Universidade de Aveiro Departamento de Biologia
2019

Ludmilla Paixão Blaschikoff

Dogs *(Canis lupus familiaris)* from the Iberian Peninsula dated to the Chalcolithic period: a genomic approach

Cães *(Canis lupus familiaris)* da Península Ibérica do período Calcolítico: uma abordagem genómica

DECLARAÇÃO

Declaro que este relatório é integralmente da minha autoria, estando devidamente referenciadas as fontes e obras consultadas, bem como identificadas de modo claro as citações dessas obras. Não contém, por isso, qualquer tipo de plágio quer de textos publicados, qualquer que seja o meio dessa publicação, incluindo meios eletrónicos, quer de trabalhos académicos.

Universidade de Aveiro Departamento de Biologia
2019

Ludmilla Paixão Blaschikoff

Dogs *(Canis lupus familiaris)* from the Iberian Peninsula dated to the Chalcolithic period: a genomic approach

Cães *(Canis lupus familiaris)* da Península Ibérica do período Calcolítico: uma abordagem genómica

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Ecologia Aplicada, realizada sob a orientação científica da Professora Doutora Ana Elisabete Godinho Pires, Investigadora do Centro de Investigação em Biodiversidade e Recursos Genéticos – Rede de Investigação em Biodiversidade e Biologia Evolutiva (CIBIO-InBIO/Laboratório de Arqueociências da Direcção Geral do Património Cultural) e Professora Auxiliar da Faculdade de Medicina Veterinária da Universidade Lusófona de Humanidades e Tecnologias (FMV, ULHT) e coorientação do Doutor Eduardo Manuel Silva Loureiro Ferreira, Investigador do Centro de Estudos do Ambiente e do Mar e do Departamento de Biologia da Universidade de Aveiro e do Doutor Octávio Manuel Ribeiro Serra, Investigador do Instituto Nacional de Investigação Agrária e Veterinária (INIAV).

Dissertação de Mestrado desenvolvida no âmbito do Projeto Woof - Tracing the origins and evolutionary paths of the Iberian and the Maghreb Dog, financiado pela FCT/MCTES e FEDER.

o júri

presidente

Professor Doutor Carlos Manuel Martins Santos Fonseca
Professor Associado com Agregação, do Departamento de Biologia da Universidade de Aveiro

Professora Doutora Catarina Ginja
Professora Auxiliar Convidada da Faculdade de Ciências da Universidade do Porto e Investigadora do Centro de Investigação em Biodiversidade e Recursos Genéticos – Rede de Investigação em Biodiversidade e Biologia Evolutiva (CIBIO)

Professora Doutora Ana Elisabete Godinho Pires
Investigadora do Centro de Investigação em Biodiversidade e Recursos Genéticos – Rede de Investigação em Biodiversidade e Biologia Evolutiva (CIBIO-InBIO/Laboratório de Arqueociências da Direcção Geral do Património Cultural) e Professora Auxiliar da Faculdade de Medicina Veterinária da Universidade Lusófona de Humanidades e Tecnologias (FMV, ULHT)

agradecimentos

Agradeço a toda a minha família pelo apoio e presença, em especial aos meus pais e meu irmão. Ao meu pai, Ruslan, que sempre colocou o estudo dos seus filhos em primeiro lugar, me apoiou e abriu-me as portas para estudar na Europa. À minha mãe Tamara que sempre me ensinou os valores e virtudes da vida; mãe e melhor amiga, enxugou minhas lágrimas nos momentos de aflição e me acalmou com seu jeito doce que só quem a conhece sabe. Ao meu irmão, Nikolai, que sempre me inspirou com sua inteligência. Eu amo-vos muito.

Ao meu avô, João Gonçalves, que sempre disse aos seus netos estudarem para depois não estarem a carregar caixas nas costas. Conselho este que nunca esquecerei. Durante o meu mestrado e distante da minha terra natal, meu querido e amado avô partiu desse mundo. Mas àqueles que permanecem vivos no coração de alguém nunca morrem. À minha avó, Maria José, mulher forte, guerreira e muito religiosa, que me inspira pela sua energia e por sempre colocar a família acima de tudo. À minha prima, Lucyanna, que sempre me enviava mensagens positivas e engraçadas, "colorindo" o meu dia. Aos meus tios, Tânia, Williams e Telma, e primo Bruno, que mesmo distantes estão sempre me apoiando.

Ao meu namorado, Vasco, que Portugal me apresentou e já são 3 anos de muito amor e companheirismo. Sem o seu apoio, paciência e conselhos, essa tese não teria sido escrita. E não podia esquecer da sua família muito querida, Maria Augusta, Paulo Jorge e Marta, que tenho como uma segunda família e que sempre estiveram presentes para me ajudar no que fosse necessário.

A todos os animais que já tive, em especial ao Costelinha, Bilu, Tandera, Pepita, Máximus, Liz, Gatuna, Kitty, Ling, Alejandro e Timon, que me ensinaram a amar os animais e a cuidar da natureza.

Á minha orientadora Doutora Ana Elisabete, que estuda a evolução dos cães ancestrais da Península Ibérica. Sem ela, o passado dos cães em Portugal continuaria um mistério e o desenvolvimento desta tese ainda estaria no futuro. Obrigada pelas inúmeras horas de conversas para desenvolvermos este trabalho, a paciência para tirar minhas dúvidas e por me dar a oportunidade de ingressar nesse belíssimo projeto.

Ao meu coorientador Doutor Eduardo Ferreira, que aceitou o desafio de ser meu orientador interno e partilhar o seu conhecimento em genética populacional.

Também quero agradecer a todos do INIAV – Instituto Nacional de Investigação Agrária e Veterinária de Santarém – que me acolheram com muito carinho durante os dois meses que lá estive a fazer o estágio de bioinformática. Agradeço em especial à Doutora Sílvia Alves e ao meu coorientador, o Doutor Octávio Serra. À Sílvia, uma pessoa que admiro muito, cheia de conhecimento e amante dos animais. Obrigada pelas boleias que me deu e pelas conversas "cabeça" que tínhamos durante o trajeto para o INIAV. Ao Octávio, exemplo de carisma e dedicação, que sempre esteve disponível para me ajudar quando surgiam dúvidas e me ensinou e descomplicou a complicada bioinformática. Agradeço também à Doutora Fernanda Simões (INIAV) por sugerir a participação do Doutor Octávio Serra para o ensino da bioinformática.

Por último e não menos importante, agradeço à colaboração da Doutora Sílvia Guimarães, investigadora do CIBIO-inBIO, responsável pela extração e sequenciação do DNA das amostras aqui analisadas.

Todos, cada um ajudando de uma forma, tornaram possível este estudo existir. Com o vosso apoio, cruzando e modificando a minha vida, sinto-me orgulhosa da minha tese de Mestrado.

palavras-chave

Canis lupus, zooarqueogenética, DNA ancestral, Next Generation Sequencing, bioinformática, Ibéria

resumo

Os cães existem na Península Ibérica pelo menos desde o Paleolítico Superior; o resto arqueológico mais antigo data há cerca de 16,000 AP (Erralla, Espanha). Existem diferentes teorias sobre a origem dos cães na Europa. Estudos anteriores indicam que os cães podem ter chegado à Europa a partir de uma população domesticada de lobos oriundos da Ásia Oriental, ou a partir de duas populações de lobos geneticamente distintas da Eurásia Oriental e Ocidental, domesticadas independentemente, e que mais tarde, a população de cães da Eurásia Oriental se espalhou e substituiu parcialmente a população da Eurásia Ocidental.

Um estudo recente focando na composição genética de 6 cães Ibéricos do período Mesolítico, sugeriu que uma domesticação local na Península Ibérica pode ter ocorrido na Europa pré-Neolítica. Considerando o debate mantido sobre a origem dos cães, é crucial desvendar a composição genética de populações passadas e periféricas da Europa – usando métodos específicos para recuperar e analisar o DNA ancestral, de diferentes períodos, a fim de investigar a origem e a trajetória evolutiva dos cães no seu global. Nomeadamente, pode revelar-se importante por fornecer dados sobre uma possível contribuição do lobo Ibérico para a origem dos primeiros cães Ibéricos e informação genómica potencialmente útil para a detecção de eventos de hibridação histórica entre o cão e o seu parente selvagem, o lobo Ibérico – uma subespécie endêmica e atualmente considerada "Em perigo" de extinção. Esta informação pode ser englobada aquando a definição de medidas de gestão e conservação futuras para a espécie selvagem lobo Ibérico.

Neste trabalho, uma abordagem genómica (Next Generation Sequencing, NGS) foi a escolhida para recuperar sequências do genoma mitocondrial (mt) e nuclear de *Canis* de três sítios arqueológicos Ibéricos datados do Calcolítico [ca. 5,000-4,000 anos atrás]: dois cães de Leceia em Oeiras, Portugal; dois cães de Casetón de La Era em Valladolid, Espanha; e um lobo de Penedo de Lexim em Mafra, Portugal. Utilizando as ferramentas de bioinformática actuais, esses genomas foram identificados e compilados. Além disso, para entender a relação de populações passadas/modernas, construiu-se uma rede filogenética (baseada num fragmento parcial da região controlo do mtDNA) reunindo 254 sequências de *Canis*, bem como uma árvore filogenética de 23 mitogenomas de *Canis* disponíveis em bases dados públicos.

Embora a recuperação e análise do genoma nuclear sejam um maior desafio se proveniente de amostras ancestrais, este foi investigado para a identificação do sexo molecular desses 5 espécimes.

Relativamente ao estudo dos cães pré-históricos da Ibéria, esta é a primeira tentativa de aplicar com sucesso o método NGS para investigar a sua composição genética. Neste estudo, foi possível: gerar sequências do genoma mitocondrial (com 1x a 17x de cobertura) e recuperar entre 0.09% e 3.75% do genoma nuclear endógeno das 5 amostras do Calcolítico; identificar haplótipos de DNA mitocondrial e atribuí-los a dois (A e C) dos quatro principais haplogrupos descritos para os cães (A, B, C e D); gerar dados genómicos de um lobo Ibérico do Calcolítico que, tanto quanto investiguei, constituem os primeiros dados genómicos de um espécime de lobo da Ibéria e desta cronologia. Os resultados mostram que os cães Ibéricos do Calcolítico apresentavam a mesma frequência de haplótipos do Haplogrupo A (Hg anteriormente presente neste território, em contraste com as outras regiões da Europa), bem como do Haplogrupo C (já presente em outras regiões da Europa desde o Paleolítico).

keywords

Canis lupus, zooarchaeogenetics, ancient DNA, Next Generation Sequencing, bioinformatics, Iberia

abstract

Domestic dogs exist in the Iberian Peninsula at least since the Upper Late Palaeolithic; the oldest remain dated to 16,000 BP years old (Erralla, Spain). There are different theories about the origins of European dogs. Previous studies indicated that dogs may arrived in Europe from an Eastern Asia domesticated population of wolves, or that two genetically distinct wolf populations in Eastern and Western Eurasia may have been independently domesticated, and that afterwards the Eastern dog population spread and partially replaced an indigenous Western Eurasian dog population.

A recent study focusing in the genetic composition of 6 Mesolithic Iberian dogs reported that a local domestication in the Iberia Peninsula may have occurred in pre-Neolithic Europe. Considering the debated origin of Iberian dogs, it is crucial to unravel the genetic composition of past European peripheral populations - using specific methods to recover and analyse ancient DNA, from different periods in order to further investigate their origins and evolutionary trajectories. Additionally, it may prove important to provide data on a possible contribution of the Iberian wolf to the origin of the first Iberian dogs and genomic information potentially useful for the detection of historical hybridization events between the dog and its wild relative, the Iberian wolf – a subspecies and an endemism currently considered “Endangered”. This information can be included in the definition of future management and conservation measures for the wild Iberian wolf species.

In this work, a genomic approach (Next Generation Sequencing, NGS) was carried out to recover mitogenome and nuclear genomic data of *Canis* from three Iberian archaeological sites dated to the Chalcolithic [ca. 5,000-4,000 years BP], in particular: two dogs from Leceia in Oeiras, Portugal; two dogs from Casetón de La Era in Valladolid, Spain; and one wolf from Penedo de Lexim in Mafra, Portugal. Using the most up-to-date bioinformatic tools, their mitochondrial (mt) and nuclear genomes were sequenced. In addition, to understand the relationship of past/extant populations, a phylogenetic network (based on a partial fragment of the mtDNA control region) comprising 254 *Canis* sequences, as well as a phylogenetic tree of 23 *Canis* mitogenomes, publicly available, were constructed. Furthermore, the nuclear genome, although more challenging to recover and analyse from ancient samples, was investigated to molecularly assess the sex of these 5 *Canis* specimens.

Regarding ancient Iberian dogs, this is the first attempt to successfully apply NGS methods to investigate their genomic composition. In this study, it was possible to: generate the draft of mitochondrial genomes (coverages ranged between 1x and 17x) and recover between 0.09% and 3.75% of endogenous nuclear genomic data of these 5 *Canis* specimens; identify mitochondrial DNA haplotypes and assign those to 2 (A and C) of the four major dog haplogroups described (A, B C and D); generate genetic data from a Chalcolithic wolf - to the best of my knowledge this is the first genomic data available from an Iberian wolf specimen from this chronology. The results shown that the Chalcolithic Iberian dogs had about the same frequency of Haplogroup A (previously present in this territory, but contrasting with other European regions), as well as of the Haplogroup C (already present in other European regions since the Paleolithic).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF BOXES

LIST OF APPENDICES

POSTER [1]

Blaschikoff *et al*. **"Bioinformatic tools in the study of ancient dogs – an Iberian case study"**. VIII Bioinformatics Open Days, University of Minho, Braga, Portugal, 20-22 February of 2019.

POSTER[1,2]

Blaschikoff *et al*. **"Iberian Chalcolithic *Canis*: a genomic approach to know them better"**. 8th meeting of the ICAZ Archaeozoology, Genetics, Proteomics and Morphometrics (AGPM) working group, Natural History Museum of Paris, France, 17-18 October of 2019.

[1]This communication is a result of this thesis.
[2]Poster abstract submitted to the 8th ICAZ AGPM working group meeting.

1 INTRODUCTION

1. Introduction

1.1 The history of animal domestication

Domestication of animals and plants started at the Late Pleistocene (MacHugh *et al.* 2016), beginning with the domestication of wolf (*Canis lupus,* Linnaeus 1758), followed, much later, by the ancestral of livestock species (e.g. goats, sheep, cattle and pigs) and crops (e.g. rye, wheat). Despite the existence of thousands species in the world, the number of animal species that man has been able to domesticate does not exceed four dozens (Larson & Fuller 2014), fact that proves that the necessary circumstances for this process to take place appear to have happened rarely during the past.

There is not an unanimously accepted definition of domestication because of the diverse array of different relationships between humans and animals and plants that fall within the general rubric of domestication (Vigne *et al.* 2005). The process of domestication may not have been a deliberate act, but the result of a coevolutionary process with multiple stages along three different pathways (commensal, prey and directed; see Zeder 2012b).

Nonetheless, domestication triggered a rapid and profound shift in the evolution, ecology and demography of both humans and domesticated species. After the process of domestication took place, biological changes continued through gene flow between domestic and wild populations, relaxation of natural selective pressure, and later artificial selection pressures driven by humans, resulting in the appearance of new traits by mutation (Larson & Fuller 2014). In order to decipher the key differences that allow to distinguish between the wild and domestic forms, many studies have concentrated in different approaches/methods to study archaeological records (Vigne *et al.* 2005; Zeder 2006; MacHugh *et al.* 2016).

1.2 Domestication of *Canis lupus:* one event, many theories

1.2.1 Man's oldest friend: from wolf to dog.

The mystery of resolving the complexity of the origin of dogs (*Canis lupus familiaris*, Linnaeus 1758) began formally with Charles Darwin, in 1868, when in his book entitled "The

Variation of Animals and Plants under Domestication", he wondered whether dogs had evolved from a single species or from an unusual mating between a wolf and a jackal (Darwin 1885). Part of the answer for this question only came in the late 1980s, when morphological (Olsen 1985; Benecke 1987; Clutton-Brock 2016) and genetic (Wayne *et al.* 1992; Vilà *et al.* 1997; Savolainen *et al.* 2002) analysis finally confirmed that dogs had descended from gray wolves (*Canis lupus*; both share 99.96% of their nuclear DNA (Lindblad-Toh *et al.* 2005)) – but not from the extant gray wolf (Figure 1). Instead, they would have descended from a now-extinct Late Pleistocene wolf population, as extant wolves are not closely related to the wolves that were first domesticated. Even tough, the closest living relative of the dog is the extant gray wolf (Lindblad-Toh *et al.* 2005; Freedman *et al.* 2014; Fan *et al.* 2016; Thalmann & Perri 2018).

However, some questions are still unanswered: "where dogs first appeared?"; "when this happened?"; and "what is the best way to find these answers?". Knowing the answer to these questions will not only end with an old search, but also will improve our knowledge

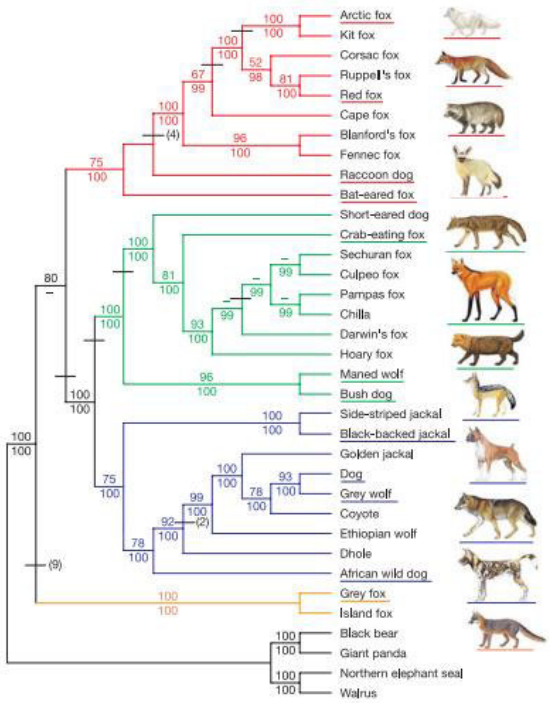


Figure 1. Canid phylogenetic tree based on 14948 bp of intron and exon sequences. Above and below the internodes are two different measures of support: bootstrap and bayesian value, respectively. The colors groups identify the red-fox-like clade (red), the South American clade (green), the wolf-like clade (blue) and the grey and island fox clade (orange). Underlined species names are represented with corresponding illustrations. Dogs and wolves are related species and descend from a common ancestor. Modified and reproduced with permission of the authors (Lindblad-Toh et al. 2005)

of pre-history humans and the development of civilization, given that dogs were the first human's domesticate – before any other plant or other animal – (Morey 1994), and had a profound influence on the course of human history, such as the transition from hunter-gatherers to farmers, the peopling of the Americas, the spread of pastoralism into Europe, and, most recently, European colonialism throughout the Americas and elsewhere (Shannon et al. 2015).

Hagner (2018), mentions that domestication is a gradual process that occurs along a *continuum*, where no clear separation line between wolves and dogs can be determined (Figure 2) and arbitrary names can be given to some stages to better understand the processes. Irving-Pease *et al.* (2018) argue that "*Timing domestication should therefore focus on questions related to the numerous changes in the way humans interacted with domesticates, how those relationships varied in time and space, the relative intentionality of human actions and the genetic and morphological effects on the taxa in question*".

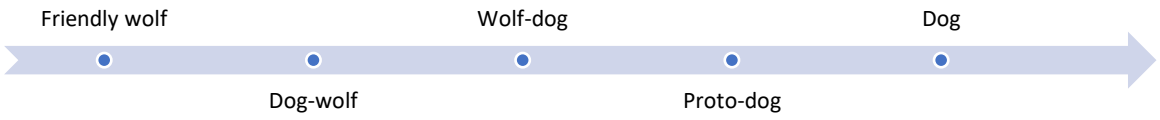


Figure 2. Timeline of dog domestication stages indicated by Hagner (2018).

Most experts believe that domestication began with a wolf initiative. In other words, they self-domesticated (Budiansky 1992; Morey 1994). In the time of human hunter-gatherers, thousands of years ago, a bolder and less suspicious population of wolves came into initial contact with humans to feed on the remained carcasses that were discarded at the edges of human-hunter settlements, founding a new niche with a convenient food supply. After only few generations, the wolves of this bolder lineage reproduced and generation after generation became associated with humans. Thus, confidence arose between the two species and man probably started to realize that the presence of the wolf could be useful as protection against other large carnivores and helping them to take down prey, eventually becoming permanent hunting partners. Therefore, the active phase of domestication - where man had the initiative – began by breeding early canines to be better hunters and guardians (Morey 1994; Grimm 2015).

Back to Palaeolithic period, humans hunted animals using heavy stone axes and spears. During Mesolithic period, the development of an improved arrow that allowed to hunt at long-distance, in partnership with dogs that could help to track down and bring wounded animals, probably enhanced the success in hunting (Vigne 2007; Clutton-Brock 2016). Prehistoric dogs may have been used to transport heavy pack on their backs by dragging carts, as a paleopathology study on Chalcolithic and Bronze age dogs skeletons revealed flattening of the dorsal tips of ancient dog vertebrae (Albizuri *et al.* 2011; Liesau Von Lettow-Vorbeck *et al.* 2014; Grimm 2015).

There were some morphological changes accepted to be associated with early dog domestication events (Figure 3):

- Reduction in body and head size: the new habitat chosen by the scavenger wolves led to changes in breeding strategies (accelerated maturation, larger littler sizes and shortened generation time) in face to changes in selective pressure (new conditions of food and water availability, low interspecific competition, relaxation in competition/predation pressures, increase in intraspecific competition), resulting in size reduction (due truncation of the growth period) (Tchernov & Horwitz 1991; Morey 1992; Clutton-Brock 2016). Paedomorphism, the retention of juvenile features in sexually mature adults, a criteria long time accepted to distinguish dog from wolf (Morey 1994; Waller *et al.* 2013), currently have been rejected (Drake 2011), however whether early dogs resembled juvenile wolves is not excluded. At a late stage of domestication, the selection of smaller dogs was also related to tameness and submission to man (Morey 1994);
- Wider Snout: increases in snout width as a consequence of a shape change of the mid-face in dogs (Drake 2011; Schmitt & Wallace 2012).
- Development of an angle between the nasal/maxilla bones and the forehead bones, called a “frontal stop” (Drake 2011).
- Carnassial size reduction: reduction of the carnassial teeth in the earliest dogs (Janssens *et al.* 2019). Mandibular tooth crowding (Morey 1992; Germonpré *et al.* 2012; Clutton-Brock 2016), however, this character is no longer a reliable indicator

to assign species identification of early dogs, since high levels of tooth crowding in ancient wolves are also reported (Ameen *et al.* 2017);

- Changes in fur color (Anderson *et al.* 2009; Ollivier *et al.* 2013; Shannon *et al.* 2015; Clutton-Brock 2016) and texture, shape of ears (floppy ears), eye color, tail length and its curvature can be also found in early dogs (Trut 1999; Wilkins *et al.* 2014).

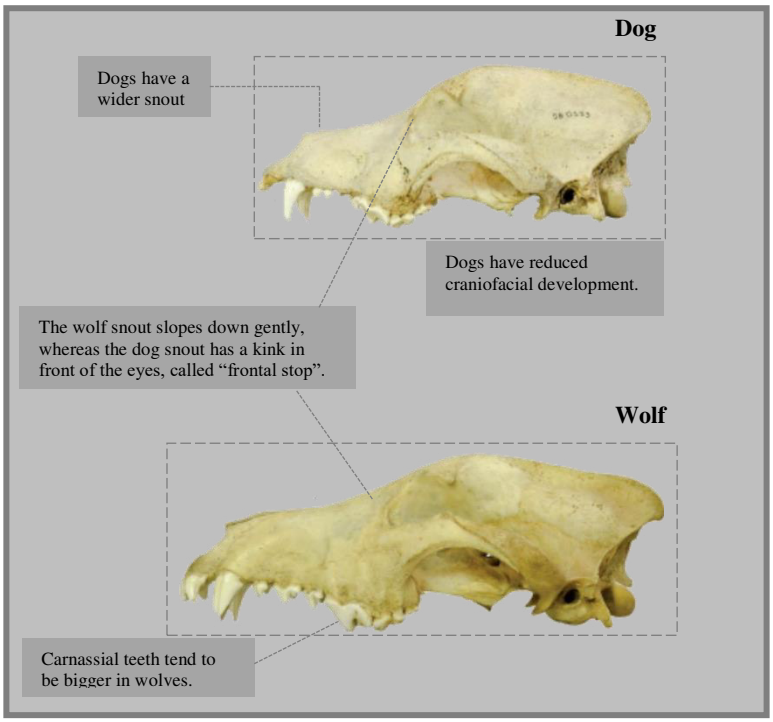


Figure 3. Cranium characteristics used to distinguish wolves from dogs. Modified and reproduced with permission of the authors (Grimm 2015)

In the year of 2005, when the first dog genome was completely sequenced and published (Lindblad-Toh *et al.* 2005), researchers could have a glimpse in the genetics basis for dog domestication. Researchers could associate some genes found in the dogs' genome to domestication. Those genes are mostly involved in nervous system development and function, supporting the hypothesis that first stage of domestication was selecting for behaviours, such as reduction in aggression, tameness and submission to live in human coexistence (Trut 1999; Akey et al. 2010; Pendleton et al. 2018). Other genes are involved in starch and fat metabolism, reflecting genetic changes an adaptation to a new diet available as a consequence of the development of agriculture (Axelsson et al. 2013).

1.2.2 Archaeological evidence – earliest dog remains

Currently, genetics and archaeological fields have come together to allow for a better understanding of the dog origins and evolution. Domestication was not abrupt, resulting in morphological differences that were not very apparent between the first domestic dogs and their wild wolf ancestors during the early stages of wolf domestication (Larson *et al.* 2012; Hagner 2018).

Before the advent of the molecular genetic toolbox, osteometric, morphometric and dating analyses were conducted to identify species. However, it is not possible to confirm when and where the domestication of the wolf happened based only on the morphological analysis of fossil remains: 1) wolves and the earliest dogs were likely very similar morphologically, making it difficult sometimes to distinguish their bones; 2) wolves in pre-historic times used to have a much broader distribution, making it difficult to classify remains solely on the basis of geography; 3) until date, few canid fossil remains have been found, resulting in a temporally and geographically fragmented record (Freedman & Wayne 2017).

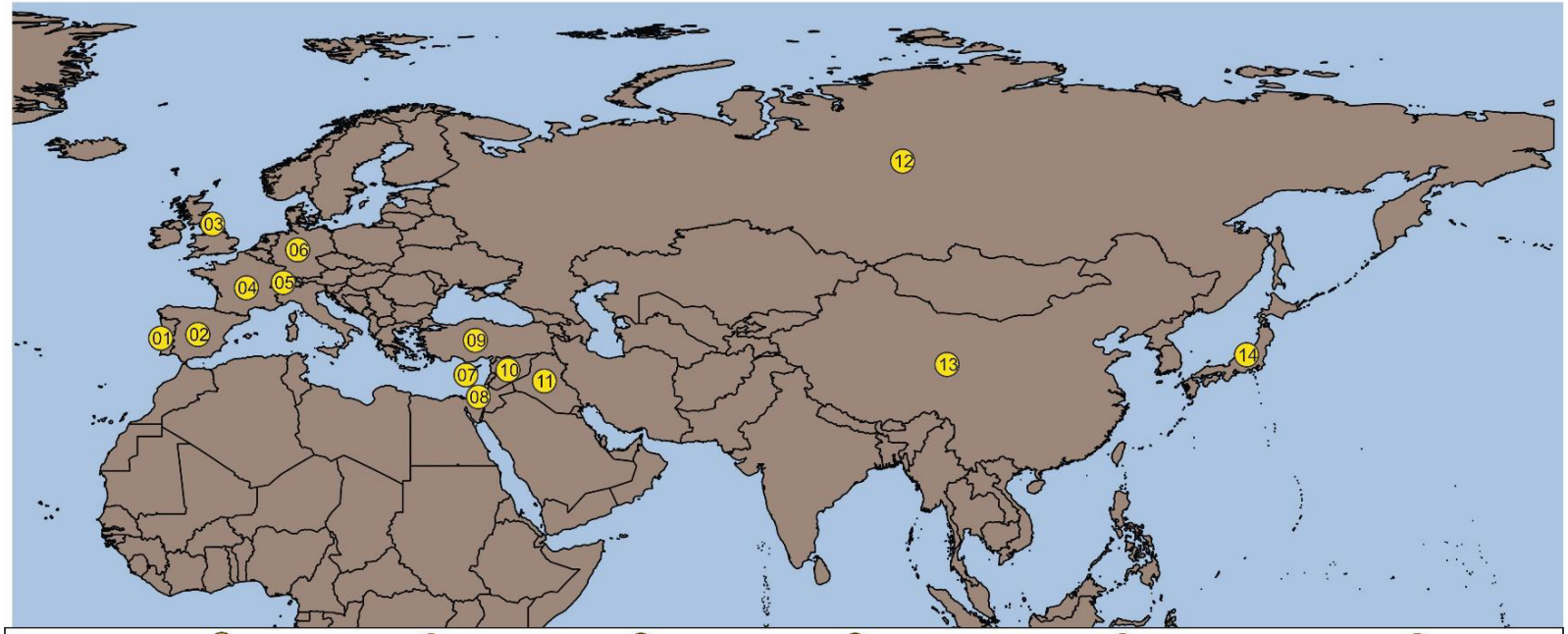
Dogs from Eurasia

Well documented remains of early domestic dogs come from the Late Pleistocene and Early Holocene periods (see Appendix I), with few disputed dogs remains dated prior to the Last Glacial Maximum (22,000-19,000 BP (Yokoyama *et al.* 2000)) (Sablin & Khlopachev 2002; Germonpré *et al.* 2009, 2012; Druzhkova *et al.* 2013). Dogs were well established across Eurasia before the end of the Late Pleistocene, before the advent of agriculture and domestication of other animals, indicating that the earliest dogs arose when humans were hunter-gatherers and not farmers (Davis & Valla 1978; Napierala & Uerpmann 2012; Freedman & Wayne 2017).

Dog remains have been found at several archaeological sites located in distinct geographic areas, raising the hypothesis of multiple and independent processes of domestication of the wolf before the Neolithic period (Vilà *et al.* 1997; Pionnier-capitan 2010; Frantz *et al.* 2016). According to zooarchaeology, the oldest archaeological evidence of domestic dog

comes from the upper Palaeolithic 30,000 years-ago (Predmostí dog; Germonpré *et al.* 2012), although the first remains confidently assigned to dogs appear in Europe and in the Middle East only by the end of the late Glacial period, ca. 14,000 and 12,000 years-ago, respectively (Kesslerloch and Ain Mallaha dogs; Davis & Valla 1978; Napierala & Uerpmann 2012).

Several studies place the first steps of wolf domestication in East Asia, Central Asia, Southeast Asia, Middle East, or Western Europe (see Table 1) before the Neolithic transition (Freedman *et al.* 2014). From the following Neolithic period, remains of dogs were abundant in archaeological sites from many parts of the world and status assignment in these remains becomes easier because dog-like features (e.g. small size and skull and mandible shortening) were fully developed (Clutton-Brock 2016). A brief map description of each Palaeolithic and Mesolithic archaeological site is shown above (Figure 4). However, a more complete register of the main archaeological finds worldwide (disputed and undisputed) can be found at supplementary material of Larson *et al.* (2012) and Ollivier *et al.* (2018).



EUROPE	01 PORTUGAL	02 SPAIN	03 ENGLAND	04 FRANCE	05 SWITZERLAND	06 GERMANY
	Poças de São Bento	Erralla	Seamer Carr	Le Clouseau	Hauterive-Champréveyres	Oelknitz
	Muge		Star Carr	Montespan	Kesslerloch Cave	Teufelsbrucke
				Pont d'Ambon		Bonn–Oberkassel
				Saint-Thibaud-de-Couz		Kniegrotte
ANATOLIA, LEVANT	07 CYPRUS	08 ISRAEL	09 TURKEY	10 SYRIA	11 IRAQ	
	Shillourokambos	Hayonim Terrace	Cafer Höyük	Tell Mureybet	Palegawra cave	
	Klimonas	Ain Mallaha	Çayönü	Tell Aswad		
EAST ASIA	12 RUSSIA	13 CHINA	14 JAPAN			
	Tumat	Nanzhuangtou	Natsushima Shell			
	Ushki I		Kamikuroiwa			
	Zhokhov					

Figure 4. Geographical Location (numbers 1 to 14) of the Palaeolithic and Mesolithic archaeological sites in Eurasia from where the earliest undisputed dogs remain were excavated. Remains excavated from each site is described in detail in Appendix I.

Dogs from the Americas

Genetic evidence indicates New World dogs, i.e. the Americas, originated from Old World dog lineages (Leonard *et al.* 2002). When the first humans who colonized America crossed the strait of Bering into the New World ca. 12,000-14,000 years ago, it is believed that they brought with them multiple lineages of Eurasian dogs (Leonard *et al.* 2002). The oldest dog remains until date was found in Danger Cave, Utah and is dated from ca 10,000-9,000 years ago (Grayson *et al.* 1988).

Canids remains with ambiguous status and morphology

- Goyet Cave, Belgium: a 36,000 cal BP skull was first indicated to belong to a Palaeolithic dog-like individual (Germonpré et al. 2009). However, its mitochondrial DNA did not match any modern wolf nor dog, placing it as an ancient sister-group rather than a direct ancestor of modern dogs. It may represent an unsuccessfully domestication event or phenotypically and genetically distinct wolves, perhaps from an extinct wolf lineage (Pionnier-Capitan *et al.* 2011; Thalmann *et al.* 2013; Drake *et al.* 2015). Crockford and Kusmin (2012) argue "that the 'Palaeolithic dogs' described by Germonpré et al. (2009, 2012) […] may simply be rather 'short-faced' wolf individuals that lived within a population of typical wolves that interacted in various ways with human hunters".
- Predmostí, Czech Republic: Three complete skulls were identified as Palaeolithic dogs estimated to be ca 27,000 BP (Germonpré *et al.* 2012). According to Larson et al 2012, although these skulls exhibit dog-like traits, they could also belong to an extinct population of wolves, as also argued above by Crockford and Kusmin (2012).
- Taimyr, Siberian: The 35,000 year-old siberian wolf, belonged to a population that diverged from a now extinct common ancestor of gray wolves and domestic dogs before the peak of the Last Glacial Maximum (22,000-19,000 BP (Yokoyama *et al.* 2000)) providing insight into wolf-dog divergence (Skoglund *et al.* 2015).
- Eliseyevichi I, Western Russia: Morphological measurements assigned two skulls found in this Upper Palaeolithic site to large dogs (Sablin & Khlopachev 2002) dated to ca. 16,945-16,190 cal BP (Pionnier-Capitan et al. 2011). New genetic and morphometric studies (Thalmann *et al.* 2013; Drake *et al.* 2015; Janssens *et al.* 2019)

questioned the validity of traditional measurements used for taxonomic identification, claiming that Eliseevichi I dogs are in fact wolves.

- Razboinichya Cave, Altai Mountains, Southern Russia: Morphological studies indicated that the "Altai dog", a ca 33,000-33,500 cal BP Palaeolithic doglike, came from a lineage that is now extinct and that was derived from a population of small wolves (also now extinct) (Ovodov *et al.* 2011; Larson *et al.* 2012). However, a later genetic analysis of mtDNA claimed that they are, in fact, early dogs (Druzhkova et al. 2013).

1.2.3 Genetic studies: when and where the first dogs arose?

The difficulty in pointing out the exact location and dating when the wolf domestication occurred, can be explained by the subsequent complex demographic processes which may have altered the patterns of genetic diversity (Shannon *et al.* 2015). Firstly, an important clue to reveal dogs' origin may rely in the genetic signatures found in their DNA. However, this is blurred by the admixture that occurred between different dog and wolf populations over the last 10,000 years (Larson *et al.* 2012; Freedman *et al.* 2014; Fan *et al.* 2016), population bottlenecks due domestication and, more recently, breed formation (Freedman *et al.* 2014; Freedman & Wayne 2017). Secondly, the divergence between the dog and the modern wolf occurred within a short period (a rapid speciation may result in incomplete lineage sorting, which is an imperfect segregation of all alleles into all lineages), making it difficult to date the separation from its wild counterpart (Freedman *et al.* 2014). Timing domestication is further complicated by the few generations that separates dogs from their ancestor, so the number of mutations between the dog and the wolf is small (Freedman & Wayne 2017)

Taking all this in account, and accepting as true that dogs were originally domesticated from an extinct wolf population (Freedman *et al.* 2014), many researchers have proposed that the direct analysis of ancient specimens might be a better approach in discovering dog's origin and in elucidating the domestication processes. Genetic studies using ancient DNA (aDNA) can provide important insights into the understanding of the past

demographic history and human-driven selection for certain traits in animals of the past (Botigue et al., 2017; Ollivier et al., 2016, 2013; Pilot et al., 2014), and can be helpful when bones are morphologically indistinct as genetic differentiation can precede morphological changes associated to domestication.

The first phylogenetic analysis was conducted by Vilá et al (1997), using mtDNA control region (CR) sequence data of dogs and wolves; four major clades containing dog haplotypes were identified (I-IV) and an estimation of dog domestication was placed at least 135,000 year ago, although according to the authors *"such estimates may be inflated by unobserved multiple substitutions at hypervariable sites"*. Forward on, the first study to attempt in pinpoint a geographic origin of dogs (Savolainen *et al.* 2002) was based in genetic diversity of 654 dogs from Europa, Asia, Africa and North America and 38 Eurasian wolves. Sequencing 582 bp of mtDNA d-loop region – a useful marker for addressing intraspecific evolutionary questions - from both dog and wolf, Savolainen *et al.* (2002) constructed a phylogeny that recovered clades I-IV of Vilá et al (1997) and added two more clades containing dog haplotypes, designating them as haplogroups A, B, C, D, E, F, indicating that dogs are derived from 6 separate lineages. The new designation given by Savolainen *et al* (2002) to the clades has been a benchmark to all subsequent studies in phylogeny of dogs.

Back to dating the origin of dogs, later studies using Whole Genome Sequencing (WGS) of modern wolves and dogs argued for a more recent origin of dogs – 30,000 years ago (Wang et al. 2013) or 16,000-11,000 years ago (Freedman *et al.* 2014). In the year of 2015, Skoglund and colleagues used the 35,000-year-old Taimyr wolf to estimate dog-wolf divergence to at least 27,000 years ago. Later, in 2017, a study compared the mitochondrial genome sequences of 3 Neolithic dogs with sequences from modern dogs and wolves, giving a dog-wolf divergence time of 36,900-41,500 years BP followed by domestication occurring between 20,000-40,000 years BP (Botigué *et al.* 2017). A more recent study analysed canids Y-chromosome sequence and revealed that the dog male lineage and the modern gray wolf genetically diverged from a common ancestor between 68,000-151,000 years PB (Oetjens et al. 2018). The reason for this disparity of dates between studies can be explained by the different age chosen to calibrate the mutation rate in the wolf. Even not reaching an agreement, up to date, the timing of domestication has been accepted on

a date in the Upper Palaeolithic, between 15,000 and 12,000 years ago, due to clear archaeological evidence of morphologically distinct modern dogs in that time (Benecke 1987; Street 2002; Thalmann *et al.* 2013; Janssens *et al.* 2018).

In order to determine whether dogs where domesticated in one or multiple places, and the precise time of these events, many researches worldwide conducted DNA analysis using different approaches. A table containing the main theories of dog domestication, until date, is presented below for a better understanding.

Table 1. Compilation of the putative geographic location of dog domestication supported by different studies.

Geographic Location	Author	Supporting evidence	Notes
Western Europe	(Thalmann et al. 2013)	• Comparison of mitogenome of 18 ancient canids from Eurasia and America, along with extant dogs and wolves from around the world; • The phylogenetically results indicated that modern dogs are more closely related to either ancient or modern canids of Europe; • Dog-wolf divergence time was estimated at 18,800–32,100 years ago.	Thalmann and his colleagues also used the Goyet and Altai dogs in their research, concluding that they may represent aborted domestication episodes.
Central Asia	(Shannon et al. 2015)	• Survey of 185,805 genotyped markers of 4,676 extant purebred dogs and 549 village dogs from 38 countries, combined with previously generated array and mtDNA data from dogs and wolves. • Results supports Central Asia origin for dogs, having undergone a strong domestication bottleneck followed by population expansion in East Asia.	1) Village dogs are relatively free of admixture, genetically diverse and geographically widespread, making them a powerful candidate to uncover dog population history. 2) Study of extant dogs cannot exclude the possibility that domestication occurred earlier elsewhere and then, either through migration or a separate domestication event, arrived and diversified in Central Asia.
East Asia	(Savolainen *et al.* 2002)	• Extant mtDNA of dogs and wolves suggests a greater antiquity of haplotypes in East Asia; • Concluded that domestication event occurred at ca. 15,000 years ago in East Asia.	A lack of dog remains dated before 12,500 years BP in this region (East Asia) rebuts this proposal (Larson *et al.* 2012)

	(Boyko *et al.* 2009)	• Similar mtDNA haplotype diversity in African and East Asian village dogs; • Hypothesis of an East Asian origin of dogs.	
	(Duleba *et al.* 2015)	• Analysis of 555 mitochondrial genome of extant dogs; • Indicates that dogs may have originated in East Asia during the Mesolithic and Upper Palaeolithic.	
Southeast Asia	(Pang *et al.* 2009)	• A study of mitochondrial genome of 169 extant dogs and mtDNA Control Region of dogs and wolves across the Old World, sampling either indigenous village dogs or breeds with known geographic origins; • Indicated the origin of dogs in the South of the Yangtze River (ASY), China, in reason of the highest diversity of mtDNA haplotypes only found there.	The problem appointed for this proposal is that no wolf remains have been found in this region and the earliest dog remains dates only to 4,200 BP (Larson *et al.* 2012). Although, this could be due to the unfavourable environmental condition for preserving fossils in this region, or less archaeological studies developed in this region.
	(Brown *et al.* 2011)	• Using mtDNA D-loop, and Y-chr markers (SNP and STR), they analysed village dogs from Middle East and Southeast Asia, along with 138 breed dogs; • Evaluate genetic evidences for a Middle East (as claimed by vonHoldt *et al.* [2010]) versus ASY dog origin; • The results supported a dog origin in Southeast Asia instead Middle East.	
	(Ding *et al.* 2012)	• Y-choromossome DNA sequences of extant dogs worldwide; • Indicates the origin of dogs in South of the Yangtze River (ASY), China, because of the	

		diversity of Ychr-DNA haplogroups found there.	
	(Wang et al. 2016)	• Based on extant mtDNA of 19 breed dogs worldwide, including 11 indigenous dogs from Southeast Asia, 12 indigenous dogs from Northern East Asian, 4 village dogs from Nigeria and 12 Eurasian gray wolves; • Suggests that dogs originated from Southeast Asia 33,000 YBP. From this population, a subset of early dogs, around 15,000 years ago, started migration to the Middle East, Africa and Europe and at least America.	
Middle East	(VonHoldt *et al.* 2010)	• Survey of 48,000 autosomal SNPs in dog breeds and wolves; • Concluded that dogs have more haplotype sharing with wolves from Middle East than with other wolf populations.	Freedman and colleagues (Freedman *et al.* 2014) refused this proposal. They argued that this genetic proximity is likely due to dog-wolf introgression in the Middle East rather than an indication of Middle Eastern origins.
West and East Eurasia (Dual Origin)	(Frantz *et al.* 2016)	• Based on 59 hypervariable mtDNA fragments from ancient European dogs, a 28x nuclear genome of an ancient dog from Ireland, 80 extant dogs' whole-genome data and 605 modern dogs (including village dogs and 48 breeds) genotyped on the CanineHD 170 K HD SNP array; • Revealed a deep splitting separating modern East Asian and Western Eurasian dogs (ca. 14,000-6,400 BP).	Botigué *et al.* (2017) refuses the hypothesis of dual origin and Late Neolithic dog population replacement. Whole genome sequencing of an Early and End Neolithic dog from Germany demonstrated continuity with each other and ancestry with modern European dogs.

