Bovine Mitochondrial Genomic Diversity and Association of the Mitochondrial Protein Transcriptome to Energy Metabolism and Feed Efficiency

Submitted by

Jigme Dorji

B.V.Sc & AH, 2006, Karnataka Vet., Animal & Fisheries Sciences University, India

Master of Agricultural Science, 2011, The University of Melbourne, Australia

A thesis submitted in total fulfilment of the requirements of the degree of

Doctor of Philosophy

School of Applied Systems Biology

College of Science, Health and Engineering

La Trobe University

Victoria, Australia

May 2021

Table of Contents

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Chapter 2

Chapter 3

Chapter 4

List of Abbreviations

ADP	Adenosine diphosphate
ALT	Alternate allele
AMOVA	Analysis of molecular variance
ADP	Adenosine triphosphate
BCS	Body condition score
BHB	β-hydroxybutyrate
BLAST	Basic local alignment search tool
bp	Base pairs
BRS	Bovine reference sequence (mitochondria)
BUN	Blood urea nitrogen
BW	Bodyweight
CSB	Conserved sequence block
CPM	Counts per million
CV	Coefficient of variation
DE	Differentially expressed
DIM	Days in milk
D-loop	Displacement loop
DMI	Dry matter intake
DNA	Deoxyribonucleic acid
DP	Read depth
DS	Alternate allele dose probability
EB	Energy balance
ECM	Energy correct milk
EDTA	Ethylenediaminetetraacetic acid
ETC	Electron transport chain
FDR	False discovery rate
gBLUP	genomic best linear unbiased prediction
GP	Genotype probability
GRM	Genomic relationship matrix
GS	Gene significance
GT	Genotype
GWAS	Genome wide association studies

H_RFI	High residual feed intake
HD	High density
Het_GT	Heteroplasmic genotype
HSP	Heavy-strand promoter
KEGG	Kyoto encyclopedia of genes and genomes
LD	Low density
LFC	Log2 fold change
LHON	Leber's hereditary optic neuropathy
L_RFI	Low residual feed intake
LSP	Light-strand promoter
ME	Module eigengene
med	median
MELAS	Myopathy, encephalopathy, lactic acidosis stroke-like episodes
MIDD	Maternally inherited diabetes and deafness
MP/MiP	Mitochondrial protein
MT	Mitochondrial
mtDNA	Mitochondrial DNA
MtMiP/MtMP	Mitochondrial genome encoded mitochondrial proteins
NAFLD	Non-alcohol fatty liver disease
NEFA	Non-esterified fatty acid
NGS	Next generation sequencing
NuMiP/NuMP	Nuclear genome encoded mitochondrial proteins
NUMTs	Nuclear mitochondrial sequences
OXPHOS	Oxidative phosphorylation
PBS	Phosphate buffer saline
PC	Principal component
PCA	Principal component analysis
POS	Position (bp) on the mitochondrial genome
PPI	Protein protein interaction
QTL	Quantitative trait loci
r	Coefficient of correlation
REF	Reference allele
RFI	Residual feed intake
RIN	RNA integrity number

RNAseq	RNA sequencing
rRNA	ribosomal RNA
scRNAseq	Single-cell RNA sequencing
SD	Standard deviation
SNP	Single nucleotide polymorphism
TAD	Topologically associating domain
TCA	Tricarboxylic acid
tRNA	transfer RNA
VCF	Variant call format
WBC	White blood cells

Abstract

Mitochondria are semi-autonomous organelles responsible for cellular energy metabolism in eukaryotic cells. Mitochondrial function is affected by mitochondrial proteins (MP) that are encoded by genes from both the mitochondrial and nuclear genomes. The potential influence of the mitochondrial genome on energy metabolism-related traits may warrant its inclusion in animal breeding methodology. It is currently unknown whether the variation on the whole mitochondrial genome is adequate to be exploited in genetic analyses, such as association and genomic prediction studies. Further, genes associated with energy metabolism that may be useful for genomic prediction can be identified with gene expression analyses. Therefore, this thesis aimed to evaluate the whole mitochondrial genome diversity in cattle and use RNA sequencing to profile MP gene expression across tissues and their association with feed efficiency in dairy cattle. Firstly, mitochondrial diversity based on whole mitochondrial genome sequences was high both within and across breeds and variants can be accurately imputed. Secondly, the MP gene expression profile in 29 adult bovine tissues showed over-expression of MP genes in heart and skeletal muscles and under-expression in thyroid and thymus tissues indicating their association with tissue energy demand. The MP genes from both genomes were co-expressed suggesting that they interact. Lastly, differential gene expression analysis in blood of lactating dairy cows revealed that 38 MP genes were under-expressed in the high feed efficiency group and enriched for the oxidative phosphorylation pathway. Overall, this thesis's findings suggest adequate variation in cattle mitochondrial genomes for genomic evaluations. It strengthens our understanding of the roles of differential MP gene expression and coexpression across tissues to suit animal metabolism and energy demands. Altogether, including variants from the mitochondrial genome for genomic prediction is possible considering the genomic variation across animals and the strong signals of MP genes in energy metabolism-related traits.

Statement of Authorship

Except where reference is made in the text of the thesis, this thesis contains no material published elsewhere or extracted in whole or in part from a thesis accepted for the award of any other degree or diploma. No other person's work has been used without due acknowledgment in the main text of the thesis. This thesis has not been submitted for the award of any degree or diploma in any other tertiary institution. The extent and nature of contributions by co-authors are specified in Appendix 1. Each statement of collaborative input has been approved by all co-authors, and their approval verified by Prof. Hans D. Daetwyler in the Authority to Submit Form.

Jigme Dorji

20th of May 2021

Funding Acknowledgements

This project was funded by DairyBio, a joint venture between Dairy Australia (Melbourne, Australia), Agriculture Victoria (Melbourne, Australia) and the Gardiner Foundation (Melbourne, Australia). Jigme Dorji was supported by a DairyBio Graduate Researcher Scholarship, an Australian Government Research Training Program, and Fee Remission Scholarship (LTUFFRS) from La Trobe University.

Animal Ethics Statement

Chapter 3 accessed the genotype data from the 1000 Bull Genome consortium, which has no direct involvement with the animals and, therefore, no need for animal ethics approval. The ethical approval, including the permission to euthanise the animals in Chapter 4, was obtained from the Department of Jobs, Precincts and Regions Ethics Committee (Application No. 2014–23). The ethical approval for the feed efficiency trial and RNA sequencing in Chapter 5 was granted by the Agricultural Research and Extension Animal Ethics Committee (Department of Jobs, Precincts and Regions, Attwood, Victoria, Australia) (Application No. 2013-14).

Acknowledgements

I cannot begin to express my gratitude to my Principal Supervisor, Prof. Hans D. Daetwyler, who has been instrumental in making my PhD program an enriching learning experience through his patience, motivation, guidance and unwavering support. I feel very fortunate to have had the opportunity to work with him. I'm also extremely grateful to Dr. Iona M. MacLeod, Senior Research Scientist, Agriculture Victoria, for her role as Daily Supervisor, who has always been there to guide, support and steepen my learning during the program. I must also thank Prof. Benjamin G. Cocks, Research Director, for insightful comments and suggestions and support as a co-supervisor.

My PhD program wouldn't have been possible without the generous DairyBio Graduate Research Scholarship from the DairyBio Initiative, funded jointly by Agriculture Victoria Research (Melbourne, Australia) and Dairy Australia, and the La Trobe University Full Fee Remission Scholarship. Thank you for the fund support. Further, I would like to take this opportunity to thank my previous funding agencies, the Australian Government, for the support to undertake a Masters program through Endeavour Postgraduate Awards 2010 at the University of Melbourne. The Government of India Scholarships through the Royal Government of Bhutan has enabled me to undertake a Bachelor degree in Bangalore, India and the Royal Government of Bhutan for providing free education till year 12 in Bhutan. Your roles were critical in launching me to the platform to undertake this PhD. To my supervisors, lecturers and teachers in these institutes and schools, I bow to you all.

One of the greatest privileges I had through this PhD program was interacting and learning from some of the leading and emerging world-class scientists at AgriBio and the CompBio group in particular. Thank you to all, and I look forward to work and collaborate with you all in the near future.

I must also thank Ms. Kendra Whiteman (Student coordinator, Agriculture Victoria), Dr. Janet Wheeler (Director, Graduate Research for the School of Applied Systems Biology), Prof. Michael Stear (Chair, Progress Review Committee) and Mr. Tony Francis (DairyBio Mentor)

who ensured that my PhD program was engaging and on tract by providing necessary support.

To my late dad *Ata*, the words you left, “study till the end” has been the inner voice in the pursuit of learning till this stage. I now realise your message is more profound and beyond academics. *Kadrinchey*, for those golden words. To my *Anyi* Chojeymo and *Ajang* Sangay who supported my schooling from Year VI to XII despite own their limitation, *Kadrinchey*, the most beautiful souls. To my amazing mum *Ana*, I am out of words to thank you for your love, care, and struggles to keep us fed, sheltered and clothed in the void left after Ata’s demise. They have not only been a source of my courage and strength to conquer those dark days but also to explore and seek uncharted domains. To *Ata* Chimi, thank you for being with mum and taking care of her. To my two beloved sisters who, while taking care of mom during my studies, had to leave the world too soon upon the beckon of the mighty heaven. Thank you and you were mum’s wonderful daughters. To my brother Kinley with whom, like a pair of oxen, ploughed through the dry and hard ground with each other’s support. Thank you for being there at all times.

Finally, to my beloved wife Tandin, thank you for your patience, love, care, and support and for taking the best care of our kids during the PhD journey. At times, life has been challenging, but we always found a way out when we faced it together. To my son Palchen and lovely daughters Choeyang and Pema, thank you for bearing with my limited or no fun time with you. I hereby, pass on to you the message of your *Memey's* words, “study till the end” and be a good human being.

Chapter 1:

General Introduction

Jigme Dorji[1,2]

[1] School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, Australia, 3083

[2] Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Victoria, Australia, 3083

1.1 Thesis background and objectives

Mitochondria are cellular organelles in eukaryotic cells primarily responsible for energy metabolism. These organelles are semi-autonomous, having their own genome encoding for 37 genes. The mitochondrial genome is haploid but occurs in multiple copies within mitochondria. Unlike nuclear DNA, the mitochondrial genome does not recombine and it is maternally inherited from an oocyte's cytoplasm (i.e. cytoplasmic inheritance) (Chinnery and Hudson, 2013). Thus, the mitochondrial DNA (mtDNA) is used to evaluate maternal lineages and mitochondrial genetic diversity in livestock species (Henkes et al., 2005, Cai et al., 2007, Di Lorenzo et al., 2018), but commonly, only a part or all of the D-loop region of the mtDNA consisting of the hypervariable sub-regions is examined. While complete mitochondrial genome-based diversity evaluations are more comprehensive (Achilli et al., 2008), the choice of D-loop or other regions may have been driven by the cost of sequencing in the past.

The mitochondrial functions depend on a set of proteins encoded by both nuclear and mitochondrial genomes called the mitochondrial proteins (MP). Over a thousand MPs from the nuclear genome (NuMPs) are synthesized and transported into the mitochondria (Schmidt et al., 2010). Further, replication and transcription of the mitochondrial genome depend on the factors and enzymes encoded by the nuclear genome (reviewed in Taanman (1999)), indicating intricate interactions between the two genomes. Thus, mutations in genes from both genomes are implicated in mitochondrial disorders in human, which are characterised by errors in metabolism (Chinnery, 2000, Taylor and Turnbull, 2005). For example, the mitochondrial disease LHON (Leber's hereditary optic neuropathy) has a non-Mendelian mitochondrial mode of inheritance and is associated with mutations in mitochondrial genes (NADH dehydrogenase subunit 1 *ND1*, NADH dehydrogenase subunit 4 *ND4* and NADH dehydrogenase subunit 6 *ND6* of the Complex I of ETC) (Stenton and Prokisch, 2020), while Leigh Syndrome is associated with the mutations of nuclear MP genes

(cytochrome c oxidase assembly factor *SURF1,* cytochrome c oxidase assembly protein *SCO1,* cytochrome c oxidase assembly protein *SCO2,* heme A: farnesyltransferase cytochrome c oxidase assembly factor *COX10,* cytochrome c oxidase assembly factor *COX14*, translational activator of cytochrome c oxidase I *TACO1*) and follow autosomal recessive inheritance (Chinnery, 2015).

While the inheritance of mitochondrial diseases in humans differs from that of quantitative traits, genetic variation in the mitochondrial genome and the interaction with the nuclear genome could influence quantitative traits in cattle. The use of mitochondrial genome in addition to the nuclear genome are likely to provide better resolution of the genetic structure and may affect association studies. According to Manolio et al. (2009), the unaccounted family information may mask the association and diminish the proportion of heritability explained and contributing to the missing heritability in complex diseases. Similarly, the low heritability estimate of complex traits in dairy cattle, in particular feed efficiency and fertility, which concern energy utilization may improve through inclusion of the mitochondrial genome in the analysis. The accuracy of genomic predictions for feed efficiency (e.g. residual feed intake RFI) is low (~ 0.4) (Pryce et al., 2014). Genomic predictions of dairy cattle traits currently exclude the mitochondrial genome (Goddard, 2009), and the exclusion may be mainly for two reasons. First, the association of whole genome mitochondrial variants to phenotypic traits have not been adequately investigated in dairy cattle to date. Previous studies investigating the effect of maternal lines and the association of the mitochondrial variants with quantitative traits are limited and inconclusive (Bell et al., 1985, Schutz et al., 1994). Second, the roles of mitochondria and associated biological pathways on most dairy cattle traits remain poorly understood. Currently, little is known of the whole genome mitochondrial variation in the cattle breeds and the feasibility to undertake mitochondrial genomic evaluations. Thus, one of the immediate needs is to evaluate whole mitochondrial genomic variation both within and

across the breeds.

With the falling cost of sequencing over the years (Wetterstrand, 2021), the use of genome sequencing has become more routine in the genomic selection of dairy cattle. As a result, complete mitochondrial genome sequences are also becoming more readily available as a spin-off of nuclear genome sequencing (i.e. with no additional cost). Therefore, it offers an opportunity to evaluate the diversity based on the whole mitochondrial genome sequences. The complete mitochondrial genome sequence compared to the D-loop region alone is likely to provide more comprehensive information on diversity. Further, the complete genome sequence enables more accurate quantification of genetic variation in the preliminary assessment of using mitochondrial variants to undertake association studies. Furthermore, having both mitochondrial and nuclear genome sequences together are ideal for studying interactions between the mitochondrial protein genes from the two genomes and associating the variants in these proteins to the traits of interest.

Another highly relevant resource to evaluate the role of the mitochondrial genome and its interaction with the nuclear genome on phenotypic traits is the transcriptomic datasets available from the modern sequencing techniques (RNAseq). Transcriptomic data are available across a wide range of tissues and specific tissues relating to production traits and contain the gene transcripts of mitochondrial protein genes from nuclear and mitochondrial genomes. Gene expression analysis offers an alternative approach to studying the association between genes and phenotypic traits (Salleh et al., 2017, van Dam et al., 2018, Scott et al., 2020). Differential gene expression, tissue-specific expression and co-expression enable the elucidation of the biological pathways within the tissue that may affect the trait. Thus, it is a novel approach to consider transcriptome from both the mitochondrial and nuclear genome to analyse the gene expression across tissues and their association to a trait (e.g. RFI).

Given the above, this project undertakes a range of studies that add to the body of knowledge required to determine whether it is feasible in the near future to incorporate mitochondrial genomics in the paradigm of genomic selection.

1.2 Overall thesis objectives

The overarching goal of the research was to understand the mitochondrial genomic diversity and impact on the dairy cattle traits to contribute to the future use of the mitochondrial genome for genomic improvement of livestock. The contribution to the goal was mainly through the three following objectives:

1) to use the entire bovine mitochondrial genome and quantify the diversity within and across breeds,
2) to use transcriptomics for tissue-specific gene expression of mitochondrial proteins, and
3) to identify genes and biological pathways associated with energy metabolism and feed efficiency.

1.3 Thesis outline

This thesis contains six chapters. Chapter 1 sets out the background of the thesis topics and outlines the thesis' main objectives. Chapter 2 reviews the information on mitochondria, mitochondrial genome, and the applications of mitochondrial genetic studies and their gaps. Chapter 3, 4 and 5 present the primary empirical research on bovine mitochondrial genomic diversity, expression of mitochondrial protein genes across a range of bovine tissues, and mitochondrial protein gene expression in two groups of dairy cattle divergent for residual feed intake, respectively. The manuscript for Chapter 3 will be submitted to Scientific Reports, while Chapter 4 and 5 are published in peer-reviewed scientific journals and are incorporated in the publishing journal format. Chapter 6 discusses the advances

and limitations of this PhD research, identifies opportunities, and provide future direction towards understanding the underlying biology and applying mitochondrial genomic data in genomic predictions.

Chapter 2. Literature Review presents the state of knowledge and gaps on mitochondrial genome, including the following topics: mitochondrial genome inheritance, nuclear mitochondrial sequences (NUMTs), mitochondrial DNA transcription, differential gene expression, and mito-nuclear genome interactions. The review also considers the applications and the gaps in mitochondrial genome studies in farm animals, particularly on mitochondrial genetic diversity analysis and the association of mitochondrial DNA to phenotypic traits and animal breeding.

Chapter 3. Cattle maternal diversity inferred from 1,883 *taurine* and *indicine* mitogenomes (Dorji et al. to be submitted to *Scientific Reports*). This chapter evaluates the whole mitochondrial genome diversity within and across modern cattle breeds (nucleotide diversity, haplotypes and haplogroups). Population structure and admixture were analysed using the standard analyses methods for mitochondrial DNA and alternate nuclear genome analysis techniques.

Chapter 4. Expression of mitochondrial protein genes encoded by nuclear and mitochondrial genomes correlate with energy metabolism in dairy cattle (Dorji et al., 2020; *BMC Genomics*). The chapter evaluates the mitochondrial protein gene expression in 29 tissues of two adult cows and six tissues from two foetuses. Gene expression was determined through RNA sequencing. A gene was considered over-expressed or under-expressed in a tissue if the expression was significantly higher or lower than the mean of gene expression across all other tissues. Similarly, co-expression of mitochondrial protein genes encoded by mitochondrial and nuclear genomes was investigated. The over-expressed genes in tissues and the co-expression cluster containing the mitochondrial protein genes were subjected to functional enrichment analysis.

Chapter 5: Mitochondrial protein gene expression and the oxidative phosphorylation pathway associated with feed efficiency and energy balance in dairy cattle (Dorji et al., 2021; *Journal of Dairy Science*). In this chapter, indicators of feed efficiency, namely residual feed intake (RFI) and energy balance were predicted from the dry matter intake. The animals were then ranked based on their RFI values, and the top 14 and bottom 14 ranking animals with RNA sequences were selected as High RFI & Low RFI groups for the study. Gene expression in the blood was determined from RNA sequencing. Differential gene expression between two RFI groups and weighted gene co-expression networks were analyzed. The functions of differentially expressed genes and the genes in co-expression clusters associated with the RFI were annotated. The key genes associated with the RFI were identified and their pathways further refined using a protein-protein interaction network.

Chapter 6: General Discussion contains a summary of the research findings in Chapter 3, 4 and 5. It also presents the limitations within these studies and highlights opportunities in using mitochondrial genome in diversity analysis and implications for effective utilisation of mitochondrial genome diversity in genomic selection.

1.4 Acknowledgements

I would like to thank Prof Hans Daetwyler and Dr Iona MacLeod for their invaluable comments and suggestions on the chapter.

1.5 References

Achilli, A., A. Olivieri, M. Pellecchia, C. Uboldi, L. Colli, N. Al-Zahery, M. Accetturo, M. Pala, B. H. Kashani, U. A. Perego, V. Battaglia, S. Fornarino, J. Kalamati, M. Houshmand, R. Negrini, O. Semino, M. Richards, V. Macaulay, L. Ferretti, H.-J. Bandelt, P. Ajmone-Marsan, and A. Torroni. 2008. Mitochondrial genomes of extinct aurochs survive in domestic cattle. Current Biology 18(4):R157-R158.

Bell, B. R., B. T. McDaniel, and O. W. Robison. 1985. Effects of Cytoplasmic Inheritance on Production Traits of Dairy Cattle1. Journal of Dairy Science 68(8):2038-2051.

Cai, X., H. Chen, C. Lei, S. Wang, K. Xue, and B. Zhang. 2007. mtDNA diversity and genetic lineages of eighteen cattle breeds from Bos taurus and Bos indicus in China. Genetica 131(2):175-183.

Chinnery, P. 2000. Mitochondrial Disorders Overview. in GeneReviews [Internet]. Vol. 2021. 1993-2021 ed. A. H. Adam MP, Pagon RA, et al.,, ed. University of Washington, Seattle, Seattle (WA).

Chinnery, P. F. 2015. Chapter 24 - Mitochondrial Disorders Due to Mutations in the Nuclear Genome. Pages 283-290. Fifth Edition ed. Elsevier Inc.

Chinnery, P. F. and G. Hudson. 2013. Mitochondrial genetics. Br Med Bull 106(1):135-159.

Di Lorenzo, P., H. Lancioni, S. Ceccobelli, L. Colli, I. Cardinali, T. Karsli, M. R. Capodiferro, E. Sahin, L. Ferretti, P. Ajmone Marsan, F. M. Sarti, E. Lasagna, F. Panella, and A. Achilli. 2018. Mitochondrial DNA variants of Podolian cattle breeds testify for a dual maternal origin. PLOS ONE 13(2):e0192567.

Goddard, M. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136(2):245-257.

Henkes, L. E., W. A. Silva Jr, J. C. F. Moraes, and T. A. Weimer. 2005. Mitochondrial control region genetic diversity and maternal ancestry of a Brangus-Ibage cattle populations. Genetics and Molecular Biology 28:60-66.

Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. C. Mackay, S. A. McCarroll, and P. M. Visscher. 2009. Finding the missing heritability of complex diseases. Nature 461(7265):747-753.

Pryce, J. E., W. J. Wales, Y. de Haas, R. F. Veerkamp, and B. J. Hayes. 2014. Genomic selection for feed efficiency in dairy cattle. Animal 8(1):1-10.

Salleh, M. S., G. Mazzoni, J. K. Höglund, D. W. Olijhoek, P. Lund, P. Løvendahl, and H. N. Kadarmideen. 2017. RNA-Seq transcriptomics and pathway analyses reveal potential regulatory genes and molecular mechanisms in high- and low-residual feed intake in Nordic dairy cattle. BMC genomics 18(1):258-258.

Schmidt, O., N. Pfanner, and C. Meisinger. 2010. Mitochondrial protein import: from proteomics to functional mechanisms. Nature Reviews Molecular Cell Biology 11(9):655-667.

Schutz, M. M., A. E. Freeman, G. L. Lindberg, C. M. Koehler, and D. C. Beitz. 1994. The effect of mitochondrial DNA on milk production and health of dairy cattle. Livestock Production Science 37(3):283-295.

Scott, M. A., A. R. Woolums, C. E. Swiderski, A. D. Perkins, B. Nanduri, D. R. Smith, B. B. Karisch, W. B. Epperson, and J. R. Blanton, Jr. 2020. Whole blood transcriptomic analysis of beef cattle at arrival identifies potential predictive molecules and mechanisms that indicate animals that naturally resist bovine respiratory disease. PLOS ONE 15(1):e0227507.

Stenton, S. L. and H. Prokisch. 2020. Genetics of mitochondrial diseases: Identifying mutations to help diagnosis. EBioMedicine 56.

Taanman, J.-W. 1999. The mitochondrial genome: structure, transcription, translation and replication. Biochimica et Biophysica Acta (BBA) - Bioenergetics 1410(2):103-123.

Taylor, R. W. and D. M. Turnbull. 2005. Mitochondrial DNA mutations in human disease. Nat Rev Genet 6(5):389-402.

van Dam, S., U. Võsa, A. van der Graaf, L. Franke, and J. P. de Magalhães. 2018. Gene co-expression analysis for functional classification and gene–disease predictions. Briefings in Bioinformatics 19(4):575-592.

Wetterstrand, K. A. 2021. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP).

Chapter 2:

Literature Review

Jigme Dorji[1,2]

[1] School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, Australia, 3083

[2] Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Victoria, Australia, 3083

Mitochondria have gathered interest over recent years, particularly in medical research because of their role in mitochondrial diseases resulting from their dysfunction. Mitochondria are the only other organelle within a eukaryotic cell besides the nucleus to have their own genome. However, the mitochondrial genome differs from the nuclear counterpart in many aspects, including ploidy, copy number, inheritance, and transcription. The mitochondrial genome closely interacts with the nuclear genome adding to complexity in understanding its function. Thus, the aspects of the mitochondrial genome requiring careful consideration in their use for genomic studies have been reviewed. Further, this review covers the application of mitochondrial DNA studies in farm animals, mainly the mitochondrial genetic diversity and the use of mitochondrial DNA towards animal breeding to set the context of the research.

2.1 Mitochondria as organelles

Mitochondria are membrane-bound organelles found in almost all eukaryotic cells and are primarily responsible for cellular energy metabolism. Up to 95% of the cellular energy is produced in mitochondria, and the organelles are known as the cell's *'powerhouses'* (Marín-García and Goldenthal, 2002, Tzameli, 2012). Other essential functions of the organelle are calcium homeostasis, cell signalling and apoptosis (cellular death) (Green and Reed, 1998, Cottrell and Turnbull, 2000, Giorgi et al., 2012, Tait and Green, 2012).

Morphologically, mitochondria are generally described as rod-shaped with slight variation among the cell types. For example, mitochondria are long and filamentous in fibroblasts, spheres or ovoids in hepatocytes and ovoid or rod-shaped in vascular smooth muscle cells (McCarron et al., 2013). Structurally, mitochondria are double membraned organelles, where the inner membrane is folded into the matrix to form cristae (Figure 1a). The matrix is the site for the tricarboxylic acid (TCA) cycle. The cristae host complexes of proteins that form the electron transport chain (ETC).

The citric acid cycle and the ETCs are intricately linked, where the former generates reduced electron carriers (NADH, $FADH_2$) which are then oxidized to release electrons that are shuttled through the protein complexes of ETC. As a result, an electrochemical gradient is generated by pumping the protons across the mitochondrial inner membrane enabling the ATP synthase (C5) to synthesise ATP (adenosine triphosphate) using ADP (adenosine diphosphate) and phosphate ((Wallace, 2018), reviewed in Zhao et al. (2019)). The whole process is known as oxidative phosphorylation (OXPHOS). Apart from 13 proteins encoded by the mitochondrial genome, all proteins of the ETC (250-300) (Table 1) and TCA cycle originate in the nuclear genome, indicating a direct role of the nuclear genome in primary mitochondrial function.

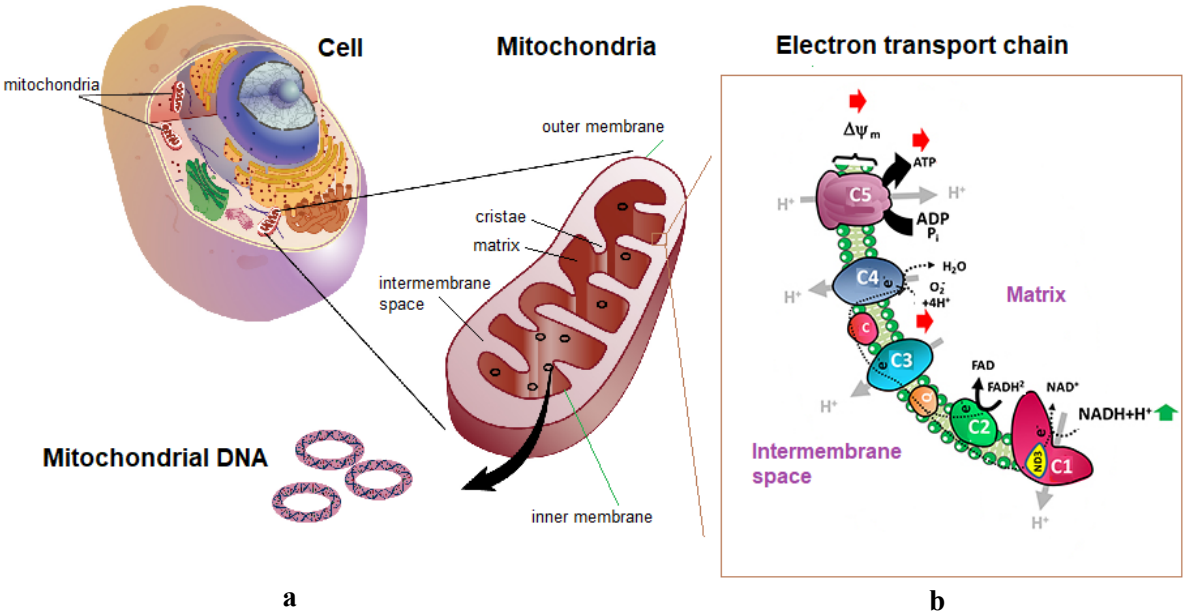


Figure 1. A eukaryotic cell showing the mitochondria, substructures and mitochondrial DNA (a) (source: adapted from https://www.genome.gov/genetics-glossary/Mitochondria) and the electron transport chain protein complexes (C1 to C5); C1 NADH dehydrogenase, C2 Succinate dehydrogenase, C3 Cytochrome c reductase, C4 Cytochrome c oxidase and C5 ATP synthase (b). (Source: adapted from Vial et al., 2019)

Table 1. Components of electron transport chain complexes and contribution from the mitochondrial genome

ETC Complex	Role	No. of subunits	Mito†	References
C1 NADH dehydrogenase	Transfer electrons from matrix NADH to ubiquinone	45	7	Formosa et al. (2018)
C2 Succinate dehydrogenase*	Reduce of FAD to $FADH_2$ (Receives electrons from succinate and then transfers the electrons to FeS clusters)	4	-	Cecchini (2003)
C3 Ubiquinone cytochrome c reductase	Transfer of electrons in Q cycles	11	1	Schägger et al. (1986)
C4 Cytochrome c oxidase	Transfers electrons from Cyt c to the terminal electron acceptor O_2 to generate water	13	3	Kadenbach and Hüttemann (2015)
C5 ATP synthase	Phosphorylate ADP to ATP	10	2	Jonckheere et al. (2012)

* component of both ETC and TCA cycle, Mito† Mitochondrial proteins encoded by mitochondrial genome

2.2 Mitochondrial genome - mitochondria as semi-autonomous organelle

Mitochondria are semi-autonomous organelles containing their own mitochondrial genome, which is haploid and occurs in multiple copies. The mitochondrial genome is small, compact, circular and in the form of a double-stranded helix. The cattle mitochondrial genome is ~16.4 kb in length and was first sequenced in the early 1980s (Anderson et al., 1982). Since then, this mitochondrial genome has been used as a mitochondrial reference genome in almost if not all studies concerning mitochondrial DNA. The mitochondrial genome is referred to as compact because it lacks introns and intergenic sequences. It has unusually small rRNA and tRNA molecules, overlapping genes and a lack of termination codons (Taanman, 1999). The two strands (double-stranded) of mtDNA due to the difference in buoyant densities in a cesium

chloride gradient are designated as heavy (H) and light (L) strands (Welter et al., 1988, Zimmerman et al., 1988).

The mitochondrial genome has a coding region encoding 13 proteins, two rRNAs, and 22 tRNAs (37 genes in total) (Figure 2a). The mitochondrial rRNAs and tRNAs are involved in the synthesis of mitochondrial proteins originating from the mitochondrial genome which constitute the ETC complexes (Figure 1b). The non-coding region is mainly the control region which contains the origin of replication of heavy strand and initiation sites for transcription of mitochondrial DNA (HSP1, HSP2, LSP) (Tully and Levin, 2000).

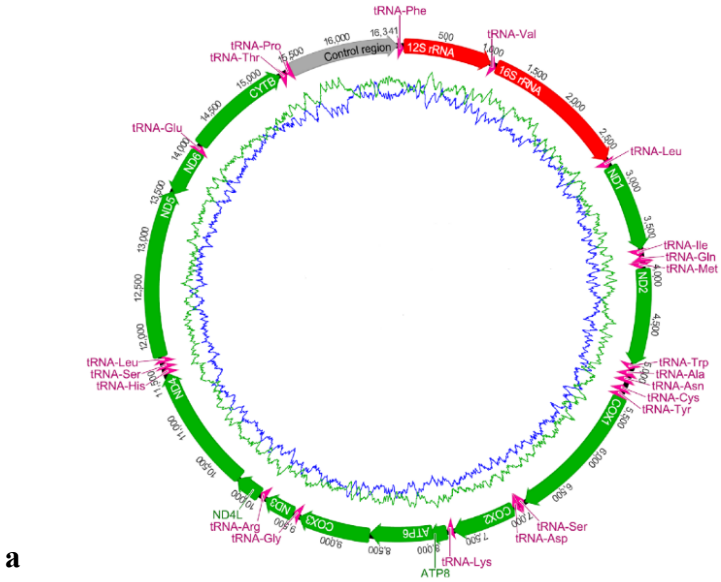


Figure 2a. A complete mitochondrial genome of *Bos taurus* showing protein-coding genes (green), transfer RNAs (pink), ribosomal RNAs (red) and the control region (grey). Source: adapted from Guo et al. (2017). Figure 2b. Mitochondrial control region showing the hypervariable region I and II, central conserved domain and conserved sequence blocks (CSBs). Source: Adapted from Avise (2000)

The hypervariable regions (I and II) of the control region are the most polymorphic region on the mitochondrial genome (Figure 2b) and explain the extensive use of D-loop for mitochondrial genetic diversity studies in livestock.

The number of mitochondrial genomes per mitochondria is nearly constant, while the number of mitochondria per cell differs across the cell types (Robin and Wong, 1988). Thus, the number of mitochondria per cell is the main factor contributing to the variation in the total number of mitochondrial genomes per cell. The number of mitochondria per cell varies greatly with cell types. For example, mitochondria are absent in the erythrocytes, but 50-75 mitochondria are present in a sperm cell with one copy of mtDNA per mitochondrion (Bahr and Engler, 1970, Hecht et al., 1984), 50-100 mitochondria are present per cardiac cell with 1-10 copies of mtDNA per mitochondria (Marín-García and Goldenthal, 2002), 500-4000 mitochondria are found per hepatocyte (Degli Esposti et al., 2012) and 92,500 mitochondria are found in mature oocytes (Pikó and Matsumoto, 1976). The number of mitochondrial genomes in a cell is thus used as a proxy for the number of mitochondria in a cell and is an indicator for mitochondrial activity (Wai et al., 2010).

Among the cells, oocytes are one of the most studied cells related to fertility. The mitochondrial content in oocytes varied during the development and physiological stages and as a consequence, the copy number of mitochondria genomes per oocyte also varied (May-Panloup et al., 2007). In mammals, the number of mitochondrial genomes increases during the early post-implantation of an embryo as a physiological mechanism to distribute mitochondria and mtDNA to dividing cells (Wai et al., 2010). The number of mitochondrial genomes is not correlated to the ATP content in the oocyte a (May-Panloup et al., 2007, Iwata et al., 2011). A low number of mitochondrial genomes is associated with poor oocyte fertilisation and infertility in humans (Reynier et al., 2001). However, the mitochondrial content does not affect the fertilisation in cattle (Iwata et al., 2011). Thus, the association of the mitochondrial

content in the oocyte to fertility is currently not clear.

The number of mitochondrial genomes has been determined through quantitative real-time PCR (Rooney et al., 2015) and, more recently, whole-genome sequencing data (Zhang et al., 2017) also provide an alternative estimate based on the relative abundance of mitochondrial DNA to nuclear DNA. Further, targeted nuclear probes and mitochondrial genome sequencing are reported to produce a more accurate and cost-effective alternative to whole-genome sequencing (Zhou et al., 2020).

2.3 Maternal inheritance of mitochondrial genome

The mitochondrial genome is inherited maternally from the cytoplasm of the oocyte (i.e. cytoplasmic inheritance). There are several mechanisms within the oocyte and spermatozoa that occur during the fertilization which favour the maternal transfer of mitochondria from the oocyte rather than the spermatozoa. For example, the number of mitochondria in the spermatozoa are drastically reduced during spermatogenesis and mitochondria are further eliminated during fertilisation through a series of processes, including endonuclease G digestion, dilution, ubiquitination, DNA digestion, and autophagy (Reynier et al., 2001, Iwata et al., 2011, Sato and Sato, 2011). In contrast, the oocyte has 10^3 fold more mitochondrial DNA than the spermatozoa which is retained during preimplantation embryogenesis (Lee et al., 2012). The maternal inheritance of the mitochondrial genome finds application in the population genetic studies, particularly in the investigation of maternal diversity and lineages (Castro et al., 1998).

However, paternal mitochondrial DNA mutations have been occasionally detected among mitochondrial DNA sequences in humans exhibiting mitochondrial disease-like symptoms (e.g. intolerance to exercises and fatigue) (Schwartz and Vissing, 2002, Luo et al., 2018). Paternal inheritance has been also reported in other species, including sheep (Zhao et al., 2004) and mice (Gyllensten et al., 1991), indicating that paternal

inheritance can occur in other species. The mechanism of preventing paternal inheritance of mitochondrial DNA is robust, but not always foolproof. The paternal mtDNA mutations detected in sheep were not related to disease, suggesting that the inherited paternal mtDNA in animals may exist unnoticed but may potentially affect the production.

2.4 Heteroplasmy – coexistence of wild and mutant mitochondrial genomes

There are multiple mitochondria within a cell, and each mitochondrion has atleast one copy of haploid genome. The result is several copies of mitochondrial genomes per cell. The multiple copies of mitochondrial genomes are generally identical sequences (i.e. homoplasmy). However, mutant and wild-type versions of the genome can co-exist to give rise to a condition known as heteroplasmy (Chinnery and Hudson, 2013). In bovines, the prevalence of heteroplasmy was reported as early as the 1980s from observing the allele switching across the maternal generations and leading to the framing of the hypothesis on the mtDNA genetic bottlenecks in female germlines (reviewed in Wallace and Chalkia, 2013).

Heteroplasmy can arise as a new mutation in a low frequency mixture of mutant and normal genomes (1:1000) which are enriched within individual cells and ultimately can predominate to influence the cellular phenotype (Wallace and Chalkia, 2013). The levels of heteroplasmy vary widely both between and within individuals, organs, and even between cells. The level of tolerance of mutated genomes in *in-vitro* studies can be high (up to 90%, without developing respiratory chain defects), depending on the types of mutation and the tissues (Chinnery et al., 1997, White et al., 1999). However, the heteroplasmic variant levels in the blood are associated with the severity of mitochondrial diseases (e.g. A3242G heteroplasmy) (Grady et al., 2018). In addition, the pathogenic heteroplasmic variant (resulting from a non-synonymous point mutation at nucleotide 8993 in ATP synthase F0 subunit 6 *ATP6*) was maternally transmitted (Holt et al., 1990), indicating the heteroplasmic

variants can be transmitted maternally.

Next-Generation sequencing (NGS) has significantly enhanced the capacity to detect heteroplasmy (Schmitt et al., 2012), but also presents a challenge in distinguishing low-level heteroplasmy from the noise resulting from sequencing errors (Li and Stoneking, 2012).

2.5 Nuclear Mitochondrial Sequences (NUMTs)

Some DNA fragments from the mitochondrial genome exist in the nuclear genome as NUMTs (Nuclear Mitochondrial Sequences) (Lopez et al., 1994), which are considered to have arisen as a part of the evolutionary process. In humans, cattle and horses, the total length of NUMTs was up to 0.0087% (263.48 kb), 0.0023% (69.86 kb) and 0.0018% (54.72 kb) of the respective genomes, with variable distribution and similarity to the mitochondrial genome (Hazkani-Covo et al., 2010). Another study in cattle showed that the size of NUMTs ranged from 37 to 1932 bp, and the homologous identity between NUMTs and their corresponding mtDNA fragments varied from 73 to 98% (i.e. not a whole mitochondrial gene) (Liu and Zhao, 2007). However, the number of NUMTs detected depends on the search strategy (e.g. BLAST, thresholds) employed and the reference genome version (Hazkani-Covo et al., 2010).

The functionality of NUMTs is unknown but they are generally believed to be non-functional considering the observations of multiple internal termination codons and imperfect folding tRNAs for active functioning (Lopez et al., 1994, Liu and Zhao, 2007). The NUMTs can be co-amplified with the mitochondrial DNA and potentially interfere in the interpretation of mtDNA results, specifically in disease diagnosis because of wrongly reporting them as mutations or heteroplasmy (Wallace et al., 1997, Thangaraj et al., 2003, Yao et al., 2008) and phylogenetic studies involving universal primers (Sorenson and Fleischer, 1996, Bensasson et al., 2001). The interference from NUMTs can be minimised by sampling tissue that is

relatively rich in mtDNA, and with careful processing, DNA extraction and purification,reverse transcription, and post-amplification checks (Bensasson et al., 2001). This indicates that samples such as semen (ratio of mitochondrial to nuclear DNA content of almost 1:1) are highly prone to interference from the NUMTs and highlights the importance of downstream bioinformatic analytical procedures.

2.6 Mitochondrial DNA transcription

The process of converting genetic codes in mtDNA to messenger RNA (i.e. transcription) depends on enzymes and transcription factors originating from the nuclear genome. The mitochondrial RNA polymerase (POLRMT) directs transcription of mtDNA and requires TFB2M (transcription factor B2, mitochondrial) and TFAM (transcription factor A, mitochondrial) as two main transcription factors (both genes are in the nuclear genome). The transcription of mtDNA takes place at three promoter sites (2 heavy strand promoters HSP1 and HSP2, and 1 light strand promoter LSP) in the D-loop region (Figure 3).

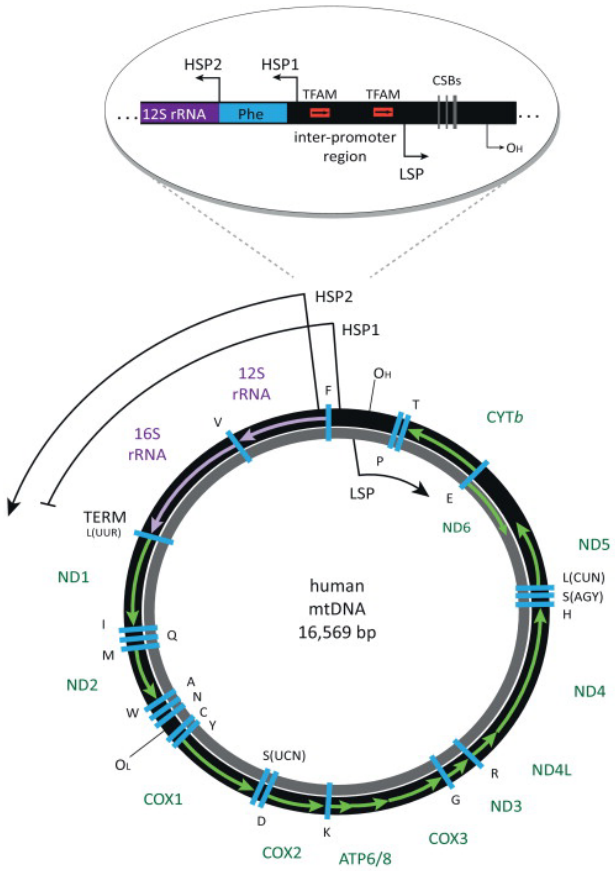


Figure 3. A schematic description of mitochondrial DNA transcription showing the initiation sites and promoters. Source: Bestwick and Shadel (2013)

The transcription at each promoter site is independent and results in polycistronic units. The transcription on HSP1 and HSP2 sites proceeds counter-clockwise. The transcription from HSP1 terminates within the tRNA-Leu (UUR) gene to primarily synthesise two rRNAs. Transcription at HSP2 follows through the rRNAs, 12 protein-coding genes and 14 transfer RNAs (tRNA) genes. On the other hand, LSP transcription proceeds clockwise, starting with RNA primers for replication, through the D-loop, eight tRNA and one mRNA to be terminated by a roadblock protein TERM (encoded by a nuclear gene), which binds downstream of rRNA genes. The long polycistronic RNA chains are cleaved precisely into separate RNA species (Taanman, 1999, Bestwick and Shadel, 2013, Hillen et al., 2018). Currently, the significance of initiation and choice of the individual sites are not understood.

2.7 Mitochondrial protein gene expression analysis

The gene expression analyses across tissues within an individual or a tissue between two phenotypically divergent groups can identify the critical genes at play for the specific function of the tissues as well as within tissue expression differences potentially contributing to the phenotypes (van Dam et al., 2018). The MP gene expression, mainly from the mitochondrial genome across tissues and diseases, has been studied in animals. For example, the differential expression of nine protein genes from the mitochondrial genome across bovine tissues (Harhay et al., 2010) and lowered the expression of *COX1* and *ND6* genes in white blood cells of dogs affected with retinal dysplasia (Appleyard et al., 2006). In another study, all 13 protein genes from the mitochondrial genome were under-expressed in the breast, oesophageal, head and neck, kidney, and liver cancers in humans (Reznik et al., 2017). In pigs, Freeman et al. (2012) used a co-expression analysis to identify a cluster of genes associated with ATP generation and metabolic pathways from the nuclear genome which were highly expressed in the heart. They reported high expression of 11 protein genes from the mitochondrial genome in all tissues. Their study covers the expression of

mitochondrial protein genes mainly related to OXPHOS from two genomes. Other studies also showed MP genes from both genomes among the list of differentially expressed genes in bovine tissues, (e.g. Chamberlain et al., 2015) and in the mammary tissues of the high and low milk protein percentage groups of Holstein cattle at peak lactation and non-lactating period (Li et al., 2016). There were at least six (glycine amidinotransferase *GATM,* glycyl-tRNA synthetase *GARS,* leucine zipper and EF-hand containing transmembrane protein 1 *LETM1,* sideroflexin 2 *SFXN2,* solute carrier family 25 member 22 *SLC25A22* and mitochondrial ribosomal protein L15 *MRPL15*) and 17 MP genes that were differentially expressed between the groups in the peak lactation and non-lactating period, respectively.

2.8 Mito-nuclear interactions

The biological function of eukaryotic cells depends on interactions that arehighlycoevolved between mitochondrial and nuclear genomes (Rand et al., 2004, Levin et al., 2014). Mitochondrial maintenance and function are dependent on proteins encoded by the nuclear genome. The mitochondrial genome codes for less than 1% of the mitochondrial proteins, and more than 1000 mitochondrial proteins originating from the nuclear genome are translated in the cytoplasm and then imported into the mitochondria (Calvo and Mootha, 2010). Some of the nuclear mitochondrial proteins include (but are not limited to) enzymes and factors involved in the electron transport chain (~250-300), mtDNA replication (e.g. mtDNA polymerase γ (POLG1), Twinkle (TWNK)), transcription (TFAM, TFBM1, TFBM2), and enzymes that maintain an appropriate balance of free nucleotides within the mitochondrion (e.g. thymidine phosphorylase (TP), thymidine kinase (TK), deoxyguanosine kinase (DGK) and adenine nucleotide translocator 1 (ANT1) (Chinnery, 2006).

The in vitro interaction between the mitochondrial and nuclear genomes are studied using Cytoplasmic hybrids (cybrids). The cybrids are cells made to have a uniform

nuclear background but different mtDNA. In a study involving cybrids of Holstein, Luxi and yak (cytoplast from the cell of these species fused with p^0 cells derived from a mammary alveolar cell line derived from a Holstein cow) (Wang et al., 2017), the cattle cybrids showed higher oxygen consumption and lowered expression of nuclear mitochondrial protein genes than the primary cells. However, the inter-species cybrids (cattle-yak) showed perturbed mitochondrial function and energy metabolism, suggesting a mito-nuclear mismatch, and indicating a more compatible intra-species mito-nuclear interaction than the inter-species cybrids. The mito-nuclear matching involving cybrids has an opportunity to select female lines for fitness traits in future.

Another form of interaction emerging between the genomes is regulating protein translation from the mitochondrial genome through microRNAs. MicroRNAs are short non-coding RNAs (~22 nucleotides), having a repressive regulatory role in gene expression by either inhibiting translation or degrading target mRNA (Bartel, 2004). The translation of the mitochondrial protein COX1 is suppressed with the overexpression of miR-181c from the nuclear genome (Das et al., 2012).

2.9 Applications of mitochondrial genetic studies in farm animals

The literature on mitochondrial genetic studies in farm animals is predominantly in the domain of mitochondrial population genetics. Another relevant application is in the identification of species of origin in meat and meat products. There have also been attempts to understand the role and use of maternal lines or effects (pre-genomic era) and mitochondrial DNA information (early genomic era) for animal breeding. The pre-genomic era was characterized by estimating cytoplasmic or maternal effects on production traits primarily based on pedigree information. The early genomics period saw attempts to associate polymorphisms of the D-loop, protein-coding genes and rRNAs with phenotypes.

2.9.1 Mitochondrial diversity and phylogenetic studies

Mitochondrial DNA studies in livestock species have been instrumental in tracing ancestry, domestication centres and dispersal (Beja-Pereira et al., 2006, Ajmone-Marsan et al., 2010, Bonfiglio et al., 2012). Such studies have been used to evaluate maternal genetic diversity and characterize livestock species and breeds (FAO, 2011). The mitochondrial haplogroup is the most common approach of linking the individual to the maternal ancestors/progenitors and origins associated with the domestication sites. A haplogroup is a group of similar haplotypes as defined by combinations of a predefined set of variants in mitochondrial DNA inherited from a common ancestor (Wallace and Chalkia, 2013).

Traditionally, the D-loop is the most extensively used region of the mitochondrial DNA in evaluating diversity within livestock species due to its hypervariability (Table 2). The high cost of whole mitochondrial genome sequencing in the past favoured the use of just this region. Occasionally, the protein-coding and rRNAs genes were also used (Table 2). Longer sequence and whole mitochondrial genome sequences provided better molecular diversity and phylogeographic resolution than shorter segments (Torroni et al., 2006, Achilli et al., 2008), indicating the advantages of having longer segments. However, with the reducing cost of sequencing over the years (Wetterstrand, 2021), large numbers of animals are routinely sequenced, and the whole mitochondrial genome sequences have become readily available. The use of whole mitochondrial genome sequences in studies can present a more comprehensive picture of mitochondrial diversity specially within breeds and subpopulations.

A further application for mtDNA is to distinguish between species by barcoding a short region of the mtDNA *COX1* gene. This region is used in the global screening and bio-identification of a wide range of species (e.g. animals, fishes, birds, insects) (Hebert et al., 2003, Ward et al., 2005, Hollingsworth et al., 2009, Jalali et al., 2015).

Table 2. Mitochondrial DNA based genetic diversity studies in cattle and the region of DNA

Breeds (Country)	mtDNA region	N	References
European, Indian, African	Complete D-loop	26	Loftus et al. (1994)
breeds Northeast Asian cattle	D-loop	104	Kim et al. (2003)
Spanish cattle	D-loop	60	Miretti et al. (2004)
Lulu cattle (Nepal)	D-loop	31	Takeda et al. (2004)
Roman and Evolène	D-loop	14	Schlumbaum et al. (2006)
European cattle	D-loop	520	Beja-Pereira et al. (2006)
American Creole cattle	D-loop	454	Lirón et al. (2006)
Bhutanese cattle	D-loop	30	Lin et al. (2007)
Lidia cattle (Spain)	D-loop	517	Cortés et al. (2008)
Southern Europe, Near East	Complete genome	56	Achilli et al. (2008)
Eurasian taurine	D-loop	268	Kantanen et al. (2009)
Zebu cattle	Control region	844	Chen et al. (2010)
Creole cattle	D-loop	413	Ginja et al. (2010)
European, African, American	D-loop	2200	Bonfiglio et al. (2012)
Chikso (Korea)	CYTB	239	Kim et al. (2013)
White Park cattle (British Isles)	D-loop	19	Ludwig et al. (2013)
Wuchuan Black cattle	D-loop	56	Yang et al. (2014b)
Indian cattle	D-loop	170	Sharma et al. (2015)
Polled and Kuchinoshima (Japan)	D-Loop,12S rRNA	57	Mannen et al. (2017)
Gaungxi (China)	Complete genome	56	Xia et al. (2019a)
Chinese cattle	D-loop	1105	Xia et al. (2019b)
12 breeds (China)	16S rRNA	251	Yan et al. (2019)

N = Number of samples

Another closely related field is identifying and tracking species of origin in meat products (Cho et al., 2014, Han et al., 2017). The fact that mtDNA occurs in many copies and is less vulnerable to exonucleases that degrade the DNA molecules, coupled with the availability of reference sequence data for multiple species from use in DNA barcoding (*COX1* region), allows discrimination of the meat from several species (Unseld et al., 1995). The most employed mtDNA regions in species identification are 16S rRNA, 12S rRNA, CYTB and COX1 (Table 3).

Table 3. Common meat contamination and adulteration which are identified using the appropriate region of mitochondrial DNA

Species identifiable	mtDNA region	References
Pork, chicken, sheep, ostrich, horse, beef	CYTB, COXI, 12S rRNA	Kitpipit et al. (2014)
Cow, chicken, turkey, sheep, pig, buffalo, camel and donkey	COX1	Haider et al. (2012)
Beef, buffalo, sheep, goat, pork	CYTB	Kumar et al. (2014)
Mammals, birds, fish, reptiles, coleopteran, crustaceans, cephalopods, avian game species	16S rRNA	Sarri et al. (2014)
Cultured fly, human, and mouse cells; commercial eel, fish, shrimp, pig, cow, rabbit, and chicken tissues, alligator, cat, deer, dog, donkey, duck, equine, pigeon, and turkey	16S rRNA, 12S rRNA	Yang et al. (2014a)
Chicken, duck, turkey, guinea fowl, quail	12S rRNA	Girish et al. (2004)
Cow, buffalo, pig, sheep, and chicken	12S rRNA	Rastogi et al. (2004)
Cattle, buffalo, sheep, goat	16S rRNA	Mane et al. (2013)

2.9.2 Animal selection and breeding

Pre-genomics era - Evaluation of maternal/cytoplasmic effects/maternal lines

There is no evidence of direct use of mtDNA in animal breeding in the literature to date. However, interests in mitochondria have been demonstrated in studies investigating the indirect role of maternal or cytoplasmic effects on milk production, reproduction, and meat quality traits. The genetic component of maternal effects (genetic and environmental) has been estimated in sheep maternal lines (Maniatis and Pollott, 2002).

One of the pilot studies to investigate the effect of cytoplasmic inheritance on production traits in Holstein cattle was by Bell et al. (1985), who found the cytoplasmic

inheritance accounted for 2.0 and 1.8% of the total variation of milk yield and fat yield in the first lactation. The cytoplasmic line was also a significant source of variation for fat and protein (10%) and milk returns (13%) in a subpopulation of Friesian cattle (Dutch Friesian, Holstein Friesian and British Friesian) (Huizinga et al., 1986). These studies suggest maternal effects are an important source of variation for milk production traits. On the contrary, Albuquerque et al. (1998) reported minimal maternal genetic effects and cytoplasmic effects on the phenotypic variances of milk yield, fat yield, and fat percentage in Holstein cattle.

The maternal effects were also evident from genetic evaluation models. Boettcher et al. (1996) observed that the exclusion of cytoplasmic effect in the genetic evaluation resulted in overestimating heritability with an animal model. Similarly, maternal effect manifested as a consistently higher heritability in milk production traits from daughter-dam than paternal half-sibs (reviewed in Gibson et al. (1997)).

In beef cattle, the maternal line was attributed as an important source of variation for birth weights, weaning weights and average daily gain (Tess et al., 1987), and for ultrasonic backfat thickness in Hereford (beef) cattle (Tess and Robison, 1990). On the other hand, Rohrer et al. (1994) found negligible effect (< 0.002%) for the same traits in purebred Brangus cattle. The cytoplasmic effect for weaning weight and age at first calving was also negligible in Afrikaner beef cattle (Neser et al., 2014). The cytoplasmic effect on bull fertility parameters (e.g. semen motility) was marginal and was not a significant source of variation for male fertility traits in Angus bulls (Garmyn et al., 2011).

Overall, the evaluation of maternal or cytoplasmic effects is not without limitations. Kennedy (1986) cautioned that the estimation models used were likely biased to overestimate the cytoplasmic effect. Other challenges include limited sample sizes, computational capacity, lack of robust models to detect mitochondrial genetic effects with certainty, and the practical application of the

information (Gibson et al., 1997). Further, most if not all cytoplasmic or maternal lines were based on the pedigree. Pedigree errors are common in livestock farms (Visscher et al., 2002). Further, limited pedigree depth could affect the correct distinction of maternal line. Despite the shortfalls of correctly assigning maternal lines (in the absence of mtDNA sequencing), non-uniform models in the estimation of effects and reliability of the results, these efforts during the pre-genomic period highlight the interest in understanding the effect of mitochondrial DNA in the form of maternal or cytoplasmic effects on the production and reproduction traits. Thus, the interest in evaluating the effect of mitochondrial DNA on the production traits has existed since the early 1980s (i.e. not very recent) but has not progressed due to the inherent limitations in the approaches. However, the investigation of the effect of maternal lines determined from the mitochondrial DNA (e.g. haplotypes) continues to be an area of interest to this day.

Genomics era - associating mitochondrial polymorphism to economically important traits

Mitochondrial DNA sequencing provides an alternative approach to evaluating the role of mtDNA on production and reproduction traits. Mitochondrial polymorphisms depending on their locations in coding or non-coding region have been postulated to affect phenotypes, mtDNA replication or altered mtDNA gene expression. The tRNA mutations could affect the ability to add amino acids to proteins leading to slow protein production, thereby affecting the ability of mitochondria to produce energy for cells. A mutation in an rRNA, on the other hand, alters the mitochondrial ribosome's function affecting the synthesis of mitochondrial protein. Such changes may affect phenotypic traits by altering the efficiency of the electron transport chain and ATP production or causing various symptoms such as muscle weakness, neurological problems, seizures, hearing loss, and diabetes (Schutz et al., 1994, Krzywanski et al., 2011). While these postulated mechanisms

remain to be tested, they are restrictive to the mitochondrial genome alone and underscore the role of mito-nuclear interaction in mitochondrial function.

Production and feed efficiency traits

The earliest genomic study in dairy cattle tested the association of 17 sequence variants of the mtDNA D-loop with production and health traits in dairy cattle using an animal model (Schutz et al., 1994). This study showed that a single base-pair substitution T16074C had a significant positive association with lactation energy and milk yield, fat yield and solids not fat yields. At the same time, C16231T was negatively associated with these traits. A single transition A169G was associated with increased milk yield by 482 kg, fat by 24 kg and SNF by 51 kg. Although their study had a considerably large dataset (1800 records, 728 cows), the substitution effect or the causality on the yields must be taken with caution for two main reasons. Firstly, they assumed that all animals within a maternal line based on pedigree had identical mtDNA and sampled only one animal for sequencing per maternal line. A similar study (Boettcher et al., 1996) associating rRNA and D-loops sequence variants with production traits employed two animal per maternal lines reported variation in the D-loop sequence within a maternal line which is plausible considering the potential pedigree errors. Secondly, only 17 D-loop sequence variants are considered in the association study, while the D-loop, particularly in the hypervariable regions, are highly polymorphic. Finally, milk production is a quantitative trait and associating a single variant to any variation must be interpreted with caution considering effects from mutations from the coding region of mitochondria and the mito-nuclear interaction of the genomes.

Other studies associated polymorphisms within D-loop and/or a gene within mitochondrial DNA with growth traits and meat quality. Three SNPs within the *ND5* gene in Nanyang cattle were associated with growth traits (Zhang et al., 2008), where the haplotype with substitutions at T12900C, A12923T, and C12924T had higher birth weight, height and length and average daily gain at six months (i.e. had a positive

effect on growth traits). In another study (Biase et al., 2007), a substitution from G to T in *tRNA^{asn}* gene in Nellore cattle was associated with maternal and individual estimated breeding values for body weight. In Japanese black cattle, based on the mixed model procedure, the mtDNA substitution G2232A in the *16S rRNA* gene was reported to have an association with meat quality (longissimus muscle area and marbling) and was suggested as a potential candidate for mitochondrial effect on meat quality (Mannen et al., 2003). In Iberian pigs (Fernández et al., 2008), the polymorphism at C9104T (*COXII) and* A715G (*12S rRNA*) affected fat (+1 g) and protein (-0.5 g) per 100 g of muscle. However, considering only a short segment/region of the mitochondrial genome was used, one can assume that these SNPs were in linkage disequilibrium with other genes.

Despite the plausible role of the mitochondrial function on feed efficiency traits (Herd et al., 2004), the effect of mitochondrial polymorphisms on feed efficiency traits has not been studied. This may be due to the complexity of the trait and the challenges to measuring the phenotypes related to feed efficiency.

Fertility traits

The exact mechanism of how the mitochondrial DNA polymorphism could affect fertility and fitness remains unknown. In general, mtDNA's association with fertility is mainly viewed from the ability to replicate mitochondrial DNA in adequate quantities in the oocyte as there are massive increases in the mtDNA content during egg development (May-Panloup et al., 2007, Benkhalifa et al., 2014). The mtDNA content differs significantly between fertilised and unfertilised oocytes (Reynier et al., 2001, Santos et al., 2006). Thus, polymorphism induced mitochondrial DNA replication and dysfunction and suboptimal function relating to poor biogenesis or poor cytoplasmic maturation may result in poor fertility according to Reynier et al. (2001).

In purebred Hereford and composite multi-breed cattle, calving rate was associated with mtDNA haplotypes based on D-loop and ND5 regions (Sutarno et al., 2002). In

number of oocytes in an ovary, which correlates to the litter size, was significantly associated with polymorphisms, haplotypes and haplogroups in the mtDNA D-loop. A SNP at C323T showed the most prominent and significant differences in oocyte numbers between the C and T allelic groups by 10^5 folds. Similarly, a significant difference in the number of oocytes was reported among the haplogroups (Liu et al., 2019). In Afec-Assaf sheep, the haplogroup based on the complete mitochondrial genome was significantly associated with variation in ewe prolificacy (lambs born/ ewes lambing) ($P < 0.0001$) (Reicher et al., 2012).

Fitness and adaptation traits

The effect of mitochondria, as a centre of energy production and thermogenesis, on the adaptation of animals to extreme environments of temperature and altitude have also been investigated. Two separate studies have compared the polymorphisms of two mitochondrial genes between Tibetan yaks inhabiting the high-altitude rangelands and cattle. Nonsynonymous mutations in *ND1* and *ND2* genes in yaks (Shi et al., 2018) and haplotypes in *ATP 6* (six haplotypes) and *ATP 8* (three haplotypes) found exclusively present in the yak (Wang et al., 2018) were reported to be positively associated with high altitude adaptation in yak compared to cattle. However, the major drawbacks of these studies are the approach of comparing two different species (cattle and yak) for adaptation and using only two genes when whole mitochondrial genome sequencing are available.

2.10 Summary

Mitochondrial DNA, primarily the D-loop region, has been widely used in mitochondrial diversity studies in livestock species. The effects of mitochondrial lines as determined from pedigree records were the cornerstone to study the role of mitochondrial DNA in production, fertility, and fitness traits. Later, the association of mtDNA polymorphism, mainly in the D-loop, with the production traits was studied. The underlying mechanism of how these polymorphisms affect the traits are generally

unknown, but most studies reported them as causal mutations. The approaches used had severe limitations and were of no or limited applicability in animal breeding. Thus, it is timely to revisit the goal of using the mitochondrial genomic information in the genomic improvement of livestock by capitalising on the sequencing and computational capabilities in the genomics era. This would mainly include the use of complete mitochondrial genome sequences for the assessment of mitochondrial genetic diversity, genome-wide association studies and consideration of genomic interactions. Further, the emerging multi-omics tools would also benefit the elucidation of the role of mitochondrial DNA in understanding complex production traits related to energy metabolism in farm animals.

2.11 Acknowledgements

I would like to thank Prof Hans Daetwyler and Dr Iona MacLeod for their invaluable comments and suggestions on the literature review.

2.12 References

Achilli, A., A. Olivieri, M. Pellecchia, C. Uboldi, L. Colli, N. Al-Zahery, M. Accetturo, M. Pala, B. H. Kashani, U. A. Perego, V. Battaglia, S. Fornarino, J. Kalamati, M. Houshmand, R. Negrini, O. Semino, M. Richards, V. Macaulay, L. Ferretti, H.-J. Bandelt, P. Ajmone-Marsan, and A. Torroni. 2008. Mitochondrial genomes of extinct aurochs survive in domestic cattle. Current Biology 18(4):R157-R158.

Ajmone-Marsan, P., J. F. Garcia, and J. A. Lenstra. 2010. On the origin of cattle: How aurochs became cattle and colonized the world. Evolutionary Anthropology: Issues, News, and Reviews 19(4):148-157.

Albuquerque, L. G., J. F. Keown, and L. D. Van Vleck. 1998. Variances of Direct Genetic Effects, Maternal Genetic Effects, and Cytoplasmic Inheritance Effects for Milk Yield, Fat Yield, and Fat Percentage1. Journal of Dairy Science 81(2):544-549.

Anderson, S., M. H. de Bruijn, A. R. Coulson, I. C. Eperon, F. Sanger, and I. G. Young. 1982. Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. J Mol Biol 156(4):683-717.

Appleyard, G. D., G. W. Forsyth, L. M. Kiehlbauch, K. N. Sigfrid, H. L. Hanik, A. Quon, M. E. Loewen, and B. H. Grahn. 2006. Differential mitochondrial DNA and gene

expression in inherited retinal dysplasia in miniature Schnauzer dogs. Investigative ophthalmology & visual science 47(5):1810-1816.

Avise, J. C. 2000. Phylogeography : the history and formation of species. Cambridge, Mass. : Harvard University Press, Cambridge, Mass.

Bahr, G. F. and W. F. Engler. 1970. Considerations of volume, mass, DNA, and arrangement of mitochondria in the midpiece of bull spermatozoa. Experimental Cell Research 60(3):338-340.

Bartel, D. P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116(2):281-297.

Beja-Pereira, A., D. Caramelli, C. Lalueza-Fox, C. Vernesi, N. Ferrand, A. Casoli, F. Goyache, L. J. Royo, S. Conti, M. Lari, A. Martini, L. Ouragh, A. Magid, A. Atash, A. Zsolnai, P. Boscato, C. Triantaphylidis, K. Ploumi, L. Sineo, F. Mallegni, P. Taberlet, G. Erhardt, L. Sampietro, J. Bertranpetit, G. Barbujani, G. Luikart, and G. Bertorelle. 2006. The origin of European cattle: evidence from modern and ancient DNA. Proc Natl Acad Sci U S A 103(21):8113-8118.

Bell, B. R., B. T. McDaniel, and O. W. Robison. 1985. Effects of Cytoplasmic Inheritance on Production Traits of Dairy Cattle1. Journal of Dairy Science 68(8):2038-2051.

Benkhalifa, M., Y. J. Ferreira, H. Chahine, N. Louanjli, P. Miron, P. Merviel, and H. Copin. 2014. Mitochondria: participation to infertility as source of energy and cause of senescence. The international journal of biochemistry & cell biology 55:60-64.

Bensasson, D., D.-X. Zhang, D. L. Hartl, and G. M. Hewitt. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends in Ecology & Evolution 16(6):314-321.

