

Bioinformatic tools in the study of ancient dogs – preliminary results of an Iberian case study

Ludmilla Blaschikoff¹ | Octávio Serra² | Silvia Guimarães³ | JL Cardoso^{4,5,6} | Fernanda Simões² | Cleia Detry⁴ | Catarina Ginja³ | Carlos Fernández-Rodríguez⁷ | Eduardo Ferreira⁸ | Ana Elisabete Pires^{3,9,10}

¹Departamento de Biologia, Universidade de Aveiro, Portugal, ²Unidade de Biotecnologia e Recursos Genéticos, Instituto Nacional de Investigação Agrária e Veterinária, I.P., Portugal, ³CIBIO-InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Portugal, ⁴UNIARQ - Centro de Arqueologia da Universidade de Lisboa, Faculdade de Letras de Lisboa, Universidade de Lisboa, Portugal, ⁵ICArEHB - Interdisciplinary Center for Archaeology and the Evolution of Human Behavior; FCHS; Universidade do Algarve; Campus de Gambelas, Portugal, ⁶Departamento de Ciências Sociais e de Gestão, Universidade Aberta, Portugal, ⁷Departamento de Historia, Facultad de Filosofia y Letras, Universidad de León, España, ⁸Departamento de Biologia & CESAM, Universidade de Aveiro, Portugal, ⁹LARC - Laboratório de Arqueociências, Direção Geral do Património Cultural, Lisboa, Portugal, ¹⁰Faculdade de Medicina Veterinária, Universidade Lusófona de Humanidades e Tecnologias, Portugal.

INTRODUCTION

- Ancestral samples has a crucial role uncovering details of dog domestication and its evolution, which can no longer be recovered by studies based only on modern samples (Larson *et al.* 2012).
- DNA extracted from tissues such as bones and/or teeth present in low quantity and highly degraded (Lindahl 1993).
- Bioinformatic tools has revealed new insights into genomics of ancient specimen (Shendure & Ji 2008).
- Domestic dogs exist in the Iberian Peninsula at least since the Upper Late Paleolithic (Vigne 2005).
- Genetic studies of Iberian extant dogs revealed the existence of four haplogroups A, B, C, D - haplogroup A is the most frequent whereas D is the least represented (Pires *et al.* 2006, 2017).
- Archaeological remains from Iberian dogs dated to the Chalcolithic period already exhibited a large phenotypic variability – indicating that human driven selection acted towards the appearance of different dog types for the performance of various tasks (e.g. hunting, livestock guarding, defense) (Pires *et al.* 2001; Arana & Rodríguez 2013).

AIM

In this study, we attempted to carry out a genomic amplification of ancient DNA from four pre-historic Iberian dogs using the Illumina sequencing method; to assign mtDNA sequences to the major dog haplogroups and to characterize some phenotypic traits of these dogs.

