EUROPEAN MASTER IN ANIMAL BREEDING AND GENETICS

Evaluation of the impact of mitochondrial variation in the estimation of breeding values for dairy cattle

Gabriela Mafra Fortuna

June, 2021

Main supervisor: Dr Martin Johnsson (SLU)

Co-supervisor(s): Dr Birgit Zumbach (UGOE)

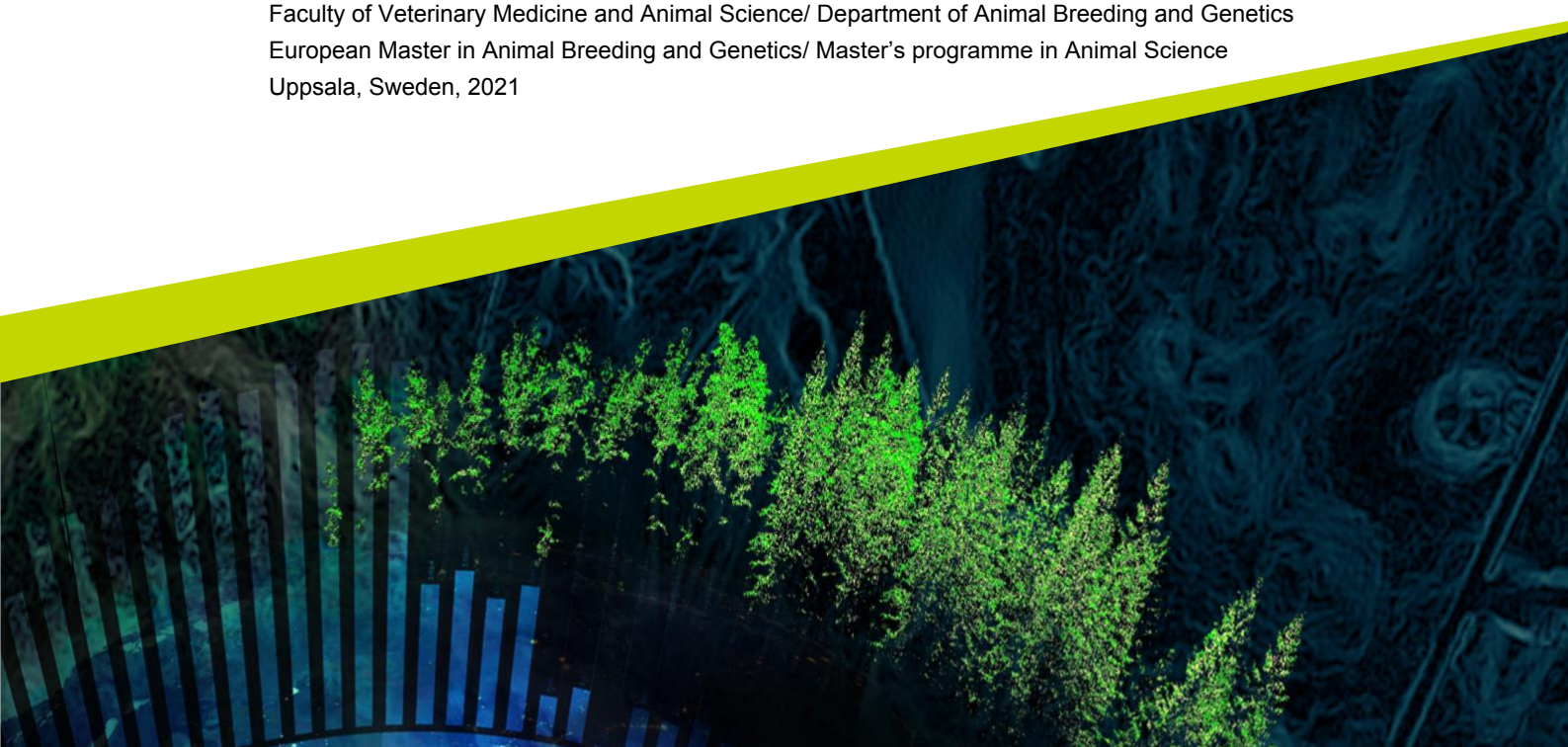
Dr Gregor Gorjanc (UoE, The Roslin Institute)

Co-funded by the
Erasmus+ Programme
of the European Union

Evaluation of the impact of mitochondrial variation in the estimation of breeding values for dairy cattle

Gabriela Mafra Fortuna

Independent project in Animal Science, EX0870 • 30 hp
Swedish University of Agricultural Sciences, SLU
Faculty of Veterinary Medicine and Animal Science/ Department of Animal Breeding and Genetics
European Master in Animal Breeding and Genetics/ Master's programme in Animal Science
Uppsala, Sweden, 2021



Evaluation of the impact of mitochondrial variation in the estimation of breeding values for dairy cattle

Gabriela Mafra Fortuna

Supervisor:	**Martin Johnsson, PhD, SLU, Department of Animal Breeding and Genetics, Uppsala, Sweden.**
Supervisor:	Birgit Jutta Zumbach, PhD, UGOE, Department of Animal Breeding and Genetics, Göttingen, Germany.
Supervisor:	Gregor Gorjanc, PhD, University of Edinburgh, The Roslin Institute, Edinburgh, United Kingdom.
Examiner:	Susanne Eriksson, PhD, SLU, Department of Animal Breeding and Genetics, Uppsala, Sweden.

Credits:	30 hp
Level:	Second cycle
Course title:	**Independent project in animal science**
Course code:	**EX0870**
Programme/education:	**European Master in Animal Breeding and Genetics/ Master's programme in Animal Science**
Course coordinating dept:	**Department of Animal Breeding and Genetics**
Place of publication:	**Uppsala, Sweden**
Year of publication:	**2021**
Keywords:	**mitogenome, variance components, milk yield, simulation**

Swedish University of Agricultural Sciences
Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics

Publishing and archiving

Approved students' theses at SLU are published electronically. As a student, you have the copyright to your own work and need to approve the electronic publishing. If you check the box for **YES**, the full text (pdf file) and metadata will be visible and searchable online. If you check the box for **NO**, only the metadata and the abstract will be visible and searchable online. Nevertheless, when the document is uploaded it will still be archived as a digital file.
If you are more than one author you all need to agree on a decision. Read about SLU's publishing agreement here: https://www.slu.se/en/subweb/library/publish-and-analyse/register-and-publish/agreement-for-publishing/.

☒ YES, I/we hereby give permission to publish the present thesis in accordance with the SLU agreement regarding the transfer of the right to publish a work.

☐ NO, I/we do not give permission to publish the present work. The work will still be archived and its metadata and abstract will be visible and searchable.

Abstract

Mitochondria are independent cellular components responsible for cellular respiration. Through oxidative phosphorylation they convert Adenosine diphosphate and inorganic phosphate into Adenosine Triphosphate, ATP, the essential molecule sourced by all intracellular metabolic processes. As a cytoplasmic component, mitochondria are transferred to offspring in a uniparental fashion. The combination of evolutionary events generated a compact, haploid, non-recombining and significantly conserved mitochondrial genome across mammalian species. In cattle for instance, it is composed by around 16 kbp presented in a circular double-strand molecule that encodes for 22 tRNAs, 2 rRNAs and 13 protein-coding genes linked to energy production. It also presents a regulatory non-coding region known as D-loop. For a while, mitochondrial genetic variation was considered under neutral equilibrium. However, an increasing number of studies are connecting mitochondrial polymorphisms to variability in phenotypical expression in many species. Nonetheless, mitochondrial DNA has been recurrently identified as source of phenotypic variation for production traits in dairy cattle. Reports indicate that up to 5% of the phenotypic variation for such traits is regarded to the mitogenome. The real impact of the mentioned findings on breeding practices are yet unknown. Reflecting that 5% is a significant share on phenotypic variation, especially when comparing the length of nuclear and mitochondrial genomes, this project was performed on an attempt to clarify whether mitochondrial effect should be accounted for in the estimation of breeding values for dairy cattle. Considering that the genetic merit is the sum of nuclear and cytoplasmic components, a dairy cattle breeding scheme selecting for one polygenic trait with multiple observations was simulated. Using the R package "AlphaSimR" both nuclear and mitochondrial genomes were obtained from a coalescent simulation and used to simulate breeding activities. Breeding values were estimated under four scenarios: (1) standard repeatability model based on progeny testing; (2) a repeatability model including mitochondrial effect as random effect and based on progeny testing; (3) standard single-step GBLUP (ssGBLUP); and (4) ssGBLUP including mitochondrial effect. Two scenarios were also tested regarding the number of causative loci in the mitochondrial DNA: (1) all segregating sites were causal; (2) only one segregating site was causal. The project highlighted discrepancies between published data and simulated inferences of mitochondrial diversity, indicating that further investigation of the population genetics of the mitochondria is necessary. Results indicate an advantage of accounting for mitochondrial effect on the estimation of breeding values for female dairy cattle, although no impact on genetic gain was observed. Including mitochondrial effect in breeding value estimations may be most beneficial for the selection of females to be used for in-vitro fertilization or embryo-transfer techniques.

Keywords: mitogenome, variance components, milk yield, simulation

Popular scientific summary

Mitochondria are cellular components responsible for cellular respiration. These components produce energy molecules, fundamental for all cellular activities and the maintenance of life. They contain an independent genome, smaller than that found in the nucleus and the transmission of mitochondria across generations happens exclusively from mother to offspring. For a while, mitochondrial genetic variation was thought to be neutral, when the appearance of new mutations does not affect the function or expression of the genome. However, an increasing number of studies have been connecting mitochondrial genetic variation to variability in the expression of different traits in many species. In dairy cattle, reports indicate that the mitochondrial genome is associated with up to 5% of the variation in production traits.

The impact of the mitochondrial genetic variation on the selection and genetic improvement of dairy cattle populations are yet unknown. For milk yield, 30% of the variation in production between cows is regarded to differences in their DNA. Reflecting that the mitochondrial DNA comprises a much smaller number of genes in comparison to nuclear DNA, 5% of the phenotypic variation is a large share. Thus, this project aimed to clarify whether the effect of mitochondrial variation should be accounted for in the estimation of breeding values for dairy cattle. Using a simulation study allowed exploring different scenarios to better understand the questions the study aimed to answer.

For the study a dairy cattle breeding scheme selecting for one trait influenced by many genes (milk yield) and with multiple observations (lactations) was simulated. Using computational tools, I simulated both nuclear and mitochondrial genomes and used them to simulate breeding activities as mating and selection. Four breeding scenarios were tested considering two breeding schemes based on progeny testing and two breeding schemes based on genome testing. For each pair of breeding schemes, two models for estimating breeding values were applied allowing the comparison between a standard model commonly used by breeders with a model accounting for the mitochondrial effect.

To account for the uncertainty regarding how many genes in the mitochondrial DNA influence milk yield, two trait scenarios were tested. The first considered that all variant sites in the mitochondrial DNA influenced the evaluated trait, while the second considered only one site having such influence.

Results indicate an advantage of accounting for mitochondrial effect in the estimation of breeding values for dairy cows, although no impact on genetic gain was observed. Including mitochondrial effect in breeding value estimations may be most beneficial for the selection of females to be used for in-vitro fertilization or embryo-transfer techniques. The project highlighted discrepancies between published data and simulated inferences of mitochondrial diversity, indicating that further investigation of the population genetics of the mitochondria is necessary.

Table of contents

List of tables

List of figures

Abbreviations

r_a	Accuracy
σ^2_a	Additive genetic variance
ADP	Adenosine Diphosphate
ATP	Adenosine Triphosphate
DNA	Desoxyribonucleic acid
N_e	Effective population size
Ebv	Estimated breeding value
ΔG	Genetic gain
mtGBLUP	Genome testing scheme with mitochondrial effect model
stdGBLUP	Genome testing scheme with standard repeatability model
UGOE	Georg-August-Universität Göttingen
G	Giga
h^2	Heritability
$histN_e$	Historical effective population size
histGen	Historical generation
Kbp	Kilo base pairs
mtDNA	Mitochondrial DNA
σ^2_m	Mitochondrial variance
N_f	Number of females
N_m	Number of males
σ^2_p	Phenotypic variance
mtPBLUP	Progeny testing scheme with mitochondrial effect model
stdPBLUP	Progeny testing with standard repeatability model
QTL	Quantitative trait loci
ROS	Reactive Oxygen species
rRNA	Ribosomal – Ribonucleic acid
i	Selection intensity
SNP	Single nucleotide polymorphism
SNV	Single Nucleotide Variant

ssGBLUP	single-step Genomic Best Linear Unbiased Prediction
SLU	Swedish University of Agricultural Sciences
tRNA	Transfer - Ribonucleic acid
Tbv	True breeding value

1. Introduction

1.1. Dairy Cattle Breeding

The Food and Agriculture Organization of the United Nations (FAO 2020) states that 852 million tonnes of milk were produced worldwide in 2019. The leading producers increased their output compared to previous years due to improvements in the efficiency of large-scale farms and feed availability and quality. An increase in demand caused by growing urbanisation also contributed to the overall rise.