Bestwick, M. L. and G. S. Shadel. 2013. Accessorizing the human mitochondrial transcription machinery. Trends in Biochemical Sciences 38(6):283-291.

Biase, F. H., F. V. Meirelles, R. Gunski, P. A. Vozzi, L. A. F. Bezerra, R. A. Vila, A. J. M. Rosa, R. B. Lôbo, and L. Martelli. 2007. Mitochondrial DNA single nucleotide polymorphism associated with weight estimated breeding values in Nelore cattle (Bos indicus). Genet. Mol. Biol 30(4):1058-1063.

Boettcher, P. J., M. T. Kuhn, and A. E. Freeman. 1996. Impacts of Cytoplasmic Inheritance on Genetic Evaluations1. Journal of Dairy Science 79(4):663-675.

Bonfiglio, S., C. Ginja, A. De Gaetano, A. Achilli, A. Olivieri, L. Colli, K. Tesfaye, S. H. Agha, L. T. Gama, F. Cattonaro, M. C. T. Penedo, P. Ajmone-Marsan, A. Torroni, and L. Ferretti. 2012. Origin and Spread of Bos taurus: New Clues from Mitochondrial Genomes Belonging to Haplogroup T1. PLOS ONE 7(6):e38601.

Calvo, S. E. and V. K. Mootha. 2010. The mitochondrial proteome and human disease. Annu Rev Genomics Hum Genet 11:25-44.

Castro, J. A., A. Picornell, and M. Ramon. 1998. Mitochondrial DNA: a tool for populational genetics studies. Int Microbiol 1(4):327-332.

Cecchini, G. 2003. Function and Structure of Complex II of the Respiratory Chain. Annual Review of Biochemistry 72(1):77-109.

Chamberlain, A. J., C. J. Vander Jagt, B. J. Hayes, M. Khansefid, L. C. Marett, C. A. Millen, T. T. T. Nguyen, and M. E. Goddard. 2015. Extensive variation between tissues in allele specific expression in an outbred mammal. BMC Genomics 16(1):993.

Chen, S., B. Z. Lin, M. Baig, B. Mitra, R. J. Lopes, A. M. Santos, D. A. Magee, M. Azevedo, P. Tarroso, S. Sasazaki, S. Ostrowski, O. Mahgoub, T. K. Chaudhuri, Y. P. Zhang, V. Costa, L. J. Royo, F. Goyache, G. Luikart, N. Boivin, D. Q. Fuller, H. Mannen, D. G. Bradley, and A. Beja-Pereira. 2010. Zebu cattle are an exclusive legacy of the South Asia neolithic. Molecular biology and evolution 27(1):1-6.

Chinnery, P. F. 2006. Mitochondrial DNA in Homo Sapiens. Pages 3-15 in Human Mitochondrial DNA and the Evolution of Homo sapiens. H.-J. Bandelt, V. Macaulay, and M. Richards, ed. Springer Berlin Heidelberg, Berlin, Heidelberg.

Chinnery, P. F., N. Howell, R. N. Lightowlers, and D. M. Turnbull. 1997. Molecular pathology of MELAS and MERRF. The relationship between mutation load and clinical phenotypes. Brain 120(10):1713-1721.

Chinnery, P. F. and G. Hudson. 2013. Mitochondrial genetics. Br Med Bull 106(1):135-159.

Cho, A.-R., H.-J. Dong, and S. Cho. 2014. Meat Species Identification using Loop-mediated Isothermal Amplification Assay Targeting Species-specific Mitochondrial DNA. Korean J Food Sci Anim Resour 34(6):799-807.

Cortés, O., I. Tupac-Yupanqui, S. Dunner, M. A. García-Atance, D. García, J. Fernández, and J. Cañón. 2008. Ancestral matrilineages and mitochondrial DNA diversity of the Lidia cattle breed. Animal genetics 39(6):649-654.

Cottrell, D. A. and D. M. Turnbull. 2000. Mitochondria and ageing. Current Opinion in Clinical Nutrition & Metabolic Care 3(6).

Das, S., M. Ferlito, O. A. Kent, K. Fox-Talbot, R. Wang, D. Liu, N. Raghavachari, Y. Yang, S. J. Wheelan, E. Murphy, and C. Steenbergen. 2012. Nuclear miRNA regulates the mitochondrial genome in the heart. Circ Res 110(12):1596-1603.

Degli Esposti, D., J. Hamelin, N. Bosselut, R. Saffroy, M. Sebagh, A. Pommier, C. Martel, and A. Lemoine. 2012. Mitochondrial roles and cytoprotection in chronic liver injury. Biochemistry research international 2012:387626.

FAO. 2011. Molecular genetic characterization of animal genetic resources. FAO Animal Production and Health Guidelines. No. 9. FAO. Published in 2011, 85 pp. ISBN: 978-92-5-107032-1. Available at http://www.fao.org/docrep/014/i2413e/i2413e00.pdf. Pages 114-114 in Animal Genetic Resources/Ressources génétiques animales/Recursos genéticos animales. Vol. 49. 2011/12/01 ed. Cambridge University Press.

Fernández, A. I., E. Alves, A. Fernández, E. de Pedro, M. A. López-García, C. Ovilo, M. C. Rodríguez, and L. Silió. 2008. Mitochondrial genome polymorphisms associated with longissimus muscle composition in Iberian pigs. J Anim Sci 86(6):1283-1290.

Formosa, L. E., M. G. Dibley, D. A. Stroud, and M. T. Ryan. 2018. Building a complex complex: Assembly of mitochondrial respiratory chain complex I. Seminars in Cell & Developmental Biology 76:154-162.

Freeman, T. C., A. Ivens, J. K. Baillie, D. Beraldi, M. W. Barnett, D. Dorward, A. Downing, L. Fairbairn, R. Kapetanovic, S. Raza, A. Tomoiu, R. Alberio, C. Wu, A. I. Su, K. M. Summers, C. K. Tuggle, A. L. Archibald, and D. A. Hume. 2012. A gene expression atlas of the domestic pig. BMC Biology 10(1):90.

Garmyn, A. J., D. W. Moser, R. A. Christmas, and J. M. Bormann. 2011. Estimation of genetic parameters and effects of cytoplasmic line on scrotal circumference and semen quality traits in Angus bulls. Journal of Animal Science 89(3):693-698.

Gibson, J. P., A. E. Freeman, and P. J. Boettcher. 1997. Cytoplasmic and mitochondrial inheritance of economic traits in cattle. Livestock Production Science 47(2):115-124.

Ginja, C., M. C. T. Penedo, L. Melucci, J. Quiroz, O. R. Martínez López, M. A. Revidatti, A. Martínez-Martínez, J. V. Delgado, and L. T. Gama. 2010. Origins and genetic diversity of New World Creole cattle: inferences from mitochondrial and Y chromosome polymorphisms. Animal genetics 41(2):128-141.

Giorgi, C., C. Agnoletto, A. Bononi, M. Bonora, E. De Marchi, S. Marchi, S. Missiroli, S. Patergnani, F. Poletti, A. Rimessi, J. M. Suski, M. R. Wieckowski, and P. Pinton. 2012. Mitochondrial calcium homeostasis as potential target for mitochondrial medicine. Mitochondrion 12(1):77-85.

Girish, P. S., A. S. Anjaneyulu, K. N. Viswas, M. Anand, N. Rajkumar, B. M. Shivakumar, and S. Bhaskar. 2004. Sequence analysis of mitochondrial 12S rRNA gene can identify meat species. Meat science 66(3):551-556.

Grady, J. P., S. J. Pickett, Y. S. Ng, C. L. Alston, E. L. Blakely, S. A. Hardy, C. L. Feeney, A. A. Bright, A. M. Schaefer, G. S. Gorman, R. J. Q. McNally, R. W. Taylor, D. M. Turnbull, and R. McFarland. 2018. mtDNA heteroplasmy level and copy number indicate disease burden in m.3243A>G mitochondrial disease. EMBO Molecular Medicine 10(6):e8262.

Green, D. R. and J. C. Reed. 1998. Mitochondria and Apoptosis. Science (New York, N.Y.) 281(5381):1309-1312.

Guo, X., X. Ding, X. Wu, P. Bao, L. Xiong, P. Yan, and J. Pei. 2017. Complete mitochondrial genome of Anxi cattle (Bos taurus). Conservation Genetics Resources 10:393-395.

Gyllensten, U., D. Wharton, A. Josefsson, and A. C. Wilson. 1991. Paternal inheritance of mitochondrial DNA in mice. Nature 352(6332):255-257.

Haider, N., I. Nabulsi, and B. Al-Safadi. 2012. Identification of meat species by PCR-RFLP of the mitochondrial COI gene. Meat science 90(2):490-493.

Han, S.-H., H.-S. Oh, and I.-C. Cho. 2017. Identifying the species of origin in commercial sausages in South Korea. Journal of Applied Animal Research 45(1):179-184.

Harhay, G. P., T. P. L. Smith, L. J. Alexander, C. D. Haudenschild, J. W. Keele, L. K. Matukumalli, S. G. Schroeder, C. P. Van Tassell, C. R. Gresham, S. M. Bridges, S. C. Burgess, and T. S. Sonstegard. 2010. An atlas of bovine gene expression reveals novel distinctive tissue characteristics and evidence for improving genome annotation. Genome Biology 11(10):R102.

Hazkani-Covo, E., R. M. Zeller, and W. Martin. 2010. Molecular Poltergeists: Mitochondrial DNA Copies (numts) in Sequenced Nuclear Genomes. PLOS Genetics 6(2):e1000834.

Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. deWaard. 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences 270(1512):313-321.

Hecht, N. B., H. Liem, K. C. Kleene, R. J. Distel, and S.-m. Ho. 1984. Maternal inheritance of the mouse mitochondrial genome is not mediated by a loss or gross alteration of the paternal mitochondrial DNA or by methylation of the oocyte mitochondrial DNA. Developmental Biology 102(2):452-461.

Herd, R. M., V. H. Oddy, and E. C. Richardson. 2004. Biological basis for variation in residual feed intake in beef cattle. 1. Review of potential mechanisms. Australian journal of experimental agriculture 44(4-5):423-430.

Hillen, H. S., D. Temiakov, and P. Cramer. 2018. Structural basis of mitochondrial transcription. Nature Structural & Molecular Biology 25(9):754-765.

Hollingsworth, P. M., L. L. Forrest, J. L. Spouge, M. Hajibabaei, S. Ratnasingham, M. van der Bank, M. W. Chase, R. S. Cowan, D. L. Erickson, A. J. Fazekas, S. W. Graham, K. E. James, K.-J. Kim, W. J. Kress, H. Schneider, J. van AlphenStahl, S. C. H. Barrett, C. van den Berg, D. Bogarin, K. S. Burgess, K. M. Cameron, M. Carine, J. Chacón, A. Clark, J. J. Clarkson, F. Conrad, D. S. Devey, C. S. Ford, T. A. J. Hedderson, M. L. Hollingsworth, B. C. Husband, L. J. Kelly, P. R. Kesanakurti, J. S. Kim, Y.-D. Kim, R. Lahaye, H.-L. Lee, D. G. Long, S. Madriñán, O. Maurin, I. Meusnier, S. G. Newmaster, C.-W. Park, D. M. Percy, G. Petersen, J. E. Richardson, G. A. Salazar, V. Savolainen, O. Seberg, M. J. Wilkinson, D.-K. Yi, and D. P. Little. 2009. A DNA barcode for land plants. Proceedings of the National Academy of Sciences 106(31):12794.

Holt, I. J., A. E. Harding, R. K. Petty, and J. A. Morgan-Hughes. 1990. A new mitochondrial disease associated with mitochondrial DNA heteroplasmy. Am J Hum Genet 46(3):428-433.

Huizinga, H. A., S. Korver, B. T. McDaniel, and R. D. Politiek. 1986. Maternal effects due to cytoplasmic inheritance in dairy cattle. Influence on milk production and reproduction traits. Livestock Production Science 15(1):11-26.

Iwata, H., H. Goto, H. Tanaka, Y. Sakaguchi, K. Kimura, T. Kuwayama, and Y. Monji. 2011. Effect of maternal age on mitochondrial DNA copy number, ATP content and IVF outcome of bovine oocytes. Reproduction, Fertility and Development 23(3):424-432.

Jalali, S. K., R. Ojha, and T. Venkatesan. 2015. DNA Barcoding for Identification of Agriculturally Important Insects. Pages 13-23 in New Horizons in Insect Science: Towards Sustainable Pest Management. A. K. Chakravarthy, ed. Springer India, New Delhi.

Jonckheere, A. I., J. A. M. Smeitink, and R. J. T. Rodenburg. 2012. Mitochondrial ATP synthase: architecture, function and pathology. Journal of Inherited Metabolic Disease 35(2):211-225.

Kadenbach, B. and M. Hüttemann. 2015. The subunit composition and function of mammalian cytochrome c oxidase. Mitochondrion 24:64-76.

Kantanen, J., C. J. Edwards, D. G. Bradley, H. Viinalass, S. Thessler, Z. Ivanova, T. Kiselyova, M. Cinkulov, R. Popov, S. Stojanovic, I. Ammosov, and J. Vilkki. 2009.

Maternal and paternal genealogy of Eurasian taurine cattle (Bos taurus). Heredity 103(5):404-415.

Kennedy, B. W. 1986. A further look at evidence for cytoplasmic inheritance of production traits in dairy cattle. J Dairy Sci 69(12):3100-3105.

Kim, J.-H., M. J. Byun, M.-J. Kim, S. W. Suh, Y.-G. Ko, C. W. Lee, K.-S. Jung, E. S. Kim, D. J. Yu, W. H. Kim, and S.-B. Choi. 2013. mtDNA Diversity and Phylogenetic State of Korean Cattle Breed, Chikso. Asian-Australas J Anim Sci 26(2):163-170.

Kim, K.-I., J.-H. Lee, S.-S. Lee, and Y.-H. Yang. 2003. Phylogenetic Relationships of Northeast Asian Cattle to Other Cattle Populations Determined Using Mitochondrial DNA D-Loop Sequence Polymorphism. Biochemical Genetics 41(3):91-98.

Kitpipit, T., K. Sittichan, and P. Thanakiatkrai. 2014. Direct-multiplex PCR assay for meat species identification in food products. Food Chemistry 163:77-82.

Krzywanski, D. M., D. R. Moellering, J. L. Fetterman, K. J. Dunham-Snary, M. J. Sammy, and S. W. Ballinger. 2011. The mitochondrial paradigm for cardiovascular disease susceptibility and cellular function: a complementary concept to Mendelian genetics. Laboratory investigation; a journal of technical methods and pathology 91(8):1122-1135.

Kumar, D., S. P. Singh, N. S. Karabasanavar, R. Singh, and V. Umapathi. 2014. Authentication of beef, carabeef, chevon, mutton and pork by a PCR-RFLP assay of mitochondrial cytb gene. J Food Sci Technol 51(11):3458-3463.

Lee, H.-S., H. Ma, Rita C. Juanes, M. Tachibana, M. Sparman, J. Woodward, C. Ramsey, J. Xu, E.-J. Kang, P. Amato, G. Mair, R. Steinborn, and S. Mitalipov. 2012. Rapid Mitochondrial DNA Segregation in Primate Preimplantation Embryos Precedes Somatic and Germline Bottleneck. Cell Reports 1(5):506-515.

Levin, L., A. Blumberg, G. Barshad, and D. Mishmar. 2014. Mito-nuclear co-evolution: the positive and negative sides of functional ancient mutations. Front Genet 5:448.

Li, C., W. Cai, C. Zhou, H. Yin, Z. Zhang, J. J. Loor, D. Sun, Q. Zhang, J. Liu, and S. Zhang. 2016. RNA-Seq reveals 10 novel promising candidate genes affecting milk protein concentration in the Chinese Holstein population. Scientific Reports 6(1):26813.

Li, M. and M. Stoneking. 2012. A new approach for detecting low-level mutations in next-generation sequence data. Genome biology 13(5):R34-R34.

Lin, B.-Z., S. Odahara, S. Sasazaki, Y. Yamamoto, T. Namikawa, K. Tanaka, T. Dorji, G. Tshering, F. Mukai, and H. Mannen. 2007. Genetic diversity of Bhutanese Cattle

analyzed by mitochondrial DNA D-loop variation. The journal of animal genetics 35:5-10.

Lirón, J. P., C. M. Bravi, P. M. Mirol, P. Peral-García, and G. Giovambattista. 2006. African matrilineages in American Creole cattle: evidence of two independent continental sources. Animal genetics 37(4):379-382.

Liu, H., W. Shi, D. Wang, and X. Zhao. 2019. Association analysis of mitochondrial DNA polymorphisms with oocyte number in pigs. Reproduction, Fertility and Development 31(4):805-809.

Liu, Y. and X. Zhao. 2007. Distribution of nuclear mitochondrial DNA in cattle nuclear genome. Journal of animal breeding and genetics = Zeitschrift fur Tierzuchtung und Zuchtungsbiologie 124(5):264-268.

Loftus, R. T., D. E. MacHugh, D. G. Bradley, P. M. Sharp, and P. Cunningham. 1994. Evidence for two independent domestications of cattle. Proceedings of the National Academy of Sciences 91(7):2757.

Lopez, J. V., N. Yuhki, R. Masuda, W. Modi, and S. J. O'Brien. 1994. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. Journal of Molecular Evolution 39(2):174-190.

Ludwig, A., L. Alderson, E. Fandrey, D. Lieckfeldt, T. K. Soederlund, and K. Froelich. 2013. Tracing the genetic roots of the indigenous White Park Cattle. Animal genetics 44(4):383-386.

Luo, S., C. A. Valencia, J. Zhang, N.-C. Lee, J. Slone, B. Gui, X. Wang, Z. Li, S. Dell, J. Brown, S. M. Chen, Y.-H. Chien, W.-L. Hwu, P.-C. Fan, L.-J. Wong, P. S. Atwal, and T. Huang. 2018. Biparental Inheritance of Mitochondrial DNA in Humans. Proceedings of the National Academy of Sciences 115(51):13039.

Mane, B. G., S. K. Mendiratta, A. K. Tiwari, and R. Narayan. 2013. Sequence analysis of mitochondrial 16S rRNA gene to identify meat species. Journal of Applied Animal Research 41(1):77-81.

Maniatis, N. and G. E. Pollott. 2002. Nuclear, cytoplasmic, and environmental effects on growth, fat, and muscle traits in suffolk lambs from a sire referencing scheme. Journal of Animal Science 80(1):57-67.

Mannen, H., M. Morimoto, K. Oyama, F. Mukai, and S. Tsuji. 2003. Identification of mitochondrial DNA substitutions related to meat quality in Japanese Black cattle. J Anim Sci 81(1):68-73.

Mannen, H., R. Yonesaka, A. Noda, T. Shimogiri, I. Oshima, K. Katahira, M. Kanemaki, T. Kunieda, Y. Inayoshi, F. Mukai, and S. Sasazaki. 2017. Low mitochondrial DNA

diversity of Japanese Polled and Kuchinoshima feral cattle. Animal science journal = Nihon chikusan Gakkaiho 88(5):739-744.

Marín-García, J. and M. J. Goldenthal. 2002. [The mitochondrial organelle and the heart]. Revista espanola de cardiologia 55(12):1293-1310.

May-Panloup, P., M. F. Chretien, Y. Malthiery, and P. Reynier. 2007. Mitochondrial DNA in the Oocyte and the Developing Embryo. Pages 51-83 in Current Topics in Developmental Biology. Vol. 77. Academic Press.

McCarron, J. G., C. Wilson, M. E. Sandison, M. L. Olson, J. M. Girkin, C. Saunter, and S. Chalmers. 2013. From structure to function: mitochondrial morphology, motion and shaping in vascular smooth muscle. Journal of vascular research 50(5):357-371.

Miretti, M. M., S. Dunner, M. Naves, E. P. Contel, and J. A. Ferro. 2004. Predominant African-derived mtDNA in Caribbean and Brazilian Creole cattle is also found in Spanish cattle (Bos taurus). The Journal of heredity 95(5):450-453.

Neser, F. W. C., J. B. v. Wyk, and M. M. Scholtz. 2014. Evaluation of cytoplasmic genetic effects for production and reproduction traits in Afrikaner cattle. South African Journal of Animal Science 44(5).

Pikó, L. and L. Matsumoto. 1976. Number of mitochondria and some properties of mitochondrial DNA in the mouse egg. Developmental Biology 49(1):1-10.

Rand, D. M., R. A. Haney, and A. J. Fry. 2004. Cytonuclear coevolution: the genomics of cooperation. Trends Ecol Evol 19(12):645-653.

Rastogi, G., M. Dharne, A. Bharde, V. S. Pandav, S. V. Ghumatkar, R. Krishnamurthy, M. S. Patole, and Y. S. Shouche. 2004. Species determination and authentication of meat samples by mitochondrial 12S rRNA gene sequence analysis and conformation-sensitive gel electrophoresis. Current Science 87(9):1278-1281.

Reicher, S., E. Seroussi, J. I. Weller, A. Rosov, and E. Gootwine. 2012. Ovine mitochondrial DNA sequence variation and its association with production and reproduction traits within an Afec-Assaf flock. J Anim Sci 90(7):2084-2091.

Reynier, P., P. May-Panloup, M.-F. Chrétien, C. J. Morgan, M. Jean, F. Savagner, P. Barrière, and Y. Malthièry. 2001. Mitochondrial DNA content affects the fertilizability of human oocytes. Molecular Human Reproduction 7(5):425-429.

Reznik, E., Q. Wang, K. La, N. Schultz, and C. Sander. 2017. Mitochondrial respiratory gene expression is suppressed in many cancers. eLife 6.

Robin, E. D. and R. Wong. 1988. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. J Cell Physiol 136(3):507-513.

Rohrer, G. A., J. F. Taylor, J. O. Sanders, and R. M. Thallman. 1994. Evaluation of line and breed of cytoplasm effects on performance of purebred Brangus cattle. J Anim Sci 72(11):2798-2803.

Rooney, J. P., I. T. Ryde, L. H. Sanders, E. H. Howlett, M. D. Colton, K. E. Germ, G. D. Mayer, J. T. Greenamyre, and J. N. Meyer. 2015. PCR based determination of mitochondrial DNA copy number in multiple species. Methods Mol Biol 1241:23-38.

Santos, T. A., S. El Shourbagy, and J. C. St. John. 2006. Mitochondrial content reflects oocyte variability and fertilization outcome. Fertility and Sterility 85(3):584-591.

Sarri, C., C. Stamatis, T. Sarafidou, I. Galara, V. Godosopoulos, M. Kolovos, C. Liakou, S. Tastsoglou, and Z. Mamuris. 2014. A new set of 16S rRNA universal primers for identification of animal species. Food Control 43:35-41.

Sato, M. and K. Sato. 2011. Degradation of Paternal Mitochondria by Fertilization-Triggered Autophagy in C. elegans Embryos. Science (New York, N.Y.) 334(6059):1141-1144.

Schägger, H., T. A. Link, W. D. Engel, and G. von Jagow. 1986. [22] Isolation of the eleven protein subunits of the bc1 complex from beef heart. Pages 224-237 in Methods in Enzymology. Vol. 126. Academic Press.

Schlumbaum, A., M. Turgay, and J. Schibler. 2006. Near East mtDNA haplotype variants in Roman cattle from Augusta Raurica, Switzerland, and in the Swiss Evolène breed. Animal genetics 37(4):373-375.

Schmitt, M. W., S. R. Kennedy, J. J. Salk, E. J. Fox, J. B. Hiatt, and L. A. Loeb. 2012. Detection of ultra-rare mutations by next-generation sequencing. Proceedings of the National Academy of Sciences 109(36):14508.

Schutz, M. M., A. E. Freeman, G. L. Lindberg, C. M. Koehler, and D. C. Beitz. 1994. The effect of mitochondrial DNA on milk production and health of dairy cattle. Livestock Production Science 37(3):283-295.

Schwartz, M. and J. Vissing. 2002. Paternal Inheritance of Mitochondrial DNA. New England Journal of Medicine 347(8):576-580.

Sharma, R., A. Kishore, M. Mukesh, S. Ahlawat, A. Maitra, A. K. Pandey, and M. S. Tantia. 2015. Genetic diversity and relationship of Indian cattle inferred from microsatellite and mitochondrial DNA markers. BMC genetics 16:73-73.

Shi, Y., Y. Hu, J. Wang, M. A. Elzo, X. Yang, and S. Lai. 2018. Genetic diversities of MT-ND1 and MT-ND2 genes are associated with high-altitude adaptation in yak. Mitochondrial DNA. Part A, DNA mapping, sequencing, and analysis 29(3):485-494.

Sorenson, M. D. and R. C. Fleischer. 1996. Multiple independent transpositions of mitochondrial DNA control region sequences to the nucleus. Proc Natl Acad Sci U S A 93(26):15239-15243.

Sutarno, J. M. Cummins, J. Greeff, and A. J. Lymbery. 2002. Mitochondrial DNA polymorphisms and fertility in beef cattle. Theriogenology 57(6):1603-1610.

Taanman, J.-W. 1999. The mitochondrial genome: structure, transcription, translation and replication. Biochimica et Biophysica Acta (BBA) - Bioenergetics 1410(2):103-123.

Tait, S. W. G. and D. R. Green. 2012. Mitochondria and cell signalling. Journal of Cell Science 125(4):807.

Takeda, K., M. Satoh, S. P. Neopane, B. S. Kuwar, H. D. Joshi, N. P. Shrestha, H. Fujise, M. Tasai, T. Tagami, and H. Hanada. 2004. Mitochondrial DNA analysis of Nepalese domestic dwarf cattle Lulu*. Animal Science Journal 75(2):103-110.

Tess, M. W., C. Reodecha, and O. W. Robison. 1987. Cytoplasmic Genetic Effects on Preweaning Growth and Milk Yield in Hereford Cattle. Journal of Animal Science 65(3):675-684.

Tess, M. W. and O. W. Robison. 1990. Evaluation of cytoplasmic genetic effects in beef cattle using an animal model3. Journal of Animal Science 68(7):1899-1909.

Thangaraj, K., M. B. Joshi, A. G. Reddy, A. A. Rasalkar, and L. Singh. 2003. Sperm Mitochondrial Mutations as a Cause of Low Sperm Motility. Journal of Andrology 24(3):388-392.

Torroni, A., A. Achilli, V. Macaulay, M. Richards, and H. J. Bandelt. 2006. Harvesting the fruit of the human mtDNA tree. Trends in genetics : TIG 22(6):339-345.

Tully, L. A. and B. C. Levin. 2000. Human Mitochondrial Genetics. Biotechnology and Genetic Engineering Reviews 17(1):147-178.

Tzameli, I. 2012. The evolving role of mitochondria in metabolism. Trends in Endocrinology & Metabolism 23(9):417-419.

Unseld, M., B. Beyermann, P. Brandt, and R. Hiesel. 1995. Identification of the species origin of highly processed meat products by mitochondrial DNA sequences. PCR methods and applications 4(4):241-243.

van Dam, S., U. Võsa, A. van der Graaf, L. Franke, and J. P. de Magalhães. 2018. Gene co-expression analysis for functional classification and gene–disease predictions. Briefings in Bioinformatics 19(4):575-592.

Vial, G., D. Detaille, and B. Guigas. 2019. Role of Mitochondria in the Mechanism(s) of Action of Metformin. Frontiers in Endocrinology 10(294).

Visscher, P. M., J. A. Woolliams, D. Smith, and J. L. Williams. 2002. Estimation of Pedigree Errors in the UK Dairy Population using Microsatellite Markers and the Impact on Selection. Journal of Dairy Science 85(9):2368-2375.

Wai, T., A. Ao, X. Zhang, D. Cyr, D. Dufort, and E. A. Shoubridge. 2010. The role of mitochondrial DNA copy number in mammalian fertility. Biology of reproduction 83(1):52-62.

Wallace, D. C. 2018. Mitochondrial genetic medicine. Nature Genetics 50(12):1642-1649.

Wallace, D. C. and D. Chalkia. 2013. Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease. Cold Spring Harb Perspect Biol 5(11):a021220-a021220.

Wallace, D. C., C. Stugard, D. Murdock, T. Schurr, and M. D. Brown. 1997. Ancient mtDNA sequences in the human nuclear genome: A potential source of errors in identifying pathogenic mutations. Proceedings of the National Academy of Sciences 94(26):14900.

Wang, J., Y. Shi, M. A. Elzo, S. Dang, X. Jia, and S. Lai. 2018. Genetic diversity of ATP8 and ATP6 genes is associated with high-altitude adaptation in yak. Mitochondrial DNA. Part A, DNA mapping, sequencing, and analysis 29(3):385-393.

Wang, J., H. Xiang, L. Liu, M. Kong, T. Yin, and X. Zhao. 2017. Mitochondrial haplotypes influence metabolic traits across bovine inter- and intra-species cybrids. Scientific Reports 7(1):4179.

Ward, R. D., T. S. Zemlak, B. H. Innes, P. R. Last, and P. D. N. Hebert. 2005. DNA barcoding Australia's fish species. Philosophical Transactions of the Royal Society B: Biological Sciences 360(1462):1847-1857.

Welter, C., E. Meese, and N. Blin. 1988. Rapid step-gradient purification of mitochondrial DNA. Molecular biology reports 13(2):117-120.

Wetterstrand, K. A. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP).

White, S. L., V. R. Collins, R. Wolfe, M. A. Cleary, S. Shanske, S. DiMauro, H.-H. M. Dahl, and D. R. Thorburn. 1999. Genetic Counseling and Prenatal Diagnosis for the Mitochondrial DNA Mutations at Nucleotide 8993. The American Journal of Human Genetics 65(2):474-482.

Xia, X., K. Qu, F. Li, P. Jia, Q. Chen, N. Chen, J. Zhang, H. Chen, B. Huang, and C. Lei. 2019a. Abundant Genetic Diversity of Yunling Cattle Based on Mitochondrial Genome. Animals 9(9).

Xia, X., K. Qu, G. Zhang, Y. Jia, Z. Ma, X. Zhao, Y. Huang, H. Chen, B. Huang, and C. Lei. 2019b. Comprehensive analysis of the mitochondrial DNA diversity in Chinese cattle. Animal genetics 50(1):70-73.

Yan, L., Y. She, M. A. Elzo, C. Zhang, X. Fang, and H. Chen. 2019. Exploring genetic diversity and phylogenic relationships of Chinese cattle using gene mtDNA 16S rRNA. Archives animal breeding 62(1):325-333.

Yang, L., Z. Tan, D. Wang, L. Xue, M. X. Guan, T. Huang, and R. Li. 2014a. Species identification through mitochondrial rRNA genetic analysis. Sci Rep 4:4089.

Yang, S. L., R. Y. Lin, L. X. Xu, and L. Cheng. 2014b. Analysis of polymorphisms of mitochondrial DNA D-loop and Mc1R gene in Chinese Wuchuan Black cattle. Journal of Applied Animal Research 42(4):487-491.

Yao, Y. G., Q. P. Kong, A. Salas, and H. J. Bandelt. 2008. Pseudomitochondrial genome haunts disease studies. Journal of Medical Genetics 45(12):769.

Zhang, B., H. Chen, L. Hua, C. Zhang, X. Kang, X. Wang, C. Pan, X. Lan, and C. Lei. 2008. Novel SNPs of the mtDNA ND5 gene and their associations with several growth traits in the Nanyang cattle breed. Biochem Genet 46(5-6):362-368.

Zhang, P., B. D. Lehmann, D. C. Samuels, S. Zhao, Y.-Y. Zhao, Y. Shyr, and Y. Guo. 2017. Estimating relative mitochondrial DNA copy number using high throughput sequencing data. Genomics 109(5):457-462.

Zhao, R. Z., S. Jiang, L. Zhang, and Z. B. Yu. 2019. Mitochondrial electron transport chain, ROS generation and uncoupling (Review). Int J Mol Med 44(1):3-15.

Zhao, X., N. Li, W. Guo, X. Hu, Z. Liu, G. Gong, A. Wang, J. Feng, and C. Wu. 2004. Further evidence for paternal inheritance of mitochondrial DNA in the sheep (Ovis aries). Heredity 93(4):399-403.

Zhou, K., Q. Mo, S. Guo, Y. Liu, C. Yin, X. Ji, X. Guo, and J. Xing. 2020. A Novel Next-Generation Sequencing–Based Approach for Concurrent Detection of Mitochondrial DNA Copy Number and Mutation. The Journal of Molecular Diagnostics 22(12):1408-1418.

Zimmerman, E. G., D. R. Akins, J. V. Planz, and M. J. Schurr. 1988. A rapid procedure for isolating mitochondrial DNA. Gene Analysis Techniques 5(5):102-104.

Chapter 3:

Cattle maternal diversity inferred from 1,883 *taurine* and *indicine* mitogenomes

To be submitted to Scientific Reports

Jigme Dorji [1,2], Christy J. Vander Jagt[2], Amanda J. Chamberlain[2], Benjamin G. Cocks[1,2], Iona M. MacLeod[2], Hans D. Daetwyler[1,2]

[1] School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, Australia, 3083

2 Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Victoria, Australia, 3083

Abstract

The study of maternal diversity based on a limited set of mitochondrial regions or variants is a common tool to better understand past demographic events in livestock. Additionally, there is growing evidence of mitochondrial genetic variants having a direct association with a range of phenotypes. It is therefore of interest to explore the full range of maternal diversity that exists across the entire bovine mitogenome. This study evaluated the whole mitochondrial genome diversity of 1,883 animals representing 156 globally important cattle breeds, using conventional haplogrouping as well as methods developed to assess nuclear DNA diversity. Overall, the mitochondrial genomes were diverse with 11 major haplogroups, 1309 haplotypes, nucleotide diversity (π) 0.012 and haplotype diversity 0.99. While most animals fell in the haplogroup that was expected given their breed origins, a small proportion of African *taurine* (3.5%) and *indicine* (1.3%) animals haplogroups were found among the European *taurine* breeds and their composites (N 1302). Additionally, we present evidence confirming a new *indicine* subgroup (I1a, 64 animals) characterised by two private mutations within the I1 haplogroup. Of the breeds studied, only those of Chinese origin belonged to this I1a subgroup. The results from non-conventional mitochondrial population structuring methods such as principal component analysis and hierarchical clustering were consistent with the haplogroup information. Mitogenome haplotype diversity within each breed was high and ranged from 0.932 (Buryat cattle) to 0.998 (Norwegian Red and Holstein cattle). The total genetic variation was attributed mainly to within-breed variance (96.9%), suggesting the potential for trait association studies within a breed. The accuracy of the imputation of missing genotypes was high (99.8%), except for the relatively rare heteroplasmic genotypes. Further, imputed genotypes enabled the prediction of haplogroup with an error rate of < 1.0%. The imputation error rate is in a range that is acceptable for association studies and genomic prediction for complex traits.

Introduction

Based on archaeogenetic evidence, modern day cattle originated from at least two distinct wild aurochsen (Bos primigenius) following two separate domestication events: one in the Fertile Cresent approximately 10,000 years ago and the second in the Indus Valley some 8000 years ago [1-4]. After domestication, cattle spread to Europe with human migration mainly along the Mediterranean coastline and the Danube River [5, 6] to reach the British Isles (6,500 years ago). These cattle populations also expanded to the Iberian Peninsula following the northern coastal region of Africa [5, 7]. Similarly, cattle from the Indus Valley spread to China and South-East Asia [8] and Africa (~2,500–3,500 years ago) [9-11]. The two genetically distinct major cattle sub-species from these two early domestication sites still predominate in modern day cattle as *Bos taurus taurus* and *Bos taurus indicus* along with their widespread crossbreds.

An important part of the molecular evidence for the origin of cattle has been based onmitochondrial DNA (mtDNA) studies. The mitochondrial genome is small (16.34 kb), circular, haploid, non-recombining and maternally inherited [12]. Mitochondrial genome diversity can be described at three levels: nucleotide positions, haplotypes (unique sequences of nucleotides) and haplogroups (higher level of related groups among the haplotypes). Mitochondrial haplotype clustering [13] and mitochondrial haplogroups based on a set of known and pre-defined mutations point to plausible maternal origins and evolutionary history. The compiled haplogroup trees and the corresponding mutations were based on 233 cattle previously used for haplogrouping [14-16] available from GenBank, and are publicly available as a resource called DomeTree [17].

While the mitochondrial genetic diversity of many cattle breeds has been previously characterized [18], there is increasing interest in the role of mitochondrial diversity on important traits in both humans and livestock animals. In humans, mitochondrial mutations have been associated with several condition such as LHON (Leber hereditary

optic neuropathy), MELAS (mitochondrial myopathy, encephalopathy, lactic acidosis and stroke-like episodes), MIDD (maternally inherited diabetes and deafness) as reviewed in [19]. In livestock there is no clear evidence of causality, but mitochondrial haplotypes/mutations have been associated with meat quality [20], litter size [21], and reproductive capacity [22] in pigs, as well as increased milk production in cattle [23]. At a cellular level, mitochondrial haplotypes have been shown to influence DNA methylation and gene expression in embryonic stem cells [13], as well as metabolic traits in porcine and bovine 'cybrids' (cytoplasmic hybrid cell lines containing different cytoplast against uniform nuclear background) [24, 25].

To date, most mitochondrial molecular diversity studies in cattle are primarily evaluated based on the non-coding hypervariable control region (D-loop) or involve limited breeds either within a country or within a region [14, 18, 26, 27]. While the partial or whole D-loop region is informative for population genetics because it is hyper-variable, whole mitochondrial genome sequences are more likely to reflect the full range of mitochondrial genomic diversity. Now that large sequence databases for cattle are available, it is timely to undertake a comprehensive study involving worldwide breeds, countries and continents for a holistic understanding of the mitochondrial landscape in modern cattle. One such database available for cattle is from the 1000 Bull Genomes project [28].

The nuclear DNA variants from the 1000 Bull Genomes project have been extensively used in genomic analyses, particularly for imputation, genome-wide association and genomic predictions in dairy and beef cattle [29-33]. The variants from autosomal chromosomes have also been used to determine population structure and ancestry of bulls [34]. On the other hand, the 1000 Bull Genomes project mitochondrial sequence variants have not been used in mitochondrial diversity studies. While the imputation of mitochondrial variants for population genetics studies are not recommended, it is clearly of interest to empirically test the accuracy of

imputation of mitochondrial variants. Large scale data sets of imputed mitogenomes could contribute in predicting and associating phenotypes to the mitochondrial haplotypes in parallel to the variants from autosomal chromosomes.

The use of mitochondrial variants for mitochondrial diversity from 1000 Bull Genomes project requires close attention to two key aspects of the data. First, the short-read data are not specific to the mitochondrial genome only and some nuclear mitochondrial sequences (NUMTs) can potentially be wrongly aligned back to the mitochondrial genome. This may manifest as heteroplasmy (multiple alleles observed within an animal at a given MT position) but the expectation is that for most tissues, the MT allele reads will be more numerous than NUMT alleles. Thus a read depth filter could help mitigate this issue. However, due to the low number of mitochondria in sperm cells, the wrongly aligned NUMTs will be harder to be distinguished from true mitochondrial reads due to more even read depth as heteroplasmic sites. This necessitates strict quality control and filters to minimize the impact of NUMTs on the analysis. It should also be noted that true heteroplasmy in MT genomes does exist due to multiple mitogenome copies sometimes carrying different mutations (reviewed in [35]). Further, the format in which data is presented (VCF) is not a standard input format for most of the available mtDNA analysis tools. The standard input format for most of the available mtDNA analysis tools. The format conversion must consider the attributes specific to mtDNA (haploid, missing bases, heteroplasmy, and indels). Currently, only a few tools are available to convert MT variants into a more routinely used fasta format, but these tools lack description, particularly on the handling of heteroplasmy. Furthermore, the allocation of the allelic base call at heteroplasmic positions in haploid genotypes needs careful consideration because heteroplasmy is not generally considered in diversity analyses.

We used the mtDNA variants from the 1000 Bull Genomes Project to:

- develop an approach to pre-process and filter the MT sequence data from the 1000 Bull Genomes project to remove samples that may be contaminated with NUMTs,
- evaluate cattle mitochondrial diversity, haplotypes, and haplogroups across and within cattle breeds,
- compare unsupervised clustering techniques to conventional mitochondrial grouping tools using whole mitogenomes, and
- investigate the accuracy of imputation of sporadic missing mitochondrial variants for inclusion in haplogroup assignment.

Materials and Methods

Sequence data and filtering

Our study utilised whole mitochondrial genomes from Run 8 of the 1000 Bull Genomes Project [28]. Run 8 included 4931 animals representing over 200 *taurine* and *indicine* breeds and their crosses. The mitochondrial genome was aligned to the latest Bovine Reference Genome, ARS-UCD1.2_Btau5.0.1Y.fa, which combines ARS_UCD1.2 [36] with the Y chromosome assembly from Btau5.0.1 [37], because the ARS-UCD-1.2 animal was female, and therefore includes the mitochondrial (M) genome version from the ARS-UCD-1.2 assembly. The average read coverage per animal across the mitochondrial genome was 12.34. There were over 6000 mtDNA variants: 5420 SNPs and 836 INDELS. Heteroplasmy (due to a mixture of two or more mitochondrial genomes or NUMT interference) was observed at almost all variant positions (5119 out 5943). The mean number of heteroplasmy per SNP and per animal was 253.0 and 302.2, respectively. The mean number of missing genotypes per SNP was 464.0, and the mean number of missing genotypes per animal was 160.0.

In order to obtain a high quality and reliable dataset for the analysis, we applied quality filters at both site and individual animal levels. The site quality control thresholds used

were similar to those applied in nuclear mtDNA sequences of the 1000 Bull Genomes Project [31]. We applied thresholds per site of: minimum phred-score quality of 30 (Q30), minimum mapping quality of 30 (MQ30), minimum minor allele count of 2 (AC2) and maximum read depth (DP) of mean +3 SDs using VCFtools [38] and BCFtools [39]. This preliminary filtered dataset consisted of 3394 polymorphic sites, including heteroplasmic sites and indels. We then filtered out indels and variant sites with missing genotypes because these are not efficiently handled in conventional mtDNA analysis tools and are generally discarded from the analysis. We imposed an individual animal filter based on average read depth coverage and heteroplasmic sites. The animals with low average read coverage (DP < 10) across all remaining sites were removed. Further, to develop filters to remove animals that may have excessive and/or questionable heteroplasmy due to contamination from NUMTs, animals were evaluated in two groups as

1. males whose DNA samples were either from semen or from unknown tissue (Semen group), and

2. females and males with DNA sampled from known tissues other than semen (Non-semen group).

The distribution of the number of heteroplasmic sites per individual between groups was compared. The heteroplasmic site distribution in the non-semen group approached a maximum of 150 heteroplasmic sites per individual compared to a maximum of over 700 heteroplasmic sites per individual in the semen group (Figure S2). We therefore applied a maximum threshold of 150 heteroplasmic sites per animal, removing about 300 animals from the semen group leaving 1,883 animals remaining in the data set. Further, the allelic ratios of the major to minor alleles at the heteroplasmic positions in the semen group showed an increase following the application of this filter (Figure S3) so that the major allele count was nearly twice that of the minor alleles.

Data processing

The existing mitochondrial DNA analysis tools either require a continuous stretch of mitochondrial DNA sequence from a specific region (D-loop, COX2, CYTB etc) or a whole mitochondrial genome in prescribed formats. We adopted a genotype-based allele assignment approach for the conversion to a homoplasmic variant sequence. Homoplasmic variants (0/0, 1/1, 2/2 etc.) were directly assigned the corresponding alleles, while the heteroplasmic sites (0/1, 0/2,0/3, 1/3 etc.) were assigned a homoplasmic status for the most abundant allele based on read depth. In other words, the allele (reference - REF or alternative - ALT) with a higher read depth was chosen as the most representative base for a sequence at heteroplasmic positions. In cases where the allele read depth of REF and ALT alleles were equal, the ALT allele was chosen as the base for the position in the sequence. It was assumed that this strategy would be more informative of the existing allelic diversity and would help avoid reference bias. In addition, we also generated a complete genome length sequence (16340 bp in fasta format) using bases for variant positions and inserting "N" (missing base) in non-variant positions because this full mitochondrial genome length sequence format was required to predict the haplogroups using traditional tools (maternal origin and lineages).

Analysis

Mitochondrial DNA polymorphism, diversity and haplotypes

The variant sequences (derived from VCF) were used for the description of the overall DNA polymorphism and evaluation of nucleotide and haplotype diversity using DnaSP program [40] for selected breeds ($N \geq 20$ animals). The analysis of molecular variance was conducted using Arlequin 3.5 [41]. We used the maximum likelihood tree implemented in the MEGA X program [42] to derive a phylogenetic tree among the breeds as well as a haplotype network within breed employing median joining tree in the PopART program [43].

Mitochondrial haplogroups

Mitochondrial haplogroups were predicted using the MitoToolPy program [17] using the whole mitochondrial genome sequence (fasta) prepared as described above. The tool aligned the query sequences to the bovine reference sequence V00654 (hereafter referred to as BRS) which was generated using a shotgun DNA sequencing strategy [12]. The tool derived the list of SNPs and compared them to the predetermined list of SNPs specific to haplogroups to assign a haplogroup. The tool then provided a list of variants missing in the query sequence (missing variants) for an assigned haplogroup (where haplogroups are pre-defined by MitoToolPy) and a list of SNPs not in an assigned haplogroup, but present in query sequence as private variants. The private variant output from the tool provides the opportunity to infer a new subgroup within a haplogroup and to annotate the variants specific to a haplogroup and breed.

Additionally, because of the extensive use of D-loop sequences in determining mitochondrial diversity and haplogroups in the past, mitochondrial variant sequences from only the D-loop region were also used to predict the mitochondrial haplogroups in MitoToolPy as a comparison. The outputs from the MitotoolPy (private and missing variants) had slightly altered nucleotide positions due to being aligned to an older reference genome (BRS) incorporated in the software. To enable the annotation of variants to the latest reference (ARS-UCD1.2_Btau5.0.1Y_M.fa, hereafter known as ARS-UCD1.2_M), the haplogrouping variants in the MitoToolPy were lifted over (positions and bases) to ARS-UCD1.2_M, and the reference genome for alignment within the tool was changed to the ARS-UCD1.2_M.fa. The two reference genomes differed in their length (BRS 16338 bp, ARS-UCD1.2_M 16340 bp) resulting from two deletions in the former as well as having nucleotide base differences at 12 positions (Table 1). Briefly, all haplogroup determining variants in MitoToolPy after 222 bp were incremented by +1 up to 588 bp and positions after 588 bp (BRS) by +2 to correct for the two deletions. Further, bases were changed as appropriate, i.e. five variants

among the haplogroup determining variants were the same base as ARS-UCD1.2_M and thus removed as they were no longer variant when lifted over to the ARS-UCD1.2_M (Table S1). The position changes resulting from the manual liftover were confirmed by aligning 218 complete mitochondrial DNA sequences (previously used to derive variants for haplogroups in MitoToolPy available from NCBI under the same accession number) to the ARS-UCD1.2_M and these conformed to Table S1 and showed additional variants (Table S2).

Table 1. Equivalent positions and reference (Ref) alleles differing between ARS-UCD1.2 (ARS) and Bovine reference sequence (BRS) relative to ARS-UCD1.2 and indicating whether variant positions belong to the pre-defined haplogroup variant set as defined in cattleTree_whole.txt file of MitotoolPy.

ARS Position & Ref. allele	**BRS Position & Ref. allele**	**Haplogroup Variant in MitoToolPy**	**Comments**
222 C	- (deletion)	No	deletion at position 222
364 G	363 C	No	
589 C	- (deletion)	No	deletion at position 589
2538 A	2536 C	Yes (2536 A)	ARS and HG_variants have same base
3345 G	3343 C	No	
3387 C	3385 T	No	
3541 A	3539 G	No	
4321 C	4319 T	No	
8190 C	8188 T	Yes (8188 C)	ARS and HG_variants have same base
8712 T	8710 C	No	
9684 C	9682 G	Yes (9682 C)	ARS and HG_variants have same base
12167 C	12165 T	Yes (12165 C)	ARS and HG_variants have same base
13312 C	13310 A	Yes (13310 C)	ARS and HG_variants have same base
15637 T	15635 C	No	

Private variants and annotation

We investigated private variants specific to certain individuals within a haplogroup and within individuals within a breed to investigate whether they may be biologically meaningful. The private variants specific to a particular group within a haplogroup/breed in this study were annotated using SNPeff [44]. The importance of the

coding variants was predicted by SNPeff as being either high (e.g. stop gained), moderate (missense variants) or low (synonymous variant) and non-coding variants were annotated as modifier (e.g. upstream/downstream variants).

Unsupervised clustering

The overall mitochondrial population structure was also investigated through three unsupervised clustering approaches. First, clustering based on principal components was derived from a genomic relationship matrix (GRM) generated from all filtered polymorphic variants. Fasta files were converted to .bed format using Plink ver1.9 [45] and these genotypes were used to generate a haploid GRM (*make-grm-xchr* option in GCTA [46]) to use as input for a principal component analysis (PCA) also completed with GCTA . Principal components (PCs) 1, 2 and 3 were plotted using scatterplot3d [47], and the clustering was interactively visualised using the rgl package in R [48]. Second, we determined the individual ancestry and population structure of the animals using Admixture [49]. The estimate of population subgroups was determined using the Admixture cross-validation errors approach, and then *a priori* population structure was implemented with *k* ranging from 2 to 6, where *k* is the expected number of populations. The third approach was hierarchical clustering implemented in the R package dendextend [50] using a matrix of nucleotide differences between each pair of sequences (calculated using an in-house python script). The hierarchical clusters were implemented at the highest (2 groups) as well as the lowest levels (0 nucleotide difference). To check for the concordance between these three unsupervised clustering methods, the resulting clusters/groups were annotated according to the individual's predicted haplogroups from MitoToolPy.

Imputation of missing genotypes and haplogrouping

The accuracy of imputation of sporadic missing mitochondrial genotypes and the effect of this imputation on haplogroup assignment were investigated. The empirical accuracy of imputation was tested using the filtered sequence dataset that had no

missing genotypes. The 1883 individuals in the filtered data were split into two random groups consisting of 333 (I) and 1550 (II) animals. A random 10% of the genotypes of individuals in Group I were masked (set to missing) at random sites and then imputed using Beagle 4.0 [51] following the *gt* and *ref* options and providing Group I (as *gt*) and Group II (as *ref*) accordingly. The accuracy of imputation in Group I was evaluated as the proportion of imputed genotypes in agreement with original genotypes at the masked and unmasked positions separately. The Beagle estimate of alternate allele dose probability (DS) and genotype probabilities (GP) were used to define the most likely base call at heteroplasmic positions. For example, for an imputed heteroplasmic genotype with a reference and alternate allele of 0|1, 0|2, or 0|3 etc (where 1, 2 & 3 represent alternate alleles for multi-allelic sites), if the DS is < 1, a Ref allele is assigned while DS =1 is assigned the Alt allele. In rare cases where heteroplasmy was imputed as two alternate alleles, when DS=2 (equal probabilities), the base was set to missing 'N' and when DS is < 2, the more frequent allele (summed across genotype probabilities) was assigned as the base for this position. These variant sequences with imputed sporadic missing genotypes were then reconstituted to a full genome sequence in fasta format by adding Ns at other non-variant positions and then used for re-predicting the haplogroups. The extent of agreement between an individual's haplogroup using the real and partially imputed genotypes was examined. The mean accuracies of imputation and the predicted haplogroup were calculated from 50 repeats of this cross-validation (i.e., resampling Group I and Group II animals and following the above steps). The empirical accuracy of imputation was assessed as the concordance between real and imputed genotypes.

Results

General description of variants

The raw variant call dataset was filtered to have high quality SNP genotypes,

and animals with missing genotypes were not included. This not only reduced the and animals with missing genotypes were not included. This not only reduced the overall number of animals and sites substantially but also reduced the number of heteroplasmic genotypes, improved the average read depth and retained higher quality sites (Table 2, Figure S1). However, when we compared the levels of heteroplasmy per individual separately in the Semen and Non-semen tissue groups, heteroplasmy was much higher in the Semen derived samples.

Table 2. Summary of the parameters of raw and filtered variant datasets before and after removing of sites with missing data (Site) and removing both sites and animals with missing data (Site & Ani).

Parameters	**Raw dataset (Unfiltered)**	**Dataset filtered by**	
		Site	**Site & Ani**
No. of Animals (Ani) in dataset	4931	4931	1883
Total No. of POS in dataset	5903	3394	3069
Total No. of POS with at least one Het_GT animal	5201	3394	1227
Mean No. of Ani with Het_GT per POS (med)	253 (5)	388 (12.5)	2 (0)
No. of Ani with at least one Het_GT	3934	3717	712
Mean No. of POS with Het_GT per Ani (med)	302.2 (278)	266 (245)	3.5 (0)
No. of POS_Missing GT	5903	3394	0
Mean No. of Ani with Missing GT per POS (med)	420 (409)	232.3 (175)	0
No. of Ani with Missing GT	3299	2748	0
Mean No. of POS_Missing GT across all Ani (med)	251.3 (7)	159.9 (3)	0
Mean read depth per POS (across all Ani) (med)	284.5 (299.8)	287 (299.9)	699 (723)
Mean read depth per Ani (across all POS) (med)	284 (18.9)	287 (18.9)	699 (597)

Ani=Animal, POS=nucleotide position, GT=genotype, Het_GT Heteroplasmic genotype, med=median, Site=nucleotide position, med=median

We therefore imposed a strong filter based on the maximum number of heteroplasmic sites per individual, to result in a similar distribution of heteroplasmy and allelic ratios in both the Semen and Non-semen groups (Figure S2, S3). Overall, in the final set of 1,883 individuals, the per site heteroplasmy count was considerably reduced and there was a slight improvement in the average read depth coverage in the final dataset (Figure S4).

Mitochondrial haplogroups, population structure and admixture

Haplogroups using MitoToolPy

The haplogroup membership for each of the 1,883 animals in the filtered set was predicted in MitoToolPy using the ARS-UCD1.2_M reference, and the lifted over variants that MitoToolPy uses to define haplogroups (Table S1). MitoToolPy detected 11 major pre-defined haplogroups (I1, I2, T1, T2, T3, T4, T5, T6, P, Q1, Q2) based on variants from the whole genome sequences (16,340 bp). Overall, T3 was the predominant haplogroup (N 1502 animals) with about 15 subgroups within T3. The dominant subgroups were T3 (752) and T3r (547) (Figure S5). In most cases, the predicted haplogroup of each animal was as would be expected based on the breed and sub-species (Table S3). All the African cattle breeds (Ankole, Afrikander, Ndama, Benishangul, Goffa, Kenana, Muturu) were classified as the T1 haplogroup that is fixed in African *taurine* breeds [18]. Generally, the *indicine* cattle breeds belonged to major haplogroup I and modern *taurine* cattle to haplogroup T, although there were some exceptions. As expected, the composite breeds mostly sourced from Australia were unpredictable. Notably, the haplogroups of Brahman cattle (N=18) were mostly T1 (N=12), T3 (N=5) and one *indicus* (I). In some animals of European breed origin and their composites (N=1302), the integration of T1 (3.5%), I1 and I2 (1.3%) haplogroups was also observed (Table 3 and 4). For example, several Holstein (N=5) and Jersey cattle (N=4) from Australia were of T1 origin. This was further confirmed by checking the original haploid genotypes for heteroplasmy across the haplogroup determining positions. For Jersey belonging to T1, the haplogroup determining positions were all homoplasmic (except 1 position in one animal) (Table S4). Altogether, the T1 haplogroup was observed in about 13 European *taurine* breeds and composites. Within Australian sourced cattle, T1 had considerable influence on Holstein, Jersey and composite breeds (36 animals).

Similarly, the I Haplogroup was present in Holstein animals from China (N=5/12), Herefords from New Zealand (N=3/4) and, as expected, in composite *taurus* x *indicus* breeds from Australia 4/12 (Table 4).

Table 3. Prevalence of T1 (African *taurine*) haplogroup in non-African cattle breeds and composites.

Origin of sample	Breed	Sub-species	n/N	Sex
Australia	Angus Lowline	*taurus*	2/2	2F
	Beefmaster	*taurus* X *indicus*	1/2	1M
	Brahman	*indicus* X *taurus*	12/18	1F, 10M, 1U
	Dexter	*taurus*	1/2	1F
	Holstein	*taurus*	5/5	4F, 1M
	Jersey	*taurus*	4/8	4M
	Senepol	*European taurus* X *African taurus*	5/12	5U
	Composite		6/13	6M
Germany	Holstein Red	*taurus*	1/3	1M
France	Blonde d'Aquitaine	*taurus*	2/16	1F, 1M
	Brown Swiss	*taurus*	1/1	1M
Korea	Hanwoo	*taurus*	2/21	2U
Unknown	Holstein	*taurus*	1/67	1F
	Romagnola	*taurus*	2/10	1M, 1U
	San Martinero	*taurus*	1/2	1M
	Limonero	*taurus*	1/9	1U

F= female; M=male; U=unknown; n=No. of animals showing T1 haplogroup, N=No. of animals in a breed sampled from the specified country.

Table 4. Prevalence of *indicine* haplogroup (I) in European *taurine* breeds and composites.

Origin of sample	Breeds	n/N	Sex	Haplogroup
Australia	Composite	4/12	M	I1
	Brahman	1/18	M	I1
	Belted Galloway	1/2	F	I1
China	Holstein	5/12	F	I1
New Zealand	Hereford	1/4	M	I2
	Hereford	2/4	M	I1
Unknown	Shorthorn	1	F	I1

F=female; M=male; U unknown; n=No. of animals showing I haplogroup, N=No. of animals

In the past, sequences from D-loop region (910 bp long) have been extensively used in the prediction of haplogroups [1, 18]. However, using our filtered D-loop genotype data, MitoTool.py could not differentiate between the two major I and T haplogroups likely because some variants used in previous studies were filtered out of our variant set. In our dataset prior to any filtering, there were 206 D-loop variants compared to 153 D-loop variants in the pre-defined set that MitotoolPy uses for prediction of haplogroups but only 87 variants overlapped. Further, in our filtered set, only 60 D-loop variants overlapped with the 153 MitoToolPy D-loop variants, suggesting that this was the main contributing factor resulting in poor resolution of haplogroups using only the D-loop variants. On the other hand, using our filtered set of sequence variants from the non-D-loop region, MitoToolPy could distinguish the major haplogroups (I, T, P and Q) but did not resolve haplogroup sub-levels. For example, the incidence of unresolved haplogroups was more than 60% of the animals between T1 and T3 (1280), and T3 and T4 (N 15). This indicates that the D-loop variants in our set played a key role in defining the sub-haplogroups when used together with the non-D-loop. This is not unexpected because the higher mutation rate in the D-loop region is more likely to resolve the sub-haplogroup levels (i.e. more recently diverged groups).

Principal Component Analysis

The PCA of the GRM derived from all filtered mitochondrial variants (whole sequence) revealed distinct clusters that corresponded to the I, T and Q major haplogroups after annotation with MitoToolPy results (Figure 1). However, sub-clustering within the major haplogroups T and T3 was not entirely resolved, despite the tendency to marginally separate T1 and T2's (Figure S6a), as well as T3 and T3r (Figure S6b).

A GRM of only D-loop variants was also used for PCA and revealed the same two major clusters (T and I, Figure S7a). Within the I cluster, sub-clusters of I1 and I2 were

separated to some extent while T1 and T3 did not separate clearly. Similarly, the variants from the non-D-loop region could segregate T and I haplogroups into separate clusters but did not resolve further into sub-clusters of haplogroups (Figure S7b).

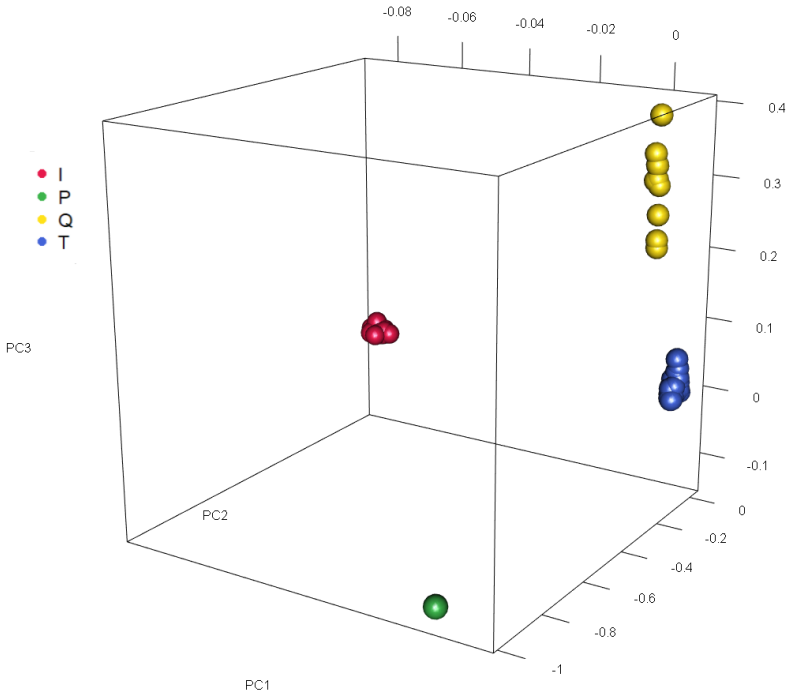


Figure 1. Principal components (PC1, 2, and 3) plot based on mitochondrial genomic relationship matrix showing the grouping of I, P, Q and T major haplogroups.

Population structure using Admixture

The population structure based on all mitochondrial sequence variants was determined using Admixture [49], where each animal is assigned a proportional membership of a predetermined number of *k* population groups (e.g. sub-species, breeds). Depending on the *k* value used (2 to 6), the major haplogroups were progressively split (Figure 2). Admixture estimated the optimal *a priori k* value to define population groups (based on the changes in cross-validation errors) was four (*k*=4) (Figure S8). When annotated with the predicted MitoToolPy haplogroups, the population structure with *k*=3 showed I separating from two further subpopulations within the T haplogroup. Further sub-groups were apparent at higher *k* values and these corresponded to sub-haplogroups within T.

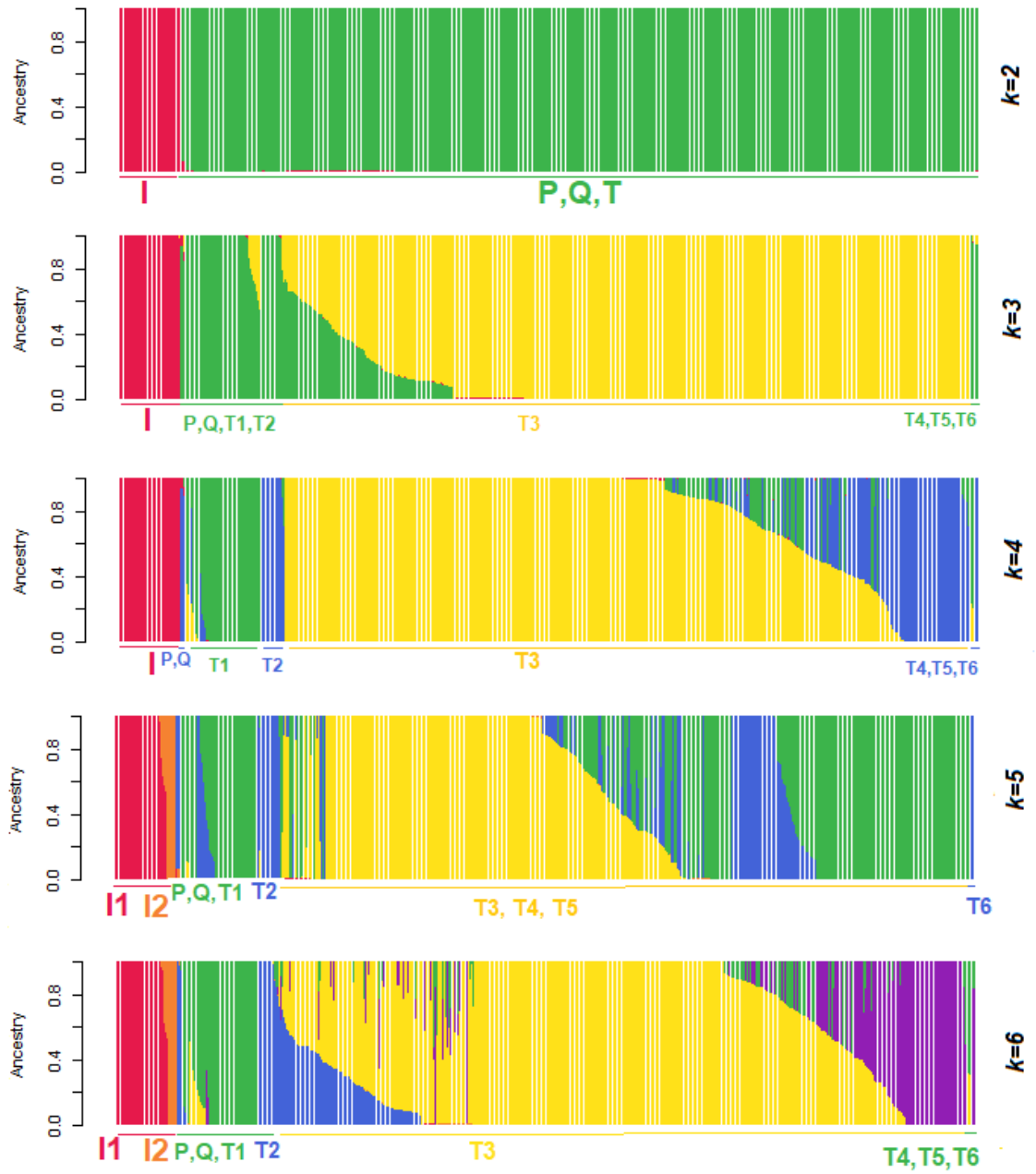


Figure 2. Population structure of cattle mitochondrial sequence variants using Admixture for a pre-defined number of populations (*k*) ranging from 2 to 6. Population structure annotated with individual animal haplogroups (I1, I2, P, Q, T1, T2, T3, T4, T5, T6) determined from MitoToolPy.

Hierarchical clustering

The nucleotide differences between each pair of whole mtDNA variant sequences was calculated using an inhouse script. The mean nucleotide difference across all pair combinations was 36 but ranged from 0 to 224. Hierarchical clustering, based on the nucleotide differences matrix between individuals, again presented two broad and

distinct clusters (Figure 3: Cluster 1 and 2). The individuals in Cluster 1 and 2 were from the major haplogroups T and I, respectively and Cluster 1 also included animals belonging to the P and Q haplogroups.