	• Indication that two genetically differentiated wolf population in Eastern and Western Eurasia may have been independently domesticated at least 15,000 years ago, and the Eastern dog population spread and partially replaced an indigenous Western European dog population.	
(Pilot *et al.* 2015)	• An analyse of genome-wide SNPs variability of free-breeding dogs (FBDs) and pure-breed dogs across Eurasia; • Suggests that modern European breeds originated locally from European FBDs and East Asian and Arctic breeds show closest affinity to East Asian FBDs; • Indicated a gradual westward expansion of East Asian indigenous dogs to the Middle East and Europe, leading afterwards to the replacement of native resident populations in Western Eurasia.	
(Deguilloux *et al.* 2009)	• Ancient mtDNA analysis of three Neolithic France dogs, compared to sequences of Swedish and Italian Neolithic dogs; • Confirmed that clade C was widespread over Western Europe and supports a maternal lineage replacement in Europe.	

1.3 The Iberian dog

In the Iberian Peninsula, the first evidence claimed to belong to a dog, comes from the Upper Late Palaeolithic, 19,000-12,500 YBP (Altuna & Mariezkurrena 1985; Vigne 2005). In Portugal, the oldest almost complete skeleton of two dogs recovered from excavations come from shell-middens located in the Muge – Cabeço da Arruda, Tagus valley and Poças de S. Bento, Sado valley - these remains are dated to the Mesolithic period (ca. 7,600 cal BP) (Detry & Cardoso 2010; Pires *et al.* 2019). These findings confirms that during the Chalcolithic period, in the Iberian Peninsula, dogs were already present in the territory a long time ago, in contrast to other domesticated animals, such as sheep, pig, goat and cattle, that were brought to the Western coast of the Iberian Peninsula only *ca* 7,500 years cal BP (Zilhão 2001; Davis & Simões 2016).

According to the morphometric records of excavated remains of dogs from the Chalcolithic Iberian Peninsula, there was already some intraspecific phenotypic variability at this time, as a result of the beginning of intentional breeding towards different objectives (human-driven selection) in parallel to accidental crosses between domestic and wild forms (Catagnano 2016).

It has been suggested that the Iberian Peninsula, during the Last Glacial Maximum (22,000-19,000 BP (Yokoyama *et al.* 2000)), served as a biodiversity refugia (Hewitt 1996). This had profound influence on the genetic structure of isolated populations (Avise *et al.* 1998) and its effects on ancestral Iberian wolves population are currently poorly understood. Curiously, a recently published study analysed the earliest Iberian dogs found and reported a high frequency (83%) of Haplogroup (Hg) A during the pre-Neolithic period (Pires *et al.* 2019), contrasting with the occurrence observed on other areas of Europe (in Frantz *et al.* (2016), the frequency of HgA haplotypes found in Europe was lower than 9% for the period 14,700 to 3,090 BP; HgC frequency was higher than 50% for the same period). Thus, despite previous studies suggest a dog population expansion from East Asia to the West during the Neolithic period as responsible for the modern pattern of predominance of HgA in dogs (Pilot *et al.* 2015; Frantz *et al.* 2016; Wang *et al.* 2016), there was already a high diversity of HgA haplotypes in Iberia before the Neolithic period.

Concerning extant Iberia dogs, they present a great diversity, carrying haplotypes belonging to the different haplogroups A, B, C and D. Haplogroup A is the most frequent, while D is the least represented (Savolainen *et al.* 2002; Pires 2006; Pires *et al.* 2017). Haplogroups E and F were never detected in Iberia, being present only in dogs from Asia (Pionnier-capitan 2010).

To date, the only study that has focused on the mtDNA of the earliest Iberian dogs and wolves is the study by from Pires et al (2019). Studies which included ancient data and tried to unravel the origin of dogs, did not include ancestral samples of Iberian dogs (Deguilloux *et al.* 2009; Thalmann *et al.* 2013; Frantz *et al.* 2016). It is crucial to unravel the genetic composition of past European peripheral populations to better understand the global evolutionary trajectories of early dogs.

1.4 Ancient DNA analysis

1.4.1 History and significance of ancient DNA (aDNA)

Paleogenomic research is a relatively recent discipline in the history of molecular biology, having as a pioneer case the extraction of 229 bp of mitochondrial (mt) DNA from the muscle of a quagga (*Equus quagga*, Boddaert, 1785) from the 19th century, a species currently extinct (Higuchi *et al.* 1984). Following, Pääbo (1985) reported a 3.4 kb fragment cloned from DNA obtained from an Egyptian human mummy with 2,400 years. These researches were the only studies conducted during the pre-PCR era. However, these studies demonstrated to be unreliable, since aDNA from these specimens was limited to low concentrations of highly degraded endogenous DNA that the isolation of bacterial clones (these studies used bacterial cloning to amplify sequences) containing similar DNA sequences was difficult, resulting in contamination (Pääbo *et al.* 2004).

After introduction of the polymerase chain reaction (PCR), it became possible to target and replicate specific DNA sequences and some improvements were made. Unfortunately, until mid-2000s, this new genomic field based in PCR technique was often discredited by science (Orlando *et al.* 2015). The degradation of aDNA together with PCR sensitivity to contamination (e.g. microorganisms or human handling) and inhibitors led to a series of

publications with false-positive results (e.g. in 1994, a DNA alleged to come from a dinosaur [Woodward et al. 1994], was actually nuclear copies of human mitochondrial DNA [numt; Zischler et al. 1995]). Only after the introduction of high-throughput sequencing (HTS) platforms, that paleogenomics studies could bloom.

Currently, it is acknowledged that to reconstruct and better understand the evolutionary processes of a past population is necessary to obtain data directly from archaeological samples. Due to human-induced decline or fragmentation of habitats and breeding selection in the recent past, the exclusive use of modern genetic data can hide important processes of population dynamics, such as changes in population size, structure and migration patterns at different time periods (Hedrick & Waits 2005; Ramakrishnan & Hadly 2009). A phylochronologic approach – in which several populations are studied over large temporal and geographical scales - has been successfully applied to inferring evolutionary history more accurately (Leonard *et al.* 2002; Botigué *et al.* 2017). Studies about the genetic composition of a species require an adequate sample size at different time periods and representing wide geographic coverage.

1.4.2 Damage patterns of aDNA

The intensity of postmortem damage of DNA is affected by the elapsed time since the death of the organism and by taphonomic processes. After death, DNA repair systems ceases whereas destructive processes (mainly spontaneous hydrolytic lesions, oxidative lesions and nonenzymatic methylation of DNA) continues, resulting in chemical modifications (Lindahl 1993; Handt *et al.* 1994; Pääbo *et al.* 2004). In addition, DNA molecules faces an "attack" from bacteria, fungi, and insects that feed on and degrade macromolecules (strand breaks; Paabo *et al*. 2004). Thus, preservation and DNA integrity are better achieved in certain environments - cold, dry and/or low oxygen, e.g. permafrost regions or temperate environments (Lindahl 1993). Therefore, amplification is crucial when working with few DNA copies from ancient material (Griffiths *et al.* 2004), allowing to recover some information from samples in which the disintegration of DNA is not yet complete (circumstances that happens when a tissue becomes rapidly desiccated after death or

when the DNA becomes adsorbed to a mineral matrix, such as teeth or bones, been fossilized; Paabo *et al*. 2004).

Even so, when performing PCR amplification, hydrolytic lesions (i.e. hydrolytic loss of amino groups, or deamination, from the nitrogenous bases: adenine -> hypoxanthine, cytosine -> uracil, 5-methyl-cytosine -> thymine, guanine -> xanthine) can cause nucleotides misincorporation during the first cycles of PCR (C»T and G»A transitions) (Hansen *et al.* 2001; Hofreiter *et al.* 2001; Pääbo *et al.* 2004). Thus, such modifications are constantly reported in aDNA studies, being C»T more frequently found at 5' end and G»A at 3' end of the sequence reads (Briggs *et al.* 2007). Experimental procedures, such as the treatment of DNA extractions with uracil N glycosylase (UNG), revealed cytosine deamination to uracil to be the most common base modification that leads to CT/GA transitions (Hofreiter *et al.* 2001).

1.4.3 Methods and Criteria for authenticity in aDNA

Obtaining adequate samples can be sometimes a hard task if the extracted DNA has suffered damages or has been contaminated. Because the techniques used in aDNA studies can contain inherent problems, during PCR era some methods and criteria have been established to avoid and identify exogenous DNA contamination and to account for sequencing inaccuracies when working with ancient DNA (Paabo 1989; Handt *et al.* 1994; Cooper & Poinar 2000; Gilbert *et al.* 2005; Pires & Ginja 2013). Some of these methods and criteria are summarized in Box 1.

Decontamination of bones and tooth is possible but, it must be used with precaution, since it can be invasive and destructive. There are physical or chemical methods, such as sandpaper polishing or electric drills or 0.1M HCl + 0.5% bleach on powdered samples (Malmstrom *et al.* 2007) or ultraviolet irradiation (O'Rourke *et al.* 2000). When working with aDNA, the environmental condition where the remains of organic material have been deposited must be also considered. Precautions should begin during and soon after excavation due contamination and destructive environmental agents (temperature increase, desalting and decrease of pH; Pruvost *et al.* 2007). Freshly excavated and nontreated material has been demonstrated to contain six times more DNA and has yielded

twice as many authentic DNA than remains treated/stored under standard procedures (washed museum-stored; Smith *et al.* 2001; Pruvost *et al.* 2007). Thus, to preserve DNA

Box 1. Methods and Criteria for authenticity in ancient DNA.
i. ***Isolation of work areas:*** to avoid contamination of endogenous DNA, it's important that DNA extractions and amplification reactions are manipulated in an environment physically isolated from other materials, such as modern samples or PCR product that might be present from other analyses (Cooper & Poinar 2000). Therefore, the use of laboratories dedicated to aDNA work is mandatory, following strict rules for the maintenance of an almost sterile environment and control for contamination[1].
ii. ***Extraction blank controls and PCR controls:*** multiple extraction procedure and negative PCR control (no template DNA is added) must be performed to detect the existence of contamination, e.g. derived from environmental microorganisms or modern human DNA that exists in laboratory and/or in reagents, or is embedded in the samples (Cooper & Poinar 2000). According to Paabo et al. (2004), three extracts may be a reasonable number of extraction attempts.
iii. ***Appropriate molecular behavior***: it's often impossible to obtain long amplification products when working with aDNA because of the fragmentation of the genetic material. PCR amplification efficiency should be inversely related to length of the amplification products, otherwise the amplification could be due to contamination with non-ancient template (Handt *et al.* 1994; Cooper & Poinar 2000). Different lengths of amplifications can be achieved from different species, however in most species the length of amplification is between 100-200 bp (Paabo *et al*. 1989).
iv. ***Reproducibility of results:*** Multiple PCR and DNA extraction from the same specimen should yield consistent results (Cooper & Poinar 2000). However, different results may be useful to identify numt (nuclear insertions of mtDNA) or contamination, when using different primer pairs to amplify partially overlapping sequences (Handt *et al.* 1996).
v. ***Cloning of amplification products and sequencing of multiple clones:*** in case the DNA amount is limiting or degraded, several amplifications and sequence of multiple clones, i.e. high-coverage sequencing are necessary. Overlapping fragments are desirable to confirm that sequence variation is authentic and not due to damage-induced errors (deamination of deoxycytidine residues), 'jumping PCR' (template switching during PCR) and contamination (Handt *et al.* 1994; Cooper & Poinar 2000).
vi. ***Independent replication:*** generation of results in independent laboratories is a common practice in high-standard aDNA research to detect contamination of chemicals or samples (Cooper & Poinar 2000).
vii. ***Biochemical preservation:*** the composition of some biomolecules is correlate with DNA survival (Poinar & Stankiewicz 1999; Cooper & Poinar 2000).
viii. ***Quantification of the number of amplifiable DNA molecules:*** By competitive PCR or Real-Time PCR to access the number of the copy of the DNA target (Cooper & Poinar 2000). Few initial template molecules (<1000 template molecules; Paabo *et al*. 2004) are more likely correlated with substitutions in the final amplification product when misincorporations happens during the early cycles of PCR (Handt *et al.* 1996).
ix. ***Associated remains:*** Associated remains can be good supporting evidence for DNA preservation and contamination (Cooper & Poinar 2000).

Some additional criteria have been subsequently included:
x. ***Use of a "carrier DNA" negative***: The addition of a control containing nonamplifiable "carrier DNA", such as nontarget DNA from a different source, should be included to avoid misleadingly clean negative controls (Handt *et al.* 1994).
xi. ***Preservation-dependent pattern of DNA damage and sequence diversity***: Sequences isolated from badly preserved samples should be more damaged than better preserved samples (as assessed via high-throughput sequencing) (Willerslev & Cooper 2005).
xii. ***Phylogenetic sense or otherwise reasonable results:*** Critical assessment of the sensibility of the results obtained from an aDNA experiment is an important aspect of aDNA research. For example, BLAST searching should be used to confirm that the sequences belong to an expected species or to find contaminants match (Gilbert *et al.* 2005; Handt *et al.,* 1994).

[1]The environment and working surfaces where DNA are handled must to be frequently decontaminated (e.g 10% bleach solution or Actril) and has its own independent air system (Pires & Ginja 2013). Disposable laboratory ware should be preferably used and non-disposable glassware should be treated with 1 N HCO and rinsed with double-distilled (dd) water before use (Handt *et al.* 1996).

molecules, it is demanded to store freshly excavated material in dry and cold conditions. Moreover, some authors believe that more criteria should be added to the list. Gilbert et al. (2005) advocate that researches should do beyond the criteria list and have a more cognitive and self-critical approach of the results and ask themselves whether the study's conclusions have sufficient evidence to support the veracity of the data.

1.4.4 Ancient DNA sequencing technologies

Currently, technical advances in sequencing DNA, i.e. high-throughput sequencing platforms [HTS] (e.g. Second Generation Sequencing, also known as Next Generation Sequencing), enhanced data authenticity by identifying contaminant and filtering and correcting aDNA damages because NGS platforms (e.g. Roche/454 FLX, Illumina HiSeq X Ten (Figure 5), Applied Biosystems SOLiDTM System, Helicos HelicospeTM and Pacific Biosciences SMRT instruments) can generates overlapping reads and multifold coverage of the target regions through emulsion PCR or solid-phase amplification and an increase recover of shorter DNA fragments (Metzker 2010; Illumina 2016), differently from traditional Pre-NGS PCR-based approaches, in which loci are individually targeted and ultrashort aDNA fragments (~30-50 bp) are unexploited. A better explanation of these NGS technologies can be found at Mardis (2008), Millar *et al.* (2008) Shendure & Ji (2008) and Metzker (2010).

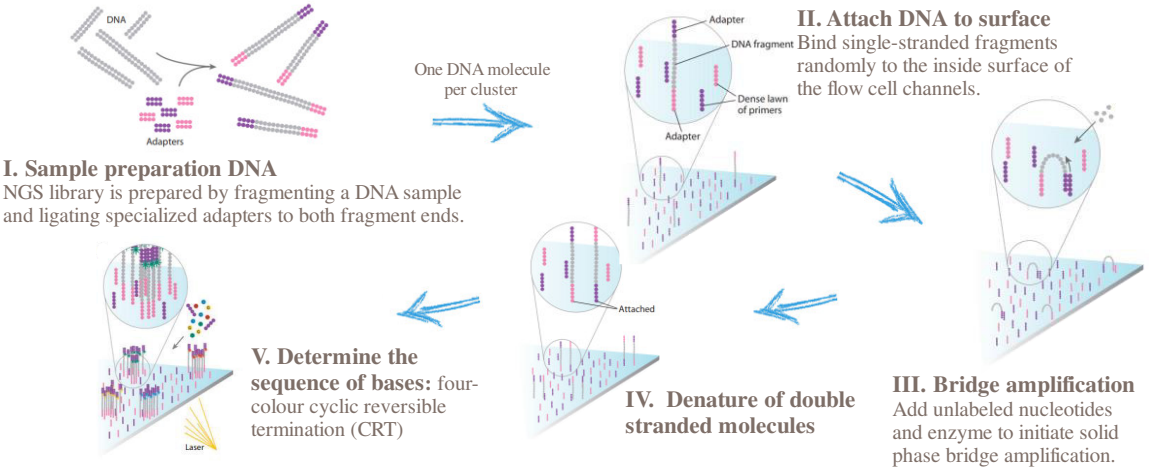


Figure 5. Representation of the workflow of the shotgun resequencing method HiSeq X Ten from Illumina and followed in this study. Adapted from Mardis (2008).

More recently, new methods in NGS to improve sequence retrieval from aDNA has been developed. For example, silica-based DNA extraction technique (Rohland & Hofreiter 2007) in combination with single-stranded library preparation (Gansauge & Meyer 2013) has allowed short molecules (50bp) to be recovered (Dabney *et al.* 2013), while conventional PCR amplification provides limited power to reconstruct sequences from short DNA fragments usually found in ancient samples, losing amounts of important information. Since then, many tens of ancient genome have been sequenced, some of them with high-coverage: 4,000-year-old paleo-eskimo genome (20x; Illumina; Rasmussen *et al.* 2010), 120,000-year-old polar bear genome (0.4x; Illumina; Miller *et al.* 2012), 700,000-year-old horse genome (Orlando *et al.* 2013), 50,000-year-old Neanderthal genome (42x; Illumina; Prüfer *et al.* 2014), 4,800-year-old dog genome (28x; Illumina; Frantz *et al.* 2016).

However, large-scale products resulted from these new methods in sequencing DNA requires advanced use of computational resources and bioinformaticians for data processing and analysis. A variety of software tools are available for analysing next-generation sequencing data, being categorized according to their different functions: (i) sequence quality scoring; (ii) alignment of sequence reads against reference genome of interest; (iii) base-calling and/or SNP detection; (iv) *de novo* assembly, from paired or unpaired reads; (v) authentication of aDNA data; and (vi) genome browsing annotation (some software packages reviewed in Shendure & Ji 2008; Orlando *et al.* 2015). As new extraction and computational methods improve, in the near future, it will expand the age range and quality of specimens from which data can reliably be obtained, bringing us new insights about population past demographic, adaptive and admixture trajectories.

A new generation of sequencers, called Third Generation Sequencing is under active development since 2010. These new platforms work by reading the nucleotide sequences at the single molecule level, not requiring breaking long strands of DNA into small segments and a very small amounts of starting DNA templates as next generation sequencers does. Moreover, it uses the total DNA isolated from the specimen, without any kind of sequence enrichment or PCR amplification, that can sometimes introduce biases into the data, thus allowing an assessment of postmortem DNA damage. However, error rate remains to be

improved before these new generation of sequencers overcome Second Generation Sequencing (Rizzi *et al.* 2012; Bleidorn 2015).

1.5 Relevance of this study

The scientific community already knows that the modern dog was not domesticated from the extant grey wolves, being species that are currently well structured, that is, there are few direct lineages between extant dogs and extant wolves that places them in the same mitochondrial haplogroup. Thus, science now looks to the past and focuses on the fossils of wolves and dogs to unravel the origin and evolution of the dog – if a single or multiple domestication of an ancient grey wolf population/populations happened, and the cases of historical hybridization between dogs and wolves.

Little is known about the origin of the Iberian dogs and a hypothesis of a local domestication from the Iberian grey wolf (*Canis lupus signatus,* Cabrera 1907) or events of admixture should not be excluded (Pires *et al.* 2019). Understanding the Iberian dog is also a way to understand the humans who lived in the Iberian Peninsula. Mutualism between the two species allowed man to be better succeeded throughout his history. Looking at past populations may also bring new discoveries (e.g past population demography) that may prove to be important for the management and conservation of wild species, such as the endemic Iberian wolf which has suffered a sharp decline due to the direct persecution of man and nowadays is considered an “Endangered” status of conservation (Queiroz *et al.* 2005).

The present study aims to make a genomic analysis of four ancestral dogs and one wolf from the Iberian Peninsula to fill the lack of information about the Iberian Canids from the Chalcolithic, providing important data to understand the origin and diversity of the Chalcolithic dogs, enriching the knowledge of *Canis lupus* in the prehistory.

1.6 General objectives of this thesis

• Use bioinformatic tools to analyse and filter DNA sequences - extracted from 4 ancient dogs and 1 ancient wolf - for screening for possible post-mortem contaminations and mutations in order to obtain a reliable consensus sequence of the samples.
• Study the genetic diversity and population structure of Chalcolithic dogs, in order to infer evolutionary trajectories and their genetic composition: identify variants in the mitochondrial genome and perform phylogenetic analyses for haplogroup assignment;
• Assess nuclear sequences to determinate the sex of these ancient dogs and wolf.

This study was developed within the scope of the project WOOF - Tracing the origins and evolutionary paths of the Iberian and the Maghreb Dog with reference PTDC/HAR-ARQ/29545/2017, supported by national funds by FCT / MCTES and co-supported by Fundo Europeu de Desenvolvimento Regional (FEDER) throughout COMPETE - POCI – Programa Operacional Competividade e Internacionalização (POCI-01-0145-FEDER-029545), in the area of Biological Sciences and sub-area of Zooarcheogenetics, headed by Ana Elisabete Pires, my supervisor.

2. Materials and Methods

2. Materials and methods

Below it is described in detail the genetic study conducted including the description of the archaeological remains dated to the Chalcolithic period, and the methods followed in the ancient DNA laboratory: sub-sampling, DNA extraction, preparation of genomic libraries, sequencing and bioinformatic analysis. My particular contribution regards the bioinformatic analysis of the generated sequences.

2.1 Archeological material

Four ancient dogs and one ancient wolf remains from Iberian Peninsula were studied for their genetic content (Table 2; Figure 6). These samples have been previously studied in Pires et al (2019) where a fragment of 181 bp of their mtDNA was investigated by a 2nd generation 454 (Roche) sequencing method (PCR based). In this study a genomic approach was attempted using a *Next Generation Sequencing* approach for the same ancient remains.

Figure 6. Photos of some Chalcolithic Iberian Canis remains. A) sample LYEP9 from Leceia, Portugal; B) sample LYEP51 from Valladolid, Spain; C) sample LYEP11 from Leceia, Portugal. Note: no picture is available for sample LYEP53 from El Casetón de la Era, Spain, nor LYEP27 (wolf). Photos by Carlos Fernandez-Rodrigues (remains from Spain) and José Paulo Ruas (remains from Portugal).

Table 2 Samples specifically analysed in this study.

Ancient *Canis* sample ID	Scientific name/ Common name	Skeletal element recovered	Origin	Chronology	Laboratory ID	Reference
LYEP9	*Canis lupus familiaris*/ domestic dog	Mandible	Leceia (PT)	ca. 5,000-4,300* BP	P9306_1034	(Pires *et al.* 2001, 2019)
LYEP11	*Canis lupus familiaris*/ domestic dog	Maxilla	Leceia (PT)	ca. 5,000-4,300 BP*	P9306_1026	(Pires *et al.* 2001, 2019)
LYEP51	*Canis lupus familiaris*/ domestic dog	Maxilla	El Casetón de la Era, Valladolid (ES)	ca. 4,000 BP*	P9306_1033	(Arana & Rodríguez 2013; Pires *et al.* 2019)
LYEP53	*Canis lupus familiaris*/ domestic dog	Tooth (3rd Incisor)	El Casetón de la Era, Valladolid (ES)	ca. 4,000 BP*	P9306_1031	(Arana & Rodríguez 2013; Pires *et al.* 2019)
LYEP27	*Canis lupus signatus*/ Iberian wolf	Tooth (1st Lower molar)	Penedo de Lexim, Mafra (PT)	4,085–3,856 cal BP**	P9306_1025	(Sousa 2010; Pires *et al.* 2019)

*Dated by archaeological context

**Indirect radiocarbon date for a specimen of *Sus* from the same stratigraphic unit.

Previous studies performing biometric analysis provide information regarding species identity for these 5 samples. For the samples LYEP9 and LYEP11, found in the archeological site of Leceia, values obtained by Pires *et al.* (2001) on the measurement of different bone parameters were compared with homologous parts of extant wolves. A dimensional reduction in the values was observed for the samples from Leceia, compared to their wild relatives, indicating that these samples probably belonged to dogs. The biometric study of Arana and colleagues (2013) confirms the status of the canids found in this archaeological site as belonging to dogs. Regarding LYEP27, osteometry analysis was conducted in order to determine the taxa of this sample. According to Moreno-Garcia and colleagues (2016) this samples probably belonged to a wolf because of the large width of this tooth (13.4 mm). Pires and colleagues (Pires *et al.* 2019) also assigned this remain to a wolf, after comparing its likelihood ratio and posterior probabilities under the hypothesis of being a wolf or a dog taking into account the archaeological, osteometric, direct dating and isotopic data.

2.2 Study areas

The number of mammals´ remains found in excavations dated to the Chalcolithic from the Iberian Peninsula is quite expressive, including a high number of domestic animal remains (Catagnano 2016). This fact can be explained by the abandonment of hunting activities in favor of intensification on agro-pastoral activities. Nevertheless, the presence of dogs does not occur equally in all archaeological sites throughout the Iberian Peninsula. Villages dogs may be present mainly in sites where the economy provided food surplus. Dogs were better tolerated in those circumstances and included in different human activities, such as hunting, property and livestock guarding and herding (in a mutualist relationship) (Pires *et al.* 2001).

Leceia (Oeiras, Portugal) is considered one of the most important archaeological excavations of a Chalcolithic site in Iberian Peninsula. Located at the Estremadura region, near the coastline and the estuary of the river Tejo (Figure 7), this region had a favorable climate for human settlement during the pre-history (Late Neolithic [3,300-2,900 cal BC], Early Chalcolithic [3,800-2600/2,500 cal BC] and middle and Late Chalcolithic [2,500-2,100

cal BC] (Cardoso & Soares 1995; Cardoso 1997, 2000, 2008)). Since 1983, 20 annual field seasons were carried out; among 122 carnivores remains excavated, a total of 81 remains of dogs were identified; some of them displayed traces of human consumption (Pires *et al.* 2001). Only two dog remains from this site were selected for this genetic study: samples LYEP9 and LYEP11. The selection of these specific samples was made after a screening of 14 samples to identify the best ones.

Penedo de Leximm (Mafra, Portugal) is a volcanic hill (223m) situated between the Ribeira da Mata and Ribeira da Laje. Next to Leceia, Penedo de Lexim is also located at the Estremadura region (Figure 7). Its occupation dates back to Neolithic, Early Chalcolithic, Late Chalcolithic, Bronze Age and Roman period, during the second half of the fourth Millennium and the third Millennium BP. Within the thousands of bones excavated and identified, only two dogs and one wolf were found (Sousa 2010). This unique Chalcolithic wolf was selected for this study: sample LYEP27.

El Casetón de la Era (Villalba de los Alcores, Valladolid, Spain) is an archaeological site located in the North of Spain (Figure 7), discovered in 1997, through the use of aerial photography. It was occupied during two distinct phases of pre-history: Chalcolithic (the first half of the third millennium cal BC) and Bronze age (1,600-1,335 BC) (Delibes de Castro *et al.* 2018). Among 27 of the carnivorous remains excavated, 16 were identified as dogs. Bone marks, related to human consumption, were not found in those specimens, indicating that dogs presence was related to other purposes rather than as a food source (Arana & Rodríguez 2013). Only two dog remains from this site were selected for this genetic study: samples LYEP51 and LYEP53. The selection of these specific samples was made after a screening of 4 samples to identify the best ones.

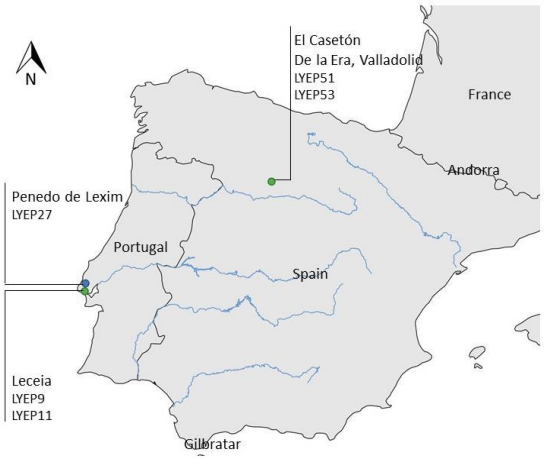


Figure 7. Location of the Iberian archaeological sites from where the studied dog/wolf remains were excavated. Dogs: Leceia, Portugal (n=2); El Casetón de la Era, Spain (n=2). Wolf: Penedo de Lexim, Portugal (n=1).

2.3 Archaeogenetics

Sub-sampling and aDNA extraction were performed in facilities dedicated exclusively to ancient DNA analysis, located at the Archaeological Research Laboratory of Stockholm University (Sweden) by a well-trained researcher, following appropriate protocols to avoid contamination as described below.

2.3.1 Sub-sampling for aDNA analysis

A thorough description can be found in Pires *et al.* (Pires *et al.* 2018) and Pires *et al.* (Pires *et al.* 2019) but, briefly, prior to the DNA extraction, samples underwent an outer surface UV sterilization, followed by removal of approximately 1 mm layer of the surface. A piece of 1 cm^2 were reduced to a fine powder using sterile scapel or a Dremel tool. Bone powder replicas were kept in a freezer at -20ºC for subsequent analyses. Other Iberian *Canis* samples which sequences were used in this thesis for comparisons were treated the same way.

2.3.2 DNA extraction, preparation of genomic libraries and sequencing

A thorough description can be found in Pires *et al.* (Pires *et al.* 2018) and Pires *et al.* (2019) but, briefly, for DNA extraction, the bone tissue powder (100-200 mg) was digested overnight with 1 mL of buffer EDTA (0-5 M, pH 8) and Urea 1 M with 10 µL of proteinase K

(20 mg/mL in water) at 38 ºC with constant agitation. For each sample, the lysate volume was concentrated down to 100 µl with Amicon columns (MilliporeAmicon Ultra-4 30 kDa) through centrifugation at 4,000 g for 10-15 min. The sample volume was then mixed with 5xPB buffer from the commercial kit QIAquick PCR Purification Kit from Qiagen. Ancient DNA was recovered, following the manufacturer's protocol, in a final volume of 100 µL. Duplicates of aDNA extracts were obtained independently from each specimen. Two negative extraction controls were included and subjected to identical procedure as powder samples.

The obtained aDNA extracts were used for preparation of blunt-end (index) Illumina genomic libraries (Meyer & Kircher 2010). The number of amplification cycles was estimated by qPCR and the libraries were amplified with indexes/right number cycles (5 PCR/per sample) using AmpliTaq Gold® DNA Polymerase (Applied BiosystemsTM). The amplified products were pooled, purified with magnetic beads (Agencourt AMPure, Beckman Coulter), quantified using DNA High Sensitivity Kit with Agilent 2100 Bioanalyzer Instrument (Agilent Technologies), followed by shotgun sequenced on Illumina HiSeq X Ten platform (High Output mode, paired-end 2x150bp; the amplified library was sequenced in a pool of 9 *Canis* samples) at the Science for Life Laboratory Sequencing Centre in Stockholm University. Raw DNA data was sorted into individual samples based on tagged sequences (de-multiplexed).

2.4 Bioinformatic processing

The script used in this study can be consulted for mtDNA analysis (Appendix II) and for nDNA analysis (Appendix III). For a better comprehension of the steps, a schematic representation of the pipeline followed in this study is available for consultation (Appendix IV). Some programs and parameters have not been commented in the pipeline as they are less crucial (e.g *samtools sort* to sort reads, *samtools index* to index BAM files, *bgzip* to compress VCF files, and *tabix* to index VCF files).

2.4.1 Raw-read processing

To inspect the quality of raw reads, we used the software Fastqc v0.11.8 (Andrews 2010). Fastqc generates a graphical report with most relevant statistics for the read set: per base sequence quality of reads, overrepresented sequences and the presence of adapter sequences, among other relevant parameters. After visual inspection, cutadapt v1.18 (Martin 2011) was used to trim any read containing adapter sequences, as well as to remove low quality bases towards read ends (quality score < 30) . Reads that had short length (<35bp) or reads containing one or more "N" were excluded. After these filters, reads were screened a second time with Fastqc in order to confirm the expected improvement in read quality. Clean paired end reads were then collapsed using AdapterRemoval v2.2.2 (Lindgreen 2012), requiring a minimum of 11 bp overlap between read pairs.

2.4.2 Mapping

Since archaeological samples tend to contain exogenous DNA, mainly due to human manipulation (Paabo 1989; Zischler *et al.* 1995), it is important to identify reads from other species and exclude them from downstream steps of analysis. The procedure accounted for this challenge. Sequenced reads were mapped against other reference genomes e.g. human, pig, chicken and cow, before being mapped against the reference dog genome.

Mitochondrial

Read mapping used BWA aln v0.7.17 (Li & Durbin 2010) with some modified parameters [seeding was disabled (-l 1024), maximum number of gaps was set to 2 (-o 2) and maximum edit distance was set to 0.03 (-n 0.03)] onto a composite reference genome consisting of human (*Homo Sapiens* [Linnaeus 1758]; NCBI Acession Number NC_012920.1), pig (*Sus scrofa* [Linnaeus 1758]; NCBI Acession Number NC_000845.1), chicken (*Gallus gallus* [Linnaeus 1758]; NCBI Acession Number NC_040902.1) and cow (*Bos taurus* [Linnaeus 1758]; NCBI Acession Number NC_006853) to detect and remove contamination (Greig *et al.* 2015).

The sequence alignment map (SAM) file generated by BWA was then filtered (mapping quality of 20 or greater) and those reads with mapping quality below the threshold were kept as contamination-free reads. These clean reads were then mapped against boxer dog reference mitochondrion (CanFam3.1; NCBI Acession Number NC_002008.4) and Eurasian wolf reference mitochondrion (NCBI Acession Number NC_009686.1) using BWA aln with the same previous parameters. The resulting SAM file was converted to BAM and only those alignments with mapping quality above 30 were maintained. AddOrReplaceReadGroups and MarkDuplicates tools from PICARD v2.18.14 (Broad Institute 2018) were used to add read groups and to remove PCR duplicates, respectively.

Nuclear

Given that nuclear genomes are much larger than mitochondrial genomes, removal of nuclear genome contaminant reads was performed sequentially for the same species as described above, instead of all at once using a composite approach. For each genome reference (human [GRCh38.p12; GenBank assembly accession GCA_000001405.27), pig [Sscrofa11.1; GenBank assembly accession GCA_000003025.6], chicken [GRCg6a; GenBank assembly accession GCA_000002315.5] and cow [ARS-UCD1.2; GenBank assembly accession GCA_002263795.2] BWA aln was employed with some modified parameters (-l 1024, -o 2 and -n 0.03). After the removal of all reads that mapped against human, pig, chicken and cow genomes (mapping quality of 20 or greater), the remaining reads were aligned against boxer dog nuclear genome reference (CanFam3.1; GenBank accession AAEX00000000.3) and Eurasian wolf denovo assembled wolf reference genome (Gopalakrishnan et al. 2017) using BWA aln with the same previous parameters. As described before for mitochondrion, good quality mappings (MQ>30) were kept and AddOrReplaceReadGroups and MarkDuplicates tools from PICARD v2.18.14 (Broad Institute 2018) were used to add read groups and to remove PCR duplicates, respectively.

2.4.3 DNA degradation

MapDamage v2.0 (Jónsson *et al.* 2013) is a software that tracks and quantifies DNA damage among ancient DNA sequencing reads generated by NGS platforms, enabling rescaling of base quality scores specific to the damage patterns of aDNA. This software was employed

using default parameters and the rescale option, in order to rescale base quality scores in bam files, attributing a low quality score to bases with signs of postmortem degradation effects, such as C>T transitions at the 5' ends of the molecule and G>A transitions at the 3' ends of the molecules. The resulting bam file with rescaled qualities for the most likely postmortem damaged bases was used for further steps.

2.4.4 Variant calling

GATK v4.0.11.0 (McKenna *et al.* 2010) HaplotypeCaller was used in order to accurately call variants, namely Indels (insertions and deletions) and SNPs (Single Nucleotide Polimorphisms). SNP calling was recalibrated by local realignment of reads in regions with candidate InDels (default settings and i) disabled soft clipped bases, ii) Don't skip calculations in ActiveRegions with no variants and iii) force active regions). Following, through GATK VariantFiltration, the SNP variants were filtered with per read SNP quality ≥20 and coverage ≥5, accepting only homozygous variants for mitochondrial genome. ReadPosRankSum and AS_BaseQRankSum tools of GATK were also used as looser parameters at the end of the reads, thus accepting edge SNPs with confidence values lower than those of the SNPs at the central positions of the reads.

Finally, all variants were then saved in variant call format (vcf) files using GATK SelectVariants for further usage.

2.4.5 Consensus mitochondrial sequence

In order to infer about the haplogroup of each ancient sample, a fasta file for each sample was created independently using the corresponding BAM alignment file.

First, Bedtools v2.27.1 (Quinlan & Hall 2010) genomecov was used to produce a bed file containing a summary of the coverage through the alignment file. An additional script was created in order to account only for Indels when there is a position with both Indels and SNPs, since Bedtools genomecov do not recognize Indels. Using Bedtools maskfasta, the positions with zero coverage or low coverage (<2x) were replaced with N's, since these regions don't offer enough reads to confidently call variants and establish haplotypes. The

result is a fasta file equal to the dog CanFam3.1 mitochondrion genome but having N's at the positions with low coverage. For those variants initially called with GATK HaplotypeCaller (see section *Variant Calling*) that did not pass filter parameters, the reference nucleotide was also replaced with N's using Bedtools maskfasta; otherwise, at the positions where SNPs passing filters were detected, the reference nucleotide was replaced by the alternative nucleotide using bcftools consensus. This tool also generates a fasta file with the consensus mitochondrial sequence, that can be used as input for any alignment software.

2.4.6 Contaminants - taxonomic assignment

In order to identify the source of contamination, publicly available sequences were downloaded from NCBI database and a taxonomic assignment using BLASTn (Altschul et al. 1990) was applied. Only the top blast hit into a subject domain level were specified for output format. For the sake of the analysis time, only 1,000,000 reads were blasted from a total of 28,383,012 number of reads, on average, per sample. These reads chosen for the analyses, are the ones that did not map against dog/wolf reference genome neither composite genomes.

2.4.7 Species identity assignment

In order to confirm the taxa of the specimens, a comparison of the coverage obtained for each alignment when mapped against dog or wolf reference was performed. Since intraspecific mappings are expected to yield better results than interspecific ones, namely for the total amount of reads mapped as well as for higher mapping qualities, this information was used to corroborate previous information (zooarchaeological information (such as dating and osteometry) regarding the species of each sample.

2.5 Multi-sequence alignment of whole mtDNA genome sequences

2.5.1 Dataset and Alignment

An exhaustive search of mitogenomes from wolves and dogs from Iberia and Eurasia – extant and ancient, was conducted to assign correctly each of the 5 ancient sequences to

an haplogroup with reference mitogenomes from dogs and wolves. Twenty-two mitogenomes was successfully retrieved (see Appendix V and Table 3).

Prior the construction of the phylogenetic tree, using Geneious v2019.1 (Kearse *et al.* 2012), sequences were aligned using Muscle algorithm. Regarding LYEP51, despite the high percentage of unknown/missing nucleotides (N) in the Control Region (CR) (see Table 5), the sequence obtained by NGS-Illumina method was used here in the complete mtDNA alignment.

Table 3. Chronology and geographic location of sequences used to construct the phylogenetic tree based on the mitogenomes available.

Context	Canid type	Cultural Period	Number of sequences
Iberia	Dog	Ancient/Chalcolithic	4 (this study)
	Wolf	Ancient/Chalcolithic	1 (this study)
	Dog	Modern	2
	Wolf	Modern	4
Eurasia	Dog	Ancient/Neolithic	1
	Dog	Ancient/Palaeolithic	2
	Wolf	Ancient/Palaeolithic	7
	Canis sp.	Ancient/Palaeolithic	1
		Total	22

2.5.2 Model Test with jModelTest2

The software jModelTest2 (Guindon & Gascuel 2003; Darriba et al. 2012) was used to select the best evolutionary model to be used in the tree construction. Using posterior probability-based criteria (e.g. BIC, which stands for Bayesian Information Criteria), the model that generated the lowest BIC score is considered to be the best evolutionary model to explain the pattern of nucleotide substitution for this specific dataset.