For over a hundred years, breeders have been improving dairy cattle populations focusing on productivity, quality, reproduction, and, more recently, health. First, with the establishment of herdbooks and milk-recording systems, dairy breeding has taken advantage of theories and technologies implemented through quantitative and population genetics (Weigel et al. 2017). The significant expansion of reproductive and progeny testing techniques and the implementation of linear mixed models to optimise accurate selection using pedigree and performance data represented the first great revolution in the field.

After a disappointing experience with marker-assisted selection caused by the failure in identifying single quantitative trait loci (QTL) with significant effects associated with milk yield (Weigel et al. 2017), the implementation of genomic selection in the early 2000s revolutionised dairy breeding once again. With accessible genotyping platforms for single nucleotide polymorphism (SNP) markers, genome selection methods and algorithms were developed and applied to dairy cattle breeding right away.

In a seven-year evaluation of the impact of genomic selection on dairy breeding, García-Ruiz et al. (2016) showed how the technology led to rapid genetic improvement, especially for low-heritability traits. In yield traits, they estimated an increase in genetic gain per year from 50 to 100%. Despite undeniable advances, a genetic improvement on the female selection pathways has been harder to achieve (García-Ruiz et al. 2016). Due to the quick dissemination of improved genetics allowed by artificial insemination, dairy breeders have focused on improving elite

bulls by comparing the performance of their daughters. Thus, many females were needed to provide essential data for accurate prediction of bulls leading to low selection pressure on the category.

Moreover, physiological barriers restrict the success of reducing generation interval in the female categories, a method that accelerated genetic gain on male populations. In the American Holstein herds, the generation interval for the Dam of cows category was reduced only from 4.2 years to 3.6 years over 25 years. On the other hand, it took only five years to see a reduction of 25 to 50% on the generation interval for Sire of bulls after the introduction of genomic selection (García-Ruiz et al. 2016). The combination of physiology, selection pressure, and the lower accuracy for estimation of breeding values penalises the female selection pathways' genetic progress over time.

A higher advantage of genomic selection can be achieved by associating it with female reproductive technologies, such as embryo transfer. Producing genomically proven embryos can effectively increase genetic gain and reduce costs on the production system (Mrode et al. 2018). However, a more extensive diffusion of female reproductive technologies require an increase in selection intensity on the female pathways (Bouquet & Juga 2013). Thus, a greater focus on selecting females will be needed, and methodologies to benefit such selection must be put in place.

A possible method for improving the selection of females is by incorporating cytoplasmic components of variance into the animal model for the estimation of breeding values. In cattle breeding, cytoplasmic components are the thought mechanism behind mitochondrial effects. They are expected to justify certain cow families being praised for delivering higher-performance offspring (Boettcher et al. 1996b; Gibson et al. 1997). According to Boettcher et al. (1996b), estimates of heritability obtained from daughter-dam regressions are higher than those obtained from paternal half-sib correlations, supporting the theory that mitochondrial effects will have an impact on the estimation of breeding values.

1.2. Mitochondria overview

Mitochondria are considered the powerhouse of the cell due to their function on oxidative phosphorylation. They constitute an essential and independent cellular component (McBride et al. 2006). Formed by an inner and outer membrane, this organelle contains its own genetic material and can divide independently from the cell where it is expressed (Alberts 2002). The mitochondrial outer membrane allows for the passage of ions and small proteins. In contrast, its inner membrane, less permeable, is populated by proteins involved in the organelle's primary function:

energy production. Through oxidative phosphorylation, electrons are produced via the citric acid cycle in the mitochondria matrix passing through the protein complexes in the inner membrane and exchanging oxygen to form water. Adenosine Diphosphate (ADP) and inorganic phosphate are then converted into Adenosine Triphosphate (ATP). ATP is the essential energy molecule sourced by all intracellular metabolic processes (Alberts 2002). As a highly dynamic structure, the mitochondria function in association with the cell's nucleus throughout the exchange of regulatory molecules. Such interaction is also specific according to the tissue of expression. Other than ATP synthesis, evidence associates mitochondria with cell signalling cascades and links it to the regulation of metabolism, cell-cycle, development, and programmed cell-death (McBride et al. 2006).

Evolutionary studies estimate the emergence of mitochondria occurring around 2.5 billion years ago (Mishmar et al. 2019). It is thought to have occurred from an endosymbiotic association between an ancestor eukaryotic host and a prokaryotic organism. The ability of the prokaryote organism to efficiently generate ATP via aerobic respiration brought benefits to the host. Thus, the ancestor eukaryote host kept mitochondria (Roger et al. 2017). From there on, the endosymbiont lost autonomy while being integrated into the host through many biological and molecular processes. Among those, the gene transfer from the ancestral mitochondria to the host nucleus was crucial to allow its incorporation as the eukaryote cell's specialised structure we know today (Roger et al. 2017).

Those evolutionary processes led to the conservation of the mitochondrial genome across mammalian species (Gissi et al. 2008). The mitochondrial DNA (mtDNA) is composed of around 16Kbp presented as a circular double-stranded molecule. In cattle, the mtDNA encodes 22 tRNAs, two rRNAs, 13 protein-coding genes from the electron transfer chain (energy producer) and shows a regulatory non-coding region (D-loop) (Srirattana et al. 2017). Despite its smaller size compared to the nuclear genome, mtDNA appears in around 1000 copies over the mitochondria in each cell (Schon et al. 2020).

1.3. Mitogenome particularities and evolution

The unusual emergence of mitochondrial DNA caused its genetic architecture to express some peculiarities, including the model of inheritance, mutation, and recombination rate.

1.3.1. Uniparental inheritance

The transmission of the mitochondrial DNA occurs in a non-mendelian fashion, derived exclusively from the oocyte as a haploid molecule, on a pattern named maternal inheritance (Sato & Sato 2013; Roger et al. 2017). It is believed in the scientific community that such a transmission mechanism is in place protecting the mitogenome from the spread of selfish genes throughout the population. Another hypothesis is that uniparental inheritance happens to prevent heteroplasmy. Heteroplasmy is the presence of multiple mitogenome variants in a single cell or tissue (Ladoukakis & Zouros 2017). Even with the paternal mitochondria's capacity to enter the oocyte in some mammalian species (Sharma & Sampath 2019), it never reaches the progeny. The mechanisms controlling the maintenance of maternal and paternal mitochondria in the oocyte are variable. In cattle, at the initial stages of pre-implementation development of fertilised cells, the sperm-mitochondria is subject to ubiquitination and degradation (Sato & Sato 2013). On the mitogenome, specific clustered sequences, that are inherited together - haplotypes, determine maternal lineages. Maternal lineages allow for tracing mitochondrial lineages back to founders and constructing the population's evolutionary history.

1.3.2. Mutation rate

The reported mitochondrial mutation rate is at least tenfold higher than the nuclear one (Allio et al. 2017). The environment can partially explain such a high occurrence of mutations in the mitochondria. Oxidative phosphorylation releases reactive oxygen species (ROS) in the mitochondria during the process of ATP synthesis. Those free radicals condition the intra-cellular environment as highly reactive, and the lack of histones constraining the mtDNA may contribute to its vulnerability to ROS (Jobling & Jobling 2013). The presence of such compounds then leads to oxidative damage of mitochondrial membranes and proteins, inducing mutations (Jobling & Jobling 2013; Sharma & Sampath 2019). The lack of repair mechanisms is another hypothesis, not excluding the first, for the mitogenome's high mutation rate. During the incorporation of the prokaryotic cell by the eukaryotic host, the organelle lost many house-keeping and regulatory genes, having its DNA down-sized. Most of its ancestral repair coding-genes might have been transferred to the nucleus, leaving the mitochondria without a direct way of controlling the emergence of new mutations (Roger et al. 2017).

1.3.3. Recombination

Regarding recombination, the general agreement is that it does not occur in the mitochondrial DNA. Although evidence indicates that animal mitochondria possess the necessary enzymatic apparatus (Ladoukakis & Zouros 2017), natural

recombination taking place in-vivo is hard to access. The lack of recombination makes the mitogenome vulnerable to the accumulation of deleterious mutations (Hill 2020). Muller's Ratchet explains a synergism between mutation and genetic drift leading to the genomic decay of populations spreading via asexual reproduction (Felsenstein 1974). According to Muller (1964), in the absence of recombination, the offspring carries the same mutations found on the parent, and new mutations are added on and transmitted to the next generation. Such a process should then cause the collapse of the population due to the accumulation of deleterious mutations.

1.3.4. Compensatory coevolution theory

Such unusual genomic structure indicates that mutation erosion should lead to the collapse of the mitogenome over time (Ladoukakis & Zouros 2017; Hill 2020). The referred meltdown, however, has never been observed (Hill 2020). The critical role to cell function held by the mitochondria does not align with its genetic makeup, suggesting that some level of control must be in place to secure that deleterious mutations will not compromise life maintenance. Purifying selection, for example, is known to occur every generation, removing harmful functional mutations from the mitogenome. Purifying selection only partially explains how the organelle survives mutational erosion, but it does not adequately clarify how the organelle deals with the accumulation of mutations.

Compensatory coevolution is one hypothesis to explain the mitogenome stability despite the accumulation of mutations. The compensatory coevolution hypothesis is based on mitochondrial and nuclear products' functional interactions to enable aerobic respiration and core energy production (Hill 2020). Reports indicate critical functional interactions between the nucleus and mitochondria happening at the mitoribosome. Those involve protein-protein interactions, nuclear-encoded aminoacyl tRNA synthases and mitochondrial encoded tRNAs, and mitochondrial encoded tRNAs and nuclear-encoded proteins (Hill 2020).

Alongside, evidence shows the existence of extensive interactions between nuclear and mitochondrial products in signalling between the mitochondria and nucleus. Hill (2020) reminds us that for any of those interactions, changes in the mitogenome sequence could affect its functional interaction with nuclear genes, causing disruptions in cell respiration.

Compensatory coevolution then theorises that the nuclear genome's evolution might compensate for the emergence of deleterious alleles in the mtDNA. In species where the mitogenome shows a high mutation rate, the rate of amino acid sequence divergence in nuclear-mitochondrial genes is accelerated but not in regular nuclear

ones (Hill 2020). The existence of nuclear protein-coding genes that directly target the mitochondrion, performing close functional association with mtDNA products indicates the opportunity for nucleus and mitochondria to engage in compensatory coevolution.

1.4. Mitogenome variation

For a while, mitochondrial genetic variation was considered neutral (Dobler et al. 2014). The neutral theory of evolution considers the variations on the genomic level as a cause of the random genetic drift of neutral mutant alleles. Because mutant variants hold the same fitness values as their wild-type counterparts, the molecule's evolution is in equilibrium.

Nevertheless, that view has changed throughout the years due to studies connecting polymorphisms of the mitogenome to variability in phenotypical expression in many species. Evidence indicates that coevolution between nuclear and mitogenome maintains favourable mtDNA variation. Moreover, positive selection within the mitochondrion and changes in fitness due to polymorphisms (Dobler et al. 2014) shed new light on understanding mitochondria evolution.

Since the mitochondrion is responsible for critical cellular functions, any mitochondrial variation is subject to severe selective pressure. Mechanisms underlying such control involve the mitogenome itself and the nuclear genome (Hill 2020). Selective mechanisms are responsible for the removal of deleterious mutations (negative selection) and the adaptation of cells to new physiological conditions (positive selection) (Shtolz & Mishmar, 2019).

Mitochondrial DNA can vary across individuals and cells and tissues in the same organism (Shtolz & Mishmar, 2019), a state called heteroplasmy. Heteroplasmy in mammals is caused mainly by variation in copy number of a repetitive sequence; a polymorphism thought to occur due to replication slippage between generations. An intergenerational genetic bottleneck occurs before the oocyte maturation to avoid heteroplasmy (Jobling & Jobling 2013; Goodwin et al. 2016). The bottleneck allows only a few mtDNA to be passed on to the offspring, reducing the chances of a mixture of wild-type and mutant molecules cohabiting (Goodwin et al. 2016). Despite the mechanisms to avoid heteroplasmy, reports indicate this state naturally occurring in some species. Furthermore, available techniques are not effective enough to detect heteroplasmy despite being commonly expressed (Ladoukakis & Zouros 2017). Because of the above-mentioned selective replication of only a restricted number of mitochondria, mutations that survive the bottleneck process

are likely to be rapidly fixed, leading to variability within the organelle's population in just one generation (Chinnery et al. 2000).

In human pathology reports, heteroplasmy is increasingly associated with mechanisms of expression of deleterious variants in the mitogenome. Schon et al. (2020) highlight that clinical disorders associated with polymorphisms of the mtDNA are commonly single-nucleotide variants (SNVs) that occur in between 60 and 80% of the molecules in affected tissues. The authors also point out that the variability in heteroplasmy levels seems to be related to the disorder expression between carriers of the same mutation. They explained that carrying the variant does not determine the observation of the related disorder. However, when the defective copies increase in proportion, the chances of expressing the undesired phenotype are higher. Thus, when the mutated copies are shown in frequencies above 60%, disorders are likely expressed.