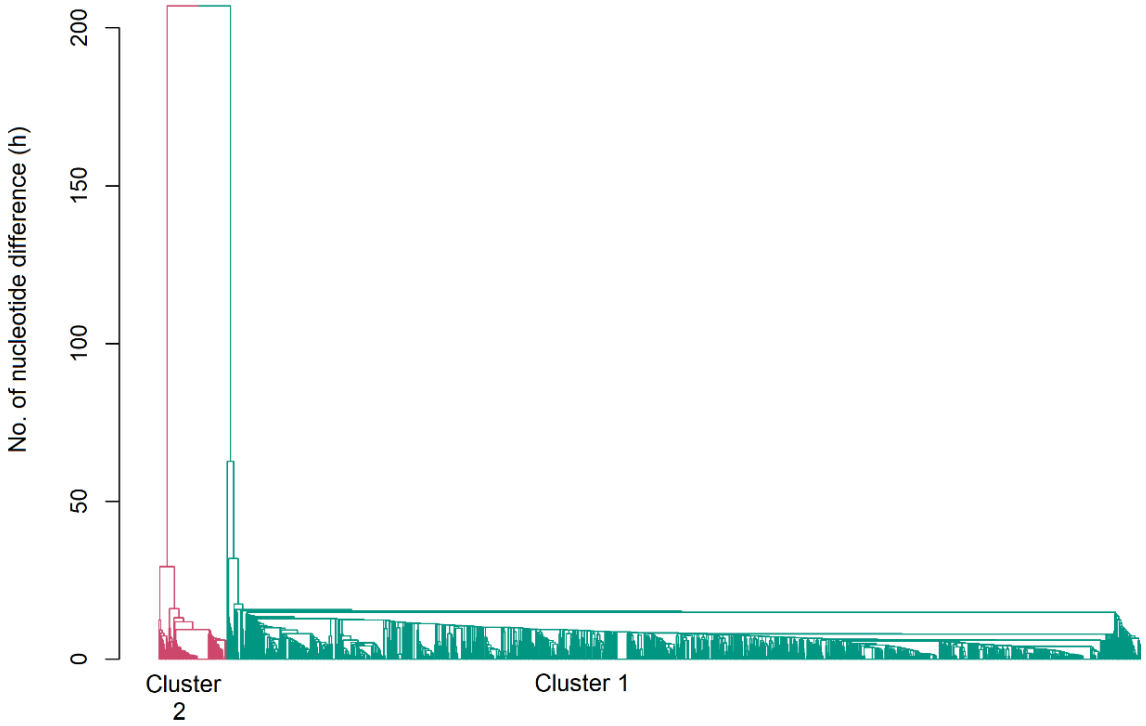


Figure 3. Hierarchical clustering of animals based on the number of nucleotide differences between the pair of mitochondrial sequences. Cluster 1 and 2 corresponded to *taurus* and *indicus* cattle, respectively.

Private variants

Private variants are additional variants present in a query mitochondrial sequence but not in the list of haplogroup determining variants. They are of interest because they can provide insights into plausible subgroups that have not been previously catalogued within the pre-defined haplogroups. We therefore examined the distribution of these private variants (output from MitoToolPy) within a haplogroup and/or breed(s). Some of the private variants were specific to a group of animals within a haplogroup (Table S5). For the most part, private variants were transition mutations from the reference allele. Four breeds had members of a sub-haplogroup that showed a specific set of private variants (Table 5). Almost 50% of private variants (N=43) were specific to particular haplogroups,

annotated either as missense (50%), upstream/downstream (30%) or synonymous (16%) gene variants (Table S5). In general, a substantial proportion of the private variants specific to a particular haplogroup included 2 SNPs in I1, 1 SNP in I2, 5 SNPs in T1, 2 SNPs in T1b1b1, 1 SNP in T1c, 1 SNP in T2 and 1 SNP in T3. Interestingly, a number of the private variants were annotated as missense, and it is therefore possible that these mutations could have downstream effects on phenotypes.

Table 5. Annotation of the private variants specific to a group of individuals within a breed showing the type of variants and affected region/gene.

Breed	**Haplo-group**	**Source of sample**	**n/N***	**Annotation: variant position (bp), type, gene**
NDama	T1	Benin, Guinea	7/12	2579, NCTE, rRNA 4714, Missense, ND2 6882, Missense, COX1 10435, Missense, ND4L
Holstein	T3	Switzerland, Canada	6/111	7948, Missense, COX2
	T3d1	United Kingdom	5/7	9807, NCTE, tRNA, 13277, Missense, ND5
Hereford	T3	Australia	2/2	5603, Synonymous, ND2
miniature Senepol	T1	Australia	3/5	6388, Synonymous, COX1

*N=total number of animals in a breed in the sub-haplogroup; n=number of animals with private variants in a breed within the haplogroup; NCTE=non-coding transcript exon

A maximum likelihood tree was constructed for whole mitogenomes for only the animals belonging to I haplogroup using MEGA X. This analysis showed four distinct clusters, one cluster corresponded to the I2 haplogroup and the three other clusters were annotated to I1 haplogroup (Figure 4a). The subclusters within I1 haplogroup were labelled as I1a, I1b and I1-Orig. The cluster I1a consisted of a group of 64 animals which were characterised by two group specific (private) variants (1497 bp and 6848 bp). The cluster I1b contained a group of 10 animals with one group specific variant (5707 bp) (Table 6).

The third I1 cluster, I1-Orig, consisted of the remaining 38 animals under I1 haplogroup in which the private variants specific to I1a and I1b were not present. The cluster I1a was mainly composed of Chinese *indicine* breeds except for two Buryat

animals (Russia), while I2, I1-Orig and I1b were mostly *indicine* breeds from the Indian subcontinent and Chinese *indicine* breeds (Table 6).

Table 6. Breed annotation and the number of animals within subclusters of the *indicus* (I) cluster based on alternate clustering techniques.

Cluster I2 (N=19) (p/q)[1]	**Cluster I1b (N=10) (p/q)**	**Cluster I1a (N=64) (p/q)**	**Cluster *I1-Orig (N=38) (p/q)**
Achai (1/4)	Bhagnari (1/4)	Bohai Black (2/5)	Achai (1/4)
Bhagnari (1/4)	Cholistani (2/5)	Buryat (2/21)	Bhagnari (2/4)
Cholistani (1/5)	Dajal (2/4)	Chaidamu Yellow (2/5)	Brahman (1/29)
Dhanni (2/5)	Dhanni (1/5)	Dabieshan (2/3)	Cholistani (2/5)
Dianzhong (1/5)	Gabrialli (1/5)	Dianzhong (1/5)	Dajal (2/4)
Gabrialli (1/5)	Hariana (1/1)	Guangfeng (3/4)	Dhanni (2/5)
Gir (1/1)		Jian (3/3)	Dianzhong (1/5)
Kangayam (1/1)	Composite (2/13)	Jiaxian Red (2/5)	Jiaxian Red (1/5)
Nari Master (1/4)		Jinjiang (3/4)	Kazakh (2/9)
Red Sindhi (1/3)		Leiqiong (3/3)	Lohani (1/1)
Sahiwal (5/7)		Lingnan (6/7)	Mongolian (1/7)
Vechur (1/1)		Luxi (5/5)	Nari Master (2/4)
		Nanyang (3/3)	Red Sindhi (2/3)
		Sichuan Indigenous (1/1	Sahiwal (2/7)
		Wandong (2/2)	Tharparkar (8/8)
Hereford (1/48)		Wannan (3/7)	Zebu Indian (1/1)
Unknown (1)		Weining (3/4)	
		Wenshan (4/6)	Composite (2)
		Xuanhan (2/5)	Galloway Belted (1/3)
		Zaobei (4/5)	Shorthorn (1/1)
		Holstein (5/330)	Hereford (2/48)
		Unknown (3)	Unknown (1)

[1] N = total number of animals in the cluster, p = No. of animals within breed in the haplogroup, q = total No. of animals within the breed, *I1-Orig =remaining animals under I1 haplogroup after assignment of other animals to I1a and I1b

Further, to explore the substructure of the I haplogroups revealed by the phylogenetic tree, the mitogenomes of only the animals assigned to haplogroup I (by MitoToolPy) were re-analysed using PCA of the GRM, Admixture and hierarchical clustering. The PC plot also showed sub-grouping of the I1 haplogroup into three well-separated clusters that were distinct from I2 (Figure 4b). Similarly, using Admixture with *k* set to 2, 3 or 4, there was distinct substructure within the I haplogroups (Figure 4c). With *k* =2, Admixture separated I2 and I1, and with *k* = 4 there was further clear separation of I2, I1a, I1b and I1-Orig, in

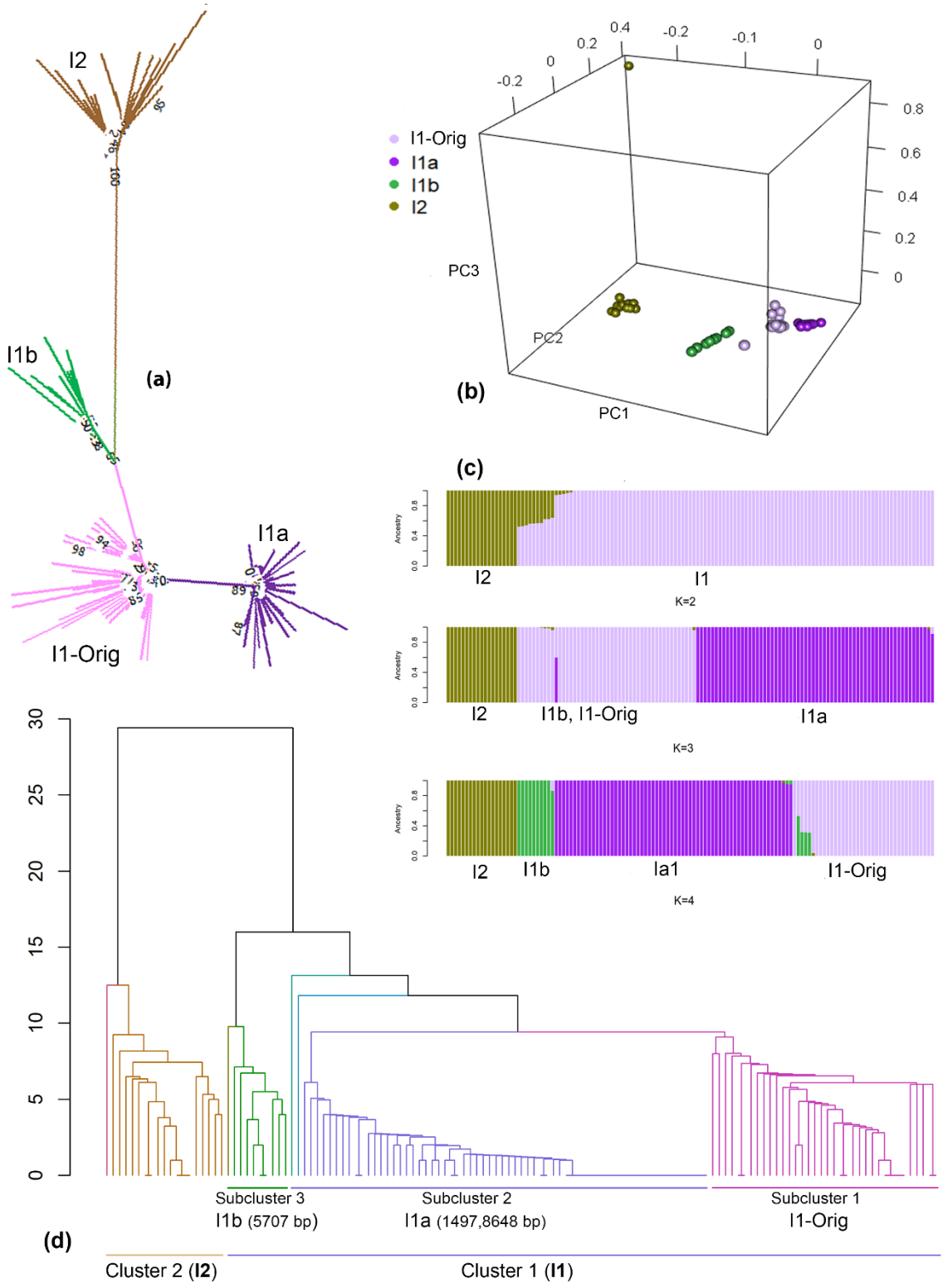


Figure 4. Subgrouping animals under I haplogroup into I2, I1 and subgroups within I1 (I1-Orig, I1a and I1b) using conventional Maximum Likelihood method (a) and alternate clustering techniques: principal component analysis (b), Admixture software (c) and hierarchical clustering based on the number of nucleotide difference between the sequences of pair of animals (d). *base pair position of private variants relative to ARS-UCD1.2_M. I1-Orig is group of animals under previous I1 haplogroup not assigned to either I1a or I1b (i.e., remaining animals in I1 Cluster1)

agreement with the PC plot. The hierarchical clustering analysis (based on animals' pairwise nucleotide differences) showed two main clusters (1 and 2 in Figure 4d). Further, distinct sub-clusters were observed within both Cluster 1 and Cluster 2, with three main subclusters under I1 haplogroups that matched those identified from the other methods (Figure 4d). The I2 cluster showed one outlier that was in agreement with the PC plot outlier (i.e. the same animal). In all the above unsupervised clustering analyses, sub-clusters I1a and I1b were comprised of the same group of animals. Interestingly, all three unconventional mitochondrial clustering methods reproduced the same grouping of these animals as with maximum likelihood method.

Mitochondrial haplotype diversity

Overall, across 1,883 animals, 1,309 whole mitochondrial genome haplotypes were identified with haplotype diversity of 0.999 (SD 0.0001). Of the 1309 haplotypes, 1010 were singletons (i.e. one animal per haplotype) indicating considerable diversity. The remaining haplotypes (299) were shared by 2 to 23 animals (Figure S9). The shared haplotypes were approximately 60% within a breed and 25% between the breeds. The haplotype diversity within breed was generally high and ranged from 0.932-0.998 (Table 7). The shared haplotypes specific to a breed were also found across animals sampled in several different countries. Two haplotypes distinct to Angus were present in animals sourced from Canada and USA. Additionally, some haplotypes were shared among several breeds and across several countries. For example, one haplotype was identified in 23 animals from a wide range of breeds including Holsteins sourced from China and a number of other breeds mostly of Asian origin (Luxi, Lingan, Zaobei, Weining, Wannan, Jian, Jinjiang, Wenshan, Nanyang, Xuanhan, Leiqiong and Bohai Black). Similarly, another haplotype was shared by 23 animals in Angus (Canada), Brown Swiss (USA), Charolais (France), Deutsches Schwarzbuntes Niederungsrind (Germany), Gelbvieh (Canada), Hereford (Australia, Russia and USA), Holsteins

Table 7. Mitochondrial DNA sequence polymorphism and diversity (standard deviation) of selected breeds with a sample size of 20 or more.

Breed	No. of Sequences	No. of Segregating Sites	Average No. of difference	No. of Haplotypes	Haplotype Diversity	Nucleotide diversity
Holstein	267	697	16.96	210	0.998 (0.001)	0.0055 (0.0010)
Jersey	27	57	9.62	16	0.937 (0.031)	0.0031 (0.0004)
Brown Swiss	84	202	8.99	64	0.993 (0.003)	0.0029 (0.0001)
Simmental	32	80	7.55	24	0.976 (0.002)	0.0025 (0.0002)
Norwegian Red	222	338	10.51	180	0.998 (0.001)	0.0034 (0.0001)
Holstein Friesians	35	121	9.94	31	0.992 (0.010)	0.0032 (0.0003)
DSN*	47	154	10.2	40	0.992 (0.007)	0.0033 (0.0002)
Angus	103	122	7.65	45	0.935 (0.014)	0.0025 (0.0001)
Yakut	35	68	9.44	15	0.938 (0.018)	0.0031 (0.0003)
Hereford	44	312	33.84	33	0.979 (0.012)	0.011 (0.0043)
Charolais	33	114	8.856	31	0.996 (0.009)	0.0028 (0.0002)
Limousin	27	100	8.80	25	0.994 (0.012)	0.0029 (0.0002)
Modern Danish Red	23	73	11.55	19	0.980 (0.020)	0.0038 (0.0002)
Hanwoo	24	146	15.59	23	0.996 (0.013)	0.0051 (0.0013)
Buryat	20	250	48.16	12	0.932 (0.035)	0.0157 (0.0070)

DSN* Deutsches Schwarzbuntes Niederungsrind

(Netherlands and USA), Rodkulla (Sweden), Red Angus (USA), Hanwoo (Korea), Belgian Blue and a composite breed (Australia). Among the breeds, Holstein was the most numerous breed in our study (N=267), therefore it was of interest to examine the network of haplotypes within Holsteins from a total of 210 haplotypes (168 singletons and 42 shared) (Figure 5). Haplotypes from T3 and subgroups formed the core of the network with side branches in agreement with MitoToolPy I and T1 haplogroup allocations.

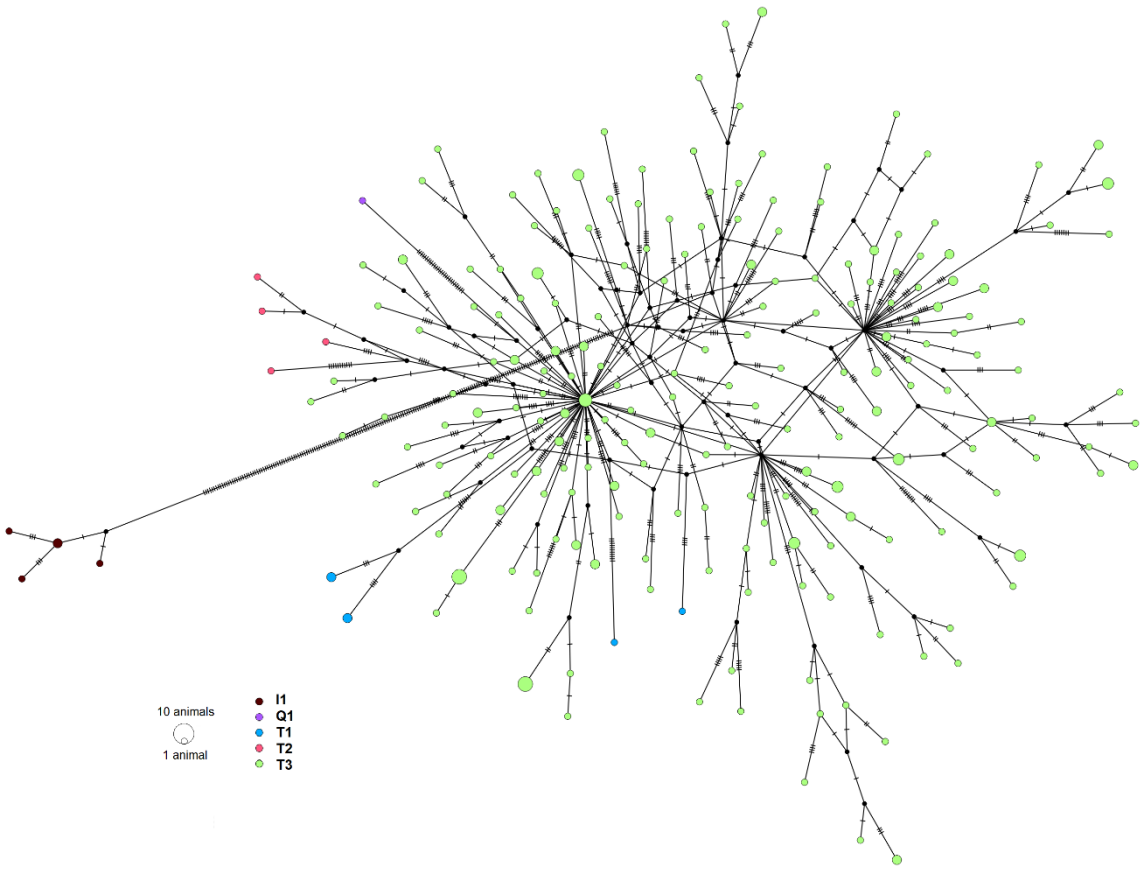


Figure 5. Haplotype network consisting of 210 haplotypes in the Holstein population (N=267) using the median-joining network in PopART and annotated with haplogroups predicted from MitoToolPy. The size of the circles is proportional to the number of animals carrying the same haplotype.

To test the ability of a naïve hierarchical clustering approach to differentiate haplotypes across all animals, we used the height of the cluster (h) corresponding to the nucleotide difference of 0 between two haplotype pairs. The resulting cluster groups were compared with the haplotypes derived from the DnaSP, mainly focussing on the non-singleton haplotypes and hierarchical clusters. There was approximately the same number of singleton clusters as singleton haplotypes (1032). At least 132 clusters and haplotypes had substantial memberships in common (Table S7). For example, Cluster 328 and Haplotype-324 (with 23 each), Cluster-7 and Haplotype-7 (21 animals each) and all other cluster-haplotype combinations with more than five animals (total 30) had 100% of the same individuals except for five groups. This demonstrates a high concordance between determination of haplotypes by hierarchical cluster and the traditionally determined haplotypes.

Mitochondrial DNA polymorphism and nucleotide diversity

We investigated mitochondrial nucleotide diversity in animals from breed groups with ≥ 20 animals. Overall, there were 1825 segregating sites, nucleotide diversity (π) was 0.012, and the average nucleotide difference between the pair of sequences was 35.5. The nucleotide diversity was high in Buryat and Hereford and other breeds had low but comparable nucleotide diversity ranging from 0.002-0.005. The analysis of molecular variance (AMOVA) showed that the percentage of genetic variation from among and within breed components was 3.1% and 96.9%, respectively, indicating high within breed genetic diversity.

Imputation and MitoToolPy haplogroup prediction

The routine practice of discarding the sites with missing genotypes from all sequences in the mtDNA analysis results in loss of information, particularly when the

proportion of missing genotypes in an individual were low. In this case, the imputation of sporadic missing genotypes could increase the number of animals and sites for analysis, but the empirical accuracy of mitochondrial imputation in cattle is unknown. To test this, we masked 10% of known genotypes (307 sites) in a random 20% of animals (333 out of 1,883). Then we imputed the masked genotypes using Beagle (version 4.0) using the *gt* and *ref* option using the remaining 1550 animals with all genotypes present as a reference for imputation. The overall concordance of this imputation was 99.8%, although concordance for heteroplasmic sites was approximately 66% (Table 8). There was a tendency for imputation to bias heteroplasmic and homoplasmic alternate genotypes towards the homoplasmic reference genotypes (0/0). The genotype likelihood '*gl*' option also produced a similar concordance of 99.5%.

To evaluate the effect of imputation on haplogroup prediction, we re-analysed the animal haplogroups in the imputed dataset and compared these to their haplogroup prediction from the original dataset. This was replicated 50 times with a new random set of animals chosen for masking genotypes for imputation. The predicted haplogroups matched in 99.7% of the individuals when compared to their haplogroup predicted from the full set of real genotypes. The accuracy of imputation and the predicted haplogroup for the masked dataset showed little variation across the 50 replications (Table S8). This suggests that missing genotypes can be imputed and used for prediction of haplogroups with high but not perfect accuracy. These results are provided for information only, that is, no imputed data was used elsewhere in this study.

Table 8. Empirical accuracy of imputing sporadic missing genotypes in mitogenomes. Number of correctly imputed genotypes (percentage correct in brackets) on the diagonals and number of genotypes wrongly imputed shown on the off-diagonals. Assessment was based on randomly masking of 10% of positions (307) per animal in 20% of animals (333).

		Original Genotype						
		0|0	0|1	0|2	1|1	2|2	3|3	
Imputed genotype	0|0	101089 (99.9%)	39		73		1	
	0|1	37	82 (65.6%)		11			
	0|2	2		3 (100%)		2		
	1|1	64	4		800 (90.4%)	1		
	2|2	4			1	16 (84.2%)		
	3|3						2 (67.0%)	
	Total	101196	125	3	885	19	3	101992/102231 (99.8%)

Discussion

Our study undertook a comprehensive analysis of mitochondrial genome sequence diversity in 1,883 cattle, including the most important global cattle breeds and sub-species in a single study. This represents one of the single largest studies of this kind demonstrating the use of short read mitochondrial sequence data from general DNA sequencing. Our use of the entire mitogenome enabled a more in-depth study of the full range of diversity, that may be important in future studies of the potential impact of mitochondrial variants on phenotypes. Our large sample size enabled subgrouping within haplogroups and breeds as well as annotation of the private variants. In addition to the conventional diversity indices, we have investigated alternate ways of analysing population structure and haplotypes of the entire mitogenome, that do not rely on predefined haplogroups and therefore capture a broader spectrum of the diversity.

Introgression of African *taurus* and *indicus* haplogroups into European *taurus*

Most breeds belonged to their anticipated haplogroup except for some animals of European breeds and composites that were mostly allocated to African *taurus* (T1) and relatively few to *indicus* (I). This is not surprising, as the T1 haplogroup has been previously reported in European cattle breeds (1-30%) from France, Spain, Portugal, Italy, Balkan and Greece [5, 18, 52, 53] and America (Creole cattle) [54]. The detection of T1 haplogroups in Iberia [55] and Sicily and southern Italy, according to [56] may be the influence of migration of African cattle into southern Europe via the Mediterranean Sea coastline. The African T1 sequence was also found in Iberian Bronze age cattle [57].

The breeds with no previous report of T1 haplogroup but found in our study are Jersey and Holstein: those showing the T1 haplotype were sourced mainly from Australia (9 out of 10). Australia has a recent history of crossbreeding European breeds with more heat tolerant imported breeds to develop cattle better adapted to the tropical environment in northern Australia [58]. Australia imported Jersey from the Channel Islands and Holsteins from the Netherlands in 1850. It is possible that some of the first African cattle arriving in Australia were Afrikander (8 bulls and 2 cows) imported from South Africa in the early 1950s [59] and other breeds (Boran, Bonsmara etc.) in the late 1980s. However, details on the sex of imported animals are not available making it difficult to confirm the most likely maternal route of T1 mitogenome transmission. In 1990, the embryos from Boran and Tuli (African) cattle were imported [60]. While the attributes of heat tolerance and tick resistance were sought after under the extensive tropical beef production system, the presence of the T1 haplogroup in dairy breeds (Holstein and Jersey) in Australia suggests these animals may be the result of upgrading from cows carrying the T1 mitochondrial lineages or sporadic cases of cross breeding to improve heat tolerance but this warrants further investigation.

The *indicine* haplogroup (I) in Holsteins in this study were largely in female samples originating from China. This is not surprising as the I haplogroup has been previously reported in Chinese Holstein [61] and at least three haplotypes were shared among Chinese Holstein and native cattle (22 animals) [62]. Imported purebred Holsteins were used to grade-up local cows as well as for the development of the Chinese Black and White cattle breed [63].

There is a possibility that the breed origin was incorrectly labelled on some samples, therefore we undertook a PCA of all the *taurus* animals (N=1451) based on a genomic relationship matrix derived from 45K autosomal SNPs. The PC plot of Holstein and Jersey breeds shows tight clustering of these animals regardless of the MT haplogroup (Figure S10) which supports that the *indicine* and the African *taurine* maternal lines in these breeds are likely due to upgrading of native cattle.

In composite breeds such as Brahman, the mitochondrial haplotypes were mostly *taurine* in this study, which is interesting because the breed's nuclear DNA is primarily of *indicus* origin in the 1000 Bull Genomes project [31]. The Australian Brahman cattle in this study were approximately 97% *taurine* (T1 47% and T3 50%) and about 3% *indicus* (I) haplogroups. Compared to our study, Brahman from China were reported with lower representation of T1 (35%) and T3 (26%), but higher in I (39%) haplogroups [27], while American Brahman showed lower T1 (30%) but higher T3 (70%) [64]. Originally, Brahman cattle were introduced into Australia from the USA in 1933 [58]. In fact, in America, Brahmans were developed from the crossing/ upgrading of *B. taurus* females (often Creole cattle) with Guzerat, Nellore, Gyr and Krishna valley cattle. As such, haplogroups in the *indicine* breeds in Americas were reported to be largely *taurine* (T3 50%, T1 48%) and rarely *indicine* (I) (1 in 66 *indicine* animals) [65].

The inter-breed introgression of haplogroups was also supported by sharing of the diverse haplotypes among breeds. This again points to the common practice of upgrading. Mitochondrial haplotypes were also shared across countries, and to a higher degree between

countries in close proximity, indicating the movement of female animals. However, the shared haplotypes between more distant countries (e.g. USA and Australia) suggests the movement of foundation females or, more recently, embryos.

Subgroup of I1 Chinese *indicus* (I1a)

The presence of both I1 and I2 *indicus* haplogroups with the predominance of I1 recorded in the current study agrees with the previous studies [66-68]. The I1 haplogroup originated in the Indus valley, while the I2 haplogroup is believed to have originated in northern India [69]. Interestingly, within the large subcluster of I1, we constantly identified a sub-cluster (I1a) comprising mainly of Chinese *indicine* breeds (19/20 breeds) (Figure 4). This sub-cluster (I1a) had two mutations specific to the subgroup, one in a *rRNA* and another within the *ATP6* genes. These mutations were annotated as non-coding transcript exonic and missense variants, respectively. There has been a previous report of specific I1 haplotype common among the Chinese breeds indicative of a nucleus of Chinese *indicus*, but this was based on D-loop sequences [67]. Another study employing whole mitochondrial genome also reported a specific group under I1 (characterised by 6 mutations) for a breed not in the current study (Yunling cattle) [27]. Two specific mutations characterising the I1a subgroup in our study were also reported in the Yunling cattle subgroup, while the other four were found non-specific to I1a group in our study. These findings, together with results from our study, suggest the presence of a unique I1 subgroup (I1a) specific to breeds emanating from China. Further, five Holstein animals in our study that originated from China also had the I1a sub-haplogroup, indicating there may have been *adhoc* or controlled upgrading of *indicus* females that carry I1a.

While I1a in this study may not be a separate haplogroup, a distinct cattle haplogroup “C” and a separate domestication event in north-eastern China during the early Holocene has been proposed by [70]. However, their proposed new

haplogroup C sequence did not match our I1a subgroup sequences. The presence of conventional I1and I2 haplogroups support the consensus among the published literature that the indicus cattle population in China is a result of migration and spread from India. The *Bos indicus* are reported to have been introduced into China between 2000-200BC and currently there is no archaeological or genetic evidence for origin of domestic cattle in ancient China [66, 71] suggesting genetic drift as a contributing factor to the formation of the subgroup (I1a). Another possibility, considering the specificity to Chinese breeds (not present in Indian *indicine* breeds), is the potential restocking of auroch female lines from the wild in China and establishing nucleus or base for the *indicine* breeds in China. There is molecular evidence of aurochs in China's northeast during the Neolithic period [72]. Therefore, these hypotheses need further investigation. The subgroup (I1b) under I1 characterised by a specific mutation (5707 bp) did not exist in any breeds from China, while the entire I1 group included breeds from both India and China. There are shared haplogroups (I1 and I2) between breeds of the two countries and also sub-groups specific to the region.

There are fewer studies within *indicine* haplogroups compared to *taurine* and previous studies classified them into only two broad haplogroups (I1 and I2). Despite several previous studies on mtDNA of the Chinese cattle, subgroupings under I1 were not reported except by [27]. One possible reason is that the location of the mutations defining subgroup under I1, is in the coding region of the mitochondrial genome, while most studies in the past were mainly based on the D-loop. Complete mitochondrial genome sequences better define the full spectrum of mitochondrial diversity compared to using the D-loop only, and may uncover mutations in coding regions that affect specific phenotypes.

Imputation of mitogenome variants

Genetic variation was mainly within breed (97%, AMOVA), and high haplotype diversity and multiple haplogroups exist within breeds. For example, Holstein animals

(N 267) belonged to at least 15 subgroups and 210 haplotypes (Table 7, Figure 5).The haplogroup and haplotypes within a breed are of strong interest for phenotype association studies and in this study we found several private variants in groups of animals that were annotated as missense in MT genes (Table S5). In humans, the association of mitochondrial haplogroups (H and R) to specific phenotypes such as risk to diseases [73-75], metabolic disorders [76, 77] and athletic endurance [78, 79], is more advanced than in domestic animals. Although two recent studies examined the relationship between mitochondrial haplogroups and litter size [21] and other phenotypic traits [22] in pigs, these studies lack sufficient power to distinguish or pinpoint specific mutations (causal) affecting the trait.

Association studies of traits could include variants from the mitochondrial genome, if it is possible to sequence and impute large numbers of animals for MT variants, but this has not yet been done in livestock. While whole-genome sequences are now regularly imputed and exploited for association studies of production traits across livestock, mitochondrial genomes are excluded from these studies and literature on mitochondrial genome imputation is scarce or non-existent for livestock. In humans, the imputation of the mitogenome from ancient remains showed that the accuracy of mitogenome imputation, like the nuclear counterpart, benefitted from having a large and diverse reference sequence [80]. Thus, utilising the existing resources such as the data from the 1000 Bull Genomes Project could potentially provide a reference set for mitogenome imputation from lower density SNP arrays. However, we would recommend following our rigorous filtering to minimize the impact of NUMTs and wherever possible using only non-semen male tissue samples or female tissue samples. The first step towards large scale imputation of MT sequence is to confirm that sporadic missing genotypes in the mitogenomes can be accurately imputed. This would provide a full reference panel of mitogenomes to impute to for animals with MT SNP genotypes from lower density panels. In the current study, the accuracy of imputation

of missing genotypes (99.8%) was comparable to results in humans that used tools specifically for imputation of the mitochondrial genome such as MitoIMP [81]. This indicates that existing tools may be applied to mitochondrial genome imputation with customization.

Applicability of unconventional mitochondrial DNA analysis tools

The current study utilised conventional tools for mtDNA analysis (DnaSP, MitoToolPy, MEGA X) but also compared these results with alternative tools such as GRM based PCA, Admixture, and hierarchical clustering based on nucleotide differences. Our primary interest in use of the alternative tools was to better quantify the full spectrum of genetic diversity across the entire mitogenome, rather than simply place animals into the higher level haplogroups. However, as expected, the results from the less conventional mitochondrial tools were mostly in agreement with haplogrouping. Therefore, alternative clustering methods, specifically the hierarchical clustering based on nucleotide differences, may be used as grouping techniques that is equivalent to haplotypes for use in trait/phenotypic association studies.

Conclusions

There is high mitochondrial genomic diversity among modern cattle and a large proportion of this genetic variation is within breeds. The introgression of African *taurine* and *indicine* mitochondrial haplogroups into European *taurine* breeds occurred at low frequencies. The patterns of population structure and haplogroups from conventional tools were very similar to results of non-traditional mitochondrial methods developed for autosomal DNA. We provide additional evidence of a new *indicus* I1 haplogroup subgroup (I1a) in Chinese *indicine* breeds. Within breed mitochondrial diversity (haplotypes/ haplogroups) is likely at a level sufficient to conduct trait association studies. Imputation of sporadic missing genotypes in the mitochondrial genome was highly

accurate with the exception of heteroplasmic sites. This could enable larger data sets to be used for population studies through recovery ofsites or animals with low levels of missing genotypes and would provide a diverse reference population for large scale imputation of mitogenomes.

Acknowledgements

I would like to thank the examiners Prof Paolo Ajmone Marson and Associate Prof Cynthia Bottema for their constructive comments in improving this chapter.

References

1. Loftus, R.T., et al., *Evidence for two independent domestications of cattle.* Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(7): p. 2757-2761.
2. Pitt, D., et al., *Domestication of cattle: Two or three events?* Evolutionary Applications, 2019. **12**(1): p. 123-136.
3. Verdugo, M.P., et al., *Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent.* Science, 2019. **365**(6449): p. 173-176.
4. Chen, S., et al., *Zebu cattle are an exclusive legacy of the South Asia neolithic.* Mol Biol Evol, 2010. **27**(1): p. 1-6.
5. Beja-Pereira, A., et al., *The origin of European cattle: evidence from modern and ancient DNA.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(21): p. 8113-8118.
6. Pellecchia, M., et al., *The mystery of Etruscan origins: novel clues from Bos taurus mitochondrial DNA.* Proceedings of the Royal Society B: Biological Sciences, 2007. **274**(1614):1175-1179.
7. Decker, J.E., et al., *Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics.* Proceedings of the National Academy of Sciences, 2009. **106**(44): p. 18644.
8. Ajmone-Marsan, P., J.F. Garcia, and J.A. Lenstra, *On the origin of cattle: How aurochs became cattle and colonized the world.* Evolutionary Anthropology: Issues, News, and Reviews, 2010. **19**(4): p. 148-157.
9. Decker, J.E., et al., *Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle.* PLOS Genetics, 2014. **10**(3): p. e1004254.
10. Hanotte, O., et al., *African Pastoralism: Genetic Imprints of Origins and Migrations.* Science, 2002. **296**(5566): p. 336.
11. Payne, W.J.A. and J. Hodges, T*ropical cattle: origins, breeds and breeding policies*. 1997, Oxford: Blackwell Science Ltd. vii + 328 pp.
12. Anderson, S., et al., *Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome*. J Mol Biol, 1982. 156(4): p. 683-717.

13. Lee, W.T., et al., *Mitochondrial DNA haplotypes induce differential patterns of DNA methylation that result in differential chromosomal gene expression patterns*. Cell DeathDiscovery, 2017. **3**(1): p. 17062.

14. Bonfiglio, S., et al., *Origin and Spread of Bos taurus: New Clues from Mitochondrial Genomes Belonging to Haplogroup T1*. PLOS ONE, 2012. **7**(6): p. e38601.

15. Achilli, A., et al., *Mitochondrial genomes from modern horses reveal the major haplogroupsthat underwent domestication*. Proceedings of the National Academy of Sciences, 2012. **109**(7):2449.

16. Achilli, A., et al., *Mitochondrial genomes of extinct aurochs survive in domestic cattle.* Current Biology, 2008. **18**(4): p. R157-R158.

17. Peng, M.-S., et al., *DomeTree: a canonical toolkit for mitochondrial DNA analyses in domesticated animals.* Molecular Ecology Resources, 2015. **15**(5): p. 1238-1242.

18. Lenstra, J.A., et al., *Meta-Analysis of Mitochondrial DNA Reveals Several Population Bottlenecks during Worldwide Migrations of Cattle.* Diversity, 2014. **6**(1).

19. Ryzhkova, A.I., et al., *Mitochondrial diseases caused by mtDNA mutations: a mini-review.* Therapeutics and clinical risk management, 2018. **14**: p. 1933-1942.

20. Fernandez, A.I., et al., *Mitochondrial genome polymorphisms associated with longissimus muscle composition in Iberian pigs.* J Anim Sci, 2008. **86**(6): p. 1283-90.

21. Wang, D., et al., *Relationship between mitochondrial DNA haplogroup and litter size in the pig.* Reproduction, Fertility and Development, 2020. **32**(3): p. 267-273.

22. St. John, J.C. and T.-S. Tsai, *The association of mitochondrial DNA haplotypes and phenotypic traits in pigs.* BMC Genetics, 2018. **19**(1): p. 41.

23. Schutz, M.M., et al., *The effect of mitochondrial DNA on milk production and health of dairy cattle.* Livestock Production Science, 1994. **37**(3): p. 283-295.

24. Yu, G., et al., *Mitochondrial Haplotypes Influence Metabolic Traits in Porcine Transmitochondrial Cybrids.* Scientific Reports, 2015. **5**(1): p. 13118.

25. Wang, J., et al., *Mitochondrial haplotypes influence metabolic traits across bovine inter- and intra-species cybrids.* Scientific Reports, 2017. **7**(1): p. 4179.

26. Xia, X., et al., *Mitogenome Diversity and Maternal Origins of Guangxi Cattle Breeds.* Animals, 2020. **10**(1).

27. Xia, X., et al., *Abundant Genetic Diversity of Yunling Cattle Based on Mitochondrial Genome.* Animals, 2019. **9**(9).

28. Hayes, B.J. and H.D. Daetwyler, *1000 Bull Genomes Project to Map Simple and Complex Genetic Traits in Cattle: Applications and Outcomes.* Annu Rev Anim Biosci, 2019. **7**: p. 89-102.

29. van Binsbergen, R., et al., *Genomic prediction using imputed whole-genome sequence data in Holstein Friesian cattle.* Genetics Selection Evolution, 2015. **47**(1): p. 71.

30. Hayes, B., et al., *Genomic prediction from whole genome sequence in livestock: the 1000 Bull Genomes Project*. 2014.

31. Daetwyler, H.D., et al., *Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle*. Nature Genetics, 2014. 46(8): p. 858-865.

32. Bouwman, A.C., et al., *Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals*. Nature Genetics, 2018. 50(3): p.362-367.

33. Xiang, R., et al., *Genome-wide fine-mapping identifies pleiotropic and functional variants that predict many traits across global cattle populations.* Nature Communications, 2021. **12**(1): p. 860.

34. Chung, N.C., et al., *Population Structure Analysis of Bull Genomes of European and Western Ancestry.* Scientific Reports, 2017. **7**(1): p. 40688.

35. Wallace, D.C. and D. Chalkia, *Mitochondrial DNA genetics and the heteroplasmy conundrumin evolution and disease.* Cold Spring Harbor perspectives in biology, 2013. **5**(11): p. a021220-a021220.

36. Rosen, B. D., et al., *De novo assembly of the cattle reference genome with single-molecule sequencing*. GigaScience, 2020. **9**(3)

37. Cattle Genome Sequencing International Consortium, *Btau_5.0.1* (Accession CM001061.2). Accessed Jan 5, 2021. https://www.ncbi.nlm.nih.gov/assembly/GCF_000003205.7/

38. Danecek, P., et al., *The variant call format and VCFtools.* Bioinformatics, 2011. **27**(15): p. 2156-2158.

39. Li, H., *A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data*. Bioinformatics, 2011. **27**(21):p.2987-2993.

40. Librado, P. and J. Rozas, *DnaSP v5: a software for comprehensive analysis of DNApolymorphism data.* Bioinformatics, 2009. **25**(11): p. 1451-1452.

41. Excoffier, L. and H.E. Lischer, *Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows.* Mol Ecol Resour, 2010.**10**(3): p.564-7.

42. Kumar, S., et al., *MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms.* Molecular Biology and Evolution, 2018. **35**(6): p. 1547-1549.

43. Leigh, J.W. and D. Bryant, *popart: full-feature software for haplotype network construction.*Methods in Ecology and Evolution, 2015. **6**(9): p. 1110-1116.

44. Cingolani, P., et al., *A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2;iso-3.* Fly, 2012. **6**(2): p. 80-92.

45. Chang, C.C., et al., *Second-generation PLINK: rising to the challenge of larger and richer datasets.* GigaScience, 2015. **4**(1).

46. Yang, J., et al., *GCTA: a tool for genome-wide complex trait analysis.* American journal of human genetics, 2011. **88**(1): p. 76-82.

47. Ligges, U., *3D Scatter Plot*. 2016.

48. Adler, D., O. Nenadf, and W. Zucchini, *Rgl: A r-library for 3d visualization with opengl*. 2003.

49. Alexander, D.H. and K. Lange, *Enhancements to the ADMIXTURE algorithm for individualancestry estimation.* BMC Bioinformatics, 2011. **12**(1): p. 246.

50. Galili, T., *dendextend: an R package for visualizing, adjusting and comparing trees ofhierarchical clustering.* Bioinformatics, 2015. **31**(22): p. 3718-3720.
51. Browning, Brian L. and Sharon R. Browning, *Genotype Imputation with Millions ofReference Samples.* The American Journal of Human Genetics, 2016. **98**(1): p.116-126.
52. Miretti, M.M., et al., *Predominant African-derived mtDNA in Caribbean and BrazilianCreolecattle is also found in Spanish cattle (Bos taurus).* J Hered, 2004. **95**(5): p.450-3.
53. Cortes, O., et al., *Ancestral matrilineages and mitochondrial DNA diversity of the Lidia cattlebreed.* Anim Genet, 2008. **39**(6): p. 649-54.
54. Lir6n, J.P., et al., *African matrilineages in American Creole cattle: evidence of two independentcontinental sources.* Animal Genetics, 2006. **37**(4): p. 379-382.
55. da Fonseca, R.R., et al., *Consequences of breed formation on patterns of genomic diversity anddifferentiation: the case of highly diverse peripheral Iberian cattle.* BMC Genomics, 2019. **20**(1): p. 334.
56. Cymbron, T., et al., *Mitochondrial sequence variation suggests an African influence in Portuguese cattle.* Proceedings. Biological sciences, 1999. **266**(1419): p. 597-603.
57. Anderung, C., et al., *Prehistoric contacts over the Straits of Gibraltar indicated by genetic analysis of Iberian Bronze Age cattle.* Proceedings of the National Academy of Sciences of theUnited States of America, 2005. **102**(24): p. 8431.
58. Tonts, M., R. Yarwood, and R. Jones, *Global Geographies of Innovation Diffusion: The Case of the Australian Cattle Industry.* The Geographical Journal, 2010.**176**(1):p90-104.
59. Ward, C. *Cattle for the tropics*. 2011. Accessed Feb 4, 2021. https:// csiropedia.csiro.au/cattle- for-the-tropics/
60. Kerton, R. *Cattle out of Africa (1990)*. 1990. Accessed Feb 4, 2021. https:// csiropedia.csiro.au/cattle-out-of-africa-1990/
61. Lai, S.-J., et al., *Genetic diversity and origin of Chinese cattle revealed by mtDNA D-loopsequence variation.* Molecular Phylogenetics and Evolution, 2006. **38**(1): p.146-154.
62. Ferreri, M., et al., *Chinese Holstein Cattle Shows a Genetic Contribution from Native AsianCattle Breeds: A Study of Shared Haplotypes and Demographic History*. Asian-Australas J Anim Sci, 2011. **24**(8): p. 1048-1052.
63. Cheng, P., *Livestock breeds of China.* FAO Animal Production and Health Paper, 1984. **46**:p. 63.
64. Ginja, C., et al., *Origins and genetic diversity of New World Creole cattle: inferences frommitochondrial and Y chromosome polymorphisms*. Animal Genetics, 2010. **41**(2):p. 128-141.
65. Ginja, C., et al., *The genetic ancestry of American Creole cattle inferred from uniparentaland autosomal genetic markers*. Scientific Reports, 2019. **9**(1): p. 11486.

66. Xia, X., et al., *Comprehensive analysis of the mitochondrial DNA diversity in Chinesecattle*. Animal Genetics, 2019. **50**(1): p. 70-73.
67. Jia, S., et al., *Genetic Variation of Mitochondrial D-loop Region and Evolution Analysis in Some Chinese Cattle Breeds.* Journal of Genetics and Genomics, 2007. **34**(6): p. 510-518.
68. Lei, C.Z., et al., *Origin and phylogeographical structure of Chinese cattle.* Animal Genetics, 2006. **37**(6): p. 579-582.
69. Magee, D.A., D.E. MacHugh, and C.J. Edwards, *Interrogation of modern and ancient genomesreveals the complex domestic history of cattle.* Animal Frontiers, 2014. **4**(3): p. 7-22.
70. Zhang, H., et al., *Morphological and genetic evidence for early Holocene cattle management in northeastern China.* Nature Communications, 2013. **4**(1): p. 2755.
71. Peng, L., et al., *Zooarchaeological and Genetic Evidence for the Origins of Domestic Cattle inAncient China.* Asian Perspectives, 2017. **56**(1): p. 92-120.
72. Cai, X., et al., *mtDNA Diversity and genetic lineages of eighteen cattle breeds from Bos taurusand Bos indicus in China.* Genetica, 2007. **131**(2): p. 175-183.
73. Abu-Amero, K.K., et al., *Association of Mitochondrial Haplogroups H and R With Keratoconus in Saudi Arabian Patients.* Investigative Ophthalmology & Visual Science, 2014.**55**(5): p. 2827-2831.
74. Farha, S., et al., *Mitochondrial Haplogroups and Risk of Pulmonary Arterial Hypertension.* PLOS ONE, 2016. **11**(5): p. e0156042.
75. Maruszak, A., et al., *Mitochondrial haplogroup H and Alzheimer's disease—Is there a connection?* Neurobiology of Aging, 2009. **30**(11): p. 1749-1755.
76. Fuku, N., et al., *Mitochondrial Haplogroup N9a Confers Resistance against Type 2 Diabetes in Asians.* The American Journal of Human Genetics, 2007. **80**(3): p.407-415.
77. Nardelli, C., et al., *Haplogroup T Is an Obesity Risk Factor: Mitochondrial DNA Haplotyping in a Morbid Obese Population from Southern Italy.* BioMed Research International, 2013. **2013**: p. 631082.
78. Castro, M.G., et al., *Mitochondrial haplogroup T is negatively associated with the status ofelite endurance athlete.* Mitochondrion, 2007. **7**(5): p. 354-357.
79. Mikami, E., et al., *Mitochondrial haplogroups associated with elite Japanese athlete status.* British Journal of Sports Medicine, 2011. **45**(15): p. 1179.
80. Mizuno, F., et al., *Imputation approach for deducing a complete mitogenome sequence from low-depth-coverage next-generation sequencing data: application to ancient remains from the Moon Pyramid, Mexico.* Journal of Human Genetics, 2017. **62**(6): p. 631-635.
81. Ishiya, K., et al., *MitoIMP: A Computational Framework for Imputation of MissingData in Low-Coverage Human Mitochondrial Genome.* Bioinformatics and biology insights, 2019. **13**: p. 1177932219873884-1177932219873884.

Supplementary Files

Supplementary figures

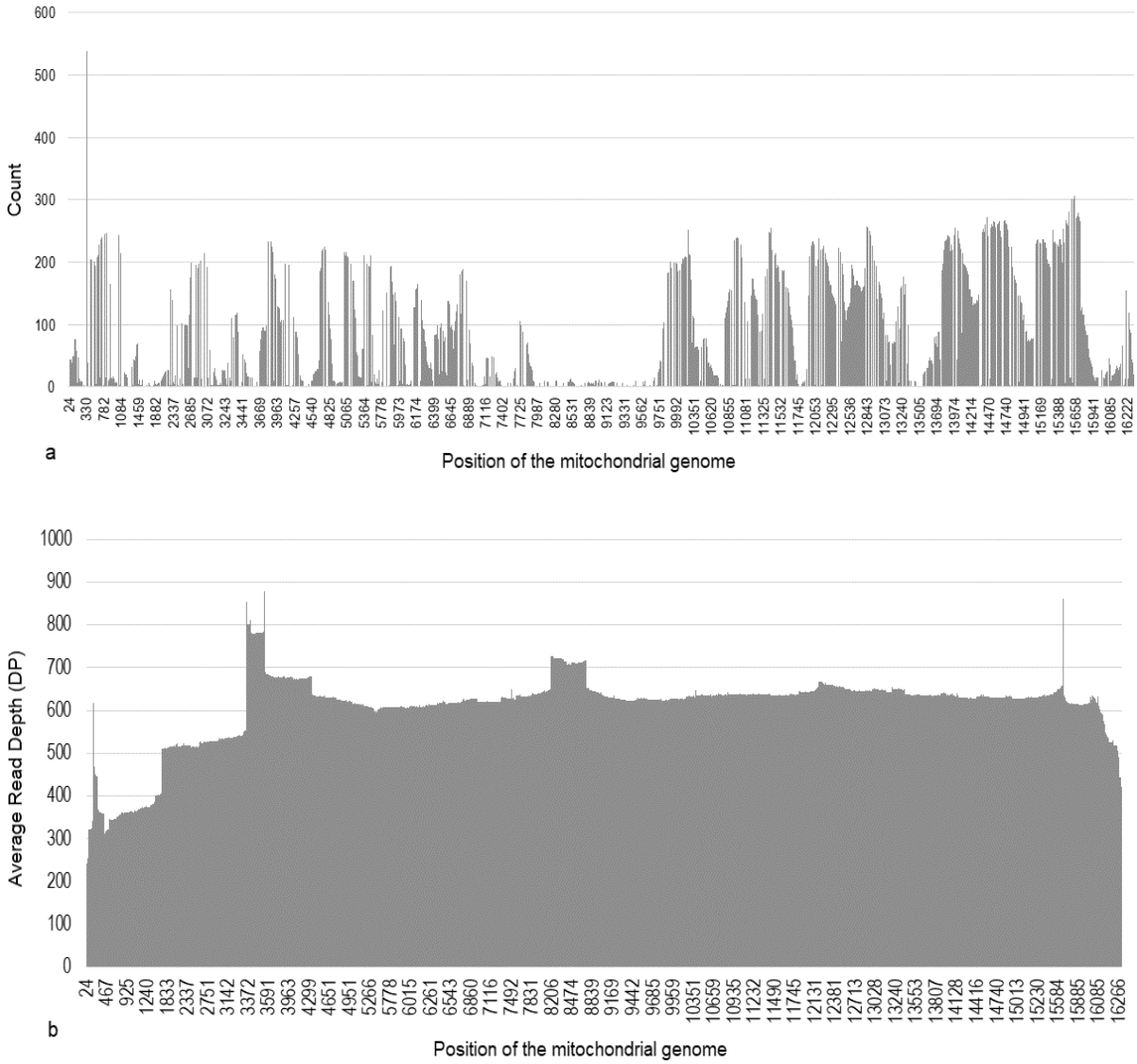


Figure S1. Distribution of heteroplasmic mitochondrial variants and average read depth in the filtered dataset of 2176 animals (sites with missing genotypes removed). Count of heteroplasmy in a position on cattle mitochondrial genome (a) and mean read depth (DP) across all animals at the positions in the mitochondrial genome (b).

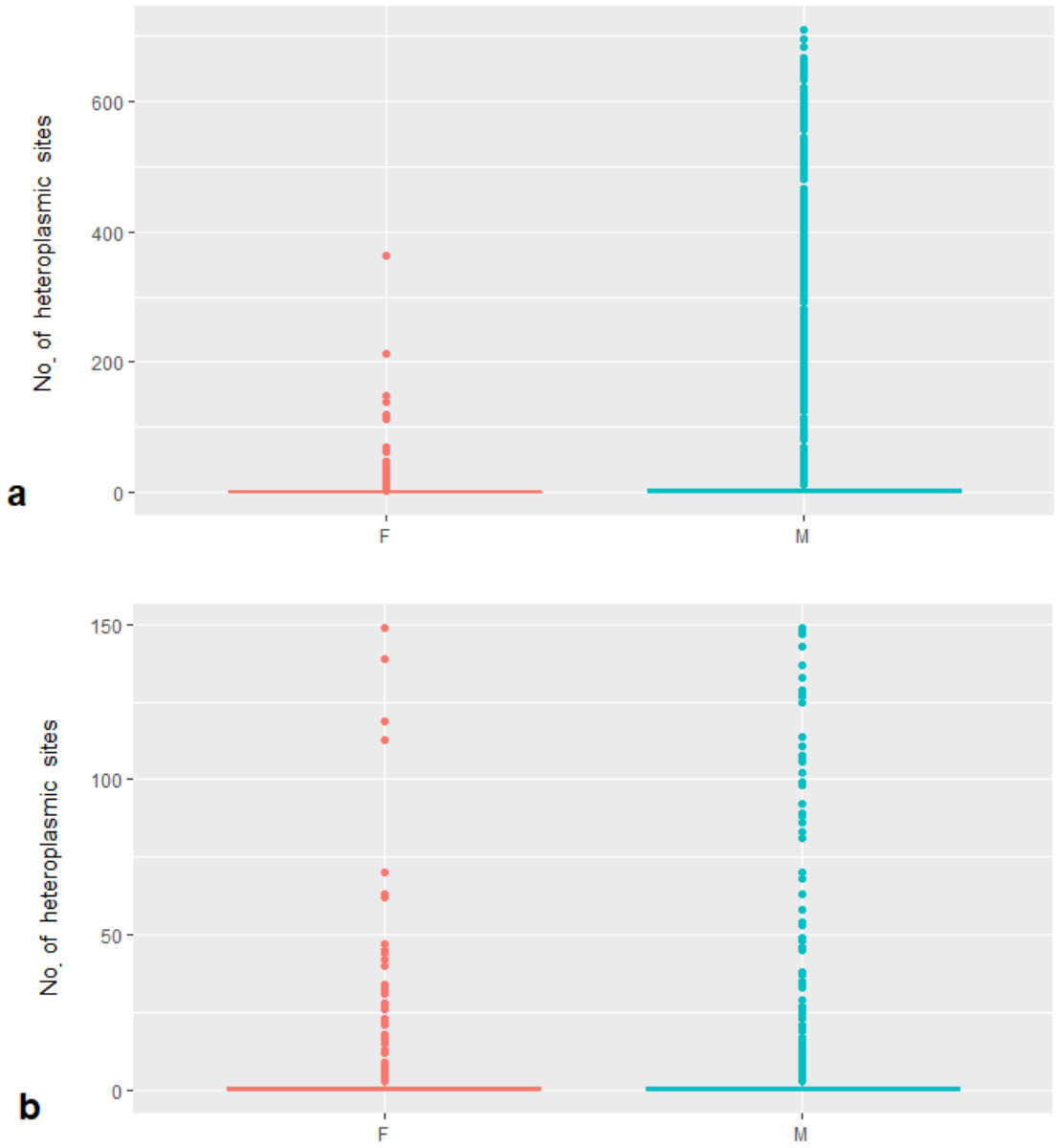


Figure S2. Box plots of the number of heteroplasmic sites per animal in individuals of two groups; F (non-semen: females and males with non-semen tissues sampled for DNA extraction) and M (males with either semen for DNA extraction or unknown tissue sample origin) before (a) and after (b) filtering out animals with > 150 heteroplasmic sites.

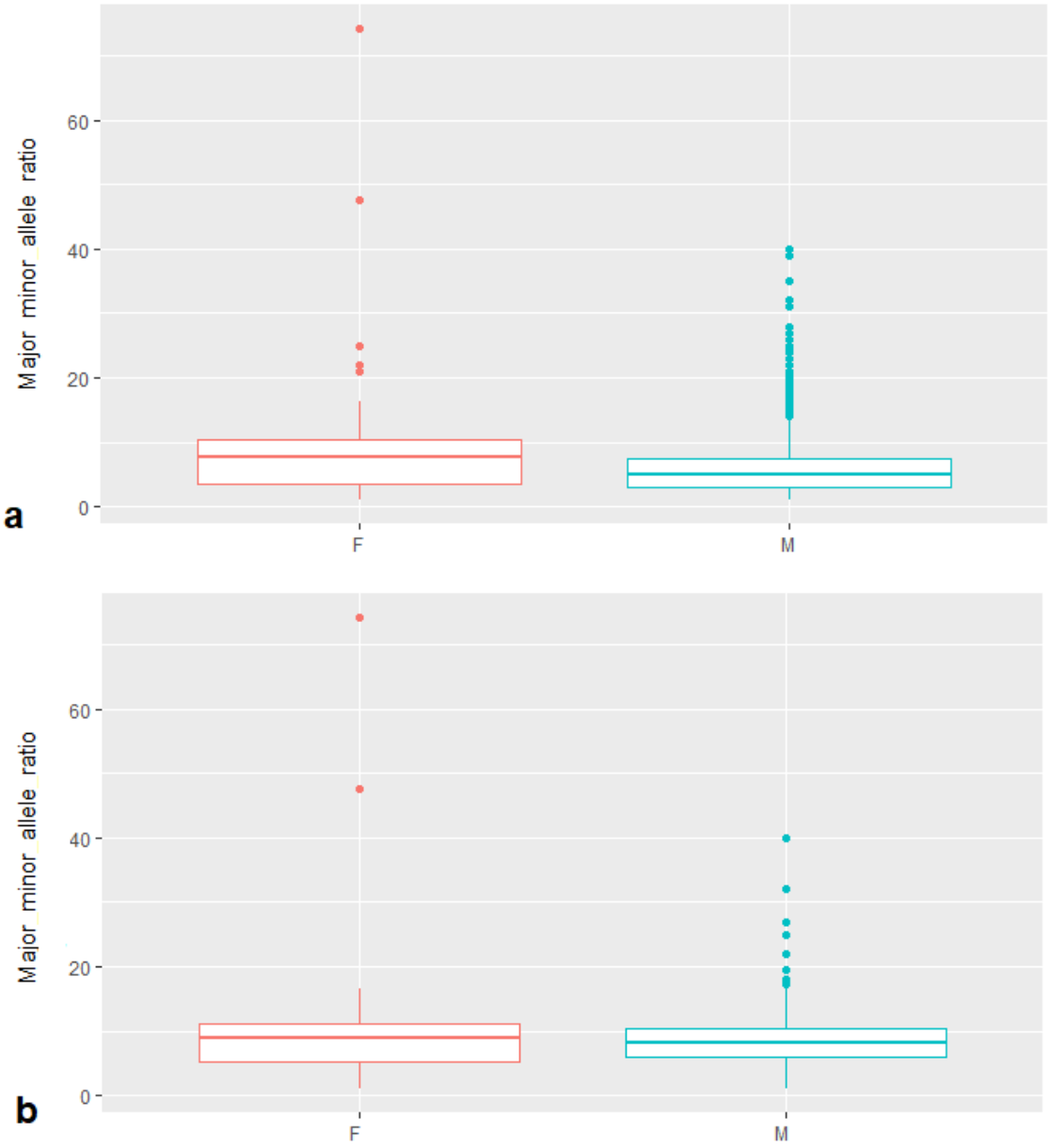
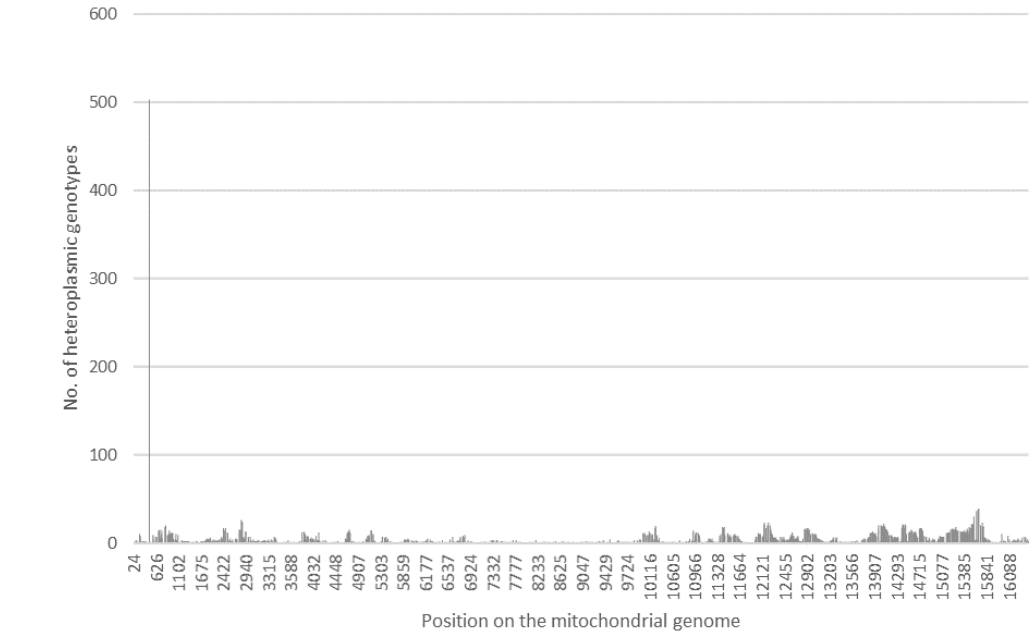
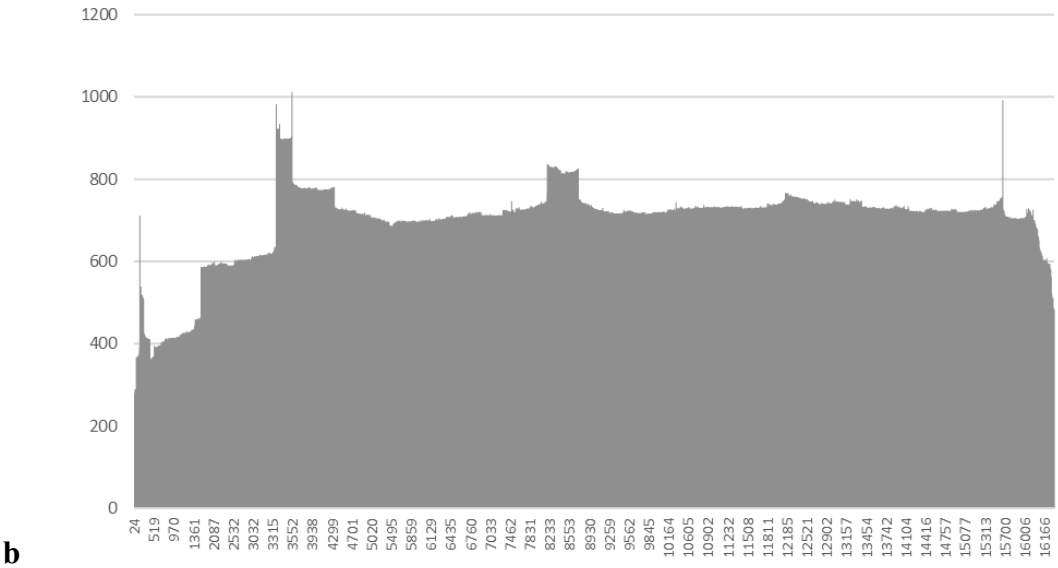


Figure S3. Box plots of allelic depth ratio of major and minor allele at the heteroplasmic sites of animals in two groups F (non-semen: females and males with non-semen tissues sampled for DNA extraction) and M (males with either semen for DNA extraction or unknown tissue sample origin) showing before (a) and after (b) filtering out animals with >150 heteroplasmic sites.



a



b

Figure S4. Distribution of mitochondrial heteroplasmic variants and average read depth in the final dataset (1,883 animals) after removing all individuals with >150 heteroplasmic MT variants. a) Count of heteroplasmy in a position on cattle mitochondrial genome (a). Mean read depth (DP) across all animals at the positions in the mitochondrial genome (b).

