades [47,117].

The results from the phylogenetic trees support the theory that the breeds originated from differentiation processes, that were caused by isolation. The breeds are well allocated in the neighbour-joining tree as the breeds' country of origin follows the same direction, from east to west, the exception being the Pyrenean Mountain dog. This breed is thought to have its main ancestry in the Maremma Sheepdog[50], both belonging to the same sub-unit of Molossers (white coat dogs), and is allocated inside the later breeds' clade in the neighbour-joining tree. Also, migration events were found between these breeds, originating in the Maremma Sheepdog. This introgression appears in both the stepping-stone model and in the whole dataset analysis, always with high migration weight. Another breed that shares strong bonds with the former breeds is the Yugoslavian Shepherd (although not having white coat)[50,167], with strong migrations being found between this breed and the Maremma Sheepdog.

Besides, those three breeds are known to descend from Asian dogs[50] and some migrations were found between them and Sage Kuchi, Kurdish dog and Turkmenistan Shepherd. Although the migration arrows point in the direction of the Asian breeds, that is highly questionable and knowing that Treemix tends to make mistakes in the migration arrow direction, one can suppose that these migration in fact happen but in the opposite direction that is, from the Asian to the European breeds.

One important migration happening is from the ancestral of all the breeds (except Tibetan Mastiff) to Molossus of Epirus which may support the "myth" that Alexander, The Great brought dogs from Asia to the Balkans[50]. These dogs probably gave birth to Molossus of Epirus which, in turn, took part in the formation of the others European breeds, through the Roman Conquests[1]. Also, a migration was found originating from Molossus of Epirus to the ancestor of Western European breeds.

A migration event originating from the grey wolf to the Leonberger appeared in both the stepping-stone and whole dataset analysis. Although being considered a Livestock guarding dog and displaying the same genetic signal, the Leonberger is not an ancient breed since it was created in the 19th century by breeding Newfoundland with St Bernard, with later admixture with Pyrenean Mountain dogs[77,198]. According to legend, the Leonberger was ostensibly bred as a "symbolic dog" that would mimic the lion in the town crest (Leonberg)[199]. Although there is no reference of crossing Leonbergers

with wolves, it is possible that after the breed almost getting extinct twice after both World Wars, breeders used grey wolves to enrich the dog's genetic pool.

The Eastern group seems to be not very well differentiated. In the Middle East, practices of nomadism are still recurrent[199,200], although nowadays more restricted because of country borders and war events thus, mattings between different breeds are likely to have occurred until recently. This belief is supported by migration events between Akbash, Sage Kuchi and the Turkmenistan Shepherd and F_{ST} value between the Kurdish dog and Akbash is among the lowest (0.0253).

The Tibetan Mastiff is a breed that inhabits the high and cold zones of the Chinese plateau and is the only canine 'living fossil' of the snowfield, unchanged by time and environment[184,186,201]. Results from phylogeny both places the Tibetan Mastiff as the basal breed, being the closest to the grey wolves and one migration was found between this breed and Sage Kuchi. Although the literature points that Tibetan Mastiff is most likely the ancestor of the LGDs, results from this work are not enough to support that theory. But the Tibetan Mastiff is, in fact, the closest breed to their wild ancestor.

Results from the Chapter 2 state that differentiation among Iberian breeds is quite low and that their evolutionary history is firmly intertwined. The results from this work point in the same direction, with individuals from some breeds being allocated with other Iberian breeds in the neighbour-joining tree. It can be easily explained by the long story of transhumance movements in the Peninsula with only ended in the late 20th century[119,188]. It is perfectly possible that the dogs that were taken with the shepherds during transhumance could breed with other breeds, thus maintaining their close genetic relationship.

One migration event found among this group was between the Spanish Mastiff and the Rafeiro of Alentejo, with geographical proximity being able to explain how the migration event happened but, being more forward, this migration could also support the common belief that Spanish Mastiff is the ancestor of Rafeiro[162]. Also, some individuals of Rafeiro of Alentejo are allocated in the Spanish Mastiff clade in the neighbour-joining tree.

Migration events and the allocation of this breed on the phylogenetic tree support the idea that Castro Laboreiro and Spanish Mastiffs were used to generate the Transmontano Mastiff. Also, results of the F3 statistics show that besides the usage of those 2 breeds in the creation of Transmontano Mastiff, Estrela Mountain dog could also have been used. Although, Rafeiro of Alentejo appeared in the f3 statistics which is strange because other results, particularly from Chapter 2, support the theory that Rafeiro of Alentejo is the main ancestor of Transmontano Mastiff.

The last migration appears from the grey wolf to the Spanish Mastiff. This breed is probably the most ancient dog breed in the Peninsula and highly resembles the Asian dogs, like Sage Kuchi. In the Iron Age, skeletons of dogs of enormous dimensions were found in the Castilian Plateau, Spain[202]. It is possible that dogs were brought from Europe to the Peninsula along with human

migrations and that these dogs bred with Iberian wolves to give birth to the first Spanish Mastiffs, although information from these times in the Peninsula are scarce.

Aidi appears in the same place in both phylogenetic trees, among the European breeds, being closer to them than with Asian or Iberian breeds. Two migratory events were inferred directed to Aidi, one from the grey wolf, which seems unlikely due to the inexistence of this species in northern Africa and the other from the Asian group. Many believe that Aidi was developed by the Phoenicians using dogs from their land (Syria and Northern Israel) during the time where these people would sail across the Mediterranean to do commerce[67,203]. This could explain the former migration event but not why Aidi seems to be more related to the European breeds.

3.6. Conclusion

It is possible that livestock guarding dogs all originated from the same stock, with differentiation mainly caused by geographical isolation leading to the birth of differentiated breeds[64]. Other factor that caused differentiation to happen may have been adaptation to habitat as well as to the type of livestock the dog was protecting or the natural predators of that region. Thus, individual breeds may have diverged by specialized skill and appearance.

Asian dogs are well known to display the highest levels of diversity among dog breeds and are representative of the basal clades in phylogenetic trees. Asian livestock guarding dogs are no exception and strongly group together in the basal clades, forming a group so cohesive that different breeds are not even able to be assigned to this private cluster in the admixture analysis and differentiation among them seems to be almost none. Genetic differentiation values are low between every single breed no matter what the geographical distance between them is. Pointing this, not only the Asian dogs, but all the LGDs form a low differentiated group.

While in previous studies it was found that non-Asian dogs present less genetic diversity, the results in this study do not show the same for livestock guarding dogs. Although a gradient in diversity loss can be perceived, the genetic diversity for western breeds is almost the same as in Asian breeds.

Other curiosity surrounding the Livestock guarding dogs is the inbreeding. Only one of the breeds here analysed, the Maremma Sheepdog, presented genetic signals of inbreeding in concordance with other dog breeds, with the remaining exhibiting low inbreeding coefficient. Since livestock guarding dogs were designed for one particular function, intense levels of selection were not necessary, and these breeds were mostly kept out of the breeding practices that modern dog breeds were under.

With this in mind, the explanation of differentiation through isolation seems to be simplistic for the occurrence of so many dog breeds, with so many different phenotypes without any major artificial events having occurred and, at the same time, maintaining such tight relationships. Events of genetic migration between breeds separated by long geographical distances were found, proving that gene flow was continued across the ages. In the Iberian Peninsula, for example, it is not fully known if dogs came when the peninsula was conquered by the Roman Empire or if they arrived earlier with Phoenicians merchants. All these human events joined with the human need to better adapt guarding dogs to different environments, support the theory of a genetic pool enrichment caused by inter-breed mating.

As a final thought, livestock guarding dogs probably appeared somewhere in the Asian continent from the available dogs already present there and the guarding breeds, as we currently know them, emerged not only because of differentiation caused by isolation but also by incorporate some signals of admixture from other breeds on their genome.

CHAPTER 4.

Final remarks

There are at least 40 extant livestock guarding breeds and many of them are today recognized by the FCI. This organization sets physical standards for dog breeds but many European and some Asian countries campaign in favour of recognition for working breeds. Unrest and war have driven many LGDs nearly to extinction and kennel clubs defend that recognition may be the key to protect endangered breeds, self-acknowledging their role as caretakers of these dogs[64].

While the official recognition may lead to an acknowledgment by the great public of certain breeds this can be also problematic. The need of different dogs that can perform diverse functions has prehistoric roots and led to the creation of distinct forms of dogs and the modern breeding practices, that only became recurrent in the 19th century, mainly focus on the creation of dogs with specific phenotypes. These practices largely require strict aesthetic requirements and closed bloodlines and are implemented by Kennel clubs everywhere.

The Tibetan Mastiff reached worldwide popularity in the early 20th century when two exemplars were brought from Tibet to offer King George V of England[52]. However, following both World Wars, the transportation of dogs from Tibet ended and the breed nearly disappeared from England. Tibetan Mastiff gained again popularity in the late 20ths and the breeding skyrocketed with the producing of a huge number of dogs, which led to the breed recognition of show organizations (FCI, AKC) and the fixed standards for the breed.

These fixed standards were only achieved by the practice of inbreeding since the exportation of dogs from Tibet was prohibited and western breeders had only access to a limited gene pool from the original stock. This resulted in dogs highly inbred and of questionable quality, somewhat different from the native dogs of Tibet that are not under breeding practices.

In the beginning of the 21st century, the Tibetan Mastiff fever continued with some breeders selling dogs at the cost of almost 2 million dollars. Since large profits were made from the sale of Tibetan Mastiffs, and exemplars bigger than the original and with exotic coats were preferred by the general public, the purebred Tibetan Mastiff got to the edge of extinction[184].

Tibetan Mastiff is only the tip of the iceberg when it comes to poor management of the domestic dogs. In Iberia, there is a clear lack of concern when it comes to native dogs with obvious preference for foreign modern breeds. The Galician Palleiro, for example, is on the verge of extinction and the official recognition could be enough to raise awareness. Although, this breed fails to meet the criteria of a pure-breed because of their low effective number.

The Tibetan Mastiff and the Galician Palleiro are not the only breeds facing a dark future. Results from the Chapter 3 demonstrated that breeds like the Maremma Sheepdog and Rafeiro of Alentejo are displaying levels of inbreeding that are not typical of livestock guarding dogs, probably due to breeding practices. Preliminary analysis using samples from, both show and working dogs, from the Estrela Mountain dog breed, confirmed that there is a diversity loss in the genetic pool of the show dogs and they can clearly be separated in a structure analysis from working dogs. Estrela

Mountain dog became one of the most popular breeds in the world and is very appreciated in Portugal but, with the breed recognition and all the fame, came the breeding practices that damaged the genomic patrimony of the breed.

