EXTENDED ABSTRACT

Journal of the World Mitochondria Society
Issue n°3, Vol. 3
DOI 10.18143/JWMS_v2i2_1999

Mitochondrial genome variability of 205 Arabian endurance horses

HEURTEAU, Alexandre (a), HOEDE, Claire (b), RICARD, Anne (c), ESQUERRE, Diane (d), MORGENTHALER, Caroline (e), MACH, Nuria (e), ROBERT, Céline (e,f), BARREY, Eric (e)

(a) Laboratoire de Biologie Moléculaire Eucaryote, UMR 5099, Toulouse, France
(b) PF Bioinfo, MIAT, Université de Toulouse, INRA, Castanet-Tolosan, France
(c) GABI, BIGE, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France; IFCE, Recherche et Innovation, 61310 Exmes
(d) GeT PlaGe, Genotoul, INRA Auzeville, Castanet Tolosan, France
(e) GABI, BIGE, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France
(f) Ecole Nationale Vétérinaire d'Alfort, Université Paris-Est, Maisons-Alfort, France

Corresponding author:
Eric Barrey
eric.barrey@inra.fr

Abstract

The objectives of this study were to propose a new strategy to call genetic variants in mitochondrial genome, and determine whether some variants are associated to endurance performance. DNA was isolated from peripheral blood and mtDNA was amplified by 5 overlapping amplicons and then sequenced by Illumina Miseq. We chose a reference genome (GenBank ID : JN398380.1) that we « circularized » to improve terminals' alignments. After quality filtering and alignment, GATK (v3.5) was used to detect SNPs, indels and heteroplasmy. Association between SNPs and performance in endurance was tested using mixed model with fixed SNP Effect and random additive genetic effect with relationship matrix. We got 590 variable positions of which 72% were known through studies of mitochondrial variability in horses (Li *et al.*, McKenna *et al.*). 80% of the protein coding variants are silent and the transition/transversion ratio is 22.5. We observed 1.5% of non-haploid genotype. We graphically described the variants and the observed heteroplasmy in terms of localization, diversity, potential effects along the mitochondrial genome. Our preliminary analysis do not allow to associate some of the variants to performance.

Introduction

Endurance horses can run up to 160 km per race. Mitochondrial DNA variations may affect the efficiency of electron transport chain and ROS production, thus contributing to endurance performance. We studied the mitochondrial genome variability among 205 endurance horses for which racing performance were known.

Material & methods

We designed 5 overlapping amplicons to specifically amplified mitochondrial genome from total DNA extracted in peripheral blood. Then we sequenced all samples with Illumina Miseq that produces 250 bp paired reads. The resulting average coverage of the genome was as huge as 6000X.
We chose a Arabian reference genome (GenBank ID JN398380.1) that we pseudo-circularized in silico to improve terminals' alignments. For this we simply linearise the

Received: 26 October 2016; Accepted: 24 February 2017

reference genome at a different location. This fig. 1 shows the two reference genome used.

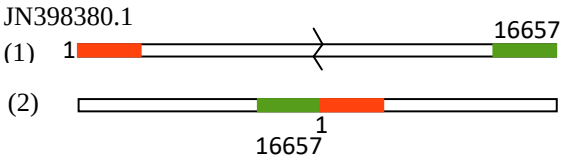

Figure 1. Reference genome 2 is built from GenBank JN398380.1 cut in the middle, then we put the extremities inside and concatenate.

The cleaning and mapping pipeline is standard: we used trimgalore v.0.4.0 (Stringency 5, Phred score < 20), bwa mem v. 0.7.12, samtools view and sort v.1.3 (1), AddReadGroups (picard-tools v1.8). We filtered reads with mapping quality < 60. Then we obtained 2 bam files (one for each reference) for each sample.
We used the GATK 3.5 package (McKenna *et al.*) to detect variations. Most parameters followed the best practices and some were fine-tuned as described in the fig. 2.

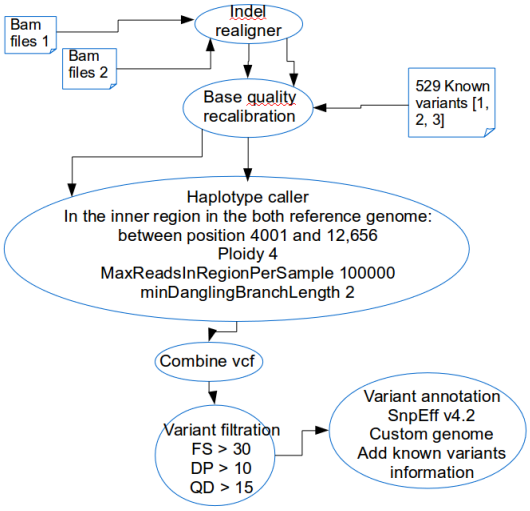

Figure 2. Variant calling pipeline implemented in Jflow (Mariette *et al.*) to launch in parallel all samples.

Association between SNPs and performance in endurance was tested using mixed model with fixed SNP Effect and random additive genetic effect with relationship matrix.

