Summer 7-25-2022

A Tale of Two Genomes: The Complex Interplay Between the Mitochondrial and the Nuclear Genomes

Abhilesh S. Dhawanjewar
University of Nebraska-Lincoln, abhilesh.dhawanjewar@huskers.unl.edu

A TALE OF TWO GENOMES: THE COMPLEX INTERPLAY BETWEEN THE MITOCHONDRIAL AND THE NUCLEAR GENOMES

by

Abhilesh S. Dhawanjewar

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Biological Sciences

(Ecology, Evolution, and Behavior)

Under the Supervision of Professors Kristi L. Montooth & Colin D. Meiklejohn

Lincoln, Nebraska

June 2022

A TALE OF TWO GENOMES: THE COMPLEX INTERPLAY BETWEEN THE MITOCHONDRIAL AND THE NUCLEAR GENOMES

Abhilesh S. Dhawanjewar, Ph.D.

University of Nebraska, 2022

Advisors: Kristi L. Montooth & Colin D. Meiklejohn

Mitochondria, the product of an ancient endosymbiotic event are pivotal to eukaryotic cells by synthesizing the majority of the cell's ATP output. However, modern-day mitochondria are completely dependent on more than one thousand nuclear-encoded products for their function and the maintenance of their genomes. The fundamentally different ways in which the mitochondrial (mtDNA) and the nuclear (nucDNA) genomes are replicated and inherited lead to captivating coevolutionary dynamics between them. The aims of this dissertation are to investigate the coevolutionary dynamics between the mitochondrial and nuclear genomes at three distinct biological scales. At the organismal level, we use a *Drosophila* strain with a characterized mitochondrial-nuclear incompatibility to test for the functional effects of mitochondrial-nuclear interactions on male reproductive fitness, in the context of both gene-environment interactions and the female-specific selective sieve that operates on mtDNA. We find that the mitochondrial-nuclear incompatibility negatively affects male fertility, although these effects are largely context-dependent. At the molecular level, using sequence and structural comparisons, we classify and characterize mutations associated with human mitochondrial disorders in

the mtDNA and nucDNA as compensated or uncompensated based on whether the mutant amino acid is observed in a non-human species. We find that mtDNA, relative to nucDNA harbors a higher proportion of compensated mutations and this pattern is likely driven by the higher mtDNA background substitution rate. At the phylogenomic level, we estimate rates of evolutionary change for mtDNA- and nucDNA-encoded genes and compare correlations between the rates of mtDNA-encoded genes and three nuclear-encoded gene sets with differing extent of overlapping interactions with the mtDNA genes. We find that the patterns of rate correlations are consistent with the extent of overlap between the mtDNA and nucDNA genes with nucDNA genes that directly interact with mtDNA exhibiting the strongest correlations. In summary, we find that the higher rate of mutation in mtDNA appears to be driving mitochondrial-nuclear coevolutionary dynamics with the effects of mitochondrial-nuclear interactions being largely context-dependent.

ACKNOWLEDGEMENTS

Life, is a series of serendipitous events, and my scientific pursuit towards understanding the intricacies of life on Earth was no different, molded by the monumentous support and influence of a multitude of people. I've been extremely fortunate to have had amazing mentors throughout my scientific career, mentors who excitingly encouraged me to follow wherever my curiosity takes me and sharpened my investigative skills to draw meaningful inferences about the natural world. I first came across the fascinating world of mitochondria during a summer research opportunity offered to me by Dr. Henrik Krehenwinkel in 2013, which set me on my journey to study the interactions between the mitochondrial and the nuclear genomes that forms the basis of this thesis. I thank Henrik for his patience and guidance during those early years. Along my undergraduate studies, I've had the pleasure to work with Dr. Priya Iyer and Dr. M.S. Madhusudhan who taught me to use complementary methods from diverse fields to answer difficult questions. Thanks a lot for the stimulating brainstorming sessions and your constant support that helped me venture out deeper into the unknown.

This thesis would not have been possible without my advisors, Dr. Kristi Montooth and Dr. Colin Meiklejohn, who put tremendous faith in me, accepted my late application for graduate school and made a lot of things happen that resulted in me starting my PhD studies at Nebraska. In addition to the scientific training, they've also imparted many valuable life lessons, for which I'm forever grateful. By way of example, they taught me how to be a good scientist and their constant encouragement and support, especially during the tough times, showed me the importance of care and empathy, even in professional settings. Thanks for all the love; I aspire to be as empathetic and caring as

you in my future appointments. My advisory committee members, Dr. Jay Storz, Dr. Etsuko Moriyama, and Dr. Jeffrey Mower instrumental in shaping the research presented in this thesis. Their complementary expertise and indispensable feedback broadened my horizons and refined the analytical methods I employed in my research. Thanks a lot for agreeing to be on my committee and for your insightful comments throughout the development of my research.

I express my deepest gratitude to my partner, Nadejda (Nadya) Mirochnitchenko who despite having finished her graduate school a couple of years ago, voluntarily agreed to weather the stresses of graduate school once again through me. Nadya, Thanks a lot for being there for me by cheering for me from the sidelines, keeping me fed, taking me on walks and reminding me to not take myself so seriously at times. You're the best Rocket League teammate I could've ever wished for, the perfect fellow traveller on our world adventures and the most agreeable companion. Our sweet pug, CeeCee Pugsley, also played a crucial role in helping me complete this monumental task by constantly accompanying me in the work-from-home sessions during the pandemic, reminding me to take pleasures in the simple things in life, and overall being an utter goofball.

One of the perks of being co-advised was that I had the fortune to be intimately involved in both Meiklejohn and Montooth labs, who were my family away from family. Omera, Cole, Justin, Katherine, Ibrahim and Joevy, thanks a lot for lending a patient ear to my grievances and helping me get through the mundane, everyday stresses of graduate school. My best friends, Rudy and Natalia were by my side during the best of times and the worst of times and I've forged countless fun memories with them, far too many to

recount, memories that I'll cherish for a long time. They've been my biggest cheerleaders and the strongest support system I could've hoped for in graduate school. Thanks a lot.

Last but definitely not least, I thank my parents, Manisha and Sunil Dhawanjewar, who like most Indian parents would've liked me to become an engineer or the other kind of doctor but were very encouraging and nurturing, instilling the courage in me to follow my dreams, wherever they might lead me. I adore their bottomless enthusiasm towards celebrating all the major milestones in my life and I constantly sense their unwavering support as it fuels me on my journey onwards.

TABLE OF CONTENTS

INTRODUCTION

The endosymbiotic origin of mitochondria is perhaps one of the most crucial evolutionary innovations that enabled eukaryotic life to diversify and flourish on earth. In addition to their primary role in aerobic respiration, present day mitochondria have proteomes of > 1000 proteins that perform crucial functions in the eukaryotic cell including protein synthesis, amino acid and nucleotide metabolism, fatty acid catabolism, apoptosis, and ion homeostasis. The endosymbiotic merger was followed by both gene loss and the transfer of mitochondrial genes to their host genomes, further tightening their symbiotic relationships. Most modern animal mitochondrial genomes are gene-dense, retaining just 37 genes—13 structural genes that function in oxidative phosphorylation (OXPHOS), 22 tRNA genes and 2 ribosomal RNA genes.

The molecular machinery for mitochondrial replication, transcription and translation is encoded by the nuclear genome. Mitochondrial function and genomic maintenance are thus dependent on nuclear-encoded products and interactions between mitochondrial genes and mitochondrially-targeted nuclear-encoded (N-mt) genes (Figure 1). These interactions are proposed to be fundamentally important in shaping evolutionary, genetic, physiological, and developmental pathways (Burton, 2022; Rand et al., 2004; Rand, 2005; Wallace, 2016), but also have important implications for human health, ranging from aging to disease to fertility.

Owing to their prokaryotic ancestry, mitochondrial genomes are fundamentally different from nuclear genomes. Mitochondrial genomes in animals lack introns and differ in their mutational biases and nucleotide composition compared to their nuclear

counterparts in ways that vary across species (Montooth & Rand, 2008). The closer proximity to redox centers in the cell in combination with the relatively lower fidelity of the mitochondrial proof-reading machinery means that mitochondrial genomes often exhibit higher mutation and substitution rates relative to nuclear genomes. Mitochondrial genomes also experience overall smaller effective population sizes (N_e) than do nuclear genomes due to a lack of recombination and maternal inheritance. The lower N_e is predicted to decrease the efficacy of natural selection on the mitochondrial genomes (Nachman, 1998; Rand & Kann, 1996, 1998). However, analyses of larger data sets infer similar efficacies of natural selection on mitochondrial and nuclear genomes in both humans and fruit flies (Cooper et al., 2015).

Genomic cooperation between these two genomes in eukaryotic cells that differ in inheritance and mutational process has motivated various hypotheses about their molecular and functional coevolution. At the organismal level, co-evolutionary pressures are expected to result in co-adapted sets of mitochondrial and nuclear genotypes that maintain physiological performance. At the molecular level, correlated compensatory substitutions in mitochondrial and nuclear-encoded proteins are expected in order to maintain close physical contact in the OXPHOS system for optimal ATP production as well as to prevent the leakage of free radicals generated during the process. At the phylogenomic level, evolutionary rates for mitochondrial and mitochondria-associated nuclear proteins are expected to be correlated to maintain mitochondrial function despite the stark differences in their respective substitution rates.

Genetic experiments have been used to combine divergent mitochondrial and nuclear genotypes in mismatched hybrids and assess their cellular and organismal effects

in contrast to naturally occurring matched genotypes. Mammalian xenomitochondrial cybrids—cultured cells combining divergent nuclear and mitochondrial genomes—have reduced OXPHOS capacity relative to native combinations, the magnitude of which positively correlates with the amount genetic divergence between the species (Kenyon & Moraes, 1997; McKenzie et al., 2003). Hybrids have reduced oxidative function and fitness at the organismal level both between divergent populations, as in the copepod species *Tigriopus californicus* (Ellison & Burton, 2006, 2008), and between some species such as the parasitoid wasps *Nasonia giraulti* and *N. vitripenni* (Ellison & Burton, 2008; Niehuis et al., 2008). However, mito-nuclear incompatibility between species may not be the rule, as it is not observed in fruit flies (Adrion et al., 2016; Montooth et al., 2010), and introgression of mtDNA between related species is commonly observed (Toews & Brelsford, 2012) .