2.5.3 Phylogenetic tree

A phylogenetic analyses using MrBayes v3.2.6 (Ronquist & Huelsenbeck 2001) for Bayesian Inference implemented in the Geneious v2019.1 software (Kearse *et al.* 2012), was used to estimate Bayesian support under the best evolutionary model determined by jModelTest2, which was the HKY85+G (Hasegawa–Kishino–Yano) model of nucleotide substitution, which accounts for variable base frequencies, different transition and transversion rates

(Hasegawa *et al.* 1985) and using default parameters (4 gamma categories; 1,100,000 chain length; 200 subsampling frequency; 100,000 burn-in length). A previously published coyote (*Canis latrans*) mtDNA complete sequence was used as an outgroup (GenBank Acession Number NC_008093.1). A Majority greedy clustering consensus tree was generated using Geneious v2019.1 software using the Bayesian Inference support values.

2.6 Multi-sequence alignment of control region (CR) mtDNA partial sequences

2.6.1 Dataset and Alignment

A Median-Joining (MJ) network method was chosen to help assigning each of the 5 ancient sequences to an haplogroup with reference sequences from extant specimens from the same region – Iberia, and to visualize genealogical relationships at the intraspecific level. Prior the construction of the networks, Mega v7 (Kumar et al. 2016) was used for the multiple alignment of a compilation of 253 sequences (Appendix VI): 27 ancient dogs from Iberia (including the 4 sequences obtained in this study), 94 ancient dogs from Eurasia (Eurasia, in this study, stands for the countries inside the perimeter considered in Figure 8), 3 ancient wolves from Iberia (including the one obtained in this study), 36 ancient wolves from Eurasia, 61 modern dogs and 18 modern wolves from Iberia and 11 sequences from wolves outside of Iberia (included in order to have wolf haplogroup 2 representatives, since in Iberia, extant specimens all belong to wolf haplogroup 1 (Bjornerfeldt *et al.* 2006; Koepfli *et al.* 2015; Koblmüller *et al.* 2016; Pires *et al.* 2018), with only one exception, an Portuguese wolf that belongs to haplogroup 2 (Pires *et al.* 2018); note that these sequences do not include all the diversity of Eurasian wolves because the search was not exhaustive for modern wolves), spanning a **43 bp** common fragment of the Control Region (CR) (position 15,610-15,652, CanFam3.1) in the Iberia and Eurasia context alignment and **66bp** common fragment of the CR (position 15,587-15,652 bp, CanFam3.1) in the Iberia context alignment (Table 4). All the sequences had to be reduced to a shorter length (43bp or 66bp) because, after the alignment, some positions were uncovered for some fragments because some sequences were shorter than others (e.g ancient Eurasian wolves sequences retrieved from Stiller *et al.* (2006) are fragments of 57 bp). Although a short portion of the Hypervariable Region 1 (HVR1) of the mitochondrial DNA was analysed, it is within a highly

informative region used to discriminate all the major Haplogroups, because it includes most of the diagnosed positions to separate the Haplogroups from each other (15595-15653 bp; Himmelberger *et al.* 2008).

Modern dogs and wolves from Eurasia were not included in this analysis due the large number of sequences available at GenBank (5,380 entries for the dog and 1315 entries for the wolf).

For the fragment analysed in this study (15,587-15,652 bp, CanFam3.1), LYEP9 and LYEP53 sequenced with the NGS-Illumina technology had unknown/missing nucleotides (N) at the 15,603 bp and 15,643 bp, respectively. In order to use these samples in the phylogenetic analysis, we complemented information about these nucleotides position, using the data retrieved by the NGS-454 method (Pires *et al.* 2019). In the case of the LYEP51 sample, the sequence obtained by NGS-Illumina could not be used because at the positions correspondent to that used for the phylogenetic analyses (15587-15652 bp, CanFam3.1) it had a high proportion of many missing genotype calls. Alternatively, for this sample, we used the sequence obtained by the NGS-454 method (Pires *et al.* 2019).
Finally, the nexus file generated in Mega with sequences alignment was imported into DnaSP v6.12.03 to collapse the sequences within identical nucleotide sequences (haplotypes) for each of the eight alignments (sub datasets; see Table 4) containing different sequence sizes (43 or 66bp).

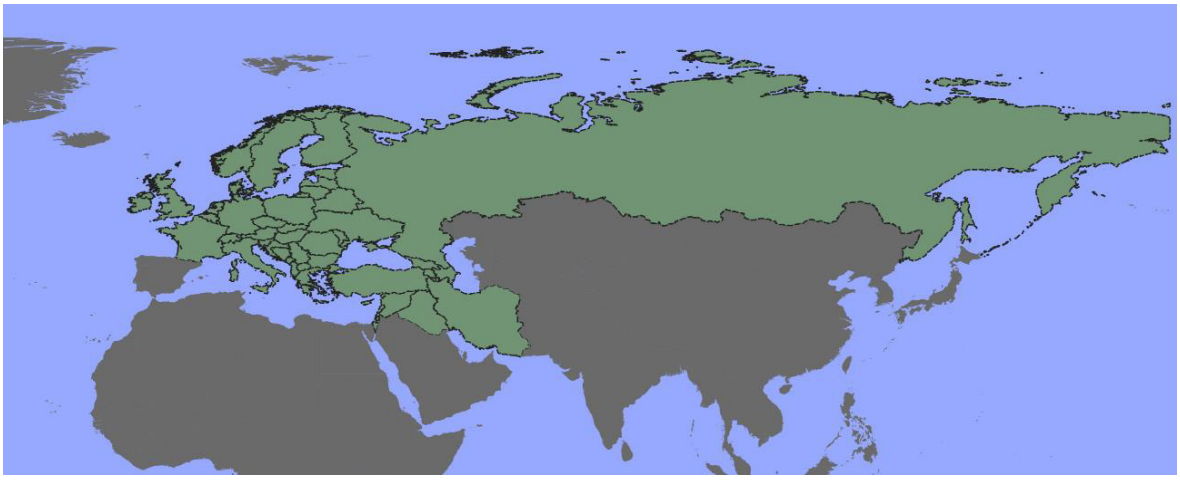


Figure 8. Extension of geographic region selected (in green) for Eurasian context to compare with Iberia.

2.6.2 Phylogenetic networks

PopART v1.7 (Leigh & Bryant 2015) was used to build MJ networks connecting mtDNA haplotypes from ancient and extant *Canis.* Eight MJ based on the different datasets comprising different nucleotide sequences (length of alignment) and different geographic and temporal representativeness (number of sequences) of the total sequences, were also generated to a better understanding of the evolution of the Iberian dog through time and space (Table 4). The decision of exclude ancient periods after the Chalcolithic period is based on the focus of this study, which is to infer the evolutionary trajectories and the genetic composition of the Chalcolithic dogs from Iberia. Artistic edition of networks was made with GIMP v2.10.8.

Table 4. Chronological and geographic information for each network.

Context	Canid type	Cultural Period	Alignment size
Iberia	Dog	Mesolithic+Neolithic+Chalcolithic+Extant	66 bp
	Wolf	Palaeolithic+Chalcolithic+Extant	66 bp
	Dog & Wolf	Palaeolithic+Mesolithic+Neolithic+Chalcolithic+Extant	66 bp
Iberia & Eurasia	Dog & Wolf	Palaeolithic	43 bp
	Dog & Wolf	Palaeolithic+Mesolithic	43 bp
	Dog & Wolf	Palaeolithic+Mesolithic+Neolithic	43 bp
	Dog & Wolf	Palaeolithic+Mesolithic+Neolithic+Chalcolithic	43 bp
	Dog & Wolf	Palaeolithic+Mesolithic+Neolithic+Chalcolithic+Extant	43 bp

2.7 Statistical analysis - Genetic distance between populations

GenAlEx v6.5 (Peakall & Smouse 2006, 2012) was used for estimating genetic distances among populations. Different populations were defined for this analysis, taking into account the geographic origin of samples – Iberia and Eurasia – and chronological period – Palaeolithic, Mesolithic, Neolithic, Chalcolithic and Extant. All the 251 sequences (with the exception of the Chalcolithic wolf, since there is only one specimen from the Chalcolithic Iberian wolf population) were maintained, considering a small fragment of 43 bp with no missing data, and a larger one of 182 bp where several missing data are present – 19% (the software GenAlEx accepts positions with missing data, assigning a special value for these positions when doing the statistical analysis). Population structure was investigated by pairwise PhiPT values, an analog of the Fixation index (Fst) (Wright 1922) parameter,

adequate for sequences analyses. PhiPT suppresses within-population variance and simply calculate population differentiation based on the haplotypic variance. The probability values estimated by 1,000 permutations were used to determine whether the partitioning of variance components was significant. Population differentiation values that ranges from 0 to 1 - where 0 means complete sharing of genetic material (high gene flow, no genetic structure) and 1 means genetic divergency (low or no gene flux and high genetic structure; limited by population's homozygosity).

According to Wright's (1978) qualitative guidelines for the interpretation of Fst, values can have the following interpretation:

- The range 0 to 0.05 may be considered as indicating little genetic differentiation;
- The range 0.05 to 0.15 indicates moderate genetic differentiation;
- The range 0.15 to 0.25 indicates great genetic differentiation;
- Values of FST above 0.25 indicate very great genetic differentiation.

Nm (Number of migrants) values based on PhiPT values were also estimated to calculate the average number of individuals migrating between populations/per generation time which in dogs and wolves are approximately 3 years (Lindblad-Toh *et al.* 2005).

2.8 Nuclear genome analysis: Sex determination of samples

Due to the very low recovery of endogenous nuclear genomes (Appendix IX), it was not possible to provide information regarding complete nuclear genomes. In this work, only sex determination analysis could be carried out. The reference CanFam3.1 genome (a female from the Boxer breed) lacks the Y chromosome, therefore reference sequences of the Y chromosome for the subsequent analysis were retrieved from different studies available at GenBank (Accession numbers KP081776; GQ366706-GQ366731; GQ366741-GQ366770; GQ366790-GQ366793; DQ973626-DQ973805). For each sample, the reads that did not map against CanFam3.1 genome were aligned against those Y sequence fragments, in order to molecularly identify the sex through their archaeological remains. Additionally, a read depth-based method was used, by comparing the ratio of reads/Mbp over all

chromosomes in order to check the read proportion on chromosomes X and/or Y compared with the remaining chromosomes. Females are expected to have a similar ratio when comparing chromosome X with autosomal chromosomes. Males are expected to present similar ratio of chromosome X and Y, but half of the ratio, when comparing sexual chromosomes with autosomal chromosomes.

For the wolf sample, it was not possible to map against a wolf reference nuclear genome, because the only de novo assembly available (Gopalakrishnan *et al.* 2017) is consisted of unplaced scaffolds; it is not known to which chromosome each scaffold belongs to. Thus, assignment of LYEP27 sex was employed using dog reference CanFam3.1 genome and the Y chromosome reference sequences mentioned above.

3. RESULTS

3. Results

3.1 Bioinformatic analysis

3.1.1 Raw-read processing

Read quality dropped most of the times towards the end of the fragments, probably due the Illumina adapters that sometimes can continue present in many reads despite the fact that reads were already filtered from the adapters. After removal of low-quality bases (≤30, as a minimum quality threshold), all the remaining reads showed at least the minimum desired quality threshold of ≥30 (Appendix VII) and the persisting adapters were removed.

3.1.2 Mapping

A table summarizing the number of reads mapping against reference genomes was generated (Appendix IX). Contaminant reads that could exist in the datasets were filtered out as described in Materials and Methods section. This resulted in the exclusion of only 0,0008% and 0,05% reads per sample, on average, for mitochondrial and nuclear genome, respectively. A highest percentage of mapped reads against endogenous nuclear in comparison with the mtDNA endogenous values is due to the shorter length of the mtDNA compares to the nuclear one. An alternative analysis (not shown) using the EAGER software (Peltzer *et al.* 2016) for the same reads generated almost the same results. Due to these results, a BLAST assignment was performed to find the source of contamination. Most of the discarded reads had a bacterial contamination origin (see section 3.1.6).

3.1.3 DNA degradation

The proportion of 5' Cytosine to Thymine (C>T) and 3' Guanine to Adenine (G>A) increases towards the end of the reads, as showed in mapDamage analysis charts (T: red line; A: blue line; Figure 9), demonstrating that all samples exhibited some level of degradation, and therefore ancient DNA's typical characteristics.

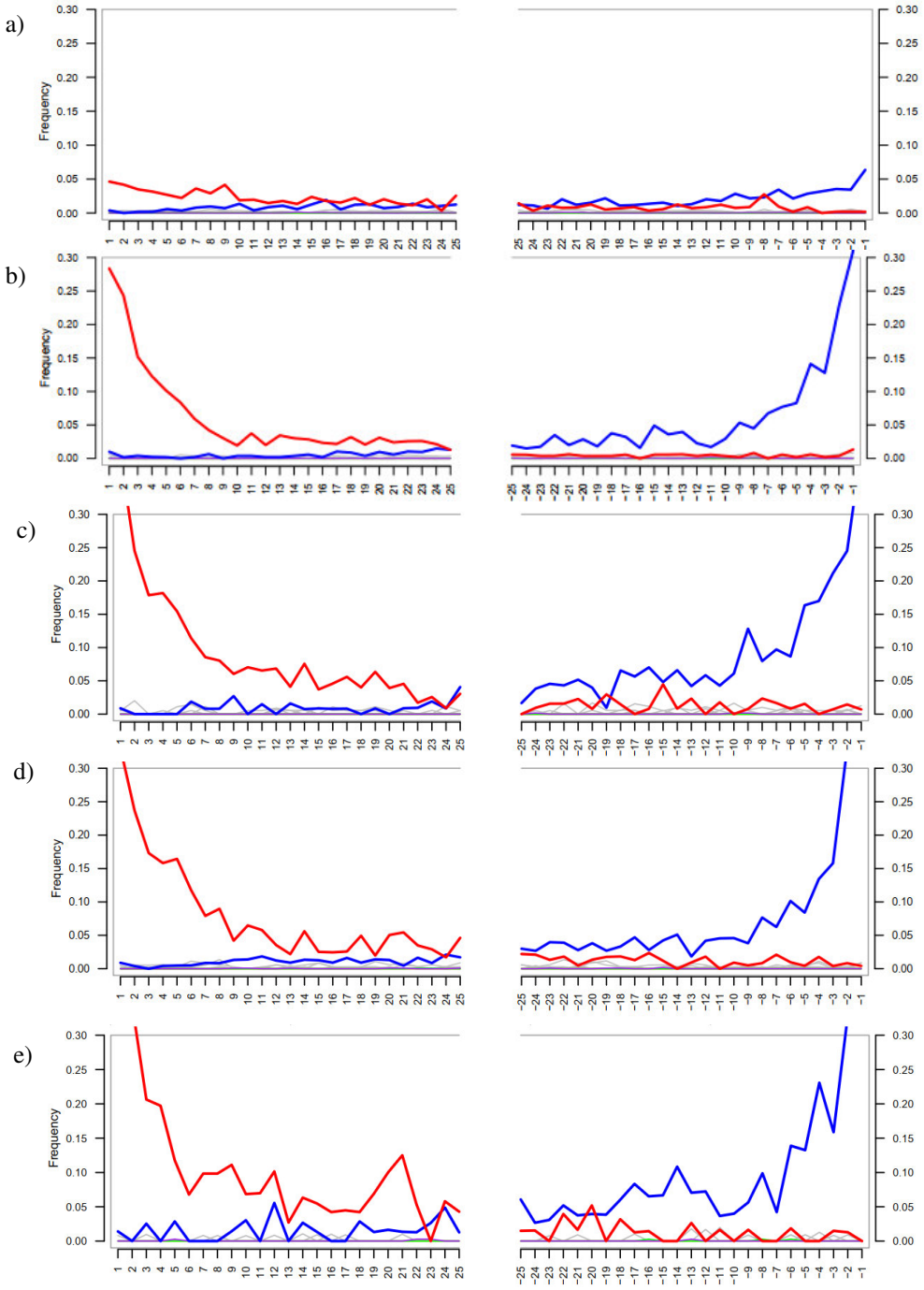


Figure 9. Substitution patterns at the 5′ and 3′ ends of the sequence typical of ancient DNA. a) *LYEP9; b) LYEP11; c) LYEP51; d) LYEP53 e) LYEP27.*

3.1.4 Consensus mitochondrial sequence

All fasta files created in this study are not available for public access. In a near future, the 5 ancient DNA samples analysed in this study will be rescreened in a specific way to increase their sequences coverage.

As expected, when analysing ancient DNA, sequences have unknown/missing bases (Ns) in different amounts (Table 5). The sequences length is in concordance with mitochondrial DNA dog reference sequence length (16,729 bp) or wolf reference sequence length (16,757 bp).

A table containing the information of mitochondrial variants (nucleotide substitution and position in genome) identified when mapping each sample against the reference mitochondrial genome can be consulted in Appendix VIII.

Table 5. mtDNA fasta files summary regarding information of sequence length and number of unknown bases for each sample.

Sample	Sequence length (bp)	Number of unknown bases (Ns)	% of unknown bases
LYEP9	16,732	268	1.60
LYEP11	16,730	143	0.85
LYEP51	16,727	4,012	23.99
LYEP53	16,731	543	3.25
LYEP27	16,757	10,231	61.06

3.1.5 Species identity assignment

The genome coverage results indicate that LYEP9, LYEP11, LYEP51 and LYEP53 samples were dogs, since these samples showed higher coverage values when they were mapped against the dog reference mtDNA genome. In the case of LYEP27, this sample belongs to a wolf (Table 6). However, in the context of ancient analysis, genetic taxonomic assignment is better used to corroborate osteometric and archaeological context analysis, due to the typical degradation of aDNA that may affect the alignment against reference genomes and lead to incorrect species identification.

Table 6. Coverage results of samples mapped against dog and wolf genome.

Mitochondrial Coverage	Sample	mapped against dog genome	% of merged reads mapped against endogenous mtDNA (%)	mtDNA % coverered >2	mean coverage of mtDNA genome	Endogenous mtDNA (%) [1]	mean coverage of mtDNA genome [1]	mapped against wolf genome	mtDNA % coverered >2	mean coverage of mtDNA genome	Species assignment
Illumina - Stockholm University	LYEP9		0.010	99.20	17x	0.010	19x		98.60	17x	Dog
Illumina - SU	LYEP11		0.012	99.28	12x	0.010	17x		97.88	12x	Dog
Illumina - SU	LYEP51		0.002	86.44	2x	0.002	3x		76.60	2x	Dog
Illumina - SU	LYEP53		0.005	98.94	5x	0.004	6x		95.20	6x	Dog
Illumina - SU	LYEP27		0.005	36.98	1x	0.002	3x		38.97	1x	Wolf

Nuclear Coverage	Sample	mapped against dog genome	% of merged reads mapped against endogenous nDNA (%)	nDNA % coverered >2	mean coverage of nDNA genome	Endogenous nDNA (%) [1]	mean coverage of nDNA genome [1]	mapped against wolf genome	nDNA % coverered >2	mean coverage of nDNA genome	Species assignment
Illumina - Stockholm University	LYEP9		3.75	0.001	0.043x	3.05	0.054x		0.001	0.043x	Dog
Illumina - SU	LYEP11		0.55	2.55E-05	0.002x	0.82	0.008x		2.81E-05	0.002x	Dog
Illumina - SU	LYEP51		0.95	4.85E-05	0.006x	0.84	0.01x		1.26E-10	0.006x	Dog
Illumina - SU	LYEP53		0.09	2.79E-06	0.0005x	0.11	0.001x		4.15E-06	0.0005x	Dog
Illumina - SU	LYEP27		0.64	1.48E-05	0.0009x	0.35	0.003x		1.52E-05	0.0009x	Wolf

[1] Using EAGER software.

3.1.6 Contaminants – taxonomic assignment

Regarding the contaminants, the majority of Blast hits have bacterial origin (Figure 10). From 1,000,000 reads analysed, match hits were found only in 287,956 reads for LYEP9, 105,302 reads for LYEP11, 89,953 reads for LYEP51, 106,580 reads for LYEP53 and 77,699 reads for LYEP27. Almost 90% of the reads analysed, on average, that do not align against reference genomes also did not find correspondence with NCBI database sequences. These

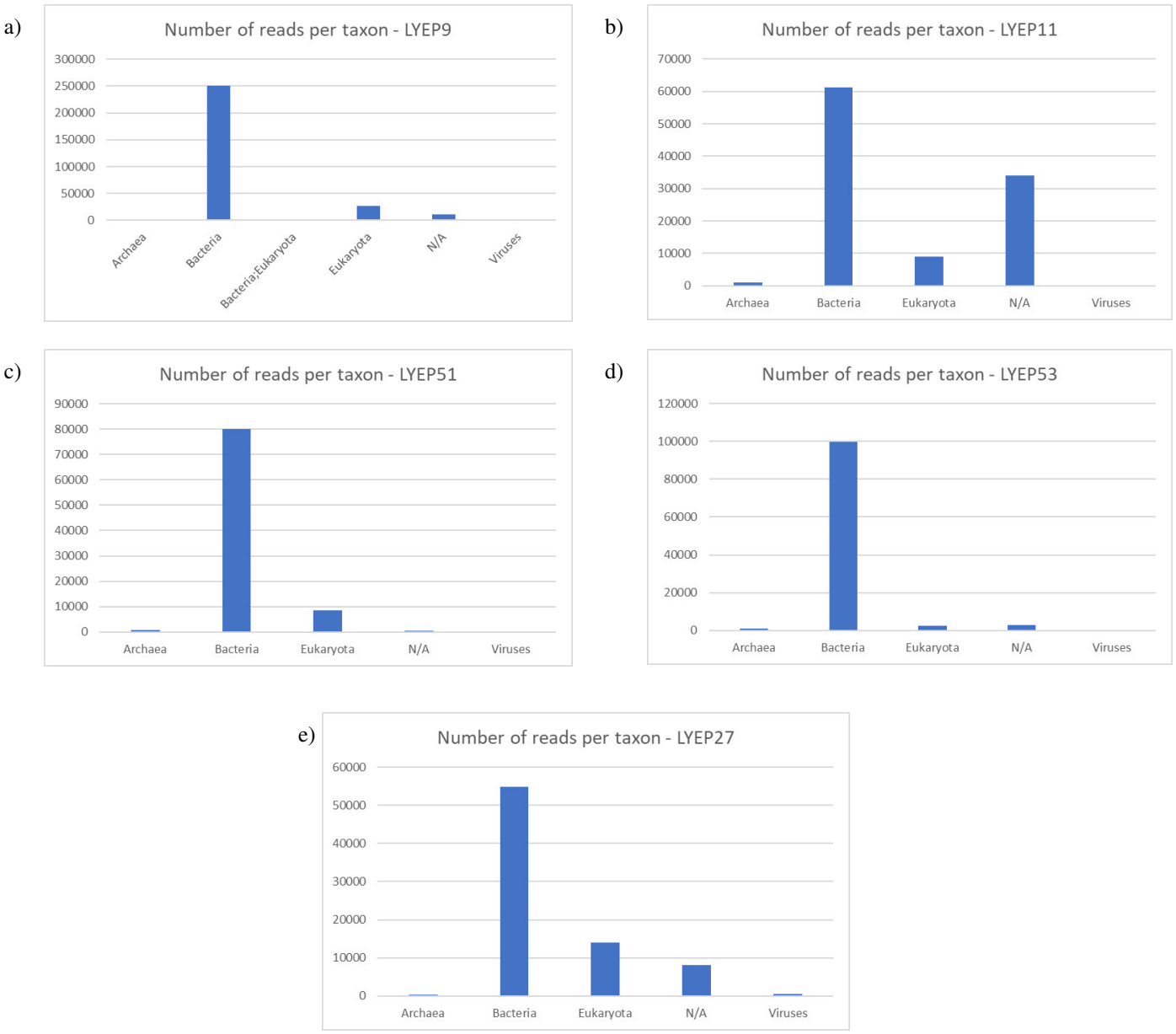


Figure 10. Number of contaminated reads out of 1M reads per taxon analysed in Blastn software. a) sample LYEP9; b) sample LYEP11; c) samples LYEP51; d) sample LYEP53; e) sample LYEP27. "N/A" results are BLAST hits of organisms not classified in NCBI database.

reads are likely sequences artefacts that are common to occur in aDNA due to the existence of few template molecules to initiate amplification (Kircher et al. 2011).

3.2 Genetic diversity

3.2.1 Phylogenetic tree

Bayesian phylogenetic tree of complete sequences (Figure 11) confirm the Haplogroup assignment of the Chalcolithic dogs: LYEP9 and LYEP11 segregates within Haplogroup A; LYEP51 and LYEP53 segregates within Haplogroup C.

One of the extant Iberian dogs (eDog_D6_ES/PT_HgD) segregates close to ancient wolves from Switzerland (aWolf26_SWI_Hg2 and aWolf40_SWI_Hg2). The other extant Iberian dog (eDog_A34_ES_HgA) segregates close to Chalcolithic dogs (aDog_LYEP9_PT and aDog_LYEP11_PT).

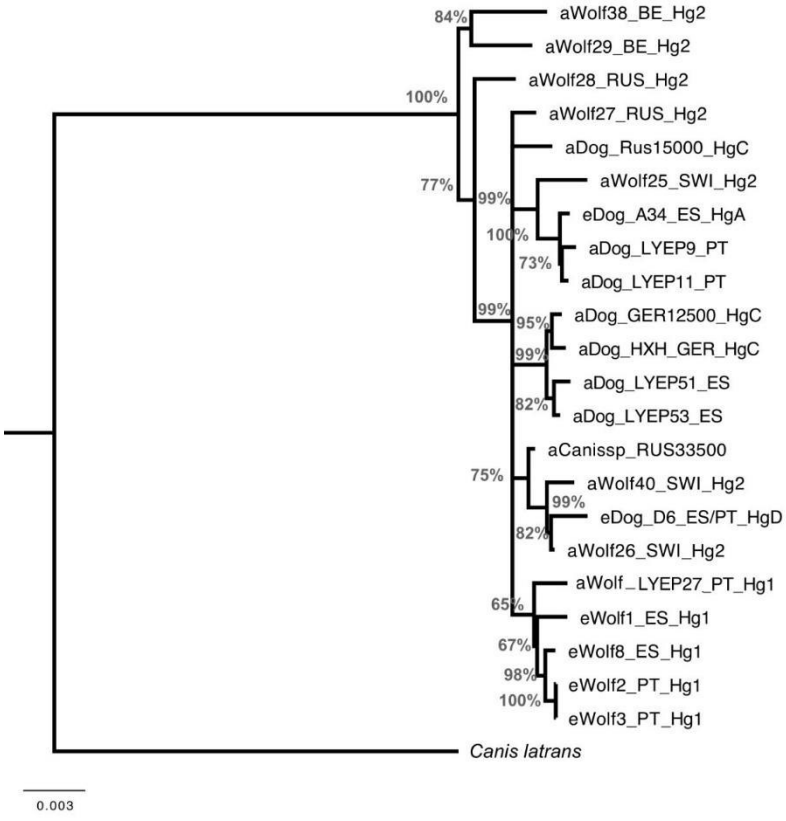


Figure 11. Bayesian phylogenetic tree (1,100,000 iterations) for complete mtDNA. Bayesian support values are indicated at each node. Ancient samples initiate with "a" and extant samples with "e". Samples without a date indicate extant samples. At each branch, the samples identifiers and geographic origin are also indicated. The scales on bottom measure evolutionary distance in substitutions per nucleotide. See Appendix V for detailed information regarding the source of the sequences displayed here.

3.2.2 Phylogenetic networks

It is important to recall here that in order to accommodate data from other authors, the mitogenomes determined for the 5 *Canis* samples of this study were trimmed to 66bp and 43 bp in the case of Iberian context and Iberian and Eurasian context, respectively (see Table 4 in Material and Methods).

To understand the origin and diversity of the Chalcolithic dogs it will be shown several networks. These will be presented firstly by canid type from the Palaeolithic to the Chalcolithic Iberian and them chronologically from the Palaeolithic to Chalcolithic periods of dogs and wolves from Iberian and Eurasian.

The following phylogenetic networks display the phylogenetic relationships among ancestral dogs from Iberia and dogs and wolves from different chronological contexts. For the Iberian context, and when only partial mtDNA sequences from dogs were used, the network generated (Figure 12) shows that only mtDNA dog Haplogroups A and C are present in Iberia since the Mesolithic until the Chalcolithic period. Among the 24 haplotypes displayed here, none of the ancient samples segregated within dog Haplogroup B or D. Haplogroup A is the most diverse Haplogroup for the Chalcolithic period (5 out of 7 haplotypes, in contrast with only two haplotypes in Haplogroup C) as happens nowadays (in extant dogs; 13 out of 22 haplotypes).

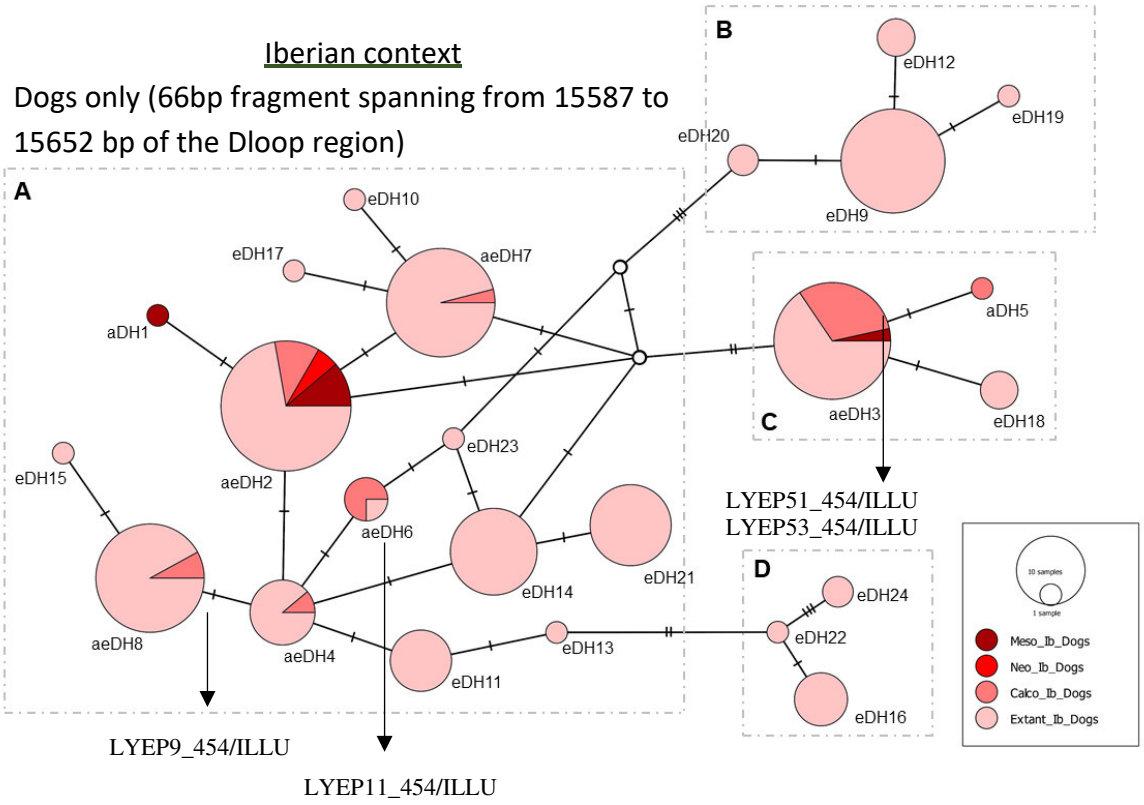
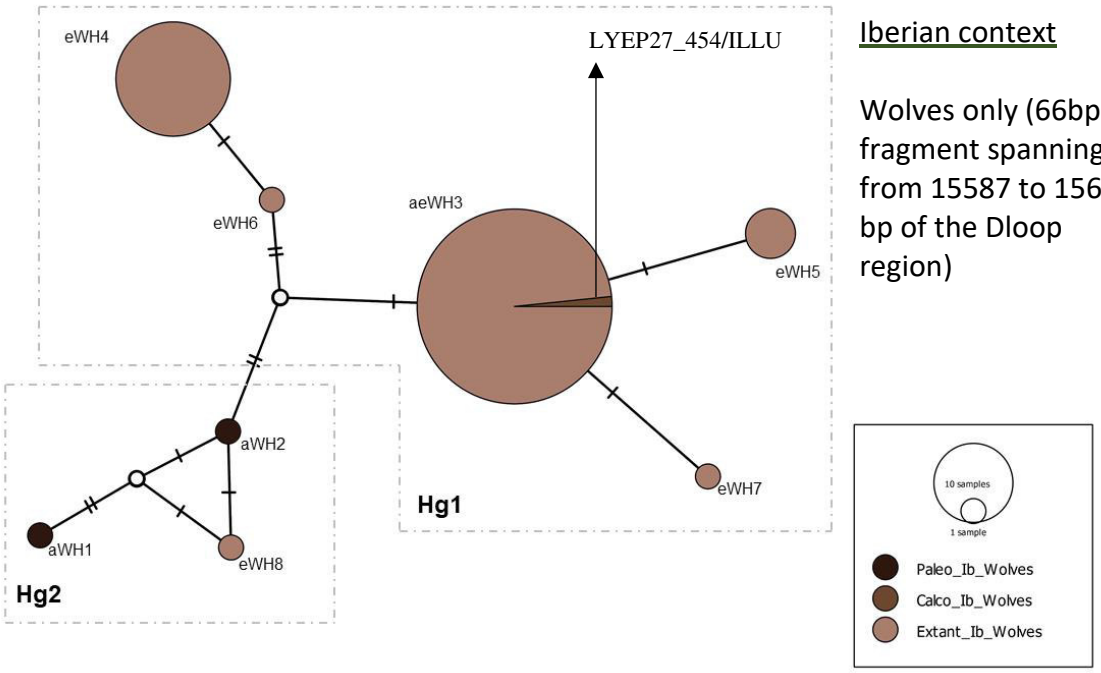


Figure 12. Median-Joining network showing genetic relationships among Iberian mtDNA dog haplotypes (partial sequences) of phylogenetic Haplogroups A, B, C and D. "a" stands for ancient; "e" stands for extant, "W" for wolf, "D" for dog and "H" for haplotype. Colored circles represent different chronological periods. White circles are median vectors (hypothetical intermediates). The sizes of colored circles are proportional to the haplotype frequency in the respective populations. Newly generated data from this thesis (NGS) are: LYEP9 and LYEP11 which segregates within HgA; LYEP51 and LYEP53 which segregates within HgC. The remaining sequences are from Pires et al (2019), Pires et al (2006) and Pang et al (2009). See Appendix VI for detailed information regarding the source and haplotype correspondence of the haplotypes displayed here.

For the Iberian context, and regarding only Iberian wolf partial mtDNA sequences, the resulting network (Figure 13) display only two wolf Haplogroups. Following Pilot *et al.* (Pilot *et al.* 2010) nomenclature, two Palaeolithic Portuguese wolves from Pires et al. (2019) study and a single extant Portuguese wolf from Pires *et al.* (Pires *et al.* 2017) segregate within Wolf_Hg2. The unique Chalcolithic wolf haplotype from Iberia (and elsewhere) segregates within Wolf_Hg1, together with haplotypes from extant Iberian wolves (Vilà *et al.* 1997; Bjornerfeldt *et al.* 2006; Koepfli *et al.* 2015; Koblmüller *et al.* 2016; Pires *et al.* 2018).

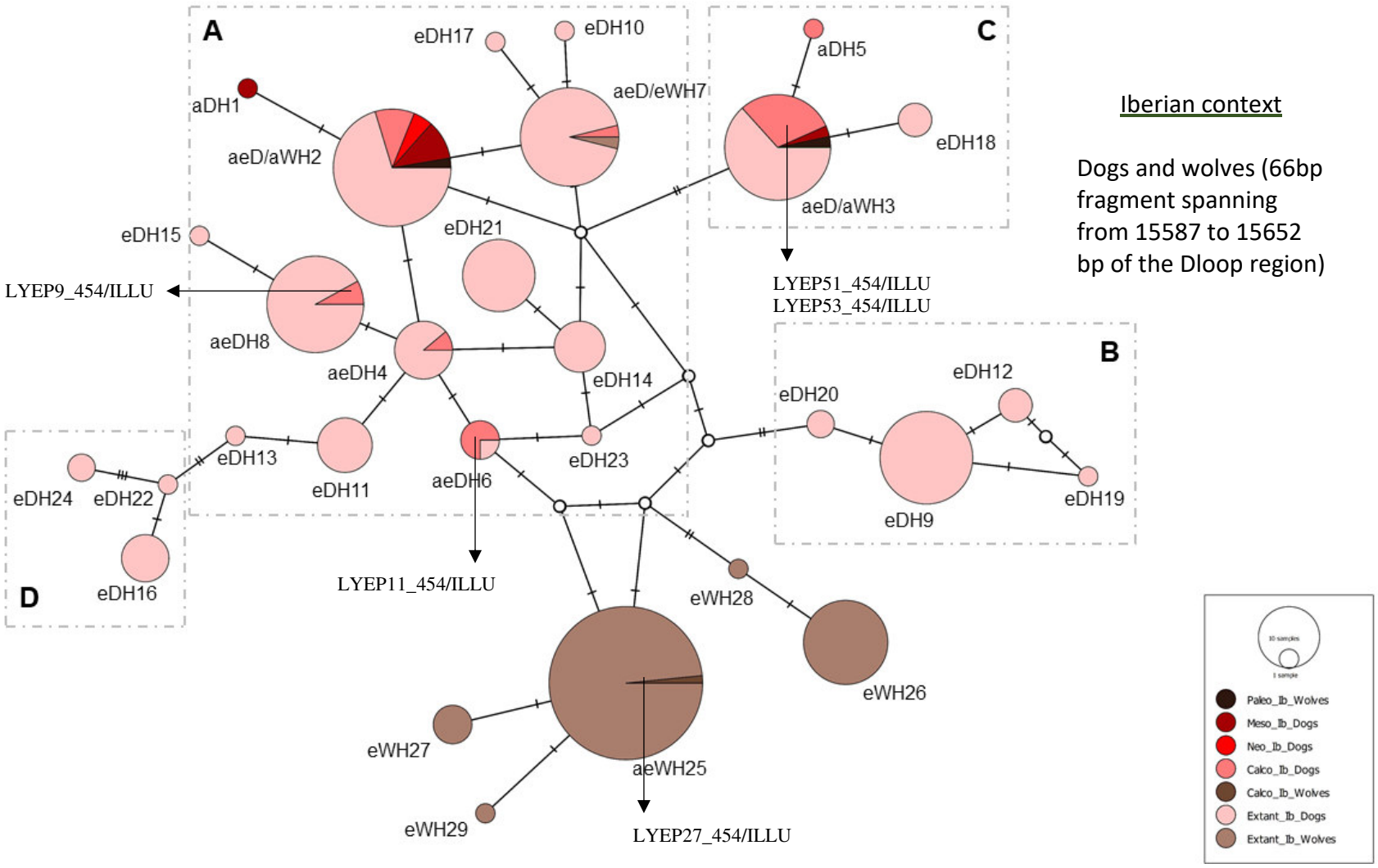


Iberian context

Wolves only (66bp fragment spanning from 15587 to 15652 bp of the Dloop region)

Figure 13. Median-Joining network showing genetic relationships among Iberian mtDNA wolf haplotypes (partial sequences) of phylogenetic Haplogroups 1 and 2. "a" stands for ancient; "e" stands for extant, "W" for wolf, "D" for dog and "H" for haplotype. Coloured circles represent different chronological periods. White circles are median vectors (hypothetical intermediates). The sizes of coloured circles are proportional to haplotype frequency in the respective populations. Newly generated data from this thesis (NGS) is: LYEP27 which segregates within Wolf_Hg1. The remaining sequences are from Bjornerfeldt et al (2006), Koblmuller et al (2016), Koeplfi et al (2015), Parra et al unpublished, Pires et al (2017), Pires et al (2019), Randi et al unpublished and Vila et al (1997). See Appendix VI for detailed information regarding the source and haplotype correspondence of the haplotypes displayed here.

When analysing a network comprising both wolf and dog haplotypes from the Iberia from different periods (Figure 14), both Palaeolithic wolves (aeD/aWH2 and aeD/aWH3) and one extant wolf from Hg1 segregates within dogs' Haplogroups. Hg2 haplotypes from extant wolves (eWH25-H29) are well segregated from dog haplotypes.



Iberian context

Dogs and wolves (66bp fragment spanning from 15587 to 15652 bp of the Dloop region)

Figure 14. Median-Joining network showing genetic relationships among Iberian mtDNA Canis haplotypes of phylogenetic Haplogroups A, B, C and D (dogs). "a" stands for ancient; "e" stands for extant, "W" for wolf, "D" for dog and "H" for haplotype. Coloured circle represents different chronological periods and species. White circles are median vectors (hypothetical intermediates). The sizes of coloured circles are proportional to haplotype frequency in the respective populations. See Appendix VI for detailed information regarding the source and haplotype correspondence of the haplotypes displayed here.