All these cases should be a warning that, while the official recognition may be an important tool to protect, not only the effective numbers of dogs within the breeds but also their genetic patrimony, it is important for people to realize that with recognition comes standards for appearance at the expense of diversity and perhaps the instincts needed for their original purpose.

Besides their cultural importance, livestock dogs have great importance in the management of wildlife because guard dogs can decrease livestock depredation by large carnivores[153]. Several on-going projects for the conservation of large carnivore populations include the use of LGDs. In fact, the usage of LGDs on those projects is an ideal option in an economical point of view since large predators tend to flee at a sight of dogs, reducing the numbers of livestock kills, in addiction of reducing the need for manpower[168]. Over 1000 LGDs are now working in the Alps and Kangal dogs were exported from Turkey to Africa to help protect Ungulate herds against cheetahs and Lions, being a “green” convervation measure[64,170,204,205].

Further work is necessary to comprehend the relations between livestock dogs and their evolutionary history. Also, a higher effort should be applied in the understanding of the real consequences of the pureness standards that kennel associations are enforcing. One proposes that genomic criteria, as described by Dreger et al.[46,115] and Talenti et al.[47], should be favoured by kennel associations to the detriment of the classical criteria which, most of the time, can always be met by breeds with inbred lineages and high effective numbers.

5. References

1. Gwatkin, R. D. S. Dog and Human Migrations. 154–165 (1933).
2. Morey, D. F. In search of Paleolithic dogs: A quest with mixed results. *J. Archaeol. Sci.* **52,** 300–307 (2014).
3. Schlebusch, C. M. *et al.* Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science (80-.).* **358,** 652–655 (2017).
4. Bocquet-Appel, J.-P. When the world's population took off: the springboard of the Neolithic Demographic Transition. *Science* **333,** 560–1 (2011).
5. Germonpré, M. *et al.* Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: osteometry, ancient DNA and stable isotopes. *J. Archaeol. Sci.* **36,** 473–490 (2009).
6. Mithen, S. J. *After the ice : a global human history, 20,000-5000 BC*. (Harvard University Press, 2004).
7. Peltenburg, E. J. & Wasse, A. *Neolithic revolution : new perspectives on southwest Asia in light of recent discoveries on Cyprus*. (Oxbow Books, 2004).
8. Lewin, R. *Human evolution: an illustrated introduction*. (Blackwell, 2005).
9. Kiple, K. F. & Ornelas, K. C. *The Cambridge world history of food*. (Cambridge University Press, 2000).
10. Donn, L. & Donn, D. *Ancient Mesopotamia/Ancient India*. (Social Studies School Service, 2004).
11. Haviland, W. A., Prins, H. E. L., Walrath, D. & McBride, B. *The essence of anthropology*. (Cengage Learning, 2013).
12. National Geographic. The Development of Agriculture. *Genographic Project* Available at: https://genographic.nationalgeographic.com/development-of-agriculture/.
13. Landry, J.-M. The use of guard dogs in the Swiss Alps: a first analysis. *Kora Bericht* **2 e,** 1–26 (1999).
14. Wayne, R. K. Cranial Morphology of Domestic and Wild Canids: The Influence of Development on Morphological Change. *Evolution (N. Y).* **40,** 243 (1986).
15. Miklosi, A. *Dog Behaviour, Evolution, and Cognition*. (Oxford University Press, 2014).
16. Jensen. *The behavioural biology of dogs*. (CABI, 2007).
17. Spady, T. C. & Ostrander, E. A. Canine behavioral genetics: pointing out the phenotypes and herding up the genes. *Am. J. Hum. Genet.* **82,** 10–8 (2008).
18. Price, E. O. Behavioral Aspects of Animal Domestication. *Q. Rev. Biol.* **59,** 1–32 (1984).
19. Lindsay, S. R. Learning and Behavioral Disturbances. in *Handbook of Applied Dog Behavior and Training, Volume One* 325–359 (Iowa State University Press).

20. Crockford, S. A commentary on dog evolution: regional variation, breed development and hybridisation with wolves. *BAR Int. Ser.* (2000).
21. Russell, N. *Social zooarchaeology : humans and animals in prehistory*. (Cambridge University Press, 2012).
22. Ucko, P. J. & Dimbleby, G. W. *The domestication and exploitation of plants and animals*. (Routledge, imprint of Taylor & Franci Group, 2017).
23. Coppinger, R. & Schneider, R. *Evolution of working dogs. In The domestic dog: its evolution, behaviour and interactions with people. Cambridge University Press, Cambridge, UK* (1995).
24. Lindblad-Toh, K. *et al.* Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438,** 803–819 (2005).
25. Vilà, C. *et al.* Multiple and ancient origins of the domestic dog. *Science* **276,** 1687–9 (1997).
26. Savolainen, P., Zhang, Y., Luo, J., Lundeberg, J. & Leitner, T. Genetic Evidence for an East Asian Origin of Domestic Dogs. *Science (80-.).* **298,** 1610–1613 (2002).
27. Wang, G.-D. *et al.* Out of southern East Asia: the natural history of domestic dogs across the world. *Cell Res.* **26,** 21–33 (2016).
28. VonHoldt, B. M. *et al.* Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* **464,** 898–902 (2010).
29. Klütsch, C. F. & de Caprona, M. D. The IGF1 small dog haplotype is derived from Middle Eastern grey wolves: a closer look at statistics, sampling, and the alleged Middle Eastern origin of small dogs. *BMC Biol.* **8,** 119 (2010).
30. Thalmann, O. *et al.* Complete Mitochondrial Genomes of Ancient Canids Suggest a European Origin of Domestic Dogs. *Science (80-.).* **342,** 871–874 (2013).
31. Freedman, A. H. *et al.* Genome Sequencing Highlights the Dynamic Early History of Dogs. *PLoS Genet.* **10,** e1004016 (2014).
32. Larson, G. & Bradley, D. G. How Much Is That in Dog Years? The Advent of Canine Population Genomics. *PLoS Genet.* **10,** e1004093 (2014).
33. Fan, Z. *et al.* Worldwide patterns of genomic variation and admixture in gray wolves. *Genome Res.* **26,** 163–173 (2016).
34. Boyko, A. R. *et al.* Complex population structure in African village dogs and its implications for inferring dog domestication history. *Proc. Natl. Acad. Sci. U. S. A.* **106,** (2009).
35. Conrad, D. F. *et al.* A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat. Genet.* **38,** 1251–1260 (2006).
36. Morris, D. *Dogs : the ultimate dictionary of over 1,000 dog breeds*. (Trafalgar Square Pub, 2002).
37. Wilcox, B. & Walkowicz, C. *Atlas of dog breeds of the world. New rev*. (TFH Publications., 1995).

38. Parker HG, Kim LV, Sutter NB, Carlson S, Lorentzen TD, Malek TB, Johnson GS, DeFrance HB, Ostrander EA, K. L. Genetic structure of the purebred domestic dog. *Science (80-.).* **1160,** 1160–1165 (2004).
39. Larson, G. *et al.* Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc. Natl. Acad. Sci.* **109,** 8878–8883 (2012).
40. Botigué, L. R. *et al.* Ancient European dog genomes reveal continuity since the Early Neolithic. *Nat. Commun.* **8,** (2017).
41. Clutton-Brock & J. *A Natural History of Domesticated Mammals*. (Cambridge University Press, 1999).
42. Gonçalves, M. A. L. Cães de protecção e cães de condução de gado: aspectos de comportamento e de endocrinologia. (Instituto Superior de Psicologia Aplicada, 2002).
43. Sierra Alfranca, I. El concepto de raza: evolución y realidad. *Arch. Zootec.* **50,** 7 (2001).
44. Coppinger, R. & Coppinger, L. *Dogs : a startling new understanding of canine origin, behavior, and evolution*. (Scribner, 2001).
45. Donna L, Morden; Seranne, Ann; Wendell J, Sammet; Gasow, J. *The joy of breeding your own show dog*. (Howell Book House Inc, 1980).
46. Dreger, D. L. *et al.* Whole-genome sequence, SNP chips and pedigree structure: building demographic profiles in domestic dog breeds to optimize genetic-trait mapping. *Dis. Model. Mech.* **9,** 1445–1460 (2016).
47. Talenti, A. *et al.* Studies of modern Italian dog populations reveal multiple patterns for domestic breed evolution. *Ecol. Evol.* **8,** 2911–2925 (2018).
48. Parker, H. G. Genomic analyses of modern dog breeds. *Mamm. Genome* **23,** 19–27 (2012).
49. Russell, N. Neolithic Human-Animal Relations. *Groniek* **48,** (2015).
50. Guardamagna, A. & Breffort, C. *Le chien de montagne des Pyrénées*. (De Vecchi, 1997).
51. Wang, G. D. *et al.* The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat. Commun.* **4,** 1860–1869 (2013).
52. Cui, T. B. & Guo, X. A Monograph of Tibetan Mastiff in China. *Shanghai Scientific & Technical Publishers* (2009).
53. Ni, Z. Tibetan Mastiff. *China Youth Press (in Chinese)* (2004).
54. Grignon, D. Le chien de berger yougoslave de Charplanninatz. *Acta Biol. Montna* **1,** 71–79 (1982).
55. Coly, J. *Le berger des Pyrénés*. (P.D. Edition, 1994).
56. Cruz, C. Another View of livestock guarding dog history. *AKC Gazette 4/95* (1995).
57. Morsiani, G. *Le St-Bernard*. (Ed. de Vecchi, 1993).
58. Tschudi, F. *Geschichte des Hundes*. (1926).
59. Grandin, T. *Livestock Handling and Transport*. *Journal of Chemical Information and Modeling*

53, (2013).

