Accounting for nuclear- and mito-genome in genetic evaluation and breeding of dairy cattle

Accounting for nuclear- and mito-genome in genetic evaluation and breeding of dairy cattle

G.M. Fortuna[1,2,3*], B.J. Zumbach[2], M. Johnsson[3], I. Pocrnic[1] and G. Gorjanc[1]
[1]The University of Edinburgh, The Roslin Institute, EH25 9RG, Edinburgh, United Kingdom;
[2]Georg-August-Universität Göttingen, Albrecht-Thaer-Weg 3, 37075, Göttingen, Germany;
[3]Swedish University of Agricultural Sciences, Gerda Nilssons väg 2, Uppsala, Sweden;
*g.mafra-fortuna@sms.ed.ac.uk

Abstract
Studies indicate that mito-genome variation impacts phenotypes in a range of species. In dairy cattle up to 5% of phenotypic variation for milk production has been associated to mito-genome variation. Bearing in mind that milk production is a very energy demanding process that inflicts systemic physiological changes, it is logical to expect that it depends on well-functioning mitochondria. Here we evaluated the impact of accounting for nuclear- and mito-genome variation in genetic evaluation and breeding of dairy cattle with the means of stochastic simulations. Results show that accounting for mito-genome variation can increase the accuracy of estimated breeding values by up to 0.04, particularly in females, and consequently increase genetic gain.

Introduction
Mitochondria are involved in many critical cellular processes via proteins encoded in nuclear-genome and mito-genome. While variation in nuclear-genome is driven by mutations, recombinations and combination of parental genomes, variation in mito-genome is driven by mutations that are inherited from dams to their offspring in haploid form without recombination. Dorji *et al.* (2021) recently reported a high level of variation in mito-genomes across globally important cattle breeds, with most of the variation attributed to within-breed component. Considering the critical role of mitochondria and the observed variation, it is expected that mito-genome variation impacts livestock production. Recent and past studies suggest that mito-genome accounts for up to five percent of phenotypic variation for yield traits in cattle (Boettcher *et al.*, 1996; Gibson *et al.*, 1997; Brajković, 2019). Previous studies have evaluated the impact of accounting for mito-genome variation in genetic evaluation in livestock using pedigree information, while this impact has not yet been evaluated with genomic information. Furthermore, given that mito-genome has a different mode of inheritance than nuclear-genome, we need to revise the definition of a breeding value that accounts for the inheritance of each genome in males and females.
The goal of this study was to evaluate the impact of accounting for nuclear-genome and mito-genome variation in estimating breeding values and using different definitions of breeding values and their estimates in a dairy cattle breeding programme with a stochastic simulation.

Materials & Methods

Simulation setting: Dairy cattle breeding scheme was simulated using AlphaSimR R-package (Gaynor *et al.*, 2020). We simulated nuclear-genome and mito-genomo independently. To generate nuclear-genome chromosomal haplotypes, we used the option "CATTLE" in "AlphaSimR" with 10 diploid chromosomes (to reduce computation time) with 10^8 base pairs each, mutation rate of $2.5*10^{-8}$, recombination rate $1*10^{-8}$, and effective population size as in MacLeod *et al.* (2013). We chose 1,000 loci per chromosome as SNP markers and another 1,000 as QTL. For the mito-genome haplotypes we considered 1 haploid chromosome with 16,202 base pairs, mutation rate of $2.5*10^{-7}$, no recombination, and effective population size of

1,000 to obtain a level of diversity that matched values found in literature (Brajković, 2019; Xia *et al.*, 2019; Dorji *et al.*, 2021). We chose all polymorphic loci (on average 1084 across 10 replicates) in mito-genome as SNP markers and as QTL at the same time. We also evaluated a setting where only one locus was a QTL. Both simulated haplotypes were randomly sampled giving rise to nuclear-genomes and mito-genomes of founding individuals. The nuclear-genomes were passed between generation with recombination and combining two parental nuclear-genomes, while mito-genomes were passed only from mothers to their progeny without recombination. We defined one polygenic trait with heritability of 0.3, partitioned between nuclear-genome ($\sigma^2_{a_n} = 0.25$) and mito-genome ($\sigma^2_{a_m} = 0.05$) components in the base population with allele substitution effects of the QTL sampled from a Gaussian distribution that gave rise to targeted genetic variances. The trait was expressed only in cows for each lactation and generated as:

