Mitochondrial genomes long read sequencing methodology applied to multispecies for the identification of genetic variants

Sophie Dhorne-Pollet, Caroline Morgenthaler, Delphine Lallias, Aurélien Capitan, Nicolas Pollet, Eric Barrey

MITOCHONDRIAL GENOMES LONG READ SEQUENCING METHODOLOGY APPLIED TO MULTI-SPECIES FOR THE IDENTIFICATION OF GENETIC VARIANTS

SOPHIE DHORNE-POLLET [1], CAROLINE MORGENTHALER [1], DELPHINE LALLIAS [1], AURELIEN CAPITAN [1], NICOLAS POLLET [2] AND ERIC BARREY [1]

[1] Université Paris Saclay; AgroParisTech; INRAE, UMR1313 Génétique Animale et Biologie Intégrative, Jouy-en-Josas, France.
[2] Evolution, Génomes, Comportement & Ecologie, CNRS, IRD, Univ. Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette, France.

INTRODUCTION

Background

In animals, mitochondrial genome variability is poorly studied, besides MT-CO1, MT-CYB, and D-Loop regions used for phylogenetic analyses. However, genetic studies aimed to predict the functional impact and the potential association with phenotypic traits require a better understanding of the whole mitochondrial DNA variations. Long read sequencing should improve the quality of mapping and genetic variants identification.

Objective

This study aimed to develop an optimized workflow for routine high-throughput sequencing of mtDNA based on long read sequencing and applicable regardless of the species analysed.

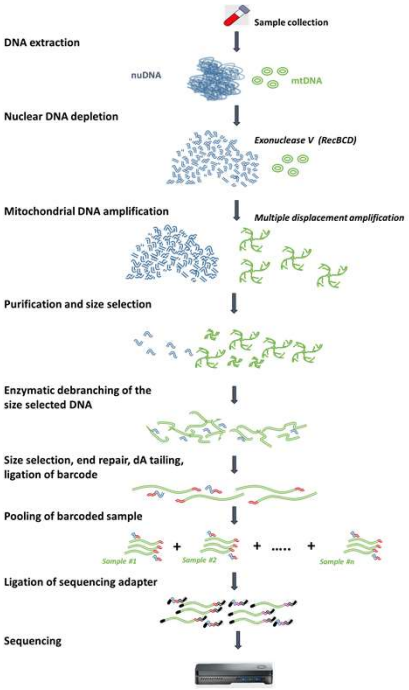
MATERIAL & METHODS

Material

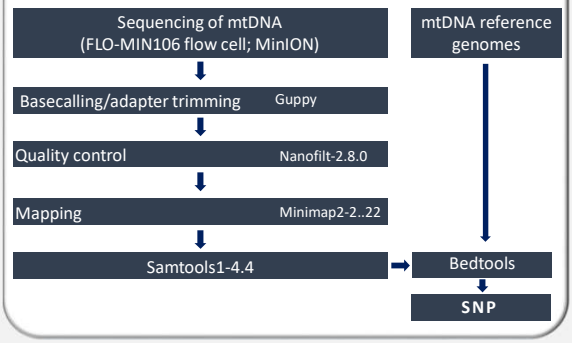
Four vertebrate species were selected to carry out this methodological project: *Equus caballus, Bos taurus, Oncorhynchus mykiss* and *Xenopus tropicalis*.

Long-read sequencing was performed using a MinION device. Sequencing protocols and libraries were previously developed for *Equus caballus* [1; 2] and generalized to the other 3 species analyzed.

Library preparation workflow

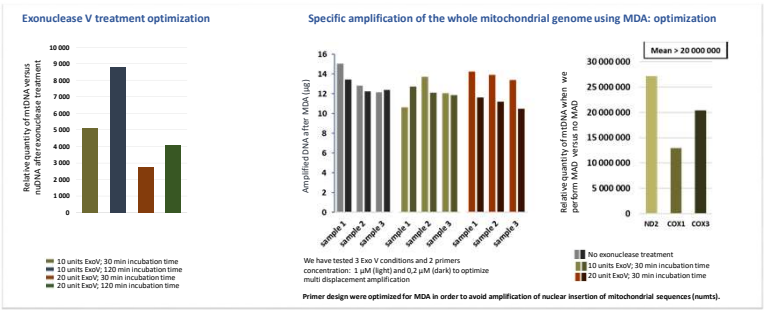


Sequencing and read analysis pipeline

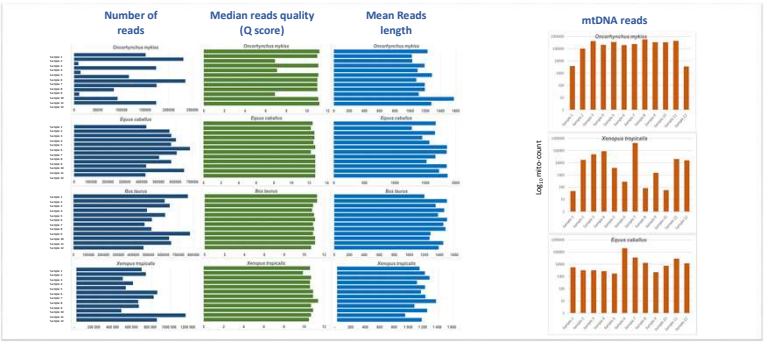


RESULTS

Optimization of mtDNA Enrichment

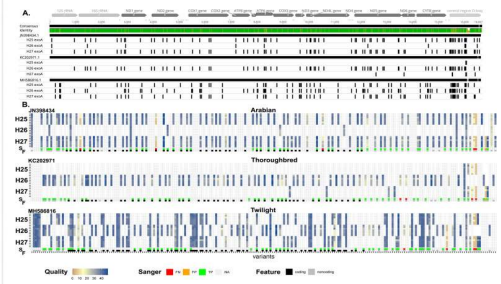


mt-DNA-seq results



Identification of genetic variants

Variant calling from long mtDNA reads: *Equus caballus*



A- The mitochondrial genome with the scale in bp and its rRNA, tRNA, gene and control region features. The three mtDNA reference sequences were aligned to visualize their differences.

B- Heatmap showing all variants identified along the mtDNA for the different samples indicated and the three different reference genomes.

Performance of Long read sequencing, MinION compared to sanger sequencing

Whatever the species, the quality of long read sequencing was good: average length of reads equal to 2.3kb; coverage of 188X to 488X.
We identified SNPs with an accuracy of 98.1%; recall of 85.2% and a F1-score of 0.912.
Long read sequencing improved the quality of the mitochondrial reference genome.

Reads produce by the MinION sequencer that can be analyzed to identify genetic variants

Electropherogram generated by laser scanner from Sanger Sequencing

CONCLUSIONS

The enrichment ratio of mtDNA reads presents variation between samples within each species, but the coverage was deep enough for variant calling. All the SNP obtained by long reads versus short reads were identical. **Long reads sequencing is accurate for genetic variants identification.**

1- Dhorne-Pollet S., Barrey E., Pollet N.. A new method for long-read sequencing of animal mitochondrial genomes: application to the identification of equine mitochondrial DNA variants. BMC Genomics, BioMed Central, 2020, 21 (1).
2- Dhorne-Pollet S, Pollet N, Barrey E. A New Method for Sequencing the Mitochondrial Genome by Using Long Read Technology. Methods Mol Biol. 2021;2277:331-343.