Sequencing *de novo* Y chromosome reference genome of the dog (in progress)

MARIS HINDRIKSON[1], PEDRO SILVA[1], RAQUEL GODINHO[1,2]

[1]CIBIO/InBio – Centro de Investigação em Biodiversidade e Recursos Genéticos, University of Porto, Campus de Vairão, Portugal
[2]FCUP – Department of Biology, Faculty of Sciences, University of Porto, Portugal

Beagle by Nora Rahumägi

Introduction

Y chromosome fulfills several important functions that have been understudied outside of model organisms. Due to inherent assembly difficulties, high repeat content, and large ampliconic regions, mammalian Y chromosomes are rarely included in the genomic analysis [1-3]. More specifically, the lack of a high quality dog Y chromosome genome prevents our complete understanding of canid genome function and evolution.

To facilitate the assembly of such complicated genomic territory, we used unamplified flow sorted Y chromosome, and combined sequencing outputs of two different high throughput technologies following Kuderna et al. [1]. This approach yields a highly continuous assembly of the dog Y chromosome. It constitutes a significant improvement over comparable previous methods, increasing continuity and simplifying the assembly of extremely large and repetitive genomes [1].

Goal

Our goal was to use a novel methodology (modified from Kuderna et al. [1]) to produce a *de novo* dog Y chromosome reference genome.

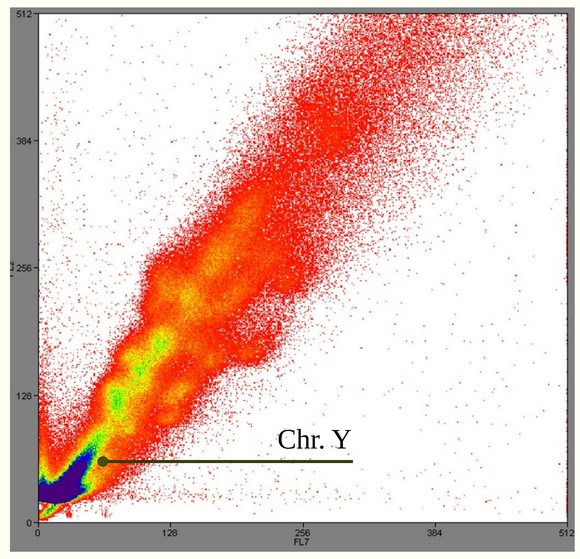

Fig. 1 – Flow-karyogram of a dog genome. The different clusters correspond to different chromosomes. The black circle delimits the cluster corresponding to the Y chromosome used for this project.

Schematic overview of laboratory works and bioinformatic analyses used to sequence *de novo* dog Y chromosome (in progress).

Flow-sorting (Fig. 1)
- 90 000 individual Y chromosomes
- beagle lung fibroblast cell line (AG08624)

Whole-genome Amplification (WGA) (Fig. 2)

Fig. 2 – Whole-genome amplification was performed using Qiagen REPLI-g Single Cell Kit.

Sequencing

(1) **Nanopore MinION** sequencer
- Ligation Sequencing kit SQK-LSK109
- FLO-MIN106D
- **generated 925 Mb of data, 201 469 bp of reads**

(2) **Illumina MiSeq** genome sequencer
- MiSeq Reagent Kit Nano v2
- (2x300 cycles), paired-end reads
- Library: Nextera Library Prep Kit

Assembly
- Initial base-calls and assembly with **Canu** v1.8[4]

Draft assembly statistics:
- **40 x coverage after error correction**
- **2085 contigs**
- **N50 10,3 kb**

Polishing with Nanopolish pipeline
- Re-perform basecalling and indexing (Nanopolish v.0.10.1)[5]

Polishing with Nanopolish pipeline
- Mapping into raw assembly (bwa mem v.0.7.17.)[6]
- Nanopolish to be corrected in chunks of 50kb

Polishing with Illumina
- Self-corrected assembly polished with Illumina reads
- Pilon (v 1.22)[7] and racon (v 1.4.3) [8]

Applications

Despite the whole genome been constructed and annotated for a female dog (CanFam3.1)[9], the dog Y chromosome reference genome is still not available. Reference Y chromosome sequence for dog will prove most valuable in the future in various research areas connected to canids such as (i) importance of paternal lineages in the domestication process of dog, (ii) development of genetic markers used to study male-specific dispersal and population genetics in canids and (iii) evaluation of sex bias during hybridization in natural populations.

REFERENCES

1. Kuderna *et al*. Nat. Comm. 10: 4 (2019).
2. Tomaszkiewicz, M. *et al*. Genome Res. 26: 530-540 (2017)
3. Li, G. *et al*. Genome Res. 23: 1486-1495. (2013).
4. Koren and Walenz *et al*. Genome Res. 27: 722-736 (2017).
5. Nanopolish (https://github.com/jts/nanopolish), 18 August 2019.
6. Li, H. & Durbin, R. Bioinform 25: 1754–1760 (2009).
7. Walker, B. J. *et al*. PLoS ONE (2014).
8. Vaser, R. *et al*. Genome Res. 27: 737–746 (2017).
9. Lindblad-Toh K. *et al*. Nature 438: 803–819 (2005).

ACKNOWLEDGEMENTS

We thank Professor Malcolm-Smith and Dr. Jorge Pereira from Cambridge Resource Centre for Comparative Genomics, University of Cambridge for flow-sorting of dog Y chromosomes.

This work is supported by Estonian Research Council from Personal Research Funding (PUT) of postdoctoral research grant (PUTJD708, awarded for M.H. in 2018). P.S. and R.G worked under research contracts from the Portuguese Foundation for Science and Technology, FCT (projects PTDC/BIA-EVF/2460/2014, PTDC/BIA-EVL/31902/2017 and DL57/2016).