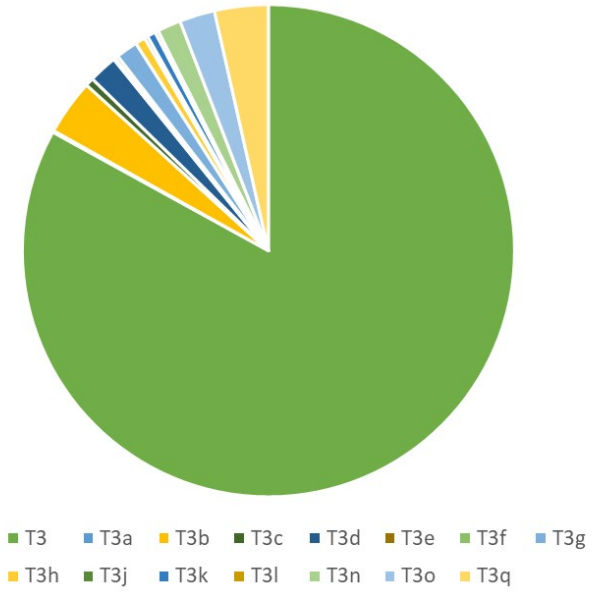
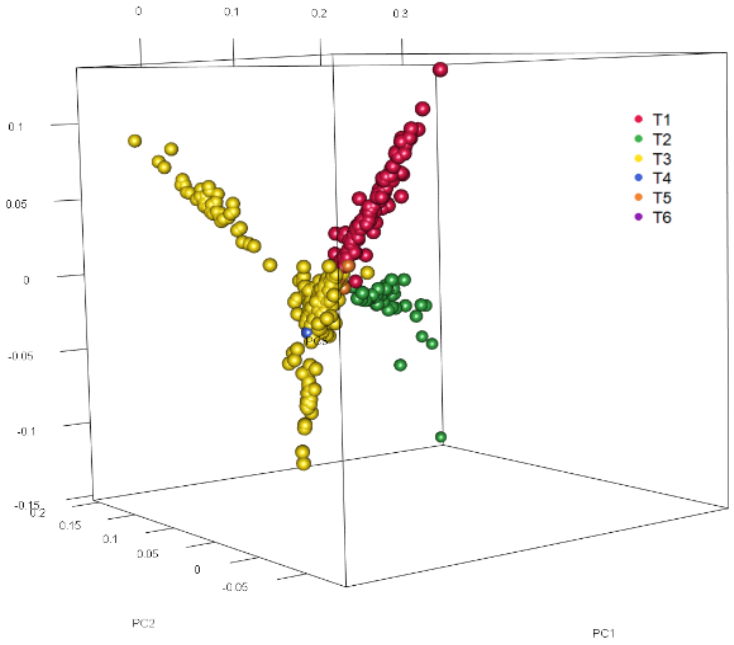
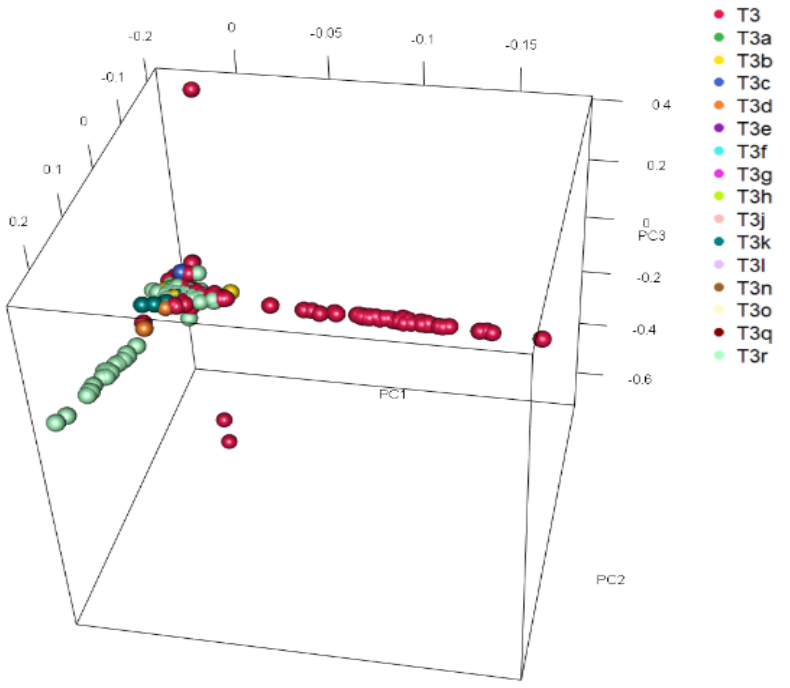


Figure S5. Pie-chart showing the subgroups within T3 major haplogroup and the composition in a total of 1502 animals.

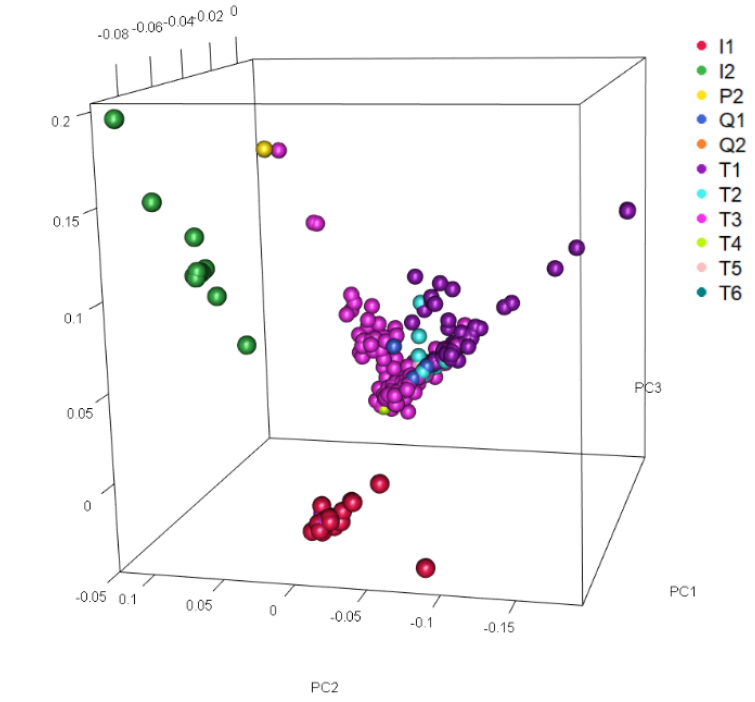


a

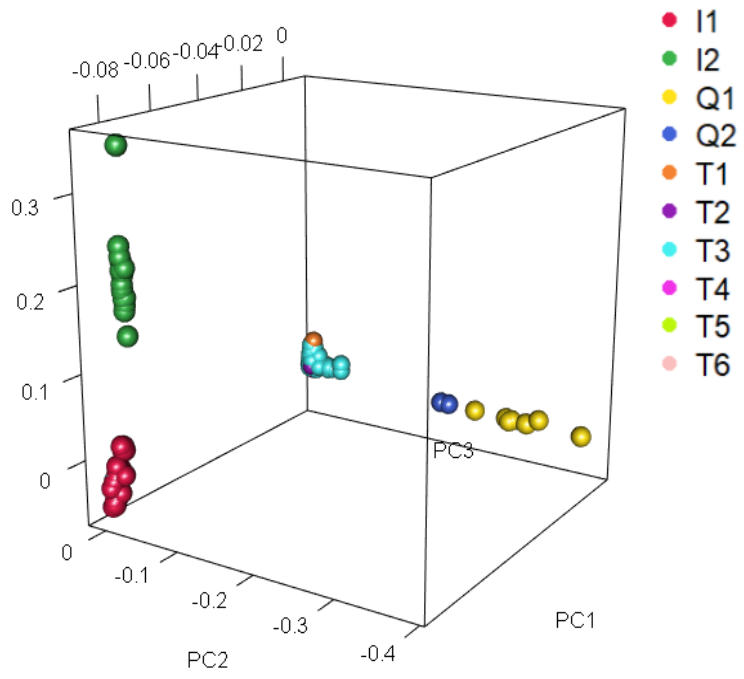


b

Figure S6. PC plot of PC1, 2 and 3 based on a genomic relationship matrix derived from whole mitochondrial DNA sequence variants annotated with haplogroup predicted from MitoToolPy. Plot annotated with T haplogroups (a) and T3 (b) sub-haplogroups.



a



b

Figure S7. PC plot based on the genomic relationship matrix of mitochondrial D-loop sequence variants (a) and variants from the non-D-loop region (b) variant sequences annotated with haplogroups predicted using MitoToolPy.

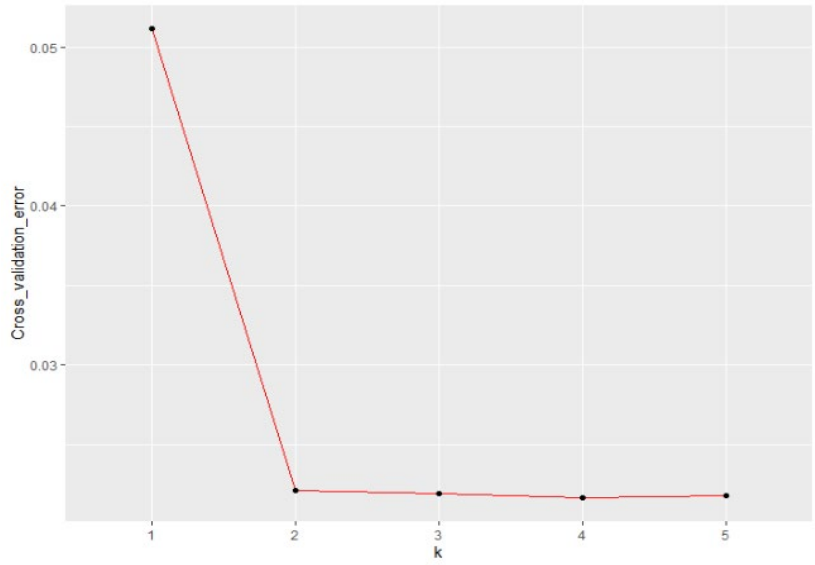


Figure S8. Determining the optimal *a priori* "k" value (i.e. number of population sub-groups) to use in Admixture software for 1,883 mitogenomes in our study. Plot shows changes in cross-validation errors using prior values of 1 to 5 for k in Admixture with the optimal value indicating *k*=4.

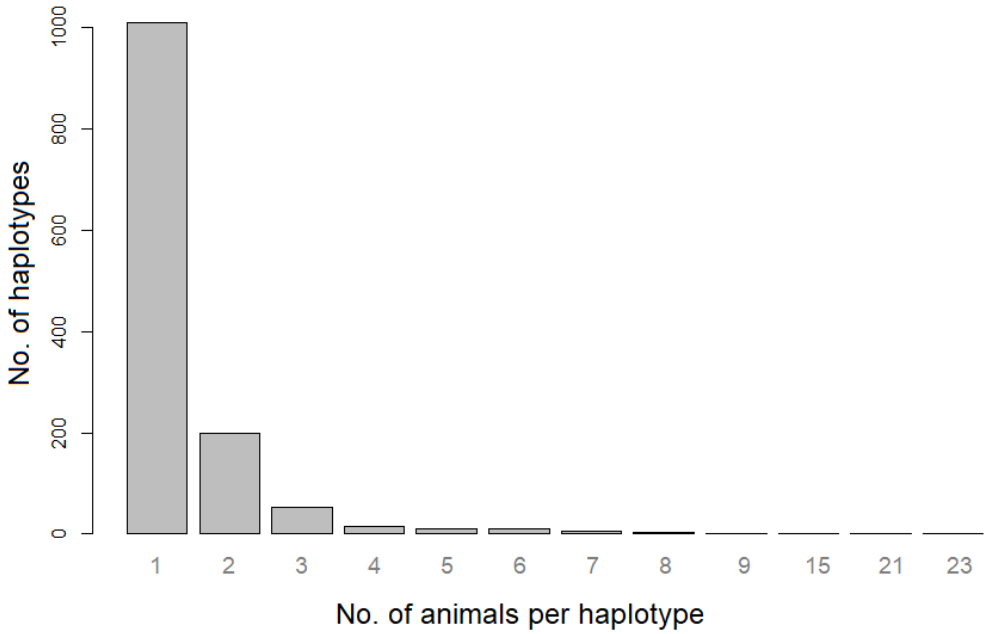


Figure S9. Frequency distribution of haplotypes with the number of animals per haplotype. Haplotypes were defined using all variant sites (excluding indels) across the mitogenome of the study animals.

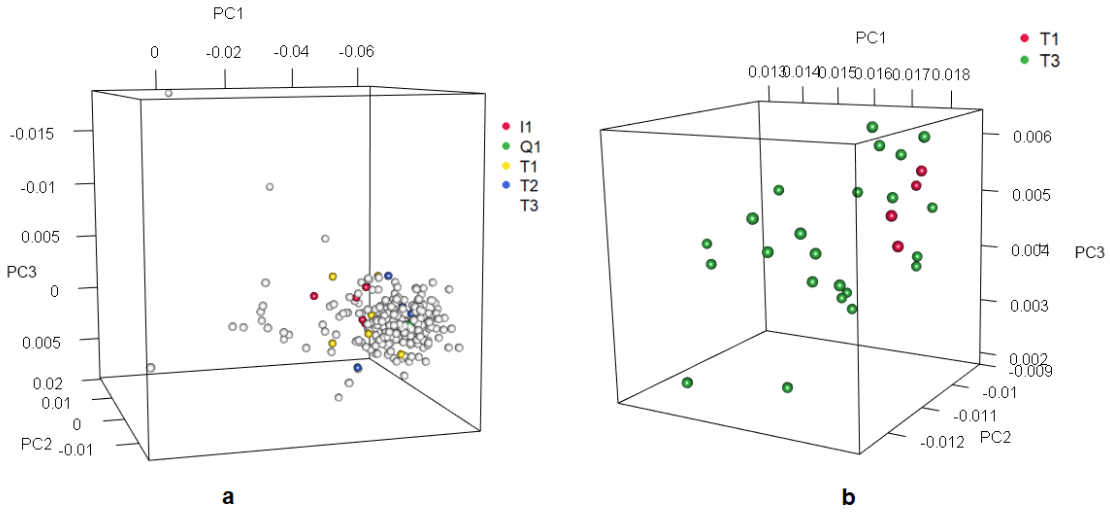


Figure S10. PC plot of Holstein (a) and Jersey (b) breeds based on a genomic relationship matrix (GRM) of the taurus only animals (N=1,451) in the study set (N=1,883). The GRM was constructed from a custom 45K autosomal SNP panel using GCTA.

Supplementary Tables

Table S1. Variants used in prediction of cattle mitochondrial haplogroup lifted over from Bovine Ref Sequence (V00654) to ARS-UCD1.2_M reference sequence (*see Appendix 3 page 156*)

Table S2. List of variants from animal sequences available at NCBI, aligned to ARS-UCD1.2_M Ref and used by MitoToolsPy to dertermine haplogroup (*see Appendix 3 page 166*)

Table S3. Haplogroups and source of samples of 1,883 animals in the filtered dataset determined from the variant sequence using MitoToolPy

Breed	Source	Haplogroups											
		N	I1	I2	P2	Q1	Q2	T1	T2	T3	T4	T5	T6
Achai	PAK	4	1	1						2			
Afrikander	AUS	3						3					
Alentejana		1								1			
Angus	AUS, CAN, NZL, USA	103								97			6
Angus German	USA	1								1			
Angus Lowline	AUS	2						2					
Angus Red	CAN, NZL, USA	10								10			
Angus Simmental		1								1			
Ankole	UGA	9						9					
Australian Lowline	GBR	1											1
Ayrshire Finnish	FIN	1								1			
Beefmaster	USA	2						1		1			
BelgianBlueHolstein		3								3			
BelgianBlueLimousin	IRE	1								1			
Benishangul	ETH	3						3					
Bhagnari	PAK	4	3	1									
BlondedAquitaine	FRA	15						2		13			
BohaiBlack		5	2							3			
Bohuskulla	SWE	3								3			
Boran	KEN	19						19					
Boskarin		1								1			
Brahman	AUS	18	1					12		5			
BrownSwiss	FRA, CHE, DEU, ITA, USA	84						1		81		2	
Buryat	RUS	20	2						2	16			
Busa	FIN	8								8			
Cabannina	ITA	2								2			
ChaidamuYellow		5	2							3			
Charolais	CAN, FRA, USA	33							1	32			
CharolaisAngus		1								1			
CharolaisRedAngus		1								1			
Chianina	ITA	5				2	2			1			
Cholistani	PAK	5	4	1									

Breed	Source	Haplogroups											
		N	I1	I2	P2	Q1	Q2	T1	T2	T3	T4	T5	T6
Cloned-polledDairyBull		2								2			
Composite	CAN, AUS	13	4					6		3			
Corriente		2								2			
CostenoConCuernos		2						1		1			
Crossbreed	CAN, FRA	21								21			
Crossbreed(HO62.5%;MO25%;JE12.5%)		1								1			
Dabieshan		3	2							1			
Dajal	PAK	4	4										
DanishRedDairy		2								2			
Dengchuan	CHN	1								1			
DSN*	DEU	47							1	46			
Dexter	AUS	2						1		1			
Dhanni	PAK	5	3	2									
Dianzhong		5	2	1						2			
Droughtmaster	AUS	1								1			
EasternFinncattle	0	5								5			
Eringer	CHE	3								3			
Evolčne	CHE	2							1	1			
Fjäll	SWE	11								11			
Fleckvieh	DEU	15							1	14			
Fogera	ETH	1						1					
Gabrialli	PAK	4	1	1						2			
Galloway		1								1			
GallowayBelted	AUS	3	1							2			
Gelbvieh	CAN	13								13			
Gir		1		1									
Goffa	ETH	3						3					
Guangfeng		4	3							1			
Guanling	CHN	1								1			
Hanwoo		24			1			2	2	19			
Hariana		1	1										
Hasake	CHN	5								5			
Hereford	AUS, CAN, NZL, RUS, USA	44	2	1						41			
HerefordMiniature	AUS	2								2			
Holstein	AUS, CAN, CHE, CHN, DEU, DNK, FRA, GBR, NLD, USA	267	5			1		6	4	251			
HolsteinCharolais	DEU,	19								19			
HolsteinFriesian	DEU, USA	35							1	34			
HolsteinHereford	USA	9								9			
HolsteinJerseyF1Crossbred	USA	1								1			
HolsteinLimousinF1Crossbred	USA	2								2			
HolsteinRed	DEU	3						1		2			
HolsteinSimmental		4								4			
IranAdmixed		9				4		1		4			
JapaneseNative		9						1		8			

Breed	Source	Haplogroups											
		N	I1	I2	P2	Q1	Q2	T1	T2	T3	T4	T5	T6
Jersey	AUS, USA	27						4		23			
Jian		4	3							1			
JiaxianRed		5	3							2			
Jinjiang		4	3							1			
Kalmykian	FIN	10							1	8		1	
Kangayam		1		1									
Kazakh		9	2					1	3	3			
KazakhWhiteheaded	RUS	5								5			
Kenana	SUD	5						5					
Lagune		1						1					
Leiqiong		3	3										
Limia		1								1			
Limonero		9						1		8			
Limousin	CHE, FRA,	27								26		1	
LimousinHereford		2								2			
LimousinHolstein	CHE	1								1			
LimousinxBrownSwiss	0	1								1			
Lingnan		7	6						1				
Lohani	PAK	1	1										
Luxi		5	5										
MaineAnjou	USA	1								1			
Maremmana	USA	1								1			
Maronesa	USA	1								1			
Menggu	CHN	11							1	10			
ModernDanishRed	DNK	23							1	22			
Mongolian		7	1						1	5			
Montbeliarde	FRA	17							1	15		1	
Muturu	NGR	10						10					
Nanyang		4	3						1				
NariMaster	PAK	4	2	1						1			
NDama	BEN, GUI,	12						12					
Nelore		2						1		1			
NordicRedCattle		2								2			
Normande	FRA	7								7			
NorthernFinncattle	FIN	9								9			
NorwegianRed	NOR	222							6	21 6			
Nuer	ETH	1						1					
Ogaden	ETH	9						9					
OriginalBraunvieh	CHE	17							3	12		2	
Ottonese	ITA	1							1				
Pajuna		1								1			
Parthenaise	FRA	2								2			
PezzataRossaItaliana		1								1			
Piedmontese	CHE	2								2			
PiedmonteseNormande		1								1			
Pinzgauer	CHE	1								1			

Breed	Source	Haplogroups											
		N	I1	I2	P2	Q1	Q2	T1	T2	T3	T4	T5	T6
Podolian-Serbia	FIN	10								4		6	
Podolica		1								1			
Qinchuan	CHN	2								2			
RedSindhi	PAK	3	2	1									
Rendena	ITA	2								2			
Ringamålako	SWE	2								2			
Rödkulla	SWE	9							1	8			
Romagnola	IRL, ITA	10				1	1	2	1	5			
RougeDesPres	FRA	3								3			
Sahiwal	PAK	5	2	3									
Salers	FRA	4								4			
SanMartinero		2						1		1			
SantaGertrudis	AUS	1								1			
Sayaguesa		1								1			
ScottishHighland	CHE, GBR	7								7			
Senepol	AUS	12						5		7			
Shaiwal	IND	2		2									
Sheko	ETH	3						3					
Shorthorn		1	1										
SichuanIndigenous		1	1										
Sikias		1							1				
Simmental	AUT, CAN, CHE	32								32			
SimmentalFleckviehPezzata rossa	AUT, CHE, ITA	3								3			
SimmentalxAngus	CHE	1								1			
SombaTongaleseModern		1						1					
SouthAnatolianRed		1							1				
SwedishRed	SWE	1								1			
SwissFleckvieh	CHE	1								1			
TexasLonghorn		1								1			
Tharparkar	PAK, IND	7	7										
TharparkerModern		1	1										
TibetanCattle		2								2			
TraditionalDanishRed	DNK	1								1			
Tuxer	CHE	1								1			
TyroleanGrey	AUT, CHE, ITA	7								7			
UgandaAdmixed		26						26					
UkrainianGrey	FIN	8								8			
Unknown		126	4	1				6	7	10 8			
Väneko	SWE	5								5			
Vechur		1		1									
Vorderwaelder	DEU	1								1			
WagyuModern		1									1		
Wandong		2	2										
Wannan		7	3							4			
Weining		5	3							2			
Wenshan		4	4										

Breed	Source	Haplogroups											
		N	I1	I2	P2	Q1	Q2	T1	T2	T3	T4	T5	T6
WesternFinncattle		4								4			
Xizang	CHN	2								2			
Xuanhan		5	2							3			
Yakut	FIN, RUS	35							6	29			
Yanbian	CHN	8						1		7			
Yaroslavl	RUS	10								10			
Zaobei		5	4							1			
Zebu	ETH	2						2					
ZebuIndian		1	1										
	Total	*1883*	*112*	*19*	*1*	*8*	*3*	*167*	*50*	*1502*	*1*	*13*	*7*

DSN* Deutsches Schwarzbuntes Niederungsrind, AUS Australia, AUT Austria, BEN Benin, CAN Canada, CHN China, DNK Denmark, ETH Ethopia, FIN Finland. FRA France, GUI Guinea, DEU Germany, IND India, IRE Ireland, NLD Netherlands, NZL New Zealand, GBR United Kingdom, NGR Nigeria, PAK Pakistan, RUS Russia, SCL Scotland, SPN Spain, SWE Sweden, CHE Switzerland, UGA Uganda, USA United States of America

Table S4. Genotypes and allele depth at key position distinguishing the haplogroup T1b1 and T3r for Jersey with T1 haplogroup.

		POS				
	Alleles	**7544**	**16024**	**16052**	**16115**	**16257**
	REF:	**G**	**G**	**C**	**T**	**T**
Animal	**ALT:**	**A, C**	**A, T**	**T**	**C**	**C**
SAMN08612491		1/1:9,234	1/1:4,302	1/1:3,309	1/1:0,272	1/1:3,201
SAMN19491856		0/1:3,32,0	1/1:0,23	1/1:0,22	1/1:0,16	1/1:0,22
SAMN08612497		1/1:5,132	1/1:1,122	1/1:0,127	1/1:0,125	1/1:2,96
SAMN19491865		1/1:1,205	1/1:0,174	1/1:0,172	1/1:0,152	1/1:1,139

Table S5. Annotation of the private variants* common to a specific groups of animals within a haplogroup.

Haplogroup (N)	bp position on ARS (BRS)	N	Annotation
I_1 (112)	1497(1495)	64	Spec, NCTE, rRNA
	5707(5705)	9	Spec, Syno, COX1
	8648(8646)	64	Spec, Miss, ATP6
	14029 (14027)	64	NonS, Miss, ND6
	16086(16084)	115	NonS, Up/down
	16112 (16110)	4	NonS, Up/down
I_2 (19)	14833 (14831)	13	Spec, Miss, CYTB
	16087 (16085)	19	NonS, Up/down
Q_1 (6)	5718 (5716)	2	NonS, Miss, COX1
	6388 (6386)	3	Syno, COX1
T_1 (57)	3684 (3682)	13	Spec, Syno, ND1
	15675 (15673)	8	NonS, NCTE,tRNA
	16262 (16260)	9	NonS, Up/Down
	2579 (2577)	7	Spec, NCTE rRNA
	4714 (4712)	7	Spec, Miss, ND2
	6882 (6880)	7	Spec, Miss, COX1
	10435 (10433)	7	Spec, Miss, ND4L
	4984 (4982)	2	NonS, Miss, tRNA
T_1b (3)	5898 (5896)	2	NonS, T Miss, G Stop_gained COX1
	10605 (10603)	3	Spec, Synon, ND4
	12416 (12414)	2	NonS, Miss, ND5
	15961 (15959)	2	NonS, Up/down
$T_1b_1b_1$(24)	12404(12402)	9	NonS, Miss, ND5
	12740 (12738)	4	Spec, Miss, ND5
	14773 (14771)	5	Spec, Miss, ND5
$T_1b_1b_1a_3$ (9)	4742(4740)	3	NonS, Syno, ND2
$T_1b_1b_1c_1$(3)	13464 (13466)	2	Miss, ND5
T_1C (11)	15462 (15460)	2	Spec, Miss,CYTB
	9569 (9567)	5	NonS, Syno, COX3
T_2 (45)	16137(16135)	12	NonS, Up/down
	5703 (5701)	13	Spec, T Miss, A Stop gained
	11370 (11368)	9	Spec, Miss, ND4
T_3 (55)	16006 (16004)	59	Spec, Up/down
	8245 (8243)	32	Spec, Syno, ATP8
	6237 (6235)	5	Miss, COX1
	16076 (16074)	15	NonS, Up/down

	16233 (16231)		~Spec, Up/down
T_3c (5)	4002 (4000)	5	NonS, Miss, ND1
T_3k (5)	16114 (16112)	3	NonS, Up/down
T_3r_5	16044 (16042)	39	~ Spec, (T3a1a), Up/down
	16234 (16232)	4	NonS, Up/down
T_5	9234 (9232)	3	NonS, Miss, COX3
	10881 (10879)	3	NonS, Miss, ND4

*variants not used for determining haplogroups in MitoToolPy,

N = No. of animals in the haplogroup,

n = number of animals showing private variants,

Spec = Specific (i.e. Private variants from MitoToolPy only found within the group of individuals within a identified haplogroup),

NonS = Nonspecific (i.e. Private variants from MitoToolPy within a group of animals within a haplogroup but also present among other individuals in other haplogroups),

Variants: up/down upstream/downstream gene variants,

NCTE = Non-coding transcript exon, Miss Missense,

Syno = Synonymous

Table S6. Haplotypes specific to breed and prevalent in samples sourced from two countries.

Shared haplotypes	**No. of animals**	**Breed**	**Country**
Hap_12	3	Angus	Canada, USA
Hap_20	8	Angus	Canada, USA
Hap_121	2	Brown Swiss	Switzerland, Germany
Hap_145	3	Brown Swiss	Italy, USA
Hap_306	6	Holstein	Canada, Switzerland
Hap_308	3	Holstein	Canada, Switzerland
Hap_332	3	Holstein	Denmark, Germany
Hap_404	3	Holstein	France, The Netherlands
Hap_415	2	Holstein	France, The Netherlands
Hap_446	2	Jersey Tyrolean	Australia, USA
Hap_901	2	Grey	Austria, Switzerland

Table S7. Number of animals in common ((≥ 5) between each specific Haplotype (determined by DnaSP) and Cluster (based on hierarchical clustering of the nucleotide difference between pairs)

Haplotype No.	No. of Anim^{+} in_hap	Cluster No.	No. of Anim^{+} in Cluster	No. of Anim^{+} common	% Concordance*
324	23	328	23	23	100%
7	21	7	21	21	100%
8	17	8	15	15	88%
197	15	198	14	14	93%
9	8	9	8	8	100%
855	8	866	8	8	100%
243	8	246	8	8	100%
20	8	20	8	8	100%
549	7	554	7	7	100%
24	7	24	7	7	100%
23	7	23	7	7	100%
19	7	19	7	7	100%
857	6	868	6	6	100%
825	6	835	6	6	100%
62	7	63	6	6	86%
505	6	510	6	6	100%
269	6	272	6	6	100%
1193	6	1209	6	6	100%
1192	6	1208	6	6	100%
1191	6	1207	6	6	100%
1080	6	1093	6	6	100%
786	5	795	5	5	100%
658	9	750	5	5	56%
544	5	549	5	5	100%
517	5	522	5	5	100%
515	5	521	5	5	100%
419	5	423	5	5	100%
306	6	310	5	5	83%
1138	5	1152	5	5	100%
10	5	10	5	5	100%
1029	5	1041	5	5	100%

Anim^{+} = Animal, * based on No. of animals in each Haplotype

Table S8. Number and percentage of genotypes correctly imputed in 307 masked sites per animal in 333 animals (102231 sites per run) and correct prediction of haplogroup* from the imputed genotypes. Results are shown for 50 independent tests of imputation.

Run†	Genotypes (out of 102231)		Haplogroups of 333 imputed animals	
	Correctly imputed	% Correctly imputed	No. correctly predicted	% of animal correct
1	101951	99.73%	330	99.10 %
2	101963	99.74%	333	100.00%
3	101987	99.76%	333	100.00%
4	101984	99.76%	333	100.00%
5	101921	99.70%	333	100.00%
6	101959	99.73%	332	99.70%
7	102036	99.81%	330	99.10%
8	101972	99.75%	333	100.00%
9	101938	99.71%	332	99.70%
10	101972	99.75%	332	99.70%
11	101983	99.76%	332	99.70%
12	101910	99.69%	333	100.00%
13	101953	99.73%	332	99.70%
14	102021	99.79%	330	99.10%
15	101963	99.74%	331	99.40%
16	101987	99.76%	332	99.70%
17	101976	99.75%	333	100.00%
18	101973	99.75%	332	99.70%
19	101991	99.77%	330	99.10%
20	101948	99.72%	328	98.50%
21	101999	99.77%	332	99.70%
22	101985	99.76%	332	99.70%
23	101964	99.74%	331	99.40%
24	101986	99.76%	332	99.70%
25	102005	99.78%	331	99.40%
26	101970	99.74%	331	99.40%
27	101978	99.75%	333	100.00%
28	101968	99.74%	333	100.00%
29	101925	99.70%	329	98.80%
30	102020	99.79%	333	100.00%
31	101993	99.77%	331	99.40%
32	101965	99.74%	331	99.40%
33	101985	99.76%	331	99.40%
34	101977	99.75%	331	99.40%
35	102000	99.77%	331	99.40%
36	101995	99.77%	331	99.40%
37	101957	99.73%	331	99.40%
38	101951	99.73%	331	99.40%
39	101995	99.77%	331	99.40%
40	101919	99.69%	331	99.40%

41	101990	99.76%	328	98.50%
42	102017	99.79%	328	98.50%
43	101973	99.75%	328	98.50%
44	101978	99.75%	328	98.50%
45	101938	99.71%	328	98.50%
46	101955	99.73%	333	100.00%
47	101960	99.73%	333	100.00%
48	102008	99.78%	333	100.00%
49	101985	99.76%	333	100.00%
50	102027	99.80%	332	99.70%
Mean	**101975**	**99.75%**	**331**	**99.49%**

† Run = results based on 50 independent tests

* Haplogroup predicted using MitoTool.Py

Chapter 4:

Expression of mitochondrial protein genes encoded by nuclear and mitochondrial genomes correlate with energy metabolism in dairy cattle

BMC Genomics **21,** 720 (2020). https://doi.org/10.1186/s12864-020-07018-7

Jigme Dorji, Christy J. Vander Jagt, Josie B. Garner, Leah C. Marett, Brett A. Mason, Coralie M. Reich, Ruidong Xiang, Emily L. Clark, Benjamin G. Cocks, Amanda J. Chamberlain, Iona M. MacLeod & Hans D. Daetwyler

Published: October 2020 (open access)

Dorji et al. BMC Genomics (2020) 21:720
https://doi.org/10.1186/s12864-020-07018-7

BMC Genomics

RESEARCH ARTICLE Open Access

Expression of mitochondrial protein genes encoded by nuclear and mitochondrial genomes correlate with energy metabolism in dairy cattle

Jigme Dorji[1,2*], Christy J. Vander Jagt[2], Josie B. Garner[3], Leah C. Marett[3], Brett A. Mason[2], Coralie M. Reich[2], Ruidong Xiang[2,4], Emily L. Clark[5], Benjamin G. Cocks[1,2], Amanda J. Chamberlain[2], Iona M. MacLeod[2] and Hans D. Daetwyler[1,2]

Abstract

Background: Mutations in the mitochondrial genome have been implicated in mitochondrial disease, often characterized by impaired cellular energy metabolism. Cellular energy metabolism in mitochondria involves mitochondrial proteins (MP) from both the nuclear (*Nu*MP) and mitochondrial (*Mt*MP) genomes. The expression of MP genes in tissues may be tissue specific to meet varying specific energy demands across the tissues. Currently, the characteristics of MP gene expression in tissues of dairy cattle are not well understood. In this study, we profile the expression of MP genes in 29 adult and six foetal tissues in dairy cattle using RNA sequencing and gene expression analyses: particularly differential gene expression and co-expression network analyses.

Results: MP genes were differentially expressed (DE; over-expressed or under-expressed) across tissues in cattle. All 29 tissues showed DE *Nu*MP genes in varying proportions of over-expression and under-expression. On the other hand, DE of *Mt*MP genes was observed in < 50% of tissues and notably *Mt*MP genes within a tissue was either all over-expressed or all under-expressed. A high proportion of *Nu*MP (up to 60%) and *Mt*MP (up to 100%) genes were over-expressed in tissues with expected high metabolic demand; heart, skeletal muscles and tongue, and under-expressed (up to 45% of *Nu*MP, 77% of *Mt*MP genes) in tissues with expected low metabolic rates; leukocytes, thymus, and lymph nodes. These tissues also invariably had the expression of all *Mt*MP genes in the direction of dominant *Nu*MP genes expression. The *Nu*MP and *Mt*MP genes were highly co-expressed across tissues and co-expression of genes in a cluster were non-random and functionally enriched for energy generation pathway. The differential gene expression and co-expression patterns were validated in independent cow and sheep datasets.

(Continued on next page)

* Correspondence: jigme.dorji@agriculture.vic.gov.au
[1]School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3083, Australia
[2]Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, VIC 3083, Australia
Full list of author information is available at the end of the article

(Continued from previous page)
Conclusions: The results of this study support the concept that there are biological interaction of MP genes from the mitochondrial and nuclear genomes given their over-expression in tissues with high energy demand and co-expression in tissues. This highlights the importance of considering MP genes from both genomes in future studies related to mitochondrial functions and traits related to energy metabolism.

Keywords: Mitochondria, Energy metabolism, Differential gene expression, Gene co-expression, Cattle

Background

There is growing evidence that mitochondrial dysfunction arises from variations in the mitochondrial genome and that their interplay with the nuclear genome has a role in mitochondrial diseases in humans, including metabolic disorders and diabetes [1–3]. Mitochondria and mitochondrial functions are critical for tissues with high energy requirement [4]. Energy is produced in mitochondria through a process of oxidative phosphorylation (OXPHOS). Besides energy production, mitochondria mediate programmed cell death (apoptosis), aging, calcium homeostasis, and signalling as reviewed in [5–7].

Mitochondrial proteins (MP) are the proteins localized in mitochondria and are key component to mitochondrial functions [8]. There are an estimated 1500 proteins in mitochondria of rats [9], participating as components of electron transport chain, metabolic pathways, and factors for replication, initiation and regulation in transcription and translation. To date, 1158 MP stand verified in human [10] and almost all MP (> 99%) are of nuclear origin (*Nu*MP) and imported into the mitochondria [11, 12] with the exception for 13 proteins (< 1%), which originate from the mitochondrial genome (*Mt*MP). Mitochondria have their own genome, which is inherited maternally [13–15]. The cattle mitochondrial genome is haploid with a small circular structure (~ 16.4 kb) with a coding region encoding for 37 genes (13 proteins, 22 tRNAs and 2 rRNAs) and a non-coding control region [16]. Mitochondrial DNA mutations in cattle have previously been shown to be associated with fertility and productivity [17–20], and environmental adaptability to high altitudes in yaks [21, 22]. Unlike nuclear DNA, mitochondrial genomes occur in multiple copies, and their numbers are relatively constant within a cell type and development stage but vary considerably among cell types [23–25].

Gene expression is referred to as one of a series of processes from gene activation to mature protein function that contributes to the expression of cellular phenotypes [26]. The expression of a gene is often specific to tissue types, and a notable example is the dominance of major milk protein transcripts in the bovine lactating mammary gland [27]. Gene expression is commonly studied using RNA sequencing (RNAseq) where the number of reads mapping to a gene (counts) is used to measure gene expression.

The characterization of gene expression, identification of gene function, gene-disease or gene-production associations from genome-wide gene expression [28] employs differential gene expression and co-expression network analyses. Differential gene expression compares the gene expression in a sample with another sample or group of samples. Gene co-expression analysis measures the correlation between the expression levels of genes and associates gene clusters with biological processes and facilitates prediction of gene function of previously unknown genes [29]. At a very local level, co-expression of small groups of genes results from being in close proximity [30, 31] and in chromosomal domains characterized by frequent internal DNA-DNA interactions known as topological association domains (TADs) [32].

Most RNAseq based gene expression analyses to date have focused on nuclear genes rather than genes from the mitochondrial genome [33]. Nonetheless, a comprehensive examination of MP genes from both genomes is central to understanding genome-genome interactions, their role in meeting specific energy demand, and development of mitochondrial diseases. Metabolic profiles and energy demands vary widely across organs and tissue types [34–36]. The varying demand for energy across tissues is possibly in part facilitated through tissue specific and differential expression of MP genes. Currently, a comprehensive study on the expression of MP genes (both *Nu*MP and *Mt*MP) across tissues is lacking in bovine, although the expression of individual or groups of *Mt*MP genes has been published as part of larger gene sets [33, 37]. Therefore, our study aimed to characterize MP gene expression across both adult and foetal tissues in dairy cattle. We used RNAseq of 35 tissues from two adult cows and two foetuses (29 adult and six foetal tissues) to investigate differential gene expression and gene co-expression. We validated our findings using publicly available RNAseq data for an additional dairy cow and three sheep.

Results

Differential expression of mitochondrial protein genes

Main cows: adult

In total, 16,166 genes including 1041 MP genes were available for analysis after filtering (out of 24,616 annotated Ensembl genes). A gene was considered as differentially expressed (DE) in a tissue if the expression was different from the average expression across all other tissues (LFC > |0.6|, FDR < 0.01). Across all genes, 13 to 40% of genes in total were DE in one or more tissues and as high as 50% each of the DE genes were over-expressed or under-expressed (Fig. 1). Table 1 provides a summary of the number of DE genes by category across tissues. The highest overall numbers of DE genes among the tissues were in blood leukocytes($N = 9218$), loin muscle ($N = 7560$), brain caudal lobe ($N = 7504$), and brain cerebellum ($N = 7161$), and the lowest in the ovary ($N = 3003$), omental fad pad ($N = 3008$), and mediastinal lymph node ($N = 3428$). The DE genes in heart, skeletal muscles and tongue were significantly enriched for OXPHOS, metabolic pathways and neurodegenerative diseases pathways, and enriched for metabolic pathways in liver and kidney cortex (Table 2).

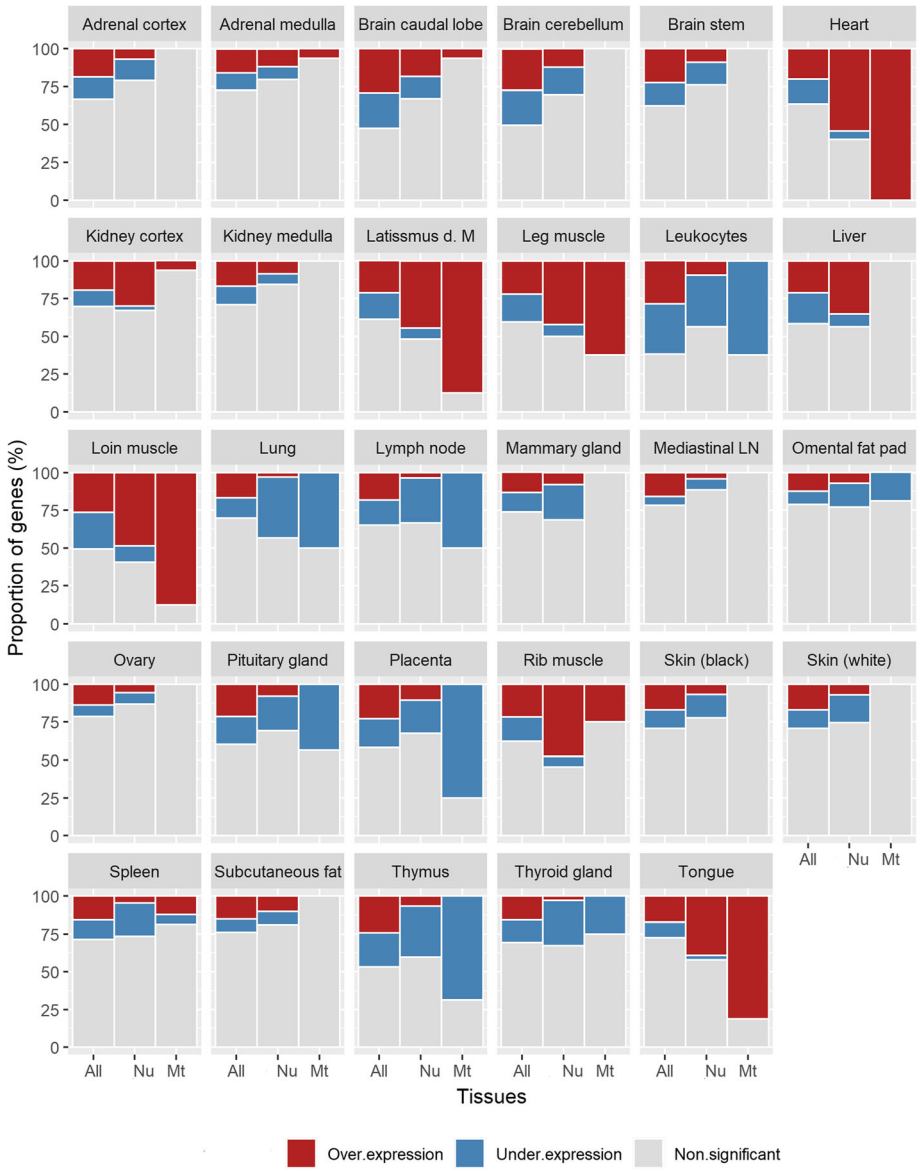


Fig. 1 Percentage of differentially expressed genes by gene categories for 29 tissues in the Main Cows dataset. m. = muscle, LN = lymph node; Gene category: All = All protein coding genes from nuclear and mitochondrial genomes, Nu = Mitochondrial protein coding genes from the nuclear genome (*Nu*MP), Mt = Mitochondrial protein coding genes from the mitochondrial genome (*Mt*MP)

Table 1 Number of differentially expressed (DE) genes in tissues by gene categories averaged for two cows in the Main Cows dataset

Tissue	MG ♁ (37)			***Nu***MP ⁎ (1041)			All ♨ (24,616)			***Mt***MP ♨ (13)	***Mt*** tRNA ¥	***Mt*** rRNA ♈
	Over	Under	Total DE	Over	Under	Total DE	Over	Under	Total DE			
Adrenal cortex	0	0	0	66	157	223	2018	2532	4550	0	0	0
Adrenal medulla	2	0	2	124	89	213	1620	2260	3880	0	0	2
Brain caudal lobe	2	0	2	190	151	341	3354	4150	7504	0	0	2
Brain cerebellum	1	0	1	120	187	307	3282	3879	7161	0	0	1
Brain stem	0	0	0	95	161	256	1911	3294	5205	0	0	0
Heart	17	0	17	576	53	629	2185	2833	5018	13	2	2
Kidney cortex	5	0	5	315	30	345	1465	2851	4316	3	0	2
Kidney medulla	0	0	0	93	58	151	1618	2444	4062	0	0	0
Latissimus dorsi M.	14	0	14	481	74	555	2668	2970	5638	13	1	0
Leg muscle	14	0	14	456	80	536	2773	3083	5856	13	1	0
Leukocytes	0	9	9	99	368	467	5270	3948	9218	9	0	0
Liver	0	0	0	360	82	442	2914	3124	6038	0	0	0
Loin muscle	14	0	14	514	114	628	3853	3707	7560	13	1	0
Lung	0	5	5	31	430	461	1793	2565	4358	5	0	0
Lymph node	0	7	7	38	306	344	2649	2562	5211	7	0	0
Mammary	0	0	0	86	250	336	1805	1864	3669	0	0	0
Mediastinal LN	0	0	0	46	77	123	993	2432	3425	0	0	0
Omental fat	0	2	2	78	163	241	1107	1901	3008	0	0	2
Ovary	0	0	0	58	70	128	985	2018	3003	0	0	0
Pituitary gland	0	5	5	73	251	324	2522	3033	5555	5	0	0
Placenta	0	10	10	110	232	342	2842	3267	6109	10	0	0
Rib muscle	7	0	7	500	70	570	2325	2933	5258	7	0	0
Skin black	0	0	0	79	176	255	1814	2574	4388	0	0	0
Skin white	0	0	0	79	205	284	1780	2554	4334	0	0	0
Spleen	2	0	2	47	260	307	2003	2449	4452	0	0	2
Subcutaneous fat	0	0	0	105	95	200	1299	2238	3537	0	0	0
Thymus	0	10	10	63	363	426	3509	3208	6717	10	0	0
Thyroid	0	2	2	33	329	362	2335	2335	4670	2	0	0
Tongue	14	0	14	413	30	443	1369	2365	3734	13	1	0

() total number of genes in a category; Over = Over-expression, Under = under-expression; MG ♁ Genes from mitochondrial genome including tRNA and rRNAs, *Nu*MP ⁎ Mitochondrial protein genes encoded by the nuclear genome, *Mt*MP ♨ Mitochondrial protein genes encoded by the mitochondrial genome, *Mt* tRNA ¥ Mitochondrial transfer RNA, *Mt* rRNA ♈ Mitochondrial ribosomal RNA, All ♨ all genes from nuclear and mitochondrial genomes, *M* Muscle, *LN* Lymph node

More than 99% of the MP genes originate from the nuclear genome (*Nu*MP). The proportion of DE *Nu*MP genes across the tissues varied from 12 to 60% with higher proportions (> 50%) in heart and skeletal muscles. The proportion of under or over-expressed DE *Nu*MP genes within tissues varied considerably. A relatively greater proportion of *Nu*MP genes were over-expressed in heart, kidney cortex, skeletal muscles, and tongue, and under-expressed in blood leukocytes, lymph nodes, placenta, lungs, mammary, and thymus (Fig. 2). The expression of *Nu*MP genes was similar between animals in the Main Cows as indicated by the clustering together of same tissues, with the exception of five tissues (Fig. 2; adipose, ovary, kidney cortex, and leukocytes).

In contrast to *Nu*MP, differential expression of *Mt*MP genes were observed in less than 50% of tissues (14 out of 29 tissues). The proportion of DE *Mt*MP genes within tissues ranged widely from 0 (no genes) to 100% (all 13 *Mt*MP genes). Specifically, *Mt*MP genes were 100% DE in heart, leg muscle, latissimus dorsi muscle, loin muscle, and tongue, and ranged between 50 and 75% in other tissues (leukocytes, placenta, thymus, rib muscle,

Table 2 KEGG functional annotation of overall differentially expressed genes of selected tissues with the largest number of genes averaged across two cows in the Main Cows dataset

Tissues	Enrichment	No. of genes (Overlap ***Nu***MP[a])	Adj. p	Other pathways
Heart	Oxidative phosphorylation	101 (93)	$3.3e^{-53}$	Parkinson's disease, Alzheimer's disease, Huntington's disease, NAFLD, carbon metabolism, Metabolic pathways, cardiac muscle contraction, TCA cycle,
Leg muscle	Oxidative phosphorylation	83 (38)	$1.0e^{-32}$	Parkinson's disease, Alzheimer's disease, NAFLD, Huntington's disease, Carbon metabolism, metabolic pathways, Proteasome, cardiac muscle contractionBiosynthesis of antibiotics
Loin muscle	Oxidative phosphorylation	89 (86)	$5.8e^{-39}$	Parkinson's disease, Alzheimer's disease, NAFLD, Huntington's disease, Huntington's disease, carbon metabolism, Proteosome, Cardiac muscle contraction
Rib muscle	Oxidative phosphorylation	86 (39)	$1.8e^{-35}$	Parkinson's disease, Alzheimer's disease, NAFLD, Huntington's disease, Metabolic pathways, carbon metabolism, Proteasome
Tongue	Oxidative phosphorylation	98 (91)	$3.0e^{-50}$	Parkinson's disease, Alzheimer's disease, NAFLD, Huntington's disease, Metabolic pathways, carbon metabolism, cardiac muscle contraction,
Kidney cortex	Metabolic pathways	312 (118)	$3.9e^{-28}$	Biosynthesis of antibiotics, Carbon metabolism, Valine, leucine and isoleucine degradation, Glycine, serine and threonine metabolism, tryptophane metabolism, Fatty acid metabolism
Kidney medulla	Focal adhesion	67 (0)	$3.0e^{-10}$	Tight junction, calcium signalling pathway, Gastric acid secretion. ECM-receptor interaction, cGMP-PKG signalling pathway, Gastric acid secretion, Dilated cardiomyopathy
Liver	Metabolic pathways	387 (123)	$4.1e^{-49}$	Biosynthesis of antibiotics, Peroxisome, valine, leucine and isoleucine degradation, complement and coagulation cascades, fatty acid degradation, tryptophan metabolism, carbon metabolism
Brain caudal lobe	Axon guidance	57 (0)	$8.0e^{-18}$	Glutamatergic Synapse, Domaminergic synapse, MAPK signalling pathway, Adrenergic signalling in cardiomyocytes, Retrograde endocannabinoid signalling, cAMP signalling pathway, Synaptic vesicle cycle, GABAergic synapse, Morphine addiction, Glutamatergic synapse, Circadian entrainment, Dopaminergic synapse
Brain cerebellum	Glutamatergic synapse	49 (2)	$3.1e^{-14}$	GABAergic synapse, Retrograde endocannabinoid signalling, Morphine addiction, Circadian entrainment, Dopaminergic synapse, cAMP signalling pathway, Adrenergic signalling in cardiomyocytes, axon guidance,
Brain stem	GABAergic synapse	36 (2)	$3.1e^{-9}$	Glutamergic synapse, Morphine addiction, Retrograde endocannabinoid signalling, Dopaminergic synapse, Circadian entrainment
Leukocytes	Chemokine signalling pathways	74	$4.7e^{-15}$	Focal adhesion, leukocyte transendothelial migration, Rap1 signalling pathway, natural killer cell mediated cytotoxicity, regulation of actin cytoskeleton, B cell receptor signalling pathway

[a]we show the number of genes in the top enriched pathway that overlap mitochondrial proteins

and lymph node). Unlike *Nu*MP genes, all DE *Mt*MP genes were expressed in a single direction (i.e. either all over-expressed or all under-expressed) meaning every DE *Mt*MP gene was over-expressed in heart, tongue, muscles and kidney cortex, and under-expressed in blood leukocytes, placenta, lymph node, pituitary, thymus, and thyroid (Fig. 3). Further, there were similarities between the expression of DE *Mt*MP and *Nu*MP genes within a tissue. For instance, every tissue showing over-expression of DE *Mt*MP genes invariably showed predominant over-expression of *Nu*MP genes and similarly for under-expression. In addition to *Mt*MP genes, some of the non-protein coding genes from the mitochondrial genome were also DE in several tissues (*16 s rRNA*, *12 s rRNA*, *tRNA-Pro* and *tRNA-Ser)*.

Within groups of tissues with either over-expression of MP genes (heart, skeletal muscles, liver and kidney cortex) or under-expression (leukocytes, thymus, placenta and lymph node), we examined all overlapping genes and their functional enrichment. In heart and skeletal muscles, there were 1088 over-expressed genes in common including 320 *Nu*MP and seven *Mt*MP genes. Altogether across these 1088 genes, there was significant enrichment for OXPHOS, metabolic pathways and neurodegenerative disease pathways as in these individual tissues. Similarly, liver and kidney cortex had 1249 over-expressed genes in common including 223 *Nu*MP genes (0 *Mt*MP genes) and these were significantly enriched for metabolic pathways and peroxisome, valine, leucine and isoleucine degradation. In contrast, the DE genes in common for tissues in the under-expression group was low (63 genes) with only 20 *Nu*MP genes (0 *Mt*MP genes). Across all 63 genes, there was enrichment for adrenergic signalling in cardiomyocytes, dilated cardiomyopathy, cardiac muscle contraction and hypertrophic cardiomyopathy. Altogether, these results indicated a significant role of the over-expressed MP genes contributing to the enriched pathways in the over-expression tissue group, while this pattern was not observed in the under-expression group.

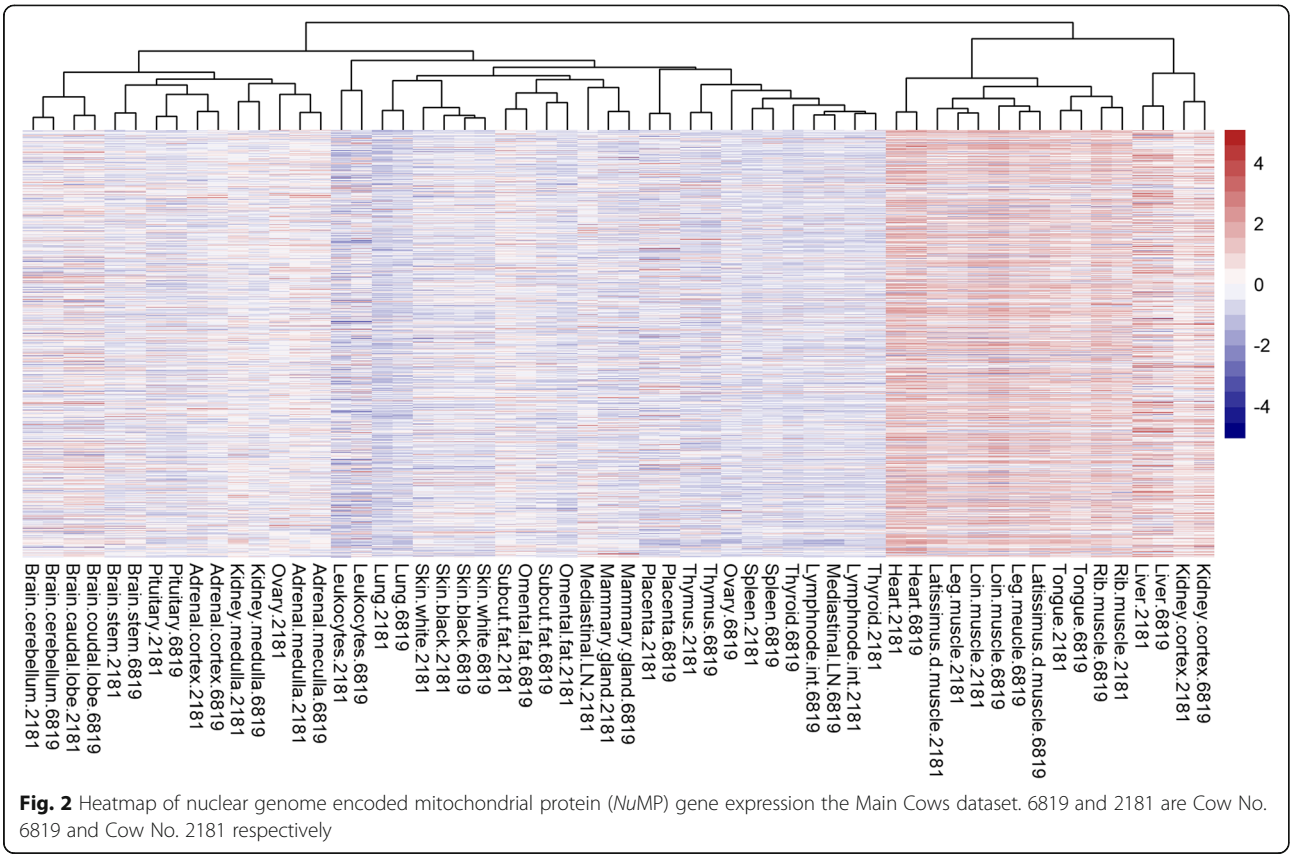


Fig. 2 Heatmap of nuclear genome encoded mitochondrial protein (*Nu*MP) gene expression the Main Cows dataset. 6819 and 2181 are Cow No. 6819 and Cow No. 2181 respectively

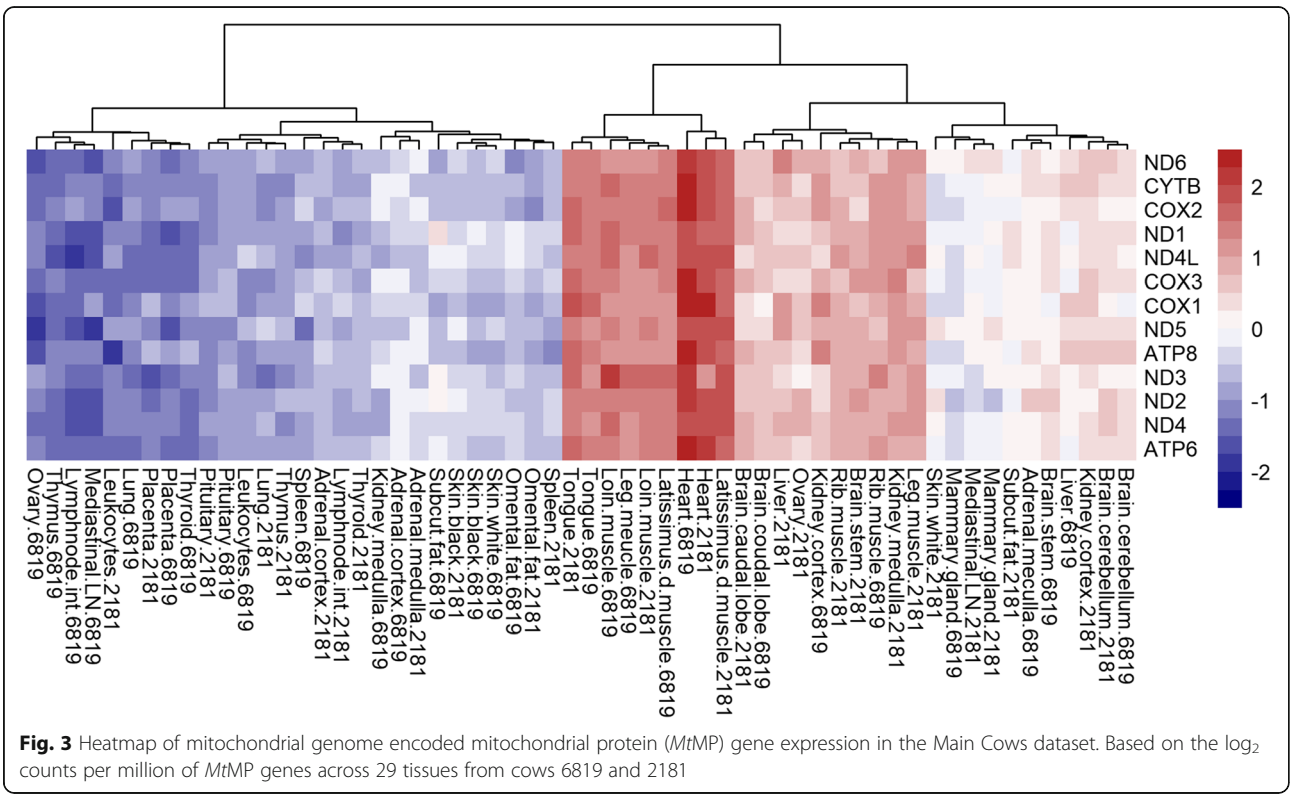


Fig. 3 Heatmap of mitochondrial genome encoded mitochondrial protein (*Mt*MP) gene expression in the Main Cows dataset. Based on the $\log_2$ counts per million of *Mt*MP genes across 29 tissues from cows 6819 and 2181

Main cows: Foetuses

The analysis for functional enrichment of overall DE genes in six foetal tissues showed significant enrichment of OXPHOS and metabolic pathways only in heart and lungs but not in leg muscle (Additional file 6). The *Nu*MP genes were over-expressed in the heart and under-expressed in the remaining tissues, including leg muscles (Additional file 7). Similarly, the *Mt*MP genes were prominently over-expressed in heart, under-expressed in the lungs and not significant in the remaining tissues (Additional file 8). Higher expression of *Nu*MP genes was observed in liver of the male foetus and it did not cluster with liver of the female foetus.

Co-expression network analysis of mitochondrial protein genes

The gene co-expression network constructed based on the affinity matrix from genes correlated in expression > |0.95| in adult cows had altogether 3643 genes clustered into four major network clusters I-IV (Fig. 4). The *Nu*MP genes were concentrated in two main clusters (I and IV) indicating co-expression among *Nu*MP genes and the remaining *Nu*MP genes were sparsely scattered across all other clusters. Similarly, *Mt*MP genes were all grouped in cluster I. Clusters I and IV containing sub-groups of highly co-expressed *Nu*MP and *Mt*MP and *Nu*MP genes are referred to as *Nu*MP-*Mt*MP and *Nu*MP clusters respectively. Within the *Nu*MP-*Mt*MP cluster, the MP genes from the respective genomes were highly co-expressed. The *Nu*MP-*Mt*MP cluster was significantly enriched for OXPHOS, metabolic pathways and mitochondrial diseases' pathways. Similarly, the *Nu*MP cluster (cluster IV) was over-represented for signalling pathways, contraction, metabolic pathways and myopathies related to the heart (Table 3). The gene functions of the non-mitochondrial protein genes (Non-MP) in the *Nu*MP-*Mt*MP cluster were associated with heart and muscle functioning, signalling, and contraction (Additional files 9,10).

We tested if the co-expression of *Nu*MP genes in the *Nu*MP-*Mt*MP cluster was due to random chance using a Chi-square goodness of fit test. The frequency of *Nu*MP genes in the cluster was significantly higher than random ($^2 = 307.6$, $p < 0.01$), supporting that the cluster was enriched with co-expressed MP.

Further, we investigated the effect of TAD on the co-expression by comparing the number of 651 TAD mapped genes in the *Nu*MP-*Mt*MP cluster with the

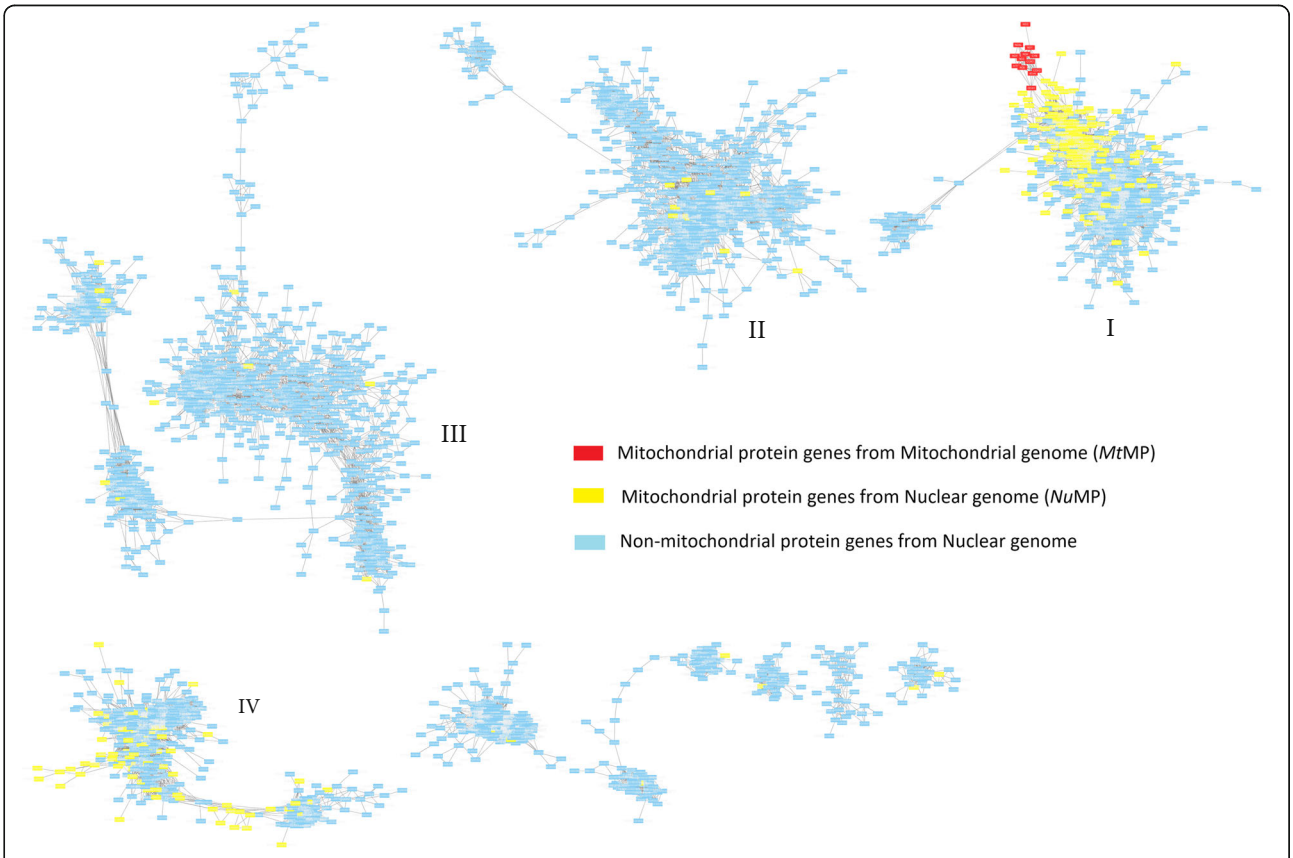


Fig. 4 Gene co-expression network clusters across tissues in the Main Cows based on similarity matrix computed using Pearson correlations > |0.95|

Table 3 Summary of gene, composition, and functional enrichment of KEGG pathways of genes in co-expression clusters (FDR <1e-05) in the Main Cows dataset

Cluster	No. of ***Mt***MP genes	No. of ***Nu***MP Genes	No. of Non-MP genes	Total No. of genes	Enrichment of pathways
I (*Nu*MP-*Mt*MP Cluster)	13	216	584	813	Parkinson's disease, Oxidative phosphorylation, Alzheimer's diseases, Huntington diseases, Non-alcohol fatty liver diseases, metabolic pathways, Citrate cycle, carbon metabolism, Cardiac muscle contraction, Proteosome
II	0	10	871	881	Retrograde endocannabinoid signaling, GABAergic synapse, Nicotine addiction, Morphine addictions, Glutamatergic synapse, Dopaminergic synapse, Synaptic vesicle cycle, Neuroactive ligand-receptor interaction
III	0	12	923	935	Cell adhesion molecules (CAMs), *Staphylococcus aureus* infection, intestinal immune network for IgA production, Leishmaniasis, Antigen processing and presentation, viral myocarditis, Allograft rejection, primary immunodeficiency, Hematopioetic cell lineage, Natural killer cell-mediated cytotoxicity
IV (*Nu*MP cluster)	0	79	466	545	Chemical carcinogenesis, Complement and coagulation cascade, Drug metabolism – cytochrome p450 metabolism, steroid hormone biosynthesis, retinol metabolism, Metabolic pathways, Complement and coagulation cascades, bile secretion, primary bile acid biosynthesis, tryptophan metabolism, carbon metabolism, fatty acid metabolism

mean from 100 randomly generated samples of 651 genes from 3022 TAD mapping genes across the clusters. It showed involvement of *Nu*MP-*Mt*MP genes in a similar number of TADs (472 ± 10 vs 484), but *Nu*MP-*Mt*MP were more likely to be present in groups of two or more within TADs. The total number of genes occurring in a two or more in a TAD was 282 and 116 (±6) in *Nu*MP-*Mt*MP cluster and random samples respectively. This indicated that co-expression in the *Nu*MP-*Mt*MP cluster was enriched within TADs.