A network based on Palaeolithic dogs and wolves haplotypes from Iberian and Eurasian (Figure 15), apparently do not show a clearly structure that separate haplotypes among Haplogroups. However, using data from other authors and reference haplotypic data, it was possible to assign Haplogroups. This network displays the segregation of LYEP46* – haplotype aW/DH4 together with two Palaeolithic dogs from Germany and France (both HgC), and one Palaeolithic wolf from Italy (Wolf_Hg2). Sample LYEP44** - haplotype aWH17, is a unique haplotype from the Palaeolithic of Iberian and Eurasian context. Regarding the specimens with ambiguous assignment (*Canis from Goyet Cave, Belgium* and

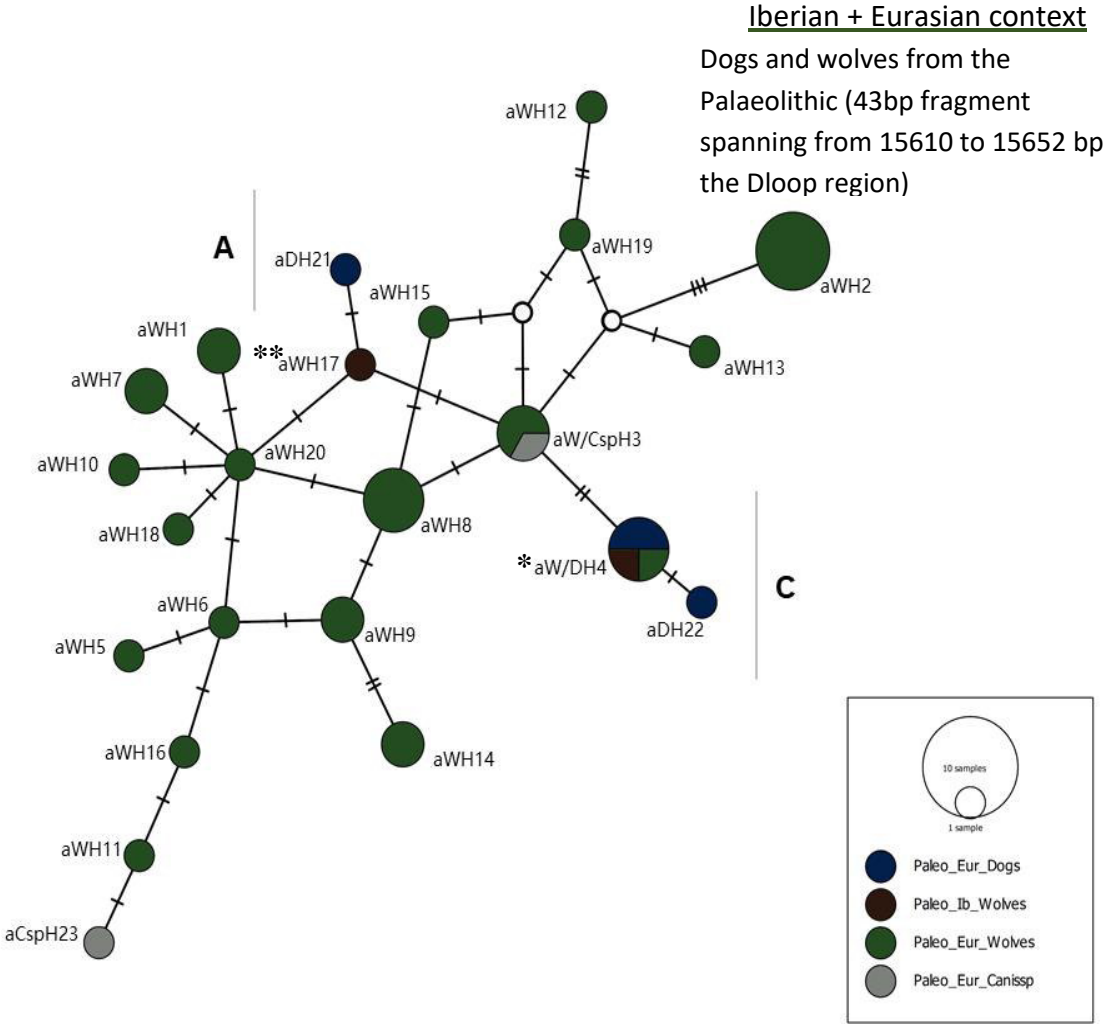


*Figure 15. Median-Joining network showing genetic relationships among Iberian and Eurasian mtDNA Canis haplotypes segregating within Haplogroups A and C (dogs). "a" stands for ancient; "e" stands for extant, "W" for wolf, "D" for dog and "H" for haplotype. Coloured circles represent different Palaeolithic Canis. White circles are median vectors (hypothetical intermediates). The sizes of coloured circles are proportional to haplotype frequency in the respective populations. None of these sequences were generated under this study. See Appendix VI for detailed information regarding the source and haplotype correspondence of the haplotypes displayed here. *Haplotype where LYEP46 is included; **Haplotype where LYEP44 is included.*

Razboinichya Cave, Russia), they segregate with other wolves, in this study. So, thereafter these samples are considered as wolves.

After adding Mesolithic period to the network (Figure 16), it can be observed that HgC is more present in Eurasia (more haplotypes, higher frequencies), in contrast to HgA that is more present in Iberia. Palaeolithic wolf LYEP46 and Mesolithic dog LYEP68_A (aW/DH4*), segregate with Mesolithic dogs from France, Romania and Estonia and one Palaeolithic dog from Germany and one Palaeolithic wolf from Italy (in Haplogroup C). Regarding wolf

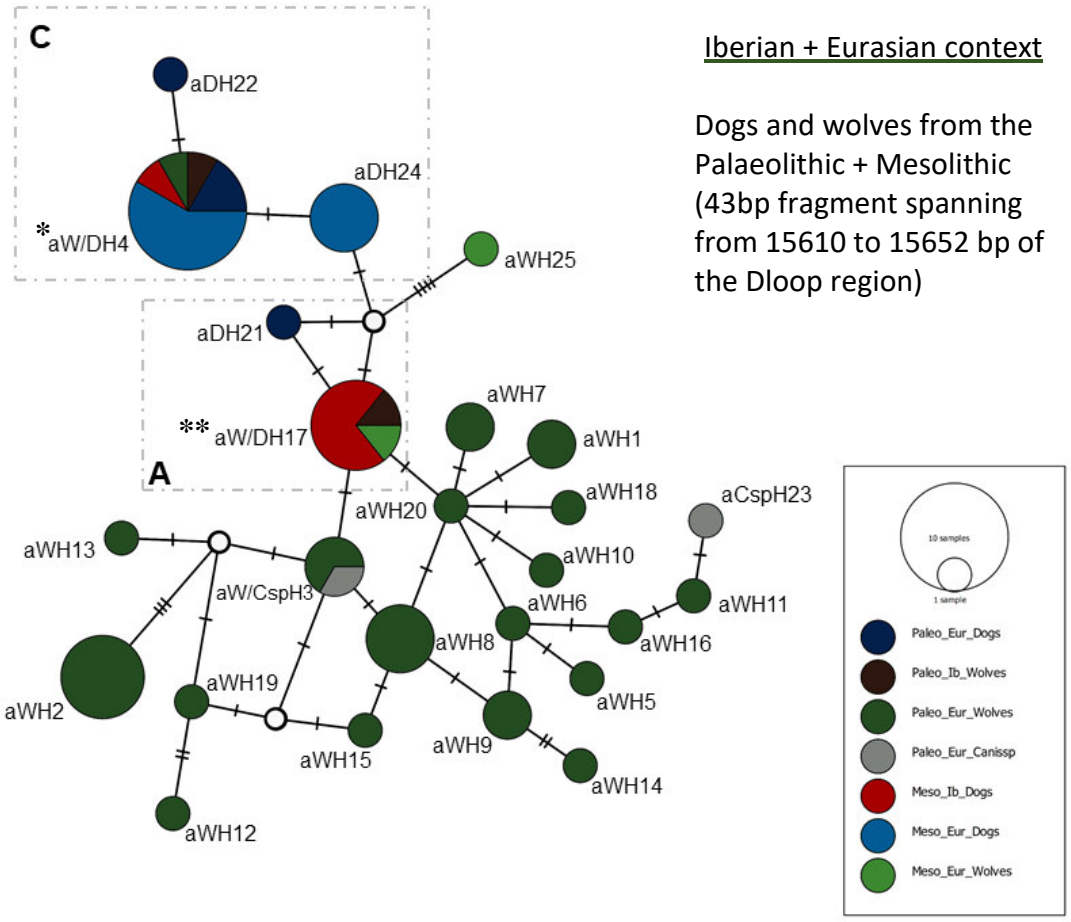


*Figure 16. Median-Joining network showing genetic relationships among Iberian and Eurasia mtDNA Canis haplotypes. "a" stands for ancient; "e" stands for extant, "W" for wolf, "D" for dog and "H" for haplotype. Coloured circles represent different chronological periods and species. White circles are median vectors (hypothetical intermediates). The sizes of coloured circles are proportional to haplotype frequency in the respective populations. None of these sequences were generated under this study. See Appendix VI for detailed information regarding the source and haplotype correspondence of the haplotypes displayed here. *Haplotype where LYEP46 and LYEP68_A are included; **Haplotype where LYEP44 and Mesolithic Iberian dogs are included.*

LYEP44 and most of the Mesolithic Iberian dogs (aW/DH17**), they share the same haplotype with a Mesolithic Italian wolf (in clade A).

In the Palaeolithic+Mesolithic+Neolithic network (Figure 17), it is described the appearance of dogs carrying haplotypes segregating in Haplogroups D and B in the Neolithic but outside Iberia. Only two Iberian dog specimens are dated to the Neolithic period, they share the same haplotype which clusters within Haplogroup A. A larger sampling is necessary for this period.

Iberian + Eurasian context

Dogs and wolves from the Palaeolithic + Mesolithic + Neolithic (43bp spanning from 15610 to 15652 bp of the Dloop region)

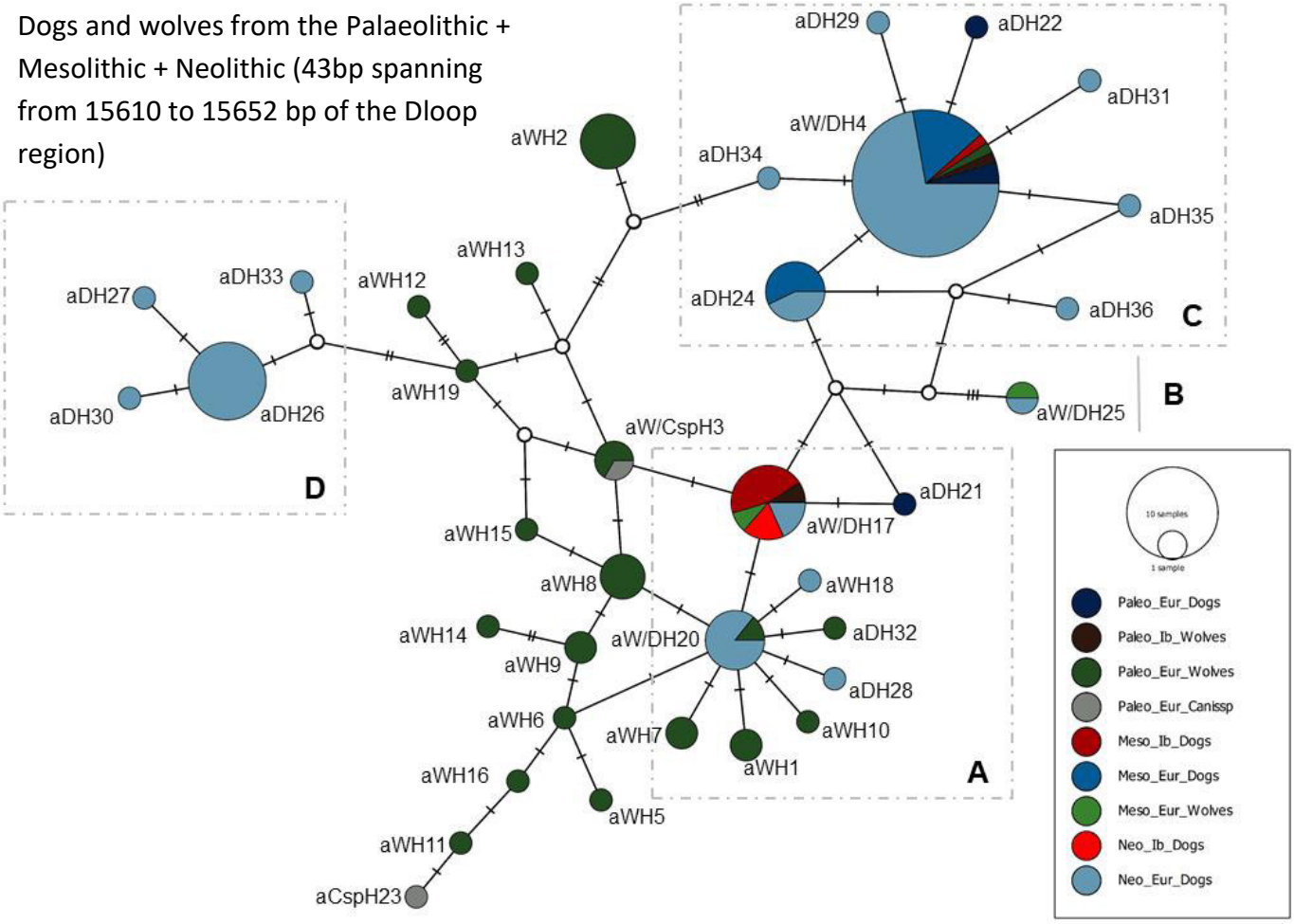


Figure 17. Median-Joining network showing genetic relationships among Iberian and Eurasia mtDNA Canis haplotypes of phylogenetic Haplogroups A, B, C and D (dogs). "a" stands for ancient; "e" stands for extant, "W" for wolf, "D" for dog and "H" for haplotype. Coloured circles represent different chronological periods and species. White circles are median vectors (hypothetical intermediates). The sizes of coloured circles are proportional to haplotype frequency in the respective populations. See Appendix VI for detailed information regarding the source and haplotype correspondence of the haplotypes displayed here.

According to the next network (Figure 18), in the Chalcolithic period, the single Iberian wolf sample analysed (LYEP27, one of the samples from this study) held a unique haplotype (aWH40^{+}). Chalcolithic Iberian dog haplotypes (n=7) are segregating within Haplogroups A and C. The samples from this study are LYEP9, LYEP11, LYEP51, LYEP53 and LYEP27 -

Iberian + Eurasian context

Dogs and wolves from the Palaeolithic + Mesolithic + Neolithic + Chalcolithic (43bp fragment spanning from 15610 to 15652 bp of the Dloop region)

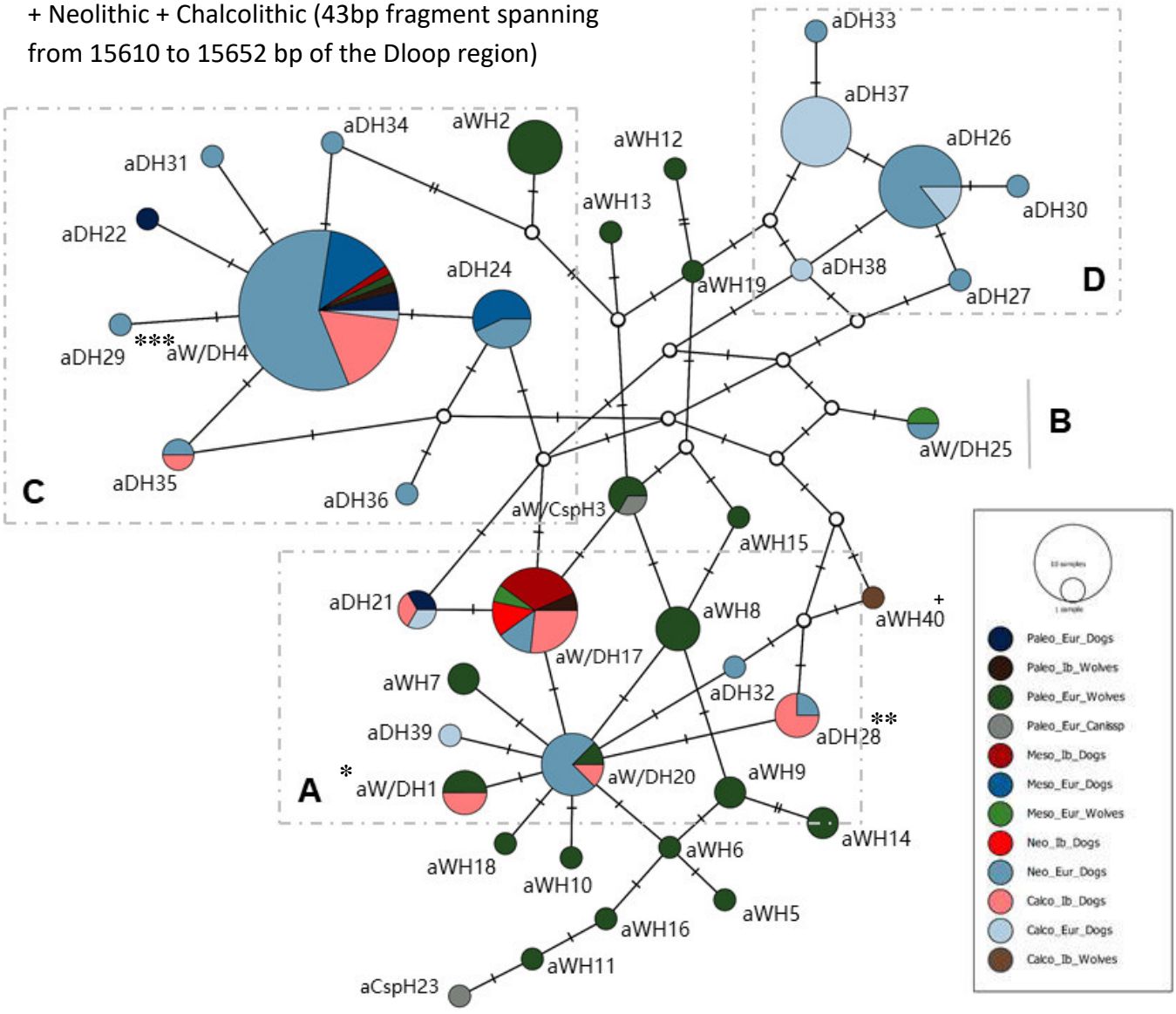


*Figure 18. Median-Joining network showing genetic relationships among Iberian and Eurasia mtDNA Canis haplotypes of phylogenetic Haplogroups A, B, C and D (dogs), spanning Palaeolithic, Mesolithic, Neolithic and Chalcolithic periods. "a" stands for ancient; "e" stands for extant, "W" for wolf, "D" for dog and "H" for haplotype. Coloured circles represent different chronological periods and species. White circles are median vectors (hypothetical intermediates). The sizes of coloured circles are proportional to haplotype frequency in the respective populations. See Appendix VI for detailed information regarding the source and haplotype correspondence of the haplotypes displayed here. *Haplotype where LYEP9 is included; **Haplotype where LYEP11 is included; ***Haplotype where LYEP51 and LYEP53 are included; +Haplotype where LYEP27 is included.*

haplotypes aW/DH1*, aDH28**, aWDH4*** and aWH40^{+} respectively, from HgA and HgC. Haplotype aW/aeDH4 (HgC) is the most frequently found and is detected in every period.

Below, the next network represents mtDNA haplotypes detected in Iberian and Eurasian dogs and wolves from Palaeolithic, Mesolithic, Neolithic, Chalcolithic and Extant periods (Figure 19). Haplotype aDH35 no longer exists in extant Iberian dogs.

Extant Iberian wolves are well structured/separated from wolves of other European region, with exception of eWolf25 (Bulgaria) and aWolf4 (Portugal) that share the same haplotype aeW/aeDH17 (in the network this haplotype can be found segregating within dog Haplogroup A), and eWolf18 (Latvia/Russia/Sweden) and eWolf28 (Portugal) that also share the same haplotype eW/aeDH21 (in the network this haplotype can be found segregating within dog Haplogroup A). In the most geographically and temporally best represented haplotypes (aW/aeDH4 and H17) are included the Palaeolithic Iberian wolves (samples aWolf3 and aWolf4).

Regarding extant Iberian wolf Haplogroups, there is not a clear separation between wolf_Hg1 and wolf_Hg2, since 94% (n=16) belongs to wolf_Hg1 and 6% (n=1) to wolf_Hg2. As mentioned in Materials and Methods, only some representativeness of Eurasian wolves from Haplogroup 2 were used in this analysis. For this reason, beside the known existence of two distinct mtDNA haplotypes in extant Italian wolves (Randi *et al.* 2000; Boggiano *et al.* 2013; Montana *et al.* 2017), here only one haplotype appears (eWH54).

Iberian + Eurasian context

Dogs and wolves from the Palaeolithic + Mesolithic + Neolithic + Chalcolithic + Extant (43bp spanning from 15610 to 15652 bp of the Dloop region)

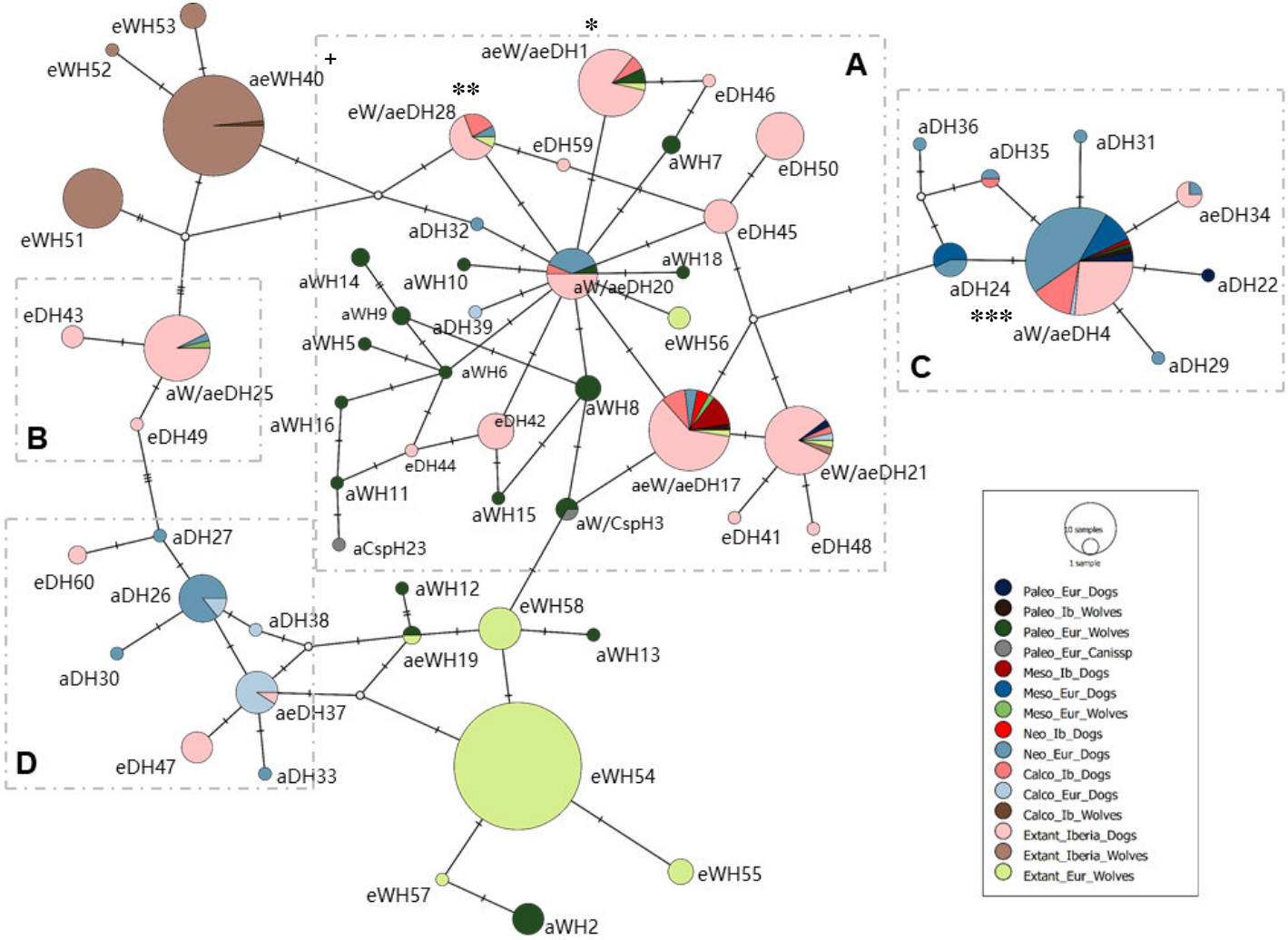


*Figure 19. Median-Joining network showing genetic relationships among Iberian and Eurasia mtDNA Canis haplotypes of phylogenetic Haplogroups A, B, C and D (dogs), spanning Palaeolithic, Mesolithic, Neolithic and Chalcolithic and Extant periods. "a" stands for ancient; "e" stands for extant, "W" for wolf, "D" for dog and "H" for haplotype. Coloured circles represent different chronological periods and species. White circles are median vectors (hypothetical intermediates). The sizes of coloured circles are proportional to haplotype frequency in the respective populations. See Appendix VI for detailed information regarding the source and haplotype correspondence of the haplotypes displayed here. *Haplotype where LYEP9 is included; **Haplotype where LYEP11 is included; ***Haplotype where LYEP51 and LYEP53 are included; +Haplotype where LYEP27 is included.*

There is a genetic continuity regarding dog mtDNA Haplogroup representativeness in Iberian dogs, although the Iberian Neolithic period is sub-represented with only 2 samples. In contrast, in Eurasia, a sharp difference is noted by the Neolithic period with the introduction of dogs with a different mtDNA genetic composition (Figure 20).

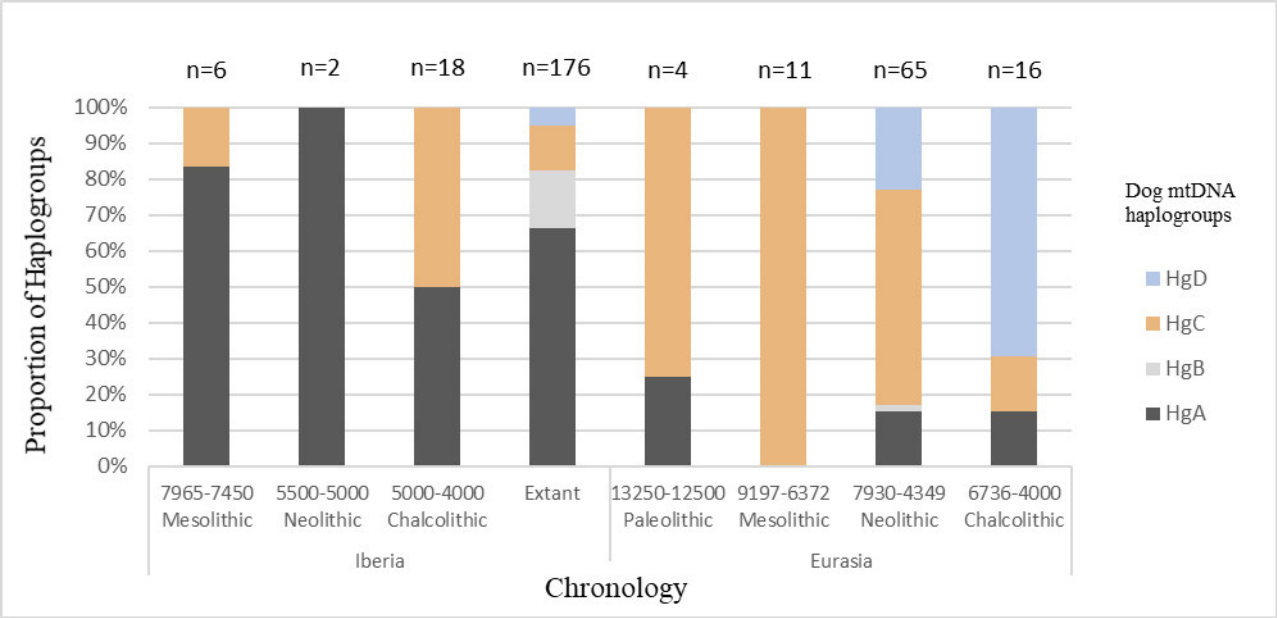


Figure. 20. Frequency of the main dog mtDNA-Haplogroups (A, B, C and D) across time in Iberia and Eurasia. A high frequency of HgA dogs can be detected continuously in Iberia since the Mesolithic. Dates below charts represents the oldest sample and the earliest sample from dataset for each cultural period.

3.2.3 Chalcolithic Iberian dogs - PhiPT statistics

Iberian dogs show almost no intraspecific genetic differentiation among the populations over time – regarding the ancient periods considered: values range from 0 (not significant; p>0,05; Mesolithic vs Neolithic) to 8% (not significant; p>0,05; Mesolithic vs Chalcolithic) when the 43 bp fragment is used. If a fragment of 182 bp containing missing data is used, this value is 0% for both comparisons (Mesolithic vs Neolithic and Mesolithic vs Chalcolithic), but not significant (p>0,05). The evolutionary trajectory of the Chalcolithic Iberian dogs is of a genetic continuity over time as evidenced by the PhiPT values below (Table 7). Values estimated from 43 bp fragments are on the left side of the bar symbol. On the right side of the bar symbol are the values estimated from 182 bp fragments.

Table7. Iberian dogs PhiPT values over periods for two different fragments lengths (43bp and 182bp).

Cultural Period	PhiPt values (p-value)
Mesolithic vs Neolithic (43 bp/182bp)	0% (0.262) / 0% (0.352)
Mesolithic vs Chalcolithic (43 bp/182bp)	8% (0.135) / 0% (0.332)

Regarding genetic distance between Iberian dogs and their European counterparts over time, PhiPT values increases substantially (all PhiPT values ≥0.20), which indicate great/very great genetic diferentiation according to Wright's qualitative guidelines (see Materials and Methods section). Accounting for an average generation time of three years in dogs and wolves (Lindblad-Toh *et al.* 2005), to achieve these PhiPT values of 0.22/0.27 migrant individual every 3 years (or one migrant individual every 15/12 years) for the Mesolithic period, 1.92/1.19 migrant individual every 3 years (or one migrant individual every 6/6 years) for the Neolithic period and 0.34/0.29 migrant individual every 3 years (or one migrant individual every 9/12 years) for the Chalcolithic period would have migrated (bred with) between these populations of dogs (Table 8). The lower time estimated for the Mesolithic period may be explained by the contrasting number of samples from Iberia and Eurasia.

Table 8. PhiPT and Nm (Number of migrants) values of Iberian and Eurasian dogs from different chronological periods.

		Iberian Dogs (n)	Eurasian Dogs (n)	PhiPT Ib* Eur (p value) (43bp)/ (182bp)	Nm (43bp)/Nm (182bp)
Period	meso	6	11	0.69 (0.001)/ 0.65 (0.001)	0.22/0.27
	neo	2	65	0.21 (0.09) /0.30 (0.09)	1.92/1.19
	chalco	18	14	0.59 (0.001) / 0.62 (0.001)	0.34/0.29

As mentioned in Materials and Methods section, in this study, Chalcolithic Iberian wolf structure could not be estimated confidently since there was only one Chalcolithic Iberian wolf in the database.

3.3 Nuclear DNA data: Sex determination of samples

For the LYEP9 sample, once the reads have been normalized by the length of each respective chromosome, only 255.3 reads per 1 million base pairs (0.026%) aligned to the chromosome X with mapping quality of 60 or above (255.3 reads(Q60)/Mbp), compared for instance, with 485.3 reads(Q60)/mbp that aligned to chromosome 1, resulting in half the ratio when comparing chromosome X with autosomal chromosome (255.3/485.3=52.6%); while 357.3 reads(Q60)/mbp aligned to chromosome Y, resulting in a slightly higher proportion of reads when comparing to chromosome X (255.3/357.3=71.6%) (Figure 21;Table 9), an acceptable result due to the fact that there is no reference Y chromosomes; its actual size is not known and may have been underestimated. This result is similar to a previously reported ancient male dog (Frantz *et al.* 2016).

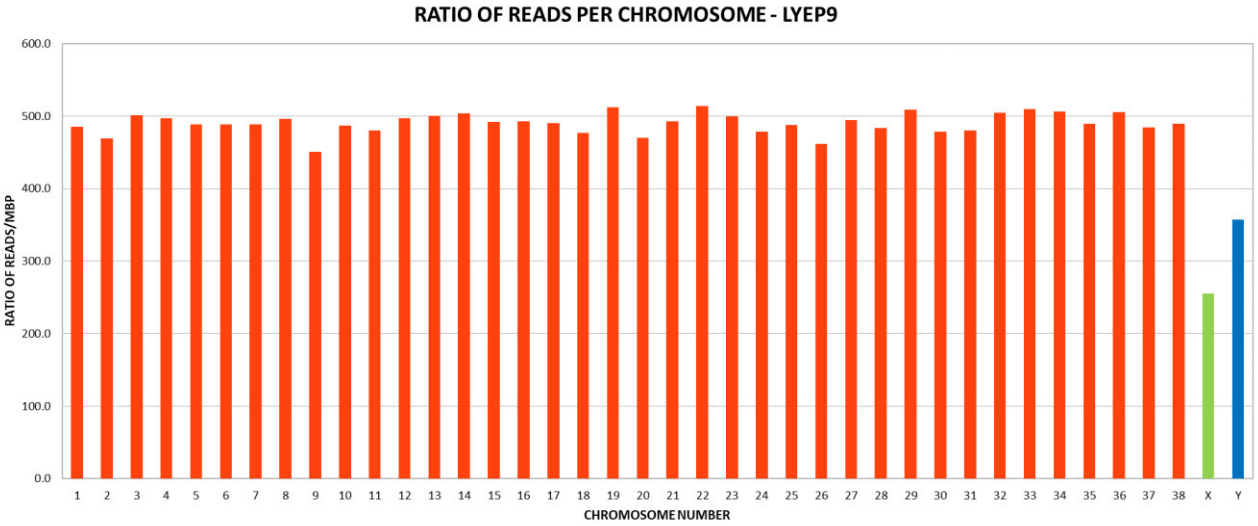


Figure 21. Histograms representing the proportions of sequencing reads mapping each chromosome for LYEP9. Chromosome X is represented by the green bar and chromosome Y by the blue bar.

Applying the same depth based method, LYEP11 and LYEP53 were classified as males as well (Table 9), where only 26.1 reads(Q60)/mbp and 4.4 reads(Q60)/mbp, respectively, aligned to the chromosome X compared with 49.3 reads(Q60)/mbp and 8.6

reads(Q60)/mbp, respectively, that aligned to chromosome 1. This results in half the ratio when comparing chromosome X with autosomal chromosome for LYEP11 (26.1/49.3=52.95%) and LYEP53 (4.4/8.6=50.73%). Regarding chromosome Y, for both LYEP11 and LYEP53 samples, a higher proportion of reads aligned to chromosome Y when compared to chromosome X - 26.1/52.0=50% and 4.4/6.9=63.62%, respectively (Figure 22 and Figure 23). As mentioned above, the high proportion of reads aligned to chromosome Y is an acceptable result due to the fact that there are no reference Y chromosomes; its actual size is not known and may have been underestimated.

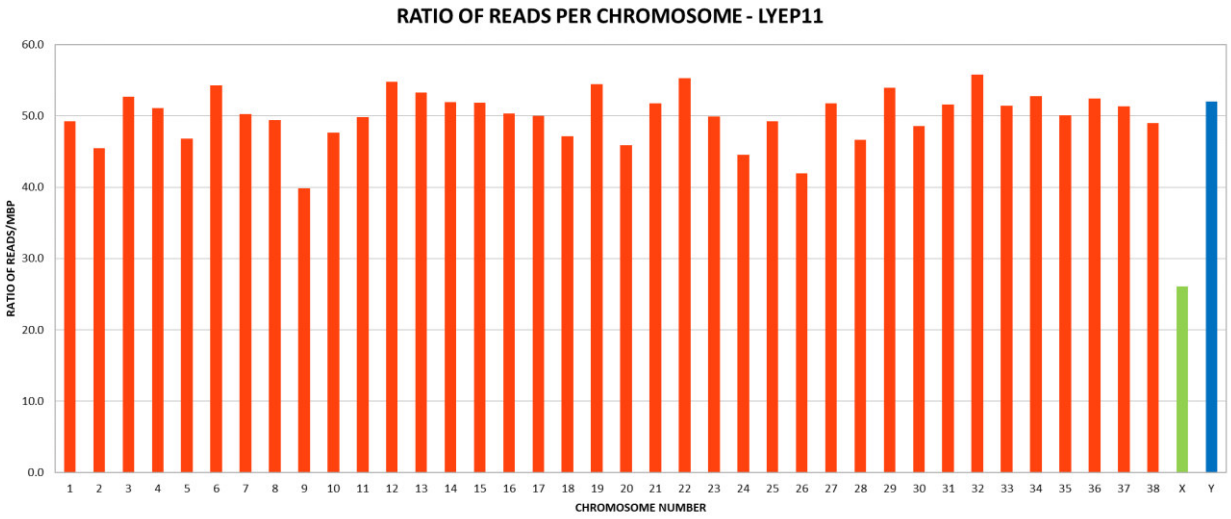


Figure 22. Histograms representing the proportions of sequencing reads mapping each chromosome for LYEP11. Chromosome X is represented by the green bar and chromosome Y by the blue bar.

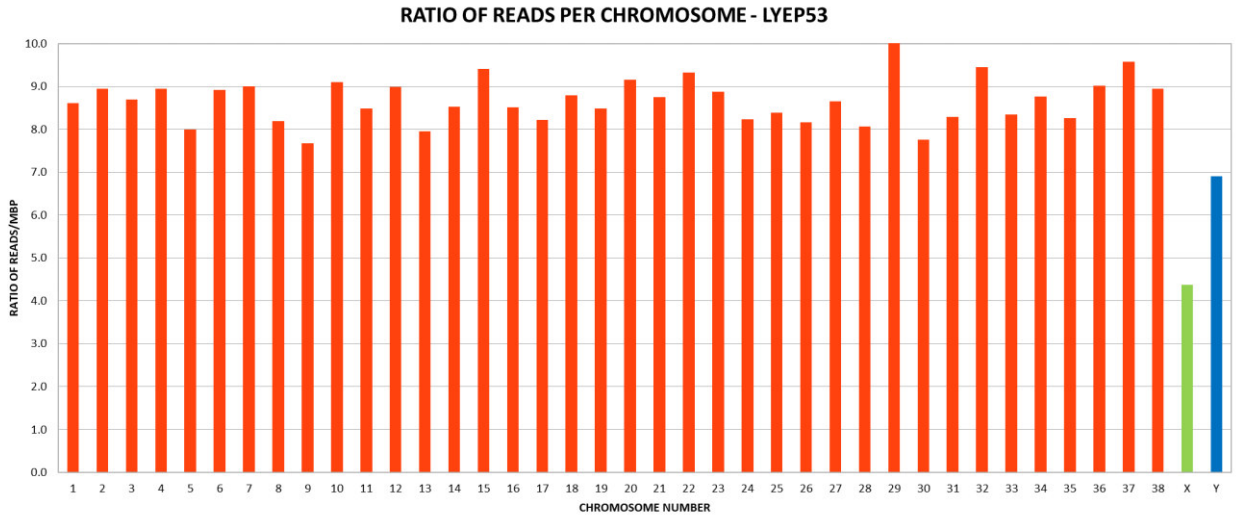


Figure 23. Histograms representing the proportions of sequencing reads mapping each chromosome for LYEP53. Chromosome X is represented by the green bar and chromosome Y by the blue bar.

The sample LYEP51 is the only female (Table 9). A similar ratio of X and autosomal chromosome was found, where 93.8 reads(Q60)/mbp aligned to chromosome X and 93.4 reads(Q60)/mbp aligned to chromosome 1. Some of the reads that did not map against dog genome (41.9 reads(Q60)/mbp) aligned to chromosome Y, however it is possibly a result of contaminant reads that aligned to fragments of chromosome Y, not endogenous DNA (Figure 24).

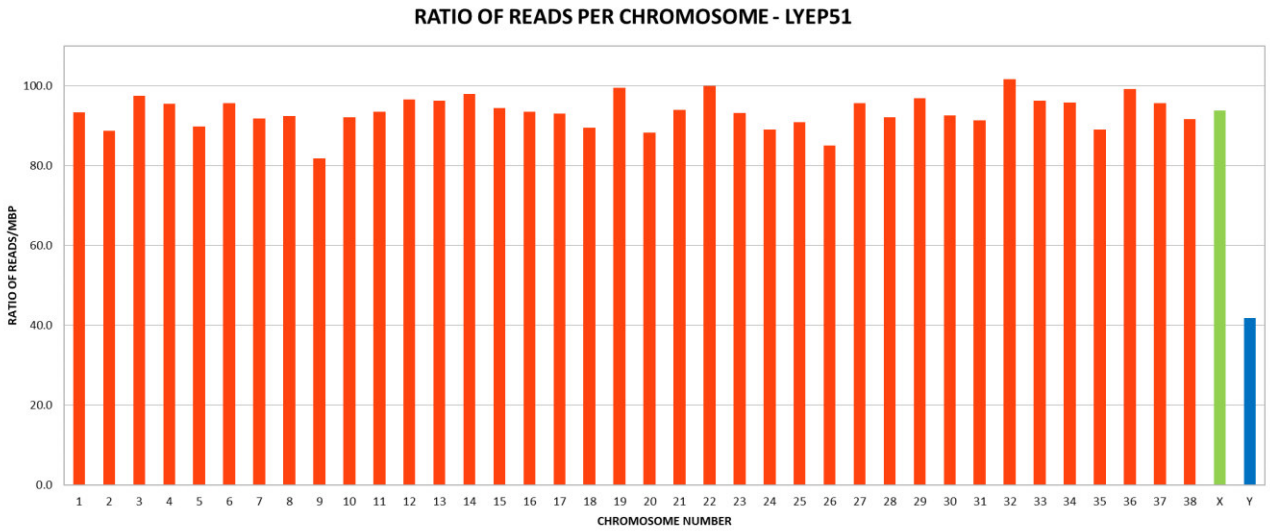


Figure 24. Histograms representing the proportions of sequencing reads mapping each chromosome for LYEP51. Chromosome X is represented by the green bar and chromosome Y by the blue bar.

