

# The Domestication and Natural History of Horses: a Genomic Approach

**Maddalena Cella**

mc2820@ic.ac.uk

**Main Supervisor:** Dr Matteo Fumagalli, m.fumagalli@imperial.ac.uk

**Co-supervisor:** Prof Vincent Savolainen, v.savolainen@imperial.ac.uk

**keywords:** bioinformatics, genomics, direct to consumer genetics, population genetics, domestication, animal agriculture

## 1 Introduction

Direct-to-consumer genetic (DCTG) tests allow individuals to obtain information regarding ancestry and other traits of interest at a low cost. DTCG is normally used in humans, but has been recently introduced for domestic animals such as cats and dogs (Lyons and Buckley, 2020; Deane-Coe et al., 2018). Animal agriculture is another area that could benefit from DTCG as it allows breeders to have easy access to health and disease information about individuals at a low cost. DTCG normally uses Single Nucleotide Polymorphism (SNP) arrays to target SNP associated with traits of interest (Khan and Mittelman, 2018), however when a good reference panel is available, low coverage Whole Genome Sequencing (WGS) could have a similar performance to SNP arrays. Recent improvements in horse genomics include the release of a new improved horse reference genome assembly (EquCab3.0) (Raudsepp et al., 2019; Kalbfleisch et al., 2018). The high quality of the assembly has allowed the development of SNP arrays and WGS of thousands of individual horses and the identification of many SNPs related to health and performance (Raudsepp et al., 2019). Given the amount genome data available for horses and the interest in them for entertainment, agriculture and the petting industry, they represent the ideal species for testing the use of low-coverage WGS to infer ancestry and phenotype traits of interest. In particular, having at my disposal a hair DNA sample from a mixed-breed horse sequenced at low-coverage, I aim to:

1. predict its ancestry from a reference panel of horses breeds;
2. predict its phenotype for some traits of interest;
3. write a comprehensive pipeline for the workflow used to achieve 1. and 2.

## 2 Methods

I will obtain WGSs from NCBI and ENA to create a reference panel of horse breeds. The DNA sample of a mixed-breed horse sequenced at low-coverage will be mapped against the reference panel to infer ancestry and individual admixture levels. I will also infer the phenotype of the sample by looking at the presence or absence of SNPs associated with some Mendelian traits of interest (Raudsepp

et al., 2019). In order to infer ancestry and phenotype, I will utilise a variety of scripts and programmes already available, as well as develop new tools when necessary. The pipeline developed will be written in *Nextflow*, which ensures reproducibility of the analysis by avoiding numerical instability through multi-scale containerisation (Di Tommaso et al., 2017).