$$y_{ij} = \mu_i + a_{n,j} + a_{m,j} + p_j + e_{ij} \quad (1)$$

where y_{ij} is the phenotype of cow j in lactation i, μ_i is the population mean for lactation i, $a_{n,j}$, the value of nuclear-genome for animal j, $a_{m,j}$ the value of mito-genome for animal j, which was the same as for its mother, maternal grand-mother, etc., p_j the permanent environment effect of animal j ~$\mathrm{N}(0, 0.10)$, and e_{ij} is the environmental effect ~$\mathrm{N}(0, 0.60)$.

Genetic evaluation: We analysed the phenotype data using the generative model (1) with pedigree- and genome-based information accounting for the mito-genome or not. In the pedigree-based model we assumed $\boldsymbol{a}_n \sim \mathrm{N}(\mathbf{0}, \mathbf{A}0.25)$ with $\mathbf{A}$ being pedigree relationship matrix for nuclear-genome and $\boldsymbol{a}_m \sim \mathrm{N}(\mathbf{0}, \mathbf{I}0.05)$ with $\mathbf{I}$ being an identity matrix of dimension equal to the number of different maternal founder lineages. In the genomic-based model we assumed $\boldsymbol{a}_n \sim \mathrm{N}(\mathbf{0}, \mathbf{H}_n 0.25)$ with $\mathbf{H}_n$ being the "single-step" joint pedigree and genomic (SNP marker) relationship matrix for nuclear-genome (Aguilar et al., 2010) and $\boldsymbol{a}_m \sim \mathrm{N}(\mathbf{0}, \mathbf{G}_m 0.05)$ with $\mathbf{G}_m$ being genomic (SNP marker) relationship matrix for mito-genome with dimension equal to the number of distinct mito-genomes, n_m. We calculated it as $\mathbf{G}_m = \mathbf{M}\mathbf{M}^{\mathrm{T}}/k$, where $\mathbf{M}$ is an $n_m \times n_s$ matrix of 0 (ancestral allele) and 1 (mutation) for n_m mito-genomes and its n_s polymorphic loci (we assume we know all these loci by sequencing mtDNA within pedigrees), $k = \sum_{l=1}^{n_s} p_l(1 - p_l)$, and p_l is the frequency of mutations. For both models we firstly assumed that $\mathbf{p} \sim \mathrm{N}(\mathbf{0}, \mathbf{I}0.10)$ and $\mathbf{e} \sim \mathrm{N}(\mathbf{0}, \mathbf{I}0.60)$ during a burn-in phase. Following we estimated the components and used the new estimates as fixed parameters for the remainder of simulation. We fitted the models with the BLUPF90 suite (Misztal et al., 2018).

Breeding value definition: We defined breeding value of individual j in two ways: (i) as nuclear breeding value $a_{n,j}$; or (ii) as the sum of nuclear and mitochondrial breeding value, $a_{n,j} + a_{m,j}$. The definition (ii) is the correct definition for females because they transmit both nuclear-genome and mito-genome to the next generation. However, males potentially express the effect of mito-genome they inherited from their mother, but they only transmit their nuclear genome to the next generation. Hence, for males, the definition (i) is the correct definition.

Breeding Scenarios: We evaluated sixteen scenarios driven by three factors across 10 replicates. The first factor was a breeding scheme: (i) progeny testing-based selection (PT) and (ii) genomic selection (GS). The second factor was a modelling and selection scenario: (a) using a model without mito-genome effect and selecting both females and males on their nuclear breeding values (STANDARD); (b) using a model with mito-genome effect and selecting both females and males on their nuclear breeding values (BASELINE); (c) using a model with mito-

genome effect and selecting both females and males on their nuclear plus mito breeding values (EXTREME); and (d) using a model with mito-genome effect and selecting females on their nuclear plus mito breeding values and selecting males on their nuclear breeding values (OPTIMAL). The third factor was assuming that all or just one polymorphic locus in mito-genome is a QTL. We evaluated all the scenarios with (i) accuracy of genetic evaluation (as correlation between true breeding value and estimated breeding value) in scenarios STANDARD vs BASELINE and (ii) genetic gain in nuclear breeding value over 20 years in scenarios STANDARD, BASELINE, EXTREME, and OPTIMAL.