The only ancient wolf analysed in this study was a male (Table 9), according to the 9.1 reads(Q60)/mbp aligned to chromosome X against 17.9 reads(Q60)/mbp that aligned to chromosome 1. Moreover, 28.5 reads(Q60)/mbp aligned to chromosome Y, a higher proportion of reads when comparing to chromosome X ratio (28.5/9.1=31.5%) (Figure 25). Again, the high proportion of reads aligned to chromosome Y is an acceptable result due to the fact that there are no reference Y chromosomes; its actual size is not known and may have been underestimated. This results confirm the result obtained by Moreno-Gárcia and colleagues (2016) using osteometry approach to infer about the sex of this ancient wolf, as a male.

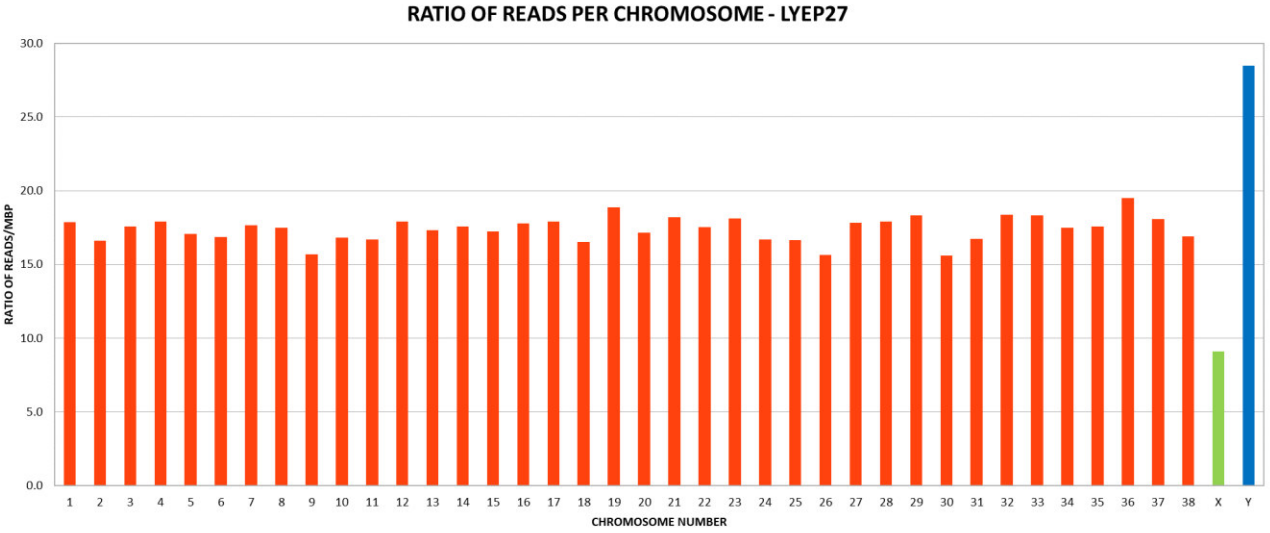


Figure 25. Histograms representing the proportions of sequencing reads mapping each chromosome for LYEP27. Chromosome X is represented by the green bar and chromosome Y by the blue bar. Note: this sample was not mapped against the de novo wolf genome (Gopalakrishnan et al. 2017) because it consisted only of unplaced scaffolds.

Table 9. Results of molecular sex determination for each Canis sample.

Sample ID	Sex assignment
LYEP9	Male
LYEP11	Male
LYEP51	Female
LYEP53	Male
LYEP27	Male

4. DISCUSSION

4 Discussion

This is the first attempt to reconstruct mitochondrial and nuclear genomes from ancient *Canis* from Iberia with NGS methods. The sequences obtained in this study showed the typical pattern of ancient DNA, such as low endogenous DNA (Appendix IX), nucleotide substitutions at the 5' and 3' ends (Figure 9), bacterial contaminant reads (Figure 10) and unknown nucleotides (N's; Table 5). However, it was possible to reconstruct more than 90% of the mitogenome for LYEP9, LYEP11 and LYEP53; unfortunately, LYEP51 and LYEP27 presented a higher percentage of unknown bases, allowing the reconstruction of approximately 76% and 39 % of the mitogenome, respectively (Table 5).

From a network analysis and using reference sequences from extant dogs from other studies (including the 4 ancient dog sequences which sequences were generated by the NGS-454 method by Pires *et al* 2019 for the samples of this study) for which their dog mtDNA haplogroup was known, I was able to assign each of the newly generated sequence (this study) to its closest Haplogroup. The presence of reference sequences in a study of ancient specimens has shown to be important to clearly assign sequences to the closest Haplogroup and to identify unique haplotypes.

With regard to genetic diversity and population structure of Iberian dogs, a genetic continuity regarding Haplogroup representativeness is observed since Mesolithic period, with Hg A being prevalent. In contrast, in the other regions of Eurasia, Hg C is the most prevalent during pre-Neolithic period (Deguilloux *et al.* 2009; Frantz *et al.* 2016; Ollivier *et al.* 2018), being partially replaced by HgA and HgD after the Neolithic (Botigué *et al.* 2017; Ollivier *et al.* 2018). Neolithic Eurasian haplotypes were mainly Haplogroup A, C, or D type, but also one dog carried mtDNA of the Haplogroup B type (Botigué *et al.* 2017; Ollivier *et al.* 2018) (Figure20). The introduction of mtDNA Hg B in Iberia appears to have occurred later than Chalcolithic since this Hg was never detected in 29 dogs analysed between Mesolithic and Chalcolithic periods.

According to Pires *et al.* (Pires *et al.* 2018), Hg D may have arrived in Iberia during late Roman occupation (ca. 1,600 years ago). This alteration of the genetic composition could

be explained by the migration of dogs from other regions of Europe into Iberia. During the Chalcolithic period in Iberia, mtDNA Hg C dogs increased in frequency (Figure 20). When comparing Iberian dogs with their contemporaneous Eurasian counterparts, genetic differentiation is statistically significant, showing great genetic divergency between populations over periods (all pairwise PhiPt values are >0.20). These results suggest a strong genetic isolation of Iberian dogs from the rest of Europe until the Chalcolithic period. PhiPt values based on alignments of sequences with different lengths (smaller fragments with no missing data versus longer sequences with missing data), differ little. Equivalent values are accommodated within the same partition of Wright's (1978) guideline.

It is important to highlight the high frequency and distribution of haplotype H2/H17 (Figure 14/Figure 19). This haplotype accommodates one of the oldest Iberian wolf analysed in this study, LYEP44, and its presence is maintained in dogs in the following periods. Genetic data suggests is a genetic continuity between paleolithic wolves and early dogs in Iberia, but not among paleolithic and extant wolves as has been identified before (Pires et al 2019), probably due to a bottleneck as consequence of human persecution and habitat fragmentation during the past few hundred years (Álvares 2011). Another haplotype that is well represented in different chronology of Iberian ancient dogs is the haplotype H3/H4 (Figure 14/Figure 19). Also, in this haplotype, a genetic continuity between wolf (LYEP46) and dog's haplotype can also be noted. This haplotype is also the most frequent (9 out of 21 dogs) in Chalcolithic dogs.

Wolf Haplogroup 2 is described to be more frequent in ancient wolf population. However, during the last several thousand years, Haplogroup 2 became outnumbered by Haplogroup 1 (Pilot *et al.* 2010). The oldest Iberian wolves – LYEP44 and LYEP46 belonged to Hg2, while LYEP27- the Chalcolithic Iberian wolf of this study, belonged to Haplogroup 1. All haplotypes of extant Iberian wolves fall exclusively within Haplogroup 1, with the exception of a Portuguese wolf, eWolf28, that is the only extant Iberian wolf that belonged to Haplogroup 2 (Pires *et al.* 2017). This wolf is considered a relic, since, in extant Eurasia, Haplogroup 2 can only be detected in Italian wolves (Pilot *et al.* 2010). Pires *et al.* (Pires *et al.* 2019) conclude that this change in Iberian wolf mitochondrial composition (Haplogroup

1 outnumbering Haplogroup 2), associated with the genetic continuity of Paleolithic wolf and extant dog haplotype, suggests that Mesolithic dogs kept the genetic signature of ancient Iberian wolves, transmitting it up to present-days dogs.

It is necessary to sequence more ancient Iberian wolves to investigate the genetic composition and structure over time and clarify this turnover. There is a possibility that the early Iberian dogs were locally domesticated from the Iberian wolves. This study also emphasizes the importance of including ancient wolves and early dogs from European periphery in more global studies of the domestication evolution of dogs. It has been suggested that the Iberian Peninsula, during the Last Glacial Maximum (21,000-17,000 years BP), served as a biodiversity refugia (Hewitt 1996). This pre-historic episode had a profound influence on the genetic structure of isolated populations (Avise *et al.* 1998) such as the Iberian wolf and consequently the Iberian dogs .

The occurrence of events of admixture between wolves and dogs along the past has been widely accepted in the literature (Vilà *et al.* 1997; Vila *et al.* 1999; Leonard *et al.* 2007), in particular as an explanation for the different mitochondrial lineages present in dogs. In the past, many haplotypes were shared between wolves and dogs. Currently, as shown in Figure 14, the modern population of Iberian wolves is well segregated from modern Iberian dogs including village dogs , and it is not currently common the occurrence of hybridization between domestic dogs and wild Iberian wolves, with the exception of certain remote areas and occasions where and when feral dogs may contact with Iberian wolves in the wild (Godinho *et al.* 2011; Torres *et al.* 2017) . This situation deserves constant monitoring and special attention to avoid the loss of the Iberian wolf genetic patrimony. Presently it is a subject of intense research.

Since the nuclear DNA is present in less quantity than mitochondrial DNA within a cell (Ho & Gilbert 2010; Chinnery & Hudson 2013), the recovery of endogenous nuclear DNA in an ancient sample becomes even more difficult. In that way, the low coverage and quantity of endogenous nuclear DNA led this study to focus only in determining the sex of the specimens. To identify the sex of ancestral specimens, which are often found only bone fragments (some of which do not allow inferring sex), the genomic approach presents itself

as a good solution. I had to opt for a non-direct approach, calculating and comparing the rate of the reads that mapped against the sex chromosomes and the autosomes. This strategy has also been used successfully in Frantz *et al.* (2016) to determinate the sex of a Neolithic dog.

This paleogenomic study provided important results to contributing to a better understanding of the origin of Iberian Chalcolithic dogs. However, it is important to highlight some limitations that prevented the access to more genetic information of Iberian dogs from the Chalcolithic period. The low quality of sequencing is a constraint that occurs due to the fact that we are handling/analysing ancient DNA, characterized by high degradation pattern and consequent *post-mortem* nucleotides alterations (equivalent to mutations) that difficult or even invalidates the sequences' quality reading. Nucleotide bases that are sequenced with poor reading quality are replaced by N's (unknown nucleotides) characters in the consensus sequence. The presence of unknown bases at the Hyper Variable Region of the the D-loop, the "hot spot" of nucleotide variability of the mtDNA in *Canis*, may impair the haplogroup assignment, species identity and inferences of evolutionary processes.

Other constraints lay on the molecular marker used in this study. Mitochondrial DNA markers have been widely used to investigate phylogeographic of animals, including dogs, due to its characteristics: easy amplification, maternally inherited, lack of chromosome recombination and high rate of mutation (Chinnery & Hudson 2013). However, these analyses consist on a limited approach to reconstruct the past and should be complemented with other independent data sources (e.g. chromosome Y and autosomal SNPs).

Despite the great percentage of bacterial reads contaminants, mostly reads were assigned to potential artefacts. New molecular strategies, such as capture-enrichment approach (e.g biotinylated RNA baits) (Cruz-Dávalos *et al.* 2017) that are designed to capture specific genomic regions, is recommended for contaminated and fragmented DNA, increasing the endogenous nuclear DNA recovery and reducing level of artefacts. In order to improve the sequence quality, it is recommended for future research to sequence all the 5 ancient DNA

samples together or even less in a single lane (not all together with other samples as it was done before).

The mitochondrial genomes sequences recovered in this study were not submitted to GenBank database since a new sequencing is planned to be carried out with considered changes to reduce missing nucleotides. The improvement of the coverage will allow the identification of the nucleotide bases with more confidence. The impact in gene function of certain mtDNA SNPs that were identified in this study should be further investigated.

Finally, whatever is the direction of future investigations, this study provides basic data for a better understanding of the evolutionary trajectories and genomic composition of the Chalcolithic Iberian dogs, a population, so far, little investigated.

5. FINAL CONSIDERATIONS

5 Final considerations

Directly analysing specimens from the past reduces erroneous conclusions over the genetic background of past populations, especially when working with a species that has been strongly selected and has had its genetic makeup altered over time. The development of new sequencing technologies, which allowed a higher sequencing power, has proven to be crucial in the analyses of entire genomes from old specimens – the oldest specimens analyzed so far had ca. 300,000 years-old (Meyer *et al.* 2014) and ca. 700,000 years-old (Orlando *et al.* 2013). It is important to emphasize that the success of analyzing such old samples are exceptions, as in the case of samples that are found inside caves or permafrost ground. On the order hand, warm climates, e.g. tropical and mediterranean regions, are not the best environment to preserve DNA. Therefore, it presents a high failure rate in the extraction of endogenous DNA, due to the modifications that happen over time in DNA.

Bioinformatics, a field of biology that has experienced an explosive growth in the last decade, allows the analysis and management of megadata generated by Next Generation Sequencing, something that would have been impossible without the advance of software used by bioinformaticians. In Portugal, even more Universities are adapting their *curriculum* to include the study of Bioinformatics in formation. Some Institutes and Organizations also have been offering training program and computing facilities, as well as consulting services in data analysis and management.

Working with ancestral samples differs from working with modern samples. Different software and filters were used in this study, following the best practices for ancient DNA. The Script generated here can be repeated for other ancient samples if one wishes, with the caution to install all the software needed before.

Here I conclude that despite the environmental factor from where these samples were retrieved (warm climate) and age of samples (ca. 5,000-4,000 BP) this study was successfully carried out. A multidisciplinary approach, where zooarchaeology and genetic is integrated (zooarchaeogenetic) is to uncover the evolution of domestication of the Iberian dogs. In addition, the results of the genetic composition of ancient Iberian wolves

presented here may be important as an auxiliary tool in the study of conservation and ecology of extant Iberian wolves. Knowing its genetic composition through the time, allows to better understand the events that modulated the genetic variability and evolutionary path of the Iberian wolf. This species has resisted against various obstacles (environmental and anthropogenic), being currently an important genetic patrimony.

REFERENCES

Akey, J.M., Ruhe, A.L., Akey, D.T., Wong, A.K., Connelly, C.F., Madeoy, J., *et al.* (2010). Tracking footprints of artificial selection in the dog genome. *Proc. Natl. Acad. Sci.*, 107, 1160–1165.

Albizuri, S., Fernández, M. & Tomás, X. (2011). *Evidencias sobre el uso del perro en la carga durante el Bronce Inicial en la Península Ibérica: el caso de Can Roqueta II (Sabadell, Barcelona)*. *Archaeofauna*. La Asociación.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.*, 215, 403–410.

Altuna, J.K. & Mariezkurrena, K. (1985). Bases de subsistencia de los pobladores de Erralla: Macromamíferos. *Munibe*, 37, 87–117.

Álvares, F.J. (2011). Ecologia e conservação do lobo (Canis lupus, L.) no Noroeste de Portugal. PhD thesis, 191.

Ameen, C., Hulme-Beaman, A., Evin, A., Germonpré, M., Britton, K., Cucchi, T., *et al.* (2017). A landmark-based approach for assessing the reliability of mandibular tooth crowding as a marker of dog domestication. *J. Archaeol. Sci.*, 85, 41–50.

Anderson, T.M., VonHoldt, B.M., Candille, S.I., Musiani, M., Greco, C., Stahler, D.R., *et al.* (2009). Molecular and Evolutionary History of Melanism in North American Gray Wolves. *Science (80-.).*, 323, 1339–1343.

Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data.* Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Last accessed .

Arana, M.C. & Rodríguez, C.F. (2013). Los carnívoros del recinto de fosos calcolítico de el casetón de la era (Villalba de los alcores, valladolid), 277–282.

Arias, P., Diniz, M.T., Araújo, A.C., Armendariz, Á. & Teira, L.C. (2015). At the Edge of the Marshes: New Approaches to the Sado Valley Mesolithic (Southern Portugal). In: *Muge 150th: The 150th Anniversary of the Discovery of Mesolithic Shellmiddens* (eds. Bicho, N., Detry, C., Price, D. & Cunha, E.). pp. 301–319.

Avise, J.C., Walker, D. & Johns, G.C. (1998). Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proc. R. Soc. London. Ser. B Biol. Sci.*, 265, 1707–1712.

Axelsson, E., Ratnakumar, A., Arendt, M.-L., Maqbool, K., Webster, M.T., Perloski, M., *et al.* (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature*, 495, 360–364.

Bark to the future: Ice Age puppies may reveal canine evolution. (2016). *Guard.* Available at: https://www.theguardian.com/science/2016/mar/28/bark-future-ice-age-puppies-may-reveal-canine-evolution-yakutia-russia. Last accessed 2 February 2019.

Benecke, N. (1987). Studies on early dog remains from Northern Europe. *J. Archaeol. Sci.*, 14, 31–49.

Bicho, N., Cascalheira, J. & Marreiros, J. (2012). On the (l)edge: the case of Vale Boi rockshelter (Algarve, Southern Portugal). In: *Caves in Context. The Economical, Social, and Ritual Importance of Caves and Rockshelters.* pp. 65–81.

Bjornerfeldt, S., Webster, M., Vilà, C., Björnerfeldt, S., Bjornerfeldt, S., Webster, M., *et al.* (2006). Relaxation of selective constraint on dog mitochondrial DNA following domestication. *Genome Res.*, 16, 990–994.

Bleidorn, C. (2015). Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. *Syst. Biodivers.*, 14, 1–8.

Boggiano, F., Ciofi, C., Boitani, L., Formia, A., Grottoli, L., Natali, C., *et al.* (2013). Detection of an East European wolf haplotype puzzles mitochondrial DNA monomorphism of the Italian wolf population. *Mamm. Biol.*, 78, 374–378.

Botigué, L.R., Song, S., Scheu, A., Gopalan, S., Pendleton, A.L., Oetjens, M., *et al.* (2017). Ancient European dog genomes reveal continuity since the Early Neolithic. *Nat. Commun.*, 8, 16082.

Boyko, A.R., Boyko, R.H., Boyko, C.M., Parker, H.G., Castelhano, M., Corey, L., *et al.* (2009). Complex population structure in African village dogs and its implications for inferring dog domestication history. *Proc. Natl. Acad. Sci.*, 106, 13903–13908.

Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., Prufer, K., *et al.* (2007). Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci.*, 104, 14616–14621.

Broad Institute. (2018). *Picard Tools*. Available at: http://broadinstitute.github.io/picard/. Last accessed 15 November 2018.

Brown, S.K., Pedersen, N.C., Jafarishorijeh, S., Bannasch, D.L., Ahrens, K.D., Wu, J.-T., *et al.* (2011). Phylogenetic Distinctiveness of Middle Eastern and Southeast Asian Village Dog Y Chromosomes Illuminates Dog Origins. *PLoS One*, 6, e28496.

Budiansky, S. (1992). The covenant of the wild: why animals chose domestication. *Choice Rev. Online*, 30, 30-0999-30–0999.

Cardoso, J.L. (1997). *O povoado de Leceia (Oeiras): sentinela do Tejo no terceiro milénio antes de Cristo*. 1ª

edição. IPM, CMO.
Cardoso, J.L. (2008). The chalcolithic fortified site of Leceia (Oeiras, Portugal). *Verdolay*, 11, 49–66.
Cardoso, J.L. & Soares, A.M.M. (1995). Cronologia absoluta para asocupações do neolítico final e do calcolítico inicial do Povoado Pré-histórico de Leceia (Oerias). *Estud. Arqueol. Oeiras*, 5.
Catagnano, V. (2016). Aproximación morfométrica y paleogenética al estudio de la variabilidad de Canis l. familiais en la Península Ibérica desde el Neolítico hasta época Romana y su contextualización en el ámbito del mediterráneo occidental. Universitat Autònoma de Barcelona.
Célérier, G. & Delpech, F. (1978). Un chien dans l'Azilien de "Pont d'Ambon" (Dordogne) ? *Bull. la Société préhistorique française*, 75, 212–215.
Célérier, G., Tisnerat, N. & Valladas, H. (1999). Données nouvelles sur l'âge des vestiges de chien à Pont d'Ambon, Bourdeilles (Dordogne)/New data on the age of Canis remains at Pont d'Ambon, Bourdeilles (Dordogne, France). *Paléo*, 11, 163–165.
Chaix, L. (2000). A preboreal dog from the Nothern Alps (Savoie, France). In: *Dogs through time: an archaeological perspective* (ed. Crockford, S.J.). Oxford: Archaeopress, pp. 49–59.
Chinnery, P.F. & Hudson, G. (2013). Mitochondrial genetics. *Br. Med. Bull.*, 106, 135–159.
Clutton-Brock, J. (2016). Origins of the dog: The archaeological evidence. In: *The Domestic Dog* (ed. Serpell, J.). Cambridge University Press, pp. 7–21.
Clutton-Brock, J. & Noe-Nygaard, N. (1990). New osteological and C-isotope evidence on mesolithic dogs: Companions to hunters and fishers at Star Carr, Seamer Carr and Kongemose. *J. Archaeol. Sci.*, 17, 643–653.
Cooper, A. & Poinar, H. (2000). Ancient DNA: Do It Right or Not at All. *Science (80-.).*, 289, 1139b – 1139.
Crockford, S.J. & Kuzmin, Y. V. (2012). Comments on Germonpre et al., Journal of Archaeological Science 36, 2009 "Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: Osteometry, ancient DNA and stable isotopes", and Germonpre, Lazkickova-Galetova, and Germonpre, L. *J. Archaeol. Sci.*, 39, 2797–2801.
Cruz-Dávalos, D.I., Llamas, B., Gaunitz, C., Fages, A., Gamba, C., Soubrier, J., *et al.* (2017). Experimental conditions improving in-solution target enrichment for ancient DNA. *Mol. Ecol. Resour.*, 17, 508–522.
Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., *et al.* (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci.*, 110, 15758–15763.
Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods*, 9, 772–772.
Darwin, C. (1885). *The variation of animals and plants under domestication*. J. Murray, London :
Davis, J.M. & Valla, F.R. (1978). Evidence for domestication of the dog 12,000 years ago in the Natufian of Israel. *Nature*, 276, 608–610.
Davis, S.J.M.S. & Simões, T. (2016). The velocity of Ovis in prehistoric times: the sheep bones from early neolithic Lameiras, Sintra, Portugal. In: *O Neolítico em Portugal antes do Horizonte 2020: Perspectivas em debate. AAP Monografias: 2* (eds. Diniz, M., Neves, C. & Martins, A.). Associação dos Arqueólogos Portugueses, Lisboa, pp. 51–66.
Degerbøl, M. (1961). On a find of a Preboreal domestic dog (Canis familiaris L.) from Star Carr, Yorkshire, with remarks on other Mesolithic dogs. *Proc. Prehist. Soc.*, 27, 35–55.
Deguilloux, M.F., Moquel, J., Pemonge, M.H. & Colombeau, G. (2009). Ancient DNA supports lineage replacement in European dog gene pool: insight into Neolithic southeast France. *J. Archaeol. Sci.*, 36, 513–519.
Delibes de Castro, G., Crespo Díez, M., Martín Vela, R. & Rodíguez Marcos, J.A. (2018). A ditched enclosure in the Middle Douro Valley: El Casetón de La Era (Villalba de los Alcores, Valladolid). In: *Archaeology in the River Duero Valley*. Cambridge Scholars Publishing.
Detry, C. & Cardoso, J.L. (2010). On some remains of dog (Canis familiaris) from the Mesolithic shell-middens of Muge, Portugal. *J. Archaeol. Sci.*, 37, 2762–2774.
Dikov, N. (1996). The Ushki sites, Kamchatka Peninsula. In: *American Beginnings: the Prehistory and Palaeoecology of Beringia* (ed. West FH (University of Chicago Press, C.). pp. 244–250.
Ding, Z.-L., Oskarsson, M., Ardalan, A., Angleby, H., Dahlgren, L.-G., Tepeli, C., *et al.* (2012). Origins of domestic dog in Southern East Asia is supported by analysis of Y-chromosome DNA. *Heredity (Edinb).*, 108, 507–514.
Drake, A.G. (2011). Dispelling dog dogma: An investigation of heterochrony in dogs using 3D geometric morphometric analysis of skull shape. *Evol. Dev.*, 13, 204–213.
Drake, A.G., Coquerelle, M. & Colombeau, G. (2015). 3D morphometric analysis of fossil canid skulls contradicts the suggested domestication of dogs during the late Paleolithic. *Sci. Rep.*, 5, 8299.

Druzhkova, A.S., Thalmann, O., Trifonov, V.A., Leonard, J.A., Vorobieva, N. V., Ovodov, N.D., *et al.* (2013). Ancient DNA Analysis Affirms the Canid from Altai as a Primitive Dog. *PLoS One*, 8, e57754.

Duleba, A., Skonieczna, K., Bogdanowicz, W., Malyarchuk, B., Grzybowski, T., Duleba, A., *et al.* (2015). Complete mitochondrial genome database and standardized classification system for Canis lupus familiaris. *Forensic Sci. Int. Genet.*, 19, 123–129.

Fan, Z., Silva, P., Gronau, I., Wang, S., Armero, A.S., Schweizer, R.M., *et al.* (2016). Worldwide patterns of genomic variation and admixture in gray wolves. *Genome Res.*, 26, 163–173.

Frantz, L.A.F., Mullin, V.E., Pionnier-Capitan, M., Lebrasseur, O., Ollivier, M., Perri, A., *et al.* (2016). Genomic and archaeological evidence suggests a dual origin of domestic dogs. *Science (80-.).*, 352, 1228–1231.

Freedman, A.H., Gronau, I., Schweizer, R.M., Ortega-Del Vecchyo, D., Han, E., Silva, P.M., *et al.* (2014). Genome Sequencing Highlights the Dynamic Early History of Dogs. *PLoS Genet.*, 10, e1004016.

Freedman, A.H. & Wayne, R.K. (2017). Deciphering the Origin of Dogs: From Fossils to Genomes. *Annu. Rev. Anim. Biosci.*, 5, 281–307.

Gansauge, M.-T. & Meyer, M. (2013). Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.*, 8, 737–748.

Germonpré, M., Lázničková-Galetová, M. & Sablin, M. V. (2012). Palaeolithic dog skulls at the Gravettian Předmostí site, the Czech Republic. *J. Archaeol. Sci.*, 39, 184–202.

Germonpré, M., Sablin, M. V., Stevens, R.E., Hedges, R.E.M., Hofreiter, M., Stiller, M., *et al.* (2009). 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.

Gilbert, M.T.P., Bandelt, H.-J., Hofreiter, M. & Barnes, I. (2005). Assessing ancient DNA studies. *Trends Ecol. Evol.*, 20, 541–544.

Godinho, R., Llaneza, L., Blanco, J.C., Lopes, S., Álvares, F.F., García, E.J., *et al.* (2011). Genetic evidence for multiple events of hybridization between wolves and domestic dogs in the Iberian Peninsula. *Mol. Ecol.*, 20, 5154–5166.

Gopalakrishnan, S., Samaniego Castruita, J.A., Sinding, M.-H.S.H.S., Kuderna, L.F.K.K., Räikkönen, J., Petersen, B., *et al.* (2017). The wolf reference genome sequence (Canis lupus lupus) and its implications for Canis spp. population genomics. *BMC Genomics*, 18, 495.

Gourichon, L. & Helmer, D. (2008). Étude archéozoologique de Mureybet. *Le site néolithique Tell Mureybet (Syrie du Nord. en hommage à Jacques Cauvin*, 115–228.

Grayson, D.K., Parmalee, P.W., Lyman, R.L. & Mead, J.I. (1988). Danger Cave, Last Supper Cave, and Hanging Rock Shelter : the faunas. Anthropological papers of the AMNH ; v. 66, pt. 1.

Greig, K., Boocock, J., Prost, S., Horsburgh, K.A., Jacomb, C., Walter, R., *et al.* (2015). Complete Mitochondrial Genomes of New Zealand's First Dogs. *PLoS One*, 10, e0138536.

Griffiths, A., Wessler, S., Lewontin, R., Gelbart, W., Suzuki, D. & Miller, J. (2004). *An Introduction to Genetic Analysis*. 8th edn. W. H. Freeman & Co Ltd.

Grimm, D. (2015). Dawn of the dog. *Science (80-.).*, 348, 274–279.

Guindon, S. & Gascuel, O. (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst. Biol.*, 52, 696–704.

Hagner, D. (2018). *Undeniable Solidarity: How Dogs and Humans Domesticated One Another*. AuthorHouse.

Handt, O., Hoss, M., Krings, M., Paabo, S., Höss, M., Krings, M., *et al.* (1994). Ancient DNA: Methodological challenges. *Experientia*, 50, 524–529.

Handt, O., Krings, M., Ward, R.H., Paabo, S., Pääbo, S. & Paabo, S. (1996). The retrieval of ancient human DNA sequences. *Am. J. Hum. Genet.*, 59, 368–76.

Hansen, A.J., Willerslev, E., Wiuf, C., Mourier, T. & Arctander, P. (2001). Statistical Evidence for Miscoding Lesions in Ancient DNA Templates. *Mol. Biol. Evol.*, 18, 262–265.

Hasegawa, M., Kishino, H. & Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.*, 22, 160–74.

Hedrick, P. & Waits, L. (2005). Conservation genetics: What ancient DNA tells us. *Heredity (Edinb).*, 94, 463–464.

Helmer, D. & Gourichon, L. (2008a). Premières données sur les modalités de subsistance à Tell Aswad (Syrie, PPNB moyen et récent, Néolithique céramique ancien) – Fouilles 2001-2005. *MOM Éditions*, 49, 119–151.

Helmer, D. & Gourichon, L. (2008b). Révision de la faune de Cafer Hoyük (Malatya, Turquie) : apports des méthodes de l'analyse des mélanges et de l'analyse de Kernel à la mise en évidence de la domestication. In: *Archaeozoology of the near East VIII : proceedings of the eighth International symposium on the archaeozoology of Southwestern Asia and adjacent areas : Aswa VIII, Lyon, June 28-July 1, 2006*. Persée

- Portail des revues scientifiques en SHS, pp. 169–195.
Hewitt, G.M. (1996). Some genetic consequences of ice ages, and their role, in divergence and speciation. *Biol. J. Linn. Soc.*, 58, 247–276.
Higuchi, R., Bowman, B., Freiberger, M., Ryder, O.A. & Wilson, A.C. (1984). DNA Sequences from the Quagga, an Extinct Member of the Horse Family. *Nat. Publ. Gr.*, 312, 282–284.
Himmelberger, A.L., Spear, T.F., Satkoski, J.A., George, D.A., Garnica, W.T., Malladi, V.S., *et al.* (2008). Forensic Utility of the Mitochondrial Hypervariable Region 1 of Domestic Dogs, in Conjunction with Breed and Geographic Information. *J. Forensic Sci.*, 53, 81–89.
Ho, S.Y.W. & Gilbert, M.T.P. (2010). Ancient mitogenomics. *Mitochondrion*, 10, 1–11.
Hofreiter, M., Jaenicke, V., Serre, D., Haeseler, A. Von & Pääbo, S. (2001). DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res.*, 29, 4793–4799.
Illumina. (2016). Nextera® DNA Library Preparation Kits.
Irving-Pease, E.K., Frantz, L.A.F., Sykes, N., Callou, C. & Larson, G. (2018). Rabbits and the Specious Origins of Domestication. *Trends Ecol. Evol.*, 33, 149–152.
Janssens, L., Giemsch, L., Schmitz, R., Street, M., Van Dongen, S. & Crombé, P. (2018). A new look at an old dog: Bonn-Oberkassel reconsidered. *J. Archaeol. Sci.*, 92, 126–138.
Janssens, L., Perri, A., Crombé, P., Van Dongen, S. & Lawler, D. (2019). An evaluation of classical morphologic and morphometric parameters reported to distinguish wolves and dogs. *J. Archaeol. Sci. Reports*, 23, 501–533.
Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F. & Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, 29, 1682–1684.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., *et al.* (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28, 1647–1649.
Kircher, M., Heyn, P. & Kelso, J. (2011). Addressing challenges in the production and analysis of illumina sequencing data. *BMC Genomics*, 12, 382.
Koblmüller, S., Vilà, C., Lorente-Galdos, B., Dabad, M., Ramirez, O., Marques-Bonet, T., *et al.* (2016). Whole mitochondrial genomes illuminate ancient intercontinental dispersals of grey wolves (Canis lupus). *J. Biogeogr.*, 43, 1728–1738.
Koepfli, K.-P., Pollinger, J., Godinho, R., Robinson, J., Lea, A., Hendricks, S., *et al.* (2015). Genome-wide Evidence Reveals that African and Eurasian Golden Jackals Are Distinct Species. *Curr. Biol.*, 25, 2158–2165.
Kumar, S., Stecher, G. & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.*, 33, 1870–1874.
Larson, G. & Fuller, D.Q. (2014). The Evolution of Animal Domestication. *Annu. Rev. Ecol. Evol. Syst.*, 45, 115–136.
Larson, G., Karlsson, E.K., Perri, A., Webster, M.T., Ho, S.Y.W., Peters, J., *et al.* (2012). Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc. Natl. Acad. Sci.*, 109, 8878–8883.
Leigh, J.W. & Bryant, D. (2015). Popart: full-feature software for haplotype network construction. *Methods Ecol. Evol.*, 6, 1110–1116.
Leonard, J.A., Vilà, C., Fox-Dobbs, K., Koch, P.L., Wayne, R.K. & Van Valkenburgh, B. (2007). Megafaunal Extinctions and the Disappearance of a Specialized Wolf Ecomorph. *Curr. Biol.*, 17, 1146–1150.
Leonard, J.A., Wayne, R.K., Wheeler, J., Valadez, R., Guillén, S. & Vilà, C. (2002). Ancient DNA Evidence for Old World Origin of New World Dogs. *Science (80-.).*, 298, 1613.
Li, H. & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, 26, 589–595.
Liesau Von Lettow-Vorbeck, C., Esparza Arroyo, Á. & Sánchez Polo, A. (2014). ¿Huesos en la basura o depósito ritualizado? Los perros descuartizados de La Huelga (Dueñas, Palencia). *Zephyrvs*, 74, 89.
Lindahl, T. (1993). Instability and Decay of the Primary Structure of DNA. *Nat. Publ. Gr.*, 362, 709–715.
Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., *et al.* (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, 438, 803–819.
Lindgreen, S. (2012). AdapterRemoval: Easy Cleaning of Next Generation Sequencing Reads. *BMC Res. Notes*, 5, 337.
MacHugh, D.E., Larson, G. & Orlando, L. (2016). Taming the Past: Ancient DNA and the Study of Animal Domestication. *Annu. Rev. Anim. Biosci.*, 5, 329–351.
Malmstrom, H., Svensson, E.M., Gilbert, M.T.P.. T.P., Willerslev, E., Gotherstrom, A. & Holmlund, G. (2007).

More on Contamination: The Use of Asymmetric Molecular Behavior to Identify Authentic Ancient Human DNA. *Mol. Biol. Evol.*, 24, 998–1004.

Mardis, E.R. (2008). Next-Generation DNA Sequencing Methods. *Annu. Rev. Genomics Hum. Genet.*, 9, 387–402.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17, 10.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., *et al.* (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, 20, 1297–1303.

Metzker, M.L. (2010). Sequencing technologies — the next generation. *Nat. Rev. Genet.*, 11, 31–46.

Meyer, M., Fu, Q., Aximu-Petri, A., Glocke, I., Nickel, B., Arsuaga, J.-L., *et al.* (2014). A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature*, 505, 403–406.

Meyer, M. & Kircher, M. (2010). Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb. Protoc.*, 2010, pdb.prot5448-pdb.prot5448.

Millar, C.D., Huynen, L., Subramanian, S., Mohandesan, E. & Lambert, D.M. (2008). New developments in ancient genomics. *Trends Ecol. Evol.*, 23, 386–393.

Miller, W., Schuster, S.C., Welch, A.J., Ratan, A., Bedoya-Reina, O.C., Zhao, F., *et al.* (2012). Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proc. Natl. Acad. Sci.*, 109, E2382–E2390.

Montana, L., Caniglia, R., Galaverni, M., Fabbri, E. & Randi, E. (2017). A new mitochondrial haplotype confirms the distinctiveness of the Italian wolf (Canis lupus) population. *Mamm. Biol.*, 84, 30–34.

Morel, P. & Müller, W. (1997). *Hauterive-Champréveyres, 11. Un campement magdalénien au bord du lac de Neuchâtel étude archéozoologique (secteur 1).*

Moreno-Garcia, M., Pimenta, C., Martinez-Sanchez, R., Barroso, I., Pimenta, V. & Santos, N. (2016). Caracterização osteométrica das populações atuais de lobo-ibérico Canis lupus signatus - um alicerce para o seu reconhecimento no registo arqueozoológico (Poster). *IV Congr. Ibérico do Lobo*, 2–3.

Morey, D. (1994). The Early Evolution of the Domestic Dog. *Am. Sci.*, 82, 336–347.

Morey, D.F. (1992). Size, shape and development in the evolution of the domestic dog. *J. Archaeol. Sci.*, 19, 181–204.

Mussil, R. (2000). Evidence for the domestication of wolves in central European Magdalenian sites. In: *Dogs through time: an archaeological perspective* (ed. Crockford, S.J.). Oxford: Archaeopress, pp. 21–28.

Napierala, H. & Uerpmann, H.-P. (2012). A 'new' palaeolithic dog from central Europe. *Int. J. Osteoarchaeol.*, 22, 127–137.

Nishinakagawa, H., Matsumoto, M., Otsuka, J. & Kawaguchi, S. (1992). Sekeletal remains of domestic dogs from Jomon and Yayoi sites in Kagoshima Prefecture. *J. Anthr. Soc. Nippon*, 100, 485–498.

O'Rourke, D.H.., Hayes, M.G. & Carlyle, S.W. (2000). Ancient DNA Studies in Physical Anthropology. *Annu. Rev. Anthr.*, 29, 217–42.

Oetjens, M.T., Martin, A., Veeramah, K.R. & Kidd, J.M. (2018). Analysis of the canid Y-chromosome phylogeny using short-read sequencing data reveals the presence of distinct haplogroups among Neolithic European dogs. *BMC Genomics*, 19, 350.

Ollivier, M., Tresset, A., Frantz, L.A.F., Bréhard, S., Bălăşescu, A., Mashkour, M., *et al.* (2018). Dogs accompanied humans during the Neolithic expansion into Europe. *Biol. Lett.*, 14, 20180286.

Ollivier, M., Tresset, A., Hitte, C., Petit, C., Hughes, S., Gillet, B., *et al.* (2013). Evidence of Coat Color Variation Sheds New Light on Ancient Canids. *PLoS One*, 8, e75110.

Olsen, S.J. (1985). Origins of the domestic dog: the fossil record. *J. Mammal.*, 67, 218–219.

Orlando, L., Gilbert, M.T.P. & Willerslev, E. (2015). Reconstructing ancient genomes and epigenomes. *Nat. Rev. Genet.*, 16, 395–408.

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., *et al.* (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, 499, 74–78.

Ovodov, N.D., Crockford, S.J., Kuzmin, Y. V., Higham, T.F.G., Hodgins, G.W.L. & van der Plicht, J. (2011). A 33,000-Year-Old Incipient Dog from the Altai Mountains of Siberia: Evidence of the Earliest Domestication Disrupted by the Last Glacial Maximum. *PLoS One*, 6, e22821.

Paabo, S. (1989). Ancient DNA : Extraction , characterization , molecular cloning , and enzymatic amplification. *Proc. Natl. Acad. Sci.*, 86, 1939–1943.

Pääbo, S. (1985). Molecular cloning of Ancient Egyptian mummy DNA. *Nature*, 314, 644–645.

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Després, V., Hebler, J., Rohland, N., *et al.* (2004). Genetic Analyses from Ancient DNA. *Annu. Rev. Genet.*, 38, 645–679.

Pang, J.-F., Kluetsch, C., Zou, X.-J., Zhang, A. -b., Luo, L.-Y., Angleby, H., *et al.* (2009). mtDNA Data

Indicate a Single Origin for Dogs South of Yangtze River, Less Than 16,300 Years Ago, from Numerous Wolves. *Mol. Biol. Evol.*, 26, 2849–2864.

Peakall, R. & Smouse, P.E. (2006). genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes*, 6, 288–295.

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

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., *et al.* (2016). EAGER: efficient ancient genome reconstruction. *Genome Biol.*, 17, 60.

Pendleton, A.L., Shen, F., Taravella, A.M., Emery, S., Veeramah, K.R., Boyko, A.R., *et al.* (2018). Comparison of village dog and wolf genomes highlights the role of the neural crest in dog domestication. *BMC Biol.*, 16, 64.

Pilot, M., Branicki, W., Jędrzejewski, W., Goszczyński, J., Jędrzejewska, B., Dykyy, I., *et al.* (2010). Phylogeographic history of grey wolves in Europe. *BMC Evol. Biol.*, 10, 104.

Pilot, M., Malewski, T., Moura, A.E., Grzybowski, T., Oleński, K., Ruść, A., *et al.* (2015). On the origin of mongrels: evolutionary history of free-breeding dogs in Eurasia. *Proc. R. Soc. B Biol. Sci.*, 282, 20152189.

Pionnier-capitan, M. (2010). La Domestication Du Chien En Eurasie : etude de la diversite passee, approches osteoarcheologiques, morphometriques et paleogenetiques. Ecole Normale Superieure de Lyon.

Pionnier-Capitan, M., Bemilli, C., Bodu, P., Célérier, G., Ferrié, J.-G., Fosse, P., *et al.* (2011). New evidence for Upper Palaeolithic small domestic dogs in South-Western Europe. *J. Archaeol. Sci.*, 38, 2123–2140.

Pires, A.E. (2006). Mitochondrial DNA Sequence Variation in Portuguese Native Dog Breeds: Diversity and Phylogenetic Affinities. *J. Hered.*, 97, 318–330.