At the molecular level, the higher mutation rate and smaller N_e of the mitochondrial genome relative to the nuclear genome posits a compensatory model of mitochondrial-nuclear coevolution. Under the compensatory model, deleterious mutations fix in the mitochondrial genome due to the reduced efficacy of purifying selection on the mitochondrial genome. Subsequent compensatory mutations that arise in the nuclear genome and alleviate the deleterious effects are then favored by positive selection. While the compensatory coevolutionary model is intuitively appealing, direct evidence for the model has been hard to come by. Osada & Akashi (2012) provide the most compelling evidence for compensatory molecular coevolution in the nuclear subunits of Complex IV (cytochrome c oxidase) in primates. While they found no evidence for adaptive molecular evolution in the mitochondrial-encoded subunits, the

nuclear-encoded subunits had elevated rates of amino acid substitution specifically in positions that interact with the mitochondrial-encoded subunits. Further, the temporal pattern of amino acid substitutions on the phylogeny revealed that mitochondrial amino acid substitutions preceded nuclear substitutions, lending strong support for the compensatory model with mtDNA as the driver. Over longer periods of evolutionary time, this dynamic is expected to generate correlated rates of evolution at the phylogenomic scale, particularly for those proteins that closely interact in the mitochondria.

My thesis investigates mito-nuclear dynamics at all three of these scales. We use mitonuclear genetics to test for functional effects of mito-nuclear interactions on male reproductive fitness, in the context of both gene-environment interactions and the female-specific selective sieve that operates on mtDNA (Montooth et al., 2019; Chapter 1). We then show that mtDNA, relative to nuclear DNA, harbor more compensated human disease-associated residues (Chapter 2). Finally, we show the proteins that directly interact in the mitochondria have the most highly correlated rates of amino acid substitution at the phylogenomic level across mammals (Chapter 3).

Male *Drosophila melanogaster* exhibit sharp declines in their fertility at high temperatures, while the decline in female fertility in response to high temperatures is more gradual. Furthermore, this heat-induced male sterility is reversible and is correlated with geographical distribution and local climatic conditions. Being primarily energy generators, mitochondria play a significant role in shaping energetically sensitive phenomenon and the thermal male sterility phenotype presents an excellent opportunity to test the role of genetic and environmental variation in shaping the evolution of an

important organismal fitness phenotype. The differences in response to thermal stress between males and females also enables testing hypotheses concerning sex-specific patterns of evolution for the mitochondrial and the nuclear genomes (Montooth et al., 2019). We used six hybrid fruit fly genotypes that combined the mitochondrial and nuclear genomes from *Drosophila melanogaster* and its sister species *D. simulans*. One of these hybrid genotypes harbors a well-characterized genetic incompatibility between a single nucleotide polymorphism in the mitochondrial-encoded mt-tRNA^{Tyr} and an amino acid polymorphism in the nuclear-encoded tyrosyl-tRNA synthetase that disrupts multiple phenotypes, including reduced female fecundity that is exacerbated at higher temperatures (Hoekstra et al., 2013; Zhang et al., 2017). We found that the mitochondrial-nuclear incompatibility negatively affects male fertility, but only when males are developed at high temperatures. This heat-induced male sterility is partially rescued by diet, suggesting an underlying energetic basis.

Amino acid substitutions known to cause disease in humans have been observed to occur as wild type in non-human species (Kondrashov et al., 2002; Waterston et al., 2002). A compensatory epistatic hypothesis is often invoked to explain the existence of these disease-associated mutations in non-humans wherein interacting residues present in the non-human species alleviate the deleterious effects of the mutations. Disorders affecting mitochondrial function are unique as they can result from defects in either or both the mitochondrial and nuclear genomes (Gorman et al., 2016). Mitochondrial diseases thereby represent a dual-genome complex disease and provide us with a unique system to test out compensatory hypotheses for mitochondrial-nuclear co-evolution.

Dysfunction of the OXPHOS system represents the primary cause of mitochondrial disorders, which can result from mutations in the mtDNA-encoded OXPHOS structural subunits or in the nucDNA-encoded OXPHOS structural subunits as well as due to mutations in nucDNA-encoded assembly factors essential for mitochondrial respiratory chain assembly (Figure 2). We focused on mutations in the human OXPHOS structural subunits to investigate the molecular compensation of mitochondrial diseases. By mapping human Disease-Associated Residues (DARs) onto sequence alignments of mammalian OXPHOS proteins, we estimated the proportion of compensated mutations in mtDNA and nucDNA. We further mapped these amino acid mutations onto three-dimensional protein structures to understand the structural basis of their compensation. We found different levels of compensation for the two genomes, with mtDNA harboring a significantly higher proportion of compensated DARs than nucDNA. Compensated DARs across both genomes tend to disrupt protein structure and function to a lesser extent and also exhibit stronger signatures of covariation with their structural neighborhoods.

mtDNA often evolves at a faster rate than nuclear DNA, though the ratio of the mitochondrial over the nuclear mutation rate is highly variable among animal taxa. In nonvertebrates, such as insects and arachnids, the ratio of mtDNA over nucDNA mutation rate varies between 2 and 6, whereas it is above 20, on average, in vertebrates such as birds, mammals and scaled reptiles (Allio et al., 2017). This difference in mutation rates is also reflected in mtDNA exhibiting a faster rate of sequence evolution than nucDNA (Brown et al., 1979). The reduced efficacy of natural selection for mtDNA relative to nucDNA further amplifies this difference. However, given the myriad of ways

in which nuclear gene products interact with the mitochondria, the difference in mtDNA and nucDNA evolutionary rates is likely to be different for different sets of nuclear gene products.

Nuclear gene products that physically interact with the mitochondrial proteins should experience the strongest pressure to "keep up" with the faster evolution of the mitochondrial genome. Nuclear proteins that are indirectly involved in mitochondrial maintenance and function would experience this pressure to a lesser extent followed by nuclear proteins that do not interact with the mitochondria, which would be subject to even weaker pressures. We estimated the evolutionary rates for mtDNA and nucDNA mammalian proteins and computed evolutionary rate correlations to test this hypothesis. For this we defined four categories of genes: mtDNA-encoded OXPHOS genes, nucDNA-encoded OXPHOS genes, mitochondria-associated nucDNA-encoded genes and glycolysis genes. We then estimated the evolutionary rate correlations between the mtDNA-encoded OXPHOS genes and each of the nucDNA-encoded gene categories. We found that evolutionary rate correlations between the mtDNA-encoded and nucDNA-encoded proteins are strongest in nucDNA-encoded proteins that physically interact with mtDNA-encoded products, followed by proteins that are encoded in the nucDNA but function in the mitochondria and finally proteins from the glycolysis pathway which do not interact with the mitochondria showed the weakest correlations. The ERCs are strongest for branches with higher mtDNA substitution rates suggesting faster mtDNA evolution is shaping the coevolutionary dynamics in these proteins.

The energy-processing pathways in some of the most successful organisms on Earth rely on ancient interactions resulting from an endosymbiotic merger and have since

undergone constant refinement for the past 2 billion years. The coevolutionary dynamics between the host and the endosymbiont shape fundamental macro- and micro-evolutionary processes that affect organismal physiology, population dynamics and speciation. We investigate these interactions at three biological scales to elucidate the coevolutionary dynamics between the two cellular genomes. Overall, we find that incompatibilities between the mitochondrial and the nuclear genomes can negatively affect organismal fitness, although these effects are largely context-dependent. We also find that the mitochondrial genome harbors a higher proportion of compensated disease-causing mutations, although this pattern seems to stem from a higher background mutation rate for the mitochondria which might lead to a greater proportion of these mutations finding themselves in a permissive genetic background where their deleterious effects are masked. Finally, we find that higher mitochondrial mutation rates drive a corresponding acceleration in the substitution rates of nuclear proteins that interact with the mitochondria.

FIGURES

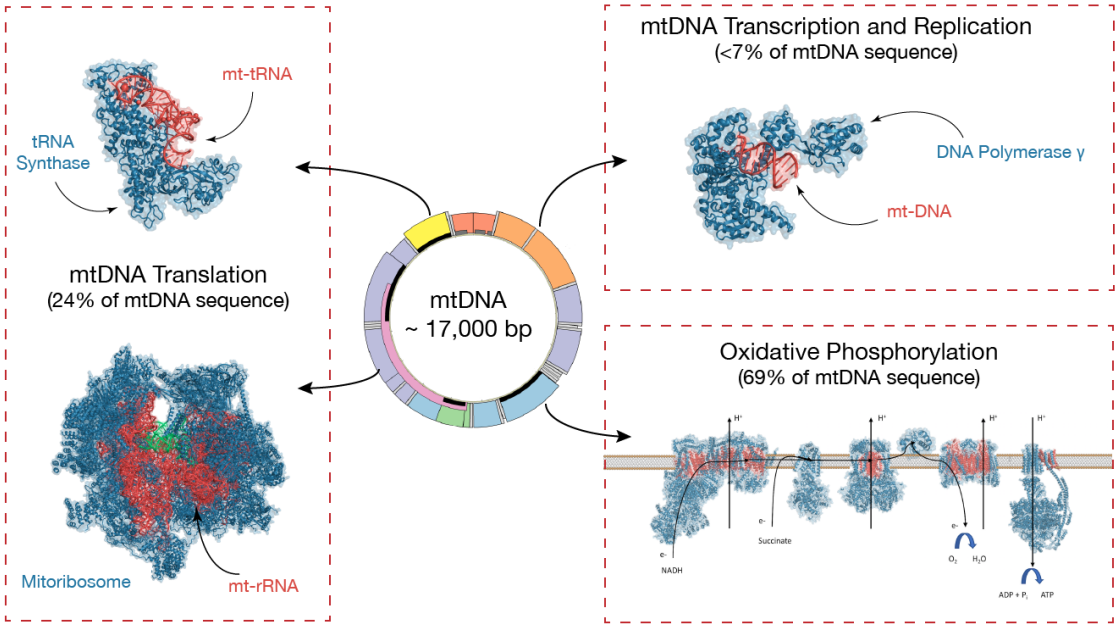

Figure 1. The mitochondrial genome is completely dependent on products of the nuclear genome for its function. The animal mitochondrial genome comprises of only 37 genes - 13 structural genes that are part of the oxidative phosphorylation enzyme complexes in the inner mitochondrial membrane; 22 tRNA genes that interact with nuclear-encoded tRNA synthases and required for mitochondrial translation; 2 structural rRNA genes that form the mitoribosome along with a multitude of nuclear encoded proteins. Further, mitochondrial replication is completely dependent on nuclear-encoded proteins. Figure modified from (Burton & Barreto, 2012).