In conclusion, variations in mitochondrial DNA might lead to implications for fitness. Mishmar et al. (2003) showed how human mitochondrial variation might have allowed the adaptation to environmental and dietarian changes. According to the authors, variants found in populations habiting the extremes north and south of the globe are less efficient in producing ATP. However, they show an increased ability to generate heat. Mishmar et al. (2003) discuss that such features might have represented an evolutionary advantage, contributing to the adaptation of the ancestral humans to the challenging environment.

Moreover, they state that the mitochondrial high mutation rate and central metabolic role make the mtDNA indispensable for the rapid adaptation of animals to new environments. Mitochondrial uniparental inheritance allows rapid segregation, expression, and adaptative selection of advantageous new mutations. Simultaneously, the lack of recombination ensures that beneficial mutations increase the frequency of the whole haplotype via hitchhiking. As environmental conditions that affect metabolic processes influence such variations, they also may impact breeding activities expressed by the appearance of mitochondrial effects.

1.5. Role in lactation

Little is known about the role of the mitochondria on the mammary gland and lactogenesis (Hadsell et al. 2011; Weikard & Kuehn 2018). Nonetheless, it is acknowledged that the mitochondria play a central role in metabolic adaptation (Weikard & Kuehn 2018). Transcriptional and translational factors regulate the copy number of the mitochondria according to the demand of the tissue where it is expressed. A study in mice (Hadsell et al. 2011) showed that during early lactation,

the mammary cell experiences an increase in ATP synthesis activity due to changes in a small number of proteins.

The abundance of mitochondria in a cell is then dependent on the cell's specialisation, age, and health. Since lactation is a highly energy-demanding physiological process, the metabolic modifications induced by it within the body are expected to impact the availability of mitochondria. Moreover, the organelle's capacity to respond to such changes may also influence the progression of lactation. Weikard & Kuehn (2018) discussed that mtDNA copy number is expected to reflect the capacity of tissues to generate energy. Their study demonstrates differences in the modulation of mitochondria biogenesis according to tissue demand, and that modulation also varies depending on the cows' milk yield level. Laubenthal et al. (2016) proposed that the physiological changes ignited by the cow's pre-lactating status might influence not only the mitogenome copy number but also its gene regulation and pathways, along with mitochondria turnover.

When mitochondria number, structure, or function is abnormal, energy production is disturbed. Thus, cows inheriting specific mitochondrial lineages may be equipped with better apparatus to support the demands of lactation, leading to better fitness translated into higher production.

1.6. mtDNA as a component of genetic merit

Dairy breeders have, so far, overlooked cytoplasmic inheritance when estimating breeding values, nonetheless making significant progress considering exclusively nuclear additive genetic effects (Boettcher et al. 1996a). Gibson et al. (1997), however, defines genetic merit as the sum of nuclear and cytoplasmatic components. Southwood et al. (1989) performed a simulation to determine the 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. Accounting for cytoplasmic effect should thus reflect residual additive genetic effects neglected by standard statistical models (Kennedy 1986).

Besides, evidence shows that maternal sources of inheritance also play a role in the expression of yield traits (Boettcher et al. 1996a). Deleterious polymorphisms in mitochondrial genome sequences can disrupt the oxidative phosphorylation chain leading to impaired function, compromising milk production. Thus, selecting for females carrying favourable mitochondrial variants may be beneficial to dairy cattle breeding.

To support the argument that mitochondrial DNA contributes to phenotypic variation, Gibson et al. (1997) point out the organelle's critical role in cellular functions, mitochondrial copy number and its genetic particularities. The mtDNA's mutagenic state has been increasingly associated with early-stage and age-related diseases in humans (Sharma & Sampath 2019). According to Boettcher et al. (1996b), maternal siblings tend to be more alike than paternal ones, which may also suggest that mitochondrial variance increases similarities within families and should also affect economic traits.

Gibson et al. (1997) emphasise the importance of mitochondrial variation in genomic estimations, especially when selecting donor females for reproductive techniques. The authors argue that from small shares on the phenotypic variation, it is possible to obtain significant differences in performance across families impacting the selection of cows for in-vitro fertilisation and embryo transfer.

1.7. Expected implications of accounting for the mitochondrial effect

The female selection pathways are more likely to be impacted when accounting for mtDNA variation in the estimation of breeding values in dairy cattle breeding. However, the category's low selection intensity makes it more challenging to observe mitochondrial effects (Boettcher et al. 1996b). The increased use of female reproductive techniques contributes to higher selection intensity, especially on the Dam of cows' selection pathway. With higher selection intensity, mitochondrial effects will also become more pronounced, making their consideration throughout the selection process more significant.

Bell et al. (1985) found mitochondrial effects significant for dairy cattle production traits in the USA. All evaluated cows were linked to a founder female by using pedigree to trace cytoplasmic inheritance origin. Lines with less than five females were discarded from the study, leaving 102 maternal lines. Applying an animal model, a sire model, and a maternal grandsire model, they found differences between maternal lines to be significant for milk yield and the estimate of cytoplasmic component of variance to be of 2% for that trait. However, they suggest that some cytoplasmic effects on yield traits may occur due to their influence on reproduction.

Using pedigree records to identify maternal lineages and fitting them as a random effect to the animal model, researchers (Spehar et al. 2017) found mitochondrial effects to be significant for yield traits in dairy cattle in Croatia. With 8,583 maternal lines that comprised at least three females, the authors were able to

attribute 3% of the phenotypic variation on milk yield to the cytoplasmic component. They highlight the importance of a positive mitochondrial effect for the Dams of cows category. In contrast, such an effect is negligible for the Dam of bulls' category as the organelle will not be transmitted down that pathway. Thus, the authors discuss that the inclusion of a maternal lineage effect when estimating breeding values for females' selection for multiple ovulation and embryo transfer would be beneficial.

In another study, sequencing 109 complete mitogenomes of Holstein cattle, researchers (Brajković 2019) estimated the proportion of variance explained by mitochondrial inheritance. Using four distinct models for variance component estimation, the authors found the cytoplasmic component to respond for 5% of the phenotypic variation for milk yield in the first three lactations. Their results indicate that the mitochondrial genome explains a considerably high proportion of the phenotypic variation for milk yield in dairy cattle, raising the question of its implications on breeding practices.

On the other hand, researchers also failed to identify mitochondrial effects associated with dairy cattle yield traits. Despite identifying polymorphisms on the mitochondrial DNA regions responsible for replication and transcription, Brown et al. (1989) did not find significant mitochondrial effects linked to dairy cattle yield traits. According to the authors, pedigree records are not sufficient to distinguish true maternal lineages. Their findings indicate that molecular markers and genomic information may correctly assess the component of variance associated with the mitochondria.

Studies regarding mitochondrial effects on yield traits on dairy cattle have strictly taken into consideration pedigree records. As highlighted by Brown et al. (1989), genomic data should be more effective in assessing the impact of mitochondrial variation on estimating breeding values. To this date, no study has been done in such a fashion. Therefore, the present work aims to evaluate the impact of accounting for mtDNA variation on the estimation of breeding values for dairy cattle breeding considering milk yield selection. This study compares estimations using a standard progeny testing scheme and a genomic selection scheme to their counterparts accounting for mitochondrial effects.

2. Materials and Methods

For this study, I designed a dairy cattle breeding scheme aimed to improve a single polygenic trait (milk yield) with multiple observations (lactations). The study was performed in R environment using the package "AlphaSimR" (Gaynor et al. 2020). Breeding values were estimated with the software BLUPF90 (Misztal et al. 2014) using two different models, giving rise to two breeding schemes: (i) progeny testing-based selection and (ii) genome testing-based selection. I started the dairy breeding programme simulations by running 20 generations of conventional progeny testing selection. Following, I evaluated four Breeding Scenarios, each running for extra 20 generations. The evaluation scenarios were: (1) conventional progeny testing (stdPBLUP), (2) progeny testing accounting for mitochondrial effects (mtPBLUP), (3) conventional genome selection (stdGBLUP), and (4) genome selection accounting for mitochondrial effects (mtGBLUP). Due to lack of information regarding the average number of causative loci in the mitogenome I have also tested two Trait Scenarios covering extreme possibilities for such parameter. The first trait scenario considered that all segregating sites on the mitochondrial genome were causal (maxQTL), while the second considered only one segregating site on the mtDNA as causal (minQTL). For validation of the results, I estimated accuracies, bias, and inflation for the estimations, whilst following genetic gain and genetic variance trends over generations.

2.1. Simulation parameters

First, I performed a coalescent simulation to obtain the two founder genomes: (1) to mimic the nuclear genetic structure and (2) to mimic the mitochondrial genetic structure. For the nuclear simulation, I established the number of diploid chromosomes as 10. Although the cattle genome is comprised of 30 chromosomes, this restriction was in place to reduce the computational burden and is expected not to influence the results. The physical length of the genome was defined as 3G base pairs. To generate whole-genome sequences, the following parameters were used: recombination rate of 1e-8 per base pair, the mutation rate of 2.5e-8 per base pair, and effective population size of 90. To simulate the historical effective population

size in taurine cattle, I used the "CATTLE" setting in "AlphaSimR" which considers the variation throughout time as described by MacLeod et al. (2013). For each of the simulated sequences, 1,000 loci were randomly selected as markers and other 1,000 loci as causal. The additive effect associated with each causative loci was defined from a normal distribution as ~N(0, σ^2_a). The simulations were performed considering one polygenic trait, milk yield, with heritability (h^2) of 0.3 and phenotypic standard deviation (sd) of 1,890 Kg. Since assuming that genetic merit is the sum of nuclear and cytoplasmic components, heritability was partitioned between additive and mitochondrial effects. Thus, the additive genetic variation (σ^2_a) was defined as considering additive effects accounting for 25% of the phenotypic variation (σ^2_p), $\sigma^2_a = 0.25\sigma^2_p$. The components of variance in ratio and their absolute values are expressed in Table 1.

Table 1. Variance components ratios (1st row) and absolute values (2nd row) for trait milk yield, repeatability model - 1st to 5th lactation.

Trait	σ^2_p	σ^2_a	σ^2_m	σ^2_{pe}	σ^2_e
Milk	sd^2	0.25	0.05	0.60	0.10
Yield (Kg)	3572100	893025	178605	2143260	357210

Sd = standard deviation; σ^2_p = phenotypic variance; σ^2_a = additive variance; σ^2_m = mitochondrial variance; σ^2_{pe} = permanent environment; σ^2_e = residual variance.

To generate the mitochondrial genetic profile, I simulated one haploid chromosome. This chromosome had 16,202 base pairs of physical length. The mitochondrial mutation rate was considered 2.5e-07 (Allio et al. 2017).

The definition of the mitochondrial effective population size (Ne) was dependent on reports regarding the diversity in mitochondrial populations across cattle breeds. When estimating the proportion of phenotypic variance associated to the mitochondrial DNA, Brajković (2019) identified 96 unique haplotypes in a population of 109 founders. This number was similar to that reported by other authors (Sharma et al. 2015; Xia et al. 2019b).

To be able to obtain a level of diversity that matched the referred data, the mitochondrial effective population size (Ne) was set to 1,000. The historical effective population size (histNe) and their respective historical generation (histGen) - information necessary to reconstruct the coalescent population evolutionary history, were as following: histNe = 1500, 2000, 2500, 3500, 7000, 10000, 17000, 62000 and histGen = 25, 155, 455, 655, 1755, 2355, 3355, 33155 (MacLeod et al. 2013).

Regarding the number of segregating sites in the mitogenome, a threshold of 400 was established based on the findings of Brajković (2019). Because mitochondrial DNA is a more concise genome than the nuclear, without introns and with only a strict regulatory non-coding region (Srirattana et al. 2017), most segregating sites could be causal. Since the actual number of genes influencing the trait of interest is unknown, two extreme trait scenarios were tested: i) all segregating sites are causal and marker density is equal to the number of causal loci (maxQTL); ii) one segregating site is causal and marker density is equal to the number of segregating sites minus one (minQTL).

A mitochondrial effect was linked to every causal locus and defined as well from a normal distribution, ~N(0, σ^2_m). The mitochondrial genetic variation (σ^2_m) was assumed to account for 5% of the phenotypic variation and therefore $\sigma^2_m = 0.05\sigma^2_p$.

Both simulated haplotype populations were randomly sampled to give rise to founder nuclear and mitochondrial genomes. The mitochondria founder population accounted for around 100 unique haplotypes. Each unique haplotype represented one maternal lineage.

The nuclear base population was composed of 2,000 individuals, half males, and half females. For each female from the nuclear population, one maternal lineage was randomly assigned. The genetic merit of an individual was considered as the sum of the effects at all its causative loci, additive and mitochondrial. To simulate phenotypes for the females the following model was applied:

$$y_{ij} = \mu_i + nGV_j + mGV_j + pe_{ij} + e_{ij}$$

Where y_{ij} is the phenotypic mean of animal j on lactation i, μ_i is the population mean for lactation i, nGV_j, is the additive genetic value of animal j, mGV_j the mitochondrial genetic value of animal j, pe_{ij} the permanent environment effect for animal j on lactation i, and e_{ij} is the random error for animal j on lactation i. The permanent environment effect was obtained by sampling from a normal distribution with mean zero and variance σ^2_{pe}. Estimated errors were obtained following the same procedure, however considering the variance as σ^2_e. Lactation means were defined as (Brajković 2019): 6,733 kg, 7,440 kg, 7,344 kg, 7,482 kg and 7,168 kg, respectively.