Pires, A.E., Amorim, I.R., Borges, C., Simões, F., Teixeira, T., Quaresma, A., *et al.* (2017). New insights into the genetic composition and phylogenetic relationship of wolves and dogs in the Iberian Peninsula. *Ecol. Evol.*, 7, 4404–4418.

Pires, A.E., Detry, C., Chikhi, L., Rasteiro, R., Amorim, I.R., Simões, F., *et al.* (2019). The curious case of the Mesolithic Iberian dogs: An archaeogenetic study. *J. Archaeol. Sci.*, 105, 116–129.

Pires, A.E., Detry, C., Fernandez-Rodriguez, C., Valenzuela-Lamas, S., Arruda, A.M., De Grossi Mazzorin, J., *et al.* (2018). Roman dogs from the Iberian Peninsula and the Maghreb – A glimpse into their morphology and genetics. *Quat. Int.*, 471, 132–146.

Pires, A.E. & Ginja, C. (2013). Why look back ? Methods and relevance of ancient DNA studies. *Cad. do GEEvH*, 1, 7–23.

Pires, F., Cardoso, J.L. & Petrucci-Fonseca, F. (2001). *Estudo arqueozoológico dos carnívoros do povoado pré-histórico de Leceia (Oeiras).*

Pitulko, V. V. & Kasparov, A.K. (2017). Archaeological dogs from the Early Holocene Zhokhov site in the Eastern Siberian Arctic. *J. Archaeol. Sci. Reports*, 13, 491–515.

Poinar, H.N. & Stankiewicz, B.A. (1999). Protein preservation and DNA retrieval from ancient tissues. *Proc. Natl. Acad. Sci.*, 96, 8426–8431.

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., *et al.* (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, 505, 43–49.

Pruvost, M., Schwarz, R., Correia, V.B., Champlot, S., Braguier, S., Morel, N., *et al.* (2007). Freshly excavated fossil bones are best for amplification of ancient DNA. *Proc. Natl. Acad. Sci.*, 104, 739–744.

Queiroz, A. (coord. ., Alves, P., Barroso, I., Beja, P., Fernandes, M., Freitas, L., *et al.* (2005). Canis lupus Lobo in Livro Vermelho dos Vertebrados de Portugal, 517–518.

Quinlan, A.R. & Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841–842.

Ramakrishnan, U. & Hadly, E.A. (2009). Using phylochronology to reveal cryptic population histories : review and synthesis of 29 ancient DNA studies. *Mol. Ecol.*, 18, 1310–1330.

Randi, E., Lucchini, V., Christensen, M.F., Mucci, N., Funk, S.M., Dolf, G., *et al.* (2000). Mitochondrial DNA Variability in Italian and East European Wolves : Detecting the Consequences of Small Population Size and Hybridization. *Conserv. Biol.*, 14, 464–473.

Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J.S., Albrechtsen, A., Moltke, I., *et al.* (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*, 463, 757–762.

Rizzi, E., Lari, M., Gigli, E., De Bellis, G. & Caramelli, D. (2012). Ancient DNA studies: new perspectives on old samples. *Genet. Sel. Evol.*, 44, 21.

Rohland, N. & Hofreiter, M. (2007). Ancient DNA extraction from bones and teeth. *Nat. Protoc.*, 2, 1756–1762.

Ronquist, F. & Huelsenbeck, J.P. (2001). MRBAYES: Bayesian inference of phylogenetic trees.

Bioinformatics, 17, 754–755.
Sablin, M. & Khlopachev, G. (2002). *The Earliest Ice Age Dogs: Evidence from Eliseevichi 1*. St. Petersburg.
Savolainen, P., Zhang, Y. ping, Luo, J., Lundeberg, J., Leitner, T., Vila, C., *et al.* (2002). Genetic Evidence for an East Asian Origin of Domestic Dogs. *Science (80-.).*, 298, 1610–1613.
Schmitt, E. & Wallace, S. (2012). Shape Change and Variation in the Cranial Morphology of Wild Canids (Canis lupus, Canis latrans, Canis rufus) Compared to Domestic Dogs (Canis familiaris) Using Geometric Morphometrics. *Int. J. Osteoarchaeol.*, 24, 42–50.
Shannon, L.M., Boyko, R.H., Castelhano, M., Corey, E., Hayward, J.J., McLean, C., *et al.* (2015). Genetic structure in village dogs reveals a Central Asian domestication origin. *Proc. Natl. Acad. Sci.*, 112, 13639–13644.
Shendure, J. & Ji, H. (2008). Next-generation DNA sequencing. *Nat. Biotechnol.*, 26, 1135–1145.
Skoglund, P., Ersmark, E., Palkopoulou, E., Dalén, L., Pontus Skoglund, Erik Ersmark, *et al.* (2015). Ancient Wolf Genome Reveals an Early Divergence of Domestic Dog Ancestors and Admixture into High-Latitude Breeds. *Curr. Biol.*, 25, 1515–1519.
Smith, C.I., Chamberlain, A.T., Riley, M.S., Cooper, A., Stringer, C.B. & Collins, M.J. (2001). Not just old but old and cold? *Nature*, 410.
Sousa, A.C. (2010). O Penedo Do Lexim E a Sequência Do Neolítico Final E Calcolítico Da Peninsula De Lisboa. Universidade de Lisboa.
Stiller, M., Green, R.E., Ronan, M., Simons, J.F., Du, L., He, W., *et al.* (2006). Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *PNAS*, 103.
Street, M. (1991). Bedburg-Königshoven: A Pre-Boreal Mesolithic site in the Lower Rhineland, Germany. In: *The Late Glacial in north-west Europe: human adaptation and environmental change at the end of Pleistocene* (eds. Barton, N., Roberts, A.J. & Roe, D.A.). British Archaeology.
Street, M. (2002). Ein Wiedersehen mit dem Hund von Bonn-Oberkassel. *Bonn. zool. Beitr.*, 50, 269–290.
Tchernov, E. & Horwitz, L.K. (1991). Body size diminution under domestication: Unconscious selection in primeval domesticates. *J. Anthropol. Archaeol.*, 10, 54–75.
Tchernov, E. & Valla, F.F. (1997). Two New Dogs, and Other Natufian Dogs, from the Southern Levant. *J. Archaeol. Sci.*, 24, 65–95.
Thalmann, O. & Perri, A.R. (2018). Paleogenomic Inferences of Dog Domestication. pp. 273–306.
Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V.J., Sawyer, S.K., Greenfield, D.L., *et al.* (2013). Complete Mitochondrial Genomes of Ancient Canids Suggest a European Origin of Domestic Dogs. *Science (80-.).*, 342, 871–874.
Torres, R.T., Ferreira, E., Rocha, R.G. & Fonseca, C. (2017). Hybridization between wolf and domestic dog: First evidence from an endangered population in central Portugal. *Mamm. Biol.*, 86, 70–74.
Trut, L. (1999). Early Canid Domestication: The Farm-Fox Experiment. *Am. Sci.*, 87, 160–169.
Turnbull, P.F. & Reed, C.A. (1974). The Fauna from the Terminal Pleistocene of Palegawra Cave, a Zarzian Occupation Site in Northeastern Iraq. *Fieldiana Anthropol.*, 63(3), 81–146.
Vigne, J.-D. (2005). L ' humérus de chien magdalénien de Erralla (Gipuzkoa , Espagne) et la domestication tardiglaciaire du loup en Europe: The Magdalénian dog humerus of Erralla (Gipuzkoa , Spain) and the Late Glacial wolf domestication in Europe. *MUNIBE*, 57, 279–287.
Vigne, J.-D., Carrère, I., Briois, F. & Guilaine, J. (2011). The Early Process of Mammal Domestication in the Near East. *Curr. Anthropol.*, 52, S255–S271.
Vigne, J.-D., Helmer, D. & Peters, J. (2005). New archaeozoological approaches to trace the first steps of animal domestication: general presentation, reflections and proposals. In: *First Steps of Animal Domestication: New archaeozoological approaches* (eds. Vigne, J., Peters, J. & Helmer, D.).
Vigne, J. (2007). Les débuts néolithiques de l'élevage des bovidés et de l'exploitation laitière dans l'ancien monde. *Les Cah. l'Ocha*, 12, 45–57.
Vila, C., Amorim, I.R., Leonard, J.A., Posada, D., Castroviejo, J., Petrucci-Fonseca, F., *et al.* (1999). Mitochondrial DNA phylogeography and population history of the grey wolf Canis lupus. *Mol. Ecol.*, 8, 2089–2103.
Vilà, C., Savolainen, P., Maldonado, J.E., Amorim, I.R., Rice, J.E., Honeycutt, R.L., *et al.* (1997). Multiple and Ancient Origins of the Domestic Dog. *Science (80-.).*, 276, 1687–1689.
VonHoldt, B.M., Pollinger, J.P., Lohmueller, K.E., Han, E., Parker, H.G., Quignon, P., *et al.* (2010). Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature*, 464, 898–902.
Waller, B.M., Peirce, K., Caeiro, C.C., Scheider, L., Burrows, A.M., McCune, S., *et al.* (2013). Paedomorphic facial expressions give dogs a selective advantage. *PLoS One*, 8, e82686.

Wang, G.-D., Zhai, W., Yang, H.-C., Wang, L., Zhong, L., Liu, Y.-H., *et al.* (2016). Out of southern East Asia: the natural history of domestic dogs across the world. *Cell Res.*, 26, 21–33.

Wang, G., Zhai, W., Yang, H., Fan, R., Cao, X., Zhong, L., *et al.* (2013). The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat. Commun.*, 4, 1860.

Wayne, R.K., Lehman, N., Allard, M.W. & Honeycutt, R.L. (1992). Mitochondrial DNA Variability of the Gray Wolf: Genetic Consequences of Population Decline and Habitat Fragmentation. *Conserv. Biol.*, 6, 559–569.

Wilkins, A.S., Wrangham, R.W. & Fitch, W.T. (2014). The "Domestication Syndrome" in Mammals: A Unified Explanation Based on Neural Crest Cell Behavior and Genetics. *Genetics*, 197, 795–808.

Woodward, Weyand, N. & Bunnell, M. (1994). DNA sequence from Cretaceous period bone fragments. *Science (80-.).*, 266, 1229–1232.

Wright, S. (1922). *Coefficients of Inbreeding and Relationship*. *Source Am. Nat.*

Wright, S. (1978). *Evolution and the genetics of populations. Vol. 4: Variability within and among Natural Populations*. University of Chicago Press, Chicago.

Yokoyama, Y., Lambeck, K., De Deckker, P., Johnston, P. & Fifield, L.K. (2000). Timing of the Last Glacial Maximum from observed sea-level minima. *Nature*, 406, 713–716.

Zeder, M.A. (2006). Archaeological Approaches to Documenting Animal Domestication. In: *Documenting domestication: new genetic and archaeological paradigms* (eds. Zeder, M.A., Bradley, D.G., Emshwiller, E. & Smith, B.D.). University of California Press.

Zeder, M.A. (2012a). Pathways to Animal Domestication. In: *Biodiversity in Agriculture* (eds. Gepts, P., Famula, T.R., Bettinger, R.L., Brush, S.B., Damania, A.B., McGuire, P.E., et al.). Cambridge University Press, Cambridge, pp. 227–259.

Zeder, M.A. (2012b). The Domestication of Animals. *J. Anthropol. Res.*, 68, 161–190.

Zilhão, J. (2001). Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. *Proc. Natl. Acad. Sci.*, 98, 14180–14185.

Zischler, H., Hoss, M., Handt, O., von Haeseler, A., van der Kuyl, A., Goudsmit, J., *et al.* (1995). Detecting Dinosaur DNA. *Science (80-.).*, 268, 1192–1193.

APPENDICES

Appendix I. Description of each remain, regarding its cultural period, which can be linked with map at Figure 4 using the numbers 01 to 14 for the site location.

		Number on the map	Geographic location	Dates (cal BP)	Elements	Dog Reference	Dating Reference	Notes
PALAEOLITHIC EARLY DOG-LIKE REMAINS	EUROPE	02	Erralla, Spain	19,000-12,500	1 humerus	(Altuna & Mariezkurrena 1985)	(Vigne *et al.* 2005)	Identified to belongs to a small dog.
		04	Le Closeau, France	14,999-14,055	7 fragments including mandible, meta carpal, metapodial and phalanxes	(Pionnier-Capitan *et al.* 2011)	(Pionnier-Capitan *et al.* 2011)	
		04	Montespan, France	15,500 - 13,500	1 atlas, 1 femur, 1 baculum	(Pionnier-Capitan *et al.* 2011)	(Pionnier-Capitan *et al.* 2011)	Identified to belong to a small male dog.
		04	Pont d'Ambon, France	12,952-12,451	39 skull, limb, mandible,vertebal, and tooth fragments	(Célérier & Delpech 1978)	(Célérier *et al.* 1999)	
		04	Saint-Thibaud-de-Couz, France	12,027-11,311	skull, right mandible, atlas, axis, some teeth, left humerous	(Chaix 2000)	(Pionnier-Capitan *et al.* 2011)	Morphological measurements of a *Canis* remains found in this site were assigned to an individual of reduced size very close to that observed in Neolithic dogs from Switzerland.
		05	Hauterive-Champréveyres, Switzerland	15,200-13,900	metatarsal and two teeth, second phalanx	(Morel & Müller 1997)	(Pionnier-Capitan *et al.* 2011)	Bone fragments were identified to belong to a dog based on measurements made on the upper canine.
		05	Kesslerloch Cave, Switzerland	14,600-14,100	partial maxillary fragment with teeth	(Napierala & Uerpmann 2012)	(Napierala & Uerpmann 2012)	This remains is considered **the earliest undisputed evidence of a domestic dog.**

Continued on the next page

PALAEOLITHIC EARLY DOG-LIKE REMAINS	**EUROPE**	06	Oelknitz, Germany	15,770-13,957	small phalanges, short metapods and part of distal humerus and tibia	(Mussil 2000)	(Pionnier-Capitan *et al.* 2011)	
		06	Teufelsbrucke, Germany	15,770-13,957	proximal metapodial fragment and first phalanx	(Mussil 2000)	(Pionnier-Capitan *et al.* 2011)	Similar remains in size to the small dog from Kniegrotte was found in this site.
		06	Kniegrotte, Germany	16,700 - 13,800	partial maxillary fragment with teeth	(Mussil 2000)	(Pionnier-Capitan et al. 2011)	Morphological measurements assigned bone remains found in this Upper Palaeolithic site to a small dog.
		06	Bonn–Oberkassel, Germany	14,708-13,874	maxillary, vertebrae, ulna and humerus fragments	(Street 2002)	(Pionnier-Capitan et al. 2011)	Initially thought as a wolf, archaeological (Benecke 1987; Street 2002; Janssens *et al.* 2018) and genetic analysis (Thalmann *et al.* 2013) show that the remains of a dog buried beside humans, to be **the first undisputed domestic dog skeleton**.
	ANATOLIA, LEVANT, CENTRAL ASIA	07	Shillourokambos, Cyprus	12,400-12,300	multiple elements	(Vigne *et al.* 2011)	(Vigne *et al.* 2011)	Multiple *Canis* remains were found at this site and associated with very small dogs.
		07	Klimonas, Cyprus	11,120-10,615	one phalanx	(Vigne *et al.* 2011)	(Vigne *et al.* 2011)	
		08	Hayonim Terrace	12,000-11,000	co-burials with humans, Complete skeleton	(Tchernov & Valla 1997)	(Tchernov & Valla 1997)	

Continued on the next page

PALAEOLITHIC EARLY DOG-LIKE REMAINS		08	Ain Mallaha, Israel	11,500	co-burials with humans, one skeleton juvenile and one adult, and one partial mandible	(Davis & Valla 1978; Tchernov & Valla 1997)	(Tchernov & Valla 1997)	Skeletons of a human and a small dog buried together are **the earliest accepted evidence** of the human-canine bond. This found supports that dogs were independently domesticated in Middle East from a lighter Southwest Asian wolf form, *Canis lupus arabs*, just before human became farmer.
		10	Tell Mureybet, Syria	11,500-11,300	skull and left and right mandibles	(Gourichon & Helmer 2008)	Évin J & Stordeur D (2008) cited in Larson *et al.* (Larson *et al.* 2012)	
		11	Palegawara Cave, Iraq	13,000	mandible	(Turnbull & Reed 1974)	(Turnbull & Reed 1974)	An early dog that shows clear evidence of cranial morphology, such as tooth size reduction and crowding in a smaller jaw (Zeder 2012a).
	EAST ASIA	12	Tumat, Eastern Russia	12,400	Two complete mummified dogs	("Bark to the future: Ice Age puppies may reveal canine evolution" 2016)	("Bark to the future: Ice Age puppies may reveal canine evolution" 2016)	Two well preserved dogs (the oldest mummified dogs in the world) turned up in permafrost Siberia, giving scientists hope to obtain high quality DNA and pinpoint the origin of domestic dogs.
		12	Ushki-I, Eastern Russia	12,900-12,600	complete skeleton	(Dikov 1996)	(Dikov 1996)	
		13	Nanzhuangtou, China	12,790 - 10,747	>31 fragments including a complete mandible	Jing Y 2010 cited in Larson et al 2012)	Jing Y 2010 cited in Larson et al 2012)	

Continued on the next page

MESOLITHIC EARLY DOG-LIKE REMAINS	**EUROPE**	01	Poças de São Bento, Portugal	6,866	almost complete skeleton	(Arias *et al.* 2015)	(Pires *et al.* 2019)	
		01	Muge, Portugal	7,070	Aamost complete skeleton	(Detry & Cardoso 2010)	(Pires *et al.* 2019)	
				7015-6930	fragments			
		01	Vale Boi, Portugal	7,080	one tooth	(Bicho *et al.* 2012)	(Pires *et al.* 2019)	
		03	Star Carr, England	11,658-10,633	skull fragment, single tooth, femur, tibia	(Degerbøl 1961; Clutton-Brock & Noe-Nygaard 1990)	(Degerbøl 1961; Clutton-Brock & Noe-Nygaard 1990)	It's possible that both bones came from the same dog or from unrelated dogs of the same size and age. Similar remains in size and proportions have also been found in another Mesolithic sites in Bedburg-Köningshoven, Germany (Street 1991) and in Denmark: these similarities found in these early dogs may indicate that they were the result of dispersal from a single founder population (by the time of early Mesolithic, the sea level had not yet separate Britain from the Continent). To sustain that propose
		03	Seamer Carr, England	11,866-11,246	6 vertebrae			

Continued on the next page

MESOLITHIC EARLY DOG-LIKE REMAINS								Bedburg-Köningshoven dog skull is more closer in size and morphology to the dog skulls found in Western Asia than it is to the Mesolithic wolf skull from Star Carr (Clutton-Brock 2016).
	ANATOLIA, LEVANT, CENTRAL ASIA	09	Cafer Höyük, Turkey	9,500-8,300	complete dog skull and 14 other elements	(Helmer & Gourichon 2008b)	(Calvin *et al*. 1999 cited in Larson et al 2012)	
		09	Çayönü, Turkey	9,200-9,100		(Özdoğan 1999 cited in Larson et al 2012)	(Özdoğan 1999 cited in Larson et al 2012)	
		10	Tell Aswad, Syria	10,200-9,400	tens of elements	(Helmer & Gourichon 2008a)	(Helmer & Gourichon 2008a)	
	EAST ASIA	12	Zhokhov, Russia	8480-8175	2 mandibles, maxilla, canine, radius, ribia	(Pitulko & Kasparov 2017)	(Pitulko & Kasparov 2017)	Suggests that sled dogs could have been used in Siberia around 15,000 years ago
		14	Natsushima Shell, Japan	9,300	complete skeleton	(Nishinakagawa et al. 1992).	(Nishinakagawa et al. 1992).	
		14	Kamikuroiwa, Japan	8,500 - 8,000	complete skeleton	(Nishinakagawa et al. 1992).	(Nishinakagawa et al. 1992).	

Appendix II. Script used for mtDNA analysis.

```
#!/bin/bash
#Script to run entire aDNA analysis at once

display_usage() {
echo '1st argument must be the path to the sample fastq file read1 [and
read2 if paired] in the following format: "path/to/read1:" OR
"path/to/read1:path/to/read2"
2nd argument is the mininum base quality for trimming
3rd argument is the minimum read length after adapter removal
4th argument is the path to the reference genome index
5th argument is the minimum mapping quality
6th argument is the path to the reference_composite genome index
7th argument is the number of threads available to use. example "15"
8th argument is the complete path to the directory were results must be
saved
9th argument is the sample name
10th argument is the minimum snp coverage
11th argument is the minimum snp quality
12th argument is the name of the snpEff database (ex: canis_mt or
canfam3.1)
13th argument is the maximum read size (ex: 150bp)'
}


#check if required arguments are there and display usage message
if [ -z "$1" ] || [ -z "$2" ] || [ -z "$3" ] || [ -z "$4" ] || [ -z "$5"
] || [ -z "$6" ] || [ -z "$7" ] || [ -z "$8" ] || [ -z "$9" ] || [ -z
"$10" ] || [ -z "$11" ] || [ -z "$12" ] || [ -z "$13" ]; then
        printf "Please provide the arguments required for the
script.\n\n"
        display_usage
        exit 1
fi

read1=$(echo $1 | cut -d ":" -f1)
read2=$(echo $1 | cut -d ":" -f2)

pair='true'

if [ ${#read2} -eq 0 ]; then
        pair='false'
fi

#variantes
trimqual="$2"
minreadlength="$3"
reference="$4"
minmapqual="$5"
reference_composite="$6"
threads="$7"
base_output="$8"
base_name="$9"
snp_coverage="${10}"
snp_quality="${11}"
snpEffDB="${12}"
readsize="${13}"
```

```
#########################
##### PREPROCESSING #####
#########################
```

Continued on the next page

```
mkdir $base_output'/preprocessing'


#FASTQC raw reads

mkdir $base_output'/preprocessing/fastqc_raw_reads'

fastqc -t "$threads" -o "$base_output/preprocessing/fastqc_raw_reads"
"$read1"

if [ $pair = 'true' ]; then
        fastqc -t "$threads" -o
"$base_output/preprocessing/fastqc_raw_reads" "$read2"
fi

#cutadapt

mkdir $base_output'/preprocessing/cutadapt'

if [ $pair = 'true' ]; then
        cutadapt -b
"AGATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNATCTCGTATGCCGTCTTCTGCTTG" -B
"AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT" -q
"$trimqual","$trimqual" -j "$threads" -m "$minreadlength" --max-n 0 -o
"$base_output"/preprocessing/cutadapt/"$base_name"_R1_trimmed.fq -p
"$base_output"/preprocessing/cutadapt/"$base_name"_R2_trimmed.fq "$read1"
"$read2"
else
        cutadapt -b
"AGATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNATCTCGTATGCCGTCTTCTGCTTG" -q
"$trimqual","$trimqual" -j "$threads" -m "$minreadlength" --max-n 0  -o
"$base_output"/preprocessing/cutadapt/"$base_name"_trimmed.fq "$read1"
fi

#FASTQC clean reads

mkdir $base_output'/preprocessing/fastqc_clean_reads'

if [ $pair = 'true' ]; then
        fastqc -t "$threads" -o
"$base_output"/preprocessing/fastqc_clean_reads
"$base_output"/preprocessing/cutadapt/"$base_name"_R1_trimmed.fq
        fastqc -t "$threads" -o
"$base_output"/preprocessing/fastqc_clean_reads
"$base_output"/preprocessing/cutadapt/"$base_name"_R2_trimmed.fq
else
        fastqc -t "$threads" -o
"$base_output"/preprocessing/fastqc_clean_reads
"$base_output"/preprocessing/cutadapt/"$base_name"_trimmed.fq
fi

#Merge clean reads
if [ $pair = 'true' ]; then
        mkdir $base_output'/preprocessing/merge_reads'
```

```
        AdapterRemoval --file1
"$base_output"/preprocessing/cutadapt/"$base_name"_R1_trimmed.fq --file2
"$base_output"/preprocessing/cutadapt/"$base_name"_R2_trimmed.fq --
basename
"$base_output"/preprocessing/merge_reads/"$base_name"_output_paired --
collapse --minlength 48
fi

#########################
######## MAPPING ########
#########################

#bwa aln - mapping against composite_genome

mkdir $base_output'/mapping'

if [ $pair = 'true' ]; then
        bwa aln -t "$threads" -l 1024 -o 2 -n 0.03 "$reference_composite"
"$base_output"/preprocessing/merge_reads/"$base_name"_output_paired.colla
psed > "$base_output"/mapping/"$base_name"_composite.sai
        bwa samse "$reference_composite"
"$base_output"/mapping/"$base_name"_composite.sai
"$base_output"/preprocessing/merge_reads/"$base_name"_output_paired.colla
psed > "$base_output"/mapping/"$base_name"_composite.sam
else
        bwa aln -t "$threads" -l 1024 -o 2 -n 0.03 "$reference_composite"
"$base_output"/preprocessing/cutadapt/"$base_name"_trimmed.fq >
"$base_output"/mapping/"$base_name"_composite.sai
        bwa samse "$reference_composite"
"$base_output"/mapping/"$base_name"_composite.sai
"$base_output"/preprocessing/cutadapt/"$base_name"_trimmed.fq >
"$base_output"/mapping/"$base_name"_composite.sam
fi

#Filter only unmapped reads (samtools view)

samtools view -bh -q 20 "$base_output"/mapping/"$base_name"_composite.sam
-U "$base_output"/mapping/"$base_name"_no_contamination.bam
samtools bam2fq "$base_output"/mapping/"$base_name"_no_contamination.bam
> "$base_output"/mapping/"$base_name"_no_contamination.fq

#bwa aln - mapping against dog reference

if [ $pair = 'true' ]; then
        bwa aln -t "$threads" -l 1024 -o 2 -n 0.03 "$reference"
"$base_output"/mapping/"$base_name"_no_contamination.fq >
"$base_output"/mapping/"$base_name"_no_contamination.sai
        bwa samse "$reference"
"$base_output"/mapping/"$base_name"_no_contamination.sai
"$base_output"/mapping/"$base_name"_no_contamination.fq >
"$base_output"/mapping/"$base_name"_no_contamination.sam
else
        bwa aln -t "$threads" -l 1024 -o 2 -n 0.03 "$reference"
"$base_output"/mapping/"$base_name"_no_contamination.fq >
"$base_output"/mapping/"$base_name"_no_contamination.sai
        bwa samse "$reference"
"$base_output"/mapping/"$base_name"_no_contamination.sai
```

Continued on the next page

Continued on the next page

```
"$base_output"/mapping/"$base_name"_no_contamination.fq >
"$base_output"/mapping/"$base_name"_no_contamination.sam
fi

#samtools

cd "$base_output"/mapping

samtools view -bh "$base_name"_no_contamination.sam -o "$base_name".bam
samtools sort "$base_name".bam > "$base_name"_sorted.bam
samtools index "$base_name"_sorted.bam

#picartools (add read groups)
java -XX:ParallelGCThreads="$threads" -XX:ConcGCThreads="$threads" -jar
$PICARD AddOrReplaceReadGroups VALIDATION_STRINGENCY="LENIENT"
ID="$base_name" SM="$base_name" PU="PU" LB="LB" PL="illumina"
I="$base_name"_sorted.bam O="$base_name"_RG.bam

#picartools (remove duplicates)

samtools sort "$base_name"_RG.bam > "$base_name"_sorted_RG.bam
samtools index "$base_name"_sorted_RG.bam
java -XX:ParallelGCThreads="$threads" -XX:ConcGCThreads="$threads" -jar
$PICARD MarkDuplicates VALIDATION_STRINGENCY="LENIENT"
REMOVE_DUPLICATES="true" I="$base_name"_sorted_RG.bam
O="$base_name"_no_dups.bam M="marked_dup_metrics.txt"

#samtools (filtering by flag and qual)

samtools sort "$base_name"_no_dups.bam > "$base_name"_sorted_no_dups.bam
samtools index "$base_name"_sorted_no_dups.bam
samtools view -q "$minmapqual" -bh "$base_name"_sorted_no_dups.bam -o
"$base_name"_filtered.bam
samtools sort "$base_name"_filtered.bam >
"$base_name"_filtered_sorted.bam
samtools index "$base_name"_filtered_sorted.bam

#########################
####### mapDamage #######
#########################

mkdir $base_output'/mapDamage'

mapDamage -l "$readsize" -d $base_output'/mapDamage' --rescale --rescale-
out "$base_name"_filtered_sorted_mapdamage.bam -i
"$base_name"_filtered_sorted.bam -r "$reference"

samtools index "$base_name"_filtered_sorted_mapdamage.bam

rm $base_name'_sorted.bam' $base_name'_sorted.bam.bai'
$base_name'_RG.bam' $base_name'_sorted_RG.bam'
$base_name'_sorted_RG.bam.bai' $base_name'_no_dups.bam'
$base_name'_sorted_no_dups.bam' $base_name'_sorted_no_dups.bam.bai'

#########################
#### VARIANT CALLING ####
#########################

mkdir $base_output'/variant_calling'
```

```
gatk HaplotypeCaller -R "$reference" -I
"$base_name"_filtered_sorted_mapdamage.bam --bam-output
"$base_name"_GATK_out.bam --pcr-indel-model CONSERVATIVE --dont-use-soft-
clipped-bases true --active-probability-threshold 0.002 --disable-
optimizations true --dont-trim-active-regions true  -O
"$base_output"/variant_calling/"$base_name".vcf

#vcf filter using GATK

cd $base_output'/variant_calling'
gatk VariantFiltration -R "$reference" -V
"$base_output"/variant_calling/"$base_name".vcf --filter-name
'FAILED_qual' --filter-expression "QD < $snp_quality" --filter-name
'FAILED_read_pos' --filter-expression "ReadPosRankSum < -1.0 &&
ReadPosRankSum > 1.0" --filter-name 'FAILED_base_rank' --filter-
expression "AS_BaseQRankSum < -1.0 && AS_BaseQRankSum > 1.0" --genotype-
filter-name 'FAILED_DP' --genotype-filter-expression "DP < $snp_coverage"
--genotype-filter-name "FAILED_HOMO" --genotype-filter-expression "isHet
== 1" --set-filtered-genotype-to-no-call true -O
"$base_output"/variant_calling/"$base_name"_filtered.vcf

gatk SelectVariants -V
"$base_output"/variant_calling/"$base_name"_filtered.vcf --exclude-
filtered true --exclude-non-variants true --remove-unused-alternates true
--restrict-alleles-to BIALLELIC -O
"$base_output"/variant_calling/"$base_name"_PASS_ONLY.vcf

gatk SelectVariants -V
"$base_output"/variant_calling/"$base_name"_filtered.vcf --exclude-
filtered true --exclude-non-variants true --remove-unused-alternates true
--select-type-to-include INDEL --restrict-alleles-to BIALLELIC -O
"$base_output"/variant_calling/"$base_name"_INDELS.vcf

########################
###### SNP effects #####
########################

java -jar /opt/anaconda3/share/snpeff-4.3.1t-1/snpEff.jar "$snpEffDB"
"$base_output"/variant_calling/"$base_name"_PASS_ONLY.vcf >
"$base_output"/variant_calling/"$base_name"_effects.vcf

#########################
####CONSENSUS SEQUENCE####
#########################

mkdir $base_output'/consensus_sequence'

bedtools genomecov -ibam
"$base_output"/mapping/"$base_name"_filtered_sorted_mapdamage.bam -bga >
"$base_output"/consensus_sequence/"$base_name"_cov_regions.bed

python /DATA/SCRIPTS/resolver_depth_nas_delecoes.py -v
"$base_output"/variant_calling/"$base_name"_INDELS.vcf -b
"$base_output"/consensus_sequence/"$base_name"_cov_regions.bed -o
"$base_output"/consensus_sequence/"$base_name"_cov_regions_DELok.bed
```

Continued on the next page

```
cat "$base_output"/consensus_sequence/"$base_name"_cov_regions_DELok.bed
| awk '$4 < 2' >
"$base_output"/consensus_sequence/"$base_name"_lowcov_regions.bed

bedtools maskfasta -fi "$reference" -bed
"$base_output"/consensus_sequence/"$base_name"_lowcov_regions.bed -fo
"$base_output"/consensus_sequence/"$base_name"_MT_REFERENCE_with_Nscov.fa

bgzip -c "$base_output"/variant_calling/"$base_name"_filtered.vcf >
"$base_output"/variant_calling/"$base_name"_filtered.vcf.gz

tabix -p vcf "$base_output"/variant_calling/"$base_name"_filtered.vcf.gz

grep "^#" "$base_output"/variant_calling/"$base_name"_filtered.vcf >
"$base_output"/variant_calling/"$base_name"_FAILED_ONLY.vcf

grep -v "^#" "$base_output"/variant_calling/"$base_name"_filtered.vcf |
grep "FAILED" >>
"$base_output"/variant_calling/"$base_name"_FAILED_ONLY.vcf

bedtools maskfasta -fi
"$base_output"/consensus_sequence/"$base_name"_MT_REFERENCE_with_Nscov.fa
-bed "$base_output"/variant_calling/"$base_name"_FAILED_ONLY.vcf -fo
"$base_output"/consensus_sequence/"$base_name"_MT_REFERENCE_with_Ns.fa

bgzip -c "$base_output"/variant_calling/"$base_name"_PASS_ONLY.vcf >
"$base_output"/variant_calling/"$base_name"_MT_PASS_ONLY.vcf.gz

tabix -p vcf
"$base_output"/variant_calling/"$base_name"_MT_PASS_ONLY.vcf.gz

bcftools consensus -f
"$base_output"/consensus_sequence/"$base_name"_MT_REFERENCE_with_Ns.fa -o
"$base_output"/consensus_sequence/"$base_name"_MT_REFERENCE_with_Ns_and_S
NPs.fa "$base_output"/variant_calling/"$base_name"_MT_PASS_ONLY.vcf.gz
```

Continued on the next page

Appendix III. Script used for nDNA analysis.

```
#!/bin/bash
#Script to run entire aDNA analysis at once

display_usage() {
echo '1st argument must be the path to the sample fastq file read1 [and
read2 if paired] in the following format: "path/to/read1:" OR
"path/to/read1:path/to/read2"
2nd argument is the mininum base quality for trimming
3rd argument is the minimum read length after adapter removal
4th argument is the path to the reference genome index
5th argument is the minimum mapping quality
6th argument is the path to the human genome index
7th argument is the number of threads available to use. example "15"
8th argument is the complete path to the directory were results must be
saved
9th argument is the sample name
10th argument is the minimum snp coverage
11th argument is the minimum snp quality
12th argument is the name of the snpEff database (ex: canis_mt or
canfam3.1)
13th argument is the maximum read size (ex: 150bp)
14th argument is the path to the pig genome index
15th argument is the path to the chicken genome index
16th argument is the path to the cow genome index'
}


#check if required arguments are there and display usage message
if [ -z "$1" ] || [ -z "$2" ] || [ -z "$3" ] || [ -z "$4" ] || [ -z "$5"
] || [ -z "$6" ] || [ -z "$7" ] || [ -z "$8" ] || [ -z "$9" ] || [ -z
"$10" ] || [ -z "$11" ] || [ -z "$12" ] || [ -z "$13" ] || [ -z "$14" ]
|| [ -z "$15" ] || [ -z "$16" ]; then
        printf "Please provide the arguments required for the
script.\n\n"
        display_usage
        exit 1
fi

read1=$(echo $1 | cut -d ":" -f1)
read2=$(echo $1 | cut -d ":" -f2)

pair='true'

if [ ${#read2} -eq 0 ]; then
        pair='false'
fi

#variantes
trimqual="$2"
minreadlength="$3"
reference="$4"
minmapqual="$5"
reference_human="$6"
threads="$7"
base_output="$8"
base_name="$9"
snp_coverage="${10}"
```

Continued on the next page

```
snp_quality="${11}"
snpEffDB="${12}"
readsize="${13}"
reference_pig="${14}"
reference_chicken="${15}"
reference_cow="${16}"

#########################
##### PREPROCESSING #####
#########################

mkdir $base_output'/preprocessing'


#FASTQC raw reads

mkdir $base_output'/preprocessing/fastqc_raw_reads'

fastqc -t "$threads" -o "$base_output/preprocessing/fastqc_raw_reads"
"$read1"

if [ $pair = 'true' ]; then
        fastqc -t "$threads" -o
"$base_output/preprocessing/fastqc_raw_reads" "$read2"
fi

#cutadapt

mkdir $base_output'/preprocessing/cutadapt'

if [ $pair = 'true' ]; then
        cutadapt -b
"AGATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNATCTCGTATGCCGTCTTCTGCTTG" -B
"AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT" -q
"$trimqual","$trimqual" -j "$threads" -m "$minreadlength" --max-n 0 -o
"$base_output"/preprocessing/cutadapt/"$base_name"_R1_trimmed.fq -p
"$base_output"/preprocessing/cutadapt/"$base_name"_R2_trimmed.fq "$read1"
"$read2"
else
        cutadapt -b
"AGATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNATCTCGTATGCCGTCTTCTGCTTG" -q
"$trimqual","$trimqual" -j "$threads" -m "$minreadlength" --max-n 0  -o
"$base_output"/preprocessing/cutadapt/"$base_name"_trimmed.fq "$read1"
fi

#FASTQC clean reads

mkdir $base_output'/preprocessing/fastqc_clean_reads'

if [ $pair = 'true' ]; then
        fastqc -t "$threads" -o
"$base_output"/preprocessing/fastqc_clean_reads
"$base_output"/preprocessing/cutadapt/"$base_name"_R1_trimmed.fq
        fastqc -t "$threads" -o
"$base_output"/preprocessing/fastqc_clean_reads
"$base_output"/preprocessing/cutadapt/"$base_name"_R2_trimmed.fq
else
```


Continued on the next page

```
        fastqc -t "$threads" -o
"$base_output"/preprocessing/fastqc_clean_reads
"$base_output"/preprocessing/cutadapt/"$base_name"_trimmed.fq
fi

#Merge clean reads
if [ $pair = 'true' ]; then
        mkdir $base_output'/preprocessing/merge_reads'

        AdapterRemoval --file1
"$base_output"/preprocessing/cutadapt/"$base_name"_R1_trimmed.fq --file2
"$base_output"/preprocessing/cutadapt/"$base_name"_R2_trimmed.fq --
basename
"$base_output"/preprocessing/merge_reads/"$base_name"_output_paired --
collapse --minlength 48
fi

#########################
######## MAPPING ########
#########################

#bwa aln - mapping against human reference

mkdir $base_output'/mapping'

if [ $pair = 'true' ]; then
        bwa aln -t "$threads" -l 1024 -o 2 -n 0.03 "$reference_human"
"$base_output"/preprocessing/merge_reads/"$base_name"_output_paired.colla
psed > "$base_output"/mapping/"$base_name"_human.sai
        bwa samse "$reference_human"
"$base_output"/mapping/"$base_name"_human.sai
"$base_output"/preprocessing/merge_reads/"$base_name"_output_paired.colla
psed > "$base_output"/mapping/"$base_name"_human.sam
else
        bwa aln -t "$threads" -l 1024 -o 2 -n 0.03 "$reference_human"
"$base_output"/preprocessing/cutadapt/"$base_name"_trimmed.fq >
"$base_output"/mapping/"$base_name"_human.sai
        bwa samse "$reference_human"
"$base_output"/mapping/"$base_name"_human.sai
"$base_output"/preprocessing/cutadapt/"$base_name"_trimmed.fq >
"$base_output"/mapping/"$base_name"_human.sam
fi

#Filter only unmapped reads from human alignment

samtools view -bh -q 20 "$base_output"/mapping/"$base_name"_human.sam -U
"$base_output"/mapping/"$base_name"_no_human_contamination.bam
samtools bam2fq
"$base_output"/mapping/"$base_name"_no_human_contamination.bam >
"$base_output"/mapping/"$base_name"_no_human_contamination.fq

#bwa aln - mapping against pig reference

if [ $pair = 'true' ]; then
        bwa aln -t "$threads" -l 1024 -o 2 -n 0.03 "$reference_pig"
"$base_output"/mapping/"$base_name"_no_human_contamination.fq >
"$base_output"/mapping/"$base_name"_pig.sai
        bwa samse "$reference_pig"
"$base_output"/mapping/"$base_name"_pig.sai
```


Continued on the next page

```
"$base_output"/mapping/"$base_name"_no_human_contamination.fq >
"$base_output"/mapping/"$base_name"_pig.sam
else
        bwa aln -t "$threads" -l 1024 -o 2 -n 0.03 "$reference_pig"
"$base_output"/mapping/"$base_name"_no_human_contamination.fq >
"$base_output"/mapping/"$base_name"_pig.sai
        bwa samse "$reference_pig"
"$base_output"/mapping/"$base_name"_pig.sai
"$base_output"/mapping/"$base_name"_no_human_contamination.fq >
"$base_output"/mapping/"$base_name"_pig.sam
fi

#Filter only unmapped reads from pig alignment

samtools view -bh -q 20 "$base_output"/mapping/"$base_name"_pig.sam -U
"$base_output"/mapping/"$base_name"_no_pig_human_contamination.bam
samtools bam2fq
"$base_output"/mapping/"$base_name"_no_pig_human_contamination.bam >
"$base_output"/mapping/"$base_name"_no_pig_human_contamination.fq

#bwa aln - mapping against chicken reference

if [ $pair = 'true' ]; then
        bwa aln -t "$threads" -l 1024 -o 2 -n 0.03 "$reference_chicken"
"$base_output"/mapping/"$base_name"_no_pig_human_contamination.fq >
"$base_output"/mapping/"$base_name"_chicken.sai
        bwa samse "$reference_chicken"
"$base_output"/mapping/"$base_name"_chicken.sai
"$base_output"/mapping/"$base_name"_no_pig_human_contamination.fq >
"$base_output"/mapping/"$base_name"_chicken.sam
else
        bwa aln -t "$threads" -l 1024 -o 2 -n 0.03 "$reference_chicken"
"$base_output"/mapping/"$base_name"_no_pig_human_contamination.fq >
"$base_output"/mapping/"$base_name"_chicken.sai
        bwa samse "$reference_chicken"
"$base_output"/mapping/"$base_name"_chicken.sai
"$base_output"/mapping/"$base_name"_no_pig_human_contamination.fq >
"$base_output"/mapping/"$base_name"_chicken.sam
fi

#Filter only unmapped reads from chicken alignment

samtools view -bh -q 20 "$base_output"/mapping/"$base_name"_chicken.sam -
U
"$base_output"/mapping/"$base_name"_no_chicken_pig_human_contamination.ba
m
samtools bam2fq
"$base_output"/mapping/"$base_name"_no_chicken_pig_human_contamination.ba
m >
"$base_output"/mapping/"$base_name"_no_chicken_pig_human_contamination.fq

#bwa aln - mapping against cow reference

if [ $pair = 'true' ]; then
        bwa aln -t "$threads" -l 1024 -o 2 -n 0.03 "$reference_cow"
"$base_output"/mapping/"$base_name"_no_chicken_pig_human_contamination.fq
> "$base_output"/mapping/"$base_name"_cow.sai
        bwa samse "$reference_cow"
"$base_output"/mapping/"$base_name"_cow.sai
```