### 3 Anticipated Outputs and Results

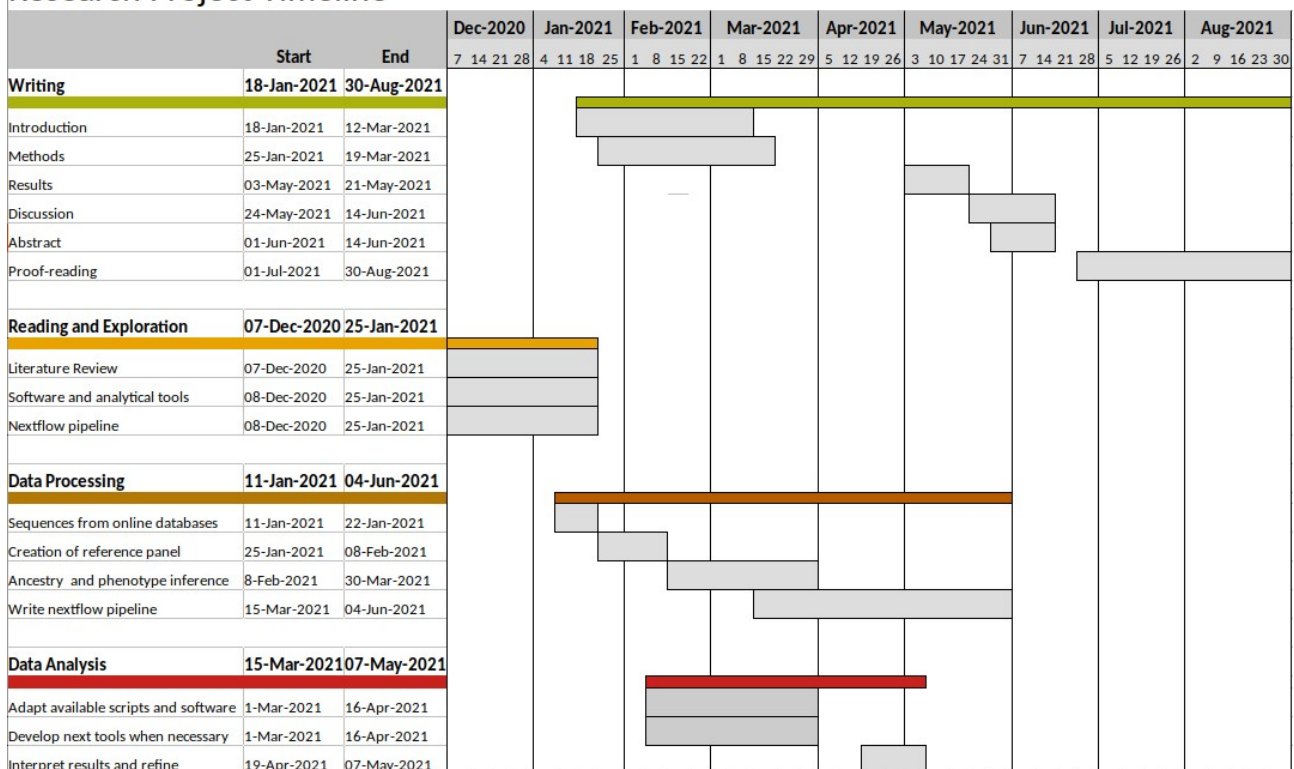
I expect to successfully infer ancestry and admixture levels of the mixed-breed sample as well as its phenotype for some Mendelian traits of interest. I also anticipate the development of a reproducible and comprehensive workflow for the use of low-coverage WGS in replacement of SNP arrays for DCTG.

### 4 Budget

The project as described above does not require additional costs.

### 5 Project Feasibility

Research Project Timeline



## 6 References

- Deane-Coe, P.E., Chu, E.T., Slavney, A., Boyko, A.R. and Sams, A.J., 2018. Direct-to-consumer dna testing of 6,000 dogs reveals 98.6-kb duplication associated with blue eyes and heterochromia in siberian huskies. *Plos genetics*, 14, pp.1–15. Available from: <https://doi.org/10.1371/journal.pgen.1007648>.
- Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E. and Notredame, C., 2017. Nextflow enables reproducible computational workflows. *Nature biotechnology*, 35(4), pp.316–319.
- Kalbfleisch, T., Rice, E., Jr, D., Walenz, B., Hestand, M., Vermeesch, J., O’Connell, B., Fiddes, I., Vershinina, A., Saremi, N., Petersen, J., Finno, C., Bellone, R., McCue, M., Brooks, S., Bailey, E., Orlando, L., Green, R., Miller, D. and MacLeod, J., 2018. Improved reference genome for the domestic horse increases assembly contiguity and composition. *Communications biology*, 1. Available from: <https://doi.org/10.1038/s42003-018-0199-z>.
- Khan, R. and Mittelman, D., 2018. Consumer genomics will change your life, whether you get tested or not. *Genome biology*, 19(1), p.120.
- Lyons, L. and Buckley, R., 2020. Direct-to-consumer genetic testing for domestic cats. *Veterinary clinics of north america: Small animal practice*, 50. Available from: <https://doi.org/10.1016/j.cvsm.2020.05.004>.
- Raudsepp, T., Finno, C., Bellone, R. and Petersen, J., 2019. Ten years of the horse reference genome: insights into equine biology, domestication and population dynamics in the post-genome era. *Animal genetics*, 50(6), pp.569–597.

78 I have seen and approved the proposal and the budget.

79

80 Dr. Matteo Fumagalli, 18/12/2020.

81 