2.2. Breeding schemes & Population structure

A representation of the breeding scheme used throughout simulations is shown in Figure 1.

I designed the breeding scheme considering overlapping generations and selection based on breeding value estimation and phenotypic performance. The population size was defined in a way to generate five elite sires every year. A 10% selection rate was imposed to obtain the required population of sires.

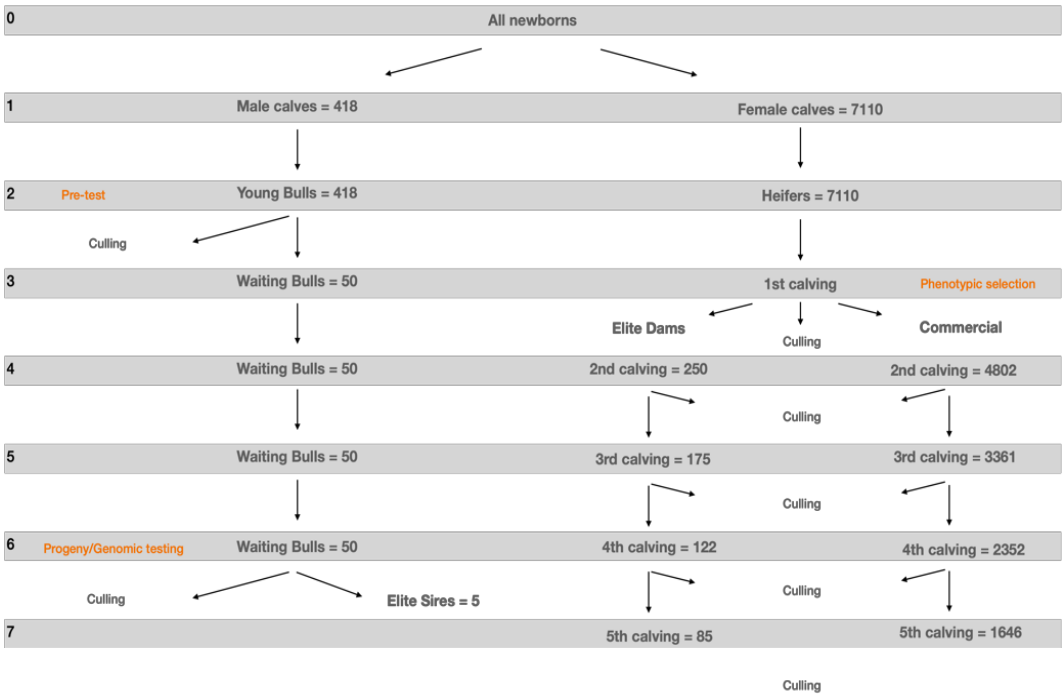


Figure 1. Dairy cattle breeding scheme.

The male selection pathway had a generation interval of 6 years, from which 4 years consisted of data gathering for the progeny test. This method secured all waiting bulls had a minimum of 100 phenotyped daughters when of the estimation of their breeding values (progeny test). The female selection pathways had a generation interval of 7 years. The simulated population was divided into six categories: (1) Elite Dams, composed by the best performing 250 females at first-lactation; (2) Commercial, composed by the best 70% first-lactation females after the selection of Elite Dams; (3) Heifers, composed by 7110 females before the closure of the first-lactation; (4) Elite Sires, composed by the 5 males showing higher breeding values; (5) Waiting Bulls, composed by the best 50 males selected from the Young Bulls category based on their breeding values; and (6) Young Bulls, composed by 97% of the male offspring obtained from the mating between Elite Sires and Elite Dams. A culling rate of 30% was applied to both Elite Dams and Commercial at the end of each lactation, moreover, by the end of the 5th lactation, all females were involuntarily culled. All Elite Sires were also culled after 5 years in the category. A summary of the population can be found in Table 2. Breeding values were estimated using the software BLUPF90 (Misztal et al. 2014).

Table 2. Population Structure summary. Total amount of animals in each category per year spent in the category (1–4).

Category	1	2	3	4	Total
Heifers	7110	7110	7110	-	21330
Elite Dams	250	175	122	85	632
Commercial	4802	3361	2352	1646	12161
Young Bulls	418	418	-	-	836
Waiting Bulls	50	50	50	50	200
Elite Sires	5	5	5	5	20
Yearly population					**35179**

2.2.1. Progeny Testing Scheme (stdPBLUP & mtPBLUP scenarios)

For the progeny testing scheme, 50 young bulls were pre-selected based on their breeding values at age 2 to enter progeny testing. After 4 years, breeding values were estimated based on pedigree data, and the best 5 proven bulls selected to become Elite Sires. All non-selected males in both selection steps were culled.

Females were selected as Elite Dams based on their performance at the end of the 1st lactation. Seventy percent of the non-selected females were moved to the Commercial group and the worst 30% culled. Commercial, together with heifers, were mated with waiting bulls to generate data for progeny test. Breeding values were estimated based on pedigree and performance data using the repeatability model defined by Mrode (2014):

$$y = \mathrm{Xb} + Za + Spe + e$$

Where: y is a vector of observations, b a vector of fixed effects (lactation order), and X the incidence matrix that links it to the records; a is a vector of additive random effects (animal), related to pedigree records by the incidence matrix Z; pe is a vector of random environment effects and S its incidence matrix. The vector of random residual effects is defined as e.

The model considered multiple measurements of the same trait for each female (multiple lactation measures). The permanent environment effect was used to account for covariances between repeated measurements caused by environmental factors acting on successive records securing accurate predictions.

2.2.2. Fitting the mitochondrial effect to the mixed model

A factor to be considered when accounting for maternal lineages on breeding values estimations is whether to fit it in the mixed model as a random or fixed effect. Fixed effects refer to those variables related to the individual which do not change or change at a continuous rate over time. On the other hand, random effects refer to variables that change randomly according to a sample from a population of variables.

Boettcher et al. (1996b) bases his argument for dealing with the maternal lineage as a fixed effect on the biology of the mitochondrial DNA. When considered that the mitogenome is not subject to recombination and that recent mutations cause a small impact on solutions, maternal lineage effects can be considered repeatable over time and, therefore, treated as a fixed effect. However, from a statistical perspective, maternal lineages are sampled from a random population and should be fitted as a random effect. Although Boettcher et al. (1996b) did not find a significant difference in correlations between real and estimated mitochondrial effects when comparing the two approaches, Gibson et al. (1997) states that considering it as a random effect is crucial to secure unbiased predictions. Accuracies and precision of predictions are also strongly associated with the variance of the effect and the size of the maternal lineage groups. The greater the number of cows per maternal lineage the better tend to be the predictions (Boettcher et al. 1996b).

Following the indications of Gibson et al. (1997), when accounting for the mitochondrial contribution to phenotypic variation, mitochondrial effects were fitted in this study as a random effect. The model used was adapted from (Mrode & Thompson 2005):

$$y = \mathrm{Xb} + Za + Wm + Spe + e$$

Where y is a vector of observations, b a vector of fixed effects (lactation order), and X the incidence matrix that links it to the records; a is a vector of additive random effects (animal), related to pedigree records by the incidence matrix Z; m is the vector of mitochondrial effects and W its incidence matrix; pe is a vector of permanent environment effects and S its incidence matrix. The vector of random residual effects is defined as e.

2.2.3. Genome Testing Scheme (stdGBLUP & mtGBLUP)

To simulate the genome testing breeding scheme, breeding values were estimated based on pedigree, performance and genotypic data via the single-step Genomic Best Linear Unbiased Prediction (ssGBLUP) methodology.

A summary of the genomic information used is found in Table 3. A total of 10,200 genotypes were used every generation to perform the genomic selection. The reference population was constructed at the beginning of the evaluation scenario by genotyping 8,944 phenotyped females selected at random. Every generation the genotyped population was updated with SNP data from new-borns. The updated population constituted 613 males and 2,461 females, all of them offspring of the nucleus population (mating between Elite Sires and Elite Dams). To maintain the size of the genomic record older genotypes were excluded every time new ones were added.

Table 3. Summary of genomic data.

Total population (N)	Reference population (N)	Genotyped males (N)	Genotyped females (N)	Mitochondrial haplotypes*(N)
35179	8944	613	2461	104

** mean of ten replicates*

For the ssGBLUP method model, a matrix H was implemented, defining the relationship between genotyped and non-genotyped animals. By replacing the inverse of the relationship matrix with H^{-1} during computations it was possible to efficiently derive genomic breeding values for selection-candidates. Therefore, H^{-1} was defined as (Aguilar et al. 2010):

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathrm{G}^{-1} - A^{-1}{}_{22} \end{bmatrix}$$

Where A is the pedigree relationship matrix whilst G the genomic relationship matrix; subscript 2 indicates genotyped animals.

Our decision to fit the mitochondrial effect as a random effect to the repeatability model required manually generating the mitochondrial genetic relationship matrix. The relationship matrix was created only once, before starting the evaluation scenarios, and was defined as (Aguilar et al. 2010):

$$G_m = \frac{ZZ'}{k}$$

Where Z is a matrix of SNP markers and k is the sum of the frequencies of heterozygous loci. Because the mitogenome is haploid, k was defined as $k = \sum p(1-p)$, where p implies the major allele frequency (Aguilar et al. 2010).

2.3. Validation

2.3.1. Bias and Inflation

Taking into consideration that the analysis was based on simulated data and, therefore, true-breeding values were known; accuracies were obtained by correlating them to their estimations ($r_a = cor(Tbv, Ebv)$). Accuracies were calculated for the following categories: (1) heifers – females before first calving, (2) 1st lactation – females that concluded their first lactation, (3) cows – all females with closed second lactation and beyond, (4) young bulls – young males' candidate for selection to enter progeny test, and (5) proven bulls – waiting bulls that finished progeny test candidates for selection. To determine the inflation and bias of the predictions a linear regression was used as follows:

$$y = 1b_0 + b_1\hat{a} + e$$

Where y is a vector of true breeding values, â a vector of the solutions (estimated breeding values), b_0 and b_1 are unknown regression coefficients and e is the residual. Therefore, the bias of the estimations was evaluated by the observation of b_0 whilst results from b_1 allowed an interpretation of their inflation. Unbiased predictions are expected to return $b_0 = 0$, $b_1 = 1$.

2.3.2. Genetic parameters

The correlation between nuclear and mitogenome was tracked throughout time in the whole population as means to understand the impact of selection on the mitogenome. Likewise, genetic mean for the nuclear genome and variances for both nuclear and mitogenome were observed.

2.4. Statistical analysis

All simulations were repeated ten times. Results are expressed as mean and 95% confidence interval of the repeated observations. Analyses were performed in R environment (Plummer et al. 2006).

3. Results

Results for the nuclear and mitochondrial genome and the total genetic merit (the sum of nuclear and mitochondrial loci) were obtained by observing the progression of the breeding programme over time. Simulations were divided into two periods, the first 20 generations constituting a burn-in stage in which selection was performed using the stdPBLUP model. The further 20 generations constituted the evaluation stage in which the four breeding scenarios were tested. Each generation was equivalent to one breeding cycle. The evaluated parameters were estimated for the whole population or according to categories. Differences regarding population categorisation and timeframe are stated when convenient. For all following figures, lines represent repetitions' mean and shaded areas indicate 95% confidence intervals.

3.1. Variance

The variation proportion attributed to mitochondrial and nuclear DNA throughout generations followed that defined at the beginning of the simulations via σ^2_a and

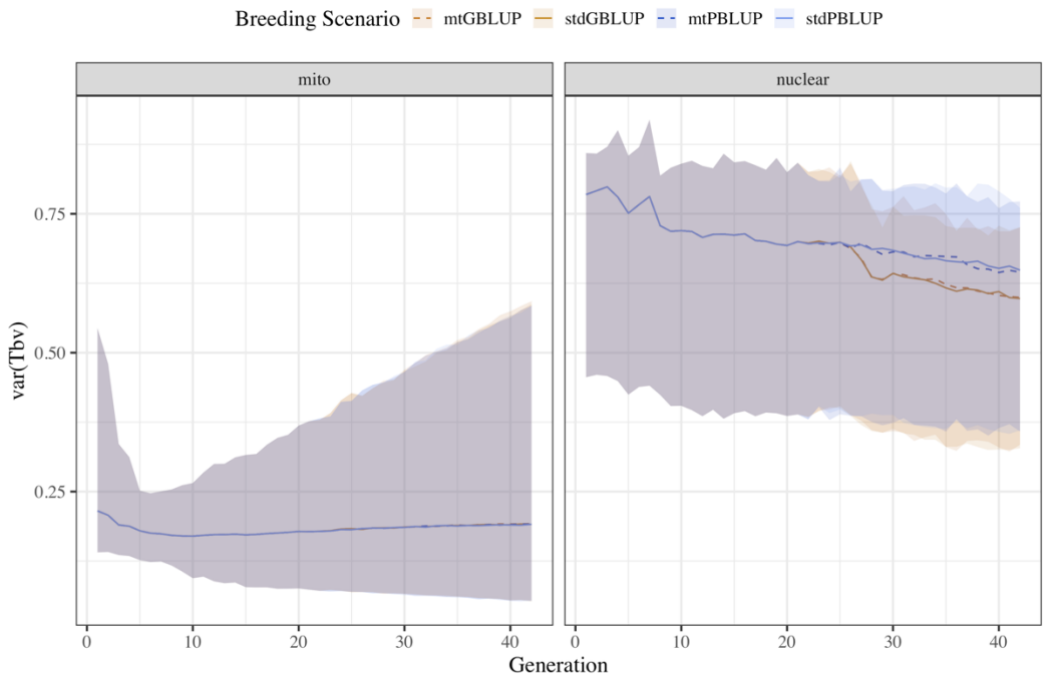


Figure 2. Genetic variance trend for mitochondrial and nuclear genomes comparing Breeding Scenarios. Results are shown as the mean of ten replicates (lines) and 95% confidence interval (shade).