METHODS

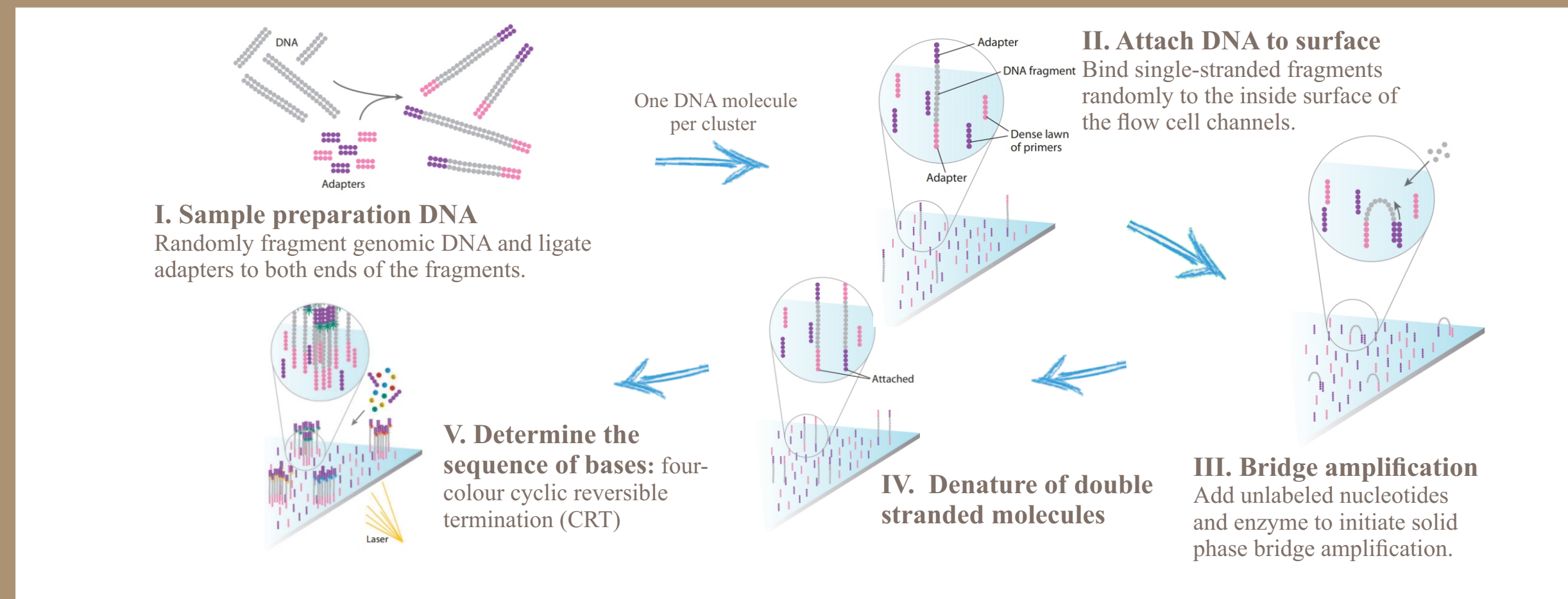


Fig. 3. Representation of the workflow of the second-generation sequencing method followed in this study. Adapted from Mardis (2008).

- We carried out Next-Generation Sequencing - Illumina (Fig. 3), and used the most up-to-date bioinformatic tools (Fig. 4) to recover and analyze mitochondrial and nuclear ancient DNA for 4 Chalcolithic Iberian dogs.
- Consensus sequence from NGS analysis for each Chalcolithic dog sample was used to infer evolutionary relationships by the construction of an haplotype network including reference data (Pires *et al.*, 2006; Randi *et al.*, 2000; Koblmüller *et al.*, 2016; Koepfli *et al.*, 2015; Björnerfeldt *et al.*, 2016).
- Additional *Canis* Iberian mtDNA sequences were used for comparisons: from the same samples previously obtained by the 1st generation NGS method – 454, as well as from a Chalcolithic Iberian wolf; and from extant specimens' dogs and wolves.
- The reference canFam3.1 genome (a female from the Boxer breed) lacks the Y chromosome, therefore sequences of the Y chromosome were retrieved from different studies available at Genbank (KP081776; GQ366706-GQ366731; GQ366741-GQ366770; GQ366790-GQ366793; DQ973626-DQ973805) and for each sample we aligned the reads that didn't map against canFam3.1 genome against those Y fragments. Additionally, we used a read depth based method, comparing the ratio of reads/Mbp over all chromosomes in order to check the read dosage on chromosome X compared with the remaining chromosomes. Females are expected to have a similar ratio and males are expected to present half the ratio, when compared with autosomal chromosomes.

MATERIALS



Fig. 1. 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).