Continued on the next page

```
"$base_output"/mapping/"$base_name"_no_chicken_pig_human_contamination.fq
> "$base_output"/mapping/"$base_name"_cow.sam
else
        bwa aln -t "$threads" -l 1024 -o 2 -n 0.03 "$reference_cow"
"$base_output"/mapping/"$base_name"_no_chicken_pig_human_contamination.fq
> "$base_output"/mapping/"$base_name"_cow.sai
        bwa samse "$reference_cow"
"$base_output"/mapping/"$base_name"_cow.sai
"$base_output"/mapping/"$base_name"_no_chicken_pig_human_contamination.fq
> "$base_output"/mapping/"$base_name"_cow.sam
fi

#Filter only unmapped reads from cow alignment

samtools view -bh -q 20 "$base_output"/mapping/"$base_name"_cow.sam -U
"$base_output"/mapping/"$base_name"_no_cow_chicken_pig_human_contaminatio
n.bam
samtools bam2fq
"$base_output"/mapping/"$base_name"_no_cow_chicken_pig_human_contaminatio
n.bam >
"$base_output"/mapping/"$base_name"_no_cow_chicken_pig_human_contaminatio
n.fq


#bwa aln - mapping against dog reference

if [ $pair = 'true' ]; then
        bwa aln -t "$threads" -l 1024 -o 2 -n 0.03 "$reference"
"$base_output"/mapping/"$base_name"_no_cow_chicken_pig_human_contaminatio
n.fq > "$base_output"/mapping/"$base_name"_only_dog.sai
        bwa samse "$reference"
"$base_output"/mapping/"$base_name"_only_dog.sai
"$base_output"/mapping/"$base_name"_no_cow_chicken_pig_human_contaminatio
n.fq > "$base_output"/mapping/"$base_name"_only_dog.sam
else
        bwa aln -t "$threads" -l 1024 -o 2 -n 0.03 "$reference"
"$base_output"/mapping/"$base_name"_no_cow_chicken_pig_human_contaminatio
n.fq > "$base_output"/mapping/"$base_name"_only_dog.sai
        bwa samse "$reference"
"$base_output"/mapping/"$base_name"_only_dog.sai
"$base_output"/mapping/"$base_name"_no_cow_chicken_pig_human_contaminatio
n.fq > "$base_output"/mapping/"$base_name"_only_dog.sam
fi

#samtools

cd "$base_output"/mapping

samtools view -bh "$base_name"_only_dog.sam -o "$base_name"_dog.bam
samtools sort "$base_name"_dog.bam > "$base_name"_sorted.bam
samtools index "$base_name"_sorted.bam

rm $base_name'_human.sai' $base_name'_human.sam' $base_name'_pig.sai'
$base_name'_pig.sam' $base_name'_chicken.sai' $base_name'_chicken.sam'
$base_name'_cow.sai' $base_name'_cow.sam'

#picartools (add read groups)
java -XX:ParallelGCThreads="$threads" -XX:ConcGCThreads="$threads" -jar
$PICARD AddOrReplaceReadGroups ID="$base_name" SM="$base_name" PU="PU"
```


Continued on the next page

```
LB="LB" PL="illumina" VALIDATION_STRINGENCY="LENIENT"
I="$base_name"_sorted.bam O="$base_name"_RG.bam

samtools sort "$base_name"_RG.bam > "$base_name"_sorted_RG.bam
samtools index "$base_name"_sorted_RG.bam
java -XX:ParallelGCThreads="$threads" -XX:ConcGCThreads="$threads" -jar
$PICARD MarkDuplicates VALIDATION_STRINGENCY="LENIENT"
REMOVE_DUPLICATES="true" I="$base_name"_sorted_RG.bam
O="$base_name"_no_dups.bam M="marked_dup_metrics.txt"

#samtools (filtering by flag and qual)

samtools sort "$base_name"_no_dups.bam > "$base_name"_sorted_no_dups.bam
samtools index "$base_name"_sorted_no_dups.bam
samtools view -F4 -q "$minmapqual" -bh "$base_name"_sorted_no_dups.bam -o
"$base_name"_filtered.bam
samtools sort "$base_name"_filtered.bam >
"$base_name"_filtered_sorted.bam
samtools index "$base_name"_filtered_sorted.bam


#########################
####### mapDamage #######
#########################

mkdir $base_output'/mapDamage'

mapDamage -l "$readsize" -d $base_output'/mapDamage' --rescale --rescale-
out "$base_name"_filtered_sorted_mapdamage.bam -i
"$base_name"_filtered_sorted.bam -r "$reference"

samtools index "$base_name"_filtered_sorted_mapdamage.bam

rm $base_name'_sorted.bam' $base_name'_sorted.bam.bai'
$base_name'_RG.bam' $base_name'_sorted_RG.bam'
$base_name'_sorted_RG.bam.bai' $base_name'_no_dups.bam'
$base_name'_sorted_no_dups.bam' $base_name'_sorted_no_dups.bam.bai'

#########################
#### VARIANT CALLING ####
#########################

mkdir $base_output'/variant_calling'

gatk HaplotypeCaller -R "$reference" -I
"$base_name"_filtered_sorted_mapdamage.bam --bam-output
"$base_name"_GATK_out.bam --pcr-indel-model CONSERVATIVE --dont-use-soft-
clipped-bases true --active-probability-threshold 0.002 --disable-
optimizations true --dont-trim-active-regions true  -O
"$base_output"/variant_calling/"$base_name".vcf

#vcf filter using GATK

cd $base_output'/variant_calling'
gatk VariantFiltration -R "$reference" -V
"$base_output"/variant_calling/"$base_name".vcf --filter-name 'quality' -
-filter-expression "QD < $snp_quality" --filter-name 'readpos' --filter-
expression "ReadPosRankSum < -1.0 && ReadPosRankSum > 1.0" --filter-name
'baserank' --filter-expression "AS_BaseQRankSum < -1.0 && AS_BaseQRankSum
```


Continued on the next page

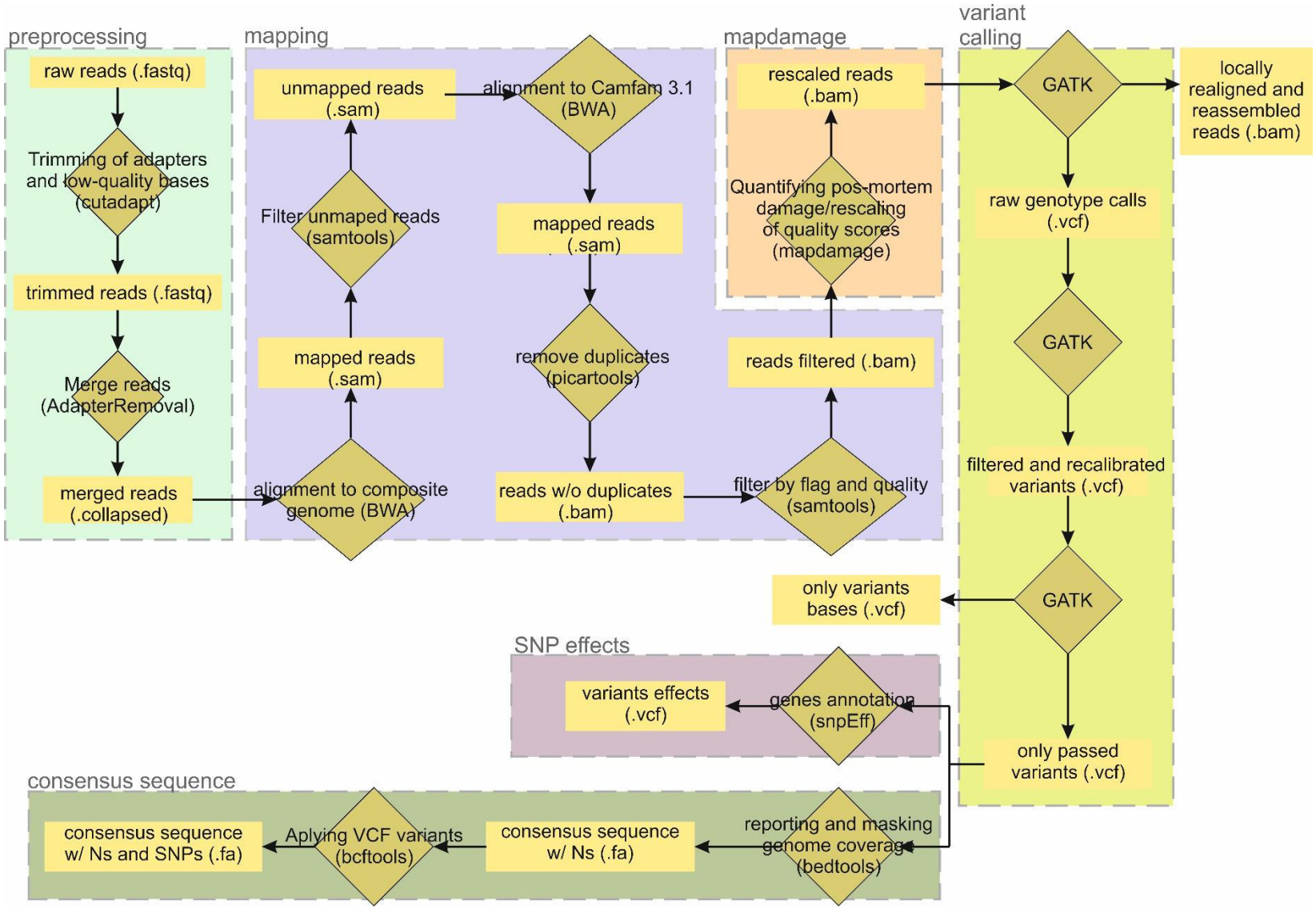
```
> 1.0" --genotype-filter-name 'DP' --genotype-filter-expression "DP <
$snp_coverage" --set-filtered-genotype-to-no-call true -O
"$base_output"/variant_calling/"$base_name"_filtered.vcf

gatk SelectVariants -V
"$base_output"/variant_calling/"$base_name"_filtered.vcf --exclude-
filtered true --exclude-non-variants true --remove-unused-alternates true
--restrict-alleles-to BIALLELIC -O
"$base_output"/variant_calling/"$base_name"_PASS_ONLY.vcf


########################
###### SNP effects #####
########################

java -jar /opt/anaconda3/share/snpeff-4.3.1t-1/snpEff.jar "$snpEffDB"
"$base_output"/variant_calling/"$base_name"_PASS_ONLY.vcf >
"$base_output"/variant_calling/"$base_name"_effects.vc
```

Appendix IV. Schematic representation of the pipeline carried out for sequences alignment and variant calling.



Appendix V. Description of samples used in the construction of Iberia and Eurasia phylogenetic tree.

Canine type	Age(BP)	Sub-region	Region	GenBank Accession Number	Sample Name	Sample Name (This Study)	Genome Region	Haplogroup	Haplogroup (in this study)	References
Coyote	Modern	n/a	n/a	NC_008093.1	n/a	Canis latrans	16,724 bp of mt genome	n/a	n/a	Bjornerfeldt et al 2009
Dog	Modern	Iberia Peninsula	Europe	EU789655	D6_R33	eDog_D6_ES/PT_HgD	16,195 bp of mt genome	D	D	Pang et al 2009
Dog	Modern	Spain	Europe	EU789714	A34_R34	eDog_A34_ES_HgA	16,195 bp of mt genome	A	A	Pang et al 2009
Wolf	Modern	Spain	Europe	DQ480505	n/a	eWolf8_ES	16,729 bp of mt genome	n/a	WOLF Hg1	Bjornerfeldt et al 2006
Wolf	Modern	Spain	Europe	KU644670	SpanishWolf2	eWolf1_ES	16,580 bp of mt genome	n/a	WOLF Hg1	Koblmuller et al, 2016
Wolf	Modern	Portugal	Europe	KT448278	CLU_PT	eWolf3_PT	16,729 bp of mt genome	n/a	WOLF Hg1	Koeplfi et al, 2015
Wolf	Modern	Portugal	Europe	KU644668	PortugueseWolf	eWolf2_PT	16,520 bp of mt genome	n/a	WOLF Hg1	Koblmuller et al, 2016
Dog	5000-4300	Leceia, Portugal	Europe	-	LYEP11	aDog_LYEP11_PT	16,587 bp of mt genome	n/a	A	This study
Dog	4000	Valladolid, Spain	Europe	-	LYEP53	aDog_LYEP53_ES	16,188 bp of mt genome	n/a	A	This study
Dog	4000	Valladolid, Spain	Europe	-	LYEP51	aDog_LYEP51_ES	12,715 bp of mt genome	n/a	C	This study
Dog	5000-4300	Leceia, Portugal	Europe	-	LYEP9	aDog_LYEP9_PT	16,454 bp of mt genome	n/a	C	This study
Wolf	4085–3856	Lexim, Portugal	Europe	-	LYEP27	aWolf1_PT_Hg1	6,526 bp of mt genome	WOLF Hg1	WOLF Hg1	This study
Dog	7173-6990 cal	Herxheim, Germany	Europe	KX379529	HXH	aDog_HXH_GER_HgC	16,725 bp of mt genome	C	C	Botigué et al, 2017
Dog	12500	Kartstein cave, Germany	Europe	KF661094	n/a	aDog_GER12500_HgC	16,239 bp of mt genome	C	C	Thalmann et al 2013
Canissp.	33500	Razboinichya cave, Russia	Europe	KF661092	n/a	aCanissp_RUS33500	16,411 bp of mt genome	n/a	aWolfHg2	Thalmann et al 2013
Wolf	14500	Kesslerloch cave, Switzerland	Europe	KF661087	Switzerland 1	aWolf26_SWI	16,357 bp of mt genome	n/a	aWolfHg2	Thalmann et al 2013
Wolf	18000	Medvezya cave, Russia	Europe	KF661081	n/a	aWolf28_RUS	16,414 bp of mt genome	n/a	aWolfHg2	Thalmann et al 2013
Wolf	30000	Goyet cave, Belgium	Europe	KF661080	n/a	aWolf29_BE	16,348 bp of mt genome	n/a	aWolfHg2	Thalmann et al 2013
Wolf	26000	Trou des Nutons, Belgium	Europe	KF661078	n/a	aWolf38_BE	16,170 bp of mt genome	n/a	aWolfHg2	Thalmann et al 2013
Wolf	22000	Kostenki, Russia	Europe	KF661085	n/a	aWolf27_RUS	16,397 bp of mt genome	n/a	aWolfHg2	Thalmann et al 2013
Wolf	14500	Kesslerloch cave, Switzerland	Europe	KF661095	Switzerland 3	aWolf40_SWI	16,089 bp of mt genome	n/a	aWolfHg2	Thalmann et al 2013
Dog	15000	Eliseevichi, Russia	Europe	KF661082	n/a	aDog_Rus15000	14,340 bp of mitochondrial genome	n/a	C	Thalmann et al 2013
Wolf	14500	Kesslerloch cave, Switzerland	Europe	KF661091	Switzerland 2	aWolf25_SWI	13,965 bp of mitochondrial genome	n/a	aWolfHg2	Thalmann et al 2013

Appendix VI. Description of samples used in the construction of Iberia & Eurasia phylogenetic networks.

Canine type	Age(BP)	Sub-region	Region	GenBank Accession Number	Sample Name	Sample Name (This Study)	Haplotype (only dogs - 66bp)	Haplotype (only wolves - 66bp)	Haplotype (dogs & wolves - 66bp)	Haplotype (dogs & wolves - 43bp)	Hg	Hg (In this study)	References	Frequency
Extant														
Dog	0	Portugal	Europe	AY706513	H38	eDog_H38_HgD	H16	-	H16	H47	D	D	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706504	H29	eDog_H29_HgC	H18	-	H18	H34	C	C	Pires *et al* 2006	2
Dog	0	Portugal	Europe	AY706491	H16	eDog_H16_HgC	H3	-	H3	H4	C	C	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706480	H05	eDog_H05_HgC	H3	-	H3	H4	C	C	Pires *et al* 2006	6
Dog	0	Portugal	Europe	AY706509	H34	eDog_H34_HgB	H9	-	H9	H25	B	B	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706506	H31	eDog_H31_HgB	H9	-	H9	H25	B	B	Pires *et al* 2006	2
Dog	0	Portugal	Europe	AY706505	H30	eDog_H30_HgB	H19	-	H19	H49	B	B	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706497	H22	eDog_H22_HgB	H20	-	H20	H25	B	B	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706484	H09	eDog_H09_HgB	H9	-	H9	H25	B	B	Pires *et al* 2006	2
Dog	0	Portugal	Europe	AY706517	H42	eDog_H42_HgA	H14	-	H14	H45	A	A	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706516	H41	eDog_H41_HgA	H15	-	H15	H46	A	A	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706510	H35	eDog_H35_HgA	H17	-	H17	H48	A	A	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706503	H28	eDog_H28_HgA	H14	-	H14	H45	A	A	Pires *et al* 2006	6
Dog	0	Portugal	Europe	AY706500	H25	eDog_H25_HgA	H21	-	H21	H50	A	A	Pires *et al* 2006	14
Dog	0	Portugal	Europe	AY706495	H20	eDog_H20_HgA	H2	-	H2	H17	A	A	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706493	H18	eDog_H18_HgA	H4	-	H4	H20	A	A	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706489	H14	eDog_H14_HgA	H7	-	H7	H21	A	A	Pires *et al* 2006	2
Dog	0	Portugal	Europe	AY706490	H15	eDog_H15_HgA	H6	-	H6	H28	A		Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706488	H13	eDog_H13_HgA	H8	-	H8	H1	A	A	Pires *et al* 2006	6
Dog	0	Portugal	Europe	AY706485	H10	eDog_H10_HgA	H7	-	H7	H21	A	A	Pires *et al* 2006	19
Dog	0	Portugal	Europe	AY706482	H07	eDog_H07_HgA	H2	-	H2	H17	A	A	Pires *et al* 2006	16
Dog	0	Portugal	Europe	AY706479	H04	eDog_H04_HgA	H8	-	H8	H1	A	A	Pires *et al* 2006	3
Dog	0	Portugal	Europe	AY706494	H19	eDog_H19_HgA	H11	-	H11	H42	A	A	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706501	H26	eDog_H26_HgB	H20	-	H20	H25	B	B	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706502	H27	eDog_H27_HgC	H3	-	H3	H4	C	C	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706507	H32	eDog_H32_HgC	H18	-	H18	H34	C	C	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706508	H33	eDog_H33_HgB	H9	-	H9	H25	B	B	Pires *et al* 2006	3
Dog	0	Portugal	Europe	AY706511	H36	eDog_H36_HgA	H8	-	H8	H1	A	A	Pires *et al* 2006	2
Dog	0	Portugal	Europe	AY706512	H37	eDog_H37_HgC	H3	-	H3	H4	C	C	Pires *et al* 2006	2
Dog	0	Portugal	Europe	AY706514	H39	eDog_H39_HgC	H3	-	H3	H4	C	C	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706515	H40	eDog_H40_HgD	H16	-	H16	H47	D	D	Pires *et al* 2006	4
Dog	0	Portugal	Europe	AY706518	H43	eDog_H43_HgA	H4	-	H4	H20	A	A	Pires *et al* 2006	2
Dog	0	Portugal	Europe	AY706519	H44	eDog_H44_HgA	H13	-	H13	H44	A	A	Pires *et al* 2006	1

Continued on the next page

Dog	0	Portugal	Europe	AY706520	H45	eDog_H45_HgB	H12	-	H12	H43	B	B	Pires *et al* 2006	3
Dog	0	Portugal	Europe	AY706521	H46	eDog_H46_HgA	H11	-	H11	H42	A	A	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706522	H47	eDog_H47_HgA	H10	-	H10	H41	A	A	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706523	H48	eDog_H48_HgA	H7	-	H7	H21	A	A	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706524	H49	eDog_H49_HgB	H9	-	H9	H25	B	B	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706476	H01	eDog_H01_HgA	H11	-	H11	H42	A	A	Pires *et al* 2006	6
Dog	0	Portugal	Europe	AY706477	H02	eDog_H02_HgD	H22	-	H22	H37	D	D	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706478	H03	eDog_H03_HgC	H3	-	H3	H4	C	C	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706481	H06	eDog_H06_HgD	H16	-	H16	H47	D	D	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706483	H08	eDog_H08_HgA	H2	-	H2	H17	A	A	Pires *et al* 2006	2
Dog	0	Portugal	Europe	AY706486	H11	eDog_H11_HgB	H9	-	H9	H25	B	B	Pires *et al* 2006	4
Dog	0	Portugal	Europe	AY706487	H12	eDog_H12_HgB	H9	-	H9	H25	B	B	Pires *et al* 2006	1
Dog	0	Portugal	Europe	AY706492	H17	eDog_H17_HgB	H9	-	H9	H25	B	B	Pires *et al* 2006	9
Dog	0	Portugal	Europe	AY706496	H21	eDog_H21_HgA	H8	-	H8	H1	A	A	Pires *et al* 2006	12
Dog	0	Portugal	Europe	AY706498	H23	eDog_H23_HgC	H3	-	H3	H4	C	C	Pires *et al* 2006	7
Dog	0	Portugal	Europe	AY706499	H24	eDog_H24_HgA	H4	-	H4	H20	A	A	Pires *et al* 2006	5
Dog	0	Spain	Europe	D83627	A11_m430	eDog_A11a_HgA	H2	-	H2	H17	A	A	Pang *et al* 2009	1
Dog	0	Spain	Europe	D83627	A11_R79	eDog_A11b_HgA	H2	-	H2	H17	A	A	Pang *et al* 2009	1
Dog	0	Spain	Europe	AB007385	A11_m343	eDog_A11c_HgA	H2	-	H2	H17	A	A	Pang *et al* 2009	1
Dog	0	Spain	Europe	AB007385	A11_m344	eDog_A11d_HgA	H2	-	H2	H17	A	A	Pang *et al* 2009	1
Dog	0	Spain	Europe	AB007385	A11_m401	eDog_A11e_HgA	H2	-	H2	H17	A	A	Pang *et al* 2009	1
Dog	0	Spain	Europe	AB007385	A11_R77	eDog_A11f_HgA	H2	-	H2	H17	A	A	Pang *et al* 2009	1
Dog	0	Spain	Europe	AY656744	A11_m498	eDog_A11g_HgA	H2	-	H2	H17	A	A	Pang *et al* 2009	1
Dog	0	Portugal	Europe	AB007396	A19_m440	eDog_A19_HgA	H9	-	H7	H21	A	A	Pang *et al* 2009	1
Dog	0	Spain	Europe	U96639	A20_R32	eDog_A20_HgA	H9	-	H7	H21	A	A	Pang *et al* 2009	1
Dog	0	Spain	Europe	AY656751	A34_R34	eDog_A34_HgA	H23	-	H23	H59	A	A	Pang *et al* 2009	1
Dog	0	Portugal	Europe	DQ480502	D6_m464	eDog_D6a_HgD	H24	-	H24	H60	D	D	Pang *et al* 2009	1
Dog	0	Spain	Europe	DQ480502	D6_R33	eDog_D6b_HgD	H24	-	H24	H60	D	D	Pang *et al* 2009	1
Wolf	0	Spain	Europe	DQ480505	1	eWolf8_Hg1	-	H3	H25	H40	n/a	WolfHg1	Bjornerfeldt *et al* 2006	1
Wolf	0	Iberia	Europe	EF380226	ClupMIT1	eWolf7_Hg1	-	H4	H26	H51	n/a	WolfHg1	Parra *et al*, unpublished	17
Wolf	0	Iberia	Europe	EF380227	ClupMIT2	eWolf6_Hg1	-	H6	H28	H51	n/a	WolfHg1	Parra *et al*, unpublished	1
Wolf	0	Iberia	Europe	EF380228	ClupMIT3	eWolf5_Hg1	-	H5	H27	H52	n/a	WolfHg1	Parra *et al*, unpublished	2
Wolf	0	Iberia	Europe	EF380229	ClupMIT4	eWolf4_Hg1	-	H3	H25	H40	n/a	WolfHg1	Parra *et al*, unpublished	6
Wolf	0	Portugal	Europe	KT448278	CLU_PT	eWolf3_Hg1	-	H3	H25	H40	n/a	WolfHg1	Koeplfi *et al*, 2015	1
Wolf	0	Portugal	Europe	KU644668	PortugueseWolf	eWolf2_Hg1	-	H3	H25	H40	n/a	WolfHg1	Koblmuller *et al* 2016	1

Continued on the next page

Wolf	0	Spain	Europe	KU644670	SpanishWolf2	eWolf1_Hg1	-	H4	H26	H51	n/a	WolfHg1	Koblmuller *et al* 2016	1
Wolf	0	Iberia	Europe	same as DQ480505	wH-1A	eWolf31	-	H3	H25	H40	n/a	WolfHg1	Pires *et al* 2017	46
Wolf	0	Portugal	Europe	JX845621	wH-4	eWolf11_Hg1	-	H7	H29	H53	n/a	WolfHg1	Pires *et al* 2017	1
Wolf	0	Portugal	Europe	JX845622	Wh-2	eWolf12_Hg1	-	H4	H26	H51	n/a	WolfHg1	Pires *et al* 2017	3
Wolf	0	Portugal	Europe	JX845625	wH-1C	eWolf15_Hg1	-	H3	H25	H40	n/a	WolfHg1	Pires *et al* 2017	3
Wolf	0	Spain	Europe	JX845623	wH-3	eWolf14_Hg1	-	H5	H27	H52	n/a	WolfHg1	Pires *et al* 2017	1
Wolf	0	Portugal	Europe	JX845624	wH-1B	eWolf13_Hg1	-	H3	H25	H40	n/a	WolfHg1	Pires *et al* 2017	1
Wolf	0	Portugal	Europe	not different from Pires et al. (2006) GenBank AY706485 and AY706523 (887 bp, dog haplotypes H10 and H48, respectively)	w5	eWolf28_Hg2	-	H8	H7	H21	n/a	WolfHg2	Pires *et al* 2017	1
Wolf	0	Iberia	Europe	AF008137	W1	eWolf17_Hg1	-	H3	H25	H40	Wolf Hg1	Wolf Hg1	Vila *et al* 1997	1 *[2]
Wolf	0	Finland	Europe	AF115698	W13	eWolf21_Hg2	-	-	-	H56	Wolf Hg2*[1]	Wolf Hg2*[1]	Randi et al, unpublished	3
Wolf	0	Bulgaria	Europe	AF115700	W15	eWolf22_Hg2	-	-	-	H57	Wolf Hg2*[1]	Wolf Hg2*[1]	Randi et al, unpublished	1
Wolf	0	Bulgaria	Europe	AF115701	W16	eWolf23_Hg2	-	-	-	H58	Wolf Hg2*[1]	Wolf Hg2*[1]	Randi et al, unpublished	10
Wolf	0	Slovakia/Bulgaria	Europe	AF115707	W17/D5	eWolf24_Hg2	-	-	-	H1	Wolf Hg2*[1]	Wolf Hg2*[1]	Randi et al, unpublished	5
Wolf	0	Bulgaria	Europe	AF115714	W18/D13	eWolf25_Hg2	-	-	-	H17	Wolf Hg2*[1]	Wolf Hg2*[1]	Randi et al, unpublished	1
Wolf	0	Spain	Europe	AF115702	W19	eWolf10_Hg1	-	H5	H27	H52	Wolf Hg1	Wolf Hg1	Randi et al, unpublished	1
Wolf	0	Spain	Europe	AF115703	W20	eWolf9_Hg1	-	H3	H25	H40	Wolf Hg1	Wolf Hg1	Randi et al, unpublished	2
Wolf	0	Italia	Europe	AF115699	W14	eWolf16_Hg2	-	-	-	H54	Wolf Hg2*[1]	Wolf Hg2*[1]	Randi et al, unpublished	101
Wolf	0	Latvia/Russia/Sweden	Europe	AF098123	W9	eWolf18_Hg2	-	-	-	H21	Wolf Hg2*[1]	Wolf Hg2*[1]	Koop *et al* 1999	1*[2]
Wolf	0	Ukraine	Europe	DQ421803	W15	eWolf19_Hg2	-	-	-	H28	Wolf Hg2*[1]	Wolf Hg2*[1]	Pilot *et al* 2006	8
Wolf	0	Bulgaria/Greece	Europe	DQ421804	W17	eWolf20_Hg2	-	-	-	H55	Wolf Hg2*[1]	Wolf Hg2*[1]	Pilot *et al* 2006	4
Wolf	0	Ukraine	Europe	DQ421805	W18	eWolf26_Hg2	-	-	-	H19	Wolf Hg2*[1]	Wolf Hg2*[1]	Pilot *et al* 2006	1
Wolf	0	Romania	Europe	AF338810	W8	eWolf27_Hg2	-	-	-	H58	Wolf Hg2*[1]	Wolf Hg2*[1]	Valiere *et al* 2003	1
							*[1]Haplotypes included are not representative of the whole Eurasia							
							*[2] Frequency not found; it is assumed presence/absence							
Chalcolithic														
Dog	6736	Romania	Europe	dryad.8gp06	aEurA23	aDog_aEurA23_HgD	-	-	-	H37	A/D	D	Pionnier-Capitan 2010	1
Dog	6736	Romania	Europe	dryad.8gp06	aEurA24	aDog_aEurA24_HgD	-	-	-	H37	A/D	D	Pionnier-Capitan 2010	1
Dog	6736	Romania	Europe	dryad.8gp06	aEurA25	aDog_aEurA25_HgD	-	-	-	H37	A/D	D	Pionnier-Capitan 2010	1
Dog	6736	Romania	Europe	dryad.8gp06	aEurA26	aDog_aEurA26_HgD	-	-	-	H37	A/D	D	Pionnier-Capitan 2010	1
Dog	6736	Romania	Europe	dryad.8gp06	aEurA27	aDog_aEurA27_HgD	-	-	-	H37	A/D	D	Pionnier-Capitan 2010	1
Dog	6736	Romania	Europe	dryad.8gp06	aEurA28	aDog_aEurA28_HgD	-	-	-	H37	A/D	D	Pionnier-Capitan 2010	1
Dog	6736	Romania	Europe	dryad.8gp06	aEurA29	aDog_aEurA29_HgD	-	-	-	H37	A/D	D	Pionnier-Capitan 2010	1
Dog	6200	Romania	Europe	dryad.8gp06	aEurA35	aDog_aEurA35_HgD	-	-	-	H38	D	D	Pionnier-Capitan 2010	1
Dog	6200	Romania	Europe	dryad.8gp06	aEurA36	aDog_aEurA36_HgD	-	-	-	H37	D	D	Pionnier-Capitan 2010	1
Dog	6200	Romania	Europe	dryad.8gp06	aEurA37	aDog_aEurA37_HgD	-	-	-	H37	A/D	D	Pionnier-Capitan 2010	1

Continued on the next page

Dog	6200	Romania	Europe	dryad.8gp06	aEurA38	aDog_aEurA38_HgD	-	-	-	H26	D	D	Pionnier-Capitan 2010	1
Dog	4800	Ukraine	Europe	dryad.8gp06	aEurA50	aDog_aEurA50_HgD	-	-	-	H26	D	D	Pionnier-Capitan 2010	1
Dog	4800	Moldova	Europe	dryad.8gp06	aEurA51	aDog_aEurA51_HgD	-	-	-	H37	A/D	D	Pionnier-Capitan 2010	1
Dog	5000-4300	Portugal	Europe	KY014680	LYEP72	aDog_LYEP72_HgA	H4	-	H4	H20	A	A	Pires *et al* 2019	1
Dog	5000-4300	Spain	Europe	KY014671	LYEP54	aDog_LYEP54_HgC	H3	-	H3	H4	C	C	Pires *et al* 2019	1
Dog	5000-4300	Spain	Europe	KY014668	LYEP50	aDog_LYEP50_HgC	H5	-	H5	H35	C	C	Pires *et al* 2019	1
Dog	5000-4300	Portugal	Europe	KY014666	LYEP23	aDog_LYEP23_HgC	H3	-	H3	H4	C	C	Pires *et al* 2019	1
Dog	5000-4300	Portugal	Europe	KY014665	LYEP22	aDog_LYEP22_HgA	H6	-	H6	H28	A	A	Pires *et al* 2019	1
Dog	5000-4300	Portugal	Europe	KY014664	LYEP20	aDog_LYEP20_HgC	H3	-	H3	H4	C	C	Pires *et al* 2019	1
Dog	5000-4300	Portugal	Europe	KY014663	LYEP17	aDog_LYEP17_HgA	H2	-	H2	H17	A	A	Pires *et al* 2019	1
Dog	5000-4300	Portugal	Europe	KY014662	LYEP16	aDog_LYEP16_HgA	H7	-	H7	H21	A	A	Pires *et al* 2019	1
Dog	5000-4300	Portugal	Europe	KY014661	LYEP15	aDog_LYEP15_HgC	H3	-	H3	H4	C	C	Pires *et al* 2019	1
Dog	5000-4300	Portugal	Europe	KY014660	LYEP14	aDog_LYEP14_HgA	H2	-	H2	H17	A	A	Pires *et al* 2019	1
Dog	5000-4300	Portugal	Europe	KY 014659	LYEP13	aDog_LYEP13_HgC	H3	-	H3	H4	C	C	Pires *et al* 2019	1
Dog	5000-4300	Portugal	Europe	KY 014658	LYEP12	aDog_LYEP12_HgA	H2	-	H2	H17	A	A	Pires *et al* 2019	1
Dog	5000-4300	Portugal	Europe	KY014656	LYEP9	aDog_LYEP9_NGS_454_HgA	H8	-	H8	H1	A	A	Pires *et al* 2019	1
Dog	5000-4300	Portugal	Europe	KY014657	LYEP11	aDog_LYEP11_NGS_454_HgA	H6	-	H6	H28	A	A	Pires *et al* 2019	1
Dog	4000	Spain	Europe	KY014669	LYEP51	aDog_LYEP51_NGS_454_HgC	H3	-	H3	H4	C	C	Pires *et al* 2019	1
Dog	4000	Spain	Europe	KY014670	LYEP53	aDog_LYEP53_NGS_454_HgC	H3	-	H3	H4	C	C	Pires *et al* 2019	1
Dog	5000-4300	Portugal	Europe	KY014655	LYEP8	aDog_LYEP8_HgC	H3	-	H3	H4	C	C	Pires *et al* 2019	1
Dog	5000-4300	Portugal	Europe	KY014654	LYEP7	aDog_LYEP7_HgA	H2	-	H2	H17	A	A	Pires *et al* 2019	1
Wolf	4085–3856	Portugal	Europe	KY014649	LYEP27	aWolf1_Hg1	-	H3	H25	H40	WOLF Hg1	WOLF Hg1	Pires *et al* 2019	1
Dog	5000-4300	Portugal	Europe	n/a	LYEP9	aDog_LYEP9_NGS_Illu	H8	-	H8	H1	n/a	A	This study	1
Dog	5000-4300	Portugal	Europe	n/a	LYEP11	aDog_LYEP11_NGS_Illu	H6	-	H6	H28	n/a	A	This study	1
Dog	4000	Spain	Europe	n/a	LYEP51	aDog_LYEP51_NGS_Illu	-	-	-	-	-	-	This study	1
Dog	4000	Spain	Europe	n/a	LYEP53	aDog_LYEP53_NGS_Illu	H3	-	H3	H4	n/a	C	This study	1
Dog	4110	Italy	Europe	AY741669	aEurA08	aDog_aEurA08_HgC	-	-	-	H4	C	C	Verginelli *et al* 2005	1
Neolithic														
Dog	5300–4500 BP	Sweden	Europe	AY673655	12	aDog_SE12_HgC	-	-	-	H36	C	C	Malmström *et al* 2008	1
Dog	5300–4500 BP	Sweden	Europe	AY673656	13	aDog_SE13_HgC	-	-	-	H4	C	C	Malmström *et al* 2008	1
Dog	5300–4500 BP	Sweden	Europe	AY673658	15	aDog_SE15_HgC	-	-	-	H4	C	C	Malmström *et al* 2008	1
Dog	5300–4500 BP	Sweden	Europe	AY673659	16	aDog_SE16_HgA	-	-	-	H17	A	A	Malmström *et al* 2008	1
Dog	5300–4500 BP	Sweden	Europe	AY673660	17	aDog_SE17_HgC	-	-	-	H4	C	C	Malmström *et al* 2008	1
Dog	5300–4500 BP	Sweden	Europe	AY673661	18	aDog_SE18_HgC	-	-	-	H4	C	C	Malmström *et al* 2008	1

Continued on the next page

Dog	5300–4500 BP	Sweden	Europe	AY673649	2	aDog_SE2_HgC	-	-	-	H24	C	C	Malmström *et al* 2008	1
Dog	5300–4500 BP	Sweden	Europe	AY673664	21	aDog_SE21_HgC	-	-	-	H4	C	C	Malmström *et al* 2008	1
Dog	5300–4500 BP	Sweden	Europe	AY673665	22	aDog_SE22_HgC	-	-	-	H4	C	C	Malmström *et al* 2008	1
Dog	5300–4500 BP	Sweden	Europe	AY673668	25	aDog_SE25_HgC	-	-	-	H4	C	C	Malmström *et al* 2008	1
Dog	5300–4500 BP	Sweden	Europe	AY673650	3	aDog_SE3_HgA	-	-	-	H17	A	A	Malmström *et al* 2008	1
Dog	5300–4500 BP	Sweden	Europe	AY673651	4	aDog_SE4_HgA	-	-	-	H20	A	A	Malmström *et al* 2008	1
Dog	5300–4500 BP	Sweden	Europe	AY673666	23	aDog_SE23_HgC	-	-	-	H4	C	C	Malmström *et al* 2008	1
Dog	5300–4500 BP	Sweden	Europe	AY673662	19	aDog_SE19_HgC	-	-	-	H4	C	C	Malmström *et al* 2008	1
Dog	6315	Hungary	Europe	dryad.8gp06	aEurA01	aDog_aEurA01_HgC	-	-	-	H4	C	C	Frantz *et al* 2016	1
Dog	6315	Hungary	Europe	dryad.8gp06	aEurA02	aDog_aEurA02_HgC	-	-	-	H4	C	C	Frantz *et al* 2016	1
Dog	6315	Hungary	Europe	dryad.8gp06	aEurA03	aDog_aEurA03_HgD	-	-	-	H26	D	D	Frantz *et al* 2016	1
Dog	6315	Hungary	Europe	dryad.8gp06	aEurA04	aDog_aEurA04_HgC	-	-	-	H4	C	C	Frantz *et al* 2016	1
Dog	6315	Hungary	Europe	dryad.8gp06	aEurA05	aDog_aEurA05_HgA	-	-	-	H20	A	A	Frantz *et al* 2016	1
Dog	6315	Hungary	Europe	dryad.8gp06	aEurA06	aDog_aEurA06_HgC	-	-	-	H4	C	C	Frantz *et al* 2016	1
Dog	6315	Hungary	Europe	dryad.8gp06	aEurA07	aDog_aEurA07_HgC	-	-	-	H4	C	C	Frantz *et al* 2016	1
Dog	6315	Hungary	Europe	dryad.8gp06	aEurA17	aDog_aEurA17_HgC	-	-	-	H24	C	C	Frantz *et al* 2016	1
Dog	6315	Hungary	Europe	dryad.8gp06	aEurA18	aDog_aEurA18_HgC	-	-	-	H4	C	C	Frantz *et al* 2016	1
Dog	6315	Hungary	Europe	dryad.8gp06	aEurA19	aDog_aEurA19_HgC	-	-	-	H4	C	C	Frantz *et al* 2016	1
Dog	4900-4700 cal	Ireland	Europe	dryad.8gp06/PRJEB13070	Iri4000	aDog_Iri4000_HgC	-	-	-	H4	C	C	Frantz *et al* 2016	1
Dog	6093	France	Europe	dryad.8gp06	aEurA20	aDog_aEurA20_HgC	-	-	-	H4	C	C	Pionnier-Capitan 2010	1
Dog	6093	France	Europe	dryad.8gp06	aEurA21	aDog_aEurA21_HgC	-	-	-	H4	C	C	Pionnier-Capitan 2010	1
Dog	6093	France	Europe	dryad.8gp06	aEurA22	aDog_aEurA22_HgC	-	-	-	H4	C	C	Pionnier-Capitan 2010	1
Dog	6000	France	Europe	dryad.8gp06	aEurA33	aDog_aEurA33_HgD	-	-	-	H27	D	D	Pionnier-Capitan 2010	1
Dog	6000	France	Europe	dryad.8gp06	aEurA34	aDog_aEurA34_HgA	-	-	-	H28	A	A	Pionnier-Capitan 2010	1
Dog	6200	France	Europe	dryad.8gp06	aEurA32	aDog_aEurA32_HgC	-	-	-	H4	C	C	Pionnier-Capitan 2010	1
Dog	6570	Germany	Europe	dryad.8gp06	aEurA40	aDog_aEurA40_HgC	-	-	-	H4	C	C	Pionnier-Capitan 2010	1
Dog	6570	Germany	Europe	dryad.8gp06	aEurA41	aDog_aEurA41_HgC	-	-	-	H29	C	C	Pionnier-Capitan 2010	1
Dog	6570	Germany	Europe	dryad.8gp06	aEurA42	aDog_aEurA42_HgD	-	-	-	H26	D	D	Pionnier-Capitan 2010	1
Dog	7379	Romania	Europe	dryad.8gp06	aEurA47	aDog_aEurA47_HgD	-	-	-	H30	D	D	Pionnier-Capitan 2010	1
Dog	7379	Romania	Europe	dryad.8gp06	aEurA48	aDog_aEurA48_HgC	-	-	-	H24	C	C	Pionnier-Capitan 2010	1
Dog	6200	France	Europe	dryad.8gp06	aEurA56	aDog_aEurA56_HgC	-	-	-	H31	C	C	Pionnier-Capitan 2010	1
Dog	6200	France	Europe	dryad.8gp06	aEurA57	aDog_aEurA57_HgC	-	-	-	H4	C	C	Pionnier-Capitan 2010	1
Dog	6098	Russia	Nothern Asia	dryad.8gp06	aEurA62	aDog_aEurA62_HgA	-	-	-	H20	A	A	Pionnier-Capitan 2010	1
Dog	6098	Russia	Nothern Asia	dryad.8gp06	aEurA63	aDog_aEurA63_HgA	-	-	-	H32	A	A	Pionnier-Capitan 2010	1
Dog	6700	Romania	Europe	dryad.8gp06	aEurA49	aDog_aEurA49_HgD	-	-	-	H26	D	D	Pionnier-Capitan 2010	1