σ^2_m. A ratio of 4:1 (nuclear to mitochondrial) was maintained during the burn-in and evaluation stages, as observed in Figure 2.

No significant difference between Trait Scenarios in the genetic variation trend was present nor between Breeding Scenarios. Therefore, Figure 2 represent the mean observation for both maxQTL and minQTL scenarios. The genetic variation for mitochondrial and nuclear genomes are shown over 40 years of selection comparing the four Breeding Scenarios tested.

Mitochondrial genetic variation was reduced in the first ten generations of selection, moving from 0.223 to 0.177 (20.63%). With the progression of the breeding programme, however, variation was recovered and maintained around 0.20.

The genetic variation for the nuclear genome decreased throughout the 40 generations of selection. The first ten generations faced an intense reduction in genetic variation of 9.39%. In the second half of the burn-in stage, variation loss was restricted, and just 1.42% change was observed. The implementation of genomic selection, whether via the mtGBLUP or the stdGBLUP scenarios, led to more genetic diversity loss. During the 20 generations, 13.81% of the variation was lost. In comparison, using the progeny testing scenarios (mtPBLUP or stdPBLUP) caused a reduction of 7.12%.

The correlation between nuclear and mitochondrial loci was also tracked over time, and it is shown in Figure 3.

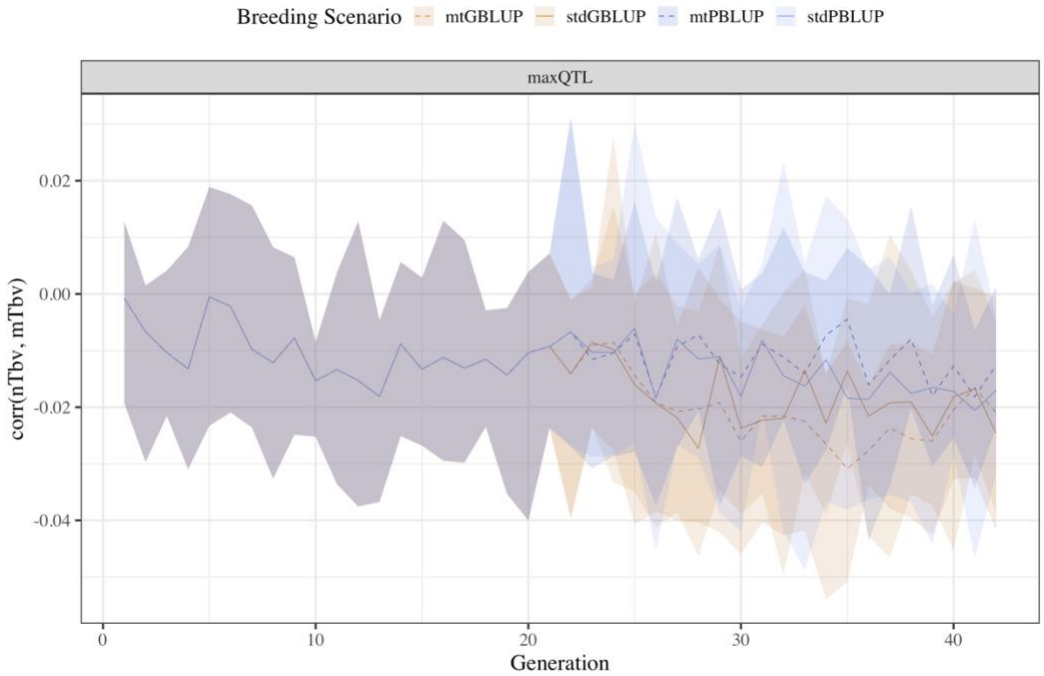


Figure 3. Correlation between nuclear and mitochondrial genomes comparing Breeding Scenarios. Results are shown as mean of ten replicates (line) and 95% confidence interval (shade).

There was no significant difference in correlation of nuclear and mitochondrial loci between Trait Scenarios or Breeding Scenarios. The genetic correlation started close to zero. With the implementation of selection, this association shifted, becoming negative. During the course of selection, the correlation between the two genomes was maintained negative with an average of -0.013.

3.2. Genetic gain

The genetic gain was estimated for the population as a whole, shown in Figure 4.

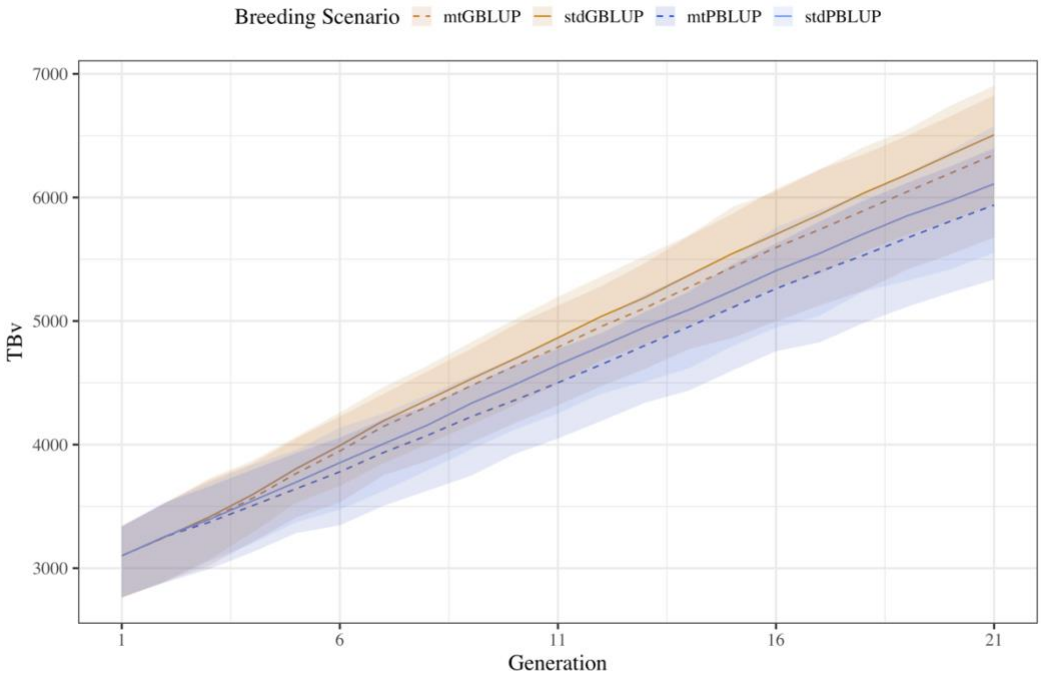


Figure 4. Population genetic gain over 20 years of selection comparing Breeding Scenarios. Results are shown as mean of ten replicates (lines) and 95% confidence interval (shade).

No influence of Trait Scenario or applying the mitochondrial effect models (mtPBLUP and mtGBLUP) was captured. An average 110% gain was obtained with 20 generations of selection. The genetic gain was accelerated in the genome testing breeding schemes (stdGBLUP and mtGBLUP) in comparison with the progeny testing breeding schemes (stdPBLUP and mtPBLUP). At the start of the evaluation scenario, the mean breeding value was 2971 kg. The mean was increased to ~6426 kg with the genome testing schemes and to ~6024 kg with the progeny testing schemes.

3.3. Nuclear Genome

3.3.1. Accuracy

Figure 5 presents the accuracy for nuclear breeding values estimations under the Trait Scenario maxQTL.

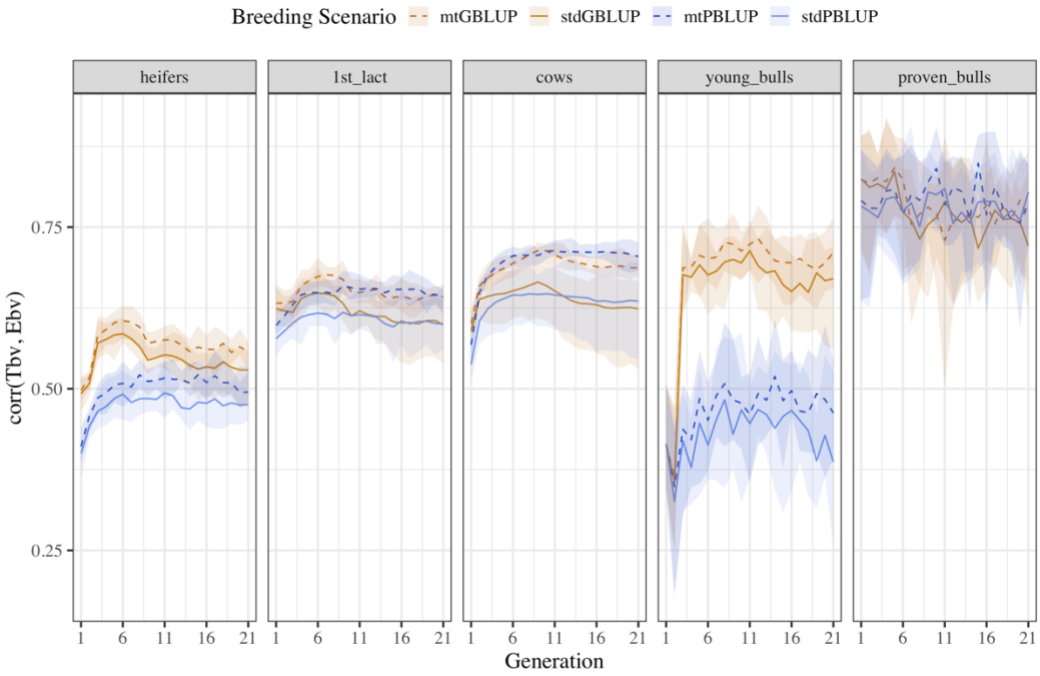


Figure 5. Correlation between nuclear true and estimated breeding values under maxQTL Scenario comparing Breeding Scenarios. Results are shown as mean of ten replicates (lines) and 95% confidence interval (shade).

The results show that genomic selection increased accuracies for the categories heifers and young bulls. The average nuclear accuracy went from 0.49 to 0.56 for heifers and 0.45 to 0.66 for young bulls. Regarding the implementation of the mitochondrial effect model, improvement in nuclear accuracy was significant for all female categories. In the category young bulls, despite the Breeding Scenarios mtGBLUP and mtPBLUP tending to show higher nuclear accuracy than its counterparts, the difference was not significant. Table 4 summarises the average nuclear accuracy for the female categories during the evaluation period.

The category cows showed the most significant gain in applying the mitochondrial effect model. Moving from the Breeding Scenario stdPBLUP to mtPBLUP led to a 0.07 gain in nuclear accuracy. A similar gain was obtained when moving from the stdGBLUP to the mtGBLUP Scenario.

Table 4. Twenty generations average of accuracy of nuclear estimations for female categories.

Category	stdPBLUP	mtPBLUP	stdGBLUP	mtGBLUP
Heifers	0.47	0.50	0.55	0.57
1st Lactation	0.61	0.65	0.62	0.65
Cows	0.63	0.70	0.64	0.69

stdPBLUP = Progeny testing breeding scheme with standard model for estimating breeding value; mtPBLUP = Progeny testing breeding scheme accounting for mitochondrial effect on breeding value estimation; stdGBLUP = Genome testing breeding scheme with standard breeding value estimation model; mtGBLUP = Genome testing breeding scheme accounting for mitochondrial effect on breeding value estimations.

3.3.2. Validation

The bias of the nuclear estimated breeding values is presented in Figure 6. No difference was observed between Trait Scenarios and, thus, Figure 6 shows only results for maxQTL Scenario.