60. Thomas, K. *Man and the natural world : a history of the modern sensibility*. (Pantheon Books, 1983).
61. Schmitt, R. Chiens de protection des troupeaux. *Ethnozootechnie* 51–58 (1989).
62. Barnes, E. Working like a dog. *Mother Earth News* **168,** (1998).
63. Coppinger, R., Glendinning, J., Torop, E., Matthay, C. & Sutherland, M. & Smith, C. Degree of behavioural neoteny differentiates canid polymorphs. *Ethology* **75,** 89–108 (1987).
64. Gehring, T. M., VerCauteren, K. C. & Landry, J.-M. Livestock Protection Dogs in the 21st Century: Is an Ancient Tool Relevant to Modern Conservation Challenges? *Bioscience* **60,** 299–308 (2010).
65. Fédération Cynologique Internationale. *Portuguese Sheepdog*. (2008).
66. Fédération Cynologique Internationale. *Pyrenean Mountain dog*. (2001).
67. Fédération Cynologique Internationale. *Aidi*. (2003).
68. United Kennel Club. *Akbash dog*. (1998).
69. Fédération Cynologique Internationale. *Anatolian Shepherd dog*. (2017).
70. Fédération Cynologique Internationale. *Castro Laboreiro dog*. (2008).
71. Fédération Cynologique Internationale. *Caucasian Shepherd dog*. (2010).
72. Fédération Cynologique Internationale. *Central Asia Shepherd dog*. (2010).
73. Fédération Cynologique Internationale. *Estrela Mountain dog*. (2008).
74. Fédération Cynologique Internationale. *Kangal dog*. (2018).
75. Fédération Cynologique Internationale. *Komondor*. (2000).
76. Molosserdogs.com. Kurdish Shepherd Dog.

Available at: http://molosserdogs.com/m/articles/view/2738-kurdish-shepherd-dog.

77. Fédération Cynologique Internationale. *Leonberger*. (1996).
78. Fédération Cynologique Internationale. *Maremma Sheepdog*. (2015).
79. Kynologikós Omilos Elládos. *Molossus of Epirus*.
80. molosserdogs.com. Persian Mastiff.

Available at: http://molosserdogs.com/m/articles/view/Persian-Mastiff.

81. Fédération Cynologique Internationale. *Pyrenean Mastiff*. (1982).
82. Fédération Cynologique Internationale. *Rafeiro of Alentejo*. (2008).
83. molosserdogs.com. Sage Kuchi. Available at: http://molosserdogs.com/m/articles/view/1003-afghan-mountain-dog.
84. Fédération Cynologique Internationale. *Spanish Mastiff*. (1996).
85. Fédération Cynologique Internationale. *St Bernard*. (2016).
86. Fédération Cynologique Internationale. *Tibetan Mastiff*. (2015).
87. Clube Português de Canicultura. *Transmontano Mastiff*. (2017).

88. molosserdogs.com. Turkmenistan Shepherd Dog.
Available at: http://molosserdogs.com/m/articles/view/3110-turkmenistan-shepherd-dog.
89. Fédération Cynologique Internationale. *Yugoslavian Shepherd dog*. (1970).
90. Real Sociedad Canina de España. *Basque Shepherd*.
91. Club Can de Palleiro (clubcandepalleiro.com). Galician Palleiro.
Available at: http://www.clubcandepalleiro.com/origenes-la-raza/.
92. molosserdogs.com. Galician Palleiro.
Available at: http://molosserdogs.com/m/articles/view/3067-palleiro-cattledog.
93. Real Sociedad Canina de España. *Leonese Shepherd*.
94. Fédération Cynologique Internationale. *Saint Miguel Cattle dog*. (2007).
95. mastinesibericos.es. Spanish Wolfdog. Available at: http://mastinesibericos.es/otras-razas-caninas/perro-lobo-español/.
96. Clube Português de Canicultura. *Terceira Cattle dog*.
97. Fédération Cynologique Internationale. *Portuguese Water dog*. (2008).
98. Fédération Cynologique Internationale. *Spanish Water dog*. (1996).
99. molosserdogs.com. Villano Bulldog.
Available at: http://molosserdogs.com/m/articles/view/2853-villano-bulldog.
100. Arif, I. A. *et al.* DNA marker technology for wildlife conservation. *Saudi J. Biol. Sci.* **18,** 219–225 (2011).
101. Richard, G.-F., Kerrest, A. & Dujon, B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol. Mol. Biol. Rev.* **72,** 686–727 (2008).
102. Lougheed, S. C., Gibbs, H. L., Prior, K. A. & Weatherhead, P. J. A comparison of RAPD versus microsatellite DNA markers in population studies of the massasauga rattlesnake. *J. Hered.* **91,** 458–63 (2000).
103. Spencer, C. C., Neigel, J. E. & Leberg, P. L. Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. *Mol. Ecol.* **9,** 1517–1528 (2000).
104. Nielsen, R. Molecular Signatures of Natural Selection. *Annu. Rev. Genet.* **39,** 197–218 (2005).
105. Slatkin, M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139,** (1995).
106. GUICHOUX, E. *et al.* Current trends in microsatellite genotyping. *Mol. Ecol. Resour.* **11,** 591–611 (2011).
107. Weeks, D. E., Conley, Y. P., Ferrell, R. E., Mah, T. S. & Gorin, M. B. A tale of two genotypes: consistency between two high-throughput genotyping centers. *Genome Res.* **12,** 430–5 (2002).
108. Kruglyak, L. & Nickerson, D. A. Variation is the spice of life. *Nature* (2001).
109. Vignal, A., Milan, D., SanCristobal, M. & Eggen, A. A review on SNP and other types of molecular markers and their use in animal genetics. *Genet. Sel. Evol.* **34,** 275–305 (2002).

110. Helyar, S. J. *et al.* Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Mol. Ecol. Resour.* **11,** 123–136 (2011).
111. Morin, P. A., Luikart, G. & Wayne, R. K. SNPs in ecology, evolution and conservation. *Trends Ecol. Evol.* **19,** 208–216 (2004).
112. Lachance, J. & Tishkoff, S. A. SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it. *BioEssays* **35,** 780–786 (2013).
113. Gärke, C. *et al.* Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations. *Anim. Genet.* **43,** 419–428 (2012).
114. Choi, B. H. *et al.* Genome-wide analysis of the diversity and ancestry of Korean dogs. *PLoS One* **12,** 1–20 (2017).
115. Dreger, D. L. *et al.* Commonalities in Development of Pure Breeds and Population Isolates Revealed in the Genome of the Sardinian Fonni's Dog. *Genetics* **204,** 737–755 (2016).
116. Huson, H. J. *et al.* Breed-specific ancestry studies and genome-wide association analysis highlight an association between the MYH9 gene and heat tolerance in Alaskan sprint racing sled dogs. *Mamm. Genome* **23,** 178–194 (2012).
117. Parker, H. G. *et al.* Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development. *Cell Rep.* **19,** 697–708 (2017).
118. Akey, J. M. *et al.* Tracking footprints of artificial selection in the dog genome. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 1160–5 (2010).
119. Ruiz, M. & Ruiz, J. P. Ecological history of transhumance in Spain. *Biol. Conserv.* **37,** 73–86 (1986).
120. Martín PG; Angioni G; Raverdy Ch; Ojanen U; Estébanez NL; Martínez BG; Chamorro GV; Lenberg JG; Ilera FA; Alonso AC; Albaladejo PF; Bernal AM; Toledo AA; Díez LM; Canada AM; García MM; Pascual MR; Cardoso JJ. *Por los Caminos de la Trashumancia [On the Tanshumance paths].* (1994).
121. Eeden, L. M. Van *et al.* Carnivore conservation needs evidence-based livestock protection. *PLoS Biol.* **In press,** 1–8 (2018).
122. Jordana, J., Piedrafita, J., Sanchez, A. & Puig, P. Comparative F Statistics Analysis of the Genetic Structure of Ten Spanish Dog Breeds. *J. Hered.* **83,** 367–374 (1992).
123. van Asch, B. *et al.* MtDNA diversity among four Portuguese autochthonous dog breeds: a fine-scale characterisation. *BMC Genet.* **6,** 37 (2005).
124. Pires, A. E. *et al.* Mitochondrial DNA Sequence Variation in Portuguese Native Dog Breeds: Diversity and Phylogenetic Affinities. *J. Hered.* **97,** 318–330 (2006).
125. Pires, A. E. *et al.* Molecular structure in peripheral dog breeds: Portuguese native breeds as a case study. *Anim. Genet.* **40,** 383–392 (2009).
126. Pires, A. E. *et al.* New insights into the genetic composition and phylogenetic relationship of

wolves and dogs in the Iberian Peninsula. *Ecol. Evol.* **7,** 4404–4418 (2017).

127. Miller, S. A., Dykes, D. D. & Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16,** 1215 (1988).
128. Bannasch, D. L., Bannasch, M. J., Ryun, J. R., Famula, T. R. & Pedersen, N. C. Y chromosome haplotype analysis in purebred dogs. *Mamm. Genome* **16,** 273–280 (2005).
129. Holmes, N. G. *et al.* Isolation and characterization of microsatellites from the canine genome. *Anim. Genet.* **24,** 289–292 (1993).
130. Holmes, N. G. *et al.* Eighteen canine microsatellites. *Anim. Genet.* **26,** 132a–133 (1995).
131. Lingaas, F. *et al.* Towards construction of a canine linkage map: Establishment of 16 linkage groups. *Mamm. Genome* **8,** 218–221 (1997).
132. Ostrander, E. A., Sprague, G. F. & Rine, J. Identification and Characterization of Dinucleotide Repeat (CA)n Markers for Genetic Mapping in Dog. *Genomics* **16,** 207–213 (1993).
133. Ostrander, E. A., Mapa, F. A., Yee, M. & Rine, J. One hundred and one new simple sequence repeat-based markers for the canine genome. *Mamm. Genome* **6,** 192–195 (1995).
134. Breen, M. *et al.* 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 Res.* **11,** 1784–1795 (2001).
135. Fredholm, M. & Winter, A. K. Variation of short tandem repeats within and between species belonging to the Canidae family. *Mamm. Genome* **6,** 11–18 (1995).
136. Kerns, J. A. *et al.* Characterization of the dog Agouti gene and a nonagoutimutation in German Shepherd Dogs. *Mamm. Genome* **15,** 798–808 (2004).
137. Francisco, L. V, Langsten, A. A., Mellersh, C. S., Neal, C. L. & Ostrander, E. A. A class of highly polymorphic tetranucleotide repeats for canine genetic mapping. *Mamm. Genome* **7,** 359–362 (1996).
138. Mariat, D., Kessler, J. L., Vaiman, D. & Panthier, J. J. Polymorphism characterization of five canine microsatellites. *Anim. Genet.* **27,** 434–435 (1996).
139. Neff, M. W. *et al.* A second-generation genetic linkage map of the domestic dog, Canis familiaris. *Genetics* **151,** 803–20 (1999).
140. Guyon, R. *et al.* A 1-Mb resolution radiation hybrid map of the canine genome. *Proc. Natl. Acad. Sci.* **100,** 5296–5301 (2003).
141. Shibuya, H., Collins, B. K., Huang, T. H.-M. & Johnson, G. S. A polymorphic (AGGAAT), tandem repeat in an intron of the canine von Willebrand factor gene. *Anim. Genet.* **25,** 122 (1994).
142. Blacket, M. J., Robin, C., Good, R. T., Lee, S. F. & Miller, A. D. Universal primers for fluorescent labelling of PCR fragments-an efficient and cost-effective approach to genotyping by fluorescence. *Mol. Ecol. Resour.* **12,** 456–463 (2012).