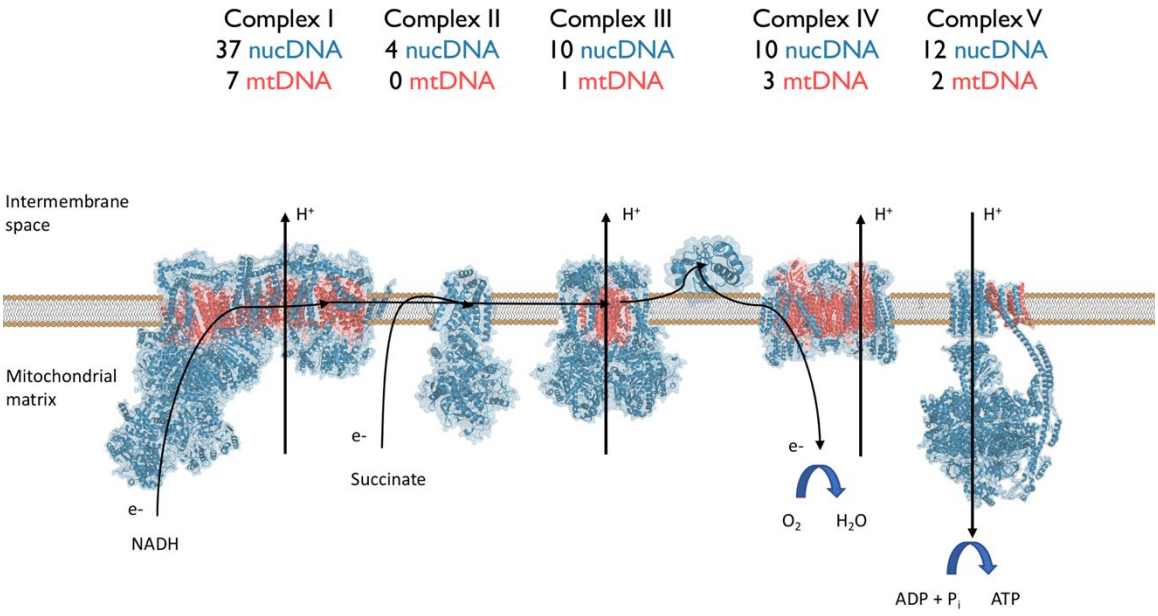

Figure 2. The molecular machinery for mitochondrial oxidative phosphorylation (OXPHOS) comprises of five multi-subunit protein complexes embedded in the inner mitochondrial membrane. These enzyme complexes comprise proteins encoded by both the mitochondrial genome (colored red) and the nuclear genome (colored blue). Complexes I-IV transport electrons from NADH or succinate to molecular oxygen thereby establishing a protein gradient across the inner mitochondrial membrane. Complex V then uses this gradient for ATP synthesis from ADP and inorganic phosphate.

REFERENCES

Adrion, J. R., White, P. S., & Montooth, K. L. (2016). The roles of compensatory evolution and constraint in aminoacyl trna synthetase evolution. *Molecular Biology and Evolution*, *33*(1), 152–161. https://doi.org/10.1093/molbev/msv206

Allio, R., Donega, S., Galtier, N., & Nabholz, B. (2017). Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: Implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Molecular Biology and Evolution*, *34*(March), 2762–2772. https://doi.org/10.1093/molbev/msx197

Brown, W. M., George, M., & Wilson, A. C. (1979). Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA*, 5.

Burton, R. S. (2022). The role of mitonuclear incompatibilities in allopatric speciation. *Cellular and Molecular Life Sciences*, *79*(2), 103. https://doi.org/10.1007/s00018-021-04059-3

Burton, R. S., & Barreto, F. S. (2012). A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Molecular Ecology*, *21*(20), 4942–4957. https://doi.org/10.1111/mec.12006

Cooper, B. S., Burrus, C. R., Ji, C., Hahn, M. W., & Montooth, K. L. (2015). Similar Efficacies of Selection Shape Mitochondrial and Nuclear Genes in Both Drosophila melanogaster and Homo sapiens. *G3 (Bethesda, Md.)*, *5*(October), 1–45. https://doi.org/10.1534/g3.114.016493

Ellison, C. K., & Burton, R. S. (2006). Disruption of mitochondrial function in interpopulation hybrids of Tigriopus californicus. *Evolution; International Journal of Organic Evolution*, *60*(7), 1382–1391. https://doi.org/10.1111/j.0014-3820.2006.tb01217.x

Ellison, C. K., & Burton, R. S. (2008). Interpopulation hybrid breakdown maps to the mitochondrial genome. *Evolution*, *62*(3), 631–638. https://doi.org/10.1111/j.1558-5646.2007.00305.x

Gorman, G. S., Chinnery, P. F., DiMauro, S., Hirano, M., Koga, Y., McFarland, R., Suomalainen, A., Thorburn, D. R., Zeviani, M., & Turnbull, D. M. (2016). Mitochondrial diseases. *Nature Reviews Disease Primers*, *2*(1), 1–22. https://doi.org/10.1038/nrdp.2016.80

Hoekstra, L. A., Siddiq, M. A., & Montooth, K. L. (2013). Pleiotropic Effects of a Mitochondrial-Nuclear Incompatibility Depend upon the Accelerating Effect of Temperature in Drosophila. *Genetics*, *195*(3), 1129–1139. https://doi.org/10.1534/genetics.113.154914

Kenyon, L., & Moraes, C. T. (1997). Expanding the functional human mitochondrial DNA database by the establishment of primate xenomitochondrial cybrids. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(17), 9131–9135.

Kondrashov, A. S., Sunyaev, S., & Kondrashov, F. A. (2002). Dobzhansky-Muller incompatibilities in protein evolution. *Proc Natl Acad Sci U S A*, *99*(23), 14878–14883. https://doi.org/10.1073/pnas.232565499

McKenzie, M., Chiotis, M., Pinkert, C. A., & Trounce, I. A. (2003). Functional respiratory chain analyses in murid xenomitochondrial cybrids expose coevolutionary constraints of cytochrome b and nuclear subunits of complex III. *Molecular Biology and Evolution*, *20*(7), 1117–1124. https://doi.org/10.1093/molbev/msg132

Montooth, K. L., Dhawanjewar, A. S., & Meiklejohn, C. D. (2019). Temperature-Sensitive Reproduction and the Physiological and Evolutionary Potential for Mother's Curse. *Integrative and Comparative Biology*, *59*(4). https://doi.org/10.1093/icb/icz091

Montooth, K. L., Meiklejohn, C. D., Abt, D. N., & Rand, D. M. (2010). Mitochondrial-nuclear epistasis affects fitness within species but does not contribute to fixed incompatibilities between species of Drosophila. *Evolution; International Journal of Organic Evolution*, *64*(12), 3364–3379. https://doi.org/10.1111/j.1558-5646.2010.01077.x

Montooth, K. L., & Rand, D. M. (2008). The Spectrum of Mitochondrial Mutation Differs across Species. *PLoS Biology*, *6*(8), e213. https://doi.org/10.1371/journal.pbio.0060213

Nachman, M. W. (1998). Deleterious mutations in animal mitochondrial DNA. *Genetica*, *102–103*(1–6), 61–69. https://doi.org/10.1023/A:1017030708374

Niehuis, O., Judson, A. K., & Gadau, J. (2008). Cytonuclear genic incompatibilities cause increased mortality in male F2 hybrids of Nasonia giraulti and N. vitripennis. *Genetics*, *178*(1), 413–426. https://doi.org/10.1534/genetics.107.080523

Osada, N., & Akashi, H. (2012). Mitochondrial-nuclear interactions and accelerated compensatory evolution: Evidence from the primate cytochrome C oxidase complex. *Molecular Biology and Evolution*, *29*(1), 337–346. https://doi.org/10.1093/molbev/msr211

Rand, D. M. (2005). Mitochondrial Genetics of Aging: Intergenomic Conflict Resolution. *Science of Aging Knowledge Environment*, *2005*(45). https://doi.org/10.1126/sageke.2005.45.re5

Rand, D. M., Haney, R. a, & Fry, A. J. (2004). Cytonuclear coevolution: The genomics of cooperation. *Trends in Ecology & Evolution*, *19*(12), 645–653. https://doi.org/10.1016/j.tree.2004.10.003

Rand, D. M., & Kann, L. M. (1996). Excess amino acid polymorphism in mitochondrial DNA: Contrasts among genes from Drosophila, mice, and humans. *Molecular Biology and Evolution*, *13*(6), 735–748. https://doi.org/10.1093/oxfordjournals.molbev.a025634