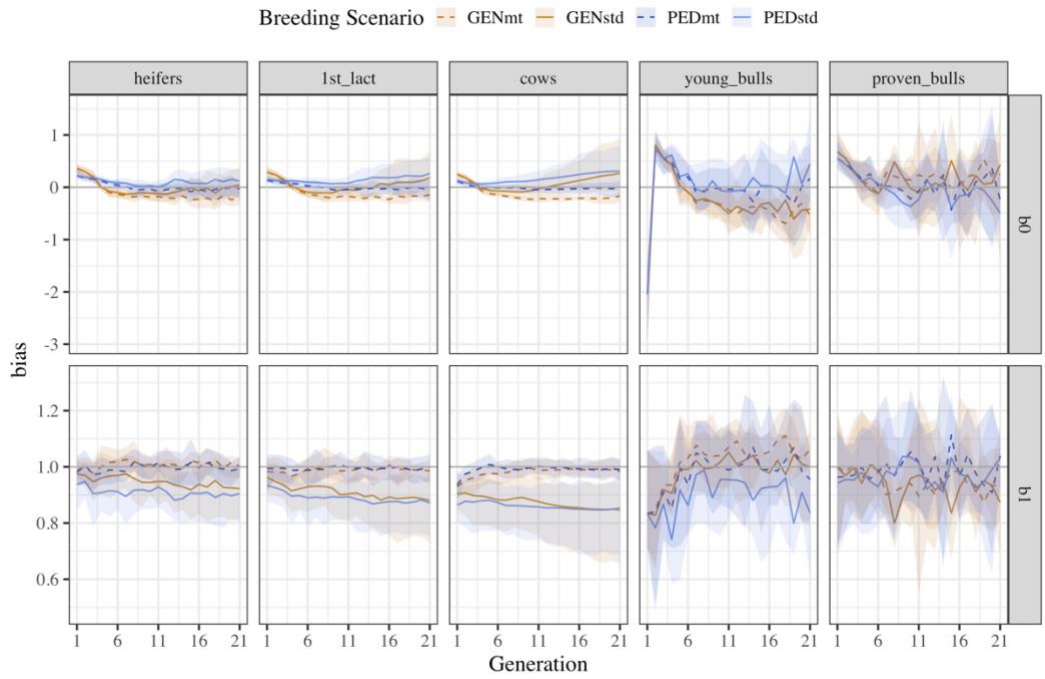


Figure 6. Bias for nuclear breeding values obtained as the intercept (b0, bias) and the slope (b1, inflation) of the regression of true on estimated breeding values. Comparison of Breeding Scenarios on the maxQTL Trait Scenario. Results are shown as mean of ten replicates (lines) and 95% confidence interval (shade).

For all categories, except proven bulls, the use of genomic selection induced bias. For the category first lactation and cows, the mtGBLUP Scenario caused more bias on the estimated nuclear breeding values than the other scenarios. Moreover, the Progeny testing scheme accounting for mitochondrial effect (mtPBLUP) was the only scenario leading to unbiased predictions.

For all female categories, the implementation of the mitochondrial effect model (mtPBLUP and mtGBLUP) seemed to correct the estimations' scale, leading the results for the slope (lower graphs) to be closer to 1 than that observed for the scenarios stdPBLUP and stdGBLUP. No significant difference between Breeding Scenarios was observed for the male categories, and results tended to 1. This result indicates no inflation on nuclear breeding value estimations for these categories.

3.4. Mitochondrial Genome

3.4.1. Accuracy

Figure 7 show the results for accuracy of mitochondrial estimated breeding values. Results are compared across the four Breeding Scenarios and two Trait Scenarios.

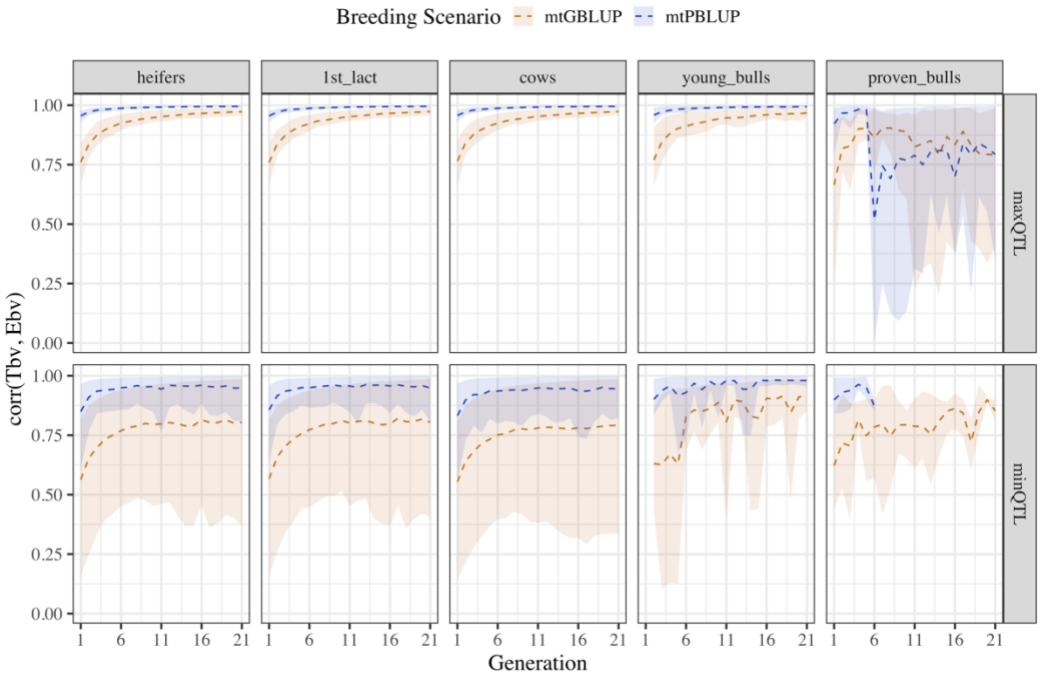


Figure 7. Correlation between mitochondrial true and estimated breeding values, comparison of Trait Scenarios. Results are shown as mean of ten replicates (lines) and 95% confidence interval (shade).

The pedigree model, mtPBLUP, was highly effective in predicting mitochondrial breeding values for all categories except proven bulls.

Mitochondrial accuracies were around one during all evaluation stages in both Trait Scenarios under the mtPBLUP Breeding Scenario due to correctness of the pedigree. The genomic model presented lower accuracy for estimating

mitochondrial breeding values. For the maxQTL Trait Scenario, average accuracies started ~0.75 and moved closer to one by the end of the simulations. Having only one causative locus in the mitogenome impacted the predictions. For the minQTL Scenario, the confidence interval of repetitions was higher, and average mitochondrial accuracies ranged around 0.65.

3.4.2. Validation

Regarding the validation of the mitochondrial genome results, all the female categories results show similar behaviour. Thus, Figure 8 present only the categories cows, young bulls and proven bulls.

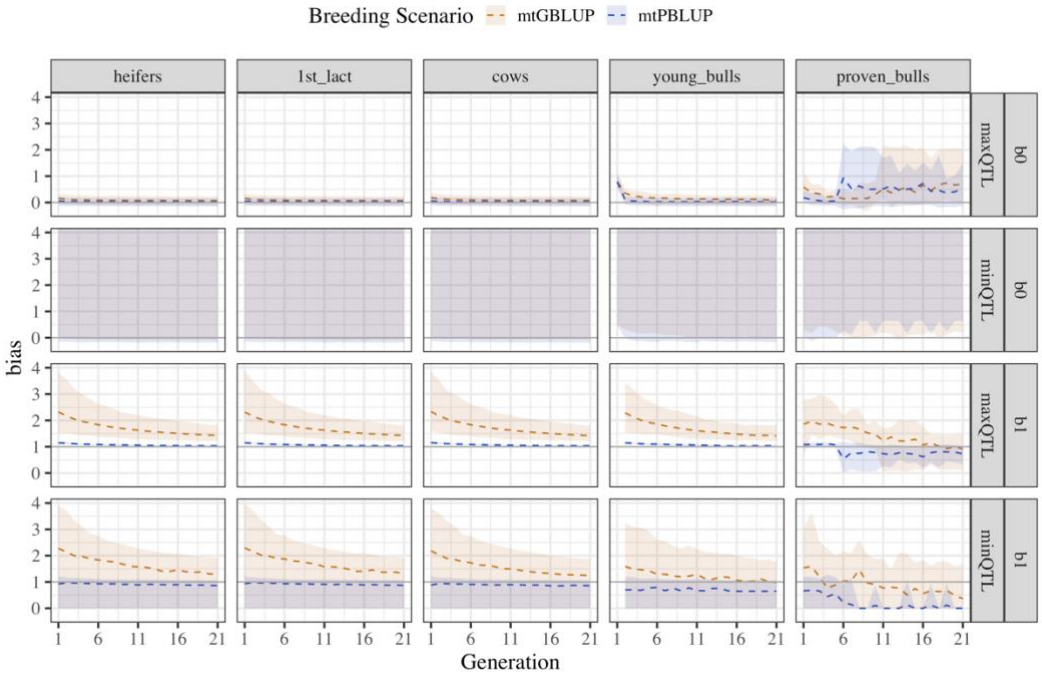


Figure 8. Bias for the mitochondrial breeding values obtained as the intercept (b0, bias) and the slope (b1, inflation) of the regression of true on estimated breeding values. Results are shown as mean of ten replicates (lines) and 95% confidence interval (shade).

The Figure shows that females' mitochondrial estimations were unbiased on the maxQTL Scenario, although the mtGBLUP Scenario induces inflation of the estimations. When testing the minQTL Scenario, the regression intercept results became distant from zero, and predictions tended to be deflated on the mtPBLUP Scenario. For the male categories, young bulls' results were very similar to that observed for females. However, the prediction of proven bulls mitochondrial breeding values was slightly biased.

3.5. Total genetic merit

3.5.1. Accuracy

When considering the genetic merit as the sum of nuclear and cytoplasmic components, the mitochondrial effect significantly increased the accuracy of estimated breeding values for all categories except proven bulls (Figure 9).

The magnitude of the change in accuracy and its significance was dependent on the number of causative loci in the mitochondrial genome. Table 4 summarises the average accuracy for all categories in each Breeding Scenario under the maxQTL Trait Scenario.

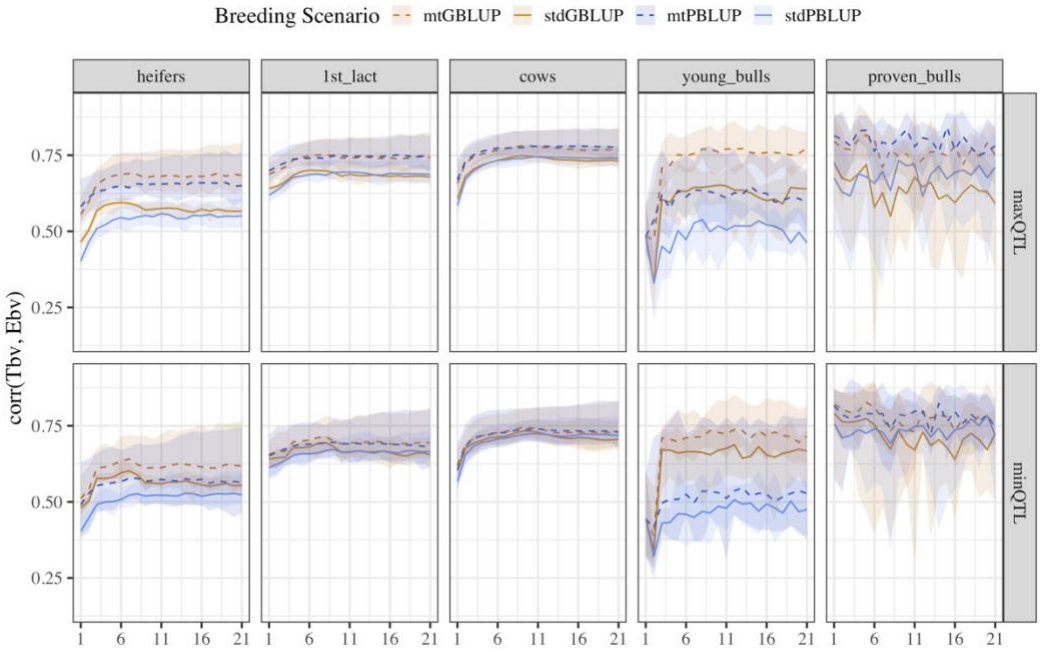


Figure 9. Correlation between total true and estimated breeding values comparing Trait Scenarios. Results are shown as mean of ten replicates (lines) and 95% confidence interval (shade).

The categories heifers and young bulls were the most benefited from the implementation of mitochondrial effect models. Both categories faced an 0.1 increase in total accuracy when comparing the stdPBLUP with mtPBLUP. The same average gain was observed when comparing the scenarios stdGBLUP and mtGBLUP.

Table 5. Twenty generations average of accuracy of total genetic merit for all categories.

Category	stdPBLUP	mtPBLUP	stdGBLUP	mtGBLUP
Heifers	0.53	0.65	0.57	0.67
1st Lactation	0.68	0.74	0.68	0.74
Cows	0.73	0.77	0.73	0.76
Young Bulls	0.49	0.61	0.61	0.73
Proven Bulls	0.68	0.79	0.64	0.76

stdPBLUP = Progeny testing breeding scheme with standard model for estimating breeding value; mtPBLUP = Progeny testing breeding scheme accounting for mitochondrial effect on breeding value estimation; stdGBLUP = Genome testing breeding scheme with standard breeding value estimation model; mtGBLUP = Genome testing breeding scheme accounting for mitochondrial effect on breeding value estimations.