Results

We present the results separately for the following five categories: (1) heifers, (2) 1st lactation cows (cows1), (3) cows with 2 to 5 lactation records (cows2-5), (4) young bulls, males without progeny; and (5) proven bulls, males that have undertaken a progeny test or have been used as sires. For the GS scenarios, the female categories were expanded to show the difference between genotyped and non-genotyped animals. We present only the scenario where all polymorphic loci in mito-genome were QTL (the other scenario was qualitatively the same).

Table 1. Accuracy of nuclear breeding values for different animal categories in the model with nuclear-genome (Model N) or nuclear-genome and mito-genome (Model N&M) with the pedigree-based model in progeny testing-based selection (PT) or with genomic-based model in genomic selection (GS).

Scenario	Category	Model N	Model N&M	Difference
PT	Heifers	0.47±0.02	0.48±0.01	0.02±0.02
	Cows1	0.61±0.02	0.63±0.01	0.03±0.02
	Cows2-5	0.54±0.04	0.58±0.01	0.04±0.04
	Young bulls	0.37±0.03	0.38±0.03	0.01±0.02
	Proven bulls	0.74±0.01	0.76±0.01	0.02±0.01
GS	Heifers (non-geno.)	0.43±0.01	0.44±0.01	0.01±0.01
	Heifers (geno.)	0.72±0.03	0.75±0.01	0.02±0.02
	Cows1 (non-geno.)	0.58±0.02	0.60±0.00	0.02±0.02
	Cows1 (geno.)	0.76±0.02	0.78±0.01	0.02±0.02
	Cows2-5 (non geno.)	0.52±0.03	0.56±0.00	0.04±0.04
	Cows2-5 (geno.)	0.71±0.03	0.73±0.01	0.02±0.03
	Young bulls (geno.)	0.69±0.02	0.72±0.02	0.03±0.03
	Proven bulls (geno.)	0.78±0.01	0.79±0.01	0.01±0.01

The accuracy of mito breeding values in both PT and GS was close to one for all animal categories (results not shown) which is expected given the lack of recombination. When evaluating the accuracy of the nuclear breeding values (Table 1), accounting for mito-genome variation increased accuracy from 0.01 to 0.04, depending on the animal category. In the PT scenario, the highest increase in accuracy was observed for “Cows2-5” (+0.04). In the GS scenario, the “Cows2-5 (not geno.)” was the category showing the best improvement (+0.04), followed by “Young-bulls (geno.)” (+0.03).

The results presented in Figure 1 show that considering mitochondrial effect when estimating breeding values can increase genetic gain when both males and females are selected on their nuclear breeding values (BASELINE scenarios). When using the OPTIMUM scenario, results are comparable to that observed for the STANDARD scenario. Lastly, the EXTREME scenario seems to induce less gain than all other scenarios in both genomic and progeny-testing settings.