143. Excoffier, L., Laval, G. & Schneider, S. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10,** 564–567 (2010).
144. Goudet, J. *FSTAT (Version 1.2): A computer program to calculate F-Statisitics. Journal of Heredity* **86,** (1994).
145. Holsinger, K. E. & Weir, B. S. Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nat. Rev. Genet.* **10,** 639–50 (2009).
146. Herrnstadt, C. *et al.* Reduced-Median-Network Analysis of Complete Mitochondrial DNA Coding-Region Sequences for the Major African, Asian, and European Haplogroups. *Am. J. Hum. Genet.* **70,** 1152–1171 (2002).
147. Bandelt, H. J., Forster, P., Sykes, B. C. & Richards, M. B. Mitochondrial portraits of human populations using median networks. *Genetics* **141,** 743–53 (1995).
148. Bandelt, H. J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16,** 37–48 (1999).
149. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164,** 1567–87 (2003).
150. Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. & Mayrose, I. Clumpak : a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* **15,** 1179–1191 (2015).
151. Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N. & Catch, F. GENETIX 4.0. 5.2., Software under Windows^{TM} for the genetics of the populations. *Univ. Montpellier, Montpellier, Fr.* (2004).
152. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* **14,** 2611–2620 (2005).
153. Rigg, R. Livestock guarding dogs: their current use worldwide. *Canid Spec. Gr.* 133 (2001).
154. Vasconcelos, R. C. *Raças de Cães Portugueses [Portuguese dog breeds]*. (Editorial Presença, 1995).
155. Gomes, M. S. Raças Caninas Autóctones Portuguesas. Contributo para o seu Estudo Genético e Demográfico [Portuguese autochthonous dog breeds. A contribution to their genetic and demographic study]. (Escola Superior Agrária de Santarém, 2003).
156. Pedersen, N., Liu, H., Theilen, G. & Sacks, B. The effects of dog breed development on genetic diversity and the relative influences of performance and conformation breeding. *J. Anim. Breed. Genet.* **130,** 236–248 (2013).
157. Petrucci-Fonseca, F. O lobo (Canis lupus signatus Cabrera, 1907) em Portugal. Problemática da sua conservação. (1990).
158. Iacolina, L. *et al.* Y-chromosome microsatellite variation in Italian wolves: A contribution to the study of wolf-dog hybridization patterns. *Mamm. Biol. - Zeitschrift für Säugetierkd.* **75,** 341–347

(2010).

159. Club del Mastín del Pirineo de España ©. Pyrenean Mastiff Story (in spanish). Available at: http://www.cmpe.es/index.php/en/mastin-del-pirineo/mastin-history. (Accessed: 27th September 2018)
160. Geraldes, A. Brandas e inverneiras: particularidades do sistema agro-pastoril crastejo [Particularities of the Castro Laboreiro's agricultural and pastoral system]. *Instituto da Conservação da Natureza/Parque Nacional da Peneda-Gerês.* 471–476 (1996). doi:10.1201/b15685-82
161. Amaro, A. Contribuição para a Caracterização da Estrutura Genética de Raças Caninas Autóctones Portuguesas. Estudo do Grupo de Ligação da Transferrina em Raças Caninas Tipo Mastim [Contribution to the genetic characterization of the Portuguese autochthonous dog bre. (University of Porto, 2001).
162. Alpoim, J. A. *O Rafeiro do Alentejo - Monografia da raça*. (1999).
163. Koskinen, M. T. Individual assignment using microsatellite DNA reveals unambiguous breed identification in the domestic dog. *Anim. Genet.* **34,** 297–301 (2003).
164. Bigi, D., Marelli, S. P., Randi, E. & Polli, M. Genetic characterization of four native Italian shepherd dog breeds and analysis of their relationship to cosmopolitan dog breeds using microsatellite markers. *animal* **9,** 1921–1928 (2015).
165. Udell, M. A. R. & Wynne, C. D. L. A review of domestic dogs' (Canis familiaris) human-like behaviors: or why behavior analysts should stop worrying and love their dogs. *J. Exp. Anal. Behav.* **89,** 247–61 (2008).
166. Amaral, A. & Veiga, V. O Cão de Fila de S. Miguel [The Azores cattle dog]. (2004).
167. Dohner, J. V. *Farm dogs : a comprehensive breed guide to 93 guardians, herders, terriers, and other canine working partners*.
168. Dohner, J. V. *Livestock guardians : using dogs, donkeys, and llamas to protect your herd*. (Storey Pub, 2007).
169. Neff, M. W. *et al.* Breed distribution and history of canine mdr1-1 , a pharmacogenetic mutation that marks the emergence of breeds from the collie lineage. *Proc. Natl. Acad. Sci.* **101,** 11725–11730 (2004).
170. Linnell, J. D. C. & Lescureux, N. *Livestock guarding dogs – cultural heritage icons with a new relevance for mitigating conservation conflicts*. (Norwegian Institute for Nature Research, 2015).
171. Cruz, C. M. O. As raças portuguesas de cães de gado e de pastoreio: aspectos morfológicos e comportamentais. (Universidade Técnica de Lisboa., 2007).
172. Illumina. CanineSNP20 BeadChip. *Customer Solutions* (2007).
173. Wayne, R. K. & Vonholdt, B. M. Evolutionary genomics of dog domestication. *Mamm. Genome*

23, 3–18 (2012).

174. Wayne, R. K. & Ostrander, E. A. Lessons learned from the dog genome. *Trends Genet.* **23,** 557–567 (2007).
175. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
176. Alexander, D. H., Novembre, J. & Lange, K. Fast Model-Based Estimation of Ancestry in Unrelated Individuals. (2009).
177. Felsenstein, J. PHYLIP: Phylogeny Inference Package. *Cladistics* **5,** 164–166 (1989).
178. FigTree. Available at: http://tree.bio.ed.ac.uk/software/figtree/.
179. Pickrell, J. K. & Pritchard, J. K. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genet.* **8,** e1002967 (2012).
180. Neuwirth, E. RColorBrewer: ColorBrewer Palettes. (2014). Available at: https://cran.r-project.org/package=RColorBrewer.
181. Neuwirth, E. & Brewer, R. ColorBrewer palettes. (2005).
182. Reich, D., Thangaraj, K. K. K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian Population History. *Nature* **461,** 489–494 (2009).
183. Vychodilova, L. *et al.* Genetic diversity and population structure of African village dogs based on microsatellite and immunity-related molecular markers. *PLoS One* **13,** e0199506 (2018).
184. Li, Q. *et al.* Origin and phylogenetic analysis of Tibetan Mastiff based on the mitochondrial DNA sequence. *J. Genet. Genomics* **35,** 335–340 (2008).
185. Li, Y., Li, Q., Zhao, X., Xie, Z. & Xu, Y. Complete sequence of the Tibetan Mastiff mitochondrial genome and its phylogenetic relationship with other Canids (Canis, Canidae). *animal* **5,** 18–25 (2011).
186. Li, Y. *et al.* The origin of the Tibetan Mastiff and species identification of Canis based on mitochondrial cytochrome c oxidase subunit i (COI) gene and COI barcoding. *Animal* **5,** 1868–1873 (2011).
187. FCI. *Cane de Pastore Maremmano Abruzzese*. (2016).
188. Silva, R. F. Transumância no Portugal central: diversidade e organizacão do território [Transhumance in central Portugal: diversity and land organization]. *Assoc. da Def. do património Arouquense, Ed.* (2000).
189. Pickrell, J. K. & Pritchard, J. K. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genet.* **8,** e1002967 (2012).
190. Fadel, F. R. *et al.* Differences in Trait Impulsivity Indicate Diversification of Dog Breeds into Working and Show Lines. *Sci. Rep.* **6,** 22162 (2016).
191. Chen, R., Irwin, D. M. & Zhang, Y.-P. Differences in Selection Drive Olfactory Receptor Genes in Different Directions in Dogs and Wolf. *Mol. Biol. Evol.* **29,** 3475–3484 (2012).

192. Tacher, S. *et al.* Olfactory Receptor Sequence Polymorphism Within and Between Breeds of Dogs. *J. Hered.* **96,** 812–816 (2005).
193. Akkad, D. A., Gerding, W. M., Gasser, R. B. & Epplen, J. T. Homozygosity mapping and sequencing identify two genes that might contribute to pointing behavior in hunting dogs. *Canine Genet. Epidemiol.* **2,** 5 (2015).
194. Sundman, A.-S., Johnsson, M., Wright, D. & Jensen, P. Similar recent selection criteria associated with different behavioural effects in two dog breeds. *Genes, Brain Behav.* **15,** 750–756 (2016).
195. Cagan, A. & Blass, T. Identification of genomic variants putatively targeted by selection during dog domestication. *BMC Evol. Biol.* **16,** 10 (2016).
196. Pfahler, S. & Distl, O. Effective Population Size, Extended Linkage Disequilibrium and Signatures of Selection in the Rare Dog Breed Lundehund. *PLoS One* **10,** e0122680 (2015).
197. Kumpulainen, M. *et al.* Founder representation and effective population size in old versus young breeds-genetic diversity of Finnish and Nordic Spitz. *J. Anim. Breed. Genet.* **134,** 422–433 (2017).
198. Lusby, M. *Leonberger*. (Kennel Club Books, 2005).
199. Patai, R. Nomadism: Middle Eastern and Central Asian. *Southwest. J. Anthropol.* **7,** 401–414 (1951).
200. Spooner, B. The Status of Nomadism as a Cultural Phenomenon in the Middle East. *J. Asian Afr. Stud.* **7,** 122 (1972).
201. Ren, D. R. *et al.* Strong heterozygote deficit in Tibetan Mastiff of China based on microsatellite loci. *animal* **3,** 1213–1215 (2009).
202. Daza Perea, A. Preliminary Studies of Late Prehistoric Dog (Canis lupus f. Familiaris Linnaeus, 1758) Remains from the Iberian Peninsula: Osteometric and 2D Geometric Morphometric Approaches. *Pap. from Inst. Archaeol.* **275334,** 1–21 (2017).
203. Aidi Breed Information: History, Health, Pictures, and more. Available at: http://easypetmd.com/doginfo/aidi.
204. VerCauteren, K. C., Lavelle, M. J., Gehring, T. M. & Landry, J. M. Cow dogs: Use of livestock protection dogs for reducing predation and transmission of pathogens from wildlife to cattle. *Appl. Anim. Behav. Sci.* **140,** 128–136 (2012).
205. Coppinger, R. P., Smith, C. K. & Miller, L. Observations on Why Mongrels May Make Effective Livestock Protecting Dogs. *J. Range Manag.* **38,** 560–1 (1985).