3.5.2. Validation

The Trait Scenario did not influence the validation of estimations for the total genetic merit (Figure 10).

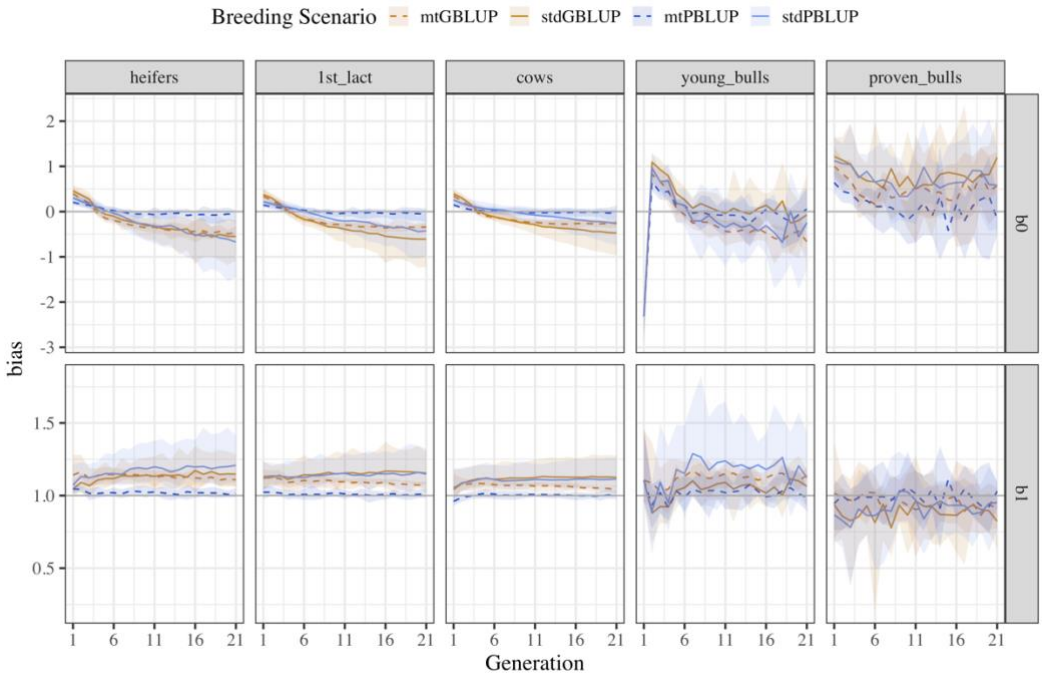


Figure 10. Bias for the total breeding values obtained as the intercept (b0, bias) and the slope (b1, inflation) of the regression of true on estimated breeding values.Results are shown as mean of ten replicates (lines) and 95% confidence interval (shade).

For the female categories, the genome testing breeding schemes introduced bias to the estimations. Standard selection schemes had the results for the slope deviating from 1. With the introduction of the mitochondrial effect models for estimation of

breeding values (mtPBLUP and mtGBLUP), results for the slope tended to approximate to 1, indicating no inflation on estimations. A tendence to correction on the estimations for the intercept (moving closer to zero) was also observed with the mitochondrial effect models in comparison to the standard breeding scenarios. No differences between breeding scenarios for b0 or b1 were observed for the categories young bulls and proven bulls. For these categories both results were considered unbiased.

4. Discussion

The results suggest that accounting for mitochondrial effect on the estimation of breeding values for milk yield in dairy cattle improves accuracy of predictions. The increase in accuracy seems to be more significant for female selection candidates which is expected considering the mitochondrial model of inheritance. Although males do not transmit their mitochondria values to offspring, accounting for mitochondrial effect on the estimation of their breeding values indicates the quality of the dam line they belong to.

Since the simulations relied on long and correct pedigree, using only maternal lineages to obtain mitochondrial effect would have been enough to derive correct estimations. However, in a practical scenario, where pedigree faults could lead to incorrect association of maternal linages, genotyping the mtDNA is advantageous to secure correctness of the relationships.

The combination of mitochondrial effect and genome selection appears to be a good strategy to improve accuracy of estimated breeding values for selection candidates without performance records (heifers and young bulls). For these categories, considering the genetic merit as the sum of nuclear and cytoplasmic components also led to higher accuracy estimations. Since breeding values for individuals with no own performance are dependent on the estimations of their relatives, including mitochondrial value on the estimations for females contributes to improve the predictions for heifers and young bulls. Improvement in accuracies, however, did not impact the overall genetic gain. For these simulations, in the mtGBLUP and mtPBLUP scenarios, males were selected on their total estimated breeding values. Moreover, their inability to transmit mitochondria and their role on driving genetic gain in the population justify the limited the impact of accounting for mitochondrial effect on the overall genetic gain. To improve genetic gain, bulls should be selected on their nuclear estimated breeding values while females on their total estimated breeding values.

The development of this study pointed to some gaps in the knowledge about the population genetics of the mitochondria. Since the theoretical assumption of mitochondrial diversity seems to contradict empirical data, violations had to be put in place to secure the simulated data was realistic. To better understand the results presented here, it is critical to discuss mitochondrial diversity. In this section will also be discussed the correlation between nuclear and mitochondrial genomes and the impact of selection on their relationship. Moreover, the consequences of the number of causative loci to the observation of mitochondrial effect and the impact

of mitochondrial effect on the estimation of breeding values for dairy cattle. Finally, the implications of the results on dairy breeding practices will be covered.

4.1. Mitochondrial diversity and the violation of assumptions

Studies show significant diversity in mtDNA across cattle populations. When assessing 2373 samples from the Croatian Holstein population, Brajković (2019) traced them back to 109 founders that harboured 96 distinct mitochondrial haplotypes. For a mixed population of 107 Indian cattle, Sharma et al. (2015) reported 60 unique mitochondrial haplotypes. Xia et al. (2019b) found 338 haplotypes in various 1105 Chinese cattle and 47 when analysing 109 sequences from Yunling cattle (Xia et al. 2019a), a composite beef breed.

The first results for the current simulations lacked diversity, with the number of mitochondrial haplotypes falling well below the values mentioned above. Using the parameters found in literature: mitochondrial mutation rate of 2.5e-07, no recombination, and effective population size of 90 (actual N_e for the autosomal genome of Holstein cattle, supposedly higher than expected for mtDNA), to perform the coalescent simulation of the mitochondrial genome, no more than 20 unique haplotypes were generated [results not shown].

Besides representing a substantial divergence from published data, mitochondrial diversity significantly impacted the observation of mitochondrial effect. Boettcher et al. (1996b) have already brought to attention that the impact of cytoplasmic inheritance on genetic evaluations are dependent on the number of maternal lineages and the size of the groups. With the initial 20 unique mitochondrial haplotypes, no difference was observed in accounting for mitochondrial effect on breeding value estimations [results not shown].

According to the concept of Mendelian sampling, the uniparental inheritance of the mitochondrial DNA should lead to a lower effective population size than that observed for the nuclear genome. For the autosomal genome, every generation, gametes are produced by drawing a possible DNA strand copy out of a pool of 4. Because the autosomal genome is diploid and both parents transmit their copies to the offspring, N_e is dependent on the number of males and the number of females available for reproduction as defined by Falconer & Mackay (1995):

$$N_e = 4 \times N_m \times N_f / N_m + N_f$$

However, for the mitochondrial genome, the strand passed on to the offspring is always a copy of that found in the mother, as the mtDNA is a haploid molecule, uniparentally inherited. Therefore, the mitochondrial effective population size is expected to equal the female effective population size, $N_e = {}^{1}/_{4} N_e$.

There is no explicit agreement on the literature indicating what the actual value is for the mutation rate on the mitogenome. It is thought that, because of the structure of the mtDNA and its environment, mutations are more likely to happen. The absence of histones surrounding the molecule is viewed as a lack of regulatory apparatus and a liability, exposing the molecule to the many reactive agents derived from the oxidative phosphorylation process (Jobling & Jobling 2013). Thus, it is sensible to consider that the mutation rate is at least tenfold higher (Allio et al. 2017) than the rate for the nuclear genome.

Because the mitogenome does not recombine, the only sources of variation are mutations. Such variation will then be impacted by the effective population size and more subjective to drift.

The assumption of effective population size was violated to enable higher diversity on the mitochondrial base population, and the parameter set to $N_e = 1000$. The mutation rate was considered ten times the nuclear mutation rate ($2.5e^{-07}$) and the nonrecombining state of the mitogenome preserved. With these considerations, an average of 104 unique haplotypes was obtained for the base population of 1000 females. A result better aligned with the observed number of haplotypes in real populations (Brajković 2019).

These observations lead to questioning the current understanding of mitochondrial evolutionary history and the correct methods to simulate its demographic history. The number of unique mitochondrial haplotypes clearly influence the magnitude of mitochondrial effect and, therefore, the results obtained from this simulation might be underestimated.

4.2. Impact of the number of causative loci in the observation of mitochondrial effect

The total number of causative loci ruling a certain trait of interest is expected to impact breeding values prediction when under genomic selection. Therefore, accuracies on the mtGBLUP scenario were measured to assess the impact of causative loci on the estimations. The results for accuracies were presented in Figures 5, 7 and 9.

For predictions of nuclear breeding values, presented in Figure 5, the means for both trait scenarios behave very similar and, no significant difference exists.

However, with a greater number of causative loci (maxQTL), the prediction of mitochondrial genetic merit becomes more accurate for all categories except proven bulls (Figure 7). A similar result is observed when analysing the total genetic merit in Figure 9. Despite higher means for all categories, no significant difference is observed between the two tests.

These results indicates that, despite mitochondrial effect being relevant for the estimation of breeding values, it is dependent on the genetic architecture of the trait. More information is needed to ensure the influence of the mitogenome on lactation is govern by multiple or single genes. Genome-wide-association studies have been used for the past decade helping to partially elucidate the polygenic profile of milk yield regarding the autosomal genome of dairy cattle (Jiang et al. 2019). A similar approach would contribute to the understanding of the role mtDNA plays on the trait and facilitate the investigation of mitochondrial effect.

Significant difference between Trait Scenarios was hard to obtain, despite extreme scenarios being tested – maxQTL considering all segregating sites as causal compared to minQTL having only one causal locus. Because the number of maternal lineages varied significantly across replicates, ranging from a minimum of 81 to a maximum of 114, confidence intervals were large. Setting a smaller range for maternal lineages over replicates may contribute to the observation of significant difference between scenarios in future studies.

4.3. Impact of accounting for mitochondrial effect on the estimation of breeding values

The study confirmed the expectations that accounting for mitochondrial effect on breeding value estimations would impact the accuracy of predictions for female candidates (Boettcher et al. 1996b). Results were presented in Figures 5 (nuclear breeding values) and 9 (total genetic merit).

The Figures shows no difference between nuclear and total genetic merit accuracies for the categories first lactation and cows. From that is possible to conclude that neglecting the cytoplasmic component in the composition of genetic merit does not influence accuracies for selection candidates that have some performance records associated to them.

On the other hand, for the category heifers, which include only females that are yet to finish their first lactation neglecting the cytoplasmic component leads to lower accuracy altogether.

Considering that the accurate selection of these young females is crucial for dairy breeding programmes, especially in the elite herd level, taking into consideration cytoplasmic components and mitochondrial effect may be beneficial.

4.4. Implications for breeding practices

The results did not indicate an impact of mitochondrial effect on genetic gain. This can be related to the fact that, in this simulations, selection intensity of females is rather small and their contribution to the genetic improvement of the population is less than that of the males. As no significant impact in accuracy estimations was observed with the introduction of mitochondrial effects for proven bulls, genetic gain was also not influenced.

However, some impact of accounting for mitochondrial effect on the accuracy of breeding value estimations for young bulls was captured. Although males do not transmit mitochondria to their offspring, because their breeding values are determined based on relatives (females with lactation records), the mitochondrial contribution to their female-relatives phenotypes cause an indirect improvement on their own predictions.

If considering strictly the selection and improvement of female populations, accounting for mitochondrial effect may be beneficial for dairy breeders. As accuracy is directly related to genetic gain, methods that lead to better estimation of breeding values will lead to higher genetic gain in the population. So far, improvement in dairy cattle populations has been made mostly through the intense selection of males and the genetic improvement in the female populations has been lower (García-Ruiz et al. 2016).

Considering the growing dissemination of female reproductive technologies applied to dairy cattle, the consideration of mitochondrial effect on the estimation of breeding values can positively contribute to a more accurate selection of egg donors. The selection of female carriers of mitochondrial lines better adapted to high energy production can favour the maintenance of lactation. Further studies are needed to secure this observation.