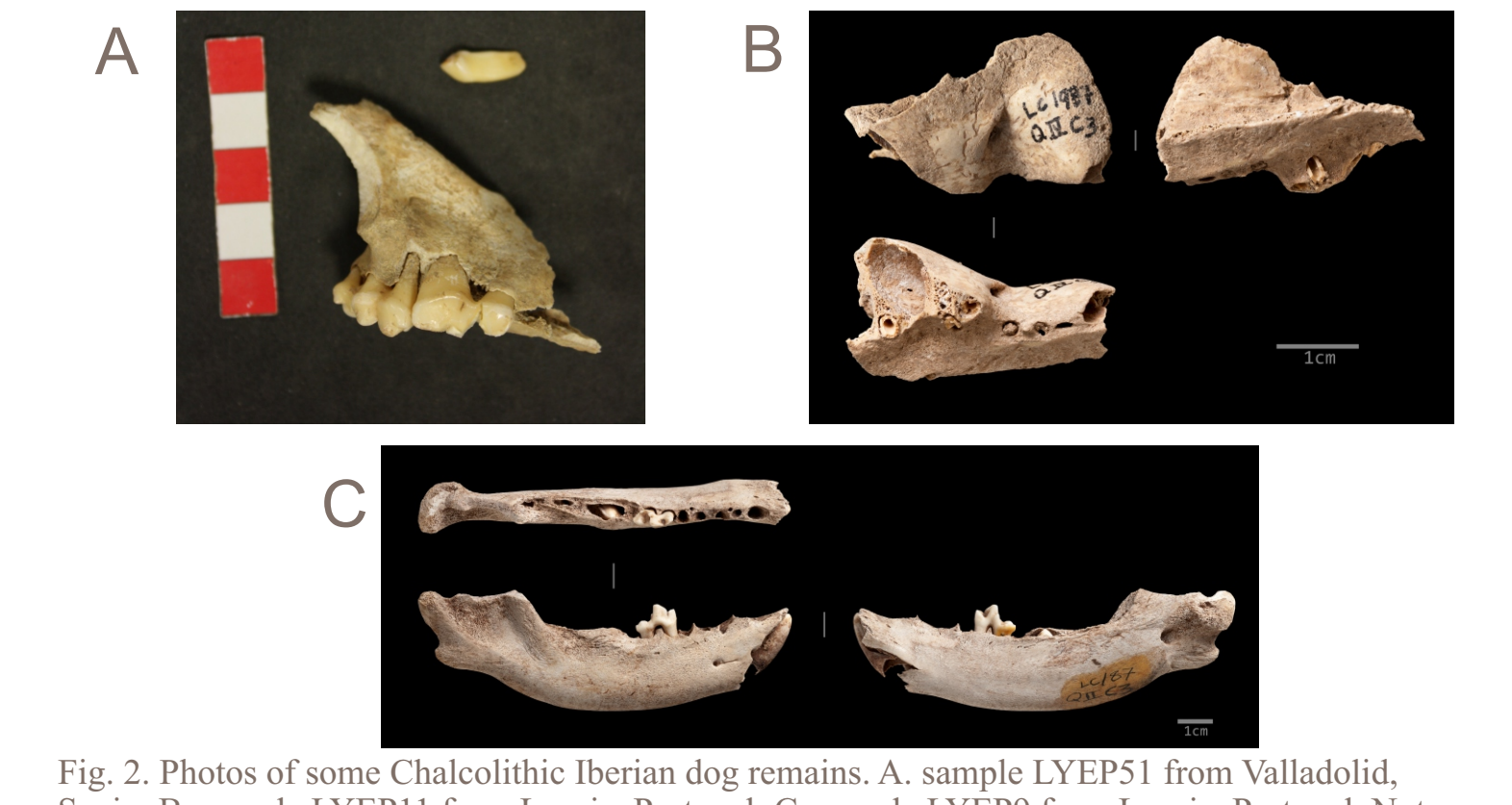


Fig. 2. Photos of some Chalcolithic Iberian dog remains. A: sample LYEP51 from Valladolid, Spain; B: sample LYEP11 from Leceia, Portugal; C: sample LYEP9 from Leceia, Portugal. Note: no picture is available for sample LYEP53 from El Casetón de la Era, Spain, nor LYEP27 (wolf). Photos by Carlos Fernandez-Rodriguez (remains from Spain) and José Paulo Ruas (remains from Portugal).