6. Attatchments

6.1 – Breed description

Aidi – Atlas Mountain dog[67]

Aidi existed in the mountains and on the plateaux of Northern Africa for ages and, nowadays, can be found in large numbers in the Atlas Mountains of Morocco. Besides the role in protecting the flocks from predators, Aidi has also acquired the function of defending the shepherds' belongings. It is a solid, very hardy dog, noted for its power and has a thick coat which protects him from big temperature fluctuations and provides a protective armour in the fights against jackals.

Akbash dog[68]

Akbash is a white livestock guardian breed native to the plains and mountains of western Turkey. In Turkey, Akbash Dogs are usually owned and bred by villagers and shepherds to protect their sheep from wolves and other predators. Akbash have the size, strength, and courage to challenge large predators and the speed and agility to chase fleet predators. Gender differences can be striking in this breed and typically the male is proportionately taller and heavier than the female. Although there is no difference in the ability of males or females to perform as guardians.

Anatolian Shepherd Dog – Coban Köpegi[69]

Being one of the most ancient guard dogs in the history, the Anatolian Shepherd is believed to have descended from the large hunting dogs existing in Mesopotamia. They guard flocks, travelling great distances on the Central Anatolian Plateau and are physically prepared for the Turkish climate (which varies from hot and very dry summers to very cold winters). Although being large, tall and having a dense double coat, the Anatolian shepherd is also capable of great speed. As other breeds of livestock guarding dogs, they are wary on strangers when on duty but also loyal and affectionate to owners, independent and intelligent.

Basque Shepherd – Pastor Vasco[90]

The Basque Shepherd Dog is a breed originated from the Basque Country, which official name is Euskal Artzain Txakurra. It is believed that they originated from Central European herding dogs. These dogs are well proportioned, with strong bodies. There are two varieties, the Gorbeiakoa and the Iletsua. Their coat varies from a smooth red to a rough fawn,

according to their type (Cobeiakoa or Iletsua). As others herding dogs they are very energetic, have territorial instinct, but maintains a strong bond with the shepherd.

Castro Laboreiro dog – Cão de Castro Laboreiro[70]

One of the most ancient breeds in the Iberian Peninsula, it owes its name to the village of Castro Laboreiro, located in the Melgaço municipality in the extreme north of Portugal, where it comes from. Although not as big as other Portuguese mastiffs, it is as intelligent. Establishes a strong bond with his family, being docile, and very patient with children. It has a characteristic alarm bark, starting with a deep sound, rising to low-pitched and ending in prolonged high-pitched sounds. Due to being isolated this breed show high levels of consanguinity.

Caucasian Shepherd Dog – Kavkazskaïa Ovtcharka[71]

This breed's covers territories from the Caucasian Range and the steppe regions of Southern Russia. Historically Caucasian Shepherd dogs were used for guarding and safe-keeping of herds, flocks and dwellings from predators, and has its origin in ancient Caucasian dogs. In contrast with other dog breeds, in the Caucasian Shepherd, the sexual dimorphism is well pronounced with males being more massive, bigger and often shorter in body than females. Like others LGDs, their behaviour is steady, active, self-confident, fearless and independent.

Central Asia Shepherd dog – Sredneasiatskaya Ovtcharka[72]

The Central Asia Shepherd is thought to have been formed as a breed more than four thousand years in the vast territory of Central Asia, which spreads nowadays from the Caspian Sea to China and from Southern Ural to Afghanistan. Physically it has a large stature and robust and muscular body. In the Caucasian Shepherd, sexual dimorphism is well defined, with males being more massive and courageous than females. They are balanced quiet and independent as well as fearless towards predators.

Estrela Mountain Dog – Cão da Serra da Estrela[73]

Estrela Mountain Dog is a mastiff originated in the Portuguese mountain from where the breed got its name. It is the most ancient Portuguese breed and is very appreciated in the country. The progressive recognition of its aptitudes has led to its diffusion throughout the world since the second half of the 20th century, as a company dog. There are two varieties of coat: long and short, being the last one extremely rare. The breed has a lively, calm and expressive look.

Galician Palleiro – Can de Palleiro[91]

Originating in Galicia (Spain), at present, this dog breed is in potential danger of extinction due to its limited number. It's a medium size dog with harmonic proportions and strong constitution, wide bones, characteristic of its rusticity. This dog shows great fidelity towards its master and with the people in the house with whom he becomes sweet and calm. It is caught in the process of being recognized by the RSCE and the FCI, since it still does not meet the minimum number of specimens.

Kangal dog – Kangal Çöban Köpeği[74]

Kangals are a Turkish breed and are officially considered the country's national dog breed. While the Kangal Shepherd Dog is often referred to as a sheep dog, it is not a herding dog, but rather a flock guardian that lives with the flock of sheep to actively fend off predators of all sizes. Although not usually legal, some dogs have been exported to African countries like Namibia and Kenya in more recent years. The Kangal is a molossus type dog with robust constitution. Steady and bold without aggression, naturally independent, very intelligent and tractable.

Komondor[75]

Komondor is an old-established Hungarian breed of Asian origin. His original ancestors almost certainly came with the migrating Old Magyars, living as stock-breeding Nomads, to the Carpathian basin. This breed is commonly large with his robust body being covered by matted, corded, throughout dense, long hair. Although his big size, the Komondor is known for its silent attacks when a predator is endangering the herds, and for his suspicious nature

Kurdish Shepherd Dog – Pshdar[76]

Kurdish Shepherd Dog or Kurd mastiff is a type of molosser, often used as guard dogs and for protection of flocks of sheep and goats from predators. It is said this breed is the oldest Asian breed, with inscriptions made from the Babylonian and Assyrian civilizations (kept in the Louvre and the British Museum) claiming this. Their typical coat varies from yellow, white, red, burgundy, blue and black colours. These dogs are known to be quite cautious and courteous in dealing with their owner and owner's family, including children.

Leonberger[77]

This breed originated from controlled breeding in a monastery in South Germany. This resulted in very large dogs with predominantly long, white coats. The first dogs really called "Leonbergers" were born in 1846. It is claimed that this breed originated from crossings

between New found lands and St. Bernard with later genetic enrichment of Pyrenean Mountain dogs. In both World Wars and the needy post war times, the numbers of breeding stock reduced dramatically. Males, in particular, are powerful and strong. They are not aggressive and are willing to be submissive.

Leonese Shepherd – Carea Leonés[93]

The Leonese Shepherd is a breed from León, Castile and León, Spain. For centuries, they tended flocks of Churra (sheep) in the mountains of the historical region of León. It's a medium size dog. This breed is used as a herding dog and as a companion. Characterized by its austerity and strength. Always in its work before the most adverse and variable climatic conditions. Because of its intelligence, the Leonese Shepherd, like most sheepdogs, is easy to train.

Maremma (and the Abruzzes) Sheepdog – Cane Da Pastore Maremmano Abruzzese[78]

This ancient breed of dogs who guard flocks comes from shepherd dogs actually still used in the Abruzzes where the breeding of sheep is still thriving nowadays. The Maremma and Abruzzes sheepdog is a big dog and strongly built of a rustic appearance. The principal function as a guard and defence dog of flocks and property, in general, asserts the manner in accomplishing these tasks with devotions to his master.

Molossus of Epirus – Μολοσσός της Ηπείρου[79]

Used as a war dog, guardian and fighter in the past, the Molossus is today most commonly employed as a property watchdog and livestock protector. It is a pure Greek breed and is believed to be one of the main ancestors of today's molosser breeds, dispersing across Eurasia during the age of Alexander, The Great. Also, because of its long existence and development, the breed is extremely healthy and enduring. Large sized, medium-sized dog, compact, strong torso with strong bones, large head and medium-length mantle. Stable but also dynamic, very competent when he needs to protect the herd or property he preserves.

Persian Mastiff – Persian Sarabi Dog[80]

The Persian Sarabi dog also known as Persian Mastiff, Persian Shepherd and Iranian Mastiff is a breed of livestock guardian dog originally from the north of Iran. Like other livestock guarding dogs, it has been used for centuries by Iranian shepherds to guard sheep from wolves, bears, jackals and other animals. They have a solid and muscular build, thick necks and sturdy bodies. They are able to tolerate harsh conditions, little food and bad weather.

Portuguese Sheepdog – Cão de Serra De Aires[65]

The Portuguese Sheepdog is a herding dog native from Portuguese Sheepdogs (Monforte, Alentejo) easily distinguishable by his own skilled way of herding the cattle and how he finds lost animals in pastures. The breed is nicknamed the “cão macaco” (monkey dog) for its furry face and simian-like attitude. This breed has been used nowadays, not only for herding but also for watching different kinds of livestock; sheep, cattle, horses, goats and pigs. This dog is medium size but possesses a great agility and speed. His coat is normally long, straight or wavy. It is known for its intelligence and liveliness and, although very devoted to the shepherd and the herd, it can be somewhat wary of strangers and vigilant at night.

Portuguese Water Dog – Cão De Água Português[97]

The Portuguese Water is originally from the Portuguese region of the Algarve, from where the breed expanded to all around Portugal’s coast and is a fairly rare breed. Its original purpose was to herd fish into fishermen's nets, to retrieve lost tackle or broken nets, and to act as couriers from ship to ship, or ship to shore. A Portuguese Water Dog is first described in 1297 in a monk’s account of a drowning sailor who was pulled from the sea, but its origin remount to that ancient water dogs came to the peninsula from the Asian steppes with the Goths, a confederation of German tribes. These theories explain the similarities and how the German Poodle and the Portuguese Water Dog may have developed from the same ancient genetic pool. Their hair is either wavy or curly. Besides being a gun dog, Portuguese Waters are also capable of displaying herding abilities and have been used for that for a lot of time.

Pyrenean Mastiff – Mastín del Pirineo[81]

Pyrenean Mastiff is a large breed originally from the Aragonese Pyrenees in Spain. They have a heavy white coat with large darker spots. His dark bark comes from deep within his chest. Friendly towards humans and protective with children, calm, noble and very intelligent. When wolves disappeared the breed almost got extinct. To save them breeders crossed them with St. Bernard and Spanish Mastiff giving origin to the modern breed, which has now long coat instead of the short coat they used to have.