5. Conclusions

Simulations and modelling allow the exploration of scenarios, helping to identify gaps in knowledge and to gain insight about variables that are more or less relevant to the system modelled. On the other hand, simulations are highly dependable on the accuracy of inputs and can be challenging to perform for complex systems. With a simplified model taking into consideration only one productive trait, we were able to recognise issues with the available information on mitochondrial DNA and its evolutionary history. The bottlenecks highlighted by this project indicate the need for more accurate and detailed information regarding the molecular profile of the mitochondrial DNA and its population genetics, along with better methods to perform coalescent simulations for inferring populations evolutionary history. The considerably high share of variation attributed to the mitogenome still leads to the conclusion that mitochondrial effect is relevant for breeding practices, despite this study not being able to secure this affirmation. The lack of impact of the mitochondrial effect on genetic gain, despite increasing accuracy for certain selection candidate's categories can raise arguments against its implementation on dairy practices. On the other hand, accounting for mitochondrial effect can bring positive impacts on the selection of females for embryo-transfer and in-vitro fertilization. An expansion of this study considering a better determination of the mitochondrial haplotypes, is needed to allow drawing more assertive conclusions regarding the impact of mitochondrial effect on breeding values estimation in dairy cattle.

References

Aguilar, I., Misztal, I., Johnson, D.L., Legarra, A., Tsuruta, S. & Lawlor, T.J. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score1. *Journal of Dairy Science*, 93 (2), 743–752. https://doi.org/10.3168/jds.2009-2730

Alberts, B. (ed.) (2002). *Molecular biology of the cell*. 4th ed. New York: Garland Science.

Allio, R., Donega, S., Galtier, N. & Nabholz, B. (2017). Large Variation in the Ratio of Mitochondrial to Nuclear Mutation Rate across Animals: Implications for Genetic Diversity and the Use of Mitochondrial DNA as a Molecular Marker. *Molecular Biology and Evolution*, 34 (11), 2762–2772. https://doi.org/10.1093/molbev/msx197

Bell, B.R., McDaniel, B.T. & Robison, O.W. (1985). Effects of Cytoplasmic Inheritance on Production Traits of Dairy Cattle. *Journal of Dairy Science*, 68 (8), 2038–2051. https://doi.org/10.3168/jds.S0022-0302(85)81066-3

Boettcher, P.J., Freeman, A.E., Johnston, S.D., Smith, R.K., Beitz, D.C. & McDaniel, B.T. (1996a). Relationships Between Polymorphism for Mitochondrial Deoxyribonucleic Acid and Yield Traits of Holstein Cows. *Journal of Dairy Science*, 79 (4), 647–654. https://doi.org/10.3168/jds.S0022-0302(96)76410-X

Boettcher, P.J., Kuhn, M.T. & Freeman, A.E. (1996b). Impacts of Cytoplasmic Inheritance on Genetic Evaluations. *Journal of Dairy Science*, 79 (4), 663–675. https://doi.org/10.3168/jds.S0022-0302(96)76412-3

Bouquet, A. & Juga, J. (2013). Integrating genomic selection into dairy cattle breeding programmes: a review. *Animal*, 7 (5), 705–713. https://doi.org/10.1017/S1751731112002248

Brajković, V. (2019). IMPACT OF MITOGENOME ON MILK TRAITS IN CATTLE. DOCTORAL DISSERTATION, Zagreb. https://core.ac.uk/download/pdf/222819316.pdf [2021-04-15]

Brown, D.R., Koehler, C.M., Lindberg, G.L., Freeman, A.E., Mayfield, J.E., Myers, A.M., Schutz, M.M. & Beitz, D.C. (1989). Molecular Analysis of Cytoplasmic Genetic Variation in Holstein Cows. *Journal of Animal Science*, 67 (8), 1926. https://doi.org/10.2527/jas1989.6781926x

Chinnery, P.F., Thorburn, D.R., Samuels, D.C., White, S.L., Dahl, H.-H.M., Turnbull, D.M., Lightowlers, R.N. & Howell, N. (2000). The inheritance of mitochondrial DNA heteroplasmy: random drift, selection or both? *Trends in Genetics*, 16 (11), 500–505. https://doi.org/10.1016/S0168-9525(00)02120-X

Dobler, R., Rogell, B., Budar, F. & Dowling, D.K. (2014). A meta-analysis of the strength and nature of cytoplasmic genetic effects. *Journal of Evolutionary Biology*, 27 (10), 2021–2034. https://doi.org/10.1111/jeb.12468

Falconer, P.D.S. & Mackay, P.T.F.C. (1995). *Introduction to Quantitative Genetics*. 4 edition. Harlow: Longman.

FAO (2020). *Dairy Market Review*. Food and Agriculture Organization of the United Nations. http://www.fao.org/3/ca8341en/CA8341EN.pdf

Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics*, 78 (2), 737–756

García-Ruiz, A., Cole, J.B., VanRaden, P.M., Wiggans, G.R., Ruiz-López, F.J. & Van Tassell, C.P. (2016). Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proceedings of the National Academy of Sciences*, 113 (28), E3995–E4004. https://doi.org/10.1073/pnas.1519061113

Gaynor, C., Gorjanc, G., Wilson, D., Money, D. & Hickey, J.*AlphaSimR: Breeding Program Simulations*.

Gibson, J.P., Freeman, A.E. & Boettcher, P.J. (1997). Cytoplasmic and mitochondrial inheritance of economic traits in cattle. *Livestock Production Science*, 47 (2), 115–124. https://doi.org/10.1016/S0301-6226(96)00023-1

Gissi, C., Iannelli, F. & Pesole, G. (2008). Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity*, 101 (4), 301–320. https://doi.org/10.1038/hdy.2008.62

Goodwin, S., McPherson, J.D. & McCombie, W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17 (6), 333–351. https://doi.org/10.1038/nrg.2016.49

Hadsell, D.L., Olea, W., Wei, J., Fiorotto, M.L., Matsunami, R.K., Engler, D.A. & Collier, R.J. (2011). Developmental regulation of mitochondrial biogenesis and function in the mouse mammary gland during a prolonged lactation cycle. *Physiological Genomics*, 43 (6), 271–285. https://doi.org/10.1152/physiolgenomics.00133.2010

Hill, G.E. (2020). Mitonuclear Compensatory Coevolution. *Trends in Genetics*, 36 (6), 403–414. https://doi.org/10.1016/j.tig.2020.03.002

Jiang, J., Ma, L., Prakapenka, D., VanRaden, P.M., Cole, J.B. & Da, Y. (2019). A Large-Scale Genome-Wide Association Study in U.S. Holstein Cattle. *Frontiers in Genetics*, 10, 412. https://doi.org/10.3389/fgene.2019.00412

Jobling, M.A. & Jobling, M.A. (eds.) (2013). *Human evolutionary genetics*. 2nd ed. New York: Garland Science.

Kennedy, B.W. (1986). A Further Look at Evidence for Cytoplasmic Inheritance of Production Traits in Dairy Cattle. *Journal of Dairy Science*, 69 (12), 3100–3105. https://doi.org/10.3168/jds.S0022-0302(86)80773-1

Ladoukakis, E.D. & Zouros, E. (2017). Evolution and inheritance of animal mitochondrial DNA: rules and exceptions. *Journal of Biological Research-Thessaloniki*, 24 (1), 2. https://doi.org/10.1186/s40709-017-0060-4

Laubenthal, L., Hoelker, M., Frahm, J., Dänicke, S., Gerlach, K., Südekum, K.-H., Sauerwein, H. & Häussler, S. (2016). Mitochondrial DNA copy number and biogenesis in different tissues of early- and late-lactating dairy cows. *Journal of Dairy Science*, 99 (2), 1571–1583. https://doi.org/10.3168/jds.2015-9847

MacLeod, I.M., Larkin, D.M., Lewin, H.A., Hayes, B.J. & Goddard, M.E. (2013). Inferring Demography from Runs of Homozygosity in Whole-Genome Sequence, with Correction for Sequence Errors. *Molecular Biology and Evolution*, 30 (9), 2209–2223. https://doi.org/10.1093/molbev/mst125

McBride, H.M., Neuspiel, M. & Wasiak, S. (2006). Mitochondria: more than just a powerhouse. *Current biology: CB*, 16 (14), R551-560. https://doi.org/10.1016/j.cub.2006.06.054

Mishmar, D., Levin, R., Naeem, M.M. & Sondheimer, N. (2019). Higher Order Organization of the mtDNA: Beyond Mitochondrial Transcription Factor A. *Frontiers in Genetics*, 10, 1285. https://doi.org/10.3389/fgene.2019.01285

Misztal, I., Tsuruta, S., Lourenco, D.A.L., Masuda, Y., Aguilar, I., Vitezica, Z. & Legarra, A. (2014). Manual for BLUPF90 family of programs. *Athens: University of Georgia*, 199

Mrode, R., Ojango, J.M.K., Okeyo, A.M. & Mwacharo, J.M. (2018). Genomic Selection and Use of Molecular Tools in Breeding Programs for Indigenous and Crossbred Cattle in Developing Countries: Current Status and Future Prospects. *Frontiers in Genetics*, 9, 694. https://doi.org/10.3389/fgene.2018.00694

Mrode, R.A. & Thompson, R. (eds.) (2005). *Linear models for the prediction of animal breeding values*. 2. ed Wallingford: CABI. https://doi.org/10.1079/9780851990002.0000

Muller, H.J. (1964). THE RELATION OF RECOMBINATION TO MUTATIONAL ADVANCE. *Mutation Research*, 106, 2–9. https://doi.org/10.1016/0027-5107(64)90047-8

Plummer, M., Best, N., Cowles, K. & Karen, V. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R news*, 6 (1), 7–11

Roger, A.J., Muñoz-Gómez, S.A. & Kamikawa, R. (2017). The Origin and Diversification of Mitochondria. *Current biology: CB*, 27 (21), R1177–R1192. https://doi.org/10.1016/j.cub.2017.09.015

Sato, M. & Sato, K. (2013). Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA. *Biochimica Et Biophysica Acta*, 1833 (8), 1979–1984. https://doi.org/10.1016/j.bbamcr.2013.03.010

Schon, K.R., Ratnaike, T., van den Ameele, J., Horvath, R. & Chinnery, P.F. (2020). Mitochondrial Diseases: A Diagnostic Revolution. *Trends in genetics: TIG*, 36 (9), 702–717. https://doi.org/10.1016/j.tig.2020.06.009

Sharma, P. & Sampath, H. (2019). Mitochondrial DNA Integrity: Role in Health and Disease. *Cells*, 8 (2). https://doi.org/10.3390/cells8020100

Sharma, R., Kishore, A., Mukesh, M., Ahlawat, S., Maitra, A., Pandey, A.K. & Tantia, M.S. (2015). Genetic diversity and relationship of Indian cattle inferred from microsatellite and mitochondrial DNA markers. *BMC Genetics*, 16 (1), 73. https://doi.org/10.1186/s12863-015-0221-0

Southwood, O.I., Kennedy, B.W., Meyer, K. & Gibson, J.P. (1989). Estimation of Additive Maternal and Cytoplasmic Genetic Variances In Animal Models. *Journal of Dairy Science*, 72 (11), 3006–3012. https://doi.org/10.3168/jds.S0022-0302(89)79453-4

Spehar, M., Ferencakovic, M., Brajkovic, V. & Curik, I. (2017). Variance estimation of maternal lineage effect on milk traits in Croatian Holstein cattle. *Agriculturae Conspectus Scientificus*, 82, 263–266

Srirattana, K., McCosker, K., Schatz, T. & St John, J.C. (2017). Cattle phenotypes can disguise their maternal ancestry. *BMC genetics*, 18 (1), 59. https://doi.org/10.1186/s12863-017-0523-5

Weigel, K.A., VanRaden, P.M., Norman, H.D. & Grosu, H. (2017). A 100-Year Review: Methods and impact of genetic selection in dairy cattle-From daughter-dam comparisons to deep learning algorithms. *Journal of Dairy Science*, 100 (12), 10234–10250. https://doi.org/10.3168/jds.2017-12954

Weikard, R. & Kuehn, C. (2018). Different mitochondrial DNA copy number in liver and mammary gland of lactating cows with divergent genetic background for milk production. *Molecular Biology Reports*, 45 (5), 1209–1218. https://doi.org/10.1007/s11033-018-4273-x

Xia, X., Qu, K., Li, F., Jia, P., Chen, Q., Chen, N., Zhang, J., Chen, H., Huang, B. & Lei, C. (2019a). Abundant Genetic Diversity of Yunling Cattle Based on Mitochondrial Genome. *Animals*, 9 (9), 641. https://doi.org/10.3390/ani9090641

Xia, X., Qu, K., Zhang, G., Jia, Y., Ma, Z., Zhao, X., Huang, Y., Chen, H., Huang, B. & Lei, C. (2019b). Comprehensive analysis of the mitochondrial DNA diversity in Chinese cattle. *Animal Genetics*, 50 (1), 70–73. https://doi.org/10.1111/age.12749

Acknowledgements

I would like to thank my supervisors, Dr Gregor Gorjanc for inviting me to perform this research, and Dr Martin Johnsson and Dr Birgit Zumbach for the patience and inputs that led to the conclusion of this Thesis. Also, to acknowledge the members of the HighlanderLab, of which Dr Gregor Gorjanc is the leader, for the help throughout this research and for the useful and enriching discussions.