Rand, D. M., & Kann, L. M. (1998). Mutation and selection at silent and replacement sites in the evolution of animal mitochondrial DNA. *Genetica*, *102–103*(0), 393–407. https://doi.org/10.1023/A:1017006118852

Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, *21*(16), 3907–3930. https://doi.org/10.1111/j.1365-294X.2012.05664.x

Wallace, D. C. (2016). Mitochondrial DNA in evolution and disease. *Nature*, *535*(7613), 498–500. https://doi.org/10.1038/nature18902

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., … Lander, E. S. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, *420*(6915), 520–562. https://doi.org/10.1038/nature01262

Zhang, C., Montooth, K. L., & Calvi, B. R. (2017). Incompatibility between mitochondrial and nuclear genomes during oogenesis results in ovarian failure and embryonic lethality. *Development*, *144*(13), 2490–2503. https://doi.org/10.1242/dev.151951

CHAPTER 1

TEMPERATURE-SENSITIVE REPRODUCTION AND THE PHYSIOLOGICAL AND EVOLUTIONARY POTENTIAL FOR MOTHER'S CURSE

This chapter represents the contents of Kristi L. Montooth, Abhilesh S. Dhawanjewar and Colin D. Meiklejohn; Temperature-Sensitive Reproduction and the Physiological and Evolutionary Potential for Mother's Curse in *Integrative and Comparative Biology*, volume 59, number 4, pp. 890–899 doi:10.1093/icb/icz091

ABSTRACT

Strict maternal transmission of mitochondrial DNA (mtDNA) is hypothesized to permit the accumulation of mitochondrial variants that are deleterious to males but not females, a phenomenon called mother's curse. However, direct evidence that mtDNA mutations exhibit such sexually antagonistic fitness effects is sparse. Male-specific mutational effects can occur when the physiological requirements of the mitochondria differ between the sexes. Such male-specific effects could potentially occur if sex-specific cell types or tissues have energy requirements that are differentially impacted by mutations affecting energy metabolism. Here we summarize findings from a model mitochondrial–nuclear incompatibility in the fruit fly *Drosophila* that demonstrates sex-biased effects, but with deleterious effects that are generally larger in females. We present new results showing that the mitochondrial–nuclear incompatibility does negatively affect male fertility, but only when males are developed at high temperatures.

The temperature-dependent male sterility can be partially rescued by diet, suggesting an energetic basis. Finally, we discuss fruitful paths forward in understanding the physiological scope for sex-specific effects of mitochondrial mutations in the context of the recent discovery that many aspects of metabolism are sexually dimorphic and downstream of sex-determination pathways in *Drosophila*. A key parameter of these models that remains to be quantified is the fraction of mitochondrial mutations with truly male-limited fitness effects across extrinsic and intrinsic environments. Given the energy demands of reproduction in females, only a small fraction of the mitochondrial mutational spectrum may have the potential to contribute to mother's curse in natural populations.

INTRODUCTION

The evolutionary potential for a female-specific selective sieve on mitochondrial mutations

Strict maternal inheritance of the mitochondrial genome (mtDNA) has the potential to function as a sex-specific selective sieve, in which the fate of mitochondrial mutations is governed solely by their selective effects in females (Frank & Hurst, 1996; Gemmell et al., 2004; Lewis, 1941). Mitochondrial mutations that are neutral in females but deleterious in males can fix in a population by drift, while sexually antagonistic mutations that benefit females and are deleterious in males will be fixed by positive selection. Frank & Hurst, 1996 offered this population genetic explanation, later termed mother's curse (Gemmell et al., 2004), for the observation that some human diseases linked to mitochondrial mutations tend to disproportionately affect males.

If, as population genetics dictates, natural selection is powerless to halt the decline of male-specific mtDNA functions, what prevents species from extinction due to mtDNA-linked male sterility? The hypothesis that has received the most attention proposes that compensatory evolution at nuclear loci restores male function (Ågren et al., 2019; Frank, 1989; Frank & Hurst, 1996). Mitochondrial replication, transcription, translation, and oxidative phosphorylation (OXPHOS) require hundreds of gene products encoded in the nuclear genome that, in theory, could reverse mother's curse. Predicting the significance and dynamics of a female-specific selective sieve on mitochondrial mutations therefore requires understanding the population genetics of interactions between alleles at mtDNA and nuclear loci (hereafter mito–nuclear interactions). Theory predicts that mito–nuclear polymorphisms will rarely be maintained in populations (Clark, 1984; Gregorius & Ross, 1984), with the exception of X-linked alleles, as the coupled transmission of the X chromosome and the mtDNA in females can sustain mito–nuclear fitness variation, particularly when this variation is sexually antagonistic (Rand et al., 2001). Supporting this theory, work in the fruit fly *Drosophila* has shown that X chromosomes from a single population generate sexually antagonistic fitness variation when combined with different mtDNAs (Montooth et al., 2010; Rand et al., 2001).

If male-detriment mtDNAs fix in populations, then recovery of male fitness could require subsequent evolution at nuclear-encoded loci. Simulations predict that the male-specific mtDNA mutational load arising from this dynamic could be substantial and exceed mutational load across the nuclear genome (Connallon et al., 2018) and compensatory nuclear alleles that alleviate the effects of mother's curse mutations are predicted to be enriched on the Y chromosome (Ågren et al., 2019). However, population

and transmission genetic parameters such as the degree of heteroplasmy, inbreeding, and even low levels of paternal mtDNA transmission can reduce the scope for a female-specific mtDNA selective sieve. For example, moderate inbreeding, where sisters mate with brothers with whom they share a mtDNA, is predicted to reverse mother's curse by exposing mutations with deleterious effects in males to natural selection (Unckless & Herren, 2009; Wade & Brandvain, 2009). Thus, while male-detrimental mutations in mtDNA can potentially accumulate in populations, this potential may vary across populations and species as a function of population genetic parameters such as the level of inbreeding. Nonetheless, across all populations the prevalence of male-detriment mtDNA polymorphisms will ultimately depend on the scope for mtDNA mutations to have male-specific or sexually antagonistic fitness effects.

Empirical support for a female-specific selective sieve

Direct empirical support for a female-specific selective sieve that permits the accumulation of male deleterious mitochondrial mutations comes from the discovery of mitochondrial polymorphisms with male-specific effects segregating within populations. Male-sterile mtDNAs have been recovered from both wild and mutagenized *Drosophila* (Clancy, 2008; Clancy et al., 2011; Patel et al., 2016; Xu et al., 2008). Quantitative effects of mtDNAs on male fertility and sperm competition have been also been measured in fruit flies (Camus & Dowling, 2018; Yee et al., 2013). Male-sterile mtDNAs can segregate cryptically in populations if their sterility effects are masked by nuclear alleles that restore male function (Clancy, 2008; Clancy et al., 2011), a dynamic that has been well documented to underlie cytoplasmic male sterility in plants (Budar et al., 2003;

Case et al., 2016; Delph et al., 2007; Frank, 1989; Lewis, 1941). However, mtDNA variants that affect only males may be the exception, and their discovery and characterization in animals have thus far been limited to a handful of organisms, including *Drosophila*. In the marine copepod *Tigriopus californicus*, mito–nuclear interactions compromise both female and male reproductive fitness in inter-population hybrids (Willett, 2008). In humans, the evidence supporting stronger effects of mitochondrial disease mutations in males versus females is sparse (Beekman et al., 2014). While mitochondrial mutations were originally implicated in human male infertility with no detected effects in females (Moore & Reijo-Pera, 2000; Ruiz-Pesini et al., 2000), these findings have been called into question, and it remains unclear the extent to which male infertility is associated with mitochondrial haplotype (Mossman et al., 2012). Even for human mitochondrial diseases where male-biased effects have been documented (e.g., Leber's Hereditary Optic Myopathy), it is not clear whether this is caused by interactions with recessive X-linked alleles that are exposed in hemizygous males or because of male-specific effects of the mitochondrial mutations (Beekman et al., 2014; Chinnery & Schon, 2003). Thus, while mtDNA variants with the sex-specific selective effects required for mother's curse do exist, whether they are exceptional or the norm remains an open question (Dowling & Adrian, 2019).

Several patterns of genome evolution have been proposed as indirect evidence of mother's curse. In *Drosophila*, many nuclear-encoded mitochondrial genes are part of gene families, with evolutionarily young paralogs that are expressed solely in testes (Gallach et al., 2010). These gene duplication events are hypothesized to reverse mother's curse by providing a nuclear target for mutations with effects that are restricted

to males and can be acted on by selection (Gallach & Betrán, 2011). Although this model is intriguing, to date no evidence beyond the existence and expression pattern of these duplicates supports or refutes their role in reversing mother's curse. Indeed, this pattern is not evident in humans (Eslamieh et al., 2017), and there is no molecular signature of recurrent positive selection on male-specific OXPHOS paralogs in *Drosophila* (Havird & McConie, 2019) — a signature that might be expected if molecular changes at these gene duplicates restored male fitness. Additionally, the *Drosophila* genome harbors many evolutionarily young paralogs with testis-specific expression that do not have mitochondrial function (Belote & Zhong, 2009; White-Cooper, 2010). A model of nuclear restoration of sexually antagonistic effects of mtDNA mutations has also been proposed to favor the movement of nuclear genes with mitochondrial function off of X chromosomes, which spends two-thirds of their time in females and are co-transmitted with the mtDNA (Drown et al., 2012). However, this pattern of chromosomal distribution of nuclear genes with mitochondrial function is taxonomically limited (Ågren et al., 2019; Dean et al., 2014; Drown et al., 2012; Hough et al., 2014) and could be driven by other phenomenon involving sex chromosomes such as dosage compensation. Thus, it remains unclear the extent to which mother's curse mutations accumulate in natural populations and generally impact genome evolution.