Pyrenean Mountain Dog – Chien de Montagne des Pyrénées[66]

Present in the Pyrenees from ages, known in the Middle Ages and used as a guardian of castles, it is mentioned by Gaston Phoebus in the 14th century. The main coat colour is white, and this dog is confident, gentle (especially with children), and affectionate. While territorial and protective of its flock or family when necessary, its general patience and loyal. The Pyrenean Mountain protects its flock by barking and being nocturnal.

Rafeiro of Alentejo – Rafeiro do Alentejo[82]

Rafeiro of Alentejo is a breed some authors say was originated from the crossing between Estrela Mountain and Spanish Mastiff, almost as ancient as them, while others believe it descends from molosser dogs from the Middle East. The breed got its name from the region of Alentejo where it appears in abundance. Rafeiro of Alentejo tends to be more vigilant at night, being very serious when guarding territory or any other property entrusted to it. It´s one of the biggest Portuguese breeds, being large-sized, powerful, rustic, sober and calm. For unknown reasons, this breed has low fertility.

Saint Miguel Cattle Dog – Cão de Fila de S. Miguel[94]

Robust and hardy, it is a cattle dog originating in the island of Saint Miguel in the Azores, also known as the “Cow Dog”. Its history is linked to that of the now extinct Terceira Sheep Dog. The coat is a brindle of brown (pale brown is described as fawn) or grey, with black. Although being a cattle dog it is equally good at guarding properties and people. Very intelligent and very receptive.

Sage Kushi[83]

Sage Kuchi or Afghan Shepherd dog is an Afghan livestock guardian dog, taking its name from the Kuchi people of Afghanistan, Afghan nomads. They are used, not only to protect flocks of sheep, goats, camels and other livestock from wild predators but also to protect the caravans from thieves. They possess a very rich gene pool, and the dogs adapt well to varying environments. They are large, often giant dogs, with a coat that can be short, medium, or long, backed by thick undercoat. The Kuchi dogs can roughly be divided into three types: The Mountain-type, the Steppe-type, and the Desert-type.

Spanish Mastiff – Mastín Español[84]

The Spanish Mastiff is a giant breed, originating in Spain. The breed is closely related to the seasonal moving in the livestock, and especially the Merino livestock which he already accompanied at the time of the “Mesta” (association, in the Middle-age, of breeders of the wandering herds) by defending them from wolves and other predators, all along their journey from one location to other and on the grazing pastures, in all seasons and sites. It is a very large and powerful dog, similar in appearance to the other Mastiff breeds, one of the biggest in the world. It is devoted to its family and may politely accept strangers. His bark is raucous, low pitched and deep, very sonorous, audible from a considerable distance.

Spanish Water Dog – Perro de Agua Español[98]

The origin of this breed remounts to the old “Barbet”, the French water dog, dates back several hundred years. Regardless of its exact origin, a woolly coated Shepherd Dog is documented to have been on the Iberian Peninsula by around 1100 CE. The densest population of Spanish Waters resides in Andalusia, Spain, where it has been used, not only for their main purpose of being a helper to the hunters of waterfowl and fishermen but also as a shepherd dog. Medium size, athletic, robust dog that is slightly longer than tall. Their learning ability is outstanding owing to his extraordinary mental grasp.

Spanish Wolf Dog – Perro Lobo Español[95]

Spanish Wolf Dog is a medium-big size dog originated from Spain. This breed is not yet recognized by any kennel club but is very appreciated in the northern regions of Spain, being used as a herding dog, guarding dog and even as a hunting dog. Hot-tempered dog, very intelligent, who is very attached to his owner, but wary of strangers. Their fur is wolf-like with different shades of grey, similar to the wolf with black spots along the legs. It is also used for hunting and very appreciated by the poachers for hunting in silence.

St. Bernard – St. Bernhardshund, Bernhardiner[85]

St. Bernard were first used as companion dogs and especially as rescue dogs for travellers lost in snow and fog, which they are famous for. The direct ancestors of the St. Bernard were the large farm dogs common in the mountains in Switzerland. There are two varieties of the St. Bernard, one with short and other with long hair. Both varieties are of considerable size, powerful, sturdy and muscular body, with impressive head and an alert facial expression. Their temperament changes from calm to lively, but he is always watchful.

Terceira Cattle Dog – Barbado da Terceira[96]

The Terceira Cattle probably evolved from dogs used in gathering wild cattle brought by the settlers as early as the 15th century. These dogs were taken to the Azores because of the need to control and gather several species of the newly introduced cattle. The Terceira Cattle Dog is a medium-sized dog, with a voluminous and robust look. Despite its joyful look, mostly due to its coat, the Terceira Cattle Dog is a strong-willed and assertive dog.

Tibetan Mastiff – Do-Khyi[86]

The Tibetan Mastiff is an ancient working breed of the nomad herders of the Himalaya and a traditional guardian of the Tibetan monasteries. It has been surrounded by great myth since its first discovery in antiquity, with historical reports praising its natural strength and

impressiveness - both physically and mentally. Even its bark has been described as a unique and highly treasured feature of the breed. The Tibetan Mastiff combines strength, robustness and endurance (in any climatic conditions). Slow to mature, only reaching its best at 2-3 years in females and at least 4 years in males.

Transmontano Mastiff – Cão de Gado Transmontano[87]

Transmontano Mastiff is settled in the Portuguese highlands, namely in Trás-os-Montes. In this mountainous area, the breed adjusted to the region's conditions and sheep and goat flocks, evolving until its morphological traits were defined, in perfect symbiosis with the conditions and work demanded. It is an exceptional guard when protecting flocks against wolf attacks, always attentive in its duty. Despite its size, it is a docile but reserved dog.

Turkmenistan Shepherd Dog - Alabai[88]

The legendary Alabai is a respected working breed in its native Turkmenistan, where it is highly valued for its purity and personality traits, as well as its superb resilience and guarding abilities. It is considered to be one of the oldest mastiffs in the world, having remained virtually unchanged for nearly 5000 years, with only occasional limited interference from humans. Like the other molossers, it is a muscular and agile mastiff, with a large head, powerful muzzle, well-adapted to the harsh climates of its native country and long working hours of guarding livestock.

Villano Bulldog – Villano de las Encartaciones[99]

They are used since ancestral times by cattle owners who own bovines of free ranging nature, to capture them in the steep valleys where they live. Many of the today's animals have little to do with those of formerly, many have Spanish Alano blood in their veins. Of rustic aspect and marked sexual dimorphism, with medium size, long and compact structure.

Yugoslavian Shepherd dog - Sharplanina[89]

It was named after the Sharplanina Mountain range where it is the most common and its origin remains a controversial issue in the scientific community. It seems likely that it came to Europe from Asia in the course of the prehistoric mass migrations. The original type of the breed has been maintained solely in such parts of the country where intense cattle breeding is still prevailing. Physically, the Sharplanina is a robust, well-proportioned dog. Of sturdy constitution, even disposition, good temperament, reliable, protective but not snappy.

6.2. Supplementary methodology

Salting out DNA extraction procedure from Miller et al. (1988) with some modifications used to extract genomic DNA from blood samples, in CIBIO/Inbio.

1) Place a small volume of well-ground tissue or 150 µL of blood in 600 µL of lysis solution in an eppendorf (temporary tube);
2) Add 10 µL (7-10 µL) of proteinase K;
3) Vortex;
4) Incubate at 55ºC to digest tissue (at least 3h);
5) Add 700µL of ammonia acetate (precipitation solution);
6) Vortex very well;
7) Centrifuge for 14min to 14,000rpm;
8) Remove 1mL of supernatant to another eppendorf (definitive tube) with 600µL of cold isopropanol (corresponds to 2/3 of the volume);
9) Mix well, inverting tubes several times;
10) Centrifuge for 12min to 14,000rpm;
11) Discard the supernatant with care not to reject the pellet (usually in the background);
12) Add 1000µL 70% cold ethanol and mix well by inverting the tubes;
13) Centrifuge 15min to 14000rpm;
14) Carefully reject ethanol (be careful not to lose DNA). If necessary, do a small centrifugation to reject the ethanol as much as possible and remove its excess with a yellow tip;
15) Allow to dry until all ethanol is evaporated with the tubes facing down;
16) Hydrate the DNA with TE / AE. Usually hydrated with 40-60 µL. This depends on the size of the pellet.

Supplementary Table 1 – Primer Mix for the Y chromosome (prepared in the lab).

Mix primers CrY	
MS34AF*-FAM (undiluted)	0.3 µL
MS34BF-VIC (1:10)	2.1 µL
MS34R (undiluted)	2.4 µL
650.79.2F-FAM (1:10)	0.8 µL
650.79.2R (undiluted)	0.8 µL
650.79.3F*-PET (undiluted)	1.25 µL
650.79.3R (undiluted)	0.7 µL
990.35F*-NED (undiluted)	0.7 µL
990.35R (undiluted)	3 µL
MS41BF-PET (1:10)	3 µL
MS41R (undiluted)	2.1 µL
VIC (undiluted)	4.25 µL
PET (undiluted)	0.8 µL
FAM (undiluted)	26.55 µL

Supplementary Table 2 – Primer Mix for the Kit AP in the autosomes (prepared in the lab).

Mix primers AP1		Mix primers AP2	
PEZ1	1 µL	FH2010	0.25 µL
PEZ5	0.8 µL	AHT132	0.65 µL
FH2079	0.2 µL	PEZ3	4.5 µL
PEZ8	0.65 µL	C27.442	0.4 µL

Supplementary Table 3 – Primer Mix for the Kit 2 in the autosomes (prepared in the lab).

Mix primers Kit 2	
AHT111	0.8 µL
FH2001	0.8 µL
CPH14	1.2 µL
C13.758	1 µL
C14.866	5 µL
AHT103	2.5 µL
C09.173	0.8 µL
VWF	0.7 µL
C20.253	3.2 µL
C04.140	0.8 µL

Supplementary Table 4 – Primer Mix for the Canine Genotypes Panel 1.1 Kit (designated as Kit 3).

Mix primers Kit 3	
AHT121	Canine Genotypes Panel 1.1 Kit (Thermo Fisher Scientific)
AHT137	
AHTh171	
AHTh260	
AHTk253	
AHTk211	
C22.279	
FH2054	
FH2848	
INRA21	
INU005	
INU030	
INU055	
REN162C04	
REN169D01	
REN169O18	
REN247M23	
REN54P11	
Amelogenin	

Supplementary Table 5 – Primer Mix for the Kit 4 in the autosomes (prepared in the lab).