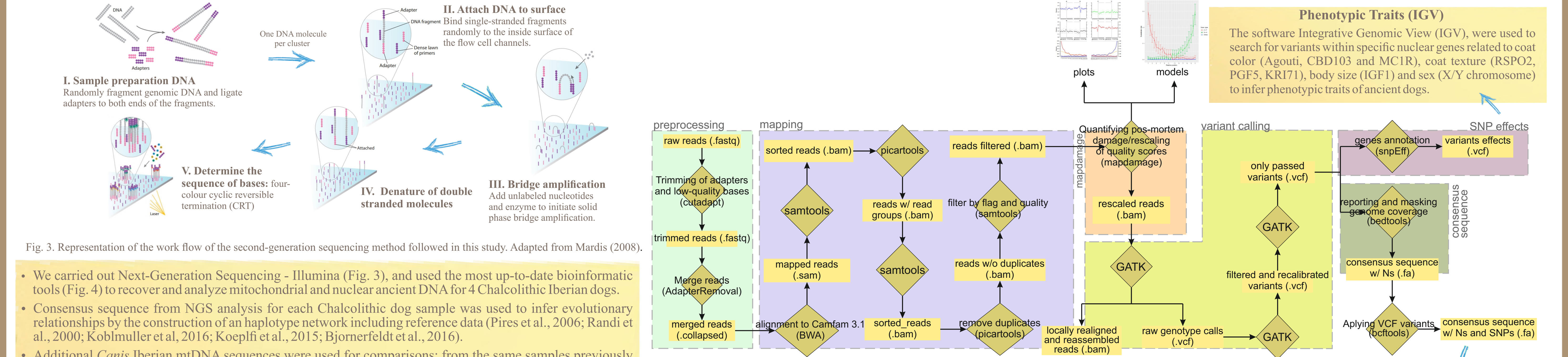


Fig. 4. Schematic representation of the pipeline carried out for sequences alignment and their genotyping.

- Mega v.7 was used to align the four consensus sequences obtained from this process with others 69 *Canis* Iberian haplotypes comprising the d-loop region, available at Genbank [accession numbers: AY706476–524 (Pires *et al.*, 2006); KY014649, KY014656–57, KY014669–70 (unpublished); JX845621–25 (Unpublished); EF380226–29 (unpublished); AF115702–03 (Randi *et al.*, 2000); KU644668, KU644670 (Koblmüller *et al.*, 2016); KT448278 (Koepfli *et al.*, 2015); DQ480505 (Björnerfeldt *et al.*, 2016)].
- A 99 bp region encompassing D-loop position was chosen for analysis as it was available from all data sources.
- DnaSP v.5 was used to collapse identical haplotypes for that 99bp region.
- PopArt was used to construct the Median-Joining (MJ) networks connecting mtDNA haplotypes from ancient and extant *Canis*.

PRELIMINARY RESULTS AND DISCUSSION

- Median-joining network (Fig. 5) depicting phylogenetic relationships among ancient/extant Iberian dogs and wolves show that Chalcolithic dogs carried mtDNA variants that segregate within haplogroups A and C.
- Some mismatches (1-4) between NGS_454 and NGS_Illumina sequences were found at each ancient dog sample. An optimization of the NGS-Illumina analysis parameters is necessary to recover previous haplogroups assignment for ancient samples.
- The Chalcolithic wolf sequence segregate together with extant wolf sequence within the H1 wolf haplogroup. Although few samples were analysed we suspect that, by the Chalcolithic time period, Iberian dogs and wolves were already structured populations based on their mtDNA genetic composition, as happens for current populations.