Sex-specific effects of a model mito–nuclear incompatibility in Drosophila

Sex-specific and sexually antagonistic fitness effects of mtDNA mutations may result from 1) sex-specific effects on physiology and development that impact fitness or 2) sex-specific fitness consequences of mtDNA mutations with similar physiological

effects in males and females. The fruit fly *Drosophila melanogaster* has emerged as a powerful model for manipulating natural mtDNA and nucDNA variation to test for the presence of sexually antagonistic effects on life-history traits related to fitness (Camus et al., 2012, 2015; Camus & Dowling, 2018; Montooth et al., 2010; Rand, 2001). Here we synthesize results from studies dissecting a model mito–nuclear incompatibility in *Drosophila* across different levels of biological organization to understand its potential to generate sex-specific fitness effects.

Mito–nuclear coevolution to maintain mitochondrial function is predicted to result in the accumulation of mito–nuclear incompatibilities between divergent populations and closely related species. We have previously identified, mapped, and functionally characterized a specific mito–nuclear incompatibility between a mitochondrial polymorphism in the mt-tRNA^{Tyr} from *Drosophila simulans* and a nuclear amino acid polymorphism in the mitochondrial-targeted tyrosine aminoacyl tRNA synthetase protein from *D. melanogaster* (Fig. 1). Each polymorphism has little to no effect on its own, but the combination of these polymorphisms in individuals with the *(simw^{501}); OreR* genotype produces a suite of deleterious effects due to compromised mitochondrial protein synthesis. Adult male and female flies with the *(simw^{501}); OreR* genotype have decreased activity of OXPHOS complexes that require mitochondrial protein translation (Complexes I, III, IV, and V of the electron transport chain), relative to control genotypes (Meiklejohn et al., 2013). Embryonic survival is compromised in this genotype, but surviving embryos have normal larval and pupal mortality, although larval development is severely delayed and this delay is exacerbated in environments that normally accelerate growth and increase energy demand (Buchanan et al., 2018; Hoekstra et al., 2013, 2018).

The ability to grow and accumulate the energy stores required for metamorphosis in spite of compromised OXPHOS activity is enabled by compensatory upregulation of the TCA cycle, glycolytic ATP production, and respiration rates in larvae with this mito-nuclear incompatible genotype (Hoekstra et al., 2013; Matoo et al., 2019).

The compensatory physiological changes that facilitate development in larvae with the mito–nuclear incompatible genotype are associated with fitness costs later in life. Relative to genetic controls, larvae with this mito–nuclear incompatibility accumulate greater levels of hydrogen peroxide, a reactive oxygen species, and have lower mitochondrial membrane potentials (Matoo et al., 2019). We hypothesize that the lower membrane potential may result from mitochondrial uncoupling as a physiological defense mechanism to prevent additional free radical production (Matoo et al., 2019). While adults with this mito–nuclear incompatibility have normal metabolic rates, they suffer a number of other defects associated with fitness. They have smaller and more brittle sensory bristles (Meiklejohn et al., 2013), decreased male and female survival when infected with a natural bacterial pathogen (Buchanan et al., 2018), and compromised female fecundity (Hoekstra et al., 2018; Meiklejohn et al., 2013; Zhang et al., 2017).

Females with the mito–nuclear incompatibility suffer reduced fecundity at 25°C that is exacerbated by stress. Females with the *(simw^{501}); OreR* genotype that survive bacterial infection have significantly decreased fecundity relative to their sham-infected sisters, revealing a potential trade-off between immunity and fecundity that is not observed in control genotypes (Buchanan et al., 2018). Females with the *(simw^{501}); OreR* genotype that develop at 28°C are sterile, while control genotypes maintain fertility at this temperature. The temperature-dependent sterility in females with the *(simw^{501}); OreR*

genotype results from the combined effects of compromised ovarian development, loss of germline stem cells, and possible underprovisioning of embryos (Zhang et al., 2017). Males have normal fecundity at 25°C (Hoekstra et al., 2018), but our inability to culture this mito–nuclear genotype at 28°C when using either female or males of this genotype as parents suggested that there may also be temperature-dependent male sterility caused by this mito–nuclear incompatibility (Hoekstra et al., 2013). While 28°C is far from critical thermal maxima in adult *D. melanogaster* (Hoffmann, 2010; Sgrò et al., 2010), development at 28°C is significantly accelerated and nearer to thermal limits (~ 32°C) for robust egg-to-adult viability (Petavy et al., 2001). We have hypothesized that this increased rate of development may place significant demand on organismal energy supply systems (Hoekstra et al., 2018).

Here we report new experiments that revealed a temperature-dependent effect of the mito–nuclear incompatibility on male fertility, similar to that previously observed in females. Temperature dependent sterility in males with the mito–nuclear incompatibility is due to defects in spermatogenesis rather than in sperm function at high temperatures and can be partially rescued by diet, as might be expected if the sterility had an energetic basis.

MATERIALS AND METHODS

We used six genotypes that combined mtDNA from *D. melanogaster* (mtDNA: *ore*) and *D. simulans* (mtDNA: *sm21, simw*^{501}) with two inbred, wildtype nuclear genomes from *D. melanogaster* (nucDNA: *OreR* and *Aut*) and included the *(simw*^{501}*); OreR* mito–nuclear incompatible genotype. These genotypes enabled us to test for

phenotypic effects of an interaction between genetic variation in the mtDNA and nuclear genomes (Fig. 1). The *sm21* mtDNA differs from the *ore* mtDNA at over 600 nucleotide sites, but differs from the *simw*^{501} mtDNA at only six sites in the coding region of the mtDNA that include the tRNA mutation that causes the mito–nuclear incompatibility (Meiklejohn et al., 2013). When the phenotypes of individuals with the *sm21* mtDNA and *ore* mtDNA are similar, but differ from individuals with the *simw*^{501} mtDNA and the *OreR* nuclear background, this suggests that the characterized mito–nuclear interaction underlies the phenotypic effect.

All genotypes were maintained at a permissive room temperature (~ 22°C). Twenty-five males and 25 females from each genotype were allowed to mate and oviposit for 6h at 25°C and then were removed from the vials. All genotypes were robustly fertile when developed at 25°C (Buchanan et al., 2018; Hoekstra et al., 2018). Eggs hatched and larvae developed at 25°C until the early third-instar, when they were transferred to 28°C. Upon eclosion, five males from each genotype were individually housed with three virgin females from an outbred population. The outbred population was generated by combining isofemale genetic strains—two from Ithaca, NY, two from the Netherlands, and three from Zimbabwe. Females of this outbred population were robustly fertile at 22-28°C and used for all experiments. Experimental males were transferred to new vials with three new virgin females every 2 days, and the old vials were moved to 25°C to minimize the effect of temperature on the development of any offspring. This was repeated five times over a10-day period, giving each experimental male ample opportunity to demonstrate fertility. Fertility for each male was measured as the average number of offspring sired per vial to account for males who died before the

end of the 10 days. Because offspring inherited their mtDNA and half of their nuclear alleles from their outbred mothers, differences in the number of offspring can be attributed to fertility of the paternal genotype, and not effects of the mito–nuclear incompatibility on offspring survival.

To test for an effect of diet, we used a standard maltose diet and three approximately isocaloric diets that differed in their protein:carbohydrate ratios (High P:C = 452 kJ/100 g, 7.1% protein, 17.9% carbohydrate; Equal P:C = 456 kJ/100 g, 4.3% protein, 21.2% carbohydrate; Low P:C = 469 kJ/100 g, 2.5% protein, 24.6% carbohydrate; values from Matzkin et al., 2011). The relative abundances of protein and carbohydrates were achieved by altering the ratio of yeast:sucrose in the diet (Table1). All genotypes were raised on the experimental diets for at least two generations before the fertility trials. We repeated these experiments across three blocks for a sample size of 15 males per genotype per diet. The average reproductive output for each genotype on each diet was calculated by first averaging the number of offspring sired by each male across the number of time points the male was alive, and then averaging across the 15 replicate males. The data were analyzed using linear models that tested for the fixed effects of mtDNA, nuclear genome, diet, and the interactions between these factors.

RESULTS

Temperature-dependent male sterility effects of a mito–nuclear incompatibility

At 25°C, males with the mito-nuclear incompatibility did not have compromised fertility and there was no significant effect of the mito–nuclear interaction on fertility (Fig.2A; Hoekstra et al., 2018). However, when males were developed at 28°C there was

a significant effect of the mito–nuclear interaction on male fertility (Table2; mtDNA × nuclear, $P < 0.0001$). Males with the *(simw^{501}); OreR* incompatibility had greatly reduced fertility, relative to their nuclear genetic controls (Fig.2B). This temperature-dependent loss of fertility was not due to impaired sperm function at 28°C; males developed at 25°C and mated at 28°C had robust fertility (Table3). Supporting the hypothesis that this temperature-sensitive effect on male fertility has an energetic basis, diet modified both fertility generally (diet, $P = 0.019$) and modified the effects of the mtDNA (mtDNA × diet, $P = 0.025$) (Table4). Increasing protein relative to carbohydrates in the diet tended to have a positive effect on fertility, and also produced the largest rescue of temperature-dependent sterility in males with the mito–nuclear incompatibility (Fig.3).

A striking pattern in this model system is the strong sexually antagonistic effect of the nuclear genotype on female and male reproductive fitness. Females with the *Aut* nuclear genotype have much higher fecundity relative to females with the *Ore* nuclear genotype (Fig.1B; Hoekstra et al., 2018; Meiklejohn et al., 2013). In contrast, males with the *Aut* nuclear genotype have lower fertility than males with the *Ore* nuclear genotype (Fig.2A; Hoekstra et al., 2018). Males with the *Aut* nuclear genotype were also particularly thermally sensitive, showing almost complete sterility when developed at 28°C (Fig.2B). Thus, male flies with the *Aut* nuclear background were sterile at 28°C due to effects of the nuclear background, while male flies with the *(simw^{501}); OreR* incompatibility are sterile at 28°C due to the mito–nuclear interaction in this genotype. In contrast, females with the *(ore); Ore* genotype did not show temperature-dependent sterility (Zhang et al., 2017). Thus, temperature appears to magnify the sexually

antagonistic effect of the nuclear genome, but through temperature-dependent sterility in males rather than females.