Mix primers Kit 4	
C08.410	1 µL
CXX.459	0.8 µL
C08.618	0.8 µL
FH2161	1.4 µL
CPH9	0.8 µL
CPH2	1 µL
C20.446	2.1 µL
C09.474	1.4 µL
C22.763	1.8 µL
CPH5	0.8 µL
REN64E19	0.8 µL

Supplementary Table 6 – PCR program used for Kit AP.

Kit AP		
Temperature	**Time**	**Nº cycles**
95ºC	15'	1X
95ºC	30"	1X
58ºC	45"	20X (-0,1ºC/cycle)
72ºC	45"	1X
95ºC	30"	1X
56ºC	45"	15X
72ºC	45"	1X
95ºC	30"	1X
53ºC	45"	10X
72ºC	45"	1X
60ºC	30'	1X
10ºC	FOREVER	-

Supplementary Table 7 – PCR program used for Kit 2.

Kit 2		
Temperature	**Time**	**Nº cycles**
95ºC	15'	1X
95ºC	30"	1X
56ºC	45"	35X
72ºC	45"	1X
95ºC	30"	1X
53ºC	45"	8X
72ºC	45"	1X
60ºC	30'	1X
10ºC	FOREVER	-

Supplementary Table 8 – PCR program used for Kit 3.

Kit 3		
Temperature	**Time**	**Nº cycles**
98ºC	3'	1X
98ºC	15"	1X
60ºC	1'15"	40X
72ºC	45"	1X
72ºC	5'	1X
10ºC	FOREVER	-

Supplementary Table 9 – PCR program used for Kit 4.

Kit 4		
Temperature	**Time**	**Nº cycles**
95ºC	15'	1X
95ºC	30"	1X
60ºC	45"	7X (-0,5ºC/cycle)
72ºC	45"	1X
95ºC	30"	1X
57ºC	45"	22X
72ºC	45"	1X
95ºC	30"	1X
53ºC	45"	8X
72ºC	45"	1X
60ºC	30'	1X
10ºC	FOREVER	-

Supplementary Table 10 – PCR program used for Kit Cry.

Kit CrY		
Temperature	**Time**	**Nº cycles**
95ºC	15'	1X
95ºC	20"	1X
60ºC	20"	32X
72ºC	20"	1X
95ºC	20"	1X
53ºC	20"	8X
72ºC	20"	1X
60ºC	30'	1X
10ºC	FOREVER	-

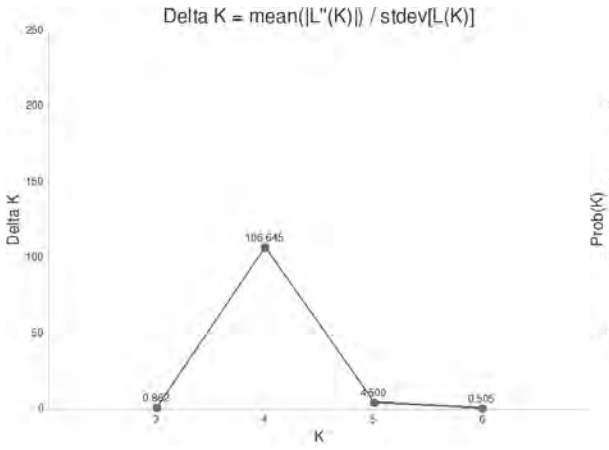
Supplementary Table 11 – PCR solution of the reagents in µL per tube.

PCR Solution	
MM (Qiagen)	5
MP	1
DNA	1
H2O	3
Total	10 µL/tube

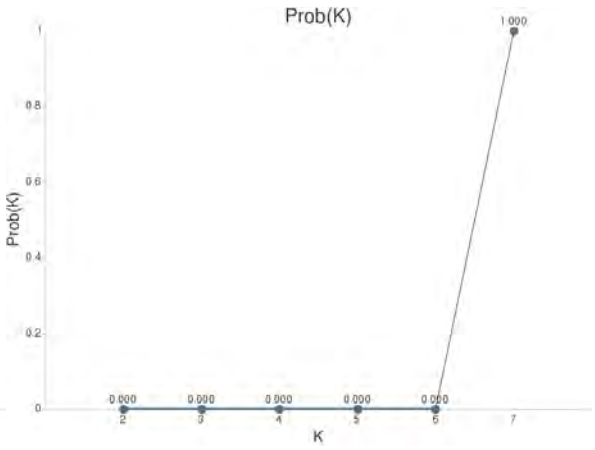
6.3. Supplementary results

Supplementary Table 12 – Exclusive and shared haplotypes among livestock guarding and herding breeds.

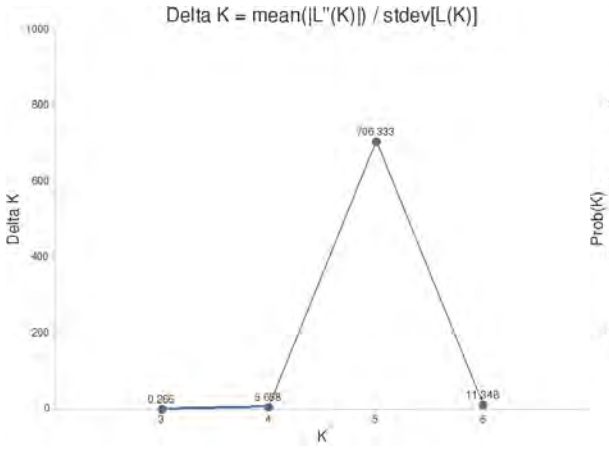
	LGDs	LHDs	Shared		LGDs	LHDs	Shared
HY1			X	HY11			X
HY2	X			HY12			X
HY3	X			HY13	X		
HY4	X			HY14			X
HY5		X		HY15			X
HY6			X	HY16	X		
HY7	X			HY17		X	
HY8		X		HY18	X		
HY9			X	HY19			X
HY10	X						



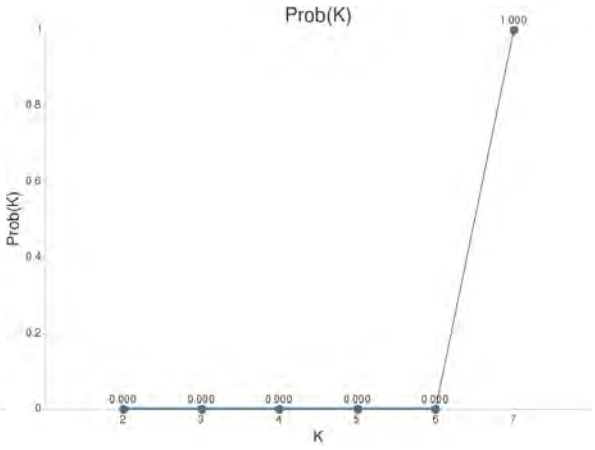
Supplementary Fig. 2 – DeltaK graph for the LGDs dataset analysis using STRUCTURE. The highest probability by Evanno method is achieved for K=2.



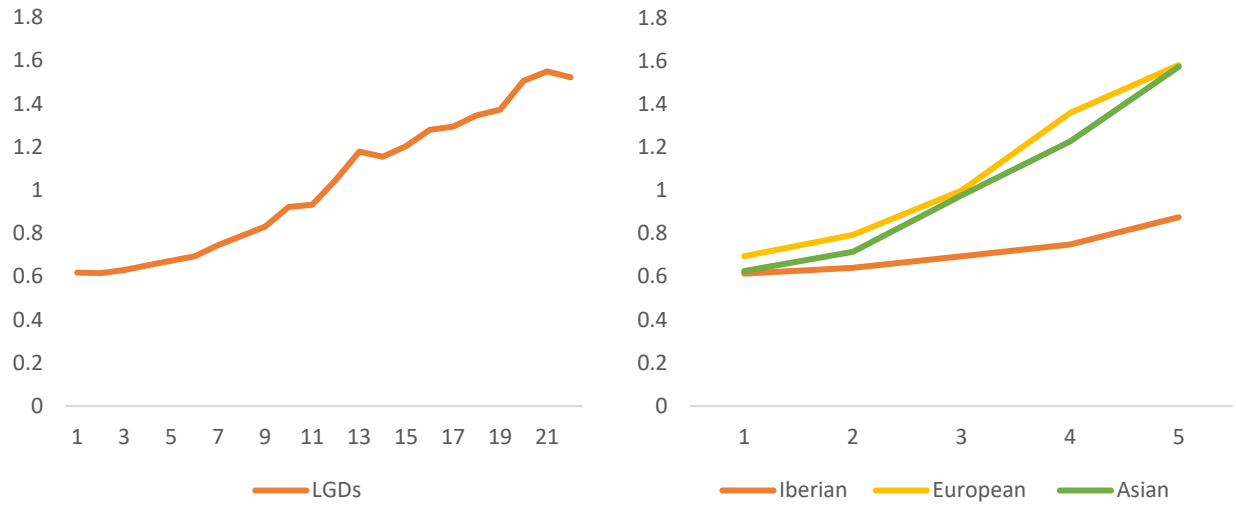
Supplementary Fig. 1 – Log likelihood of the data for each K, for the LGDs dataset analysis using STRUCTURE. The highest probability of the data is K=8.



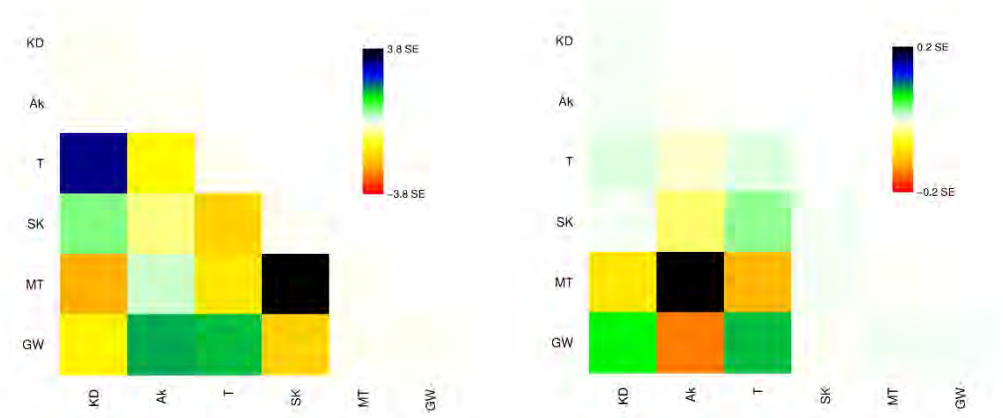
Supplementary Fig. 4 – DeltaK graph for the LHDs dataset analysis using STRUCTURE. The highest probability by Evanno method is achieved for K=4.