Results

For the 205 horses, we found 590 combined variable positions (one position varies every 27 bps in average), see Fig. 3 A. Only 5 contain indels. 72% have already been found in previous studies (Aken et al., Achilli et al., Cardinali et al.). 80% of the protein coding variants are silent and the transition/transversion ratio is 22.5, see Fig. 3 B. Furthermore, we could observe 1.5% of non-haploid genotype (potential heteroplasmy).

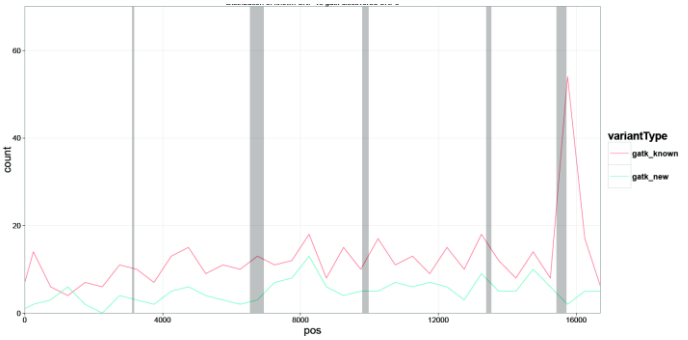

Figure 3A. Known and unknown variants along the mitochondrial genome

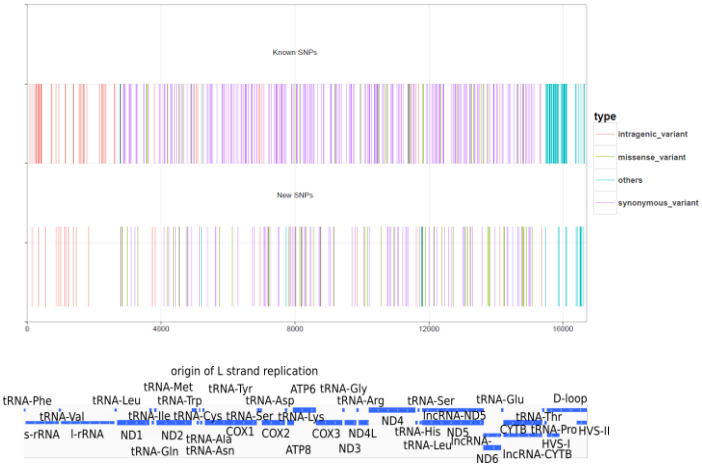

Figure 3B. Variants functional annotation

Figure 3. A: variants detected along the mitochondrial genome. In red : known variants, in blue unknown variants. GATK fails to detect new variants in the D-loop region because it's hypervariable region and the mapping is very difficult. So GATK in most of case does not call variant in this region. **B**: Functional annotation along the mitochondrial genome. In red: variants positioned in gene coding ARN, light blue

Received: 26 October 2016; Accepted: 24 February 2017

variants in non-coding region and purple and green affect genes coding proteins. Purple variants are synonymous, green are non-synonymous. Except in the D-loop control region, we are relatively confident about the quality of SNPs detected.

Preliminary study of statistical analysis can not yet link any variants to performance in endurance racing. Further analysis is still required.

Conclusion

We obtained 590 mitochondrial variants in 205 horses. We highlight 1.5 non-haploid genotypes, but we need to check if it's heteroplasmy or numts (nuclear sequences of mitochondrial origine (Nergadze *et al.*)). Further statistical analysis is still required.

References

Achilli, A., A. Olivieri, P. Soares, H. Lancioni, B. H. Kashani, U. A. Perego, S. G. Nergadze, V. Carossa, M. Santagostino, S. Capomaccio, M. Felicetti, W. Al-Achkar, M. C. T. Penedo, A. Verini-Supplizi, M. Houshmand, S. R. Woodward, O. Semino, M. Silvestrelli, E. Giulotto, L. Pereira, H-J. Bandelt, and A. Torroni. 2012. Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *PNAS*. 109 (7), 2449-2454.

Aken, B.L., S. Ayling, D. Barrell, L. Clarke, V. Curwen, S. Fairley, J. F. Banet, K. Billis, C. G. Girón, T. Hourlier, K. Howe, A. Kähäri, F. Kokocinski, F. J. Martin, D. N. Murphy, R. Nag, M. Ruffier, M. Schuster, Y. A. Tang, J-H. Vogel, S. White, A. Zadissa, P. Flicek, S. M. J. Searle. 2016. The Ensembl gene annotation system. *Database*.

Cardinali, I., H. Lancioni, A. Giontella, M.R. Capodiferro, S. Capomaccio, L. Buttazzoni. 2016. An Overview of Ten Italian Horse Breeds through Mitochondrial DNA. *PLoS ONE*. 11(4).

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25: 2078-9.

Mariette, J., F. Escudié, P. Bardou, I. Nabihoudine, C. Noirot, M-S. Trotard, C. Gaspin, C. Klopp. 2015. Jflow: a workflow management system for web applications. *Bioinformatics*. 32 (3): 456-458.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M.A. DePristo. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*. 20:1297-303.

Nergadze, S.G., M. Lupotto, P. Pellanda, M. Santagostino, V. Vitelli and E. Giulotto. 2010. Mitochondrial DNA insertions in the nuclear horse genome. *Animal Genetics*. 41: 176–185.

Received: 26 October 2016; Accepted: 24 February 2017