DISCUSSION

We found that an incompatible mito–nuclear genotype generates sterility in both sexes as a result of exposure to elevated temperatures during larval development. While there may be differences in the underlying causes of female and male sterility in this system, suggesting some scope for sex-specific physiological outcomes in gametogenesis, the broader pattern is that the compromised energetics caused by this mito–nuclear genotype negatively impacts both female and male reproduction when energy demand exceeds supply. In females, demand appears to exceed supply even at normal developmental temperatures. In males, reproductive capacity is compromised when demand exceeds supply during development and spermatogenesis at higher temperatures.

Moving forward: the physiological potential for a female-specific selective sieve

The shared role of mitochondria in females and males to supply energy via OXPHOS would seem to provide limited scope for sex-specific effects of mtDNA mutations on physiological performance and fitness. However, recent advances in developmental biology reveal that there may be substantial sexual dimorphism in metabolism and development of organs that have historically been viewed as sexually monomorphic (Millington & Rideout, 2018). For example, gut physiology and intestinal stem cell proliferation differs between the sexes, is influenced by juvenile hormone, and

can affect lipid metabolism and adult reproductive output (Hudry et al., 2016; Reiff et al., 2015). Given the role of the gut in nutrient absorption and energy homeostasis, sexual dimorphism in gut physiology and development may cause genetic variation in energy metabolism to have different physiological consequences in males and females. Some of the sexual dimorphism in fruit fly physiology and development is directly regulated by sex-determination genes (Hudry et al., 2016; Millington & Rideout, 2018; Rideout et al., 2015; Sawala & Gould, 2017). For example, activity of the sex-determination gene transformer in the larval fat body regulates differences in body size between males and females via insulin signaling (Rideout et al., 2015), which influences rates of larval growth that establish sexual dimorphism in adult body size (Okamoto et al., 2013; Sawala & Gould, 2017; Testa et al., 2013). Particularly relevant to mitochondrial function, males and females differ in their oxidative stress biology (Pomatto et al., 2017). Sexual dimorphism established early in development and persisting in adult organ homeostasis and cellular maintenance has clear potential to guide investigation of the physiological potential for mitochondrial mutations to have sex-specific effects.

The genetically and developmentally independent processes of oogenesis and spermatogenesis may lend themselves to sex-specific effects of mtDNA mutations, particularly if one of these processes tends to be more sensitive to mutations affecting energy metabolism. Frank & Hurst, (1996) suggested that male reproduction may be particularly susceptible to the effects of mtDNA mutations; for example, in mammals, mitochondrial function is thought to provide energy for swimming sperm. However, the extent to which glycolysis and OXPHOS share the labor of ATP production is an active area of investigation in mammals (du Plessis et al., 2015; Mukai & Okuno, 2004) and

appears to vary among the insects that have been studied (Werner & Simmons, 2008). A division of labor proposed by (du Plessis et al., 2015) in which OXPHOS-generated ATP is used for mammalian sperm development, maturation, and some forms of motility, while glycolysis-generated ATP is used for hyperactivated motility, capacitation and the acrosome reaction, is consistent with our findings in *Drosophila*; male flies with the mito–nuclear incompatibility that developed at 25°C produce sperm that function at 28°C, while males that develop at 28°C, a temperature that enhances the OXPHOS energetic defect (Hoekstra et al., 2013), develop sperm that do not successfully fertilize females even when those fertilized females are placed at 25°C. Female reproduction is also energetically costly; a meta-analysis across a wide range of animal taxa estimated that, on average, females have a gamete biomass production rate that is approximately two to four orders of magnitude higher than males (Hayward & Gillooly, 2011), suggesting that the energy demands of gametogenesis are usually greater in females than in males. In both male and female *Drosophila*, gametogenesis is sensitive to energy and nutrient availability, and restricting dietary protein slows germline cell proliferation which is reversible upon restoring dietary protein in both sexes (Drummond-Barbosa & Spradling, 2001; McLeod et al., 2010; Yang & Yamashita, 2015). Thus, while gametogenesis may be the most likely aspect of animal biology to show sex-specific effects of mtDNA mutations, it remains unclear whether males will more often be negatively impacted than females. The significance of the mother's curse hypothesis ultimately rests on the proportion of mtDNA mutations that affect male reproductive fitness with no deleterious effects in females.

Even if mitochondrial variation has similar effects on male and female physiology or gamete production, this variation could nonetheless generate sexually antagonistic fitness effects if males and females differ significantly in their life history or ecology. For example, sexual dimorphism in dispersal patterns, in the intensity of intrasexual competition for resources, territories, or mates, and in the energetic investment in life-history traits, can all potentially impact the extent to which female or male fitness is determined by genetic variation in energy metabolism. An energetic framework that considers energy supply-demand balance in the context of development, physiology, life history, and ecology may be a powerful approach to predict where on the tree of life we expect significant accumulation of male-detrimental mitochondrial mutations that make it through the female-specific selective sieve. mtDNA and mito-nuclear genetic effects are often environmentally sensitive (Hoekstra et al., 2013; Mossman et al., 2016; Patel et al., 2016; Willett & Burton, 2003; Zhu et al., 2014), and extrinsic and intrinsic conditions that elevate energy demand are expected to reveal effects of mutations that compromise energy supply (Hoekstra et al., 2018). An important question is whether mtDNA mutations with male-specific effects remain male-specific across a range of developmental stages and ecological conditions. Thus, integrative physiology and eco-physiology have an important role to play in our growing understanding of mito–nuclear evolutionary and ecological dynamics.

ACKNOWLEDGMENTS

We thank David Rand for his pioneering contributions to the field of mito–nuclear ecology and for supporting this work from its inception. We thank the organizers

and speakers of the symposium, the Dowling lab and Jim Mossman for their insights into mother's curse in *Drosophila*.

FUNDING

This work was supported by funding from NSF IOS CAREER award 1149178, NSF EPSCoR Track 2 award 1736249, and a Rosemary Grant Award from the Society for the Study of Evolution. Participation in the symposium was supported by funding from the National Science Foundation [IOS-1839203]; the Company of Biologists [EA1694]; the Crustacean Society; and the SICB Divisions of Comparative Physiology and Biochemistry, Invertebrate Zoology, and Phylogenetics and Comparative Biology.

FIGURES AND TABLES

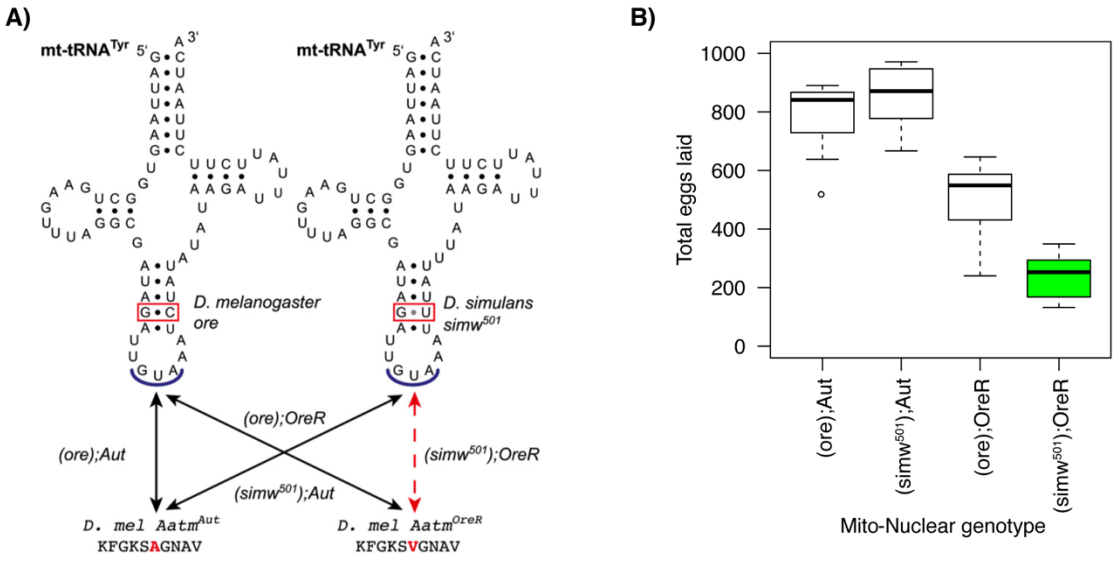

Figure 1. A model of mito–nuclear incompatibility in fruit flies arises from interactions between a mtDNA polymorphism in the anticodon stem of the mt-tRNA^{Tyr} and a nuclear amino acid polymorphism in the mitochondrially targeted aminoacyl tRNA^{Tyr} synthetase that charges the mt-tRNA^{Tyr} for mitochondrial protein synthesis. A) The mitochondrial and nuclear SNP genotypes of the six strains used in this study. The (simw^{501}); OreR genotype combines incompatible mitochondrial and nuclear SNPs, while the other (mtDNA); Nuclear genotypes serve as genetic controls. Figure modified from (Hoekstra et al. 2013) with permission from the Genetics Society of America. B) Females with the mito–nuclear incompatible genotype, highlighted in green, lay significantly fewer eggs at 25°C. The y-axis is total eggs laid by a female over 10 days. This figure is from Hoekstra et al. (2018).