Validation of patterns of mitochondrial protein gene expression

The key findings on the expression of MP genes from the Main Cows dataset were validated using two independent datasets (Validation Cow and Validation Sheep). Both validation sets confirmed the general trends of MP gene expression and co-expression in tissues.

Firstly, both validation datasets confirmed the over-expression of MP genes in heart and skeletal muscles, and under-expression in blood leukocytes as in the adult tissues of the Main Cows dataset (Additional file 11-18). Further, expression of MP genes within tissues, as indicated by LFC values between Main Cows and Validation Cow (Fig. 5), were highly correlated (R^2 0.67–0.96) except for thyroid (R^2 0.01). Similarly, the correlation of LFC values between Main Cows and Validation Sheep was high (R^2 0.6–0.87), except for mammary and lungs (R^2 0.36, 0.34) (Additional file 19). We investigated the poor correlation of gene expression in thyroid between the Main Cows and the Validation Cow. At least 35 DE *Nu*MP genes in common between the datasets were expressed in opposite directions. These genes were mainly enriched for metabolic pathways, pyruvate metabolism and synthesis of antibiotics. Interestingly, the expression of *Nu*MP genes between Validation Cow and Validation Sheep were moderately correlated including thyroid (R^2 0.59) except for lung and mammary tissues (Additional file 20).

Secondly, the *'either all over-expression or all under-expression'* of DE *Mt*MP genes within tissues was supported by findings from both validation datasets. Further, the expression of *Mt*MP genes in the direction of the dominant DE *Nu*MP genes also remained evident across datasets.

Thirdly, the co-expression of *Mt*MP and *Nu*MP genes in a cluster were reproduced in the Validation Cow (Additional file 21), and to some extent in Validation Sheep (Additional file 22). The co-expression of MP genes in the *Nu*MP-*Mt*MP cluster in the Validation Cow was more than expected by random chance (^{2}= 207.847, $p < 0.01$) showing that the enrichment of the cluster for co-expression of MP genes.

Finally, the overlap of genes in *Nu*MP-*Mt*MP clusters across the Main Cow and validation datasets was higher than would be expected if genes were randomly allocated to clusters. In particular, the occurrence of *Mt*MP genes were almost coincidental (13/13) between cow datasets and 12/13 genes in common between cow and sheep datasets. Similarly, a considerable proportion of *Nu*MP genes and also non-mitochondrial protein genes, were in common across datasets (Fig. 6).

Discussion

This study described and validated specific patterns of differential expression for over 1000 mitochondrial protein genes, encoded by the nuclear and mitochondrial genomes, in bovine across different tissues. The study also presented strong evidence of co-expression between *Nu*MP and *Mt*MP genes.

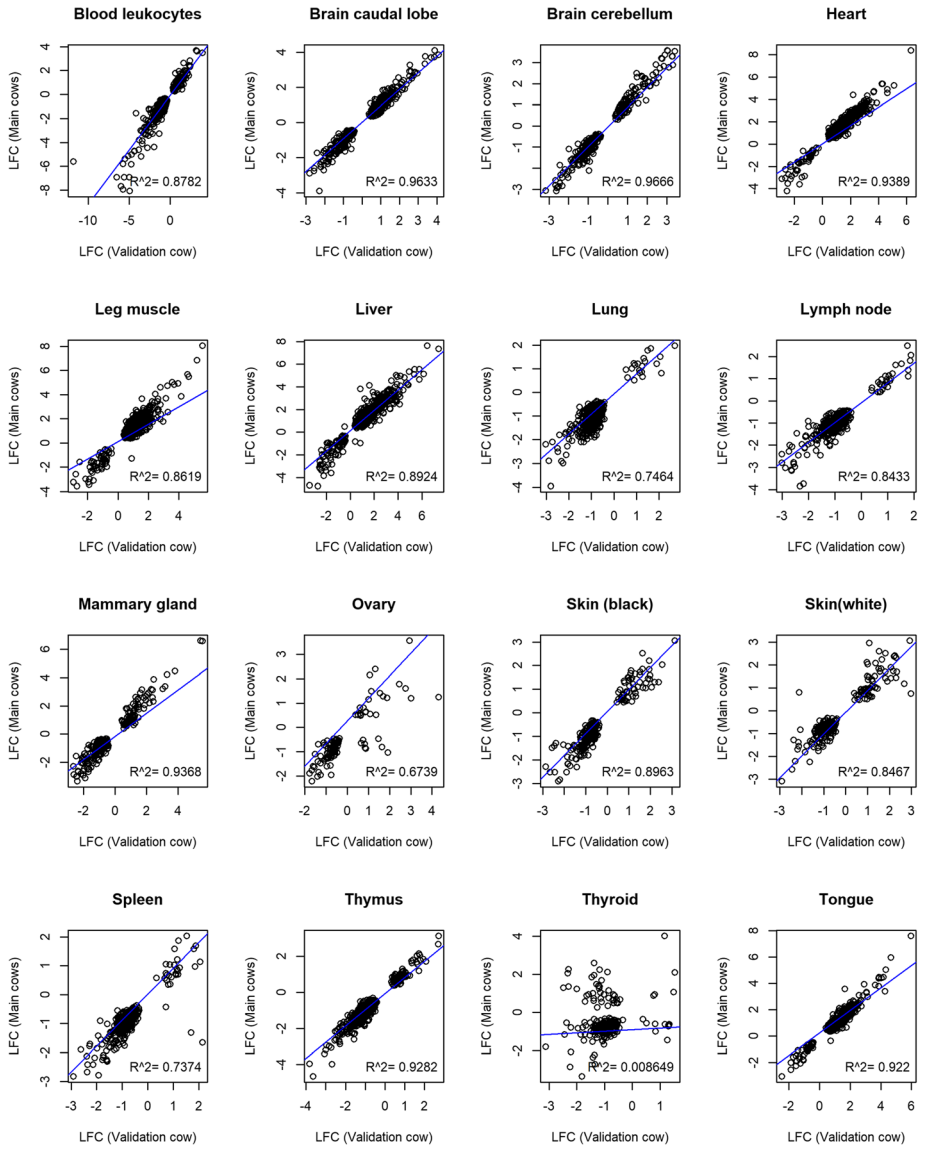


Fig. 5 Scatterplot of log2 fold change values of differentially expressed mitochondrial protein genes from nuclear genome in the Main Cows against the Validation Cow

Differential expression of mitochondrial protein genes

Overexpression of MP gene in more metabolically active tissues

The observed patterns of differential expression of MP genes within tissue, where the proportion of DE MP genes exceeded 40%, appears to correlate with the known metabolic demand of tissues. MP genes were over-expressed in tissues with high reported metabolic demand (heart, skeletal muscles, tongue, and kidney cortex: Table 4), and under-expressed in tissues with low reported energy demand (adipose tissue and blood leukocytes: Table 4). In humans, about 60–70% of the total resting energy expenditure is accounted for by kidney, brain, liver, and heart, which altogether constitute less

Table 4 Specific metabolic rates of organs and tissues across species (kcal/kg/day)

Species	Heart	Kidney	Brain	Liver	Skeletal muscle[a]	Adipose	Reference
Cattle	429	412	185	130	–	–	[36]
Sheep	588	496	255	200	–	–	[36]
Human	440	440	240	200	10–15	4.5	[4, 38]

[a]Rate for resting muscle

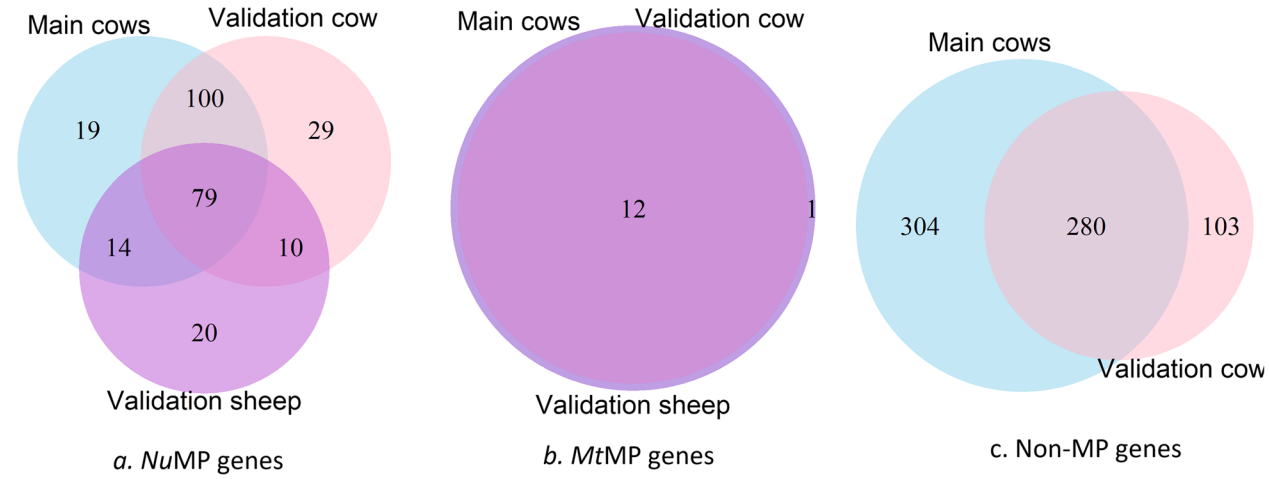


Fig. 6 Venn diagram showing the number of genes in common among the *Nu*MP-*Mt*MP co-expression clusters. a. Mitochondrial protein genes encoded by the nuclear genome (*Nu*MP genes), b. Mitochondrial protein genes encoded by the mitochondrial genome (*Mt*MP genes) and c. non-mitochondrial protein genes from nuclear genomes in common between the Main Cows and Validation Cow

than 6% of adult body weight, whereas skeletal muscle (40–50% of body weight) accounts for 20–30% of resting energy expenditure [4, 39], which altogether account up to 80% of energy expenditure.

The heart meets almost the entire energy demand through the mitochondrial OXPHOS pathway (95%) [40]. A higher expression of selected *Mt*MP genes (*ND1, ND5, ATP6, CYTB*) were reported in the heart compared to other tissues (brain, kidney, liver and skeletal muscle) in mice [41], which supports heart as the tissue with the highest *Mt*MP gene expression. Similarly, skeletal muscles, which has low resting energy demand, are capable of spiking by almost 1000-fold during short intensive exercise [42, 43]. OXPHOS is highlighted as an important pathway for generating energy during the exercise/muscular activity in both short intensive as well as prolonged exercise [44]. The observed higher expression of MP genes in the tongue seems likely because the tongue is a muscular organ. Furthermore, results from the heart and skeletal muscles group reinforced the importance of OXPHOS and metabolic pathways in energy metabolism in these tissues.

A high expression of MP genes specifically in kidney cortices may be attributed to energy generation taking place at the proximal and distal convoluted tubules, which are also the site for active reabsorption of metabolites [45, 46]. In kidney and liver, the enrichment for metabolic pathways but not for OXPHOS, despite their high energy demand, is suggestive of dominance of non-OXPHOS pathways in energy metabolism.

Tissues with under-expression of mitochondrial protein genes

Among the tissues with under-expression of MP genes, only adipose tissue in human has a published metabolic rate. In keeping with our observed low MP gene expression in adipose, the metabolic rate of human adipose tissues was low (3.2–4.6 kcal/kg/day) [38]. On the other hand, leukocytes (and other tissues with under-expression for MP) have mainly non-energy related mitochondrial functions, such as redox signalling and controlling apoptosis [47], which, in part, could explain the incidence of under-expression of MP genes in blood leukocytes. Further, results from the analysis of group of tissues showing MP gene under-expression revealed a low number of DE genes in common across these tissues and no enrichment for energy pathways support a diminished role of mitochondrial energy function in leukocytes.

As for the adult cows, the highest expression of MP genes in the foetal heart tissue was expected considering the early foetal development and establishment of the heartbeat occurs as early as 3 weeks in the bovine foetus. In contrast to adult cows, the low expression of MP genes in foetal leg muscle was likely attributable to only partial development and non-functionality of the muscle. Skeletal muscle development, mainly secondary myogenesis, is initiated in the foetal stage from 9 weeks post-fertilization to parturition [48] and our foetal calves were around 16 weeks old. Generally, the remaining foetal tissues measured in this study are reported to be functionally inert in the foetal development stage, including lungs [49] and explains the under-expression of the MP genes.

In general, the expression profiles of MP genes in a tissue were consistent as indicated by clustering of same tissue of two or more animals within the dataset. Nonetheless, some tissues were exceptions, including foetal livers may be attributed to the sampling and cellular heterogeneity of the samples because each cell type may have a specific expression profile.

To sum up, the expression of MP genes in this study concurs with the energy demand of tissues (where known) implying that the increased energy demand may be met through increased expression of MP genes. Furthermore, previous studies report that there is a specific correlation between mRNA and protein quantity across tissues [50, 51].

Mitochondrial genome encoded mitochondrial protein (*Mt* MP) gene expression

Besides, energy demand in tissues as the basis of increased transcription rates of MP genes, high *Mt*MP gene expression could also result from increased mitochondrial genome copy numbers. Mitochondrial DNA copy number differs considerably across tissue types, but remains closely regulated within a tissue type [23]. Studies in humans indicate that mitochondrial genome copy numbers are aligned with tissue energy demands: for example heart, skeletal muscle, omental fat, and blood leukocytes had 6970, 3650, 400–600 and 91 copies per diploid nuclear genome respectively [52–54]. Studies comparing copy number and gene expression of all *Mt*MP genes across tissues are scarce. A study in striated muscles (cardiac, type 1 skeletal muscle and type 2 skeletal muscles) of rabbit [55] demonstrated that the expression of *Mt*MP gene (*CYTB*) was proportional to mitochondrial copy number. Thus, it is plausible that the varying gene expression (indicating energy requirements) of tissue types are modulated through their mitochondrial DNA copy number.

Direction of expression of differentially expressed mitochondrial protein genes

There were two interesting aspects of the direction of DE *Mt*MP genes; first, the '*all over-expression or all under-expression*' of DE *Mt*MP genes within tissues and second, the directional consistency of DE *Mt*MP genes expression in the dominant direction of DE *Nu*MP genes. The first phenomenon of occurrence of DE *Mt*MP genes in single direction has not been previously reported to the best of our knowledge. A possible explanation of this phenomenon rests in the mechanism of transcription because the entire mitochondrial genome is transcribed as a near-complete polycistronic unit [56, 57], so that almost all mitochondrial genes are transcribed as one unit. The initiation of transcription, particularly at *HSP2* promotor site on the mitochondrial genome generates a near-complete polycistronic unit [58]. The second trend showing the common direction of DE *Mt*MP and *Nu*MP genes was observed in all tissues exhibiting significant DE of *Mt*MP genes. This suggests DE *Nu*MP and *Mt*MP in these tissues are co-regulated.

Co-expression of mitochondrial protein genes

The co-expression of mitochondrial protein genes was a prominent finding in the current study. Co-expression of MP gene in *Mt*MP-*Nu*MP cluster was further tested to be non-random and non-random co-expression of genes are previously reported across species [30]. Further, the significant enrichment of *Nu*MP-*Mt*MP co-expression cluster for OXPHOS and metabolic pathway supports co-functional co-expression of genes [29, 59]. Similarly, results from MP gene expression study in humans showed a significant correlation between *Mt*MP and *Nu*MP gene expression within tissues [60], suggesting close coordination between nuclear and mitochondrial genomes in relation to energy demand. The functional enrichment of our *Nu*MP-*Mt*MP cluster for the OXPHOS pathway and non-MP genes in the cluster for heart myopathies, contraction and signalling, emphasize their role in energy metabolism and supporting systems. The *Nu*MP cluster was enriched for metabolic pathways which is another important energy metabolism component of mitochondria.

The investigation of involvement of TAD on the co-expression demonstrated that the co-expressed genes in *Nu*MP-*Mt*MP cluster occurring in two or more within a TAD compared to the random sample. This indicated the potential role of TADs in co-expression of mitochondrial protein gene in our study. As such, the intra-TAD gene co-expression was not different from random for most chromosomes in another study [61].

Validation

Overall, there were high correlation and consistency evident in the expression (differential expression and co-expression) of mitochondrial protein genes in tissue across the datasets. However, we have not considered for the physiological states, number of tissues sampled, and sequencing platforms employed in our validation study. Firstly, a notable difference in expression profile of *Nu*MP genes in the thyroid between the Main Cows and the Validation Cow, and Main Cows and Validation Sheep is potentially related to pregnancy of the Main Cows, as the Validation Cow and the Validation Sheep were not pregnant. The activity of thyroid and thyroid hormone synthesis are reportedly increased during pregnancy in human [62] and thyroid hormones are known to regulate metabolism [63]. The interaction of thyroid and MP function in metabolism is an area of interest for

further investigation but beyond the scope of current work. Secondly, we based the differential gene expression of a gene in tissue to the mean expression across all other tissues where both the number of tissues and tissue types were not completely identical across the datasets (29 tissues in Main Cows, 18 tissues in Validation Cow and 15 tissues in Validation Sheep). Thereby, expression in tissue across the datasets has been compared to the mean expression of different sample sizes, which might vary across the datasets. Thirdly, the sequencing platforms used were different for each dataset: the Main Cows dataset were sequenced on HiSeq™ 3000 (Illumina), the Validation Cow was sequenced on HiSeq^{TM}2000 sequencer (Illumina) and Validation Sheep were sequenced on Illumina HiSeq^{TM} 2500.

Conclusions

Mitochondrial protein genes were differentially expressed across tissues. Tissues with high energy demand showed over-expression and under-expression was observed in tissues with low energy requirements, which suggests a link between mitochondrial protein gene expression and the energy demand of each tissue. Furthermore, mitochondrial protein genes from both genomes (*Nu*MP and *Mt*MP) were significantly co-expressed and enriched for co-functionality. This implies that it is necessary to consider mitochondrial protein genes from both genomes in studies related to mitochondrial function. Mitochondrial protein gene expression analysis may be extrapolated to production traits such as feed efficiency, heat tolerance, adaptability to cold climate, to further elucidate their role in relation to energy metabolism.

Methods

Data

The standard best practice recommendations for RNA-seq is at least three samples of each tissue (from different individuals) [64]. This study utilized RNAseq from three cows; two Holstein cows and their foetuses, and one Holstein cow from a previous study [27]. As the cows in the two datasets were physiologically different due pregnancy status and also used different sequencing platforms, we analysed them separately and the results from the two cows dataset (Main Cows) was validated in the one cow dataset (Validation Cow). Further, gene expression patterns in cattle were validated in a sheep dataset previously published [65], which is a closely related species (Validation Sheep) [37]. The Main Cows dataset had RNAseq from 29 tissues from two adult cows and six tissues from two 16 weeks old foetuses. The Validation Cow data consisted of RNAseq reads from 18 tissues, and the Validation Sheep data were gene expression counts for a subset of tissues (15 tissue types) of three Texel x Blackface adult females. The tissue-specific gene expression patterns in the Main Cows dataset were validated using the validation datasets.

Ethics, animals and tissue sampling

The ethical approval, including the permission to euthanise the animals of the Main Cows datasets were obtained from the Department of Jobs, Precincts and Regions Ethics Committee (Application No. 2014–23). Two lactating and pregnant Holstein cows and their two foetuses at 16 weeks of gestation representing a comparable physiological status from the Agriculture Victoria Research dairy herd at Ellinbank, Victoria, Australia (38°14′ S, 145°56′ E) were chosen for the study. The cows were offered 6 kg of wheat per day with perennial ryegrass pasture grazed in the paddock, supplemented with pasture silage or hay where required. Both cows were born in 2006, 16 weeks pregnant and were sampled on day 205 and 173 of their lactation (cow 2181 and 6819, respectively). Cow 2181 had a male foetus (2181F), and cow 6819 had a female foetus (6819F). Both foetuses were from the same sire (half-sibs).

Blood samples were drawn from the coccygeal vein by venipuncture before euthanasia and processed following the blood fractionation and white blood cell stabilization protocols of the RiboPure™ blood kit (Ambion by Life Technologies). Other tissues were sampled following euthanasia of the animals. The cows were euthanised individually by a trained veterinarian and not within line of sight of another deceased animal to minimise stress. The cow was restrained in a crush and given an intravenous injection of 600 mg of xylazine IV adequate to cause moderate sedation. The cow was immediately released from the crush, and once the sedation had taken effect and the cow was sitting down, 300 mg of ketamine was given intravenously. Once the cow laid down, 1 l of 25% magnesium sulphate solution was administered intravenously until pronounced deceased by the veterinarian. Once pronounced dead, all tissue types were dissected from the animal. Connective tissue was removed, and the samples were dissected into 1 cm squares, sealed in a 5 ml tube and flash-frozen in liquid nitrogen. Subcutaneous fat was sampled from the rib region. Blood (on ice) and tissues samples (in liquid nitrogen canisters) were then moved to the main laboratory and stored at − 80 °C. The metadata and RNAseq reads for all 40 tissues are available at EMBL-EBI European Nucleotide Archive (ENA) under study accession ERP118133. For this study, we generated data for 35 samples (29 tissues from adult cows and six from the foetuses) (Table 5).

Table 5 List of 35 organ-tissue sampled from the two adult cows and two foetuses in the Main Cows dataset

Tissues/organs	Tissues/organs	Tissues/organs
Adrenal gland cortex (Adrenal cortex)	Lung	Spleen
Adrenal gland medulla (Adrenal medulla)	Mammary gland (Mammary)	Thymus
Omental fat pad (Omental fat)	Heart	Thyroid gland (Thyroid)
Subcutaneous fat	Brain cerebellum	Tongue
Kidney cortex	Brain stem	Blood leukocytes (Leukocytes)
Kidney medulla	Brain caudal lobe	
Longissimus thoracic muscle (Loin muscle)	Pituitary gland	Foetal brain
Semimembranosus muscle (Leg muscle)	Placenta	Foetal kidney
Intercostal muscle (Rib muscle)	Liver	Foetal lung
Latissimus dorsi muscle (Msub)	Skin black	Foetal heart
Mediastinal lymph node (Mediastinal LN)	Skin white	Foetal liver
Lymph node	Ovary	Foetal leg muscle

RNA extraction and sequencing

RNA from blood leukocytes was extracted using the RiboPure Blood Kit (Ambion) according to the manufacturer instructions. For tissues, 100 mg of tissue was ground in a TissueLyserII (Qiagen) with liquid nitrogen, and then ~ 30 mg of ground tissue was used to extract RNA using Trizol (Invitrogen) following standard procedures. RNA was passed through a PureLink RNA Mini column (Qiagen) for clean-up and concentration and eluted in 30 μl RNase free water. RNA Integrity Numbers (RIN), which indicates the RNA quality, were determined using Agilent Tapestation (Agilent) and RNAseq libraries were prepared from all samples (Additional file 1) with RIN > 6 at which the 3′ bias level is at a minimum using the SureSelect Strand-Specific RNA Library Prep Kit (Agilent) as instructed by the manufacturer. Libraries were barcoded uniquely, assigned randomly to one of two pools and sequenced on a HiSeq™ 3000 (Illumina) in a 150-cycle paired-end run. One hundred and fifty bases paired-end reads were called with bcltofastq and output in fastq format. The quality of the libraries and alignment are as presented in Additional file 2. Poor-quality bases were filtered, and sequence reads trimmed using an in-house script. Bases with a quality score of < 20 were trimmed from the 3′ end of reads. Reads with a mean quality score < 20, > 3 N's, or final length < 50 bases were not included. Only paired reads were retained for alignment.

Read alignment and gene counting

For each library, paired-end reads were mapped to Ensembl bovine genome UMD3.1 reference [66] and annotated using STAR version 2.5.3ab [67]. Aligned reads were checked for quality using Qualimap 2 [68], and unique mapping reads for samples (Additional file 3). The R package featureCounts [69] was used to generate a count matrix of read counts per gene for every sample.

Mitochondrial protein genes

Mitochondrial protein genes in the current study were based on the list of MP identified in humans, available in Mitocarta 2.0 [10]. The official gene names of 1158 MP genes were directly converted to bovine ensemble gene IDs using a gene ID conversion function in the software DAVID (Database for Annotation, Visualization, and Integrated Discovery) version 6.8 [70, 71]. This translated into 1054 bovine MP ensemble gene IDs (1041 *Nu*MP and 13 *Mt*MP), which were used as the final list of MP genes for further analysis in this study (Additional file 4). Additionally, 24 non-protein coding genes from the mitochondrial genome (22 *tRNA*s and 2 *rRNA*s) were also included in the analysis. The mitochondrial protein gene expression profiles in tissues are expected to be similar across mammalian species because they share a very important mitochondrial function [72, 73].

Differential gene expression analysis

The lowly expressed genes were filtered out using function filterByExpr of edgeR package for differential expression analysis in R [74]. Differential expression of genes was analysed using the glmQLFit function. A model was fitted to the data with a design matrix of an overall mean of gene expression counts across all other tissues as the intercept and tissue as a fixed effect, i.e. differential expression is relative to the average expression of the same gene across all other tissues. The glmQLTest method was used to identify DE genes, specifically up or down expressed. A list of DE genes, along with their fold changes, was generated and summarized for each tissue. A gene was considered as differentially expressed (DE) in tissue if its expression was significantly higher than the mean expression of same gene across all other tissues (i.e. ≥ |0.6| $\log_2$ fold changes (LFC) = 1.5-fold difference, FDR < 0.01). The sign + and

- of LFC values of the DE gene was used to deduce the expression as either over-expression or under-expression respectively compared to the mean of expression across all other tissues. Further, log_2CPM (counts per million) values of all *Nu*MP and *Mt*MP genes (including non-DE MP genes) across tissues were visualized as heatmap using R package pheatmap [75]. In addition, we looked at the number of DE MP genes by genome (i.e. *Nu*MP and *Mt*MP), their direction of expression and the proportion of DE genes to the total genes in category. The foetal tissues were analysed separately following the procedures implemented for the adult cows.

Co-expression network analysis across tissues

The functionally associated genes tend to be co-expressed, and this is used to infer novel function as well as to identify candidate genes in diseases and their prediction [28]. To study the co-expression pattern in tissues, we used a similarity network based on a Pearson correlation coefficient of gene expression (>|0.95|) of adult cows in the Main Cows dataset, executed using a plugin ExpressionCorrelation [76] in Cytoscape 3.6.1 [77], to construct gene co-expression clusters. We analysed the co-expression cluster involving MP genes for;

i. biological significance of the cluster using functional enrichment analysis and composition of the genes,
ii. whether the co-expression of *Nu*MP genes in *Nu*MP-*Mt*MP cluster was greater than random expectations using Chi-square goodness of fit (χ^2);

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Where,O_i = observed frequency of genes $_{(i)}$ in the *Nu*MP-*Mt*MP cluster ($_i$ = *Nu*MP, Non-*NuMP* gene).

E_i = expected frequency of genes $_{(i)}$ from the overall clusters ($_i$ = *Nu*MP, Non-*NuMP* gene)

iii. the effect of TAD in co-expression of *Nu*MP genes in *Nu*MP-*Mt*MP cluster considering TAD as one of the several factors potentially influencing the co-expression of small group of genes. Briefly, we mapped the co-expressed bovine genes across the clusters (3643) to the putative bovine TADs derived from the IMR90hg18 [78] and 3022 genes mapped to 1286 TADs. Similarly, within the *Nu*MP-*Mt*MP cluster, 651 of 813 co-expressed genes mapped to 484 TADs. Of this, 282 co-expressed genes were distributed in groups of 2 or more per TAD. We compared this to averages from 100 random samples of 651 genes from TAD mapping genes across all clusters (3022). The averages for number of TADs of the random samples and genes found in group of 2 or more within a TAD were 472 (±10) and 116 (±10) respectively.

Functional enrichment analysis

The DAVID software was used to investigate the functional enrichment of differentially expressed genes within a tissue and co-expressed genes across tissues: up to 3000 genes (maximum permissible in DAVID) were selected and analysed for over-representation in KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways [79]. Up to top 10 pathways with adjusted p < 1e-5 are discussed in this study.

Validation

The patterns of MP gene expression in tissues of Main Cows were validated using two previously published datasets: a lactating Holstein cow (2 years old, 65 days in milk) with 18 tissues (additional file 17) [27] (i.e. Validation Cow); and three adult female Texel x Scottish Blackface sheep from the sheep gene expression atlas project [37], which were aged about 2 years and locally (Scotland) acquired (i.e. Validation Sheep). Depending on the number of tissues in common with cattle datasets, 15 tissues were chosen from the sheep study (Additional file 18). The Validation Cow was analysed separately due to its difference in physiological status compared to the Main Cows dataset. The RNAseq reads of the Validation Cow were processed, aligned, gene counts generated and analysed following the protocols for Main Cows. Similarly for sheep, the raw gene counts [65] were normalized and subjected to standard processing and analyses for differential expression and co-expression. In sheep, 823 MP genes were identified as overlapping the Mitocarta 2.0 Human database, using the same approach as in cattle (Additional file 5). The pattern of MP gene expression across tissues was visualized with a heatmap and co-expression networks as described for Main Cows. One of the purposes of validation was to look at the consistency of gene expression patterns across datasets. To evaluate the consistency of differential expression of MP gene expression in a tissue across the datasets, a scatterplot of the LFC values of DE *Nu*MP genes (in common between the datasets) and their coefficient of determination (R^2) was used to indicate correlation between datasets. For consistency in co-expression of MP genes, the *Nu*MP-*Mt*MP co-expression cluster was further examined for the composition and commonality of genes among the datasets.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12864-020-07018-7.

Additional file 1: Table S1. Average RIN by tissue types.

Additional file 2: Table S2. Quality of library preparation.

Additional file 3: Table S3. Read alignment quality check.

Additional file 4: Table S4. List of Mitochondrial protein genes derived from Mitocarta in cattle.

Additional file 5: Table S5. List of Mitochondrial protein genes derived from Mitocarta in Sheep.

Additional file 6: Table S6. Number of differentially expressed (DE) genes by gene categories averaged for two foetuses in the Main Cows.

Additional file 7: Figure S1. Heatmap of expression of nuclear genome encoded mitochondrial protein (*Nu*MP) in tissues of foetuses 6819F and 2181F in the Main Cows.

Additional file 8: Figure S2. Heatmap of mitochondrial genome encoded mitochondrial protein (*Mt*MP) genes in tissues of foetuses 6819F and 2181F in the Main Cows.

Additional file 9: Table S7. List of non-mitochondrial protein (Non-MP) genes clustering with the mitochondrial protein genes in cluster I (*Nu*MP-*Mt*MP cluster) in the Main Cows.

Additional file 10: Table S8. KEGG pathway enrichment of the non-mitochondrial protein (Non-MP) genes in *Nu*MP-*Mt*MP cluster in the Main Cows

Additional file 11: Figure S3. The proportion of differentially expressed gene in each gene category in 18 tissues in a Validation Cow (All=All genes encoded by nuclear and mitochondrial genome, Nu= Mitochondrial protein genes encoded by nuclear genome (*Nu*MP), Mt= Mitochondrial protein genes encoded by mitochondrial genome (*Mt*MP).

Additional file 12: Figure S4. Heatmap of expression of nuclear genome encoded mitochondrial (*Nu*MP) gene in the Validation Cow.

Additional file 13: Figure S5. Heatmap of expression of mitochondrial genome encoded mitochondrial protein (*Mt*MP) genes in the Validation Cow.

Additional file 14: Figure S6. The proportion of differentially expressed gene in each gene category and direction of gene regulation in 15 tissues in the Validation Sheep (All=All genes encoded by nuclear and mitochondrial genome, Nu=Mitochondrial protein genes encoded by nuclear genome (*Nu*MP), Mt=Mitochondrial protein genes encoded by mitochondrial genome (*Mt*MP).

Additional file 15: Figure S7. Heatmap of nuclear genome encoded mitochondrial protein genes (*Nu*MP) in the Validation Sheep (three adults Texel x Blackface female sheep AF1, AF2, and AF3).

Additional file 16: Figure S8. Heatmap of mitochondrial genome encoded mitochondrial protein genes (*Mt*MP) genes in Validation Sheep (three adults Texel x Blackface females AF1, AF2, and AF3).

Additional file 17: Table S9. Number of differentially expressed gene (DEG) s and their direction in tissues by gene categories in the Validation Cow.

Additional file 18: Table S10. Number of differentially expressed gene (DEG) s and their direction in tissues by gene categories in the Validation Sheep.

Additional file 19: Figure S9. Scatter plot of log fold changes of the Main Cows against the log-fold changes of the Validation Sheep for mitochondrial protein gene expression in tissues.

Additional file 20: Figure S10. Scatter plot of log fold changes of the Validation Cow against the log-fold changes of the Validation Sheep for mitochondrial protein gene expression in thyroid.

Additional file 21: Figure S11. Gene co-expression network constructed based similarity matrix computed using Person Correlation Coefficient of gene expression at r > |0.95| across tissues of the Validation Cow.

Additional file 22: Figure S12. Gene co-expression network constructed based similarity matrix computed using Person correlation coefficient of gene expression at r > |0.95| across tissues of the Validation Sheep (three Texel x blackface adult female sheep AF1, AF2 and AF3).

Abbreviations

CPM: Counts Per Million; DAVID: Database for Annotation, Visualization, and Integrated Discovery; DE: Differentially expressed;
EDTA: Ethylenediaminetetraacetic acid; FDR: False discovery rate; KEGG: Kyoto Encyclopedia of Genes and Genomes; LFC: Log2 fold change;
MP: Mitochondrial proteins; MPG: Mitochondrial protein genes;
Mt: Mitochondrial; *Mt*MP: Mitochondrial genome encoded mitochondrial proteins; NAFLD: Non-alcohol fatty liver disease; *Nu*MP: Nuclear genome encoded mitochondrial proteins; OXPHOS: Oxidative phosphorylation;
PBS: Phosphate Buffer Saline; RIN: RNA integrity number; RNA: Ribonucleic acid; rRNA: Ribosomal ribonucleic acid; TAD: Topologically associating domain; TCA: Tricarboxylic acid; tRNA: Transfer ribonucleic acid

Acknowledgements

The authors are grateful to the Agriculture Victoria staff at the Ellinbank and Bundoora sites that were involved in sample collection. The authors thank the DairyBio project (a joint venture between Agriculture Victoria, Dairy Australia, and the Gardiner Foundation) for funding. JD receives a La Trobe University fee remission scholarship.

Authors' contributions

The conception of the experiment: JD, IMM, BGC, AJC, CJV, HDD. Sampling, RNA sequencing and processing: AJC, CJV, JBG, LM, BM, CMR, JD. Data analysis JD, IMM, RX, CJV, AJC, ELC, HDD. Manuscript preparation JD, IMM, CJV, AJC, HDD. Manuscript revision JD, IMM, BGC, LM, CMR, JBG, HDD. Management IMM, BGC, HDD. All authors have read and approved the final manuscript.

Funding

The study was funded by DairyBio project (a joint venture between Agriculture Victoria, Dairy Australia, and the Gardiner Foundation). JD is a recipient of Full Fee Remission Scholarship from La Trobe University (LUFFRS). However, funders did not have direct role in the study.

Availability of data and materials

RNAseq datasets for two cows and their foetuses are available at EMBL EBI European Nucleotide Archive (ENA) under study accession ERP118133 at https://www.ebi.ac.uk/ena/data/search?query=ERP118133, and for Validation Cow in fastq format under study accession number ERP107617 at https://www.ebi.ac.uk/ena/data/search?query=ERP107617. The raw gene counts from three adult Texel X Blackface female 1, 2 and 3 are from the sheep gene expression atlas dataset available at https://doi.org/10.7488/ds/2616. The bovine reference genome UMD 3.1 is available at https://www.ncbi.nlm.nih.gov/assembly/GCA_000003055.5, and the list of human mitochondrial proteins (MitoCarta 2.0) is available at www.broadinstitute.org/pubs/MitoCarta. The data generated and analysed in our study supporting the conclusions of the article are included in the Additional files.

Ethics approval and consent to participate

Use of animals and ethical approval for this experiment was granted by the Department of Jobs, Precincts and Regions Animal Ethics Committee (Application No. 2014–23).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3083, Australia. [2]Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, VIC 3083, Australia. [3]Agriculture Victoria, Ellinbank Dairy Centre, Ellinbank, VIC 3822, Australia. [4]Faculty of Veterinary & Agricultural Science, University of Melbourne, Parkville, VIC 3052, Australia. [5]The Roslin Institute and Royal (Dick)

School of Veterinary Studies, University of Edinburgh, Edinburgh, Scotland, UK.

Received: 28 January 2020 Accepted: 20 August 2020
Published online: 19 October 2020

References

1. Taylor RW, Turnbull DM. Mitochondrial DNA mutations in human disease. Nat Rev Genet. 2005;6(5):389–402.
2. Gorman GS, Chinnery PF, DiMauro S, Hirano M, Koga Y, McFarland R, et al. Mitochondrial diseases. Nat Rev Dis Prim. 2016;2:16080.
3. Wallace DC. Mitochondrial diseases in man and mouse. Science. 1999; 283(5407):1482–8.
4. Elia M, Livesey G. Energy expenditure and fuel selection in biological systems: the theory and practice of calculations based on indirect calorimetry and tracer methods. World Rev Nutr Diet. 1992;70:68–131.
5. Wang C, Youle RJ. The role of mitochondria in apoptosis*. Annu Rev Genet. 2009;43:95–118.
6. Sun N, Youle RJ, Finkel T. The mitochondrial basis of aging. Mol Cell. 2016; 61(5):654–66.
7. Contreras L, Drago I, Zampese E, Pozzan T. Mitochondria: the calcium connection. Biochim Biophys Acta. 2010;1797(6):607–18.
8. Fox TD. Mitochondrial protein synthesis, import, and assembly. Genetics. 2012;192(4):1203–34.
9. Lopez MF, Kristal BS, Chernokalskaya E, Lazarev A, Shestopalov AI, Bogdanova A, et al. High-throughput profiling of the mitochondrial proteome using affinity fractionation and automation. Electrophoresis. 2000; 21(16):3427–40.
10. Calvo SE, Clauser KR, Mootha VK. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. Nucleic Acids Res. 2016;44(Database issue):D1251–D7.
11. Dudek J, Rehling P, van der Laan M. Mitochondrial protein import: common principles and physiological networks. Biochim Biophys Acta. 2013;1833(2): 274–85.
12. Chacinska A, Koehler CM, Milenkovic D, Lithgow T, Pfanner N. Importing mitochondrial proteins: machineries and mechanisms. Cell. 2009;138(4):628–44.
13. Margulis L. Origin of eukaryotic cells : evidence and research implications for a theory of the origin and evolution of microbial, plant, and animal cells on the Precambrian earth. New Haven: Yale University Press; 1970.
14. Giles RE, Blanc H, Cann HM, Wallace DC. Maternal inheritance of human mitochondrial DNA. Proc Natl Acad Sci. 1980;77(11):6715–5719.
15. Hutchison CA, Newbold JE, Potter SS, Edgell MH. Maternal inheritance of mammalian mitochondrial DNA. Nature. 1974;251(5475):536–8.
16. Anderson S, de Bruijn MHL, Coulson AR, Eperon IC, Sanger F, Young IG. Complete sequence of bovine mitochondrial DNA conserved features of the mammalian mitochondrial genome. J Mol Biol. 1982;156(4):683–717.
17. Sutarno CJM, Greeff J, Lymbery AJ. Mitochondrial DNA polymorphisms and fertility in beef cattle. Theriogenology. 2002;57(6):1603–10.
18. Holyoake AJ, McHugh P, Wu M, O'Carroll S, Benny P, Sin IL, et al. High incidence of single nucleotide substitutions in the mitochondrial genome is associated with poor semen parameters in men. Int J Androl. 2001;24(3): 175–82.
19. Chan CC, Liu VW, Lau EY, Yeung WS, Ng EH, Ho PC. Mitochondrial DNA content and 4977 bp deletion in unfertilized oocytes. Mol Hum Reprod. 2005;11(12):843–6.
20. Schutz MM, Freeman AE, Lindberg GL, Koehler CM, Beitz DC. The effect of mitochondrial DNA on milk production and health of dairy cattle. Livest Prod Sci. 1994;37(3):283–95.
21. Wang J, Shi Y, Elzo MA, Dang S, Jia X, Lai S. Genetic diversity of ATP8 and ATP6 genes is associated with high-altitude adaptation in yak. Mitochondrial DNA Part A. 2018;29(3):385–93.
22. Shi Y, Hu Y, Wang J, Elzo MA, Yang X, Lai S. Genetic diversities of MT-ND1 and MT-ND2 genes are associated with high-altitude adaptation in yak. Mitochondrial DNA Part A. 2018;29(3):485–94.
23. Robin ED, Wong R. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. J Cell Physiol. 1988;136(3):507–13.
24. Goffart S, Wiesner RJ. Regulation and co-ordination of nuclear gene expression during mitochondrial biogenesis. Exp Physiol. 2003;88(1):33–40.
25. Clay Montier LL, Deng JJ, Bai Y. Number matters: control of mammalian mitochondrial DNA copy number. J Genet Genomics. 2009;36(3):125–31.
26. San Segundo-Val I, Sanz-Lozano CS. Introduction to the gene expression analysis. In: Isidoro García M, editor. Molecular genetics of asthma. New York, NY: Springer New York; 2016. p. 29–43.
27. Chamberlain AJ, Vander Jagt CJ, Hayes BJ, Khansefid M, Marett LC, Millen CA, et al. Extensive variation between tissues in allele specific expression in an outbred mammal. BMC Genomics. 2015;16:993.
28. van Dam S, Vosa U, van der Graaf A, Franke L, de Magalhaes JP. Gene co-expression analysis for functional classification and gene-disease predictions. Brief Bioinform. 2018;19(4):575–92.
29. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998;95(25): 14863–8.
30. Michalak P. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. Genomics. 2008;91(3):243–8.
31. Hurst LD, Pal C, Lercher MJ. The evolutionary dynamics of eukaryotic gene order. Nat Rev Genet. 2004;5(4):299–310.
32. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation Centre. Nature. 2012;485(7398):381–5.
33. Harhay GP, Smith TPL, Alexander LJ, Haudenschild CD, Keele JW, Matukumalli LK, et al. An atlas of bovine gene expression reveals novel distinctive tissue characteristics and evidence for improving genome annotation. Genome Biol. 2010;11(10):R102 R.
34. Wang ZM, O'Connor TP, Heshka S, Heymsfield SB. The reconstruction of Kleiber's law at the organ-tissue level. J Nutr. 2001;131:2967–70.
35. Berg JM, Tymoczko JL, Stryer L. Biochemistry. In: Tymoczko JL, Stryer L, editors. National Center for biotechnology I. 5th ed. New York: Basingstoke: W. H. Freeman; 2002.
36. Wang Z, Zhang J, Ying Z, Heymsfield SB. Organ-tissue level model of resting energy expenditure across mammals: new insights into Kleiber's law. ISRN Zoology. 2012;2012:9.
37. Clark EL, Bush SJ, McCulloch MEB, Farquhar IL, Young R, Lefevre L, et al. A high resolution atlas of gene expression in the domestic sheep (Ovis aries). PLoS Genet. 2017;13(9):e1006997.
38. Wang Z, Ying Z, Bosy-Westphal A, Zhang J, Schautz B, Later W, et al. Specific metabolic rates of major organs and tissues across adulthood: evaluation by mechanistic model of resting energy expenditure. Am J Clin Nutr. 2010; 92(6):1369–77.
39. Gallagher D, Belmonte D, Deurenberg P, Wang Z, Krasnow N, Pi-Sunyer FX, et al. Organ-tissue mass measurement allows modeling of REE and metabolically active tissue mass. Am J Phys. 1998;275(2):E249–58.
40. Stanley WC, Chandler MP. Energy metabolism in the Normal and failing heart: potential for therapeutic interventions. Heart Fail Rev. 2002;7(2):115–30.
41. Herbers E, Kekäläinen NJ, Hangas A, Pohjoismäki JL, Goffart S. Tissue specific differences in mitochondrial DNA maintenance and expression. Mitochondrion. 2019;44:85–92.
42. Glaister M. Multiple sprint work: physiological responses, mechanisms of fatigue and the influence of aerobic fitness. Sports Med (Auckland, NZ). 2005;35(9):757–77.
43. Spriet LL. Anaerobic metabolism in human skeletal muscle during short-term, intense activity. Can J Physiol Pharmacol. 1992;70(1):157–65.
44. Hargreaves M. Skeletal muscle metabolism during exercise in humans. Clin Exp Pharmacol Physiol. 2000;27(3):225–8.
45. Forbes JM. Mitochondria–power players in kidney function? Trends Endocrinol Metab. 2016;27(7):441–2.
46. Bhargava P, Schnellmann RG. Mitochondrial energetics in the kidney. Nat Rev Nephrol. 2017;13:629.
47. Kramer PA, Ravi S, Chacko B, Johnson MS, Darley-Usmar VM. A review of the mitochondrial and glycolytic metabolism in human platelets and leukocytes: implications for their use as bioenergetic biomarkers. Redox Biol. 2014;2:206–10.
48. Yan X, Zhu M-J, Dodson MV, Du M. Developmental programming of fetal skeletal muscle and adipose tissue development. J Genomics. 2013;1:29–38.
49. Strang LB. The lungs at birth. Arch Dis Child. 1965;40:575.
50. Edfors F, Danielsson F, Hallström BM, Käll L, Lundberg E, Pontén F, et al. Gene-specific correlation of RNA and protein levels in human cells and tissues. Mol Syst Biol. 2016;12(10):883.
51. Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Global quantification of mammalian gene expression control. Nature. 2011; 473:337.

52. Miller FJ, Rosenfeldt FL, Zhang C, Linnane AW, Nagley P. Precise determination of mitochondrial DNA copy number in human skeletal and cardiac muscle by a PCR-based assay: lack of change of copy number with age. Nucleic Acids Res. 2003;31(11):e61–e.
53. Lindinger A, Peterli R, Peters T, Kern B, von Flue M, Calame M, et al. Mitochondrial DNA content in human omental adipose tissue. Obes Surg. 2010;20(1):84–92.
54. Svendsen AJ, Tan Q, Jakobsen MA, Thyagarajan B, Nygaard M, Christiansen L, et al. White blood cell mitochondrial DNA copy number is decreased in rheumatoid arthritis and linked with risk factors. A twin study. J Autoimmun. 2019;96:142–6.
55. Williams RS. Mitochondrial gene expression in mammalian striated muscle. Evidence that variation in gene dosage is the major regulatory event. J Biol Chem. 1986;261(26):12390–4.
56. Aloni Y, Attardi G. Expression of the mitochondrial genome in HeLa cells. II. Evidence for complete transcription of mitochondrial DNA. J Mol Biol. 1971; 55(2):251–67.
57. Gelfand R, Attardi G. Synthesis and turnover of mitochondrial ribonucleic acid in HeLa cells: the mature ribosomal and messenger ribonucleic acid species are metabolically unstable. Mol Cell Biol. 1981;1(6):497–511.
58. Shokolenko IN, Alexeyev MF. Mitochondrial transcription in mammalian cells. Front Biosci (Landmark edition). 2017;22:835–53.
59. Lee JM, Sonnhammer EL. Genomic gene clustering analysis of pathways in eukaryotes. Genome Res. 2003;13(5):875–82.
60. Mercer Tim R, Neph S, Dinger Marcel E, Crawford J, Smith Martin A, Shearwood A-Marie J, et al. The human mitochondrial Transcriptome. Cell. 2011;146(4):645–58.
61. Soler-Oliva ME, Guerrero-Martínez JA, Bachetti V, Reyes JC. Analysis of the relationship between coexpression domains and chromatin 3D organization. PLoS Comput Biol. 2017;13(9):e1005708.
62. Fantz CR, Dagogo-Jack S, Ladenson JH, Gronowski AM. Thyroid function during pregnancy. Clin Chem. 1999;45(12):2250.
63. Mullur R, Liu Y-Y, Brent GA. Thyroid hormone regulation of metabolism. Physiol Rev. 2014;94(2):355–82.
64. Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mole Cell Proteomic. 2014;13(2):397–406.
65. Bush SJ, Hume DA, Clark EL. Unnormalised raw count estimates for the sheep gene expression atlas,[dataset]: University of Edinburgh; 2019. https://doi.org/10.7488/ds/2616.
66. Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, et al. A whole-genome assembly of the domestic cow, Bos taurus. Genome Biol. 2009; 10(4):R42.
67. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.
68. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 2016;32(2):292–4.
69. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923–30.
70. da Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4(1):44–57.
71. da Huang W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37(1):1–13.
72. Brand MD, Orr AL, Perevoshchikova IV, Quinlan CL. The role of mitochondrial function and cellular bioenergetics in ageing and disease. Br J Dermatol. 2013;169(Suppl 2(0 2)):1–8.
73. Chinnery PF, Schon EA. Mitochondria. J Neurol Neurosurg Psychiatry. 2003; 74(9):1188–99.
74. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.
75. Kolde R. Pheatmap: pretty heatmaps. R Package Version 61 ed; 2012.
76. Hui S, Sander C, Potylitsine E, Whitaker W, Bader G, Morrison L, et al. ExpressionCorrelation. Makes a similarity network where nodes are genes, and edges denote highly correlated genes. Version 1.1.0 ed 2015.
77. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.
78. Wang M, Hancock TP, Chamberlain AJ, Vander Jagt CJ, Pryce JE, Cocks BG, et al. Putative bovine topological association domains and CTCF binding motifs can reduce the search space for causative regulatory variants of complex traits. BMC Genomics. 2018;19(1):395.
79. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.

Publisher's Note

Chapter 5:

Mitochondrial protein gene expression and the oxidative phosphorylation pathway associated with feed efficiency and energy balance in dairy cattle

J. Dairy Sci. **104**:575–587, https://doi.org/10.3168/jds.2020-18503

Jigme Dorji, Iona M. MacLeod, Amanda J. Chamberlain, Christy J. Vander Jagt, Phuong N. Ho, Majid Khansefid, Brett A. Mason, Claire P. Prowse-Wilkins, Leah C. Marett, William J. Wales, Benjamin G. Cocks, Jennie E. Pryce, and Hans D. Daetwyler

Published: January 2021

J. Dairy Sci. 104:575–587
https://doi.org/10.3168/jds.2020-18503

Mitochondrial protein gene expression and the oxidative phosphorylation pathway associated with feed efficiency and energy balance in dairy cattle

Jigme Dorji,[1,2]* Iona M. MacLeod,[2] Amanda J. Chamberlain,[2] Christy J. Vander Jagt,[2] Phuong N. Ho,[2] Majid Khansefid,[2] Brett A. Mason,[2] Claire P. Prowse-Wilkins,[2,3] Leah C. Marett,[3,4] William J. Wales,[3,4] Benjamin G. Cocks,[1,2] Jennie E. Pryce,[1,2] and Hans D. Daetwyler[1,2]
[1]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, Australia, 3083
[2]Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Victoria, Australia, 3083
[3]Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Victoria, Australia, 3010
[4]Agriculture Victoria, Ellinbank Dairy Centre, Ellinbank, Victoria, Australia, 3821

ABSTRACT

Feed efficiency and energy balance are important traits underpinning profitability and environmental sustainability in animal production. They are complex traits, and our understanding of their underlying biology is currently limited. One measure of feed efficiency is residual feed intake (RFI), which is the difference between actual and predicted intake. Variation in RFI among individuals is attributable to the metabolic efficiency of energy utilization. High RFI (H_RFI) animals require more energy per unit of weight gain or milk produced compared with low RFI (L_RFI) animals. Energy balance (EB) is a closely related trait calculated very similarly to RFI. Cellular energy metabolism in mitochondria involves mitochondrial protein (MiP) encoded by both nuclear (*Nu*MiP) and mitochondrial (*Mt*MiP) genomes. We hypothesized that MiP genes are differentially expressed (DE) between H_RFI and L_RFI animal groups and similarly between negative and positive EB groups. Our study aimed to characterize MiP gene expression in white blood cells of H_RFI and L_RFI cows using RNA sequencing to identify genes and biological pathways associated with feed efficiency in dairy cattle. We used the top and bottom 14 cows ranked for RFI and EB out of 109 animals as H_RFI and L_RFI, and positive and negative EB groups, respectively. The gene expression counts across all nuclear and mitochondrial genes for animals in each group were used for differential gene expression analyses, weighted gene correlation network analysis, functional enrichment, and identification of hub genes. Out of 244 DE genes between RFI groups, 38 were MiP genes. The DE genes were enriched for the oxidative phosphorylation (OXPHOS) and ribosome pathways. The DE MiP genes were underexpressed in L_RFI (and negative EB) compared with the H_RFI (and positive EB) groups, suggestive of reduced mitochondrial activity in the L_RFI group. None of the *Mt*MiP genes were among the DE MiP genes between the groups, which suggests a non-rate limiting role of *Mt*MiP genes in feed efficiency and warrants further investigation. The role of MiP, particularly the *Nu*MiP and OXPHOS pathways in RFI, was also supported by our gene correlation network analysis and the hub gene identification. We validated the findings in an independent data set. Overall, our study suggested that differences in feed efficiency in dairy cows may be linked to differences in cellular energy demand. This study broadens our knowledge of the biology of feed efficiency in dairy cattle.

Key words: residual feed intake, mitochondrial protein gene, differential gene expression, dairy cattle

INTRODUCTION

Feed efficiency is well recognized for its high relevance to farm economics, resource sustainability, and climate change (Herrero et al., 2013; Herrero and Thornton, 2013). Feed constitutes a significant recurring expense on-farm (Chamberlain, 2012; Yilmaz et al., 2016), and selection for feed efficiency is expected to result in animals that have reduced feed or energy requirement for maintenance (Herd et al., 2003). Further, selection for feed efficiency is estimated to reduce methane emission by 15% in a decade (de Haas et al., 2014). Feed efficiency is characterized as a complex trait with challenges and costs in measurement of phenotypes (Arthur and Herd, 2008). The definition of feed efficiency differs in growing and lactating animals and is complicated in the latter, considering the rapid mobilization of body

Received March 10, 2020.
Accepted August 20, 2020.
*Corresponding author: jigme.dorji@agriculture.vic.gov.au

reserves following calving and accumulation until next calving (Roche et al., 2009). It can be measured as residual feed intake (**RFI**), which is the difference between actual and predicted DMI (Koch et al., 1963; Berry and Crowley, 2013). Energy balance (**EB**) is another indicator trait for feed efficiency closely related to RFI; EB is the difference between energy intake and energy expenditure for lactation, growth, reproduction, and maintenance. Energy balance also shows the status of body reserve mobilization and is considered a good indicator for health, reproduction, and feed management. The EB and health status of dairy cows are also indicated by their blood metabolic profiles after calving.

Selection for feed efficiency is feasible in dairy cattle and is already included in the breeding objectives of several countries, such as "feed saved" in Australia (Pryce et al., 2014b, 2015). Moreover, selection for feed efficiency demands a better understanding of the biological mechanisms underlying variation to allow for more accurate animal evaluations for the trait in the future.

Physiologically, variation in RFI is associated with metabolism, feed intake, digestion, activity, and body heat regulation (Herd et al., 2004; Herd and Arthur, 2009). The variation in RFI, according to Korver et al. (1991), reflects differences between animals in utilization of metabolizable energy and high RFI animals are likely to have higher ATP production and utilization in their tissues (Del Bianco Benedeti et al., 2018). Further, efficiency of energy utilization varies between individuals and is closely associated with genetic variation (Nkrumah et al., 2006) as well as cattle types (Pfuhl et al., 2007). Energy balance is a trait positively and closely related to RFI (Hurley et al., 2017). This is a complication specific to lactating animals and arises from the interplay of nutrient mobilization and replenishment from body reserves during lactation (Pryce et al., 2014b). Dairy cattle may differ physiologically from beef cattle in utilization, partitioning, and conversion of nutrients. Therefore, the underlying biology of feed efficiency in lactating dairy cattle requires separate consideration.

The differences in cellular energy metabolism between high (**H_RFI**) and low (**L_RFI**) RFI may provide useful insights into feed efficiency. Cellular energy metabolism occurs in mitochondria, and energy is primarily generated through oxidative phosphorylation (**OXPHOS**) in complexes of proteins called the electron transport chain (**ETC**). The proteins of ETC (~130), along with other proteins localized in mitochondria, are referred to as mitochondrial proteins (**MiP**; Fox, 2012) and are vital for mitochondrial function. It has also been proposed that defective proteins in the ETC may lead to suboptimal mitochondrial function and reduce the overall energy efficiency of the animal (Bottje, 2019). As such, the variation of metabolic demand across different tissues was reported to be reflected through differential expression of MiP genes (Dorji et al., 2020), where a gene is considered as differentially expressed (**DE**) if the expression in a tissue differed significantly from mean expression across all other tissues. Differential gene expression has been investigated between high and low milk yielders in dairy cattle (Yang et al., 2016) and between chickens and livestock divergent for feed efficiency (Lassiter et al., 2006; Kong et al., 2016b). Given the importance of energy metabolism in feed efficiency, it is plausible that MiP genes are DE between H_RFI and L_RFI groups, and that a similar pattern would be observed in positive and negative EB groups. Previous studies on the role of MiP gene expression in feed efficiency are restricted to comparatively few selected genes (Lassiter et al., 2006; Kelly et al., 2011) and are limited in their scope of inferring meaningful biological pathways.

More recently, high-throughput RNA sequencing (**RNAseq**) of entire transcriptomes allows for the identification of biological pathways and key genes behind complex traits and diseases (Salleh et al., 2018; van Dam et al., 2018; Wang et al., 2019) using gene correlation network analysis. In recent years, RNAseq has increasingly been used to study feed efficiency in cattle based on transcriptomes of liver (Alexandre et al., 2015; Salleh et al., 2017), muscle (Zhou et al., 2015; Horodyska et al., 2018), rumen epithelium (Kong et al., 2016b; Del Bianco Benedeti et al., 2018; Elolimy et al., 2018), and blood (Khansefid et al., 2017). Blood is not a common sample for feed efficiency studies compared with other tissues, presumably due to its lower importance as a component of meat, unlike muscle and liver in beef cattle, as well as the lower mitochondrial activity in blood. On the other hand, lymphocytes, constituting about 30% of white blood cells (**WBC**), are metabolically oxidative (Kramer et al., 2014) and are easily accessible compared with other tissues. Thus, differences in metabolic activity and energy efficiency between efficient and non-efficient groups are likely to be reflected in the WBC transcriptome.

The main objectives of this study were to profile MiP gene expression in WBC of high and low feed efficiency dairy cattle, to identify key biological pathways and genes underlying feed efficiency in dairy cattle. We hypothesized that MiP genes in the WBC are DE between H_RFI and L_RFI groups as well as positive EB and negative EB groups, and that their biological pathways are related to energy metabolism.

MATERIALS AND METHODS

Animals, Residual Feed Intake Calculation, and Ranking

The Agricultural Research and Extension Animal Ethics Committee (Department of Jobs, Precincts and Regions, Attwood, Victoria, Australia) approved the protocols for this experiment (Application No. 2013-14). From 2013 to 2015, 360 animals, in batches of 40 and 3 batches each year, were run through auto-feeders for 5 wk at the Agriculture Victoria Research farm at Ellinbank, Victoria, Australia. The first week was an adaptation phase, and the remaining 4 wk were for the experiment. Each cow had the DM weight of feed offered and refused (kg), DMI (kg/d), milk yield (kg/d), fat and protein yields (kg/d), BCS, and BW (kg) recorded and used in prediction of RFI using the same model in Pryce et al. (2014a), as follows:

$$DMI = \mu + DIM + Batch + Parity + ECM + BW + BCS + RFI,$$

where μ = the overall mean DMI/intercept; DMI = average DMI during 28-d experimental period; DIM = DIM at the start of each experiment; $Batch$ was the fixed effect of the experiment (n = 9); $Parity$ is the parity group; ECM = mean ECM yield per day (kg) during the 28-d experimental period; BW was average daily BW, measured using walkover scales (automatic weigh system, model AWS100; DeLaval, Tumba, Sweden). On average, 20 BW measurements were recorded per cow over the 28-d experimental period. Body condition score was assessed twice weekly by 4 assessors, using the 8-point scale described by Earle (1976). A mean BCS of the 4 assessors was recorded per week and averaged over the experimental period. Body condition score (BCS) was included as a covariate in the model to correct for an approximation of body fat content, and RFI is the residual term from the equation.

Out of 352 animals with RFI phenotypes, we considered 109 that had RNAseq data in this study. Selected animals were ranked based on RFI values, and the top and bottom 14 animals were grouped as H_RFI and L_RFI, respectively, and used as 2 divergent feed efficiency groups for further analysis. The differences in RFI estimates between H_RFI and L_RFI were tested using independent sample t-tests for significance. A principal component analysis was performed to visualize and compare global gene expression of H_RFI and L_RFI group using the ggplot2 (Wickham, 2016) and ggfortify (Tang et al., 2016) packages in R version 3.5.2 (R Core Team, 2018). Similarly, we also predicted EB following de Vries et al. (1999), thus:

$$DMI = \mu + DIM + Batch + Parity + ECM + BW + EB.$$

Abbreviations are as described in the equation for RFI estimation, except for exclusion of BCS and inclusion of EB, which is the residual term in this equation.

The group means of the variables used in the prediction of RFI and EB are provided in Supplemental Table S1 (https://doi.org/10.3168/jds.2020-18503). The gene expression counts of the 14 top- and bottom-ranking animals for RFI and EB groups were analyzed for differential gene expression.

Further, a subset of animals in our study group had mid-infrared spectral data collected from milk samples, so this was used to predict the profile of 3 metabolites that are good indicators of energy status, BHB, nonesterified fatty acid (**NEFA**), and BUN, following Luke et al. (2019). There were 9 animals with mid-infrared data in the L_RFI group and 5 in the H_RFI group, and similarly 8 in the negative EB group and 3 in the positive EB group.

RNAseq Read Alignment, Processing, and Expression Counts

The blood sampling and processing, RNA extraction, sequencing, and quality controls of the RNAseq reads used in this study have been described by Xiang et al. (2018). In this study, we used the trimmed RNAseq reads that passed quality control (in fastq format) of H_RFI and L_RFI animals. The trimmed high-quality pair-end reads of each library were aligned to Ensembl bovine genome UMD3.1 using STAR version 2.5.3ab (Dobin et al., 2013) and checked for alignment quality using Qualimap2 (Okonechnikov et al., 2016). Only the uniquely mapped reads were used for downstream analyses. A gene count matrix for every sample was generated using the R package featureCounts (Liao et al., 2014).

Differential Gene Expression

The R package edgeR (Robinson et al., 2010) was employed for differential gene expression analysis. Raw gene counts were filtered using edgeR function filterbyExp to remove genes that were not adequately expressed, resulting in 13,469 genes passing this filter. The gene expression counts were normalized and converted to counts per million to correct for variation due to sequencing depth. The edgeR function exactTest was used to identify DE genes between the 2 RFI groups at a threshold of $\log_2$ fold change (**LFC**) $> |0.6|$ and false discovery rate of 5%. The sign of LFC was used

to deduce the direction of the gene expression as either overexpressed (positive) or underexpressed (negative). A total of 1,054 MiP genes for cattle that were derived from the human Mitocarta 2.0 (Pagliarini et al., 2008; Calvo et al., 2016) were considered for the analysis. The DE genes list was further analyzed for functional enrichment and direction of regulation, and gene composition for MiP genes.

Co-Expression Network Analysis

To build gene correlation networks and to identify a group of highly correlated gene modules associated with RFI, we used the weighted gene correlation network analysis R package, **WGCNA** (Langfelder and Horvath, 2008). The gene expression counts in samples were clustered using hclust (average method) in WGCNA to identify obvious outliers. Outlier samples were removed, and a heatmap of RFI values along with a dendrogram showing a hierarchy of cluster-based gene expression was generated. The construction of an unsigned weighted gene correlation network and module detection were performed using the automatic, one-step function network construction and module detection function blockwiseModule. Module eigengene (**ME**) represented the average expression level of genes in a module. The module membership is the degree of correlation between genes and module. Gene significance (**GS**) for each gene was calculated as the correlation between gene expression counts and RFI. The correlation between ME and RFI was calculated to identify modules associated with the trait ($r > |0.4|$, P-value < 0.05).

From the modules associated with the RFI, the genes meeting the criteria of module membership $> |0.8|$, GS $> |0.2|$ and P GS < 0.01 (Wang et al., 2017; Salleh et al., 2018; Liu et al., 2019) were identified as putative hub genes. The putative hub genes were searched for on the interacting genes (STRING) database (Szklarczyk et al., 2019), available online at https://string-db.org/, to construct a protein-protein interaction (**PPI**) network and screen hub nodes. A putative hub gene with high connectivity in a PPI network plays a critical role in the pathways associated with the module.

Functional Enrichment Analysis

The DE genes and genes in correlation network modules significantly correlated with RFI ($r > |0.4|$; $P < 0.05$) were subjected to functional enrichment of Kyoto Encyclopedia of Genes and Genomes (**KEGG**) pathways (Kanehisa and Goto, 2000) inbuilt in the online software Database for Annotation, Visualization, and Integrated Discovery (**DAVID**; Huang et al., 2009a,b), available at https://david.ncifcrf.gov/summary.jsp. The cut-off for considering the pathway as enriched was set at adjusted $P < 1.0 \times 10^{-5}$. The Benjamini-Hochberg correction method in DAVID was used for the calculation of adjusted P-values.

Validation

The RFI estimates and blood gene counts from Khansefid et al. (2017) were used to validate the findings. Briefly, the data of Khansefid et al. (2017) were independent of our data set and consisted of RFI phenotypes and gene expression counts from WBC of 19 first-lactation Holstein cows (38 ± 10 DIM). The animals were ranked based on RFI phenotype to identify the top and bottom 8 animals as a H_RFI and L_RFI group, respectively. Their study broadly estimated the correlation of gene counts on the RFI values and identified highly expressed genes that were correlated with RFI using regression models. In the present study, we employed standard differential gene expression and WGCNA approaches for validation of the findings from our primary data set.

RESULTS

The mean (SD) RFI values for the H_RFI and L_RFI groups were +1.91 (0.37) and −1.67 (0.28) kg/d, respectively, and differed significantly at $P < 0.01$. In other words, the L_RFI (i.e., feed-efficient group) consumed on average 3.55 kg/d less feed than the H_RFI group (low feed-efficient group) at the same level of production and maintenance. The principal component analysis plot of overall gene expression showed that principal component 1 explained about 81% of the variation in gene expression between the RFI groups (Figure 1A). Further, animals within an RFI group were more likely to be clustered together (Figure 1B) at the first hierarchical levels, but no clear separation occurred based on overall gene expression. Energy balance was phenotypically highly correlated with RFI ($r = +0.99$, 109 animals). Except for 2 animals, all the 28 animals in positive and negative EB group were common to the RFI group. The means (SD) of the top 14 positive EB and top 14 negative EB groups were +1.90 (0.38) and −1.64, (0.30) kg/d, respectively, and similar to RFI. The difference between the means of the positive EB and negative EB groups were statistically significant ($t = -27.25$, $P < 0.00001$).