Continued on the next page

Dog	5750	France	Europe	dryad.8gp06	aEurA67	aDog_aEurA67_HgC	-	-	-	H4	C	C	Pionnier-Capitan 2010	1
Dog	5750	France	Europe	dryad.8gp06	aEurA68	aDog_aEurA68_HgB	-	-	-	H25	B	B	Pionnier-Capitan 2010	1
Dog	4349	France	Europe	dryad.8gp06	aEurA30	aDog_aEurA30_HgC	-	-	-	H4	C	C	Pionnier-Capitan 2010	1
Dog	4349	France	Europe	dryad.8gp06	aEurA31	aDog_aEurA31_HgD	-	-	-	H26	D	D	Pionnier-Capitan 2010	1
Dog	7930	Iran	SW Asia	dryad.8gp06	aEurA71	aDog_aEurA71_HgD	-	-	-	H26	D	D	Pionnier-Capitan 2010	1
Dog	7930	Iran	SW Asia	dryad.8gp06	aEurA72	aDog_aEurA72_HgD	-	-	-	H26	D	D	Pionnier-Capitan 2010	1
Dog	7930	Iran	SW Asia	dryad.8gp06	aEurA73	aDog_aEurA73_HgD	-	-	-	H26	D	D	Pionnier-Capitan 2010	1
Dog	7930	Iran	SW Asia	dryad.8gp06	aEurA74	aDog_aEurA74_HgD	-	-	-	H26	D	D	Pionnier-Capitan 2010	1
Dog	7930	Iran	SW Asia	dryad.8gp06	aEurA75	aDog_aEurA75_HgD	-	-	-	H26	D	D	Pionnier-Capitan 2010	1
Dog	7930	Iran	SW Asia	dryad.8gp06	aEurA76	aDog_aEurA76_HgD	-	-	-	H26	D	D	Pionnier-Capitan 2010	1
Dog	7930	Iran	SW Asia	dryad.8gp06	aEurA77	aDog_aEurA77_HgD	-	-	-	H26	D	D	Pionnier-Capitan 2010	1
Dog	5775	Switzerland	Europe	dryad.8gp06	aEurA81	aDog_aEurA81_HgC	-	-	-	H4	C	C	Pionnier-Capitan 2010	1
Dog	5775	Switzerland	Europe	dryad.8gp06	aEurA82	aDog_aEurA82_HgC	-	-	-	H4	C	C	Pionnier-Capitan 2010	1
Dog	5775	Switzerland	Europe	dryad.8gp06	aEurA83	aDog_aEurA83_HgC	-	-	-	H4	C	C	Pionnier-Capitan 2010	1
Dog	5775	Switzerland	Europe	dryad.8gp06	aEurA84	aDog_aEurA84_HgD	-	-	-	H33	D	D	Pionnier-Capitan 2010	1
Dog	5775	Switzerland	Europe	dryad.8gp06	aEurA85	aDog_aEurA85_HgD	-	-	-	H26	D	D	Pionnier-Capitan 2010	1
Dog	5904	Russia	Nothern Asia	dryad.8gp06	aEurA64	aDog_aEurA64_HgA	-	-	-	H20	A	A	Pionnier-Capitan 2010	1
Dog	5904	Russia	Nothern Asia	dryad.8gp06	aEurA65	aDog_aEurA65_HgA	-	-	-	H20	A	A	Pionnier-Capitan 2010	1
Dog	5904	Russia	Nothern Asia	dryad.8gp06	aEurA66	aDog_aEurA66_HgA	-	-	-	H20	A	A	Pionnier-Capitan 2010	1
Dog	5500-5300	France	Europe	EU287462	VTC3	aDog_aEurA15_HgC	-	-	-	H4	C	C	Deguilloux *et al* 2009	1
Dog	5500-5300	France	Europe	EU287461	VTC2	aDog_aEurA14_HgC	-	-	-	H4	C	C	Deguilloux *et al* 2009	1
Dog	5500-5300	France	Europe	EU287460	VTC1	aDog_aEurA13_HgC	-	-	-	H34	C	C	Deguilloux *et al* 2009	1
Dog	5500-5000	Portugal	Europe	KY014653	LYEP5	aDog_LYEP5_HgA	H2	-	H2	H17	A	A	Pires *et al* 2019	1
Dog	7173-6990 cal	Germany	Europe	KX379529	HXH	aDog_HXH_HgC	-	-	-	H35	C	C	Botigué *et al* 2017	1
Dog	4850-4582 cal	Germany	Europe	KX379528	CTC	aDog_CTC_HgC	-	-	-	H4	C	C	Botigué *et al* 2017	1
Dog	4290 ± 40	Portugal	Europe	KY014667	LYEP28	aDog_LYEP28_HgA	H2	-	H2	H17	A	A	Pires *et al* 2019	1
Mesolithic														
Dog	7550	France	Europe	dryad.8gp06	aEurA69	aDog_aEurA69_HgC	-	-	-	H4	C	C	Pionnier-Capitan 2010	1
Dog	9197	Romania	Europe	dryad.8gp06	aEurA43	aDog_aEurA43_HgC	-	-	-	H4	C	C	Pionnier-Capitan 2010	1
Dog	9197	Romania	Europe	dryad.8gp06	aEurA44	aDog_aEurA44_HgC	-	-	-	H24	C	C	Pionnier-Capitan 2010	1
Dog	9197	Romania	Europe	dryad.8gp06	aEurA45	aDog_aEurA45_HgC	-	-	-	H4	C	C	Pionnier-Capitan 2010	1
Dog	9197	Romania	Europe	dryad.8gp06	aEurA46	aDog_aEurA46_HgC	-	-	-	H4	C	C	Pionnier-Capitan 2010	1
Dog	6372	Estonia	Europe	dryad.8gp06	aEurA52	aDog_aEurA52_HgC	-	-	-	H24	C	C	Pionnier-Capitan 2010	1
Dog	6372	Estonia	Europe	dryad.8gp06	aEurA53	aDog_aEurA53_HgC	-	-	-	H4	C	C	Pionnier-Capitan 2010	1
Dog	6372	Estonia	Europe	dryad.8gp06	aEurA54	aDog_aEurA54_HgC	-	-	-	H24	C	C	Pionnier-Capitan 2010	1

Continued on the next page

Dog	6372	Estonia	Europe	dryad.8gp06	aEurA55	aDog_aEurA55_HgC	-	-	-	H24	C	C	Pionnier-Capitan 2010	1
Dog	8921	Romania	Europe	dryad.8gp06	aEurA58	aDog_aEurA58_HgC	-	-	-	H4	C	C	Pionnier-Capitan 2010	1
Dog	7550	France	Europe	dryad.8gp06	aEurA70	aDog_aEurA70_HgC	-	-	-	H4	C	C	Pionnier-Capitan 2010	1
Dog	7845-7625	Portugal	Europe	KY014683	LYEP75	aDog_LYEP75_HgA	H2	-	H2	H17	A	A	Pires *et al* 2019	1
Dog	7680-7485	Portugal	Europe	KY014682	LYEP74	aDog_LYEP74_HgA	H1	-	H1	H17	A	A	Pires *et al* 2019	1
Dog	7915-7605	Portugal	Europe	KY014677	LYEP68B	aDog_LYEP68B_HgA	H2	-	H2	H17	A	A	Pires *et al* 2019	1
Dog	7680-7450	Portugal	Europe	KY014676	LYEP68A	aDog_LYEP68A_HgC	H3	-	H3	H4	C	C	Pires *et al* 2019	1
Dog	7835-7685	Portugal	Europe	KY014675	SEP002	aDog_SEP002_HgA	H2	-	H2	H17	A	A	Pires *et al* 2019	1
Dog	7965-7848	Portugal	Europe	KY014652	LYEP3	aDog_LYEP3_HgA	H2	-	H2	H17	A	A	Pires *et al* 2019	1
Wolf	9670	Italy	Europe	AY741668	PIC3	aWolf2_Hg2	-	-	-	H17	n/a	WOLF Hg2	Verginelli *et al* 2005	1
Wolf	9860	Italy	Europe	AY741667	PIC2	aWolf5_Hg1	-	-	-	H25	n/a	WOLF Hg2	Verginelli *et al* 2005	1
Palaeotlithic														
Dog	13229	Romania	Europe	dryad.8gp06	aEurA59	aDog_aEurA59_HgC	-	-	-	H22	C	C	Pionnier-Capitan 2010	1
Dog	12500	Germany	Europe	KF661094	Ger12500	aDog_Ger12500_HgC	-	-	-	H4	n/a	C	Thalmann *et al* 2013	1
Dog	13250	Israel	SW Asia	dryad.8gp06	aEurA39	aDog_aEurA39_HgA	-	-	-	H21	A	A	Pionnier-Capitan 2010	1
Dog	12701	France	Europe	dryad.8gp06	aEurA60	aDog_aEurA60_HgC	-	-	-	H4	C	C	Pionnier-Capitan 2010	1
Canis sp.	33500	Russia	Nothern Asia	KF661092	Russia/33,500	aCanissp_Rus33500	-	-	-	H3	n/a	n/a	Thalmann *et al* 2013	1
Canis sp.	36000	Belgium	Europe	KF661079	Belgium/36,000	aCanissp_Belgium36000	-	-	-	H23	n/a	n/a	Thalmann *et al* 2013	1
Wolf	14000	Portugal	Europe	KY014651	LYEP46	aWolf3_Hg2	-	H1	H3	H4	WOLF Hg2	WOLF Hg2	Pires *et al* 2019	1
Wolf	80886±31265	Portugal	Europe	KY014650	LYEP44	aWolf4_Hg2	-	H2	H2	H17	WOLF Hg2	WOLF Hg2	Pires *et al* 2019	1
Wolf	14670	Italy	Europe	AY741666	aEurA10	aWolf6_Hg2	-	-	-	H4	n/a	WOLF Hg2	Verginelli *et al* 2005	1
Wolf	20790	CzechRep	Europe	DQ852634	a6	aWolf7_Hg2	-	-	-	H5	WOLF Hg2	WOLF Hg2	Stiller *et al* 2006	1
Wolf	44250	CzechRep	Europe	DQ852635	a11	aWolf8_Hg2	-	-	-	H3	WOLF Hg2	WOLF Hg2	Stiller *et al* 2006	1
Wolf	41710	CzechRep	Europe	DQ852636	a14	aWolf9_Hg2	-	-	-	H6	WOLF Hg2	WOLF Hg2	Stiller *et al* 2006	1
Wolf	15199	Russia	Nothern Asia	DQ852638	a21	aWolf10_Hg2	-	-	-	H7	WOLF Hg2	WOLF Hg2	Stiller *et al* 2006	1
Wolf	32500	Russia	Nothern Asia	DQ852640	a26	aWolf11_Hg2	-	-	-	H8	n/a	WOLF Hg2	Stiller *et al* 2006	1
Wolf	48020	Russia	Nothern Asia	DQ852641	a28	aWolf12_Hg2	-	-	-	H9	WOLF Hg2	WOLF Hg2	Stiller *et al* 2006	1
Wolf	49930	Russia	Nothern Asia	DQ852642	a29	aWolf13_Hg2	-	-	-	H9	n/a	WOLF Hg2	Stiller *et al* 2006	1
Wolf	LateGlacial	Belgium	Europe	DQ852644	a33	aWolf14_Hg2	-	-	-	H7	WOLF Hg2	WOLF Hg2	Stiller *et al* 2006	1
Wolf	LateGlacial	Belgium	Europe	DQ852645	a34	aWolf15_Hg2	-	-	-	H10	WOLF Hg2	WOLF Hg2	Stiller *et al* 2006	1
Wolf	PleniGlacial	Belgium	Europe	DQ852646	a36	aWolf16_Hg2	-	-	-	H11	WOLF Hg2	WOLF Hg2	Stiller *et al* 2006	1
Wolf	PleniGlacial	Belgium	Europe	DQ852647	a37	aWolf17_Hg2	-	-	-	H12	WOLF Hg2	WOLF Hg2	Stiller *et al* 2006	1
Wolf	13681	Belgium	Europe	DQ852648	a38	aWolf18_Hg2	-	-	-	H13	WOLF Hg2	WOLF Hg2	Stiller *et al* 2006	1
Wolf	24780	Belgium	Europe	DQ852649	a42	aWolf19_Hg2	-	-	-	H14	WOLF Hg2	WOLF Hg2	Stiller *et al* 2006	1
Wolf	21810	Belgium	Europe	DQ852650	a44	aWolf20_Hg2	-	-	-	H2	WOLF Hg2	WOLF Hg2	Stiller *et al* 2006	1

Continued on the next page

Wolf	27580	Germany	Europe	DQ852653	a48	aWolf21_Hg2	-	-	-	H15	WOLF Hg2	WOLF Hg2	Stiller *et al* 2006	1
Wolf	34310	Hungary	Europe	DQ852660	a61	aWolf22_Hg2	-	-	-	H16	WOLF Hg2	WOLF Hg2	Stiller *et al* 2006	1
Wolf	27520	Ukraine	Europe	DQ852661	a17	aWolf23_Hg2	-	-	-	H8	WOLF Hg2	WOLF Hg2	Stiller *et al* 2006	1
Wolf	29650	Ukraine	Europe	DQ852662	a18	aWolf24_Hg2	-	-	-	H8	n/a	WOLF Hg2	Stiller *et al* 2006	1
Wolf	14500	Switzerland	Europe	KF661091	Switzerland 2	aWolf25_Hg2	-	-	-	H18	n/a	WOLF Hg2	Thalmann *et al* 2013	1
Wolf	14500	Switzerland	Europe	KF661087	Switzerland 1	aWolf26_Hg2	-	-	-	H19	n/a	WOLF Hg2	Thalmann *et al* 2013	1
Wolf	22000	Russia	Nothern Asia	KF661085	Russia 22,000	aWolf27_Hg2	-	-	-	H8	n/a	WOLF Hg2	Thalmann *et al* 2013	1
Wolf	18000	Russia	Nothern Asia	KF661081	Russia 18,000	aWolf28_Hg2	-	-	-	H20	n/a	WOLF Hg2	Thalmann *et al* 2013	1
Wolf	30000	Belgium	Europe	KF661080	Belgium 30,000	aWolf29_Hg	-	-	-	H14	n/a	WOLF Hg2	Thalmann *et al* 2013	1
Wolf	22285–17869	Italy	Europe	MH593822	OWW4	aWolf30_Hg2	-	-	-	H1	WOLF Hg2	WOLF Hg2	Ciucani *et al* 2019	1
Wolf	17550	Italy	Europe	MH085476	OWW16	aWolf31_Hg2	-	-	-	H2	WOLF Hg2	WOLF Hg2	Ciucani *et al* 2019	1
Wolf	23940	Italy	Europe	MH085475	OWW15	aWolf32_Hg2	-	-	-	H2	WOLF Hg2	WOLF Hg2	Ciucani *et al* 2019	1
Wolf	23940	Italy	Europe	MH085474	OWW13	aWolf33_Hg2	-	-	-	H2	WOLF Hg2	WOLF Hg2	Ciucani *et al* 2019	1
Wolf	17550	Italy	Europe	MH085473	OWW12	aWolf34_Hg2	-	-	-	H3	WOLF Hg2	WOLF Hg2	Ciucani *et al* 2019	1
Wolf	23940	Italy	Europe	MH085472	OWW11	aWolf35_Hg2	-	-	-	H2	WOLF Hg2	WOLF Hg2	Ciucani *et al* 2019	1
Wolf	24700	Italy	Europe	MH085471	OWW9	aWolf36_Hg2	-	-	-	H1	WOLF Hg2	WOLF Hg2	Ciucani *et al* 2019	1
Wolf	23940	Italy	Europe	MH085470	OWW8	aWolf37_Hg2	-	-	-	H2	WOLF Hg2	WOLF Hg2	Ciucani *et al* 2019	1

Appendix VII. Graphical report of quality score per position in read (bp).

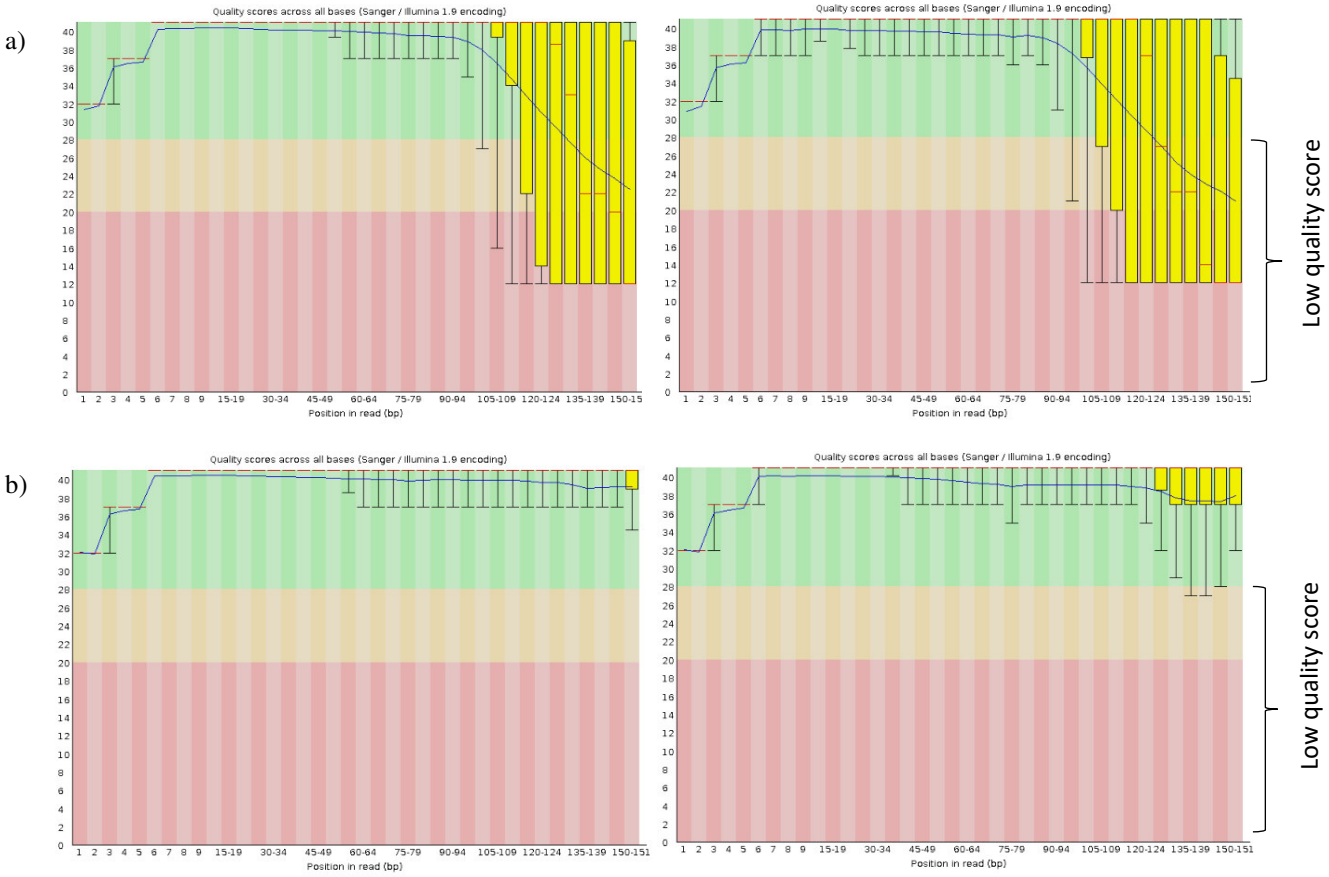
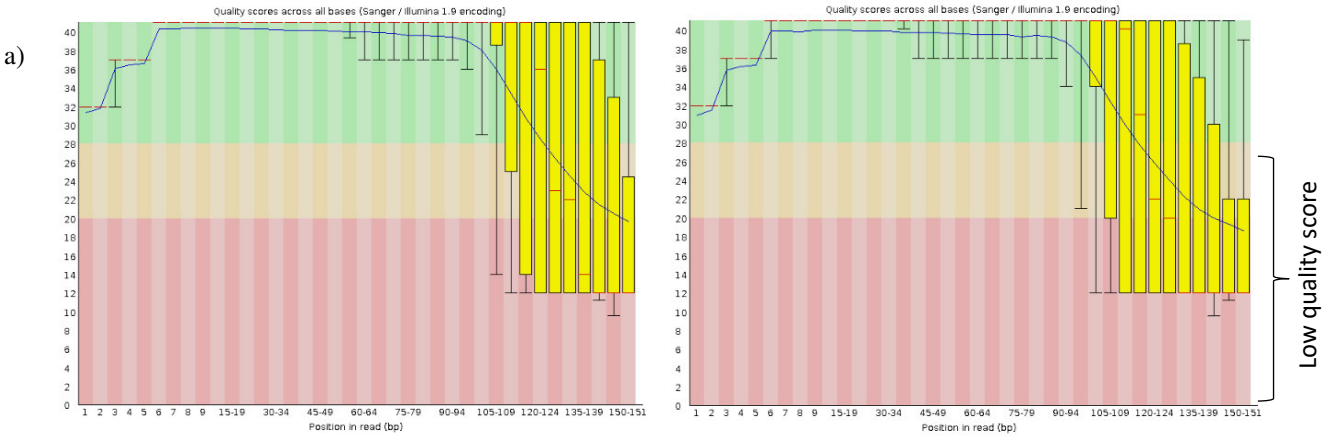


Figure 1. Graphical report of quality score per position in read (bp). a) LYEP9 read 1 and read 2 before removal of adapters; b) LYEP9 read 1 and read 2 after removal of adapters and low-quality bases (<30).



Continued on the next page

b)

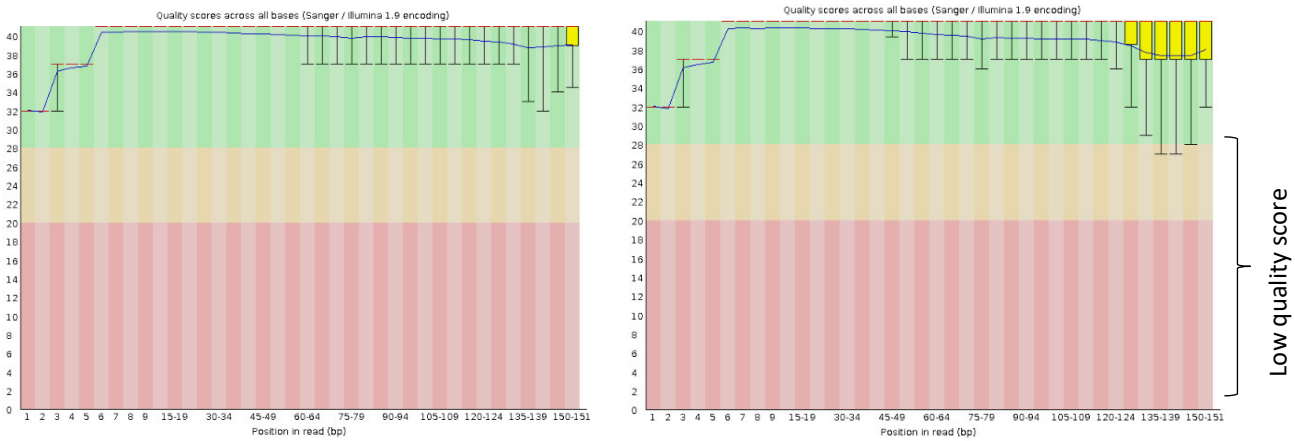


Figure 2. Graphical report of quality score per position in read (bp). a) LYEP11 read 1 and read 2 before removal of adapters; b) LYEP11 read 1 and read 2 after removal of adapters and low-quality bases (<30).

a)

b)

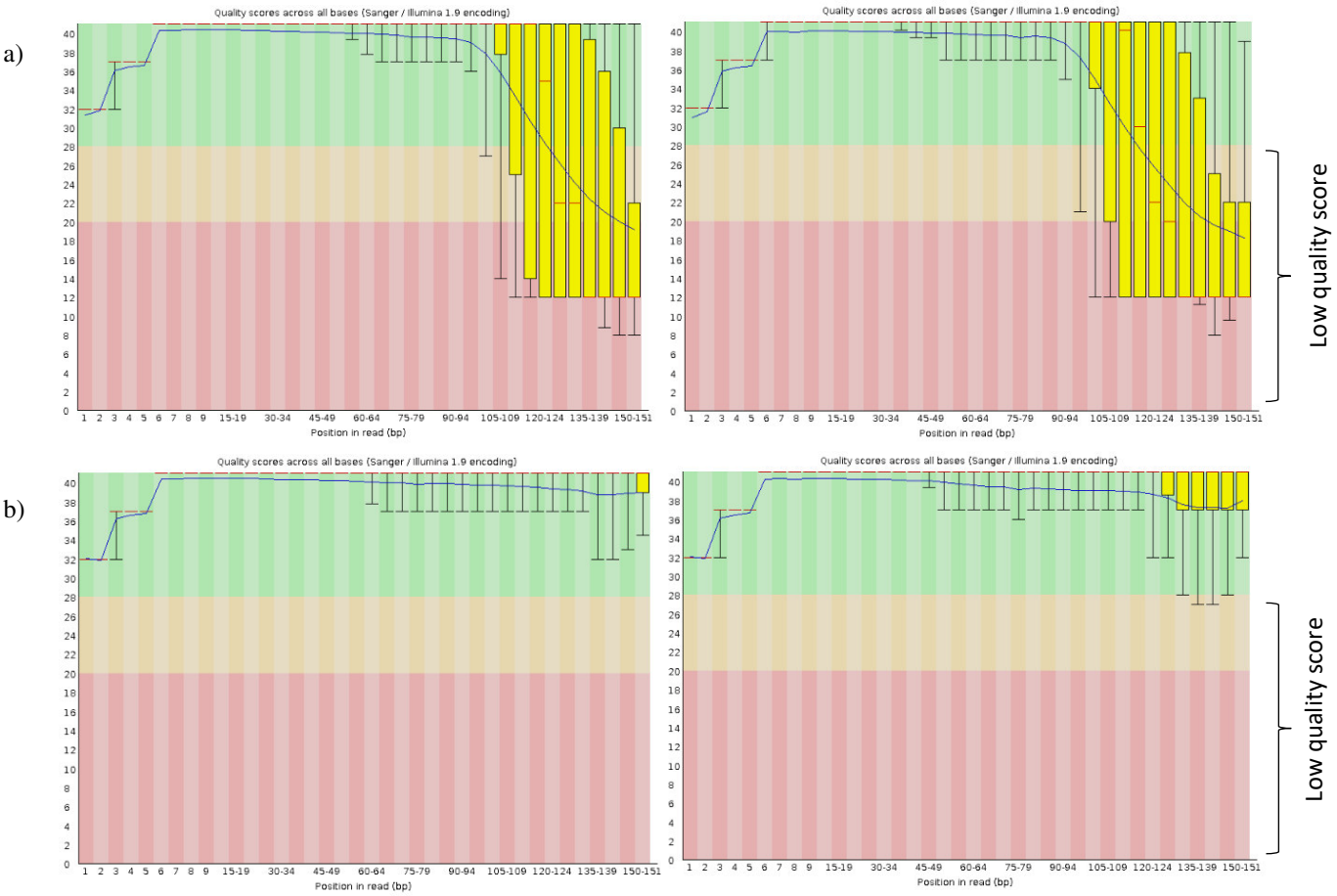


Figure 3. Graphical report of quality score per position in read (bp). a) LYEP51 read 1 and read 2 before removal of adapters; b) LYEP51 read 1 and read 2 after removal of adapters and low-quality bases (<30).

Continued on the next page

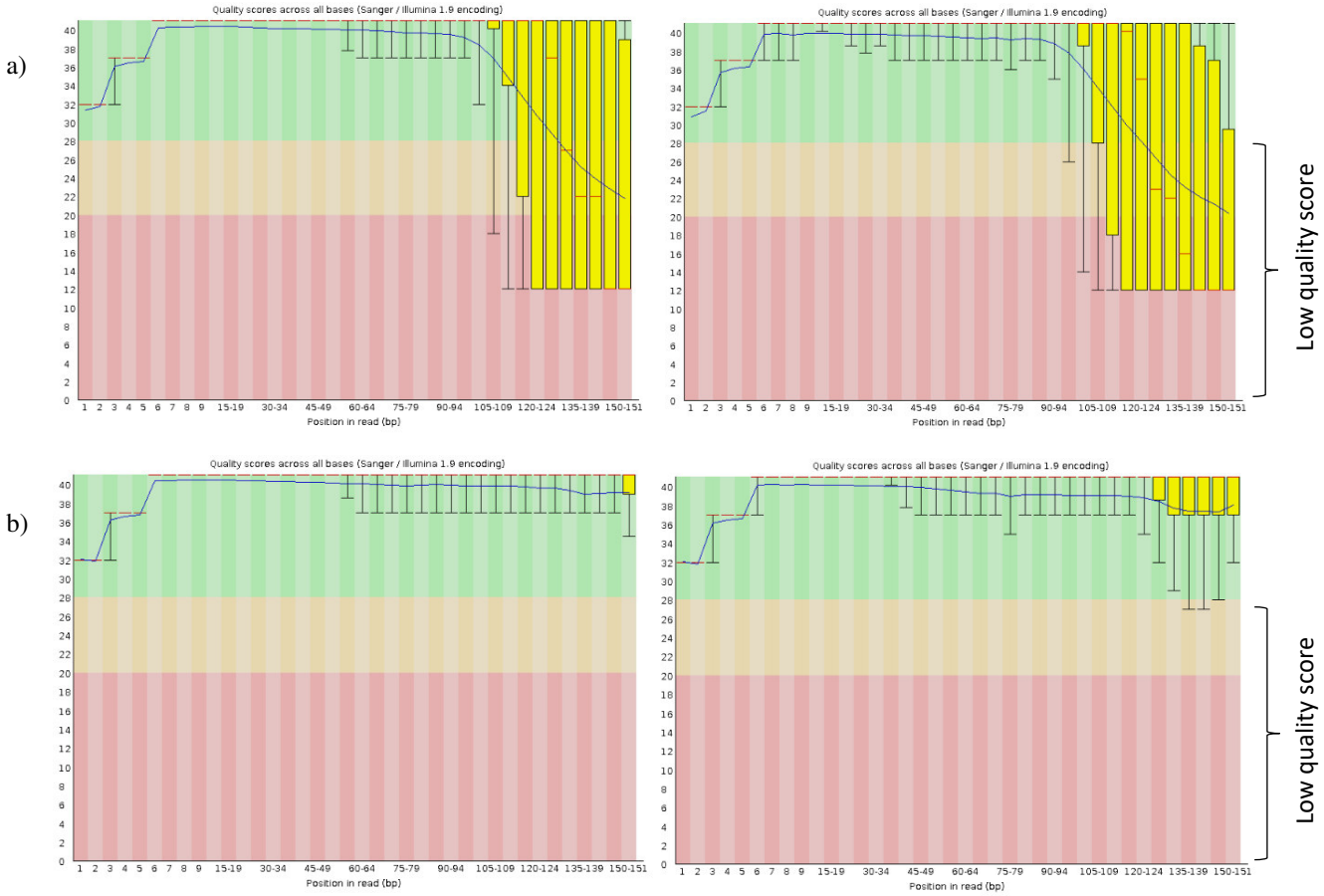


Figure 4. Graphical report of quality score per position in read (bp). a) LYEP53 read 1 and read 2 before removal of adapters; b) LYEP53 read 1 and read 2 after removal of adapters and low-quality bases (<30).

Continued on the next page

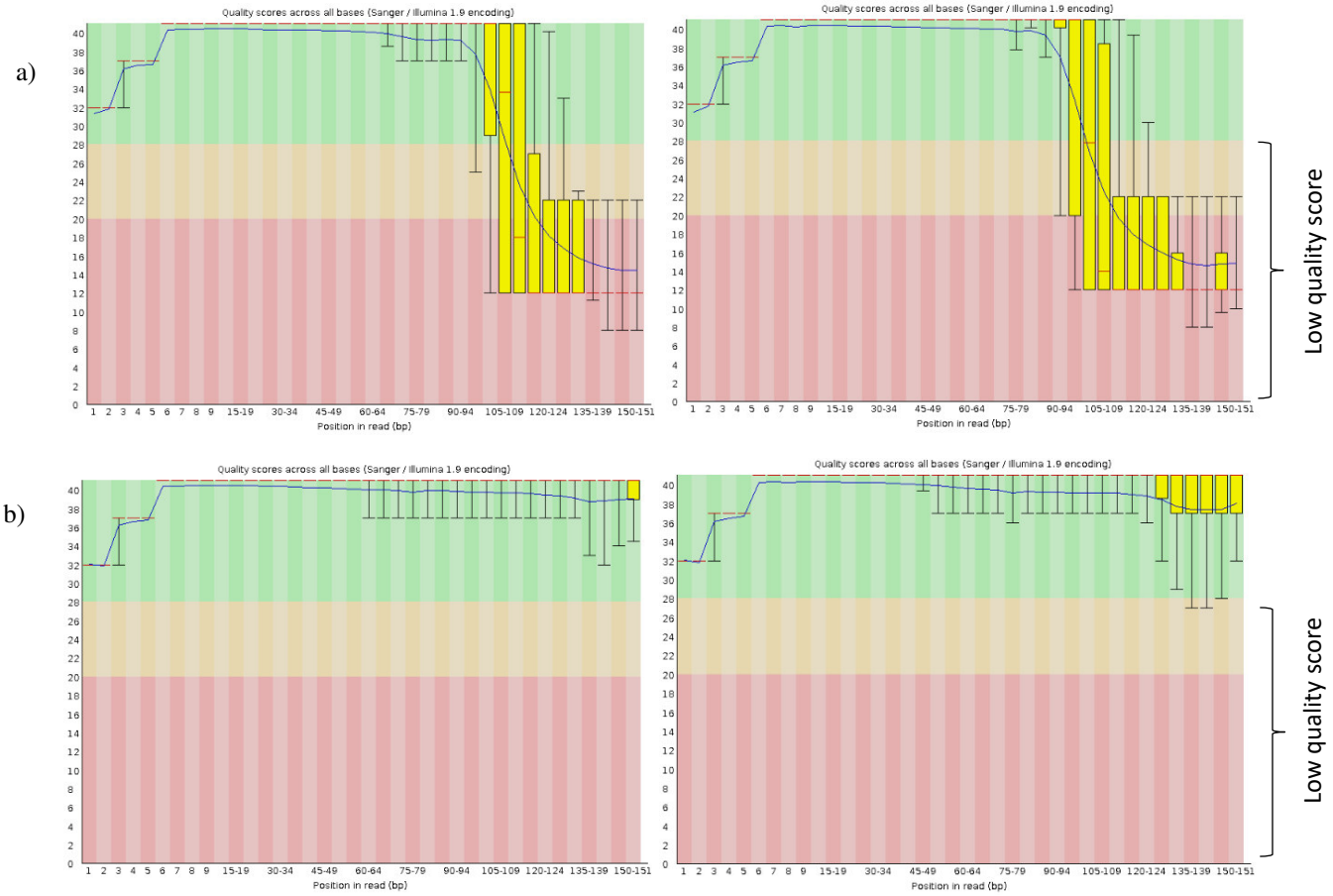


Figure 5. Graphical report of quality score per position in read (bp). a) LYEP27 read 1 and read 2 before removal of adapters; b) LYEP27 read 1 and read 2 after removal of adapters and low-quality bases (<30).

Appendix VIII. Summary of each sample variants.

LYEP9																			
Position	1351	2678	2683	2962	3196	4906	4940	5367	5444	6065	6401	6554	7593	8281	8368	8807	9911	10319	10611
Reference	A	T	G	C	T	T	T	C	T	A	C	T	T	T	C	G	A	T	A
Alternative	G	TG	A	T	C	C	C	T	C	G	T	C	C	C	T	A	ATG	C	T
Position	10992	13299	14977	15214	15620	15627	15639	15665	15814	16025	16660	16672							
Reference	G	T	T	G	T	A	T	T	C	T	T	C							
Alternative	A	A	C	A	C	G	A	C	T	C	TCC	T							
LYEP11																			
Position	388	1019	2678	2683	2962	3093	3196	5367	5444	6065	8201	8281	8368	8807	8982	9911	10992	11793	13299
Reference	A	T	T	G	C	T	T	C	T	A	G	T	C	G	G	A	G	T	T
Alternative	G	C	TG	A	T	C	C	T	C	G	A	C	T	A	A	ATG	A	C	A
Position	15214	15627	15639	15652	15814	16025													
Reference	G	A	T	G	C	T													
Alternative	A	G	A	A	T	C													
LYEP51																			
Position	381	733	1204	1748	1756	3598	4234	4503	5009	5367	5444	6470	7670	8323	8764	9222	9708	10533	10776
Reference	T	T	T	T	C	G	C	A	C	C	T	G	A	A	G	C	C	A	T
Alternative	A	C	C	C	T	A	T	G	T	T	C	A	G	G	T	T	T	T	C
Position	11250	11322	11323	11400	11402	11825	11963	12272	12636	12788	12813	13660	13708	14647	14692	15185	15508	15526	15611
Reference	T	T	C	T	T	T	C	T	T	T	G	C	C	T	G	T	C	C	T
Alternative	C	C	T	C	C	C	T	C	C	C	A	T	T	C	A	C	T	T	C

Continued on the next page

LYEP53																			
Position	381	557	733	1204	1454	1748	1756	2232	2678	2683	3196	3406	3469	5009	5367	5444	5624	5732	6065
Reference	T	A	T	T	G	T	C	A	T	G	T	C	G	C	C	T	G	A	A
Alternative	A	G	C	C	A	C	T	G	TG	A	C	T	A	T	T	C	A	T	G
Position	6257	6470	7058	8221	8225	8281	8323	8368	8703	8760	8764	8807	8991	9078	9708	9860	9911	10386	10404
Reference	G	G	T	A	T	T	A	C	G	A	G	G	A	T	C	C	A	G	C
Alternative	A	A	C	C	C	C	G	T	A	G	T	A	G	C	T	CA	ATG	A	T
Position	10533	10776	10917	10992	11250	11322	11323	11400	11402	11572	11963	12330	12788	12813	13261	13299	13319	13618	13660
Reference	A	T	G	G	T	T	C	T	T	A	C	A	T	G	C	T	C	A	C
Alternative	T	C	A	A	C	C	T	C	C	C	T	G	C	A	T	A	T	G	T
Position	13708	13777	14647	14692	15185	15435	15484	15508	15526	15611	15650	15955	16671						
Reference	C	G	T	G	T	G	A	C	C	T	T	C	T						
Alternative	T	A	C	A	C	A	G	T	T	C	C	T	C						
LYEP27																			
Position	2051	3034	3451	5520	5938	6620	7676	11042	13301	13803	14355	14672							
Reference	C	T	C	C	C	C	T	G	C	G	G	A							
Alternative	T	C	T	T	T	T	C	A	T	A	A	G							

Appendix IX. HiSeq sequencing statistics for ancient samples.

MITOCHONDRIAL								
Sample	Total reads sequenced	Number of reads after merge (forward and reverse sequence reads)	Number of retained reads after removal of human+pig+chicken+cow contamination	% of reads mapped human+pig+chicken+cow contamination genome	Number of retained reads after alignment against endogenous mtDNA	% of merged reads mapped against endogenous mtDNA (%)	mean coverage of mtDNA genome	% Duplicate reads
LYEP9	50569242	38185926	38185850	0.0002	3919	0.010	17x	0.23
LYEP11	44203964	30062441	30062342	0.0003	3537	0.012	12x	0.17
LYEP51	47366572	32212912	32212896	0.00005	767	0.002	2x	0.08
LYEP53	40676611	29837937	29837914	0.0001	1438	0.005	5x	0.10
LYEP27	39883611	11615845	11615823	0.0002	625	0.005	1x	0.28
NUCLEAR								
Sample	Total reads sequenced	Number of reads after merge (forward and reverse sequence reads)	Number of retained reads after removal of human+pig+chicken+cow contamination	% of reads mapped against human+pig+chicken+cow genome	Number of retained reads after alignment against endogenous nDNA	% of merged reads mapped against endogenous nDNA	mean coverage of nDNA genome	% Duplicate reads
LYEP9	50569242	38185926	38158803	0.0710	1430570	3.75	0,043x	0.13
LYEP11	44203964	30062441	30033302	0.0969	165224	0.55	0,002x	0.11
LYEP51	47366572	32212912	32204132	0.0273	304451	0.95	0,006x	0.07
LYEP53	40676611	29837937	29834021	0.0131	26131	0.09	0,0005x	0.07
LYEP27	39883611	11615845	11610349	0.0473	71232	0.61	0.0009x	0.25
note: endogenous DNA here is estimated based on the proportion of reads submitted to BWA that mapped without any quality score filtering.								