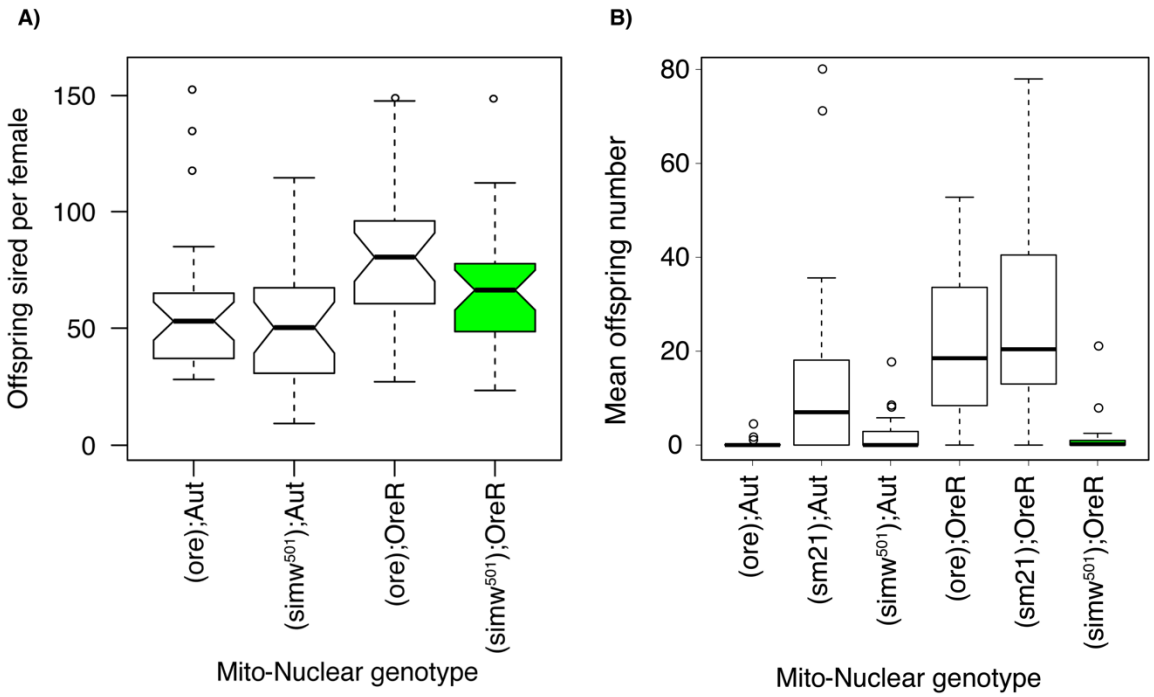

Figure 2. Males with the mito-nuclear incompatible genotype, highlighted in green, have wild-type fertility at 25°C, but are sterile at 28°C. **A)** When males are developed at 25°C. there was no evidence that mito-nuclear genotype affected male offspring production. The y-axis is the number of offspring sired by individual males divided by the number of females (up to three) that produced offspring after 48h of mating, averaged across 28-29 replicate males of each genotype. Figure from Hoesktra et al. (2018). **B)** When males were developed at 28°C, there was a significant mito-nuclear effect on fertility, with ($simw^{501}$); OreR males producing almost no offspring. The y-axis is the number of offspring sired by invidual males that were given two new females every 2 days for 10 days, averaged across 15 replicate males of each genotype.

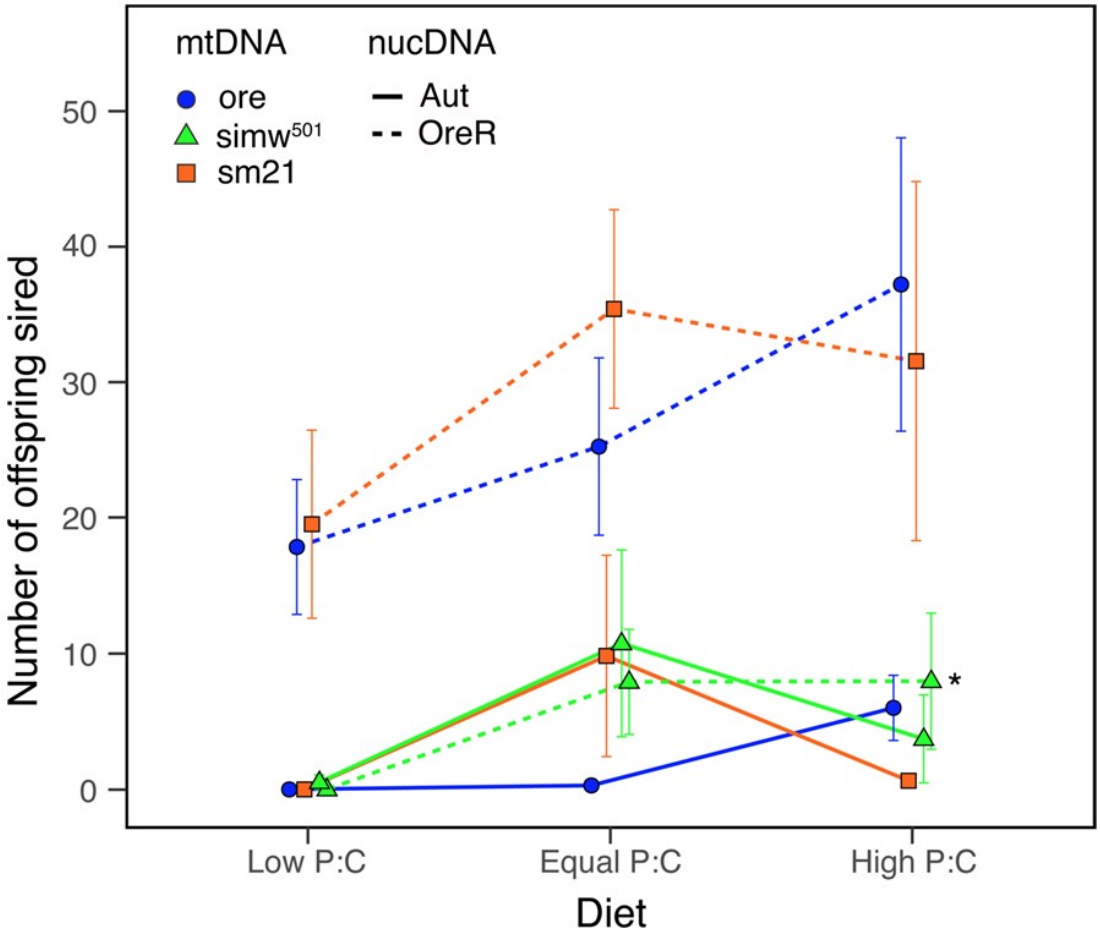

Figure 3. The protein:carbohydrate (P:C) ratio affects temperature-dependent sterility in males reared at 28°C on three approximately isocaloric diets. Increasing P:C in the diet was able to rescue some fertility in males of genotypes that were sterilized by development at 28°C, including males with the ($simw^{501}$); OreR mito–nuclear incompatible genotype, highlighted with an asterisk. The y-axis is the number of offspring sired by individual males that were given two new females every 2 days for 10 days, averaged across 15 replicate males of each genotype. Error bars are standard errors.

Table 1. Isocaloric diets varied in protein:carbohydrate (P: C) via differences in yeast:sucrose content (based on diets from Matzkin et al., 2011))

Ingredient	**Standard diet**	**High P:C diet**	**Equal P:C diet**	**Low P:C diet**
Agar (g)	1.86	1	1	1
Torula Yeast (g)	16.6	32	20	8
Cornmeal (g)	20	9	9	9
Sucrose (g)	-	8	20	32
Molasses (mL)	9.3	-	-	-
Tegosept (g)	0.6	0.45	0.45	0.45
95% ethanol (mL)	3.3	4.5	4.5	4.5
Propionic acid (mL)	1.3	-	-	-
Distilled water (mL)	200	200	200	200

Table 2. Analysis of variance of genetic effects on the number of offspring sired by males developed at 28°C on the standard diet

Factor	**$\chi 2$**	**Num DF**	***P*-value**
mtDNA	43.652	2	<0.0001
nucDNA	30.513	1	<0.0001
mtDNA:nucDNA	20.482	2	<0.0001

Table 3. The proportion of males that sired more than 100 offspring when reared at 25°C and transferred to 28°C for mating on the standard diet

Genotype[a]	**Days 1-2**	**Days 3-4**	**Days 5-6**	**Days 7-8**	**Days 9-10**
(sm21);Aut	1	1	0.71	0.14	0.14
(sm21);OreR	1	1	0.625	0.25	0
(ore);Aut	1	0.56	0.29	0	0
(ore);OreR	0.9	0.67	0.75	0.125	0
(simw^{501});Aut	0.9	0.67	0.56	0	0
(simw^{501});OreR	1	0.9	0.56	0	0

[a] Males of experimental genotypes and females from an outbred wild-type population were reared at 25°C and transferred to 28°C for mating. Individual males were housed with three virgin females for 2 days, and then transferred to new vials with three new virgin females, for a total for 10 days. After males were removed from the vials, mated females and eggs were returned to 25°C and vials were scored for fertility.

Table 4. Analysis of variance of genetic and diet effects on the number of offspring sired by males developed at 28°C on three approximately isocaloric diets

Factor	**$\chi 2$**	**Num DF**	***P*-value**
mtDNA	24.817	2	<0.0001
nucDNA	114.242	1	<0.0001
Diet	7.955	2	0.0187
mtDNA:nucDNA	52.633	2	<0.0001
mtDNA:Diet	11.168	4	0.0247
nucDNA:Diet	0.159	2	0.9233
mtDNA:nucDNA:Diet	3.377	4	0.4967

REFERENCES

Ågren, J. A., Munasinghe, M., & Clark, A. G. (2019). Sexual Conflict through Mother's Curse and Father's Curse. *Theoretical Population Biology*, *129*, 9–17. https://doi.org/10.1016/j.tpb.2018.12.007

Beekman, M., Dowling, D. K., & Aanen, D. K. (2014). The costs of being male: Are there sex-specific effects of uniparental mitochondrial inheritance? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1646), 20130440. https://doi.org/10.1098/rstb.2013.0440

Belote, J. M., & Zhong, L. (2009). Duplicated proteasome subunit genes in Drosophila and their roles in spermatogenesis. *Heredity*, *103*(1), 23–31. https://doi.org/10.1038/hdy.2009.23

Buchanan, J. L., Meiklejohn, C. D., & Montooth, K. L. (2018). Mitochondrial Dysfunction and Infection Generate Immunity–Fecundity Tradeoffs in Drosophila. *Integrative and Comparative Biology*, *58*(3), 591–603. https://doi.org/10.1093/icb/icy078

Budar, F., Touzet, P., & Paepe, R. D. (2003). The Nucleo-Mitochondrial Conflict in Cytoplasmic Male Sterilities Revisited. *Genetica*, *117*(1), 3–16.