The metabolic profiling of markers for EB (estimated from mid-infrared spectral data) found that the H_RFI group had lower mean (±SD) BHB (0.43 ± 0.06) and NEFA (0.20 ± 0.04) and higher BUN (7.24 ± 0.33) compared with L_RFI (BHB: 0.49 ± 0.08; NEFA: 0.22

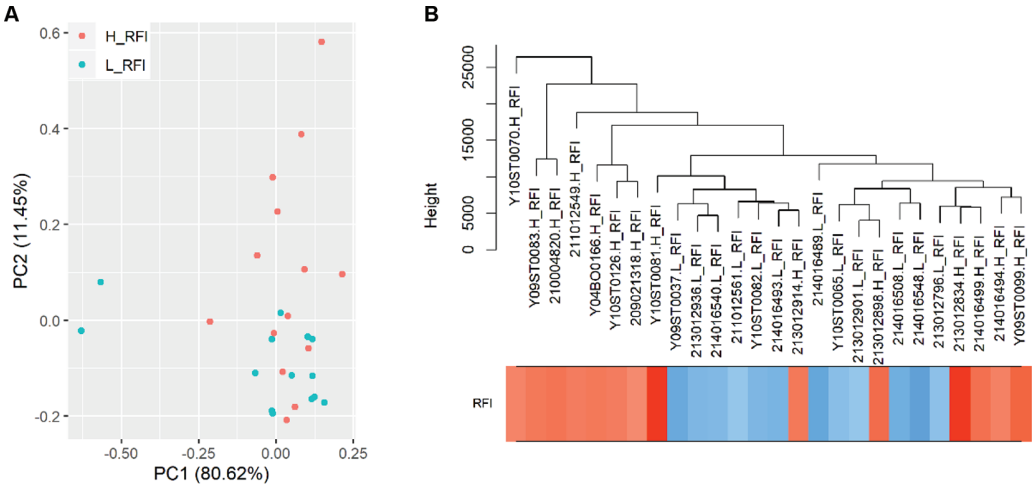


Figure 1. (A) Principal component (PC) analysis plot of global gene expression in white blood cells of 14 high residual feed intake (H_RFI) and 14 low residual feed intake (L_RFI) animals. (B) Hierarchical clustering of gene expression in the 14 H_RFI and 14 L_RFI cows with trait heatmap indicating the intensity of residual feed intake values in a scale of blue (low) to red (high).

± 0.04; BUN: 6.98 ± 0.52 mmol/L). The positive EB group had lower mean (±SD) BHB (0.45 ± 0.08) and NEFA (0.22 ± 0.04) and higher BUN (7.39 ± 0.20) compared with the negative-EB group (BHB: 0.49 ± 0.08; NEFA: 0.21 ± 0.04; BUN: 7.09 ± 0.41 mmol/L). Although these differences were not significantly different (low number of animals available with data), they were consistent in direction, demonstrating that animals with high feed efficiency (L_RFI and negative EB) also had higher levels of these metabolites (Supplemental Table S2, https://doi.org/10.3168/jds.2020-18503). Altogether, our results support that RFI and EB are highly correlated.

Differentially Expressed Genes and Functional Enrichment

Overall, 244 genes were DE between H_RFI and L_RFI groups at LFC $\geq |0.6|$) and false discovery rate 5% (Supplemental Table S3, https://doi.org/10.3168/jds.2020-18503). Functionally, DE genes were significantly enriched (number of genes, adjusted *P*) for OXPHOS (15, 2.1×10^{-7}) and ribosome (17, 1.9×10^{-9}) pathways. However, 47 DE genes were not identified in DAVID and remained unused for functional enrichment analysis. The number of overexpressed and underexpressed genes among DE genes were 64 and 180, respectively. There were 38 DE MiP genes, and all were underexpressed in the L_RFI group, with LFC ranging from −0.6 to −1.6 (Supplemental Table S4, https://doi.org/10.3168/jds.2020-18503). The DE MiP genes were all composed of *Nu*MiP and, surprisingly, contained no *Mt*MiP genes. Differential gene expression analysis was repeated with only *Mt* genes, as their mean level of expression was generally much higher than that of *Nu*MiP, which again confirmed that no *Mt* genes were DE. Further, the expression of *Mt* genes was less variable (CV 180%) compared with genes from the nuclear genome (CV 300%).

Similarly, 466 genes were DE between positive EB and negative EB groups. Among these, were 53 *Nu*MiP genes, which were all overexpressed in the positive EB group compared with the negative EB animals, except for 2 genes. The DE genes were enriched for OXPHOS (adjusted *P* 1.6×10^{-10}) and ribosome (adjusted *P* 7.5×10^{-17}) pathways. A similar expression profile between RFI and EB was expected, as these traits were highly correlated and contained almost the same animals in both RFI and EB groups. Thus, the subsequent gene network analysis was conducted only for gene expression in the RFI groups.

WGCNA

Overall, 9 correlation network modules out of 28 were significantly related to RFI ($r > |0.4|$, $P < 0.05$; Figure 2A). Of these modules, only ME2 and ME3 were significantly enriched for KEGG pathways (Figure 2B, C). Altogether, WGCNA and the functional enrichment analysis supported a strong association between OXPHOS pathway and RFI. Further, WGCNA identified additional pathways associated with RFI traits undetected with differential gene expression analysis.

Hub genes in a module have high network connectivity and are also highly associated with the corresponding traits. The 173 genes in module ME2 meeting the

threshold criteria (module membership > |0.8|, GS > |0.2|, and P GS < 0.01) were identified as putative hub genes (Supplemental Table S5, https://doi.org/10.3168/jds.2020-18503). Networks of hub genes were derived based on PPI network at minimum interaction score of 0.90. The resulting networks of hub genes were associated with KEGG pathways: ribosome, OXPHOS, spliceosome, and proteosome (Figure 3). The hub genes that were MiP genes and associated with OXPHOS pathway were *ATP5A1*, *ATP5J*, *ATP5J2*, *NDUFAB1*, *NDUFA8*, and *NDUFA10*, most of which are noncatalytic proteins in the OXPHOS complexes. For ME3, 74 putative hub genes (Supplemental Table S6, https://doi.org/10.3168/jds.2020-18503), refined based on PPI network, had 6 hub genes for OXPHOS, 5 for ribosomes, and 2 for citrate cycle (Figure 4). The refined hub genes for OXPHOS pathway and citrate cycle were MiP genes (*NDUFA5*, *NDUFA12*, *NDUFB5*, *ATP5F1*, *ATP5C1*, *SDHD*, and *PDHB*), and the dehydrogenase components were noncatalytic.

Validation

In the independent set of validation cows, the mean (SD) RFI values of H_RFI and L_RFI groups were +1.20 (0.95) and −0.97 (0.71) kg/d, respectively; consequently, as expected, the means of the RFI groups were significantly different [$t(14) = 5.20$, $P < 0.001$]. Global gene expression was not distinct but showed greater variability in the H_RFI group compared with the L_RFI group (Supplemental Figure S1, https://doi.org/10.3168/jds.2020-18503). We found 1,695 DE genes between L_RFI and H_RFI groups at LFC ≥ |0.6| and false discovery rate < 0.05, including 151 MiP genes (Supplemental Table S7, https://doi.org/10.3168/jds.2020-18503). The DE genes were significantly enriched for OXPHOS (adjusted P 2.6×10^{-7}) and ribosome (adjusted P 5.9×10^{-19}) KEGG pathways. All 151 DE MiP genes were *Nu*MiP (18 overexpressed and 133 underexpressed in the L_RFI group compared with the H_RFI group), and none were *Mt*MiP. The

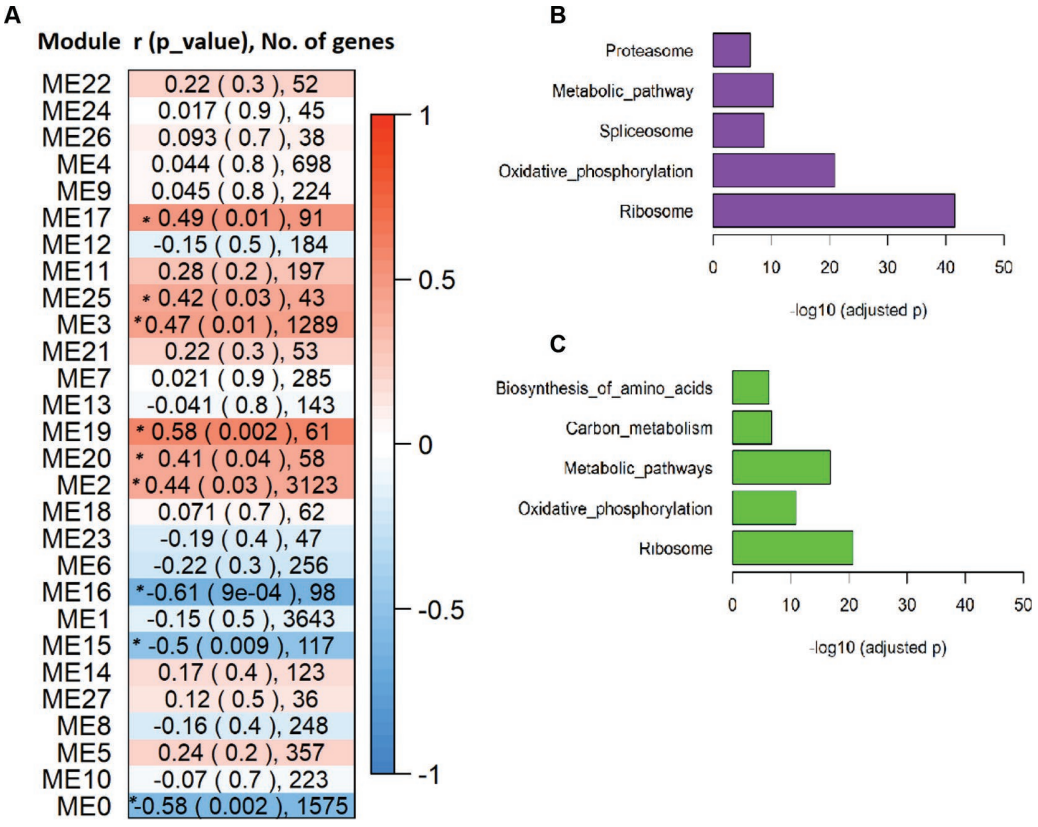


Figure 2. Gene expression modules correlated with residual feed intake (RFI) based on weighted gene correlation network analysis in 14 high-RFI and 14 low-RFI cows. (A) Module and number of genes in each module, and the relationship of module with RFI (r), with P-values in parentheses. ME = module eigengene. *Indicates module with a significant relationship (r > |0.4| and $P < 0.05$) with RFI. The direction of the relationship (correlation) is indicated by colors, where red is positive and blue is negative, and the intensity of color represents the strength of correlation. Functional enrichment of Kyoto Encyclopedia of Genes and Genomes pathways of overall genes module ME2 (B) and ME3 (C) for the main data set.

133 underexpressed MiP genes were over-represented (adjusted P) in the following KEGG pathways: OXPHOS (3.4×10^{-28}), metabolic pathways (8.6×10^{-13}), and ribosome (2.0×10^{-13}). We found 31 DE MiP genes in common between the validation and the main data sets (Supplemental Table S8, https://doi.org/10.3168/jds.2020-18503), and they showed concordant underexpression. Altogether, the differential expression of MiP genes between 2 RFI groups, the enrichment of OXPHOS pathway, underexpression of MiP genes in L_RFI group, and lack of representation of *Mt*MiP genes in the validation data set agreed with the findings of the main data set.

Five co-expression modules were significantly related to the RFI trait in WGCNA analysis: ME1, ME2, ME4, ME5, and ME8. Modules ME1 and ME5 were positively correlated with RFI, whereas ME2, ME4, and ME8 were negatively correlated with RFI (Supplemental Figure S2a, https://doi.org/10.3168/jds.2020-18503). Among the modules, only ME1 was signifi-

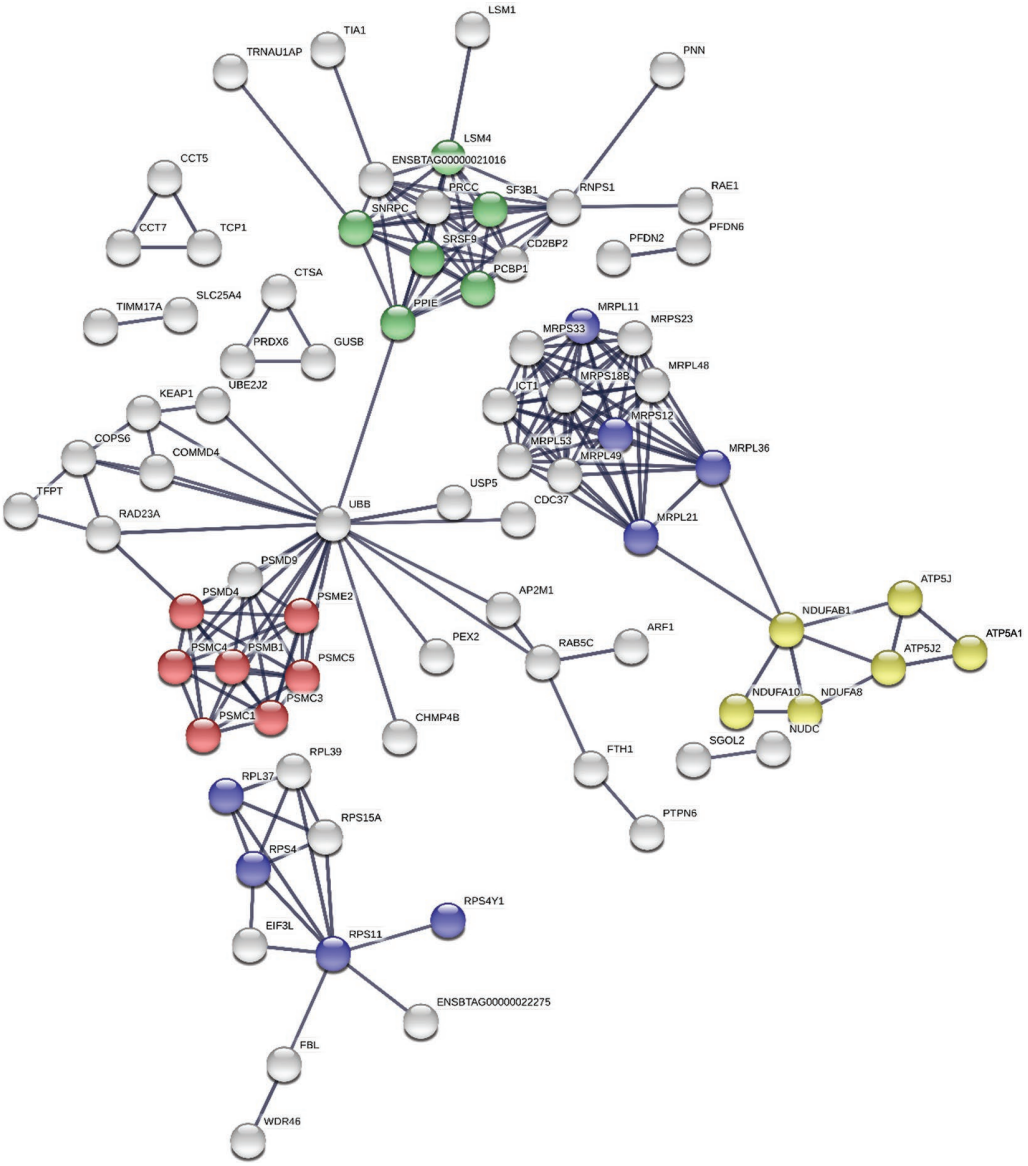


Figure 3. A protein–protein interaction network featuring the network of putative hub genes from genes in module eigengene (ME) 2 (173) with confidence of more than 0.90 from the STRING database (Szklarczyk et al., 2019; https://string-db.org/). Colors indicate different Kyoto Encyclopedia of Genes and Genomes pathways: proteosome (red), spliceosome (dark blue), ribosome (green), and oxidative phosphorylation (yellow).

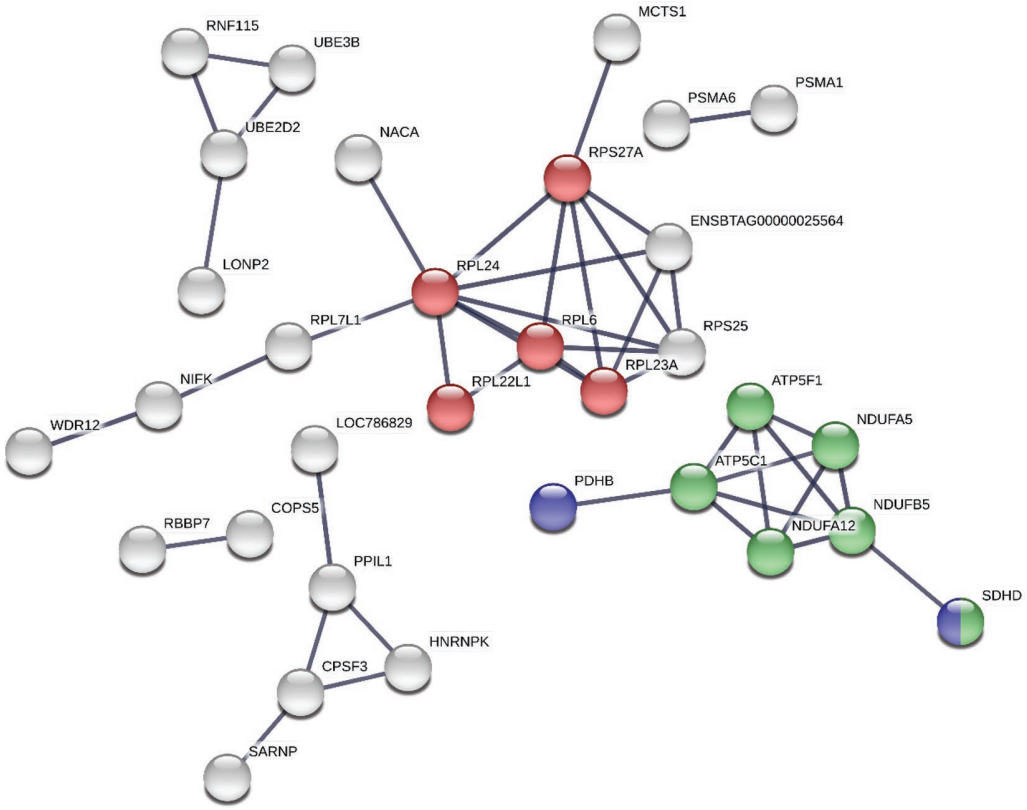


Figure 4. A protein–protein interaction network featuring putative hub genes based on genes in module eigengene (ME) 3 (74) with confidence of more than 0.90 in the STRING database (Szklarczyk et al., 2019; https://string-db.org/). Colors indicate different Kyoto Encyclopedia of Genes and Genomes pathways: ribosome (red), oxidative phosphorylation (green), and citrate cycle (blue).

cantly enriched for biological pathways. It was enriched for the ribosome, OXPHOS, spliceosome, metabolic, and proteasome pathways (Supplemental Figure S2b, https://doi.org/10.3168/jds.2020-18503). The top 402 genes were ranked in order of module connectivity and GS to RFI, as well as based on the PPI network-identified hub genes related to ribosome, OXPHOS, and peroxisome (Supplemental Figure S3, https://doi.org/10.3168/jds.2020-18503). The networked hub genes in the OXPHOS pathway that were also MiP genes were *ATP5B*, *COX4I1*, *UQCRC1*, *UQCRFS1*, *SDHA*, *NDUFS2*, *NDUFA2*, *NUDFA8*, *NDUFA9*, and *NDUFB10.* However, only *NDUFA8* overlapped with hub genes in the main data set.

DISCUSSION

Residual feed intake and EB were highly correlated and were used to cross-check gene expression patterns in high and low groups. Only 2% of genes were DE from a total of about 13,469 genes that passed the RNAseq quality filters, thus representing only a small fraction of genes. This likely explains the lack of clear clustering when considering expression levels of all genes in the 2 RFI groups in principal component analysis plot and hierarchical cluster (Figure 1A). Within the DE genes, we found that MiP genes were consistently underexpressed in the low-RFI groups, and there was considerable overlap of these DE MiP genes in an independent validation set of cows. Differentially expressed genes were over-represented for OXPHOS and ribosome pathways in both the main and validation data sets. This strongly suggests a role of the OXPHOS pathway in feed efficiency, and this was also evident from the weighted gene co-expression network analysis. However, none of the genes from the mitochondrial genome (*Mt*MiP) were DE or were found in the co-expression module associated with RFI.

Energy Balance and RFI

The model for estimation of RFI excluding the changes in BW or BCS gives an estimate of EB (Veerkamp, 2002). Furthermore, it is also very difficult to accurately measure and distinguish between EB and feed efficiency, and no consensus currently exists on the most

appropriate mathematical models (e.g., de Vries et al., 1999; Veerkamp, 2002; Hurley et al., 2016). In addition, differences in activity levels between cows have been shown to cause changes in DMI (Olijhoek et al., 2020). These measures of feed efficiency are highly correlated in mid-lactation cows (0.96), when changes in BW were almost zero (Hurley et al., 2016). Changes in BCS primarily occur in early lactation, due to mobilization of body reserves, whereas the 352 animals used in our study estimate RFI were on average approximately 100 DIM and entering mid-lactation, where little change is expected in BCS or BW (Spurlock et al., 2012; Hurley et al., 2016). Therefore, not surprisingly, our study showed a phenotypic correlation close to 1 between RFI and EB. We found similar gene expression patterns and enrichment of pathways in RFI and EB, because almost the same animals were allocated across these 2 study sets. Interestingly, the predicted serum BHB and NEFA levels (metabolite indicators for negative EB) were higher in the more feed-efficient groups (L_RFI and negative EB) compared with the less feed-efficient group (H_RFI and positive EB) but much lower than critical limits (>0.6 mmol/L and 10 mg/dL; Ospina et al., 2010) in both groups. It is therefore important to acknowledge that our observations of differential gene expression may be a reflection of genetics underpinning feed efficiency or EB. It is important in the context of genetic selection programs to further explore approaches to better distinguish between RFI and EB traits in dairy cattle (e.g., use of metabolite profiles), because if animals are inadvertently selected for poorer EB, this could affect health and fertility.

OXPHOS Pathway in Feed Efficiency

No previous studies have examined MiP genes comprehensively in relation to feed efficiency, although a few genes belonging to MiP or the OXPHOS pathway have been reported in several studies on feed efficiency of meat animal species, mainly beef cattle, pigs, and chicken, involving liver, muscle, and rumen epithelial transcriptomes (Kong et al., 2016b; Khansefid et al., 2017; Del Bianco Benedeti et al., 2018), as well as GWA studies (Khansefid et al., 2017; Li et al., 2019). The proteomic analysis of feed efficiency indicated significant differences in the abundance of MiP in the muscle of pigs (Fu et al., 2017), liver of beef cattle (Baldassini et al., 2018), and breast muscle in chicken (Kong et al., 2016a). The gene expression and protein levels of mitochondrial energy metabolism have been linked to feed efficiency in pigs (Vincent et al., 2015).

For dairy cattle, our study is perhaps among the foremost in the investigation of feed efficiency using blood transcriptomes and demonstration of the association of MiP and OXPHOS pathways in feed efficiency. Although blood has low mitochondrial activity compared with muscle and liver tissues (Dorji et al., 2020), blood transcriptome in the present study showed distinct MiP gene expression differences between high and low feed efficiency groups and enrichment of the OXPHOS pathways that have not been reported in earlier studies of dairy cattle. These findings altogether suggest that blood is potentially a good tissue for studying gene expression associated with changes in feed efficiency. More importantly, the role of MiP and the OXPHOS pathways in feed efficiency across animal and tissue types is growing more evident.

Another pathway highly associated with feed efficiency in this study was the ribosome pathway. The enrichment of the ribosome pathway was previously reported in transcriptomes of rumen epithelium (Kong et al., 2016b), muscle, and blood (Khansefid et al., 2017). Co-regulation of the ribosome with the OXPHOS pathway is expected, considering that protein synthesis is an energy-demanding process, particularly in peptide bonding, where one mole of a polypeptide bond during protein synthesis requires about 4 ATP (MacRae and Lobley, 1982).

Direction of Regulation of Mitochondrial Protein Genes

The MiP genes were underexpressed in the feed-efficient animals. The underexpression of MiP genes in blood WBC and their relation to feed efficiency may in part be explained by the lowered activity of the OXPHOS pathway, which translates into decreased energy production and probably less heat loss (Nkrumah et al., 2006). Energy saved from reduced heat loss is believed to be channeled into increased milk production (Goddard and Grainger, 2004). Similarly, when using EB, most DE MiP genes were underexpressed in the negative EB group. Given that the calculation of RFI and EB phenotypes is similar, the similarity in strong underexpression of MiP genes is, to some, expected, with most cows overlapping between the 2 phenotypes. The underexpression of MiP genes involved in energy production and use may arise simply from a shortage of energy from feed or from lower consumption of feed, potentially involving a feedback loop to reduce expression.

In addition, a lowered protein synthesis or turnover, as indicated by underexpression of genes associated with the ribosome pathway in blood of the feed-efficient group, may save energy. Underexpression of both MiP genes and MiP abundance has been reported in feed-efficient pigs (Vincent et al., 2015; Fu et al., 2017). Similarly, less feed-efficient Nellore bulls reportedly

had increased expression of genes related to OXPHOS in the rumen epithelium (Del Bianco Benedeti et al., 2018). Conversely, the transcriptome profile of rumen epithelium had overexpression of OXPHOS genes in feed-efficient beef steers (Kong et al., 2016b). The authors justified their findings by suggesting that feed efficiency can be enhanced through better absorption of nutrients from the rumen wall, facilitated by increased morphogenesis and protein turnover in the rumen epithelium. Similarly, the proteome of breast muscle had overexpression of OXPHOS proteins in feed-efficient meat chickens (Kong et al., 2016a). Considering that growth and weight gain are essential aspects of feed efficiency traits in meat animals, and protein synthesis is one of the mechanisms for muscular growth, this justifies the overexpression of genes associated with OXPHOS and ribosome pathways in the muscle of meat animals.

Overall, it is currently inconclusive whether the specific direction of the regulation of MiP genes is associated with feed efficiency. However, this has tended to depend on animal species and sampling tissue. Overexpression of MiP genes, along with genes for the ribosome pathway, may increase protein synthesis in muscles and affect weight gain in feed-efficient animals, which could be important for meat species. By contrast, underexpression of MiP genes and lower activity of OXPHOS in dairy cattle suggests lower energy production and utilization for maintenance.

Key Mitochondrial Protein Genes Related to Feed Efficiency

Mitochondrial protein genes are encoded by genes from both nuclear and mitochondrial genomes, and OXPHOS and energy metabolism are the critical functions of MiP genes and mitochondria. One of the noteworthy findings of this study was that no *Mt*MiP genes were DE between feed efficiency groups, nor were they among the genes in modules highly correlated with RFI. Further, with exceptions from the studies specifically targeting *Mt*MiP, none of the high-throughput RNAseq studies to date has indicated an association of *Mt*MiPs in feed efficiency (Kong et al., 2016b; Khansefid et al., 2017). This suggests low variability in *Mt*MiP gene expression between the high and low feed efficient groups, and that MiP variation in feed efficiency is mainly attributable to *Nu*MiP gene expression.

We found 31 out of 38 DE MiP genes of the primary set in the validation data set. Some of these MiP genes and proteins have been associated with feed efficiency in previous studies (Supplemental Table S9, https://doi.org/10.3168/jds.2020-18503). Some common gene transcripts across studies included *COX4I1*, *ATP5D*, and *UQCRQ*, and differentially abundant MiP relating to feed efficiency were COX4I1, NDUFB11, PRDX2, NME3, and UQCRQ. Overall, the 2 most common MiP and MiP genes across studies were *COX4I1* and *UQCRQ*.

The *COX4I1* gene is located on bovine chromosome 18 between 11,799,175 and 11,807,342 bp (UMD3.1) and encodes a protein that is a component of complex IV of the ETC in mitochondria. Complex IV consists of 13 proteins (Yoshikawa, 1997), of which 10 encoded by nuclear genes are regulatory, and 3 from the mitochondrial genome constitute the catalytic core in mammals (Kadenbach and Huttemann, 2015). In humans, a mutation in *COX4I1* has been associated with short stature, poor weight gain, and increased chromosomal breaks (Abu-Libdeh et al., 2017). We checked 1,646 polymorphic variants in the animals in our study to investigate whether any showed segregation patterns according to our RFI groupings within the *COX4I1* region, but no significant pattern was observed. We suggest that the role of *COX4I1* and other common genes should be further investigated.

Highly connected hub genes in a module play essential roles in biological pathways and have been suggested for use as a potential indicator of feed efficiency (Salleh et al., 2018). In this study, we looked at the top 10% of genes in the modules significantly related to feed efficiency in both the main and validation data sets. We found 62 putative hub genes in common, including 7 OXPHOS and 11 ribosome pathway genes, but only 4 in refined hub genes associated with OXPHOS between the main and the validation data sets within our study. We compared our putative hub genes with hub genes from liver in a published study in Danish Holstein and Jersey cattle (Salleh et al., 2018). Our primary data set shared only 1 putative hub gene (*LPXN*) with Holsteins, and no genes were common between the validation set and their study. It was also interesting to note that only 1 hub gene (*LCK*) was found in common between Holsteins and Jersey in their study. As such, considerable variability exists in hub genes as well as DE genes identified across studies among breeds, species, tissues, and data sets for feed efficiency. Therefore, it appears that there is still much to learn regarding the genes and mutations that underpin differences in feed efficiency, and how genomic selection programs for feed efficiency may affect other traits such as health and fertility. Our results suggest that more efficient animals have lower requirements to generate energy (as measured by MiP gene expression) and thus potentially have better metabolic efficiency of energy utilization compared with less efficient animals.

CONCLUSIONS

The findings from our study suggest that mitochondrial protein genes in the blood are differentially expressed between high and low feed efficiency groups of lactating dairy cows. However, all differentially expressed mitochondrial protein genes were from the nuclear genome and none from the mitochondrial genome. The oxidative phosphorylation pathway, which is responsible for energy production, and the ribosome function pathway were associated with feed efficiency. Mitochondrial protein genes were underexpressed in the more feed efficient group, which may suggest a lower metabolic turnover.

ACKNOWLEDGMENTS

The authors thank the DairyBio program (a joint venture between Agriculture Victoria, Dairy Australia, and the Gardiner Foundation, Melbourne, Victoria, Australia) for funding. J. D. receives a La Trobe University fee remission scholarship (Bundoora, Australia). The authors are grateful to the staff of Agriculture Victoria at the Ellinbank and Bundoora sites, who were involved in sample collection and animal husbandry. The funding agencies did not have influence on the findings of the study, and the authors have no conflict of interest.

REFERENCES

Abu-Libdeh, B., L. Douiev, S. Amro, M. Shahrour, A. Ta-Shma, C. Miller, O. Elpeleg, and A. Saada. 2017. Mutation in the COX4I1 gene is associated with short stature, poor weight gain and increased chromosomal breaks, simulating Fanconi anemia. Eur. J. Hum. Genet. 25:1142–1146. https://doi.org/10.1038/ejhg.2017.112.

Alexandre, P. A., L. J. A. Kogelman, M. H. A. Santana, D. Passarelli, L. H. Pulz, P. Fantinato-Neto, S. L. Silva, P. R. Leme, R. F. Strefezzi, L. L. Coutinho, J. B. S. Ferraz, J. P. Eler, H. N. Kadarmideen, and H. Fukumasu. 2015. Liver transcriptomic networks reveal main biological processes associated with feed efficiency in beef cattle. BMC Genomics 16:1073. https://doi.org/10.1186/s12864-015-2292-8.

Arthur, J. P. F., and R. M. Herd. 2008. Residual feed intake in beef cattle. Rev. Bras. Zootec. 37:269–279. https://doi.org/10.1590/S1516-35982008001300031.

Baldassini, W. A., S. F. M. Bonilha, R. H. Branco, J. C. S. Vieira, P. M. Padilha, and D. P. D. Lanna. 2018. Proteomic investigation of liver from beef cattle (*Bos indicus*) divergently ranked on residual feed intake. Mol. Biol. Rep. 45:2765–2773. https://doi.org/10.1007/s11033-018-4341-2.

Berry, D. P., and J. J. Crowley. 2013. Cell Biology Symposium: Genetics of feed efficiency in dairy and beef cattle. J. Anim. Sci. 91:1594–1613. https://doi.org/10.2527/jas.2012-5862.

Bottje, W. G. 2019. Board Invited Review: Oxidative stress and efficiency: The tightrope act of mitochondria in health and disease. J. Anim. Sci. 97:3169–3179. https://doi.org/10.1093/jas/skz219.

Calvo, S. E., K. R. Clauser, and V. K. Mootha. 2016. MitoCarta2.0: An updated inventory of mammalian mitochondrial proteins. Nucleic Acids Res. 44(D1):D1251–D1257. https://doi.org/10.1093/nar/gkv1003.

Chamberlain, T. 2012. Understanding the economics of dairy farming Part 1: Income, costs and profit. Livestock (Lond) 17:30–33. https://doi.org/10.1111/j.2044-3870.2012.00137.x.

de Haas, Y., J. E. Pryce, D. P. Berry, and R. F. Veerkamp. 2014. Genetic and genomic solutions to improve feed efficiency and reduce environmental impact of dairy cattle. In Proc. 10th World Congress of Genetics Applied to Livestock Production, Vancouver, Canada.

de Vries, M. J., S. van der Beek, L. M. Kaal-Lansbergen, W. Ouweltjes, and J. B. Wilmink. 1999. Modeling of energy balance in early lactation and the effect of energy deficits in early lactation on first detected estrus postpartum in dairy cows. J. Dairy Sci. 82:1927–1934. https://doi.org/10.3168/jds.S0022-0302(99)75428-7.

Del Bianco Benedeti, P., E. Detmann, H. C. Mantovani, S. F. M. Bonilha, N. V. L. Serão, D. R. G. Lopes, W. Silva, C. J. Newbold, and M. S. Duarte. 2018. Nellore bulls (*Bos taurus indicus*) with high residual feed intake have increased the expression of genes involved in oxidative phosphorylation in rumen epithelium. Anim. Feed Sci. Technol. 235:77–86. https://doi.org/10.1016/j.anifeedsci.2017.11.002.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. 2013. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635.

Dorji, J., C. J. Vander Jagt, J. B. Garner, L. C. Marett, B. A. Mason, C. M. Reich, R. Xiang, E. L. Clark, B. G. Cocks, A. J. Chamberlain, I. M. MacLeod, and H. D. Daetwyler. 2020. Expression of mitochondrial protein genes encoded by nuclear and mitochondrial genomes correlate with energy metabolism in dairy cattle. BMC Genomics 21:720. https://doi.org/10.1186/s12864-020-07018-7.

Earle, D. 1976. A guide to scoring dairy cow condition. J. Agric (Victoria) 74:228–231.

Elolimy, A. A., M. K. Abdelmegeid, J. C. McCann, D. W. Shike, and J. J. Loor. 2018. Residual feed intake in beef cattle and its association with carcass traits, ruminal solid-fraction bacteria, and epithelium gene expression. J. Anim. Sci. Biotechnol. 9:67. https://doi.org/10.1186/s40104-018-0283-8.

Fox, T. D. 2012. Mitochondrial protein synthesis, import, and assembly. Genetics 192:1203–1234. https://doi.org/10.1534/genetics.112.141267.

Fu, L., Y. Xu, Y. Hou, X. Qi, L. Zhou, H. Liu, Y. Luan, L. Jing, Y. Miao, S. Zhao, H. Liu, and X. Li. 2017. Proteomic analysis indicates that mitochondrial energy metabolism in skeletal muscle tissue is negatively correlated with feed efficiency in pigs. Sci. Rep. 7:45291. https://doi.org/10.1038/srep45291.

Goddard, M. E., and C. Grainger. 2004. A review of the effects of dairy breed on feed conversion efficiency—An opportunity lost? Science Access 1:77–80.

Herd, R. M., J. A. Archer, and P. F. Arthur. 2003. Reducing the cost of beef production through genetic improvement in residual feed intake: Opportunity and challenges to application. J. Anim. Sci. 81(Suppl. 1):E9–E17.

Herd, R. M., and P. F. Arthur. 2009. Physiological basis for residual feed intake. J. Anim. Sci. 87(Suppl. 14):E64–E71. https://doi.org/10.2527/jas.2008-1345.

Herd, R. M., V. H. Oddy, and E. C. Richardson. 2004. Biological basis for variation in residual feed intake in beef cattle. 1. Review of potential mechanisms. Aust. J. Exp. Agric. 44:423–430. https://doi.org/10.1071/EA02220.

Herrero, M., P. Havlík, H. Valin, A. Notenbaert, M. C. Rufino, P. K. Thornton, M. Blümmel, F. Weiss, D. Grace, and M. Obersteiner. 2013. Biomass use, production, feed efficiencies, and greenhouse gas emissions from global livestock systems. Proc. Natl. Acad. Sci. USA 110:20888–20893. https://doi.org/10.1073/pnas.1308149110.

Herrero, M., and P. K. Thornton. 2013. Livestock and global change: Emerging issues for sustainable food systems. Proc. Natl. Acad. Sci. USA 110:20878–20881. https://doi.org/10.1073/pnas.1321844111.

Horodyska, J., K. Wimmers, H. Reyer, N. Trakooljul, A. M. Mullen, P. G. Lawlor, and R. M. Hamill. 2018. RNA-seq of muscle from pigs divergent in feed efficiency and product quality identifies differ-

ences in immune response, growth, and macronutrient and connective tissue metabolism. BMC Genomics 19:791. https://doi.org/10.1186/s12864-018-5175-y.

Huang, D. W., B. T. Sherman, and R. A. Lempicki. 2009a. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37:1–13. https://doi.org/10.1093/nar/gkn923.

Huang, D. W., B. T. Sherman, and R. A. Lempicki. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4:44–57. https://doi.org/10.1038/nprot.2008.211.

Hurley, A. M., N. López-Villalobos, S. McParland, E. Kennedy, E. Lewis, M. O'Donovan, J. L. Burke, and D. P. Berry. 2016. Interrelationships among alternative definitions of feed efficiency in grazing lactating dairy cows. J. Dairy Sci. 99:468–479. https://doi.org/10.3168/jds.2015-9928.

Hurley, A. M., N. López-Villalobos, S. McParland, E. Lewis, E. Kennedy, M. O'Donovan, J. L. Burke, and D. P. Berry. 2017. Genetics of alternative definitions of feed efficiency in grazing lactating dairy cows. J. Dairy Sci. 100:5501–5514. https://doi.org/10.3168/jds.2016-12314.

Kadenbach, B., and M. Huttemann. 2015. The subunit composition and function of mammalian cytochrome *c* oxidase. Mitochondrion 24:64–76. https://doi.org/10.1016/j.mito.2015.07.002.

Kanehisa, M., and S. Goto. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28:27–30. https://doi.org/10.1093/nar/28.1.27.

Kelly, A. K., S. M. Waters, M. McGee, R. G. Fonseca, C. Carberry, and D. A. Kenny. 2011. mRNA expression of genes regulating oxidative phosphorylation in the muscle of beef cattle divergently ranked on residual feed intake. Physiol. Genomics 43:12–23. https://doi.org/10.1152/physiolgenomics.00213.2009.

Khansefid, M., C. A. Millen, Y. Chen, J. E. Pryce, A. J. Chamberlain, C. J. Vander Jagt, C. Gondro, and M. E. Goddard. 2017. Gene expression analysis of blood, liver, and muscle in cattle divergently selected for high and low residual feed intake 1. J. Anim. Sci. 95:4764–4775. https://doi.org/10.2527/jas2016.1320.

Koch, R. M., L. A. Swiger, D. Chambers, and K. E. Gregory. 1963. Efficiency of feed use in beef cattle. J. Anim. Sci. 22:486–494. https://doi.org/10.2527/jas1963.222486x.

Kong, B. W., K. Lassiter, A. Piekarski-Welsher, S. Dridi, A. Reverter, N. J. Hudson, and W. G. Bottje. 2016a. Proteomics of breast muscle tissue associated with the phenotypic expression of feed efficiency within a pedigree male broiler line: I. Highlight on mitochondria. PLoS One 11:e0155679. https://doi.org/10.1371/journal.pone.0155679.

Kong, R. S. G., G. Liang, Y. Chen, P. Stothard, and L. L. Guan. 2016b. Transcriptome profiling of the rumen epithelium of beef cattle differing in residual feed intake. BMC Genomics 17:592. https://doi.org/10.1186/s12864-016-2935-4.

Korver, S., E. A. M. van Eekelen, H. Vos, G. J. Nieuwhof, and J. A. M. van Arendonk. 1991. Genetic parameters for feed intake and feed efficiency in growing dairy heifers. Livest. Prod. Sci. 29:49–59. https://doi.org/10.1016/0301-6226(91)90119-B.

Kramer, P. A., S. Ravi, B. Chacko, M. S. Johnson, and V. M. Darley-Usmar. 2014. A review of the mitochondrial and glycolytic metabolism in human platelets and leukocytes: Implications for their use as bioenergetic biomarkers. Redox Biol. 2:206–210. https://doi.org/10.1016/j.redox.2013.12.026.

Langfelder, P., and S. Horvath. 2008. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics 9:559. https://doi.org/10.1186/1471-2105-9-559.

Lassiter, K., C. Ojano-Dirain, M. Iqbal, N. R. Pumford, N. Tinsley, J. Lay, R. Liyanage, T. Wing, M. Cooper, and W. Bottje. 2006. Differential expression of mitochondrial and extramitochondrial proteins in lymphocytes of male broilers with low and high feed efficiency. Poult. Sci. 85:2251–2259. https://doi.org/10.1093/ps/85.12.2251.

Li, B., L. Fang, D. J. Null, J. L. Hutchison, E. E. Connor, P. M. VanRaden, M. J. VandeHaar, R. J. Tempelman, K. A. Weigel, and J. B. Cole. 2019. High-density genome-wide association study for residual feed intake in Holstein dairy cattle. J. Dairy Sci. 102:11067–11080. https://doi.org/10.3168/jds.2019-16645.

Liao, Y., G. K. Smyth, and W. Shi. 2014. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930. https://doi.org/10.1093/bioinformatics/btt656.

Liu, Y., H.-Y. Gu, J. Zhu, Y.-M. Niu, C. Zhang, and G.-L. Guo. 2019. Identification of hub genes and key pathways associated with bipolar disorder based on weighted gene co-expression network analysis. Front. Physiol. 10:1081. https://doi.org/10.3389/fphys.2019.01081.

Luke, T. D. W., S. Rochfort, W. J. Wales, V. Bonfatti, L. Marett, and J. E. Pryce. 2019. Metabolic profiling of early-lactation dairy cows using milk mid-infrared spectra. J. Dairy Sci. 102:1747–1760. https://doi.org/10.3168/jds.2018-15103.

MacRae, J. C., and G. E. Lobley. 1982. Some factors which influence thermal energy losses during the metabolism of ruminants. Livest. Prod. Sci. 9:447–456. https://doi.org/10.1016/0301-6226(82)90050-1.

Nkrumah, J. D., E. K. Okine, G. W. Mathison, K. Schmid, C. Li, J. A. Basarab, M. A. Price, Z. Wang, and S. S. Moore. 2006. Relationships of feedlot feed efficiency, performance, and feeding behavior with metabolic rate, methane production, and energy partitioning in beef cattle. J. Anim. Sci. 84:145–153. https://doi.org/10.2527/2006.841145x.

Okonechnikov, K., A. Conesa, and F. García-Alcalde. 2016. Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32:292–294.

Olijhoek, D. W., G. F. Difford, P. Lund, and P. Løvendahl. 2020. Phenotypic modeling of residual feed intake using physical activity and methane production as energy sinks. J. Dairy Sci. 103:6967–6981. https://doi.org/10.3168/jds.2019-17489.

Ospina, P. A., D. V. Nydam, T. Stokol, and T. R. Overton. 2010. Evaluation of nonesterified fatty acids and β-hydroxybutyrate in transition dairy cattle in the northeastern United States: Critical thresholds for prediction of clinical diseases. J. Dairy Sci. 93:546–554. https://doi.org/10.3168/jds.2009-2277.

Pagliarini, D. J., S. E. Calvo, B. Chang, S. A. Sheth, S. B. Vafai, S. E. Ong, G. A. Walford, C. Sugiana, A. Boneh, W. K. Chen, D. E. Hill, M. Vidal, J. G. Evans, D. R. Thorburn, S. A. Carr, and V. K. Mootha. 2008. A mitochondrial protein compendium elucidates complex I disease biology. Cell 134:112–123. https://doi.org/10.1016/j.cell.2008.06.016.

Pfuhl, R., O. Bellmann, C. Kühn, F. Teuscher, K. Ender, and J. Wegner. 2007. Beef versus dairy cattle: A comparison of feed conversion, carcass composition, and meat quality. Arch. Tierzucht 50:59–70. https://doi.org/10.5194/aab-50-59-2007.

Pryce, J. E., O. Gonzalez-Recio, G. Nieuwhof, W. J. Wales, M. P. Coffey, B. J. Hayes, and M. E. Goddard. 2015. Hot topic: Definition and implementation of a breeding value for feed efficiency in dairy cows. J. Dairy Sci. 98:7340–7350. https://doi.org/10.3168/jds.2015-9621.

Pryce, J. E., O. Gonzalez-Recio, J. B. Thornhill, L. C. Marett, W. J. Wales, M. P. Coffey, Y. de Haas, R. F. Veerkamp, and B. J. Hayes. 2014a. Short communication: Validation of genomic breeding value predictions for feed intake and feed efficiency traits. J. Dairy Sci. 97:537–542. https://doi.org/10.3168/jds.2013-7376.

Pryce, J. E., W. J. Wales, Y. de Haas, R. F. Veerkamp, and B. J. Hayes. 2014b. Genomic selection for feed efficiency in dairy cattle. Animal 8:1–10. https://doi.org/10.1017/S1751731113001687.

R Core Team. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: A bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140. https://doi.org/10.1093/bioinformatics/btp616.

Roche, J. R., N. C. Friggens, J. K. Kay, M. W. Fisher, K. J. Stafford, and D. P. Berry. 2009. Invited review: Body condition score and its association with dairy cow productivity, health, and welfare. J. Dairy Sci. 92:5769–5801. https://doi.org/10.3168/jds.2009-2431.

Salleh, M. S., G. Mazzoni, J. K. Höglund, D. W. Olijhoek, P. Lund, P. Løvendahl, and H. N. Kadarmideen. 2017. RNA-Seq transcriptomics and pathway analyses reveal potential regulatory genes and molecular mechanisms in high- and low-residual feed intake in Nordic dairy cattle. BMC Genomics 18:258. https://doi.org/10.1186/s12864-017-3622-9.

Salleh, S. M., G. Mazzoni, P. Løvendahl, and H. N. Kadarmideen. 2018. Gene co-expression networks from RNA sequencing of dairy cattle identifies genes and pathways affecting feed efficiency. BMC Bioinformatics 19:513. https://doi.org/10.1186/s12859-018-2553-z.

Spurlock, D. M., J. C. M. Dekkers, R. Fernando, D. A. Koltes, and A. Wolc. 2012. Genetic parameters for energy balance, feed efficiency, and related traits in Holstein cattle. J. Dairy Sci. 95:5393–5402. https://doi.org/10.3168/jds.2012-5407.

Szklarczyk, D., A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, L. J. Jensen, and C. Mering. 2019. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 47(D1):D607–D613. https://doi.org/10.1093/nar/gky1131.

Tang, Y., M. Horikoshi, and W. Li. 2016. ggfortify: Unified interface to visualize statistical result of popular R packages. R J. 8:474. https://doi.org/10.32614/RJ-2016-060.

van Dam, S., U. Võsa, A. van der Graaf, L. Franke, and J. P. de Magalhães. 2017. Gene co-expression analysis for functional classification and gene–disease predictions. Brief. Bioinform. 19:575–592. https://doi.org/10.1093/bib/bbw139.

Veerkamp, R. F. 2002. Feed intake and energy balance in lactating animals. In 7th World Congr. Genet. Appl. Livest. Prod. Vol. CD-ROM Communication No. 10-01, Montpellier, France.

Vincent, A., I. Louveau, F. Gondret, C. Trefeu, H. Gilbert, and L. Lefaucheur. 2015. Divergent selection for residual feed intake affects the transcriptomic and proteomic profiles of pig skeletal muscle. J. Anim. Sci. 93:2745–2758. https://doi.org/10.2527/jas.2015-8928.

Wang, W., W. Jiang, L. Hou, H. Duan, Y. Wu, C. Xu, Q. Tan, S. Li, and D. Zhang. 2017. Weighted gene co-expression network analysis of expression data of monozygotic twins identifies specific modules and hub genes related to BMI. BMC Genomics 18:872. https://doi.org/10.1186/s12864-017-4257-6.

Wang, Z., J. Zhu, F. Chen, and L. Ma. 2019. Weighted gene coexpression network analysis identifies key genes and pathways associated with idiopathic pulmonary fibrosis. Med. Sci. Monit. 25:4285–4304. https://doi.org/10.12659/MSM.916828.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York, NY.

Xiang, R., B. J. Hayes, C. J. Vander Jagt, I. M. MacLeod, M. Khansefid, P. J. Bowman, Z. Yuan, C. P. Prowse-Wilkins, C. M. Reich, B. A. Mason, J. B. Garner, L. C. Marett, Y. Chen, S. Bolormaa, H. D. Daetwyler, A. J. Chamberlain, and M. E. Goddard. 2018. Genome variants associated with RNA splicing variations in bovine are extensively shared between tissues. BMC Genomics 19:521. https://doi.org/10.1186/s12864-018-4902-8.

Yang, J., J. Jiang, X. Liu, H. Wang, G. Guo, Q. Zhang, and L. Jiang. 2016. Differential expression of genes in milk of dairy cattle during lactation. Anim. Genet. 47:174–180. https://doi.org/10.1111/age.12394.

Yilmaz, H., M. Gül, S. Akkoyun, O. Parlakay, M. E. Bilgili, Y. Vurarak, H. Hizli, and N. Kilicalp. 2016. Economic analysis of dairy cattle farms in east Mediterranean region of Turkey. Rev. Bras. Zootec. 45:409–416. https://doi.org/10.1590/S1806-92902016000700008.

Yoshikawa, S. 1997. Beef heart cytochrome c oxidase. Curr. Opin. Struct. Biol. 7:574–579. https://doi.org/10.1016/S0959-440X(97)80124-8.

Zhou, N., W. R. Lee, and B. Abasht. 2015. Messenger RNA sequencing and pathway analysis provide novel insights into the biological basis of chickens' feed efficiency. BMC Genomics 16:195. https://doi.org/10.1186/s12864-015-1364-0.

ORCIDS

Jigme Dorji https://orcid.org/0000-0003-4057-8918
Phuong N. Ho https://orcid.org/0000-0001-9481-9612
Majid Khansefid https://orcid.org/0000-0002-5091-7293
Claire P. Prowse-Wilkins https://orcid.org/0000-0002-7506-862X
Leah C. Marett https://orcid.org/0000-0001-9698-8401
Benjamin G. Cocks https://orcid.org/0000-0001-9776-1508
Jennie E. Pryce https://orcid.org/0000-0002-1397-1282
Hans D. Daetwyler https://orcid.org/0000-0002-3018-3640

Chapter 6:

General Discussion

Jigme Dorji[1,2]

[1] School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, Australia, 3083

[2] Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Victoria, Australia, 3083

The aim of the research was to contribute to the overarching goal of utilising the mitochondrial genome for livestock genetic improvement. A novel approach of '***whole mitochondrial genome and mitochondrial proteins***' was used in the research to determine the role of the mitochondrial genome. First, whole mitochondrial genome diversity of cattle, both within and across the breeds, was evaluated to understand if there was adequate variation. Second, the association between the expression of mitochondrial protein genes from both mitochondrial and nuclear genomes in tissues with the energy metabolism was investigated. Third, the association of the mitochondrial protein transcriptome with feed efficiency (i.e. a trait related to energy metabolism and use) was examined. The findings from these studies fill in critical information gaps and advance the prospect of using the mitochondrial genome in the genomic evaluation of dairy traits.

6.1 Summary of Chapters

Chapter 3: Cattle maternal diversity inferred from 1,883 taurine and indicine mitogenomes

In this chapter, the mitochondrial genomic diversity in cattle was evaluated using whole genome sequences and piloted the imputation of the missing mitochondrial genotypes. The aims were to understand if there is adequate genetic variation in the mitochondrial genome and the likelihood of the mitochondrial genome to be imputed towards the use for mitochondrial variants for genomic evaluation in breed improvement. The study utilised the mitochondrial genome sequence from 1000 Bull Genome project to predicted haplogroups and evaluate haplotype diversity. Further, the imputation of missing genotypes was tested based on random masking of 10% of total sites and used the tools developed for nuclear DNA.

The results showed that modern cattle are genetically diverse with high mitochondrial genetic variations and haplotype diversity within a given breed (Chapter 3, Table 7). For example, Holsteins had 210 haplotypes and 15 subgroups and

within the breed (Chapter 3, Figure S5). The African taurus (T1) and indicus (I) haplogroups influenced a small proportion of European breeds (~ 5%), mainly in Australia and China (Chapter 3: Table 3 & 4). Two new subgroups (I1a and I1b) within the indicus (I1) were reported, showing additional haplogroup diversity. The I1a haplogroup was restricted mainly to the Chinese indicus breeds. The imputation of masked genotypes consistently showed high accuracy (99.8%), and the imputed sequences did not affect the haplogroup prediction (99.5% correctly predicted) (Chapter 3, Table S8).

In summary, the chapter highlights adequate mitochondrial genomic variation within the modern cattle breeds and the possibility of the imputation of the mitochondrial genome paving the way towards their use in genomic evaluations.

Chapter 4: Expression of mitochondrial protein genes encoded by nuclear and mitochondrial genomes correlate with energy metabolism in dairy cattle

Mitochondrial protein (MP) gene expression from both the mitochondrial and nuclear genomes was examined across 29 tissue types in cattle. The gene expression in tissues was quantified through RNA sequencing and analysed for the differential gene expression and co-expression in tissues.

The MP genes were differentially expressed across tissues (Chapter 4, Fig.1), typically being over-expressed in tissues with high energy demand (e.g. heart, skeletal muscles, liver, kidney) and under-expressed in the adipose, skin, white blood cells, lungs, and thymus (Chapter 4, Fig.2 & Fig.3). The expression profiles were associated with the metabolic rates of the tissues (where known) indicating that their specific energy demands may be met through the tissue-specific expression of the mitochondrial protein genes. However, it is not clear if the tissue-specific gene expression is a result of differences in the rate of transcription, mitochondrial copy number or both. Generally, the mitochondrial content of a few known cell types (e.g.

hepatocytes, heart in Chapter 2) did not correspond well with their mitochondrial protein gene expression in the current study.

The MP genes from the nuclear and mitochondrial genomes were significantly co-expressed across tissues (Chapter 4, Fig.4), indicating an interaction between the two genomes. The genes in the co-expression cluster were significantly enriched for oxidative phosphorylation (OXPHOS) – the energy production pathway.

Overall, the results provide an essential link between the MP gene expression in a tissue and the energy demand, suggesting that specific energy demand in tissue types is fulfilled through the tissue-specific expression of the MP genes. The co-expression of the mitochondrial and nuclear MP genes further suggests the role of the mitochondrial genome and the interaction with the nuclear genes in the energy metabolism and indicates that the mitochondrial genome should be considered for genetic evaluation of traits related to energy metabolism.

Chapter 5: Mitochondrial protein gene expression and the oxidative phosphorylation pathway associated with feed efficiency and energy balance in dairy cattle

This study assessed the association of the mitochondrial protein (MP) gene expression in blood with feed efficiency. Feed efficiency in cattle was measured as residual feed intake (RFI) and energy balance (EB). The animals were classified into high RFI (low feed efficiency) and low RFI (high feed efficiency) groups. The gene expression in the blood was quantified using RNA sequencing. The differential gene expression and weighted gene co-expression network analyses were used to associate gene expression with RFI.

There were 244 differentially expressed (DE) genes, including 38 MP genes between the high RFI and low RFI groups. The DE genes were significantly enriched for OXPHOS and metabolic pathways, indicating the role of energy

metabolism in feed efficiency. Specifically, the MP genes were under-expressed in the low RFI group suggesting a lower metabolic energy expenditure in feed efficient animals. Similarly, the gene clusters associated with RFI from the co-expression network analysis included the MP genes and were over-represented for OXPHOS and ribosome (i.e. protein synthesis) pathways. Notably, both DE genes between the RFI groups and genes in co-expression clusters associated with RFI were of nuclear origin (Chapter 5, Figure 2), suggesting a regulatory role of nuclear MP genes in feed efficiency.

Feed efficiency is a complex trait and the underlying biological mechanism is not currently understood. While mitochondrial function is associated with cellular energy metabolism, the link between mitochondrial function and feed efficiency is not demonstrated adequately. The results from this study provide a plausible link through the differential gene expression of MP genes in the blood (RNAseq). The under-expression of MP genes in feed efficient animals suggests a lowered mitochondrial activity, which means decreased energy production and less heat loss and channelling of energy saved into increased milk production (Goddard and Grainger, 2004, Nkrumah et al., 2006). Other notable interrelated factors within the electron transport chain associated with the efficiency of energy production, which accounts for basal metabolic rate, are proton and electron leaks (reviewed in Brand et al. (1999)). The DE genes and key genes from the co-expression clusters can be used as the potential markers for the genomic evaluation of feed efficiency in cattle. However, the regulatory role of MP genes from the nuclear genome in feed efficiency warrants further investigation.

6.2 Limitations

The empirical results reported in the three research chapters of this thesis should be considered in light of some limitations. In Chapter 3, the handling of the missing genotype and heteroplasmy, in part arising from nuclear mitochondrial sequences (NUMTs), could have slight influence on the estimated diversity. In Chapter 4,

the association of mitochondrial protein gene expression to tissue energy metabolism was based on tissues whose metabolic rate or energy demands were previously known as tissue metabolic rates were not available for most tissue. Also, there were some inconsistent expression profiles in few tissues between two animals. In Chapter 5, the association of the gene expression to the feed efficiency was based on the blood transcriptome and it was evident from Chapter 4 that the gene expression profile varies with tissues. However, overall, these limitations should not adversely affect the results and their interpretation.

In Chapter 3, the sites with missing genotype and indels were removed from the analysis. The removal of sites with missing genotypes and gaps is a standard practice in mitochondrial population genetic studies. The removal of sites could have invariably resulted in the underestimation of diversity. However, for this study, the genetic diversity estimate was still high, indicating the feasibility of undertaking genomic evaluations. Alternately, it was demonstrated that the missing genotypes would not affect the genomic evaluation as they can be imputed with high accuracy. Further, the current study considered the heteroplasmic sites in the analysis. A low level of false-positive heteroplasmy is reported as common errors in next generation sequencing. Considering this, additional analysis (not reported in Chapter 3) was undertaken to establish their authenticity through allele depths and the transmission patterns. The average allele depths of heteroplasmic alleles in the final dataset were determined as 36:10 (first:second allele), which is a reliable ratio. The transmission pattern of heteroplasmic sites on 36 dam-offspring pairs examined showed that heteroplasmy was not usually transmitted in about 50% of the cases (19 dam-offspring pairs), transmitted in about 30% (11 dam offspring pairs) and 14% had heteroplasmy not present in the dam but present in offspring, while the rest were partial transmission from dam and some newly acquired by offspring. This partially justified the use of heteroplasmic sites in the diversity estimation but could have

influenced the diversity estimates. The 1000 Bull Genomes Project contains many tissue samples with low mitochondrial content that were used for sequencing (e.g. semen), the impact of NUMTs manifesting as heteroplasmy was minimized to the level of non-semen tissues by ensuring a comparable distribution of heteroplasmy between tissues (Chapter 3, Materials and Methods). While this method provides an option to minimize heteroplasmy due to NUMTs in samples with low mitochondrial content, it is not feasible to completely eliminate the effect of NUMTs and they may continue to affect nucleotide and haplotype diversities.

In Chapter 4, the association of the MP gene expression in tissues to energy metabolism were based on the enrichment of the OXPHOS and metabolic pathways. The over-expression and under-expression in tissues were further supported by the energy demand or metabolic rates of tissues available in the literature. On the other hand, the metabolic rates of most of the tissues were not available and assumed to follow a similar pattern in other tissues. The other concern is the consistency in the gene expression within tissue in the two animals, especially among the under-expressing tissues (e.g. thyroid and ovary). The trend could have been more convincing had there been additional animals in the study.

Further, tissues are a heterogeneous mixture of cells and cell types differing in metabolic profile and sampling specific site and cell types could have been key for better consistency. Furthermore, the post-transcriptional modification and post-mortem sampling associated decay could have added noise to the already low expression in these tissues. Overall, these issues may be resolved by examining the gene expression in specific cell types using the single-cell RNA sequencing technology.

The Chapter 5, there is no prescribed tissue for RNA sequencing for feed efficiency studies, but blood is among the previously sampled tissues apart from muscle, liver and rumen epithelium (Kong et al., 2016b, Khansefid et al., 2017, Salleh et al., 2017). The expression of MP genes was highly variable among the tissue types and

was actually under-expressed in the blood (Chapter 4). Despite this, blood was sampled in the study for its relative ease for sampling and potential combination with other routine haematological assays. Nevertheless, sampling a few other tissues with over-expression of the MP genes group (e.g. skeletal muscle) could have consolidated the findings. Generally, the sample sizes of the feed efficiency studies including this study are low (109 animals). A larger sample size is expected to provide better contrast in expression between high and low RFI groups.

6.3 Opportunities and future directions

The results fill in some critical information gaps towards utilising the mitochondrial genome for genetic improvement of dairy traits. They have enhanced the understanding of variations in the whole mitochondrial genome in cattle and established that there is adequate diversity within breeds for association studies. The findings provided an important link between MP gene expression and phenotypes (tissue energy demand and feed efficiency) and the potential biological mechanisms. However, the findings could be further expanded using emerging technologies such as long-read sequencing, single-cell RNA sequencing, and proteomic analyses. Combining mitochondrial genomics, transcriptomics, and proteomics into a single study would provide more power. The imputation of the mitochondrial genome has been promising, and genomic evaluation based on the variants from the mitochondrial protein genes from the mitochondrial and nuclear genomes may be undertaken in the near future.

6.3.1 Long read genome sequencing

Long-read sequencing generates substantially longer reads than the conventional Next-Generation sequencing technologies and has several advantages over short-read sequencing. According to Logsdon et al. (2020), the Oxford Nanopore Technologies (ONT); MinION, GridION, and PromethION long-read sequencing platforms generate reads between 10-60 kb and the ultra-long reads of 100-200 kb at N50

(the sequence length of the shortest contig at 50% of the total genome length) and similar length (5-60 kb) from the PacBio Pacific Biosciences' (PacBio) RSII, Sequel, and Sequel II at N50. Some of the advantages of long-read sequencing are in eliminating amplification bias, preserving base modifications, and identifying large structural variations and transcript isoforms (Roberts et al., 2013, Depledge et al., 2019). One of the disadvantages of the long-read sequencings is the higher base error rates (10-15%), but repeated measurement of the sequences at adequate coverages can reduce the error rates (Pollard et al., 2018, Wang et al., 2019).

The advantage of long-read sequencing technologies in the context of mitochondrial genome sequencing is the ability to read the entire genome (~16 kb) as a single fragment. The mitochondrial genome as a single read would enable the detection of haplotypes based on the entire genome, which is a more accurate representation compared to using a shorter region or a gene. The whole mitochondrial genome read as a single fragment would also eliminate the interference from the nuclear mitochondrial sequences (NUMTs). The NUMTs often get co-amplified and the mitochondrial genome and may interfere with genetic analyses (see Chapter 2 Literature Review). Apart from detecting the transcript isoforms and base modification, long-read sequencing can be beneficial in the understanding of mitochondrial genome transcription. The longer reads of transcripts close to the size of the genome would elucidate the pattern of near-complete polycistronic transcription of the mitochondrial genome.

However, the obvious disadvantage of long-read sequencing of mitochondrial genomes is in detecting mitochondrial heteroplasmy. The ability to detect heteroplasmy is directly dependant on sequencing depth (Guo et al., 2013). The conventional sequencing techniques may be a better option for studies primarily looking at the heteroplasmy as in mitochondrial disease in humans. On the other hand, a low proportion of heteroplasmy is always present and not pathogenic unless in large

proportions (Payne et al., 2013). The prevalence of heteroplasmy, if being in a high proportion is important, can be detected using long-read sequencing at greater depths.

6.3.2 Imputation of missing mitochondrial genotypes

The evaluation of the mitochondrial genomic diversity from imputed sequences may not be recommended. However, imputation of the mitochondrial genome is of interest from the aspect of genomic evaluation. With filters and procedures to minimize false-positive heteroplasmy likely resulting from NUMTs, especially in DNA sequences sampled from tissues with low mitochondrial content (i.e. semen), it has been learnt that the empirical imputation by masking the sequence variants yields high prediction accuracy (Chapter 3), indicating that imputation of missing mitochondrial genotypes is feasible. Further, there is a possibility of imputing from the existing mitochondrial SNPs on SNP chips to the denser markers and even sequence. For example, BovineLDv1™ and BovineHD BeadChip™ from Illumina (www.illumina.com) have 13 and 339 mitochondrial SNPs, respectively. The historical animals on low-density (LD) panels do not have enough mitochondrial SNPs to be imputed to high-density (HD). However, the animals on the BovineHD chip are likely to be imputed well to mitochondrial sequence levels, because mitochondrial marker coverage is greater. This would also provide an opportunity to revisit genomic evaluations by including the imputed mitochondrial sequences.

Further, there are avenues to customise mitochondrial SNP panels for genotyping mitochondrial genome to evaluate genomic diversity and the imputation to sequence for genomic prediction simultaneously. This may be done by incorporating lifted over variants currently used to predict haplogroups in cattle dometree (Peng et al., 2015) (Chapter 3, Table S1). The mitochondrial SNP panels can also be independent for mitochondrial population genetic studies.

6.3.3 Single-cell RNA sequencing

RNA sequencing offers insights into the transcriptome and prospects of identifying hub genes and elucidating the critical biological pathways contributing to the phenotype. Transcriptome analysis enables the mapping of genes to phenotype. However, tissue-based RNA sequencing can be less accurate, considering the heterogeneity of cell types in tissues. The gene expression profile of tissue is the aggregated gene expression of all cell types. On the other hand, the gene expression among similar cell types is also reported heterogeneous (Huang, 2009). The genes that show mutually exclusive expression in individual cells may also be observed as genes showing co-expression in expression analyses of bulk cell populations (reviewed in Kukurba and Montgomery (2015)). Thus, a more precise understanding of the transcriptome within an individual cell type is essential for elucidating their role in cellular functions and understanding how gene expression affects traits (Hwang et al., 2018).