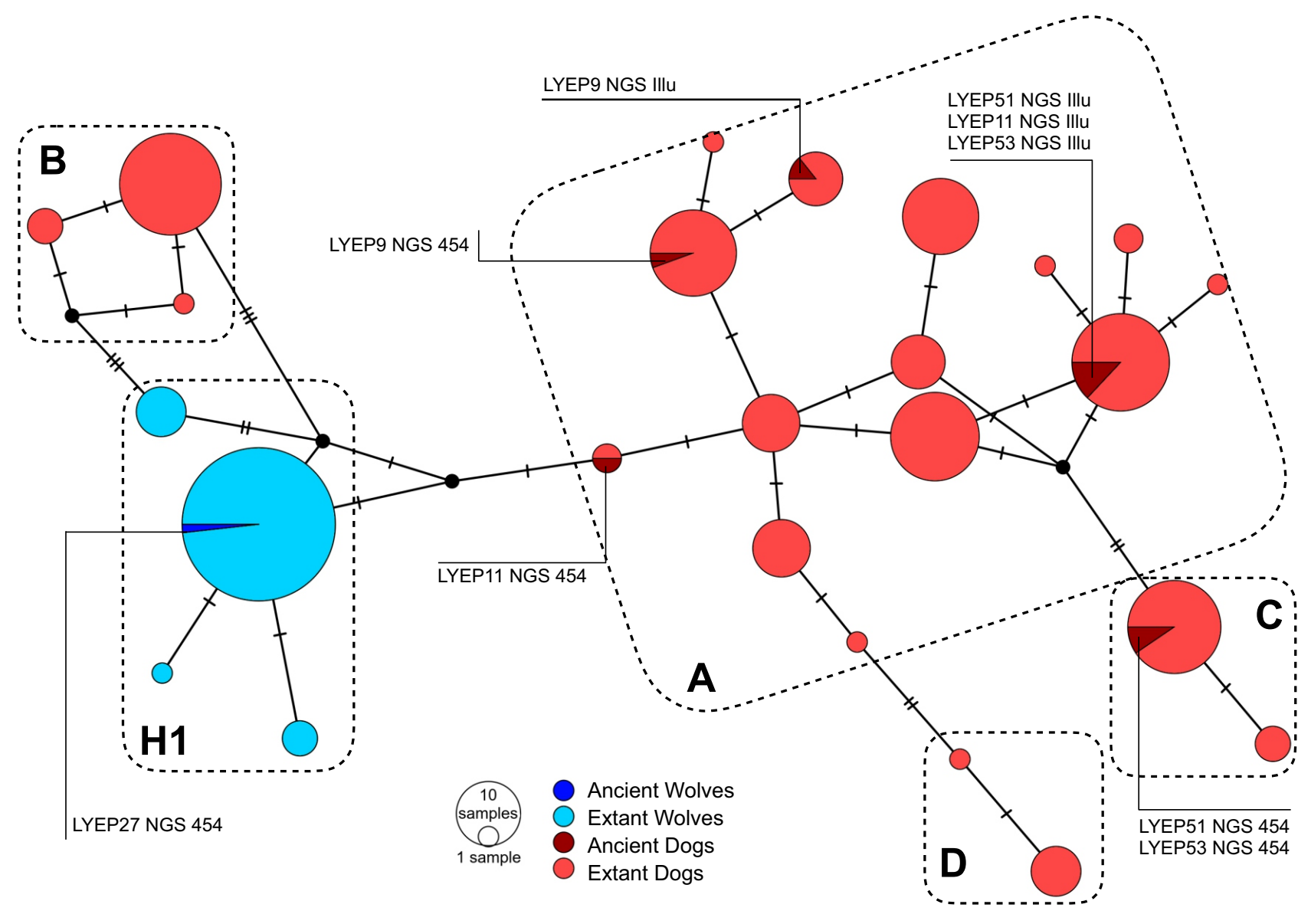


Fig. 5. Median-joining network depicting the relationship between 68 extant and 5 ancient *Canis* Iberian samples, based on a 99 base pairs (bp) long fragment of the mtDNA. Circles size is proportional to its haplotype frequency.

- Due to the endogenous DNA poor preservation few sequences were recovered for nuclear genes related to coat color, coat texture or body size.

- For sample LYEP9, 172 reads aligned against Y chromosome sequences, whereas only 4 and 3 reads aligned with this chromosome for samples LYEP11 and LYEP53, respectively.
- For sample LYEP51, no reads aligned with the Y chromosome.
- Comparing the proportion reads/Mbp (Fig. 6) for each chromosome we observed that:
 - 1) for LYEP9 and LYEP11 only half of the reads/Mbp aligned with the X chromosome, in contrast to the proportion observed for the remaining chromosomes;
 - 2) LYEP51 has the same proportion of reads/Mbp aligning along all the chromosomes, including the X chromosome;
 - 3) LYEP53 data was not conclusive, since only a few sequencing reads aligned with the nuclear genome. This evidence suggests that the LYEP9 and LYEP11 dog remains were males and that LYEP51 was a female.