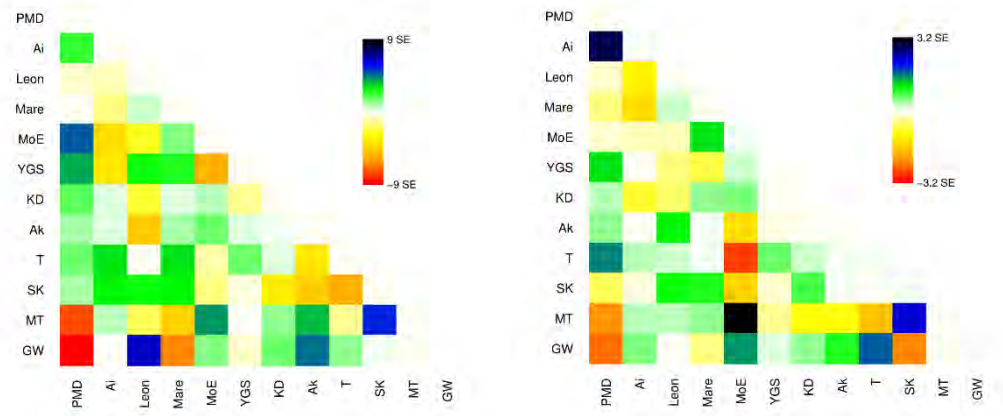
Supplementary Fig. 3 –Log likelihood of the data for each K, for the LHDs dataset analysis using STRUCTURE. The highest probability of the data is K=7.



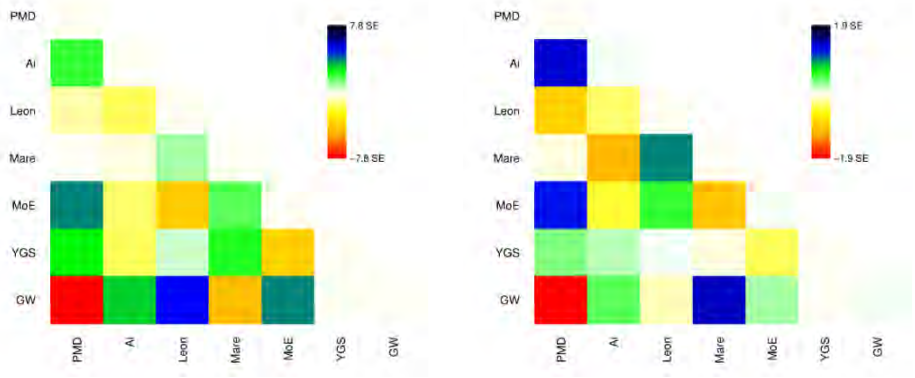
Supplementary Fig. 5 – Cross-validation plot for the whole analysis (left) and the sub-dataset analysis (right) with Admixture.



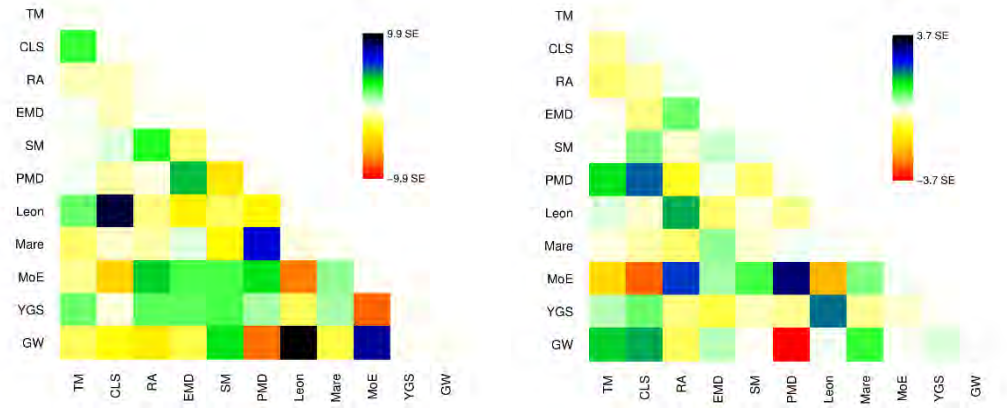
Supplementary Fig. 6 – Residual fit of the observed versus predicted squared allele frequency difference, expressed as the number of s.e. of the deviation, in the Asian group analysis. Analysis with 0 (left) and 3 (right) migrations inferred. Colours are in the palette on the right. Kurdish dog; Ak = Akbash; T = Turkmenistan Shepherd; SK = Sage Kuchi; MT = Tibetan Mastiff; GW = Grey Wolf.



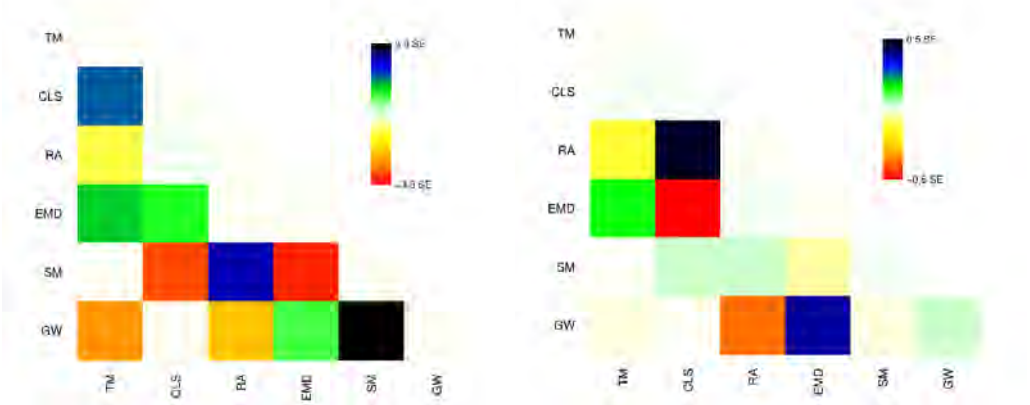
Supplementary Fig. 7 – Residual fit of the observed versus predicted squared allele frequency difference, expressed as the number of s.e. of the deviation, in the Estern group analysis. Analysis with 0 (left) and 9 (right) migrations inferred. Colours are in the palette on the right. PMD = Pyrenean Mountain Dog; Ai = Aidi; Leon = Leonberger; Mare = Maremma Sheepdog; YGS = Yugoslavian Shepherd; MoE = Molossus of Epirus; KD = Kurdish dog; Ak = Akbash; T = Turkmenistan Shepherd; SK = Sage Kuchi; MT = Tibetan Mastiff; GW = Grey Wolf.



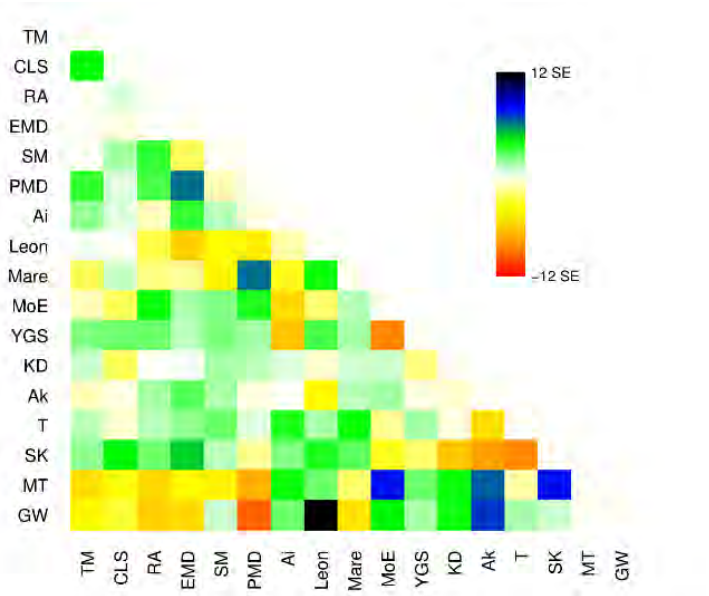
Supplementary Fig. 8 - Residual fit of the observed versus predicted squared allele frequency difference, expressed as the number of s.e. of the deviation, in the European group analysis. Analysis with 0 (left) and 4 (right) migrations inferred. Colours are in the palette on the right. PMD = Pyrenean Mountain Dog; Ai = Aidi; Leon = Leonberger; Mare = Maremma Sheepdog; YGS = Yugoslavian Shepherd; MoE = Molossus of Epirus; GW = Grey Wolf.



Supplementary Fig. 9 – Residual fit of the observed versus predicted squared allele frequency difference, expressed as the number of s.e. of the deviation, in the Western group analysis. Analysis with 0 (left) and 7 (right) migrations inferred. Colours are in the palette on the right. TM = Transmontano Mastiff; CLS = Castro Laboreiro dog; RA = Rafeiro of Alentejo; EMD = Estrela Mountain Dog; SM = Spanish Mastiff; PMD = Pyrenean Mountain Dog; Ai = Aidi; Leon = Leonberger; Mare = Maremma Sheepdog; SB = St. Bernard; YGS = Yugoslavian Shepherd; MoE = Molossus of Epirus; GW = Grey Wolf.



Supplementary Fig. 10 – Residual fit of the observed versus predicted squared allele frequency difference, expressed as the number of s.e. of the deviation, in the Iberian group analysis. Analysis with 0 (left) and 4 (right) migrations inferred Colours are in the palette on the right. TM = Transmontano Mastiff; CLS = Castro Laboreiro dog; RA = Rafeiro of Alentejo; EMD = Estrela Mountain Dog; SM = Spanish Mastiff; GW = Grey Wolf.



Supplementary Fig. 11 – Residual fit of the observed versus predicted squared allele frequency difference, expressed as the number of s.e. of the deviation, in the Whole dataset analysis, with 0 migrations inferred. Colours are in the palette on the right. TM = Transmontano Mastiff; CLS = Castro Laboreiro dog; RA = Rafeiro of Alentejo; EMD = Estrela Mountain Dog; SM = Spanish Mastiff; PMD = Pyrenean Mountain; Ai = Aidi; Leon = Leonberger; Mare = Maremma Sheepdog; SB = St. Bernard; YGS = Yugoslavian Shepherd; MoE = Molossus of Epirus; KD = Kurdish dog; Ak = Akbash; T = Turkmenistan Shepherd; SK = Sage Kuchi; MT = Tibetan Mastiff; GW = Grey Wolf.