Single-cell RNA sequencing (scRNAseq) could provide a better resolution to elucidate further the roles of MP gene expression (among cell types for energy demand and the blood in the feed efficiency) and elaborate the understanding of the biological processes. A single-cell multi-omics approach integrating the existing information from DNA, RNA and proteins (Dey et al., 2015, Genshaft et al., 2016) would be even more helpful in teasing out the detailed process of relaying the genomic information leading to phenotypic differences. The choice of cell types in scRNAseq would depend on the ease of tissue sampling and isolation of single cells and the biological relevance of the sample to the study. It is relatively straight forward in the studies comparing gene expression across cell types as in tissues in Chapter 4, where the primary cell types of the organ would be the focus cells. For example, hepatocytes in the liver, myocardiocytes in the heart, white blood cells in the blood (as mitochondria and nucleus are absent in red blood cells) and myocytes in skeletal muscles. The scRNAseq of cell types is expected to have mini-

mum variation and increase consistency for reliable measurement of gene expression.

On the other hand, the choice of cell types for scRNAseq to associate with the phenotypic and production traits (e.g. RFI as in Chapter 5) may need to consider the underlying biology of the trait. Some complex traits (such as feed efficiency, fertility) can involve multiple pathways, which can differ considerably among the cell types. Therefore, initially, multiple cell types in scRNAseq experiments of relevant tissues needs to be studied in relation to a trait. Understanding that complex traits are an outcome of several interacting pathways across tissues, the key genes identified in each specific cell types may be combined and used as the quantitative trait loci for the genomic evaluation. It would also be of interest to examine how the enriched pathways across cell types could be contributing to the trait.

6.3.4 Is increased MP gene expression a result of increased transcription or more mitochondria?

The mechanisms driving the differential expression of mitochondrial genes is an area of ongoing research. Mitochondrial gene transcription depends on several factors encoded by the nuclear genome (see Literature Review). The over-expression of MP genes from both nuclear and mitochondrial genomes in tissues with higher energy demand and also a higher expression of nuclear MP genes in blood of the low feed efficiency group were observed (Chapter 4 & 5). However, no mitochondrial protein genes from the mitochondrial genome were differentially expressed between the two feed efficiency groups. This is plausible considering the measurement of gene expression between the groups on the same tissue (blood) where the mitochondrial content is generally constant. On the other hand, the variation in the expression across the tissues is potentially a result of varying mitochondrial numbers (Chapter 4) as the mitochondrial content of the tissues varies considerably. Currently, the mitochondrial content of most of the tissues in this study remains unknown.

The information on the mitochondrial content in the tissues could have added to the explanation and relevance of differential gene expression in varying energy metabolism among the tissues. Although the determination of the mitochondrial content in tissues is an exciting aspect to consider, it was beyond the scope of the current research. The mitochondrial content in the tissues can be determined by several methods. Conventionally, the real-time quantitative PCR has been used to determine the mitochondrial content. However, lately, the whole genome or exome sequencing data are used, where the content is determined by comparing the relative abundance of the mitochondrial DNA to the nuclear DNA. There are also cost-effective alternatives using nuclear DNA probes and whole mitochondrial genome sequencing is available with reasonable accuracy (see Literature Review). The association of the mitochondrial content from multiple samples and the association with production traits is another exciting area which could disentangle the overall role of the mitochondria.

6.3.5 Mitochondrial proteomics

According to the central dogma of molecular biology (Crick, 1970), protein is the ultimate product of the genetic code in the DNA. Also, because all biological processes in a cell involve proteins, the proteomics is a relevant dataset to characterize a biological system (Cox and Mann, 2007). There is significant relationship between mRNA and proteins, but they are generally low (correlation=0.25) (Guo et al., 2008, Gry et al., 2009, Koussounadis et al., 2015). The protein levels are not only dependent on mRNAs translation but also on protein turnover (Gygi et al., 1999, Lu et al., 2007), suggesting that proteomic analysis would be more reliable in context of the biological system. Further, the ability to isolate organelles has enabled organelle-specific proteins to be investigated using organelle proteomics (Brunet et al., 2003). In fact, the initial estimate of mitochondrial proteins was based on the mitochondrial proteomic analysis from rat liver mitochondria (Lopez et al., 2000). Since then, several high-precision, robust

and cost-effective proteome analysis techniques have evolved over last two decades and offer opportunities to characterise the mitochondrial proteome across species and tissues.

Currently, there are no specific lists of mitochondrial proteins in cattle, but the mitochondrial proteome is not likely to differ significantly across species. Aligning mitochondrial protein orthologues from human to cattle and sheep mapped with good success rate (for example, 1158 human mitochondrial proteins mapped to 1054 and 822 respectively (Chapter 4, Tables S4 & S5). This is no surprise because cattle and sheep are not well annotated. Further, the current mitochondrial protein list in human is not exhaustive as about 15% of mitochondrial proteins in human (~1400) await identification and the mitochondrial protein composition is reported to vary across tissues, where about 50% are ubiquitous, while another half is tissue-specific (reviewed in Calvo and Mootha (2010)). Thus, species-specific and the completeness of the mitochondrial protein list would help in better understanding of the role of mitochondrial protein genes in the tissue-specific functions and phenotypic traits. For the bovine mitochondrial proteins, an approach would be to examine the list of 528 genes currently not analogous to human mitochondrial proteins in the gene co-expression cluster analysis (Chapter 4, Cluster I in Fig.4, Table S7). The functional enrichment of these genes was closely related to functioning the heart (e.g. contraction, adrenergic signalling, glucagon signalling). There is a possibility that the gene set can contain some previously unidentified putative (new) mitochondrial protein genes in bovine or even the bovine analogues of about 100 genes that did not map during the conversion. Similarly, the genes clusters (ME2 and ME3) in the weighted gene co-expression network analysis (Chapter 5, Figure 2) and the list of differentially expressed genes between the high_RFI and low_RFI (Chapter 5, Table S3) may be examined and validated using antibody tagging and biochemistry.

The correlation of the mitochondrial protein transcriptome and proteome is another interesting area to explore specifically the link between the multiple copies of the mitochondrial genomes, mitochondrial protein transcripts and their protein abundances. This also interesting because the proteins from the mitochondrial genome were not differentially expressed between the high and low feed efficiency groups (Chapter 5, Table S3) nor were among the gene expression cluster associated with the residual feed intake (Chapter 5, Figure 2B & 2C). These results tend to agree with previous studies involving proteomic analysis of two divergent trait groups, which showed none of the proteins from the mitochondrial genome but some mitochondrial proteins from the nuclear genome are differentially abundant in other livestock species (Kong et al., 2016a, Baldassini et al., 2018, Fonseca et al., 2019, Miller et al., 2019, Wu et al., 2020). This observation suggests that differential gene expression across the tissues could be potentially due to the differences in mitochondrial content, which is the indicator of number of mitochondria. Within a tissue, the number of mitochondria in a cell is nearly constant, and therefore, the tissue level of mitochondrial transcripts and their proteins of the trait groups would not be significantly different.

Further, this also points that the MP from the nuclear genome could be rate-limiting to mitochondrial function. Firstly, enzymes and factors for the replication, transcription and translation of the mitochondrial genome are dependent on the nuclear genome. Furthermore, some miRNAs can inhibit the translation of MP from the mitochondrial genome. Secondly, even if the rates of replication of mitochondrial DNA (37b/sec) (Johnson and Johnson, 2001) and nuclear DNA (2kb/min, ~33b/sec) (Fangman and Brewer, 1992), transcription and translation differ between the nuclear and mitochondrial genomes, the mitochondrial genome being a small genome (16 kb) means the turnover could be higher and more abundant compared to the nuclear mitochondrial proteins. Thus, transcripts and proteins from the nuclear genome could plausibly have a limiting or regulatory role on the mitochondrial function. Therefore,

the proportion of mitochondrial proteins from both genomes, and the correlation of the transcripts and proteins of the mitochondrial genome needs to be ascertained considering their massive expression in high energy demand tissues and association with traits related to mitochondrial function.

6.3.6 A way forward for the use of mitochondrial genome in genetic evaluation

One of the main goals of this project was to advance the use of the mitochondrial genome for genomic evaluation in cattle. The high variation and accuracy of the imputation of the mitochondrial genome bring us a step closer to this goal. Based on the annotation of the variants in 1000 Bull genome project (Herman and Stothard, University of Alberta, unpublished), the final dataset excluding the indels and D-loop (which are non-coding and mostly up/down stream variant) had 2935 single nucleotide variants. These variants were primarily the missense variants, nearly equal synonymous and non-coding transcript exon variants, and few were related to the gain or loss of a stop codon (Figure 6.1). The missense variants result in the substitution of the amino acid, which may or may not make the protein non-functional. This indicates

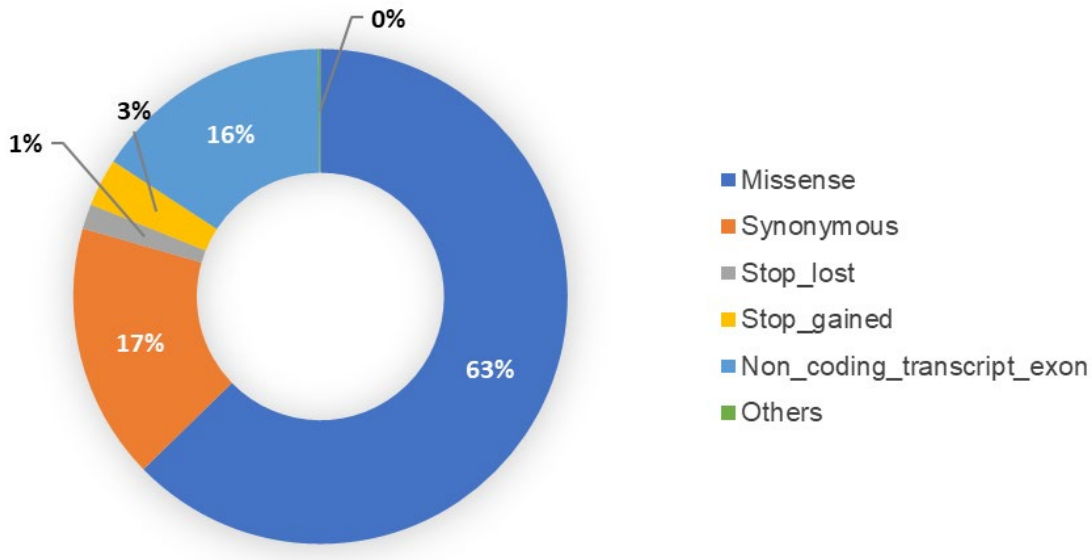


Figure 6.1. Types and composition of the single nucleotide variants in the coding region in the filtered dataset. Other variants (start lost, stop retained, incomplete terminal codon)

that variant substitutions may have a significant effect on animal production, fitness, and fertility traits.

Further, co-expression of the mitochondrial protein genes in tissues (Chapter 5) and association of the mitochondrial protein genes from the nuclear genome with the feed efficiency, an energy utilisation trait (Chapter 5), forms the basis of considering the interactions between the mitochondrial and nuclear genomes. The differentially expressed mitochondrial protein genes and the key genes from the co-expression clusters can be putative quantitative trait loci (QTL) for the trait.

The single marker linear regression model, commonly known as genome-wide association studies (GWAS), can be used to identify and validate the observed QTL. The nuclear mitochondrial protein genes exhibited prominent differential and co-expression in relation to the energy metabolism in tissues as well as in the feed efficiency (Chapter 4 & 5). The variants in these mitochondrial protein gene regions may be directly used for the GWAS and genomic predictions of feed efficiency.

A genome-wide association study including both mitochondrial and nuclear variants is feasible. For the nuclear genome, the variants from the key mitochondrial protein genes identified through differential and co-expression analyses may be used as in Kraja et al. (2019). This may be followed by conditional association analysis of mitochondrial variant to disentangle their effects using tools such as GCTA.

One of the preliminary tests for the role of the mitochondrial genome is to examine the proportion of the phenotypic variance explained by the nuclear and mitochondrial genomes and both together using genomic best linear unbiased prediction (gBLUP). While gBLUP based on the nuclear genome is routine, the genomic relationship matrix (G) of the mitochondrial genome has to consider their haploidy and maternal inheritance. The approach used for the X chromosome in male as implemented in tools such as GCTA (Yang et al., 2011) is likely to work for mitochondrial DNA (see Chapter 3

Materials and Methods). The mitochondrial genotypes may be coded as 0 (if same with the Ref) and 1 (if different from Ref) as in (Schutz et al., 1994). The G matrices, G(n) and G(m), resulting from the gBLUP can then be used for the genomic prediction based on the following equation:

$$\mathbf{y} = \mu + \mathbf{Z}\mathbf{g}_\mathbf{n} + \mathbf{Z}\mathbf{g}_\mathbf{m} + \mathbf{e}.$$

where

$\mathbf{y}$ is the vector of trait of genotyped reference animals, μ is the overall mean

$\mathbf{g_n}$ is the vector of additive genetic effects of the nuclear genome for the reference animals,

$\mathbf{g_m}$ is the vector of additive genetic effects of the mitochondrial genome for the reference animals,

$\mathbf{e}$ is the vector of random residuals, and

$\mathbf{Z}$ is the design matrix associating $\mathbf{g_n}$ and $\mathbf{g_m}$ with response variables.

$Var(g_n) = G_n\sigma^2_{gn}$; $Var(g_m) = G_m\sigma^2_{gm}$; where G_n is the GRM of nuclear genome, and G_m is the GRM of mitochondrial genome

This model would answer the question of how much variance in current phenotypic traits is due to the mitochondrial genome versus the nuclear genome. It would also result in GEBVs for both genomes that could be added together for an overall GEBV. Whether this was a valid approach leading to increased genetic gain would have to be determined by the impact on prediction accuracies. In addition, the association of the mitochondrial haplotypes to the traits is another approach (Liu et al., 2013), where the animals in a haplotype are coded 0 and different as 1. This would help identify the links between mitochondrial and MP regions to traits.

6.4 Concluding statement

This work establishes that it is possible to use the bovine mitochondrial genome in the genetic improvement of complex dairy traits. Firstly, the results have demonstrated that there is adequate mitochondrial genetic variation both within and across the breeds.Secondly, it has been shown that the mitochondrial genome can be imputed with high accuracy, so it is possible to impute previous high-density SNP chip data to sequence

level for use in genetic evaluations. Thirdly, substantial interaction of mitochondrial protein genes from the nuclear and mitochondrial genomes was demonstrated. Fourthly, mitochondrial protein genes, protein gene expression and functional pathways were associated with feed efficiency. Variants in these nuclear mitochondrial protein genes may be used in the genomic evaluation of feed efficiency. Thus, the findings will hopefully lead to more research enabling the incorporation of the mitochondrial genome in the genetic improvement of dairy and other livestock species.

6.5 Acknowledgements

I would like to thank Prof. Hans Daetwyler and Dr. Iona MacLeod for their invaluable comments and suggestions on the chapter.

6.6 References

Baldassini, W. A., S. F. M. Bonilha, R. H. Branco, J. C. S. Vieira, P. M. Padilha, and D. P. D. Lanna. 2018. Proteomic investigation of liver from beef cattle (Bos indicus) divergently ranked on residual feed intake. Molecular biology reports 45(6):2765-2773.

Brand, M. D., K. M. Brindle, J. A. Buckingham, J. A. Harper, D. F. S. Rolfe, and J. A. Stuart. 1999. The significance and mechanism of mitochondrial proton conductance. International Journal of Obesity 23(6):S4-S11.

Brunet, S., P. Thibault, E. Gagnon, P. Kearney, J. J. M. Bergeron, and M. Desjardins. 2003. Organelle proteomics: looking at less to see more. Trends in Cell Biology 13(12):629-638.

Calvo, S. E. and V. K. Mootha. 2010. The mitochondrial proteome and human disease. Annu Rev Genomics Hum Genet 11:25-44.

Cox, J. and M. Mann. 2007. Is Proteomics the New Genomics? Cell 130(3):395-398.

Crick, F. 1970. Central Dogma of Molecular Biology. Nature 227(5258):561-563.

Depledge, D. P., K. P. Srinivas, T. Sadaoka, D. Bready, Y. Mori, D. G. Placantonakis, I. Mohr, and A. C. Wilson. 2019. Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. Nature Communications 10(1):754.

Dey, S. S., L. Kester, B. Spanjaard, M. Bienko, and A. van Oudenaarden. 2015. Integrated genome and transcriptome sequencing of the same cell. Nature Biotechnology 33(3):285-289.

Fangman, W. L. and B. J. Brewer. 1992. A question of time: Replication origins of eukaryotic chromosomes. Cell 71(3):363-366.

Fonseca, L. D., J. P. Eler, M. A. Pereira, A. F. Rosa, P. A. Alexandre, C. T. Moncau, F. Salvato, L. Rosa-Fernandes, G. Palmisano, J. B. S. Ferraz, and H. Fukumasu. 2019. Liver proteomics unravel the metabolic pathways related to Feed Efficiency in beef cattle. Scientific Reports 9(1):5364.

Genshaft, A. S., S. Li, C. J. Gallant, S. Darmanis, S. M. Prakadan, C. G. K. Ziegler, M.. Lundberg, S. Fredriksson, J. Hong, A. Regev, K. J. Livak, U. Landegren, and A. K. Shalek. 2016. Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. Genome Biology 17(1):188.

Goddard, M. E. and C. Grainger. 2004. A review of the effects of dairy breed on feed conversion efficiency - an opportunity lost? Science Access 1(1):77-80.

Gry, M., R. Rimini, S. Strömberg, A. Asplund, F. Pontén, M. Uhlén, and P. Nilsson. 2009. Correlations between RNA and protein expression profiles in 23 human cell lines. BMC Genomics 10(1):365.

Guo, Y., C.-I. Li, Q. Sheng, J. F. Winther, Q. Cai, J. D. Boice, and Y. Shyr. 2013. Very Low-Level Heteroplasmy mtDNA Variations Are Inherited in Humans. Journal of Genetics and Genomics 40(12):607-615.

Guo, Y., P. Xiao, S. Lei, F. Deng, G. G. Xiao, Y. Liu, X. Chen, L. Li, S. Wu, Y. Chen, H. Jiang, L. Tan, J. Xie. Zhu, S. Liang, and H. Deng. 2008. How is mRNA expression predictive for protein expression? A correlation study on human circulating monocytes. Acta biochimica et biophysica Sinica 40(5):426-436.

Gygi, S. P., Y. Rochon, B. R. Franza, and R. Aebersold. 1999. Correlation between protein and mRNA abundance in yeast. Molecular and cellular biology 19(3):1720-1730.

Huang, S. 2009. Non-genetic heterogeneity of cells in development: more than just noise. Development (Cambridge, England) 136(23):3853-3862.

Hwang, B., J. H. Lee, and D. Bang. 2018. Single-cell RNA sequencing technologies and bioinformatics pipelines. Experimental & Molecular Medicine 50(8):1-14.

Johnson, A. A. and K. A. Johnson. 2001. Fidelity of Nucleotide Incorporation by Human Mitochondrial DNA Polymerase *. Journal of Biological Chemistry 276(41):38090-38096.

Khansefid, M., C. A. Millen, Y. Chen, J. E. Pryce, A. J. Chamberlain, C. J. Vander Jagt, C. Gondro, and M. E. Goddard. 2017. Gene expression analysis of blood, liver, and muscle in cattle divergently selected for high and low residual feed intake 1. Journal of Animal Science 95(11):4764-4775.

Kong, B.-W., K. Lassiter, A. Piekarski-Welsher, S. Dridi, A. Reverter-Gomez, N. J. Hudson, and W. G. Bottje. 2016a. Proteomics of Breast Muscle Tissue Associated with the Phenotypic Expression of Feed Efficiency within a Pedigree Male Broiler Line: I. Highlight on Mitochondria. PLOS ONE 11(5):e0155679.

Kong, R. S. G., G. Liang, Y. Chen, P. Stothard, and L. L. Guan. 2016b. Transcriptome profiling of the rumen epithelium of beef cattle differing in residual feed intake. BMC genomics 17:592-592.

Koussounadis, A., S. P. Langdon, I. H. Um, D. J. Harrison, and V. A. Smith. 2015. Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. Scientific Reports 5(1):10775.

Kraja, A. T., C. Liu, J. L. Fetterman, M. Graff, C. T. Have, C. Gu, L. R. Yanek, M. F. Feitosa, D. E. Arking, D. I. Chasman, K. Young, S. Ligthart, W. D. Hill, S. Weiss, J. Luan, F. Giulianini, R. Li-Gao, F. P. Hartwig, S. J. Lin, L. Wang, T. G. Richardson, J. Yao, E. P. Fern,ez, M. Ghanbari, M. K. Wojczynski, W. J. Lee, M. Argos, S. M. Armasu, R. A. Barve, K. A. Ryan, P. An, T. J. Baranski, S. J. Bielinski, D. W. Bowden, U. Broeckel, K. Christensen, A. Y. Chu, J. Corley, S. R. Cox, A. G. Uitterlinden, F. Rivadeneira, C. D. Cropp, E. W. Daw, D. van Heemst, L. de Las Fuentes, H. Gao, I. Tzoulaki, T. S. Ahluwalia, R. de Mutsert, L. S. Emery, A. M. Erzurumluoglu, J. A. Perry, M. Fu, N. G. Forouhi, Z. Gu, Y. Hai, S. E. Harris, G. Hemani, S. C. Hunt, M. R. Irvin, A. E. Jonsson, A. E. Justice, N. D. Kerrison, N. B. Larson, K. H. Lin, L. D. Love-Gregory, R. A. Mathias, J. H. Lee, M. Nauck, R. Noordam, K. K. Ong, J. Pankow, A. Patki, A. Pattie, A. Petersmann, Q. Qi, R. Ribel-Madsen, R. Rohde, K. S,ow, T. M. Schnurr, T. Sofer, J. M. Starr, A. M. Taylor, A. Teumer, N. J. Timpson, H. G. de Haan, Y. Wang, P. E. Weeke, C. Williams, H. Wu, W. Yang, D. Zeng, D. R. Witte, B. S. Weir, N. J. Wareham, H. Vestergaard, S. T. Turner, C. Torp-Pedersen, E. Stergiakouli, W. H. Sheu, F. R. Rosendaal, M. A. Ikram, O. H. Franco, P. M. Ridker, T. T. Perls, O. Pedersen, E. A. Nohr, A. B. Newman, A.

Linneberg, C. Langenberg, T. O. Kilpeläinen, S. L. R. Kardia, M. E. Jørgensen, T. Jørgensen, T. I. A. Sørensen, G. Homuth, T. Hansen, M. O. Goodarzi, I. J. Deary, C. Christensen, Y. I. Chen, A. Chakravarti, I. Brandslund, K. Bonnelykke, K. D. Taylor, J. G. Wilson, S. Rodriguez, G. Davies, B. L. Horta, B. Thyagarajan, D. C. Rao, N. Grarup, V. G. Davila-Roman, G. Hudson, X. Guo, D. K. Arnett, C. Hayward, D. Vaidya, D. O. Mook-Kanamori, H. K. Tiwari, D. Levy, R. J. F. Loos, A. Dehghan, P. Elliott, A. N. Malik, R. A. Scott, D. M. Becker, M. de,rade, M. A. Province, J. B. Meigs, J. I. Rotter, and K. E. North. 2019.Associations of Mitochondrial and Nuclear Mitochondrial Variants and Genes with Seven Metabolic Traits. Am J Hum Genet 104(1):112-138.

Kukurba, K. R. and S. B. Montgomery. 2015. RNA Sequencing and Analysis. Cold Spring Harb Protoc 2015(11):951-969.

Liu, C., J. Dupuis, M. G. Larson, and D. Levy. 2013. Association testing of the mitochondrial genome using pedigree data. Genet Epidemiol 37(3):239-247.

Logsdon, G. A., M. R. Vollger, and E. E. Eichler. 2020. Long-read human genome sequencing and its applications. Nature Reviews Genetics 21(10):597-614.

Lopez, M. F., B. S. Kristal, E. Chernokalskaya, A. Lazarev, A. I. Shestopalov, A. Bogdanova, and M. Robinson. 2000. High-thoroughput profiling of the mitochondrial proteome using affinity fractionation and automation. Electrophoresis 21(16):3427-3440.

Lu, P., C. Vogel, R. Wang, X. Yao, and E. M. Marcotte. 2007. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol 25(1):117-124.

Miller, B. A., A. Chapwanya, T. Kilminster, T. Scanlon, J. Milton, H. Osório, C. Oldham, J. Greeff, D. R. Bergfelt, A. M. O. Campos, and A. M. Almeida. 2019. The ovine hepatic mitochondrial proteome: Understanding seasonal weight loss tolerance in two distinct breeds. PloS one 14(2):e0212580-e0212580.

Nkrumah, J. D., E. K. Okine, G. W. Mathison, K. Schmid, C. Li, J. A. Basarab, M. A. Price, Z. Wang, and S. S. Moore.. 2006. Relationships of feedlot feed efficiency, performance, and feeding behavior with metabolic rate, methane production, and energy partitioning in beef cattle1. Journal of Animal Science 84(1):145-153.

Payne, B. A. I., I. J. Wilson, P. Yu-Wai-Man, J. Coxhead, D. Deehan, R. Horvath, R. W. Taylor, D. C. Samuels, M. Santibanez-Koref, and P. F. Chinnery. 2013. Universal heteroplasmy of human mitochondrial DNA. Human molecular genetics 22(2):384-390.

Peng, M.-S., L. Fan, N.-N. Shi, T. Ning, Y.-G. Yao, R. W. Murphy, W.-G. Yao, R. W. Murphy, W. -Z. Wang, and Y. -P Zhang. 2015. DomeTree: a canobical toolkit for mitochondrial DNA analysis in domesticated animals. Molecular Ecology Resources 15(5):1238-1242.

Pollard, M. O., D. Gurdasani, A. J. Mentzer, T. Porter, and M. S. Sandhu. 2018. Long reads: their purpose and place. Human Molecular Genetics 27(R2):R234-R241.

Roberts, R. J., M. O. Carneiro, and M. C. Schatz. 2013. The advantages of SMRT sequencing. Genome Biology 14(7):405.

Salleh, M. S., G. Mazzoni, J. K. Hoglund, D. W. Olijhoek, P. Lund, P. Lovendahl, and H. N. Kadarmideen. 2017. RNA-Seq transcriptomics and pathway analyses reveal potential regulatory genes and molecular mechanisms in high- and low-residual feed intake in Nordic dairy cattle. BMC Genomics 18(1):258.

Schutz, M. M., A. E. Freeman, G. L. Lindberg, C. M. Koehler, and D. C. Beitz. 1994. The effect of mitochondrial DNA on milk production and health of dairy cattle. Livestock Production Science 37(3):283-295.

Wang, Y.-C., N. D. Olson, G. Deikus, H. Shah, A. M. Wenger, J. Trow, C. Xiao, S. Sherry, M. L. Salit, J. M. Zook, M. Smith, and R. Sebra. 2019. High-coverage, long-read sequencing of Han Chinese trio reference samples. Scientific Data 6(1):91.

Wu, J., X. Wang, R. Ding, J. Quan, Y. Ye, T. Gu, Z. Xu, E. Zheng, G. Cai, Z. Wu, M. Yang, and J. Yang. 2020. Identification of Important Proteins and Pathways Affecting Feed Efficiency in DLY Pigs by iTRAQ-Based Proteomic Analysis. Animals (Basel) 10(2):189.

Yang, J., S. H. Lee, M. E. Goddard, and P. M. Visscher. 2011. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88(1):76-82.

APPENDICES

Appendix 1: Co-Author Contributions to Published and to be Submitted Research Articles

Jigme Dorji made significant leading contributions to the publication in line and published article listed hereafter.

Chapter 3. Cattle maternal diversity inferred from 1,883 *taurine* and *indicine* mitogenomes (to be submitted to "*Scientific Reports*")

Jigme Dorji (**JD**), Christy J. Vander Jagt (CJV), Amanda J. Chamberlain (AJC), Benjamin G. Cocks (BGC), Iona M. MacLeod (IMM), Hans D. Daetwyler (HDD)

Conceptualisation: **JD**, IMM, HDD jointly conceptualised the study

Sample acquisition: CJV, AJC, HDD acquired the samples

Animal ethics approval: Not required as it only involved genotyping

Methodology: **JD**, IMM, HDD

IMM, HDD assisted and approved the method

Data analysis: **JD**, CJV, AJC

CJV and AJC processed and made the VCF format of the data from 1000 Bull Genome available

Funding & resource acquisition: HDD, BGC

Project administration: HDD, BGC

Supervision: HDD, IMM, BGC

Writing & visualisation: **JD**

Writing (review & editing): All authors

Chapter 4: Expression of mitochondrial protein genes encoded by nuclear and mitochondrial genomes correlate with energy metabolism in dairy cattle

Jigme Dorji (**JD**), Christy J. Vander Jagt (CJV), Josie B. Garner (JBG), Leah C. Marett (LCM), Brett A. Mason (BAM), Coralie M. Reich (CMR), Ruidong Xiang (RX), Emily L. Clark (ELC), Benjamin G. Cocks (BGC), Amanda J. Chamberlain (AJC), Iona M. MacLeod (IMM) and Hans D. Daetwyler (HDD)

Conceptualisation: **JD**, IMM, HDD, BGC

Sample acquisition: CPP, BAM, CMR, JBG, CJV, LCM, ELC

Animal ethics approval: AJC, JBG, LCM

Methodology: **JD**, RX, AJC, IMM, HDD

RX assisted in the generation of gene expression counts.

AJC assisted in differential gene expression analysis

IMM, HDD assisted on topologically associating domain analysis of nuclear mitochondrial proteins and overall advice on methods

Data analysis: **JD**, AJC, CJV

AJC and CJV processed the RNAseq data and made the BAM/SAM files of the tissues available

Funding & resource acquisition: HDD, BGC

Project administration: HDD, BGC

Supervision: HDD, IMM, BGC

Writing & visualisation: **JD**

Writing (review & editing): All authors

Chapter 5: Mitochondrial protein gene expression and the oxidative phosphorylation pathway associated with feed efficiency and energy balance in dairy cattle

Jigme Dorji (**JD**), Iona M. MacLeod (IMM), Amanda J. Chamberlain (AJC), Christy J. Vander Jagt (CJV), Phuong N. Ho (PNH), Majid Khansefid (MK), Brett A. Mason (BAM), Claire P. Prowse-Wilkins (CPP), Leah C. Marett (LCM), William J. Wales (WJW), Benjamin G. Cocks (BGC), Jennie E. Pryce (JEP), and Hans D. Daetwyler (HDD)

Conceptualisation: **JD**, IMM, HDD, JEP, BGC

Sample acquisition: AJC, LCM, CPP, BAM, CJV, PNH, MK, WJW

Animal ethics approval: AJC, LCM, WJW

Methodology: **JD**, IMM, HDD, MK, PNH, JEP

MK estimated the residual feed intake and gene expression counts in the validation datasets.

JEP suggested the incorporation of energy balance and blood metabolites

IMM, HDD assisted with and approved the methods

Data analysis: **JD**, PNH

PNH assisted in the prediction of residual feed intake and energy balance

Funding & resource acquisition: HDD, JEP, BGC

Project administration: HDD, JEP, BGC

Supervision: HDD, IMM, BGC

Writing & visualisation: **JD**

Writing (review & editing): All authors

Appendix 2: Conference Proceedings

1. **J. Dorji**, C. J. Vander Jagt, J. B. Garner, L. C. Marett, B. A. Mason, C. M. Reich, C. P. Prowse-Wilkins, R. Xiang, P. N. Ho, J. Pryce, B. G. Cocks, A. J. Chamberlain, I. M. MacLeod, and H. D. Daetwyler, Mitochondrial genome diversity and association of mitochondrial protein gene expression with energy metabolism in dairy cattle. *J. Dairy Sci. 103 (Suppl. 1)*, 2020 American Dairy Science Association® Annual Meeting, Monday, June 22, to Wednesday, June 24, 2020 (PhD Oral competition). Available at https://www.journalofdairyscience.org/article/S0022-0302(20)30816-X/pdf

2. **J. Dorji**, I.M. MacLeod, C.J. Vander Jagt, A.J. Chamberlain and H.D. Daetwyler, 2019 Mitochondrial gene expression is associated with organ and tissue metabolism in dairy cattle. *Proc. Assoc. Advmt. Anim. Breed. Genet. 23:444-447* (Poster presentation).
Available at http://www.aaabg.org/aaabghome/AAABG23papers/109Dorji23444.pdf

Appendix 3: Chapter 3 Supplementary materials

Table S1: Variants used in prediction of cattle mitochondrial haplogroup lifted over from Bovine Reference Sequence (V00654) to the ARS-UCD1.2_M reference sequence.

RPQT:8,106,166,250,297,301,2560,2981,3441,3552,3602,5503,5745,5892,6438,6774,7332,7358,7516,8190,8372C,8496,9007,9980,10333,10693G,11002,11136,11421,11844,12236,12470,12471,12624,12677,12686,12752,12803,12902,13007,13058A,14038,14827,15136,15619,15629,15820,15953,15957G,16059,16123,16124,16139,16250,16257,16303

PQT:106,166,169,250,301,2560,3552,5503,5745,5892,6438,7358,8372C,10693G,11002,12470,12677,12752,13007,14038,15136,15629,15953,15955G,16124,16257

QT:169,2560,5503,8372C,11002,12470,12677,12752,13007,14038,15136,15629,15955G,16257

Q:169,2560,3240,3417T,5503,7920,8320,8372C,10719A,10929,11002,11091,12435,12470,12677,12752,13007,14038,14110,15136,15629,15955G,16257

Q1:169,1459,2560,3240,3417T,5503,7920,8320,8372C,10719A,10929,11002,11091,12435,12470,12677,12752,13007,14038,14110,15136,15629,15955G,16257

Q1a:169,244T,1459,2560,3240,3417T,5503,7920,8320,8372C,10719A,10929,11002,11091,11768,12435,12470,12677,12752,13007,13622,14038,14110,15136,15629,15955G,16114,16257

Q1a_FJ971083_HQ184034_HQ184035:169,244T,1459,2560,3240,3417T,5503,7920,8320,8372C,8619,10719A,10929,11002,11091,11768,12435,12470,12677,12752,13007,13508,13622,14038,14110,15136,15955G,16114,16257

Q1_HQ184039:169,471,1459,2560,3240,3417T,5503,5718,6660,7832,7920,8320,8372C,10719A,10929,11002,11091,12435,12470,12482,12677,12752,13007,14038,14110,15092,15100A,15136,15629,15955G,16198,16257

Q1_HQ184036_HQ184037_HQ184038_EU177866_EU177867:1459,2560,3240,3417T,5503,7920,8320,8372C,8407,10719A,10929,11002,11091,11478,12379,12435,12470,12677,12732,12752,12879,12926,13007,13821,14038,14110,15136,15629,15955G, 16060,16081,16257

Q2:169,2560,3240,3417T,5503,6438,7920,8320,8372C,10719A,10929,11002,11091,12435,12470,12677,12752,13007,14038,14110,15136,15158,15629,15923,15955G,16257

Q2_HQ184030_HQ184031_HQ184032:169,2560,3240,3417T,5503,6438,7920,8320,8372C,8881,10719A,10929,11002,11091,12435,12470,12677,12752,13007,14038,14110,15136,15158,15629,15923,15955G,16257

P:106,166,169,190,222,250,301,302,1130,1483,2560,2587,3381,3552,4678,5158,5503,5745,5892,5901,6162,6438,7358,7954,7996,8238,8360,8372C,10128,10693G,11002,11142,12018,12379,12470,12677,12752,13007,13823,14038,14131T,14875,15136,15629,15675,15953,15955G,15996,16051,16053,16060,16076,16124,16233,16257,16266

P1:169,190,222,250,301,302,1130,1483,2560,2587,3381,3552,4678,5158,5503,5745,5892,5901,6162,6438,7358,7954,7996,8238,8360,8372C,10128,10693G,11002,11142,12018,12379,12435,12470,12677,12752,13007,13823,14038,14131T,14875,15136,15629,15675,15953,15955G,15996,16051,16053,16060,16076,16124,16233,16257,16266

P1_GU985279:190,222,250,301,302,1130,1483,2147,2560,2587,3381,3552,4254,4295,4678,5158,5503,5745,5892,5901,6162,6438,7358,7954,7996,8238,8360,8372C,10128,10693G,11002,11142,11742,12018,12379,12435,12470,12471,12527,12677,12752,13007,13823,14038,14131T,14582,14875,15136,15629,15675,15953,15955G,15996,16021,16051,16053,16060,16076,16124,16143,16233,16257,16266,16303

P1_JQ437479:169,190,222,250,301,302,1130,1483,2560,2587,3190,3304,3381,3552,4678,5158,5503,5745,5892,5901,6162,6438,6453,7358,7954,7996,8238,8360,8372C,10128,10693G,11002,11142,12018,12379,12435,12470,12677,12752,13007,13541,13823,14038,14131T,14875,15136,15473,15629,15675,15953,15955G,15996,16051,16060,16076,16124,16233,16257,16266

P2:106,166,173,190,222,250,301,302,1130,1483,2173,2560,2587,3381,3552,4254,4295,4678,5158,5503,5683,5745,5892,5901,6162,6438,7358,7954,7996,8238,8360,8372C,10128,10693G,11002,11142,11470,12018,12379,12470,12677,12740,12752,13007,13823,14038,14131T,14875,15136,15675,15953,15955G,15996,16051,16053,16060,16076,16124,16233,16249,16257,16266

R:8,106,166,250,297,301,782,895,2044,2560,2570,2981,3441,3552,3602,4192,4855,5503,5745,5892,6081,6120,6438,6477,6774,6985,7332,7358,7361,7516,8190,8372C,8496,9007,9178,9661,9866,9980,10333,10693G,11002,11136,11394,11421,11757,11844,12047,12236,12470,12471,12624,12677,12686,12752,12803,12902,13007,13058A,13156,13262,13460,13508,14002,14038,14626,14827,14908,14953,15067,15136,15619,15629,15820,15902,15953,15955G,16059,16078,16123,16124,16129,16137,16139,16233,16250,16252,16266,16303

R1:8,106,166,173,201,250,297,301,782,895,2044,2560,2570,2981,3304,3441,3552,3602,4192,4363,4855,5503,5616,5745,5892,5937,6081,6120,6162,6438,6477,6774,6985,7332,7358,7361,7516,7933,8190,8372C,8496,8769,9007,9178,9482,9661,9866,9980,10333,10693G,10929,11002,11136,11394,11421,11757,11844,12047,12236,12470,12471,12624,12677,12686,12752,12803,12902,13007,13058A,13130,13156,13262,13460,13508,13719,14002,14038,14053,14459,14626,14827,14908,14953,15067,15136,15619,15626,15629,15820,15848,15902,15953,15955G,16059,16078,16123,16124,16129,16137,16139,16233,16250,16252,16266,16303

R1a:8,106,166,173,201,250,297,301,782,895,2044,2560,2570,2981,3304,3441,3552,3602,4192,4363,4855,5503,5616,5745,5892,5937,6081,6120,6162,6438,6477,6774,6985,7332,7358,7361,7516,7933,8190,8372C,8496,8769,9007,9178,9482,9661,9866,9980,10333,10693G,10929,11002,11136,11394,11421,11757,11844,12047,12236,12470,12471,12624,12677,12686,12752,12803,12902,13007,13058A,13130,13156,13262,13460,13508,13719,14002,14038,14053,14459,14626,14827,14908,14953,15067,15136,15581,15619,15626,15629,15820,15848,15902,15953,15955G,16059,16078,16123,16124,16129,16137,16139,16233,16250,16252,16266,16303

R1a_FJ971085:8,106,166,173,201,250,297,301,782,895,2044,2560,2570,2981,3304,3441,3552,3602,4192,4363,4855,5503,5616,5745,5892,5937,6081,6120,6162,6438,6477,6774,6985,7332,7358,7361,7516,7933,8190,8372C,8496,8769,9007,9178,9482,9661,9866,9980,10333,10693G,10929,11002,11136,11394,11421,11757,11844,12047,12236,12470,12471,12624,12677,12686,12752,12803,12902,13007,13058A,13130,13156,13262,13460,13508,13719,14002,14038,14053,14459,14626,14827,14908,14953,15067,15136,15581,15619,15626,15629,15820,15848,15902,15953,15955G,16059,16078,16123,16124,16129,16137,16139,16233,16250,16252,16266,16303

R1a_FJ971084:8,106,166,173,201,250,297,301,782,895,2044,2560,2570,2981,3304,3441,3552,3602,4192,4363,4855,5148,5503,5616,5745,5892,5937,6081,6120,6162,6438,6477,6774,6852,6985,7332,7358,7361,7516,7933,8190,8372C,8496,8769,9007,9178,9482,9661,9866,9980,10333,10693G,10929,11002,11136,11394,11421,11757,11844,12047,12236,12470,12471,12624,12677,12686,12752,12803,12902,13007,13058A,13130,13156,13262,13460,13508,13719,14002,14038,14053,14459,14626,14827,14908,14953,15067,15136,15581,15619,15626,15629,15820,15848,15902,15953,15955G,16059,16078,16123,16124,16129,16137,16139,16233,16250,16252,16266,16303

R2:8,24,106,166,250,297,301,782,895,2044,2560,2570,2981,3441,3552,3602,4192,4630,4855,4977,5503,5745,5892,6081,6120,6438,6477,6540,6774,6985,7332,7358,7361,7516,8190,8372C,8496,8522,9007,9178,9661,9807,9866,9980,10333,10693G,10986,11002,11136,11202,11394,11421,11751,11757,11844,12047,12236,12470,12471,12624,12677,12686,12752,12803,12902,13007,13058A,13091,13156,13262,13460,13508,14002,14038,14533,14626,14827,14843,14908,14953,15067,15136,15619,15629,15820,15902,15953,15955G,15986T,16059,16078,16123,16124,16129,16137,16139,16201+A,16202,16233,16250,16252,16266,16303

T:169,13007,16257

T1'2'3'6'7:169,16257

T1:169,16052,16115,16257

T1a:169,2057+C,16052,16115,16257

T1a1:169,2057+C,11190,16052,16115,16257

T1a1_JN817315_JN817332:8,169,2057+C,11190,16052,16115,16257

T1a1_JN817335:169,173,2057+C,4114,11190,16052,16115,16257

T1a2:169,2057+C,3406,10563,12732,16052,16115,16257

T1a2a:169,2057+C,3406,7299,10563,12732,16052,16115,16257

T1a2a_JN817342:169,2057+C,3406,7299,10563,12732,15656A,16052,16115,16257

T1a2a_JN817337:169,2057+C,3406,7299,10563,12732,16052,16115,16141,16257,16304

T1a2_JN817333:169,2057+C,3406,10563,12732,13164,16052,16115,16257

T1a3:169,2057+C,9840,16052,16115,16257

T1a3_JN817314:169,756,2057+C,6388,9840,14327,16052,16115,16257

T1a3_JN817331:169,2057+C,4167,8235,8251,9840,11791,16052,16115,16135,16233,16257

T1a4:169,2057+C,4742,7322,9604,16052,16115,16121,16169,16257

T1a4_JN817317:8,169,1300,2057+C,4742,7322,9604,16052,16115,16121,16169,16257

T1a4_JN817340:169,2057+C,4742,7322,8160,9604,11956,16052,16115,16121,16169,16257

T1a5:169,2057+C,7068,12476,16052,16115,16257

T1a5_JN817338:169,353G,2025,2057+C,7068,7511,9808,12452,12476,16052,16115,16143,16166,16257

T1a5_JN817318:169,2057+C,3546A,3942,7068,12476,16052,16115,16257

T1a6:8,169,2057+C,16052,16115,16257

T1a6_JN817336:8,169,2057+C,5017,5227,5814,13125,13673,14694,14905,15527,16052,16115,16257

T1a6_JN817308:8,169,2057+C,2080,10276,16052,16060,16110,16115,16124,16257

T1a7:106,169,2057+C,16052,16115,16257

T1a7a:106,169,2057+C,8669,16052,16115,16257

T1a7a_JN817345:106,169,1842,2057+C,8669,8861,16052,16115,16257

T1a7_EU177846:106,169,2057+C,8961,15715d,16052,16078,16115,16117,16257

T1a_JN817341:169,2057+C,16018,16052,16115,16257

T1a_GU947020:169,2057+C,4159,7029,8407,11430,14429,16052,16115,16141,16149,16257

T1a_JN817303:169,2057+C,3684,4453,11028,15064,16052,16115,16248,16249,16257

T1a_JN817347:169,2057+C,9686,12902,13425,15216,16052,16115,16257,16304

T1a_JN817316:169,2057+C,2584,10603,11024,14582,15976,16052,16115,16257

T1a_JN817312:169,665,2057+C,10519,13061,13898,16052,16115,16257,16262

T1a_JN817339:169,711,2057+C,9445,10857A,16052,16115,16257,16266

T1a_EU177843:2057+C,9920,13685,16052,16115,16257

T1a_EU177844:169,2057+C,6048,15084,16052,16115,16257

T1a_JN817313:169,267T,2057+C,13880,16052,16115,16257

T1b:169,7544,16052,16115,16257

T1b1:169,7544,16024,16052,16115,16257

T1b1a:169,178,7544,14758,16024,16052,16115,16257

T1b1a_JN817351:169,178,5638,7544,14758,16024,16052,16115,16257,16293

T1b1b:169,7544,14525T,16024,16052,16115,16257

T1b1b1:169,7544,13203,14525T,16024,16052,16115,16257

T1b1b1a:169,7544,13203,14525T,16024,16052,16058,16115,16257

T1b1b1a1:169,7544,13203,14525T,16024,16052,16058,16115,16149,16257

T1b1b1a1_JN817349:169,7544,13203,14525T,16024,16052,16058,16115,16140,16149,16257

T1b1b1a1_JN817305:169,2967,7544,9433,12362,13203,14525T,16024,16052,16058,16115,16149,16257,16304

T1b1b1a2:169,7544,12471,13203,14525T,16024,16052,16058,16115,16257

T1b1b1a2a:169,179,1483,7544,10462,12470,12471,13203,14525T,16024,16052,16058,16115,16257

T1b1b1a3:169,190,498,7544,8712,12527,13203,13302,14525T,16024,16052,16058,16115,16257,16318

T1b1b1a3a:169,190,498,7544,8712,12470,12471,12527,13203,13302,14525T,16024,16052,16058,16115,16257,16318

T1b1b1a3a1:169,190,498,7544,8712,12470,12471,12527,13203,13302,14525T,16024,16052,16058,16115,16124,16257,16318

T1b1b1a3a1_KF163080:169,190,498,6477,7544,8712,12470,12471,12527,13203,13302,14525T,16024,16052,16058,16115,16124 ,16257,16318

T1b1b1a3a2:169,190,498,3071,7544,8712,12470,12471,12527,13203,13302,14525T,16024,16052,16058,16115,16318

T1b1b1a3a2_KF163066:169,190,498,3071,7544,8190,8712,12470,12471,12527,13203,13302,14525T,16024,16052,16058,16115,16318

T1b1b1a3a2_KF163065:169,190,498,3071,4363,7544,8712,12470,12471,12527,13203,13302,14053,14525T,16024,16052,16058,16115,16318

T1b1b1a3a3:169,190,498,3272,7544,8712,12470,12471,12527,13203,13302,14525T,16024,16052,16058,16115,16257,16318

T1b1b1a3a3a:169,190,498,3272,7544,8712,12363,12470,12471,12527,13203,13302,14525T,16024,16052,16058,16115,16257,16318

T1b1b1a3a3a_KF163068:169,190,498,3272,7544,8712,12363,12470,12471,12527,13203,13302,14525T,16024,16052,16058,16115,16257,16318

T1b1b1a3a3a_KF163078:169,190,498,3272,7544,8712,12363,12470,12471,12527,13203,13302,14525T,16024,16052,16058,16115,16123,16124,16257,16318

T1b1b1a3a3a_KF163071:169,190,498,3272,7544,8712,12363,12470,12527,13203,13302,14525T,16024,16052,16058,16115,16257,16318

T1b1b1a3a3_KF163090:169,190,498,3272,3483,7544,8712,12470,12471,12527,13203,13302,14525T,16024,16052,16058,16115,16257,16318

T1b1b1a3a_KF163067:169,173,190,498,7544,8712,12470,12471,12527,13203,13302,14525T,16024,16052,16058,16115,16257,16318

T1b1b1a3a_KF163070_KF163091:169,190,498,7544,7624,8712,12470,12471,12527,13203,13262,13302,14525T,16024,16052,16058,16115,16257,16318

T1b1b1a3a_KF163085:169,190,250,498,7544,8712,12470,12471,12527,13007,13203,13302,14525T,15953,15955G,16024,16052,16058,16115,16257,16318

T1b1b1a3_KF163086:169,190,498,7544,8712,12527,13203,13302,14525T,16024,16052,16058,16115,16124,16257,16318

T1b1b1a3_KF163079:169,173,190,498,7544,8712,12527,13203,13302,14525T,16024,16052,16058,16115,16257,16318

T1b1b1a3_KF163064:169,190,498,3071,7544,8712,12527,13203,13302,14525T,16024,16052,16058,16059,16115,16318

T1b1b1a3_KF163077:169,190,498,1075C,6393,7544,8712,9097,12527,13203,13302,14525T,16024,16052,16058,16115,16123,16257,16318

T1b1b1a3_KF163083:169,190,498,3272,7516,8712,12363,12527,13203,13302,14525T,16024,16052,16058,16115,16257,16318

T1b1b1c:169,7544,13203,14525T,16024,16052,16115,16150,16257

T1b1b1c1:169,7544,12470,12471,13203,14525T,16024,16052,16115,16150,16257

T1b1b1c1_KF163073:169,7544,12470,12471,13203,14525T,15158,16024,16052,16115,16257

T1b1b_JN817302:169,7544,9102,14525T,14533,14992,15846,16024,16115

T1b1c:169,7544,16024,16115,16257

T1b1c_JN817324:169,7544,8350,12229,16024,16115,16124,16257

T1b1c_DQ124399:169,3967,7544,14749A,16024,16115,16250,16257

T1b1d:169,6052,7544,14350,16024,16052,16115,16257

T1b1d_KF163061_KF163076:169,6052,7544,12470,12471,14350,16024,16052,16115,16257

T1b1_EU177842:169,7544,15421,16024,16052,16115,16257

T1b1_KF163063:169,7544,16024,16052,16115,16257,16303

T1b1_JN817334:169,4947,7544,10579,11532,13373A,15848,16024,16052,16115,16257,16303

T1b1_JN817320:3210,7544,15149,15328,16024,16052,16115,16257

T1b1_KF163084:169,7544,13397,16024,16052,16059,16070,16115,16257

T1b1_KF163092:169,7544,9503,13262,16024,16052,16115,16257

T1b_JN817327:106,169,7544,7818,14053,15948,16115

T1b_JN817348:169,4141,5898,6705,7544,9136,9658,9731,10605,13295,13984,16052,16115,16249,16257

T1c:169,16052,16115,16124,16257

T1c1:169,16052,16115,16124,16198,16257

T1c1a:169,16052,16055,16115,16124,16198,16257

T1c1a1:169,1326,11544,16052,16055,16115,16124,16141,16198,16257

T1c1a1_JN817309:169,1326,11544,13884,16052,16055,16115,16124,16141,16198,16257

T1c1a1_JN817310:169,1326,11544,16052,16055,16115,16124,16141,16198,16257,16318

T1c1a1_JN817322:106,169,1326,1876,7657,9070,11544,16018,16052,16055,16115,16124,16141,16198,16257

T1c1a1_JN817311:169,1326,10895A,11544,15464,16052,16055,16115,16124,16141,16198,16257

T1c1a_JN817346:169,10530,10797,14459,14821,16052,16055,16115,16124,16135,16198,16257

T1c1b:169,15966,16052,16115,16124,16198,16257

T1c1b_EU177847:169,4977,6934,8893,12254,15966,16052,16115,16124,16198,16257

T1c1b_JN817300_JN817301:169,3877,11475,15966,16052,16081,16115,16124,16129,16198,16202,16257

T1c1_JN817323:169,353G,1881,4349,4882,6019,8328,9722,12407,12730,15421,15958,16052,16115,16124,16198,16257

T1c1_EU177848:169,3877,8525,8864,10327,11454,12653,16052,16115,16198,16257

T1c1_JN817307:24,169,3406,5032,9677,14386,14600G,15328,15811,16052,16115,16124,16198,16257

T1c_JN817319:169,10246A,16115,16124,16257

T1c_JN817328:169,5188,8796A,12362,13733,16052,16057,16115,16124,16257

T1c_JN817325:169,10498,10926,11791,16052,16115,16124,16137,16208,16257

T1c_JN817326:169,2192,8334,11963,15100,15158,16052,16114,16115,16124,16257

T1c_KF163072:169,4588,6027,9908,10909,15462,16052,16115,16123,16124,16257

T1d:169,6237,16052,16115,16257

T1d1:169,4858,6237,16052,16115,16257

T1d1_JN817330:169,4858,6237,11457,16052,16115,16257

T1d1_JN817321:8,169,4858,6237,9514,10516,16052,16115,16250,16257

T1d1_JN817304:169,4858,6237,6333,7762,15149,16052,16115,16257

T1d_KF163081:169,6237,10349,10659,12470,16052,16115,16123,16257

T1d_JN817298:106,169,1593,3191,6237,16052,16115,16257

T1d_JN817299:169,250,4372,6237,6453,7400,11070,16052,16115,16257

T1e:8,169,16115,16257

T1e_JN817306:8,169,15941,16115,16250,16257

T1e_EU177841:8,169,9562,15398,16115,16257

T1e_KC153973_KC153975:8,169,10599,11478,12986,13859,14327,15997+XCATTAATGTAATAAAGACATAA,16076

T1f:169,12494,16052,16257

T1f_JN817343:169,3363,9638,11850,12494,14212,14611,15412,16052,16257

T1f_JN817329:169,7757,9385,12494,14045,15092,15406,16052,16257

T1g:169,300,3684,9203,12470,12471,16052,16115,16257,16262

T1g_KF163082:169,300,3684,9203,12470,12471,16052,16115,16250,16257,16262

T1g_KF163093:169,300,3684,9203,12470,12471,16052,16115,16249,16257,16262

T1c1a1_AY526085:3561,3562G,4499G,4539,5720,5756,5855,7996,11235,12160,15512,16044,16095

T2:169,6204,6744,10710,16059C,16187,16257

T2a:169,1501,5240,6204,6744,8804,9385,10710,11289,16059C,16187,16257

T2a_DQ124393:169,1501,4358,5240,6204,6744,8804,9385,10710,11289,16059C,16187,16257

T2a_DQ124396:169,1501,5240,6744,8804,9385,10710,11289,15597,16059C,16187,16257

T2b:169,5148,6204,6744,10710,16059C,16187,16257

T2b_EU177855:169,5148,6204,6744,10710,16059C,16130,16187,16257

T2b_EU177856:169,173,5148,6744,8247,10261,10710,16051,16059C,16115,16187,16257

T2b_EU177857:169,1325,1696,2601,3720,4783,5148,6204,6744,7400,10603,10710,15159,16059C,16060,16187,16250,16257,16303

T2c:169,2087,3192,6204,6744,9989,10710,11736,14581,16059C,16187,16257

T2c_EU177860:169,2087,3192,6204,6744,9482,9989,10710,11736,13398,14581,15196,16059C,16187,16257

T2c_EU177861:2087,2749,3192,6204,6316,6744,8986G,9989,10710,11736,14581,16059C,16187,16257

T2d:169,6204,6744,10710,16059C,16076,16187,16257

T2d_EU177851:169,1461,2560,6204,6744,10710,11103,12653,15798,15987,16059C,16076,16187,16257

T2d_EU177852:169,6204,6744,10710,16059C,16069,16076,16123,16187,16257

T2_EU177853:169,170,3942,5158T,5763,6204,6744,10352,10710,13753,16052,16059C,16143,16187,16257

T2_EU177854:169,1653,6204,6744,9502,10710,14404,16059C,16064,16187,16249,16257,16303

T2_DQ124383:169,174,6204,6744,8516,10710,11370,15558,16010C,16059C,16124,16187,16257

T2_HQ025805:169,3561,3562G,4108,4327,5855,6204,6744,7960,10710,12272,14418,16059C,16187,16257

T2_EU177849:169,4955,5634C,6204,6744,10710,16059C,16133,16187,16257

T2_EU177850:169,6204,6744,7852,10710,13203,16059C,16060,16127,16187,16257

T2_EU177858:119,120,166,169,3351,3852,6204,6405,6744,8477,9310,10662,10710,11832,12294,12506,12767,13482,15885d,16059C,16082,16187,16233,16257

T2_EU177859:169,1555,4977,6204,6207,6744,8862,10710,16059C,16141,16187,16257

T2_AY676856:169,5603,6204,6744,10710,12545,15419,16059C,16079,16187,16257,16262

T3:169

T3a:169,12160

T3a1:169,12160,16044,16095

T3a1a:169,12160,15512,16044,16095

T3a1a_DQ124418:12160,15512,16044,16095

T3a1_DQ124417:9780T,12160,16044,16095

T3a2:169,8712,9482,12160,16051,16121

T3b:169,14065

T3b_DQ124385:169,3336,3852A,4771,11736,12225,14065,16129

T3b_DQ124416:169,6575,6994C,14065

T3c:169,12730

T3c_DQ124409:169,1483,8171,9022,10387,12730

T3c_DQ124410:8168A,10890,12730,14134,16250

T3d:169,13901

T3d1:169,5158,13691,13901

T3d1_AY676861:169,5158,13691,13901,16143

T3d1_AY676871:169,1134,5158,13691,13901,16304

T3d_GU947019:166,169,537,567,8927,13901

T3e:169,10884

T3e_AF492351:169,190,10884,12236,15581,16071,16143

T3e_KC153974:169,5226,8832,10884,11514,13487,16140

T3f:169,15987d,16059+A,16074A,16078

T3f1:169,9731,15987d,16059+A,16076A,16078

T3f1a:169,4333,9731,15987d,16059+A,16076A,16078

T3f1a_GU947008:169,4333,5044,9731,15987d,16059+A,16076A,16078

T3f1a_GU947014:169,4333,5226,9731,13784T,15987d,16059+A,16076A,16078

T3f1_GU947012_GU947015:169,6327,9731,12629,13280,15987d,16059+A,16076A,16078,16233

T3f2:169,190,1865,2632T,3810,5044,7241,9064,9238,15987d,16059+A,16070,16076A,16078

T3f2_GU947007:169,190,1865,2632T,3810,5044,7241,9064,9238,14218,14233,15987d,16059+A,16070,16076A,16078

T3f3:169,10462,15987d,16059+A,16076A,16078,16149

T3f3_GU947010:169,532+G,10462,15987d,16059+A,16076A,16078,16149

T3f3_GU947018:169,10462,14218,14233,15987d,16059+A,16076A,16078,16149

T3f4:5128,15196,15987d,16059+A,16076A,16078

T3g:169,5718,13520

T3g_GU947021:169,2055,4756C,4757C,5604,5718,7011,12025,12178,13091A,13520,16138,16225,16228C

T3g_AY676864:169,173,5146,5718,7125,8407,10602,12038,13520,15936

T3h:169,10743

T3h_DQ124390:169,816,2147,3304,3873,8927,10743,13403

T3h_DQ124411:106,3546,10743

T3i:169,1066+A,9449,14080,14563,15217

T3i_EU177822:169,1066+A,9449,14080,14524,14563,15217,15421

T3i_EU177823:169,1066+A,2222,9449,14080,14230,14563,15104,15217,15388,16024,16202

T3j:169,13526

T3j_EU177829:169,1292,6501,6948,7801,13526,16135,16252

T3j_EU177830:169,313A,8882,10872,13526,16044

T3k:169,12910

T3k1:169,10891,12910,15912,16057

T3k1_EU177837:169,3088,3798,9098,10891,12910,15912,16057

T3k1_DQ124376:169,1134,2571,6621,9731,10891,12910,15912,16057,16114,16169

T3k_DQ124386:106,169,3729,12910,14582,16044,16095

T3l:169,3343,7624,12025,12175,14038,15966

T3l_EU177834:169,3343,7624,12025,12175,14038,15966

T3n:169,16121

T3n1:169,5946,10085,13431,16121,16230

T3n1_DQ124391:169,5946,7781,10085,13431,16121,16230

T3f1a_DQ124397:169,3858,5946,7781,10085,13431,16121,16230

T3n_DQ124373:169,16115,16121

T3n_AY676857:169,190,4393A,5850,6636A,16055,16121

T3o:169,16124

T3o1:169,7778,11037,12686,16124

T3o1a:8,169,7778,11037,12686,16053,16124

T3o1a_DQ124388_DQ124394_DQ124398:8,169,7778,11037,12686,16053,16124,16249

T3o1_DQ124380:169,7778,8288,11037,12686,13630,16124

T3o1_DQ124382:169,7778,8668G,11037,12160,12686,13031,16124

T3o1_DQ124402:169,5683,7375G,7778,11037,12686,15760,16124

T3o_DQ124374:8,169,1195,3720,5222A,10993,15331,16069,16124

T3o_AY676873:169,8918,12803,16024,16124,16137

T3p:169,16059

T3p_EU177825:169,13496,16059

T3p_EU177824:169,5104,14596,16059,16060,16115

T3q:169,16060

T3q1:169,16060,16233

T3q1_EU177826:169,2080,2702+C,6135,8654,9517,12211,14466,16060,16231,16233

T3q1_EU177827:11674,12161,13697,13745,16060,16076,16142,16233,16252

T3q2:169,16060,16121

T3q2_DQ124384:169,9686,16060,16121

T3q2_EU177828:106,169,1554,4339,4387,5143,14293,16060,16121,16129,16249

T3r1:4903,16044

T3r1_AY676868_AY676869:173,811,3877,4903,16044,16233,16234

T3r1_AY676872:4903,10578,12060,16044

T3r2:1092

T3r2_EU177817:119,1092,4096,9070,16143

T3r2_EU177818:1092,5544+A,6570,11751,13166,13403,13832,13856,15838,16044

T3r2_AY676858:1092,3315,9731,12023,13772,15275,16064

T3r3:12707,16044

T3r3_EU177819:3729,12707,16044,16304

T3r3_EU177820:173,2029,4002,12707,15356,15962,16044,16124,16230

T3r4:12732

T3r4_DQ124408:7213,12732,15558

T3r4_DQ124413_AY676866:3345,8712,12167,12732,14908,16249

T3r4_EU177821:169,11250,12732,12881

T3r5:173

T3r5_DQ124415:173,1493,6932,9731,13175,16303

T3o1a_DQ124407:173,1493,6932,9731,13175,15597,16303

T3r5_AY676865:173,5140,7137,8384,13556,15963,16059

T3r6:353G

T3r6_GQ129208:353G,9136,12160

T3r6_KC153977:353G,10686,14418,16064,16249

T3r_AY676862:222,9076,9118C,10349,15742,16002,16143

T3r_GQ129207:166,4744,16133

T3r_HM045018:3189

T3r_AY676863:641,1483,11085,13312

T3r_EU177815:3304,6822,9259,13142,15879

T3r_EU177816:163,565,10698,11028,12452,16024

T3s:8,169

T3s_EU177836:8,169,514,4822,15094,16068,16111

T3s_JQ967333:8,169,13257,15989+GGACATAACATTAATGTAATAAAGACATAACATTAATGTAATAA,16078,16137

T3_EU177832:169,16135

T3_KF926377:169,4060,8063A,8065A,8066A,8069,13734,16053,16063,16131

T3_AY676859:169,687,14564,16114

T3_KC153972:169,177,16051,16052

T3_DQ124406:169,2127,3786,16250

T3_DQ124414:169,5473C,12236,15209,16071

T3_DQ124404:35,169,7573,7612,9178,14128,16112

T3_DQ124405:169,737,1522,5074,6082,11531C,15941

T3_AY676860:169,5047,7630,7920,9971,10462,15157,15494,15963,16149

T3_AY676855:169,6732,14849,15581,16207

T3_KC153976:169,236,3282,6336,9604

T3_DQ124387_DQ124395:169,3381,12175-12177d,14747,16088

T3_EU177831:169,1292T,6019,8900,13472,14180

T3_EU177835:169,13880,14155

T3_EU177838:169,532+G,1862+A,6690,7194,14368,16303

T3_EU177839:169,8242,11187,11802,16110,16304

T4:169,12160,15512,16044,16095,16304

T4a:169,11176,12160,15512,16044,16095,16304

T4a_DQ124372:169,6690,7705,8212,11176,12160,12494,14909C,15512,16044,16095,16304

T4a_DQ124375:169,9121,11176,12160,13233,15512,16044,16076,16095,16304

T4_DQ124377:169,7867,12160,15512,16044,16051,16059,16095,16106,16166,16304

T4_DQ124392:169,11010,12160,15512,16044,16095,16304

T4_DQ124400:169,5378,10469,12160,13511,14059,15512,16044,16095,16304

T4_DQ124401:169,353G,5599,12160,15512,16044,16095,16304

T4_DQ124412:169,12160,15512,16044,16095,16118,16304

T5:163,169,4254,12983,13007,16257

T5a:163,4254,8026,9091,9234,10881,12983,13007,15884,16199,16257

T5a_EU177863:163,4254,8026,9091,9234,10881,12782,12983,13007,15884,16199,16257

T5b:163,169,4254,9896A,12983,13007,16257

T5b_EU177864:163,169,4254,7660,9896A,12983,13007,16257

T5b_EU177865:163,169,204,1292T,4254,8047,9896A,12983,13007,16139,16257

T6:106,169,7933,11901,12925T,13376,14065,16111,16257

T7:169,3711T,8640,9305,16076,16257

I:8,106,166,206,233,234,249,250,297,301,519,739,763,818,1160,1476,1494,1679,1826,1862A,1871,2018,2101,2560,2955,2979,
2981,2990,2991,2992,3053,3073,3138,3147,3243,3327,3337,3381,3441,3537,3552,3602,3795,3831,3876,3933,3978,3987,4002,
4330,4444,4564,4732,4735,4771,4873,4939,5287,5503,5533,5616,5745,5785,5892,5919,6000,6117T,6237,6342,6369,6381,6438,
6462,6729,6774,6883,6924,7306,7332,7358,7360,7363,7516,7832,7853,8047,8170,8190,8212,8287,8310,8372C,8468,8496,
8505,8573,8751,8986,9007,9040,9070,9247,9583,9604,9769,9893,9930C,9980,10041,10068,10139,10155,10270,10324,10333,
10447,10592,10623,10693G,10851,11002,11037,11070,11136,11202,11268,11331,11409,11421,11805,11844,12137,12180,
12236,12379,12435,12470,12471,12515,12624,12674,12677,12686,12752,12803,12902,12925T,12926,13007,13058A,13100,
13106,13277,13382,13435,13439,13556,13566,13586,13679,13691,13694,13884,13908A,13911,14038,14068,14122,14131,
14140,14257,14317,14413,14418,14505,14608,14827,14860,14899,15107,15136,15148,15289,15310,15328,15565,15581,
15595,15607,15619,15629,15631,15743,15753,15820,15953,15955G,15961,15996,16024,16051,16059,16060,16076,16084,
16104,16111,16115,16118,16119,16121,16123,16124,16132,16139,16140,16145d,16149,16198,16231,16249,16250,16302,
16303