Fig. 6. Histograms representing the proportions of sequencing reads mapping each chromosome for a) LYEP9, b) LYEP11 and c) LYEP51. Chromosome X is represented by the blue bar.

CONCLUSION

- For Iberian ancient dogs, this is the first attempt to successfully apply NGS methods.
- Ancient DNA analyses and bioinformatic tools are very powerful to unravel genetic variants from past populations.
- Some optimization of the analyses is still necessary to guarantee that the generated consensus sequences are reliable.
- Regarding nuclear genes, this is an ongoing study but the preliminary results for the Chalcolithic dog samples show that DNA is poorly preserved for some genes of interest related to coat color, coat texture and body size.
- In this way, these preliminary results have shown that NGS Illumina applied on ancient DNA has potential to provide data to further investigations in domestication and evolutionary trajectories of dogs.

ACKNOWLEDGMENT

Acknowledgements to Simon Davis and Ana Catarina Sousa. This study is being developed within the Woof project - 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 Competitividade e Internacionalização (POCI-01-0145-FEDER-029545), as well as by the post-doctoral grant SFRH/BPD/112653/2015 (AEPires), SFRH/BPD/108236/2015 (CDetry) and IF/00866/2014 (CGinja).

CONTACTING AUTHORS

Ludmilla Paixão Blaschikoff:
ludmillablaskhoff@ua.pt
Ana Elisabete Pires:
ana.elisabete.pires@gmail.com

REFERENCES

Andrews, S. (2010). Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
 Arana, M.C. & Rodriguez, C.F. (2013). 277-282.
 Bandelt, H. *et al.* (1999). *Molecular Biology and Evolution*, 16(1), 37–48.
 Björnerfeldt, S. *et al.* (2006). *Genome Res.*, 16, 990-994
 Cingolani, P. *et al.* (2012). *Fly (Austin)*, 6, 80–92.
 Jonsson, H. *et al.* (2013). *Bioinformatics*, 29, 1682–1684
 Koblmüller, S. *et al.* (2016). *Journal of Biogeography*, 43(9), 1728-1738
 Koepfli, K. *et al.* (2015). *Current Biology*, 25, 2158-2165
 Kumar, S. *et al.* (2016). *Molecular Biology and Evolution*, 33, 1870-1874
 Larson, G. *et al.* (2012). *Proc. Natl. Acad. Sci.*, 109, 8878-8883.
 Leigh, J.W., Bryant D (2015). *Methods Ecol Evol* 6(9):1110–1116.
 Li, H. & Durbin, R. (2009). *Bioinformatics*, 25, 1754–1760.
 Li, H. *et al.* (2009). *Bioinformatics*, 25, 2078–2079.
 Librado, P. & Rozas, J. 2009. *Bioinformatics*, 25, 1451-1452
 Lindahl, T. (1993). *Nat. Publ. Gr.*, 362, 709-715.
 Mardis, E.R. (2008). *Annu. Rev. Genomics Hum. Genet.*, 9, 387–402.
 Martin, M. (2011). *EMBnet journal*, 17, 10–12.
 McKenna, A. *et al.* (2010). *Genome Res.*, 20, 1297–303
 Quinlan, A.R. & Hall, I.M. (2010). *Bioinformatics*, 26, 841–842.
 Randi, F. *et al.* (2000). *Conservation Biology*, 14(2), 464-473
 Pires, A.E. *et al.* (2017). *Ecol. Evol.*, 7, 4404-4418.
 Pires, A.E. *et al.* (2017). *Quaternary International*, 471, 132-146.
 Pires, A.E. *et al.* (2006). *J. Hered.*, 97, 318-330.
 Pires, F. *et al.* (2001). *Câmara Municipal de Oeiras*, 2, 183-247.
 Savolainen *et al.* (2002). *Science*, 298(5598), 1610-1613
 Schubert, M. *et al.* (2016). *BMC Research Notes*, 9(1), 1-7
 Shendure, J. & Ji, H. (2008). *Nature Biotechnology*, 26(10), 1135-1145
 Vigne, J.-D. (2005). *MUNIBE*, 57, 279-287.