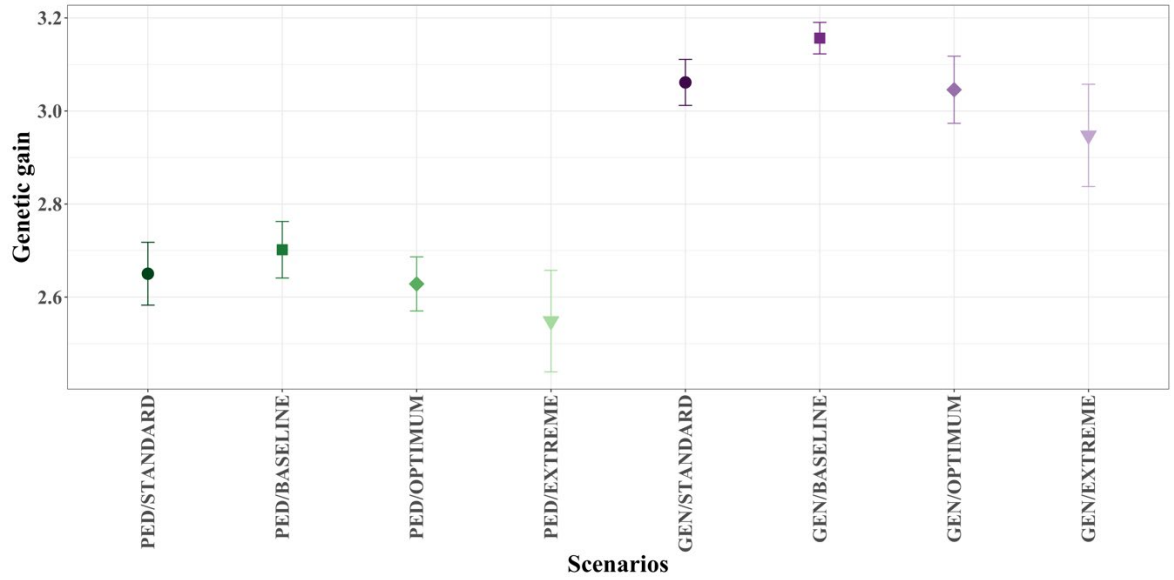

Figure 1. Average genetic gain and 95% confidence interval over 20 years of selection with progeny testing-based selection (PT) and genomic selection (GS) and four different scenarios of modelling and breeding value definition (STANDARD, BASELINE, EXTREME, and OPTIMAL)

Discussion

Dairy breeders are basing their genetic evaluation and selection solely on nuclear breeding values. Gibson et al. (1997), however, defined genetic merit as the sum of nuclear and cytoplasmatic components. Southwood et al. (1989) performed a simulation to evaluate estimation of additive maternal and cytoplasmic variance. They found inflated estimates of additive genetic variance when ignoring cytoplasmic variance or both the cytoplasmic and additive maternal effects in the analysis. Our study complemented this past work by extending it with genomic information in estimating breeding values and by clarifying the definition of breeding values in females and males when both nuclear- and mito-genome affect a trait.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S. *et al*. (2010) J. Dairy Sci. 93(2):743-752. https://doi:10.3168/jds.2009-2730.

Boettcher P.J., Kuhn M.T and Freeman A.E. (1996) J. Dairy Sci. 79(4):663-675. https://doi:10.3168/jds.S0022-0302(96)76412-3.

Brajković V. Impact of mitogenome on milk traits in cattle. Available at: https://core.ac.uk/download/pdf/222819316.pdf (Accessed: 15 April 2021).

Dorji J., Jagt C.V., Chamerlain A., Cocks B., MacLeod I. and Daetwyler H. (2021) BioRxiv. https://doi:10.21203/rs.3.rs-957964/v1.

Gaynor R.C., Gorjanc G. and Hickey J.M. (2020) G3 11(2). https://doi.org/10.1093/g3journal/jkaa017.

Gibson J.P., Freeman A.E. and Boettcher P.J. (1997) Livest. Prod. Sci. 47(2):115–124. https://doi:10.1016/S0301-6226(96)00023-1.

MacLeod I.M., Larkin D.M., Lewin H.A., Hayes B.J., Goddard M.E. (2013) Mol. Biol. 30(9):2209-2223. https://doi.org/10.1093/molbev/mst125.

Misztal I., Tsuruta S., Lourenco D. A. L., Masuda Y., Aguilar I. *et al*. Manual for BLUPF90 family programs. Available at: http://nce.ads.uga.edu/wiki/doku.php?id=documentation

Southwood O.I., Kennedy B.W., Meyer K. and Gibson J.P. (1989) J. Dairy Sci. 72(11):3006-3012. https:// doi:10.3168/jds.S0022-0302(89)79453-4.

Xia, X., Qu, K., Jia, P., Chen, Q. *et al.* (2019) Animals, 9(9):641. doi:10.3390/ani9090641.