I1:8,106,166,173,206,233,234,249,250,297,301,519,724,739,763,818,1160,1476,1494,1679,1826,1862A,1871,2018,2101,2560,
2577,2636,2955,2979,2981,2990,2991,2992,3053,3073,3138,3147,3243,3327,3337,3381,3441,3537,3552,3602,3795,3831,3876,
3933,3978,3987,4002,4330,4444,4564,4732,4735,4771,4873,4939,4990A,5287,5503,5533,5616,5745,5785,5892,5919,6000,
6117T,6237,6342,6369,6381,6438,6462,6729,6774,6883,6924,7306,7332,7358,7360,7363,7516,7832,7853,8047,8170,8190,8196,
8212,8287,8310,8372C,8468,8496,8505,8516,8573,8751,8986,9007,9040,9070,9247,9304,9482,9583,9604,9769,9893,9932C,
9980,10041,10068,10073,10139,10155,10270,10324,10333,10447,10592,10623,10693G,10851,11002,11037,11070,11136,
11202,11268,11331,11409,11421,11805,11844,12137,12180,12236,12379,12435,12470,12471,12515,12624,12674,12677,
12686,12752,12803,12902,12925T,12926,13007,13058A,13100,13103A,13106,13277,13373,13382,13435,13439,13556,13566,
13586,13630,13679,13691,13694,13884,13912A,13911,14038,14068,14122,14131,14140,14257,14317,14413,14418,14505,
14608,14827,14860,14899,15107,15136,15148,15289,15310,15328,15565,15581,15595,15607,15619,15629,15631,15743,
15753,15820,15953,15955G,15961,15996,16024,16051,16059,16060,16076,16084,16104,16111,16115,16118,16119,16121,
16123,16124,16132,16139,16140,16145d,16149,16198,16231,16249,16250,16302,16303

I1_AY126697:8,106,166,173,206,233,234,249,250,297,301,519,724,739,763,818,1160,1476,1494,1679,1826,1862A,1871,2018,
2101,2560,2577,2636,2955,2979,2981,2990,2991,2992,3053,3073,3138,3147,3243,3327,3337,3357,3381,3441,3537,3552,3602,
3795,3831,3876,3933,3978,3987,4002,4330,4444,4564,4732,4735,4771,4873,4939,4990A,5287,5503,5533,5616,5745,5785,
5892,5919,6000,6117T,6237,6342,6369,6381,6438,6462,6729,6774,6883,6924,7306,7332,7358,7360,7363,7516,7832,7853,7894,
8047,8170,8190,8196,8212,8287,8310,8372C,8468,8496,8505,8516,8573,8751,8986,9007,9040,9070,9247,9304,9482,9583,
9604,9769,9893,9932C,9980,10041,10068,10073,10139,10155,10270,10324,10333,10447,10592,10623,10693G,10851,11002,
11037,11070,11136,11202,11268,11331,11409,11421,11805,11844,12137,12180,12236,12379,12435,12470,12471,12515,
12624, 12674,12677,12686,12752,12803,12902,12925T,12926,13007,13058A,13100,13103A,13106,13277,13373,13382,13435,
13439,13556,13566,13586,13630,13679,13691,13694,13884,13912A,13911,14038,14068,14122,14131,14140,14257,14317,
14413,14418,14505,14608,14827,14860,14899,15107,15136,15148,15289,15310,15328,15565,15581,15589,15595,15607,
15619,15629,15631,15743,15753,15820,15953,15955G,15961,15996,16024,16051,16059,16060,16076,16084,16104,16111,
16115,16118,16119,16121,16123,16124,16132,16139,16140,16145d,16149,16198,16231,16249,16250,16302,16303

I1_EU177868:8,106,166,173,206,233,234,249,250,297,301,519,724,739,763,818,1160,1476,1494,1679,1826,1862A,1871,2018,
2101,2323A,2560,2577,2636,2955,2979,2981,2990,2991,2992,3053,3073,3138,3147,3243,3327,3337,3381,3441,3537,3552,
3602,3795,3831,3876,3933,3978,3987,4002,4330,4444,4564,4732,4735,4771,4873,4939,4990A,5287,5503,5533,5616,5745,
5785,5892,5919,6000,6117T,6237,6342,6369,6381,6438,6462,6729,6774,6883,6924,7306,7332,7358,7360,7363,7516,7832,7853,
8047,8170,8190,8196,8212,8287,8310,8372C,8468,8496,8505,8516,8573,8751,8986,9007,9040,9070,9247,9304,9482,9583,
9604,9769,9893,9932C,9980,10041,10068,10073,10139,10155,10270,10324,10333,10447,10592,10623,10693G,10851,11002,
11037,11070,11136,11202,11268,11331,11409,11421,11805,11844,12137,12180,12236,12379,12435,12470,12471,12515,
12624,12674,12677,12686,12752,12803,12902,12925T,12926,13007,13058A,13100,13103A,13106,13277,13373,13382,13435,
13439,13556,13566,13586,13630,13679,13691,13694,13884,13912A,13911,14038,14068,14122,14131,14140,14257,14317,
14413,14418,14505,14608,14827,14860,14899,15107,15136,15148,15289,15310,15328,15565,15581,15595,15607,15619,
15629,15631,15743,15753,15820,15953,15955G,15961,15996,16024,16051,16059,16060,16076,16084,16104,16111,16115,
16118.16119.16121.16123.16124.16132.16139.16140.16145d.16149.16198.16231.16249.16250.16302.16303

I1_FJ971088:8,106,166,173,206,233,234,249,250,297,301,519,724,739,763,818,1160,1476,1494,1679,1826,1862A,1871,2018,
2101,2560,2577,2636,2955,2979,2981,2990,2991,2992,3053,3073,3138,3147,3243,3312,3327,3337,3381,3441,3537,3552,3602,
3795,3831,3876,3933,3978,3987,4002,4330,4444,4564,4732,4735,4771,4873,4939,4990A,5287,5503,5533,5616,5745,5785,
5892,5919,6000,6117T,6237,6342,6369,6381,6438,6462,6729,6774,6883,6924,7306,7332,7358,7360,7363,7516,7832,7853,8047,
8170,8190,8196,8212,8287,8310,8372C,8468,8496,8505,8516,8573,8751,8986,9007,9040,9070,9247,9304,9482,9583,9604,
9769,9893,9932C,9980,10041,10068,10073,10139,10155,10270,10324,10333,10447,10592,10623,10693G,10851,11002,11037,
11070,11136,11202,11268,11331,11409,11421,11805,11844,12137,12139,12180,12236,12379,12435,12470,12471,12515,
12624,12674,12677,12686,12752,12803,12902,12925T,12926,13007,13058A,13100,13103A,13106,13277,13373,13382,13435,
13439,13556,13566,13586,13630,13679,13691,13694,13884,13912A,13911,14038,14068,14122,14131,14140,14257,14317,
14413,14418,14505,14608,14827,14860,14899,15107,15136,15148,15159,15289,15310,15328,15565,15581,15595,15607,
15619,15629,15631,15743,15753,15820,15953,15955G,15961,15996,16024,16051,16059,16060,16076,16084,16104,16111,
16115,16118,16119,16121,16123,16124,16132,16139,16140,16145d,16149,16198,16231,16249,16250,16302,16303

I2:8,39,106,166,206,233,234,249,250,297,301,519,739,763,818,1160,1476,1494,1679,1826,1862A,1871,2018,2101,2560,2955,
2979,2981,2990,2991,2992,3053,3073,3138,3147,3243,3318,3327,3337,3381,3441,3537,3552,3602,3795,3831,3876,3933,3978,
3987,4002,4330,4444,4564,4732,4735,4771,4873,4939,5287,5503,5533,5616,5745,5785,5892,5919,6000,6117T,6237,6342,6369,
6381,6438,6462,6729,6774,6883,6924,7306,7332,7358,7360,7363,7516,7832,7853,8047,8170,8190,8212,8287,8310,8372C,
8468,8496,8505,8573,8751,8986,9007,9040,9070,9247,9583,9604,9769,9893,9932C,9980,10041,10068,10139,10155,10270,
10324,10333,10447,10592,10623,10693G,10851,10890,10935,11002,11037,11070,11136,11202,11268,11331,11409,11421,
11805,11844,12137,12180,12236,12379,12435,12470,12471,12515,12624,12674,12677,12686,12752,12803,12902,12925T,
12926,13007,13058A,13100,13106,13277,13382,13435,13439,13556,13566,13586,13679,13691,13694,13884,13912A,13911,
14038,14068,14122,14131,14140,14257,14317,14371,14413,14418,14505,14608,14827,14860,14899,15107,15136,15148,
15289,15310,15328,15565,15581,15595,15607,15619,15629,15631,15743,15753,15820,15953,15955G,15961,15996,16024,
16051,16059,16060,16076,16084,16104,16111,16115,16118,16119,16121,16123,16124,16132,16139,16140,16143,16145d,
16149,16198,16231,16234,16249,16250,16302,16303

I2_AF492350:8,39,106,166,206,233,234,249,250,297,301,519,739,763,818,1160,1476,1494,1679,1826,1862A,1871,2018,2101,
2560,2955,2979,2981,2990,2991,2992,3053,3073,3138,3147,3243,3318,3327,3337,3381,3441,3537,3552,3602,3795,3831,3876,
3933,3978,3987,4002,4330,4444,4564,4732,4735,4771,4873,4939,5287,5503,5533,5616,5745,5785,5892,5919,6000,6117T,6237,
6342,6369,6381,6438,6462,6729,6774,6883,6924,7306,7332,7358,7360,7363,7516,7832,7853,7948,8047,8170,8190,8212,8287,
8310,8372C,8468,8496,8505,8573,8751,8986,9007,9040,9070,9247,9583,9604,9769,9893,9932C,9980,10041,10068,10139,
10155,10270,10324,10333,10447,10592,10623,10693G,10851,10890,10935,11002,11037,11070,11136,11202,11268,11331,
11409,11421,11805,11844,12137,12180,12236,12379,12435,12470,12471,12515,12624,12674,12677,12686,12752,12803,
12902,12925T,12926,13007,13058A,13100,13106,13277,13382,13435,13439,13556,13566,13586,13679,13691,13694,13884,
13912A,13911,14038,14068,14122,14131,14140,14257,14317,14371,14413,14418,14505,14608,14827,14860,14899,15107,
15136,15148,15289,15310,15328,15565,15581,15595,15607,15619,15629,15631,15743,15753,15820,15953,15955G,15961,
15996,16024,16051,16059,16060,16076,16084,16104,16111,16115,16118,16119,16121,16123,16124,16132,16139,16140,
16143.16145d.16149.16198.16231.16234.16249.16250.16302.16303

I2_EU177870:8,39,106,166,206,233,234,249,250,297,301,519,739,763,818,1160,1476,1494,1679,1826,1862A,1871,2018,2101,
2560,2955,2979,2981,2990,2991,2992,3053,3073,3138,3147,3243,3318,3327,3337,3381,3441,3537,3552,3602,3795,3831,3876,
3933,3978,3987,4002,4312,4330,4444,4564,4732,4735,4771,4873,4939,5287,5503,5533,5616,5745,5785,5892,5919,6000,6117T,
6237,6342,6369,6381,6438,6462,6729,6774,6883,6924,7306,7332,7358,7360,7363,7501,7516,7832,7853,8047,8170,8190,8212,
8287,8310,8372C,8468,8496,8505,8573,8751,8986,9007,9040,9070,9247,9583,9604,9769,9893,9932C,9980,10041,10068,10139,
10155,10270,10324,10333,10447,10592,10623,10690,10693G,10851,10890,10935,11002,11037,11070,11136,11202,11268,
11331,11409,11421,11805,11844,12137,12180,12236,12341,12379,12435,12470,12471,12515,12624,12674,12677,12686,
12752,12803,12902,12925T,12926,13007,13058A,13100,13106,13277,13382,13435,13439,13556,13566,13586,13679,13691,
13694,13884,13912A,13911,14038,14068,14122,14131,14140,14257,14317,14371,14413,14418,14505,14608,14827,14860,
14899,15107,15136,15148,15289,15310,15328,15565,15581,15595,15607,15619,15629,15631,15743,15753,15820,15953,
15955G,15961,15996,16024,16051,16059,16060,16076,16084,16104,16111,16115,16118,16119,16121,16123,16124,16132,
16139,16140,16143,16145d,16149,16198,16231,16234,16249,16250,16302,16303

I2_EU177869:8,39,106,166,173,206,233,234,249,250,297,301,519,739,763,818,1160,1476,1494,1679,1826,1862A,1871,2018,210
1,2560,2955,2979,2981,2990,2991,2992,3053,3073,3138,3147,3243,3318,3327,3337,3381,3441,3537,3552,3602,3795,3831,387
6,3933,3978,3987,4002,4330,4444,4564,4732,4735,4771,4873,4939,5275T,5287,5503,5533,5616,5745,5785,5892,5919,6000,611
7T,6237,6342,6369,6381,6438,6462,6498,6729,6774,6883,6924,7306,7332,7358,7360,7363,7516,7832,7853,8047,8170,8190,821
2,8287,8310,8372C,8468,8496,8505,8573,8751,8986,9007,9040,9070,9247,9583,9604,9769,9893,9932C,9980,10041,10068,1013
9,10155,10270,10324,10333,10447,10592,10623,10693G,10851,10890,10935,11002,11037,11070,11136,11202,11268,11331,11
409,11421,11805,11844,12137,12180,12236,12379,12435,12470,12471,12515,12624,12674,12677,12686,12752,12803,12902,1
2925T,12926,13007,13058A,13100,13106,13277,13382,13435,13439,13556,13566,13586,13679,13691,13694,13884,13912A,139
11,14038,14068,14122,14131,14140,14234,14257,14317,14371,14413,14418,14505,14608,14827,14833,14860,14899,15107,15
136,15148,15289,15310,15328,15565,15581,15595,15607,15619,15629,15631,15743,15753,15820,15953,15955G,15961,15996,
16024,16051,16052A,16059,16060,16076,16084,16104,16111,16115,16118,16119,16121,16123,16124,16132,16139,16140,1614
3,16145d,16149,16198,16231,16234,16249,16250,16302,16303

Table S2: List of variants from animal sequences available at NCBI, aligned to ARS-UCD1.2_M Ref and used by MitoToolsPy to determine haplogroup.

cattleRef_V00654:222d,364C,589d,2538C,3345,3387,3541,4321,8190,8712,9684G,12167,13312A,15637

AY126697:8,106,166,173,206,233,234,249,250,297,301,364+CC,364C,519,724,739,763,818,1160,1476,1494,1602d,1679,1826,
1862A,1871,2018,2101,2560,2577,2636,2955,2979,2981,2990,2991,2992,3053,3073,3138,3147,3243,3327,3337,3345,3357,
3381 ,3387,3441,3537,3541,3552,3602,3795,3831,3876,3933,3978,3987,4002,4321,4330,4444,4564,4732,4735,4771,4873,4939,
4990A,5287,5503,5533,5616,5745,5785,5892,5919,6000,6117T,6237,6342,6369,6381,6438,6462,6729,6774,6883,6924,7306,
7332,7358,7360,7363,7516,7832,7853,7894,8047,8170,8196,8212,8287,8310,8372C,8468,8496,8505,8516,8573,8712,8751,
8986,9007 ,9040,9070,9247,9304,9482,9583,9604,9769,9893,9932C,9980,10041,10068,10073,10139,10155,10270,10324,10333,
10447,10592,10623,10693G,10851,11002,11037,11070,11136,11202,11268,11331,11409,11421,11805,11844,12137,12167,
12180,12236,12379,12435,12470,12471,12515,12624,12674,12677,12686,12752,12803,12902,12925T,12926,13007,13058A,
13100,13103A,13106,13277,13373,13382,13435,13439,13556,13566,13586,13630,13679,13691,13694,13884,13910A,13911,
14038,14068,14122,14131,14140,14257,14317,14413,14418,14505,14608,14827,14860,14899,15107,15136,15148,15289,
15310,15328,15565 ,15581,15589,15595,15607,15619,15629,15631,15637,15743,15753,15820,15953,15955G,15961,15996,
16024,16051,16059,16060,16076,16084,16086,16104,16111,16115,16118,16119,16121,16123,16124,16132,16139,16140,
16145d,16149,16198, 16202+A,16231,16249,16250,16302,16303

AY526085:222d,364C,589d,3345,3387,3541,3561,3562G,4321,4499G,4539,5720,5756,5855,7996,8190,8712,11235,12160,
12167, 15512,15637,16044,16095

AY676855:169,364C,1602d,3345,3387,3541,4321,6732,8190,8712,12167,14849,15581,15637,16207

AY676856:169,222d,364C,3345,3387,3541,4321,5603,6204,6744,8190,8712,10710,12167,12545,15419,15637,16059C,16079,
16187,16257,16262

AY676857:169,190,364C,364+C,1602d,3345,3387,3541,4321,4393A,5850,6636A,8190,8712,12167,15637,16055,16121

AY676858:222d,364C,1092,1602+AA,3315,3345,3387,3541,4321,8190,8712,9731,12023,12167,13772,15275,15637,16064

AY676859:169,687,3345,3387,3541,4321,8190,8712,12167,14564,15637,16114

AY676860:169,222+C,364C,1602d,3345,3387,3541,4321,5047,7630,7920,8190,8712,9971,10462,12167,15157,15494,15637,
15963,16149

AY676861:169,364+C,364C,3345,3387,3541,4321,5158,8190,8712,12167,13691,13901,15637,16143

AY676862:224d,3345,3387,3541,4321,8190,8712,9076,9118C,10349,12167,15637,15742,16002,16087,16143

AY676863:222d,641,1483,1602d,3345,3387,3541,4321,8190,8712,11085,12167,15637

AY676864:169,173,222d,364C,364+CC,1602d,3345,3387,3541,4321,5146,5718,7125,8190,8407,8712,10602,12038,12167,13520,
15637,15936

AY676865:173,222d,365d,1602d,3345,3387,3541,4321,5140,7137,8190,8384,8712,12167,13556,15637,15963,16059

AY676868:173,222d,811,3345,3387,3541,3877,4321,4903,8190,8712,12167,15637,16044,16233,16234

AY676871:169,364C,1134,3345,3387,3541,4321,5158,8190,8712,12167,13691,13901,15637,16304

AY676872:222+C,364C,3345,3387,3541,4321,4903,8190,8712,10578,12060,12167,15637,16044

AY676873:169,365d,3345,3387,3541,4321,8190,8712,8918,12167,12803,15637,16024,16124,16137

DQ124372:169,364C,3345,3387,3541,4321,6690,7705,8190,8212,8712,11176,12160,12167,12494,14909C,15512,15637,16044,
16095,16304

DQ124373:169,364C,3345,3387,3541,4321,8190,8712,12167,15637,16115,16121

DQ124374:8,169,364C,1195,3345,3387,3541,3720,4321,5222A,8190,8712,10993,12167,15331,15637,16069,16124

DQ124375:169,364C,3345,3387,3541,4321,8190,8712,9121,11176,12160,12167,13233,15512,15637,16044,16076,16095,16304

DQ124376:169,364C,1134,2571,3345,3387,3541,4321,6621,8190,8712,9731,10891,12167,12910,15637,15912,16057,16114,
16169

DQ124377:169,364C,3345,3387,3541,4321,7867,8190,8712,12160,12167,15512,15637,16044,16051,16059,16095,16106,16166,
16304

DQ124380:169,364C,3345,3387,3541,4321,7778,8190,8288,8712,11037,12167,12686,13630,15637,16124

DQ124382:169,364C,3345,3387,3541,4321,7778,8190,8668G,8712,11037,12160,12167,12686Y,13031,15637,16124

DQ124383:169,174,364C,3345,3387,3541,4321,6204,6744,8190,8516,8712,10710,11370,12167,15558,15637,16010C,16059C,
16124,16187,16257

DQ124384:169,364C,3345,3387,3541,4321,8190,8712,9686,12167,15637,16060,16121

DQ124385:169,364C,3336,3345,3387,3541,3852A,4321,4771,8190,8712,11736,12167,12225,14065,15637,16129

DQ124386:106,169,364C,3345,3387,3541,3729,4321,8190,8712,12167,12910,14582,15637,16044,16095

DQ124387:169,364C,3345,3381,3387,3541,4321,8190,8712,12167,12173-12175d,14747,15637,16088

DQ124388:8,169,364C,3345,3387,3541,4321,7778,8190,8712,11037,12167,12686,15637,16053,16124,16249

DQ124390:169,364C,816,2147,3304,3345,3387,3541,3873,4321,8190,8712,8927,10743,12167,13403,15637

DQ124391:169,364C,3345,3387,3541,4321,5946,7781,8190,8712,10085,12167,13431,15637,16121,16230

DQ124392:169,364C,3345,3387,3541,4321,8190,8712,11010,12160,12167,15512,15637,16044,16095,16304

DQ124393:169,364C,1501,3345,3387,3541,4321,4358,5240,6204,6744,8190,8712,8804,9385,10710,11289,12167,15637,16059C,16187,16257

DQ124396:169,364C,1501,3345,3387,3541,4321,5240,6744,8190,8712,8804,9385,10710,11289,12167,15597,15637,16059C,16187,16257

DQ124397:169,364C,3345,3387,3541,3858,4321,5946,7781,8190,8712,10085,12167,13431,15637,16121,16230

DQ124399:169,364C,3345,3387,3541,3967,4321,7544,8190,8712,12167,14749A,15637,16024,16115,16250,16257

DQ124400:169,364C,3345,3387,3541,4321,5378,8190,8712,10469,12160,12167,13511,14059,15512,15637,16044,16095,16304

DQ124401:169,353G,364C,3345,3387,3541,4321,5599,8190,8712,12160,12167,15512,15637,16044,16095,16304

DQ124402:169,364C,3345,3387,3541,4321,5683,7375G,7778,8190,8712,11037,12167,12686,12940Y,15637,15760,16124

DQ124404:35,169,364C,3345,3387,3541,4321,7573,7612,8190,8712,9178,12167,14128,15637,16112

DQ124405:169,364C,737,1522,3345,3387,3541,4321,5074,6082,8190,8712,11531C,12167,15637,15941

DQ124406:169,364C,2127,3345,3387,3541,3786,4321,8190,8712,12167,15637,16250

DQ124407:173,364C,1493,3345,3387,3541,4321,6932,8190,8712,9731,12167,13175,15597,15637,16303

DQ124408:364C,3345,3387,3541,4321,7213,8190,8712,12167,12732,15558,15637

DQ124409:169,364C,1483,3345,3387,3541,4321,8171,8190,8712,9022,10387,12167,12730,15637

DQ124410:364C,3345,3387,3541,4321,8168A,8190,8712,10890,12167,12730,14134,15637,16250

DQ124411:106,364C,3345,3387,3541,3546,4321,8190,8712,10743,12167,15637

DQ124412:169,364C,3345,3387,3541,4321,8190,8712,12160,12167,15512,15637,16044,16095,16118,16304

DQ124413:364C,3387,3541,4321,8190,12732,14908,16249

DQ124414:169,364C,3345,3387,3541,4321,5473C,8190,8712,12167,12236,15209,15637,16071

DQ124415:173,364C,1493,3345,3387,3541,4321,6932,8190,8712,9731,12167,13175,15637,16303

DQ124416:169,364C,3345,3387,3541,4321,6575,6996C,8190,8712,12167,14065,15637

DQ124417:364C,3345,3387,3541,4321,8190,8712,9780T,12160,12167,15637,16044,16095

DQ124418:364C,3345,3387,3541,4321,8190,8712,12160,12167,15512,15637,16044,16095

EU177815:363+C,3304,3345,3387,3541,4321,6822,8190,8712,9259,12167,13142,15637,15879

EU177816:163,363+C,565,3345,3387,3541,4321,8190,8712,10698,11028,12167,12452,15637,16024

EU177817:119,363+C,1092,3345,3387,3541,4096,4321,8190,8712,9070,12167,15637,16143

EU177818:222+C,363+C,1092,1602+A,3345,3387,3541,4321,5544+A,6570,8190,8712,11751,12167,13166,13403,13832,13856,15637,15838,16044

EU177819:364C,3345,3387,3541,3729,4321,8190,8712,12167,12707,15637,16044,16304

EU177820:173,222d,364C,2029,3345,3387,3541,4002,4321,8190,8712,12167,12707,15356,15637,15962,16044,16124,16230

EU177821:169,364C,1602d,3345,3387,3541,4321,8190,8712,11250,12167,12732,12881,15637

EU177822:169,222d,363+C,1068+A,1602d,3345,3387,3541,4321,8190,8712,9449,12167,14080,14524,14563,15217,15421,15637

EU177823:169,222d,363+C,1068+A,1602d,2222,3345,3387,3541,4321,8190,8712,9449,12167,14080,14230,14563,15104,15217,15388,15637,16024,16202

EU177824:169,363+C,1602d,3345,3387,3541,4321,5104,8190,8712,12167,14596,15637,16059,16060,16115

EU177825:169,364C,1602d,3345,3387,3541,4321,8190,8712,12167,13496,15637,16059

EU177826:169,364C,2080,2702+C,3345,3387,3541,4321,6135,8190,8654,8712,9517,12167,12211,14466,15637,16060,16231,16233

EU177827:363+C,3345,3387,3541,4321,8190,8712,11674,12161,12167,13697,13745,15637,16060,16076,16142,16233,16252

EU177828:106,169,222d,364C,1554,1602d,3345,3387,3541,4321,4339,4387,5143,8190,8712,12167,14293,15637,16060,16121,16129,16249

EU177829:169,363+C,1292,1602d,3345,3387,3541,4321,6501,6948,7801,8190,8712,12167,13526,15637,16135,16252

EU177830:169,222d,313A,364C,1602d,3345,3387,3541,4321,8190,8712,8882,10872,12167,13526,15637,16044

EU177831:169,222d,364C,1292T,1602d,3345,3387,3541,4321,6019,8190,8712,8900,12167,13472,14180,15637

EU177832:169,222d,364C,1602d,3345,3387,3541,4321,8190,8712,12167,15637,16135

EU177834:169,222d,364C,3343,3345,3387,3541,4321,7624,8190,8712,12025,12167,12175,14038,15637,15966,16167Y

EU177835:169,364C,3345,3387,3541,4321,8190,8712,12167,13880,14155,15637

EU177836:8,169,222d,364C,514,1602d,3345,3387,3541,4321,4822,8190,8712,12167,15094,15637,16068,16087,16111

EU177837:169,222d,364C,1602d,3088,3345,3387,3541,3798,4321,8190,8712,9098,10891,12167,12910,15637,15912,16057

EU177838:169,222d,352+G,364C,1602d,1860+A,3345,3387,3541,4321,6690,7194,8190,8712,12167,14368,15637,16303

EU177839:169,364C,1602d,3345,3387,3541,4321,8190,8242,8712,11187,11802,12167,15637,16110,16304

EU177841:8,169,364C,3345,3387,3541,4321,8190,8712,9562,12167,15398,15637,16115,16257

EU177842:169,222d,364C,3345,3387,3541,4321,7544,8190,8712,12167,15421,15637,16024,16052,16115,16257

EU177843:222d,364C,2057+C,3345,3387,3541,4321,8190,8712,9920,12167,13685,15637,16052,16115,16257

EU177844:169,222d,364C,2057+C,3345,3387,3541,4321,6048,8190,8712,12167,15084,15637,16052,16115,16257

EU177846:106,169,364C,2057+C,3345,3387,3541,4321,8190,8712,8961,12167,15637,15715d,16052,16078,16115,16117,16257

EU177847:169,364C,1602+AA,3345,3387,3541,4321,4977,6934,8190,8712,8893,12167,12254,15637,15966,16052,16115,16124,16198,16257

EU177848:169,364C,3345,3387,3541,3877,4321,8190,8525,8712,8864,10327,11454,12167,12653,15637,16052,16115,16198,16257

EU177849:169,222d,364C,3345,3387,3541,4321,4955,5634C,6204,6744,8190,8712,10710,12167,15637,16059C,16133,16187,16257

EU177850:169,364C,3345,3387,3541,4321,6204,6744,7852,8190,8712,10710,12167,13203,15637,16059C,16060,16127,16187,16257

EU177851:169,222d,364C,1461,2560,3345,3387,3541,4321,6204,6744,8190,8712,10710,11103,12167,12653,15637,15798,15987,16059C,16076,16187,16257

EU177852:169,364C,3345,3387,3541,4321,6204,6744,8190,8712,10710,12167,15637,16059C,16069,16076,16123,16187,16257

EU177853:169,170,222d,364C,3345,3387,3541,3942,4321,5158T,5763,6204,6744,8190,8712,10352,10710,12167,13753,15637,16052,16059C,16143,16187,16257

EU177854:169,222d,364C,1602d,1653,3345,3387,3541,4321,6204,6744,8190,8712,9502,10710,12167,14404,15637,16059C,16064,16187,16249,16257,16303

EU177855:169,222d,364C,3345,3387,3541,4321,5148,6204,6744,8190,8712,10710,12167,15637,16059C,16130,16187,16257

EU177856:169,173,222d,364C,364+C,3345,3387,3541,4321,5148,6744,8190,8247,8712,10261,10710,12167,15637,16051,16059C ,16115,16187,16257

EU177857:169,222d,364C,1325,1602d,1696,2601,3345,3387,3541,3720,4321,4783,5148,6204,6744,7400,8190,8712,10603,10710,12167,15159,15637,16059C,16060,16187,16250,16257,16303

EU177858:119,120,166,169,364C,3345,3351,3387,3541,3852,4321,6204,6405,6744,8190,8477,8712,9310,10662,10710,11832,12167,12294,12506,12767,13482,15637,15885d,16059C,16082,16187,16233,16257

EU177859:169,222d,364C,1555,3345,3387,3541,4321,4977,6204,6207,6744,8190,8712,8862,10710,12167,15637,16059C,16141,16187,16257

EU177860:169R,222d,364C,2087,3192,3345,3387,3541,4321,6204,6744,8190,8712,9482,9989,10710,11736,12167,13398,14581,15196,15637,16059C,16187,16257

EU177861:222d,364C,2087,2749,3192,3345,3387,3541,4321,6204,6316,6744,8190,8712,8986G,9989,10710,11736,12167,14581,15637,16059C,16187,16257

EU177863:163,364C,3345,3387,3541,4254,4321,8026,8190,8712,9091,9234,10881,12167,12782,12983,13007,15637,15884,16199,16257

EU177864:163,169,364C,3345,3387,3541,4254,4321,7660,8190,8712,9896A,12167,12983,13007,15637,16257

EU177865:163,169,204,364C,1292T,3345,3387,3541,4254,4321,8047,8190,8712,9896A,12167,12983,13007,15637,16139,16257

EU177866:222d,364C,1459,2560,3240,3345,3387,3417T,3541,4321,5503,7920,8190,8320,8372C,8407,8712,10719A,10929,11002,11091,11478,12167,12379,12435,12470,12677,12732,12752,12879,12926,13007,13821,14038,14110,15136,15629,15637,15955G,16060,16081,16257

EU177867:222d,364C,1459,2560,3240,3345,3387,3417T,3541,4321,5503,7920,8190,8320,8372C,8407,8712,10719A,10929,11002,11091,11478,12167,12379,12435,12470,12677,12732,12752,12879,12926,13007,13821,14038,14110,15136,15629,15637,15955G,16060,16081,16257

EU177868:8,106,166,173,206,233,234,249,250,297,301,364C,519,724,739,763,818,1160,1476,1494,1602d,1679,1826,1862A,1871,2018,2101,2323A,2560,2577,2636,2955,2979,2981,2990,2991,2992,3053,3073,3138,3147,3243,3327,3337,3345,3381,3387,3441,3537,3541,3552,3602,3795,3831,3876,3933,3978,3987,4002,4321,4330,4444,4564,4732,4735,4771,4873,4939,4990A,5287,5503,5533,5616,5745,5785,5892,5919,6000,6117T,6237,6342,6369,6381,6438,6462,6729,6774,6883,6924,7306,7332,7358,7360,7363,7516,7832,7853,8047,8170,8196,8212,8287,8310,8372C,8468,8496,8505,8516,8573,8712,8751,8986,9007,9040,9070,9247,9304,9482,9583,9604,9769,9893,9932C,9980,10041,10068,10073,10139,10155,10270,10324,10333,10447,10592,10623,10693G,10851,11002,11037,11070,11136,11202,11268,11331,11409,11421,11805,11844,12137,12167,12180,12236,12379,12435,12470,12471,12515,12624,12674,12677,12686,12752,12803,12902,12925T,12926,13007,13058A,13100,13103A,13106,13277,13373,13382,13435,13439,13556,13566,13586,13630,13679,13691,13694,13884,13910A,13911,14038,14068,14122,14131,14140,14257,14317,14413,14418,14505,14608,14827,14860,14899,15107,15136,15148,15289,15310,15328,15565,15581,15595,15607,15619,15629,15631,15637,15743,15753,15820,15953,15955G,15961,15996,16024,16051,16059,16060,16076,16084,16086,16104,16111,16115,16118,16119,16121,16123,16124,16132,16139,16140,16145d,16149,16198,16202+A,16231,16249,16250,16302,16303

EU177869:8,39,106,166,173,206,233,234,249,250,297,301,364C,364+C,519,739,763,818,1160,1476,1494,1602d,1679,1826,
1862A,1871,2018,2101,2560,2955,2979,2981,2990,2991,2992,3053,3073,3138,3147,3243,3318,3327,3337,3345,3381,3387,
3441,3537,3541,3552,3602,3795,3831,3876,3933,3978,3987,4002,4321,4330,4444,4564,4732,4735,4771,4873,4939,5275T,
5287,5503,5533,5616,5745,5785,5892,5919,6000,6117T,6237,6342,6369,6381,6438,6462,6498,6729,6774,6883,6924,7306,
7332,7358,7360,7363,7516,7832,7853,8047,8170,8212,8287,8310,8372C,8468,8496,8505,8573,8712,8751,8986,9007,9040,
9070,9247,9583,9604,9769,9893,9932C,9980,10041,10068,10139,10155,10270,10324,10333,10447,10592,10623,10693G,10851,
10890,10935,11002,11037,11070,11136,11202,11268,11331,11409,11421,11805,11844,12137,12167,12180,12236,12379,
12435,12470,12471,12515,12624,12674,12677,12686,12752,12803,12902,12925T,12926,13007,13058A,13100,13106,13277,
13382,13435,13439,13556,13566,13586,13679,13691,13694,13884,13910A,13911,14038,14068,14122,14131,14140,14234,
14257,14317,14371,14413,14418,14505,14608,14827,14833,14860,14899,15107,15136,15148,15289,15310,15328,15565,
15581,15595,15607,15619,15629,15631,15637,15743,15753,15820,15953,15955G,15961,15996,16024,16051,16052A,16059,
16060,16076,16084,16086,16087,16104,16111,16115,16118,16119,16121,16123,16124,16132,16139,16140,16143,16145d,
16149,16198,16202+A,16231,16234,16249,16250,16302,16303

EU177870:8,39,106,166,206,233,234,249,250,297,301,364C,519,739,763,818,1160,1476,1494,1602d,1679,1826,1862A,1871,
2018,2101,2560,2955,2979,2981,2990,2991,2992,3053,3073,3138,3147,3243,3318,3327,3337,3345,3381,3387,3441,3537,3541,
3552,3602,3795,3831,3876,3933,3978,3987,4002,4312,4321,4330,4444,4564,4732,4735,4771,4873,4939,5287,5503,5533,5616,
5745,5785,5892,5919,6000,6117T,6237,6342,6369,6381,6438,6462,6729,6774,6883,6924,7306,7332,7358,7360,7363,7501,
7516,7832,7853,8047,8170,8212,8287,8310,8372C,8468,8496,8505,8573,8712,8751,8986,9007,9040,9070,9247,9583,9604,
9769,9893,9932C,9980,10041,10068,10139,10155,10270,10324,10333,10447,10592,10623,10690,10693G,10851,10890,10935,
11002,11037,11070,11136,11202,11268,11331,11409,11421,11805,11844,12137,12167,12180,12236,12341,12379,12435,
12470,12471,12515,12624,12674,12677,12686,12752,12803,12902,12925T,12926,13007,13058A,13100,13106,13277,13382,
13435,13439, 13556,13566,13586,13679,13691,13694,13884,13910A,13911,14038,14068,14122,14131,14140,14257,
14317,14371,14413,14418,14505,14608,14827,14860,14899,15107,15136,15148,15289,15310,15328,15565,15581,15595,
15607,15619,15629,15631,15637,15743,15753,15820,15953,15955G,15961,15996,16024,16051,16059,16060,16076,16084,
16087,16104,16111, 16115,16118,16119,16121,16123,16124,16132,16139,16140,16143,16145d,16149,16198,16202+A,16231,
16234,16249,16250, 16302,16303

FJ971083:169,222d,244T,364C,1459,2560,3240,3345,3387,3417T,3541,4321,5503,7920,8190,8320,8372C,8619,8712,10719A,
10929,11002,11091,11768,12167,12435,12470,12677,12752,13007,13508,13622,14038,14110,15136,15637,15955G,16114,
16257

FJ971084:8,106,166,173,201,215+TC,235+T,250,297,301,364C,782,895,1601-1602d, 2044,2560,2570,2981,3304, 3345,3387,
3441,3541,3552,3602,4192,4321,4363,4855,5148,5503,5616,5745,5892,5937,6081,6120,6162,6438,6477,6774,6852,6985,7332,
7358,7361,7516,7933,8372C,8496,8712,8769,9007,9178,9482,9661,9866,9980,10333,10693G,10929,11002,11136,11394,11421,
11757,11844,12047,12167,12236,12470,12471,12624,12677,12686,12752,12803,12902,13007,13058A,13130,13156,13262,
13460,13508,13719,14002,14038,14053,14459,14626,14827,14908,14953,15067,15136,15581,15619,15626,15629,15637,
15820,15848,15902,15953,15955G,16059,16078,16086,16087,16123,16124,16129,16137,16139,16202+A,16233,16250,16252,
16266,16303

FJ971085:8,106,166,173,201,215+TC,235+T,250,297,301,364C,782,895,1601-1602d,2044,2560,2570,2981,3304,3345,3387,3441,
3541,3552,3602,4192,4321,4363,4855,5503,5616,5745,5892,5937,6081,6120,6162,6438,6477,6774,6985,7332,7358,7361,7516,
7933,8372C,8496,8712,8769,9007,9178,9482,9661,9866,9980,10333,10693G,10929,11002,11136,11394,11421,11757,11844,
12047,12167,12236,12470,12471,12624,12677,12686,12752,12803,12902,13007,13058A,13130,13156,13262,13460,13508,
13719,14002,14038,14053,14459,14626,14827,14908,14953,15067,15136,15581,15619,15626,15629,15637,15820,15848,
15902,15953Y,15955S,16059,16078,16086,16087,16123,16124,16129,16137,16139,16202+A,16233,16250,16252,16266,16303

FJ971088:8,106,166,173,206,233,234,249,250,297,301,364C,519,724,739,763,818,1160,1476,1494,1602d,1679,1826,1862A,
1871,2018,2101,2560,2577,2636,2955,2979,2981,2990,2991,2992,3053,3073,3138,3147,3243,3312,3327,3337,3345,3381,3387,
3441,3537,3541,3552,3602,3795,3831,3876,3933,3978,3987,4002,4321,4330,4444,4564,4732,4735,4771,4873,4939,4990A,
5287 ,5503,5533,5616,5745,5785,5892,5919,6000,6117T,6237,6342,6369,6381,6438,6462,6729,6774,6883,6924,7306,7332,
7358, 7360,7363,7516,7832,7853,8047,8170,8196,8212,8287,8310,8372C,8468,8496,8505,8516,8573,8712,8751,8986,9007,
9040, 9070,9247,9304,9482,9583,9604,9769,9893,9932C,9980,10041,10068,10073,10139,10155,10270,10324,10333,10447,
10592, 10623,10693G,10851,11002,11037,11070,11136,11202,11268,11331,11409,11421,11805,11844,12137,12139,12167,
12180, 12236,12379,12435,12470,12471,12515,12624,12674,12677,12686,12752,12803,12902,12925T,12926,13007,13058A,
13100, 13103A,13106,13277,13373,13382,13435,13439,13556,13566,13586,13630,13679,13691,13694,13884,13910A,13911,
14038, 14068,14122,14131,14140,14257,14317,14413,14418,14505,14608,14827,14860,14899,15107,15136,15148,15159,
15289, 15310,15328,15565,15581,15595,15607,15619,15629,15631,15637,15743,15753,15820,15953,15955G,15961,15996,
16024, 16051,16059,16060,16076,16084,16086,16104,16111,16115,16118,16119,16121,16123,16124,16132,16139,16140,1
6145d, 16149,16198,16202+A,16231,16249,16250,16302,16303

GQ129207:166,363+C,3345,3387,3541,4321,4744,8190,8712,12167,15637,16087,16133

GQ129208:353G,364C,3345,3387,3541,4321,8190,8712,9136,12160,12167,15637

GU947007:169,190,222d,364C,1602d,1865,2632T,3345,3387,3541,3810,4321,5044,7241,8190,8712,9064,9238,12167,14218,
14233,15637,15987d,16059+A,16070,16076A,16078

GU947008:169,364C,1602d,3345,3387,3541,4321,4333,5044,8190,8712,9731,12167,15637,15987d,16059+A,16076A,16078

GU947010:169,352+G,364C,1602d,3345,3387,3541,4321,8190,8712,10462,12167,15637,15987d,16059+A,16076A,16078,16149

GU947012:169,364C,1602d,3345,3387,3541,4321,6327,8190,8712,9731,12167,12629,13280,15637,15987d,16059+A,16076A,
16078,16233

GU947014:169,364C,1602d,3345,3387,3541,4321,4333,5226,8190,8712,9731,12167,13784T,15637,15987d,16059+A,16076A,
16078

GU947018:169,364C,1602d,3345,3387,3541,4321,8190,8712,10462,12167,14218,14233,15637,15987d,16059+A,16076A,16078,16149

GU947019:166,169,222d,537,567,3345,3387,3541,4321,8190,8712,8927,12167,13901,15637

GU947020:169,364C,1602d,2057+C,3345,3387,3541,4159,4321,7029,8190,8407,8712,11430,12167,14429,15637,16052,16115,16141,16149,16257

GU947021:169,364C,1602d,2055,3345,3387,3541,4321,4756C,4757C,5604,5718,7011,7167R,8190,8712,12025,12167,12178,13091A,13520,15637,16138,16225,16228C

GU985279:190,224d,250,301,302,365d,1130,1483,2147,2560,2587,3345,3381,3387,3541,3552,4254,4295,4321,4678,5158,5503,5745,5892,5901,6162,6438,7358,7954,7996,8190,8238,8360,8372C,8712,10128,10693G,11002,11142,11742,12018,12167,12379,12435,12470,12471,12527,12677,12752,13007,13823,14038,14131T,14582,14875,15136,15629,15637,15675,15953,15955G,15996,16021,16051,16053,16060,16076,16087,16123R,16124,16143,16233,16257,16266,16303

HM045018:3189,3345,3387,3541,4321,8190,8712,12167,15637

HQ025805:169,222d,364C,589d,3345,3387,3541,3561,3562G,4108,4321,4327,5855,6204,6744,7960,8190,8712,10710,12167,12272,14418,15637,16059C,16187,16257

HQ184030:169,222d,364C,2560,3240,3345,3387,3417T,3541,4321,5503,6438,7920,8190,8320,8372C,8712,8881,10719A,10929,11002,11091,12167,12435,12470,12677,12752,13007,14038,14110,15136,15158,15629,15637,15923,15955G,16257

HQ184031:169,222d,364C,2560,3240,3345,3387,3417T,3541,4321,5503,6438,7920,8190,8320,8372C,8712,8881,10719A,10929,11002,11091,12167,12435,12470,12677,12752,13007,14038,14110,15136,15158,15629,15637,15923,15955G,16257

HQ184032:169,222d,364C,2560,3240,3345,3387,3417T,3541,4321,5503,6438,7920,8190,8320,8372C,8712,8881,10719A,10929,11002,11091,12167,12435,12470,12677,12752,13007,14038,14110,15136,15158,15629,15637,15923,15955G,16257

HQ184034:169,222d,244T,364C,1459,2560,3240,3345,3387,3417T,3541,4321,5503,7920,8190,8320,8372C,8619,8712,10719A,10929,11002,11091,11768,12167,12435,12470,12677,12752,13007,13508,13622,14038,14110,15136,15637,15955G,16114,16257

HQ184035:169,222d,244T,364C,1459,2560,3240,3345,3387,3417T,3541,4321,5503,7920,8190,8320,8372C,8619,8712,10719A,10929,11002,11091,11768,12167,12435,12470,12677,12752,13007,13508,13622,14038,14110,15136,15637,15955G,16114,16257

HQ184036:222d,364C,1459,2560,3240,3345,3387,3417T,3541,4321,5503,7920,8190,8320,8372C,8407,8712,10719A,10929,11002,11091,11478,12167,12379,12435,12470,12677,12732,12752,12879,12926,13007,13821,14038,14110,15136,15629,15637,15955G,16060,16081,16257

HQ184037:222d,364C,1459,2560,3240,3345,3387,3417T,3541,4321,5503,7920,8190,8320,8372C,8407,8712,10719A,10929,11002,11091,11478,12167,12379,12435,12470,12677,12732,12752,12879,12926,13007,13821,14038,14110,15136,15629,15637, 15955G,16060,16081,16257

HQ184038:222d,364C,1459,2560,3240,3345,3387,3417T,3541,4321,5503,7920,8190,8320,8372C,8407,8712,10719A,10929,11002,11091,11478,12167,12379,12435,12470,12677,12732,12752,12879,12926,13007,13821,14038,14110,15136,15629,15637,15955G,16060,16081,16257

HQ184039:169,222d,364C,471,1459,2560,3240,3345,3387,3417T,3541,4321,5503,5718,6660,7832,7920,8190,8320,8372C,8712,10719A,10929,11002,11091,12167,12435,12470,12482,12677,12752,13007,14038,14110,15092,15100A,15136,15629,15637,15955G,16198,16257

JN817298:106,169,222d,364C,1593,3191,3345,3387,3541,4321,6237,8190,8712,12167,15637,16052,16115,16257

JN817299:169,222d,250,364C,3345,3387,3541,4321,4372,6237,6453,7400,8190,8712,11070,12167,15637,16052,16115,16257

JN817300:169,222d,364C,3345,3387,3541,3877,4321,8190,8712,11475,12167,15637,15966,16052,16081,16115,16124,16129,16198,16202,16257

JN817302:169,364C,3345,3387,3541,4321,7544,8190,8712,9102,12167,14525T,14533,14992,15637,15846,16024,16115

JN817303:169,222d,364C,2057+C,3345,3387,3541,3684,4321,4453,8190,8712,11028,12167,15064,15637,16052,16115,16248,16249,16257

JN817304:169,364C,1602d,3345,3387,3541,4321,4858,6237,6333,7762,8190,8712,12167,15149,15637,16052,16115,16257

JN817305:169,364C,2967,3345,3387,3541,4321,7544,8190,8712,9433,12167,12362,13203,14525T,15637,16024,16052,16058,16115,16149,16257,16304

JN817306:8,169,364C,3345,3387,3541,4321,8190,8712,12167,15637,15941,16115,16250,16257

JN817307:24,169,222d,364C,3345,3387,3406,3541,4321,5032,8190,8712,9677,12167,14386,14600G,15328,15637,15811,16052,16115,16124,16198,16257

JN817308:8,169,222d,364C,1602+A,2057+C,2080,3345,3387,3541,4321,8190,8712,10276,12167,15637,16052,16060,16110,16115,16124,16257

JN817309:169,222d,364C,1326,3345,3387,3541,4321,8190,8712,11544,12167,13884,15637,16052,16055,16115,16124,16141,16198,16257

JN817310:169,222d,364C,1326,3345,3387,3541,4321,8190,8712,11544,12167,15637,16052,16055,16115,16124,16141,16198,16257,16318

JN817311:169,222d,364C,1326,3345,3387,3541,4321,8190,8712,10895A,11544,12167,15464,15637,16052,16055,16115,16124,16141,16198,16257

JN817312:169,222d,364C,665,2057+C,3345,3387,3541,4321,8190,8712,10519,12167,13061,13898,15637,16052,16115,16257,16262

JN817313:169,222d,267T,364C,2057+C,3345,3387,3541,4321,8190,8712,12167,13880,15637,16052,16115,16257

JN817314:169,222d,364C,756,2057+C,3345,3387,3541,4321,6388,8190,8712,9840,12167,14327,15637,16052,16115,16257

JN817315:8,169,222d,364C,2057+C,3345,3387,3541,4321,8190,8712,11190,12167,15637,16052,16115,16257

JN817316:169,222d,364C,2057+C,2584,3345,3387,3541,4321,8190,8712,10603,11024,12167,14582,15637,15976,16052,16115,16257

JN817317:8,169,364C,1300,2057+C,3345,3387,3541,4321,4742,7322,8190,8712,9604,12167,15637,16052,16115,16121,16169,16257

JN817318:169,222d,364C,2057+C,3345,3387,3541,3546A,3942,4321,7068,8190,8712,12167,12476,15637,16052,16115,16257

JN817319:169,364C,3345,3387,3541,4321,8190,8712,10246A,12167,15637,16115,16124,16257

JN817320:222d,364C,3210,3345,3387,3541,4321,7544,8190,8712,12167,15149,15328,15637,16024,16052,16115,16257

JN817321:8,169,364C,1602d,3345,3387,3541,4321,4858,6237,8190,8712,9514,10516,12167,15637,16052,16115,16250,16257

JN817322:106,169,364C,1326,1876,3345,3387,3541,4321,7657,8190,8712,9070,11544,12167,15637,16018,16052,16055,16115,16124,16141,16198,16257

JN817323:169,222d,353G,364C,1602d,1881,3345,3387,3541,4321,4349,4882,6019,8190,8328,8712,9722,12167,12407,12730,15421,15637,15958,16052,16115,16124,16198,16257

JN817324:169,364C,3345,3387,3541,4321,7544,8190,8350,8712,12167,12229,15637,16024,16115,16124,16257

JN817325:169,364C,3345,3387,3541,4321,8190,8712,10498,10926,11791,12167,15637,16052,16115,16124,16137,16208,16257

JN817326:169,364C,2192,3345,3387,3541,4321,8190,8334,8712,11963,12167,15100,15158,15637,16052,16114,16115,16124,16257

JN817327:106,169,222d,364C,3345,3387,3541,4321,7544,7818,8190,8712,12167,14053,15637,15948,16115

JN817328:169,364C,1602d,3345,3387,3541,4321,5188,8190,8712,8796A,12167,12362,13733,15637,16052,16057,16115,16124,16257

JN817329:169,222d,364C,3345,3387,3541,4321,7757,8190,8712,9385,12167,12494,14045,15092,15406,15637,16052,16257

JN817330:169,364C,1602d,3345,3387,3541,4321,4858,6237,8190,8712,11457,12167,15637,16052,16115,16257

JN817331:169,222d,364C,2057+C,3345,3387,3541,4167,4321,8190,8235,8251,8712,9840,11791,12167,15637,16052,16115,16135,16233,16257

JN817332:8,169,222d,364C,2057+C,3345,3387,3541,4321,8190,8712,11190,12167,15637,16052,16115,16257

JN817333:169,222d,364C,2057+C,3345,3387,3406,3541,4321,8190,8712,10563,12167,12732,13164,15637,16052,16115,16257

JN817334:169,222d,364C,3345,3387,3541,4321,4947,7544,8190,8712,10579,11532,12167,13373A,15637,15848,16024,16052,16115,16257,16303

JN817335:169,173,222d,364C,2057+C,3345,3387,3541,4114,4321,8190,8712,11190,12167,15637,16052,16115,16257

JN817336:8,169,222d,364C,2057+C,3345,3387,3541,4321,5017,5227,5814,8190,8712,12167,13125,13673,14694,14905,15527,15637,16052,16115,16257

JN817337:169,222d,364C,2057+C,3345,3387,3406,3541,4321,7299,8190,8712,10563,12167,12732,15637,16052,16115,16141,16257,16304

JN817338:169,222d,353G,364C,2025,2057+C,3345,3387,3541,4321,7068,7511,8190,8712,9808,12167,12452,12476,15637,16052 ,16115,16143,16166,16257

JN817339:169,222d,364C,711,2057+C,3345,3387,3541,4321,8190,8712,9445,10857A,12167,15637,16052,16115,16257,16266

JN817340:169,222d,364C,2057+C,3345,3387,3541,4321,4742,7322,8160,8190,8712,9604,11956,12167,15637,16052,16115,16121,16169,16257

JN817341:169,364C,2057+C,3345,3387,3541,4321,8190,8712,12167,15637,16018,16052,16115,16257

JN817342:169,222d,364C,2057+C,3345,3387,3406,3541,4321,7299,8190,8712,10563,12167,12732,15637,15656A,16052,16115,16257

JN817343:169,364C,3345,3363,3387,3541,4321,8190,8712,9638,11850,12167,12494,14212,14611,15412,15637,16052,16257

JN817345:106,169,222d,364C,1842,2057+C,3345,3387,3541,4321,8190,8198Y,8669,8712,8861,12167,15637,16052,16115,16257

JN817346:169,222d,364C,3345,3387,3541,4321,8190,8712,10530,10797,12167,14459,14821,15637,16052,16055,16115,16124,16135,16198,16257

JN817347:169,222d,364C,2057+C,3345,3387,3541,4321,8190,8712,9686,12167,12902,13425,15216,15637,16052,16115,16257,16304

JN817348:169,222d,364C,3345,3387,3541,4141,4321,5898,6705,7544,8190,8712,9136,9658,9731,10605,12167,13295,13984,15637,16052,16115,16249,16257

JN817349:8R,169,364C,3345,3387,3541,4321,7544,8190,8712,12167,13203,14525T,15637,16024,16052,16058,16115,16140,16149,16257

JN817351:169,178,222d,364C,3345,3387,3541,4321,5638,7544,8190,8712,12167,14758,15637,16024,16052,16115,16257,16293

JQ437479:169,190,224d,250,301,302,365d,1130,1483,2560,2587,3190,3304,3345,3381,3387,3541,3552,4321,4678,5158,5503,5745,5892,5901,6162,6438,6453,7358,7954,7996,8190,8238,8360,8372C,8712,10128,10693G,11002,11142,12018,12167,12379,12435,12470,12677,12752,13007,13541,13823,14038,14131T,14875,15136,15473,15629,15637,15675,15953,15955G,15996,16051,16060,16076,16087,16124,16233,16257,16266

KC153972:169,177,364C,1602d,3345,3387,3541,4321,8190,8712,12167,15637,16051,16052

KC153974:169,364C,1602d,3345,3387,3541,4321,5226,8190,8712,8832,10884,11514,12167,13487,15637,16140

KC153976:169,236,364C,1602d,3282,3345,3387,3541,4321,6336,8190,8712,9604,12167,15637

KC153977:222d,353G,364C,3345,3387,3541,4321,8190,8712,10686,12167,14418,15637,16064,16249

KF163061:169,215+TC,232+C,364C,1601-1602d,3345,3387,3541,4321,6052,7544,8190,8712,12167,12470,12471, 14350,15637,16024,16052,16086,16087,16115,16202+A,16257

KF163063:169,215+TC,235+T,364C,1601-1602d,3345,3387,3541,4321,7544,8190,8712,12167,15637,16024, 16052,16115,16202+A,16257,16303

KF163064:169,190,215+TC,235+T,364C,498,1601-1602d,3071,3345,3387,3541,4321,7544,8190,12167,12527,13203,13302,14525T,15637,16024,16052,16058,16059,16086,16115,16204+G,16318

KF163065:169,190,215+TC,235+T,364C,498,1601-1602d,3071,3345,3387,3541,4321,4363,7544,8190,12167,12470,12471,12527,13203,13302,14053,14525T,15637,16024,16052,16058,16086,16087,16115,16202+A,16318

KF163066:169,190,215+TC,235+T,364C,498,1601-1602d,3071,3345,3387,3541,4321,7544,12167,12470,12471,12527,13203,13302,14525T,15637,16024,16052,16058,16087,16115,16202+A,16318

KF163067:169,173,190,215+TC,235+T,364C,498,1601-1602d,3345,3387,3541,4321,7544,8190,12167,12470,12471,12527, 13203,13302,14525T,15637,16024,16052,16058,16086,16087,16115,16202+A,16257,16318

KF163068:169,190,215+TC,235+T,364C,498,1601-1602d,3272,3345,3387,3415-3454N,3541,4321,4842Y,7544,8190,12167,12363,12470,12471,12527,13203,13302,14525T,15637,16024,16052,16058,16086,16087,16115,16202+A,16257,16318

KF163070:169,190,215+TC,235+T,364C,498,1601-1602d,3345,3387,3541,4321,7544,7624,8190,12167,12470,12471,12527,13203,13262,13302,14525T,15637,16024,16052,16058,16086,16087,16115,16202+A,16257,16318

KF163071:169,190,216Y,222+CC,235+T,364C,498,1601-1602d,3272,3345,3387,3541,4321,7544,8190,12167,12363,12470,12527,13203,13302,14525T,15637,16024,16052,16058,16115,16204+G,16257,16318

KF163072:169,215+TC,235+T,364C,1601-1602d,3345,3387,3541,4321,4588,6027,8190,8712,9908,10909,12167, 15462,15637,16052,16086,16087,16115,16123,16124,16202+A,16257

KF163073:169,215+TC,235+T,364C,1601-1602d,3345,3387,3541,4321,7544,8190,8712,12167,12470,12471,13203, 14525T, 15158,15637,16024,16052,16115,16202+A,16257

KF163075:169,215+TC,235+T,364C,1601-1602d,3345,3387,3541,4321,5227,7544,8190,8712,12167,12404,12471, 13203, 14525T, 14773,15637,16024,16052,16058,16059,16115,16202+A,16257

KF163077:169,190,215+TC,235+T,364C,498,1077C,1601-1602d,3345,3387,3541,4321,6393,7544,8190,9097,12167,12527,13203,13302,14525T,15637,16024,16052,16058,16086,16087, 16115,16123,16202+A,16257,16318

KF163078:169,190,215+TC,235+T,364C,498,1601-1602d,3272,3345,3387,3541,4321,7544,8190,12167,12363,12470,12471,12527,13203,13302,14525T,15637,16024,16052,16058,16087,16115,16123,16124,16202+A,16257,16318

KF163079:169,173,190,215+TC,235+T,364C,498,1601-1602d,3345,3387,3541,4321,7544,8190,12167,12527,13203,13302, 14525T, 15637,16024,16052,16058,16115,16202+A,16257,16318

KF163080:169,190,215+TC,235+T,364C,498,1601-1602d,3345,3387,3541,4321,6477,7544,8190,12167,12470,12471,12527,13203,13302,14525T,15637,16024,16052,16058,16086,16087,16115,16124,16202+A,16257,16318

KF163081:169,215+TC,235+T,364C,1601-1602d,3345,3387,3541,4321,6237,8190,8712,10349, 10659,12167,12470, 15637,16052,16086,16087,16115,16123,16202+A,16257

KF163082:169,215+TC,235+T,300,364C,1601-1602d,3345,3387,3541,3684,4321,8190,8712,9203,12167,12470,12471,15637,16052,16115,16202+A,16250,16257,16262

KF163083:169,190,215+TC,235+T,364C,498,1601-1602d,3272,3345,3387,3541,4321,7516,8190,12167,12363,12527,13203,13302,14525T,15637,16024,16052,16058,16086,16087,16115,16202+A,16257,16318

KF163084:169,215+TC,235+T,364C,1601-1602d,3345,3387,3541,4321,7544,8190,8712,12167,13397,15637, 16024,16052, 16059,16070,16086,16087,16115,16202+A,16257

KF163085:169,190,215+TC,235+T,250,297-298N,364C,498,1601-1602d,3345,3387,3541,4321,7544,8190,12167,12470,12471,12527,13007,13203,13302,14525T,15637,15953,15955G,16024,16052,16058,16087,16115,16202+A,16257,16318

KF163086:169,190,215+TC,235+T,364C,498,1601-1602d,3345,3387,3541,4321,7544,8190,12167,12527,13203,13302,14525T,15637,16024,16052,16058,16115,16124,16202+A,16257,16318

KF163090:169,190,215+TC,235+T,364C,498,1601-1602d,3272,3345,3387,3483,3541,4321,7544,8190,12167,12470,12471,12527,13203,13302,14525T,15637,16024,16052,16058,16086,16087,16115,16202+A,16257,16318

KF163092:169,215+TC,235+T,364C,1601-1602d,3345,3387,3541,4321,7544,8190,8712,9503,12167,13262,15637,16024,16052, 16086,16087,16115,16202+A,16257

KF163093:169,215+TC,235+T,300,364C,1601-1602d,3345,3387,3541,3684,4321,8190,8712,9203,12167,12470, 12471,15637, 16052,16115,16202+A,16249,16257,16262

KF163094:169,179,215+TC,235+T,364C,1483,1601-1602d,3155Y,3345,3387,3541,4321,7544,8190,8712,10462,12167, 12470, 12471,13203,14525T,15637,16024,16052,16058,16115,16202+A,16257

KF926377:169,364C,1602d,3345,3387,3541,4060,4321,8063A,8065A,8066A,8069,8190,8712,12167,13734,15637,16053,16063, 16131

Appendix 4: Chapter 4 Supplementary materials

(available online at https://doi.org/10.1186/s12864-020-07018-7)

Additional file 1: Table S1. Average RIN by tissue types

Additional file 2: Table S2. Quality of library preparation.xls

Additional file 3: Table S3. Read alignment quality check.xls

Additional file 4: Table S4. List of Mitochondrial protein genes derived from Mitocarta in cattle.xls

Additional file 5: Table S5. List of Mitochondrial protein genes derived from Mitocarta in Sheep.xls

Additional file 6: Table S6. Number of differentially expressed (DE) genes by gene categories averaged for two foetuses in the Main Cows

Additional file 7: Figure S1. Heatmap of expression of nuclear genome encoded mitochondrial protein (NuMP) in tissues of foetuses 6819F and 2181F in Main Cows

Additional file 8: Figure S2. Heatmap of mitochondrial genome encoded mitochondrial protein (MtMP) genes in tissues of foetuses 6819F and 2181F in Main Cows.

Additional file 9: Table S7. List of non-mitochondrial protein (Non-MP) genes clustering with the mitochondrial protein genes in cluster I (NuMP-MtMP cluster) in Main Cows

Additional file 10: Table S8. KEGG pathway enrichment of the non-mitochondrial protein (Non-MP) genes in NuMP-MtMP cluster in Main Cows

Additional file 11: Figure S3. The proportion of differentially expressed gene in each gene category in 18 tissues in a Validation Cow (All=All genes encoded by nuclear and mitochondrial genome, Nu=Mitochondrial protein genes encoded by nuclear genome (NuMP), Mt=Mitochondrial protein genes encoded by mitochondrial genome (MtMP)

Additional file 12: Figure S4. Heatmap of expression of nuclear genome encoded mitochondrial (NuMP) gene in a Validation Cow

Additional file 13: Figure S5. Heatmap of expression of mitochondrial genome encoded mitochondrial protein (MtMP) genes in the Validation Cow.

Additional file 14: Figure S6. The proportion of differentially expressed gene in each gene category and direction of gene regulation in 15 tissues in Validation Sheep (All=All genes encoded by nuclear and mitochondrial genome, Nu=Mitochondrial protein genes encoded by nuclear genome (NuMP), Mt=Mitochondrial protein genes encoded by mitochondrial genome (MtMP)

Additional file 15: Figure S7. Heatmap of nuclear genome encoded mitochondrial protein genes (NuMP) in Validation Sheep (three adults Texel x Blackface female sheep AF1, AF2, and AF3)

Additional file 16: Figure S8. Heatmap of mitochondrial genome encoded mitochondrial protein genes (MtMP) genes in Validation Sheep (three adults Texel x Blackface females AF1, AF2, and AF3)

Additional file 17: Table S9. Number of differentially expressed gene (DEG)s and their direction in tissues by gene categories in a Validation Cow

Additional file 18: Table S10. Number of differentially expressed gene (DEG)s and their direction in tissues by gene categories in Validation Sheep

Additional file 19: Figure S9. Scatter plot of log fold changes of Main Cows against the log-fold changes of Validation Sheep for mitochondrial protein gene expression in tissues

Additional file 20: Figure S10. Scatter plot of log fold changes of Validation Cow against the log-fold changes of Validation Sheep for mitochondrial protein gene expression in thyroid

Additional file 21: Figure S11. Gene co-expression network constructed based similarity matrix computed using Person Correlation Co-efficient of gene expression at $r > |0.95|$ across tissues of the Validation Cow.

Additional file 22: Figure S12. Gene co-expression network constructed based similarity matrix computed using Person correlation coefficient of gene expression at $r > |0.95|$ across tissues of Validation Sheep (three Texel x blackface adult female sheep AF1, AF2 and AF3).

Appendix 5: Chapter 5 Supplementary materials

(Available at https://doi.org/10.3168/jds.2020-18503)

Supplemental Table S1. Group means of variables used in the estimation of Residual feed intake (RFI) and Energy balance (EB)

Supplementary Table S2. Beta hydroxybutyrate (BHB), Non-esterified fatty acid (NEFA)and blood urea nitrogen (BUN) of high residual feed intake (H_RFI) and low residual feed intake (L_RFI) cow groups (a); and positive energy balance (PEB) and negative energy balance(NEB) cow groups (b)

Supplementary Table S3. Differentially expressed genes in low residual fed intake (L_RFI) group compared to high residual feed intake group (H_RFI)

Supplementary Table S4. Differentially expressed mitochondrial protein genes in low residual feed intake (L_RFI) group compared to high residual feed intake (H_RFI) group

Supplementary Table S5. List of putative hub genes in Module Eigengene ME2 of the main RFI dataset

Supplementary Table S6. List of putative hub genes in Module Eigengene ME3 of the Main RFI dataset

Supplementary Table S7. Differentially expressed mitochondrial protein genes from nuclear genome in low residual feed intake (L_RFI) group compared to high residual feed intake (H_RFI) group in the validation dataset

Supplementary Table S8. List of differentially expressed mitochondrial protein genes from nuclear genome in common between low residual feed intake groups of the Main and the Validation dataset

Supplementary Table S9. Differentially expressed mitochondrial protein genes (transcripts and proteome) in current study and occurrence in other feed efficiency studies in animals

Supplemental Figure S1. a) PCA plot of gene expression in white blood cells of top and bottom 8 animals (H_RFI and L-RFI) of the validation set; b) Hierarchical clustering of gene expression in the top and bottom 8 RFI cows with trait heatmap indicating the intensity of RFI values in a scale of blue (low) to red (high)

Supplemental Figure S2. Blood gene expression modules correlated to RFI based on WGCNA in the Validation dataset. a) Module eigengene (y-axis) – RFI relationship with p-values in parenthesis. "*" indicates modules with a significant relationship ($r > |0.4|$, $p \leq 0.05$) with RFI. The relationship between the trait and the modules is used to indicate the strength of the correlation where red is positive, and blue is negative; b) KEGG pathways enrichment of the module ME1.

Supplemental Figure S3. A Protein-Protein Interaction network featuring the network of putative hub genes from genes in Module ME1 (402) with the confidence of more than 0.90 from STRING database for validation dataset. Colors indicate KEGG pathways associated with these genes: ribosome (brown), oxidative phosphorylation (purple), apoptosis (dark green) and thyroid hormone signalling (yellow).