Camus, M. F., Clancy, D. J., & Dowling, D. K. (2012). Mitochondria, Maternal Inheritance, and Male Aging. *Current Biology*, *22*(18), 1717–1721. https://doi.org/10.1016/j.cub.2012.07.018

Camus, M. F., & Dowling, D. K. (2018). Mitochondrial genetic effects on reproductive success: Signatures of positive intrasexual, but negative intersexual pleiotropy. *Proceedings of The Royal Society B: Biological Sciences*, *285*(20180187).

Camus, M. F., Wolf, J. B. W., Morrow, E. H., & Dowling, D. K. (2015). Single Nucleotides in the mtDNA Sequence Modify Mitochondrial Molecular Function and Are Associated with Sex-Specific Effects on Fertility and Aging. *Current Biology*, *25*(20), 2717–2722. https://doi.org/10.1016/j.cub.2015.09.012

Case, A. L., Finseth, F. R., Barr, C. M., & Fishman, L. (2016). Selfish evolution of cytonuclear hybrid incompatibility in *Mimulus*. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1838), 20161493. https://doi.org/10.1098/rspb.2016.1493

Chinnery, P. F., & Schon, E. A. (2003). Mitochondria. *Journal of Neurology, Neurosurgery & Psychiatry*, *74*(9), 1188–1199. https://doi.org/10.1136/jnnp.74.9.1188

Clancy, D. J. (2008). Variation in mitochondrial genotype has substantial lifespan effects which may be modulated by nuclear background. *Aging Cell*, *7*(6), 795–804. https://doi.org/10.1111/j.1474-9726.2008.00428.x

Clancy, D. J., Hime, G. R., & Shirras, A. D. (2011). Cytoplasmic male sterility in Drosophila melanogaster associated with a mitochondrial CYTB variant. *Heredity*, *107*(4), 374–376. https://doi.org/10.1038/hdy.2011.12

Clark, A. G. (1984). Natural selection with nuclear and cytoplasmic transmission. I.A deterministic model. *Genetics*, *107*(4), 679–701. https://doi.org/10.1093/genetics/107.4.679

Connallon, T., Camus, M. F., Morrow, E. H., & Dowling, D. K. (2018). Coadaptation of mitochondrial and nuclear genes, and the cost of mother's curse. *Proceedings of*

the Royal Society B: Biological Sciences, *285*(1871). https://doi.org/10.1098/rspb.2017.2257

Dean, R., Zimmer, F., & Mank, J. E. (2014). The Potential Role of Sexual Conflict and Sexual Selection in Shaping the Genomic Distribution of Mito-nuclear Genes. *Genome Biology and Evolution*, *6*(5), 1096–1104. https://doi.org/10.1093/gbe/evu063

Delph, L. F., Touzet, P., & Bailey, M. F. (2007). Merging theory and mechanism in studies of gynodioecy. *Trends in Ecology & Evolution*, *22*(1), 17–24. https://doi.org/10.1016/j.tree.2006.09.013

Dowling, D. K., & Adrian, R. E. (2019). Challenges and Prospects for Testing the Mother's Curse Hypothesis. *Integrative and Comparative Biology*, *59*(4), 875–889. https://doi.org/10.1093/icb/icz110

Drown, D. M., Preuss, K. M., & Wade, M. J. (2012). Evidence of a Paucity of Genes That Interact with the Mitochondrion on the X in Mammals. *Genome Biology and Evolution*, *4*(8), 875–880. https://doi.org/10.1093/gbe/evs064

Drummond-Barbosa, D., & Spradling, A. C. (2001). Stem Cells and Their Progeny Respond to Nutritional Changes during Drosophila Oogenesis. *Developmental Biology*, *231*(1), 265–278. https://doi.org/10.1006/dbio.2000.0135

du Plessis, S., Agarwal, A., Mohanty, G., & van der Linde, M. (2015). Oxidative phosphorylation versus glycolysis: What fuel do spermatozoa use? *Asian Journal of Andrology*, *17*(2), 230. https://doi.org/10.4103/1008-682X.135123

Eslamieh, M., Williford, A., & Betran, E. (2017). Few nuclear-encodedmitochondrial gene duplicates contribute tomale germline-specific functions in humans. *Genome Biology and Evolution*, *9*(10), 2782–2790. https://doi.org/10.1093/gbe/evx176

Frank, S. A. (1989). The Evolutionary Dynamics of Cytoplasmic Male Sterility. *The American Naturalist*, *133*(3), 345–376. https://doi.org/10.1086/284923

Frank, S. A., & Hurst, L. D. (1996). Mitochondria and male disease. *Nature*, *383*(6597), 224–224. https://doi.org/10.1038/383224a0

Gallach, M., & Betrán, E. (2011). Intralocus sexual conflict resolved through gene duplication. *Trends in Ecology & Evolution*, *26*(5), 222–228. https://doi.org/10.1016/j.tree.2011.02.004

Gallach, M., Chandrasekaran, C., & Betrán, E. (2010). Analyses of Nuclearly Encoded Mitochondrial Genes Suggest Gene Duplication as a Mechanism for Resolving Intralocus Sexually Antagonistic Conflict in Drosophila. *Genome Biology and Evolution*, *2*, 835–850. https://doi.org/10.1093/gbe/evq069

Gemmell, N. J., Metcalf, V. J., & Allendorf, F. W. (2004). Mother's curse: The effect of mtDNA on individual fitness and population viability. *Trends in Ecology & Evolution*, *19*(5), 238–244. https://doi.org/10.1016/j.tree.2004.02.002

Gregorius, H. R., & Ross, M. D. (1984). Selection with gene–cytoplasm interactions. I. Maintenance of cytoplasm polymorphisms. *Genetics*, *107*(1), 165–178. https://doi.org/10.1093/genetics/107.1.165

Havird, J. C., & McConie, H. J. (2019). Sexually Antagonistic Mitonuclear Coevolution in Duplicate Oxidative Phosphorylation Genes. *Integrative and Comparative Biology*. https://doi.org/10.1093/icb/icz021

Hayward, A., & Gillooly, J. F. (2011). The Cost of Sex: Quantifying Energetic Investment in Gamete Production by Males and Females. *PLoS ONE*, *6*(1), e16557. https://doi.org/10.1371/journal.pone.0016557

Hoekstra, L. A., Julick, C. R., Mika, K. M., & Montooth, K. L. (2018). Energy demand and the context-dependent effects of genetic interactions underlying metabolism. *Evolution Letters*, *2*(2), 102–113. https://doi.org/10.1002/evl3.47

Hoekstra, L. A., Siddiq, M. A., & Montooth, K. L. (2013). Pleiotropic Effects of a Mitochondrial-Nuclear Incompatibility Depend upon the Accelerating Effect of Temperature in Drosophila. *Genetics*, *195*(3), 1129–1139. https://doi.org/10.1534/genetics.113.154914

Hoffmann, A. A. (2010). Physiological climatic limits in *Drosophila*: Patterns and implications. *Journal of Experimental Biology*, *213*(6), 870–880. https://doi.org/10.1242/jeb.037630

Hough, J., Ågren, J. A., Barrett, S. C. H., & Wright, S. I. (2014). Chromosomal Distribution of Cytonuclear Genes in a Dioecious Plant with Sex Chromosomes. *Genome Biology and Evolution*, *6*(9), 2439–2443. https://doi.org/10.1093/gbe/evu197

Hudry, B., Khadayate, S., & Miguel-Aliaga, I. (2016). The sexual identity of adult intestinal stem cells controls organ size and plasticity. *Nature*, *530*(7590), 344–348. https://doi.org/10.1038/nature16953

Lewis, D. (1941). Male Sterility in Natural Populations of Hermaphrodite Plants the Equilibrium Between Females and Hermaphrodites to Be Expected with Different

Types of Inheritance. *New Phytologist*, *40*(1), 56–63. https://doi.org/10.1111/j.1469-8137.1941.tb07028.x

Matoo, O., Julick, C. R., & Montooth, K. L. (2019). Genetic Variation for Ontogenetic Shifts in Metabolism Underlies Physiological Homeostasis in Drosophila. *Genetics*, *212*(2), 537–552. https://doi.org/10.1534/genetics.119.302052

Matzkin, L. M., Johnson, S., Paight, C., Bozinovic, G., & Markow, T. A. (2011). Dietary Protein and Sugar Differentially Affect Development and Metabolic Pools in Ecologically Diverse Drosophila. *The Journal of Nutrition*, *141*(6), 1127–1133. https://doi.org/10.3945/jn.111.138438

McLeod, C. J., Wang, L., Wong, C., & Jones, D. L. (2010). Stem Cell Dynamics in Response to Nutrient Availability. *Current Biology*, *20*(23), 2100–2105. https://doi.org/10.1016/j.cub.2010.10.038

Meiklejohn, C. D., Holmbeck, M. a, Siddiq, M. a, Abt, D. N., Rand, D. M., & Montooth, K. L. (2013). An Incompatibility between a mitochondrial tRNA and its nuclear-encoded tRNA synthetase compromises development and fitness in Drosophila. *PLoS Genetics*, *9*(1), e1003238. https://doi.org/10.1371/journal.pgen.1003238

Millington, J. W., & Rideout, E. J. (2018). Sex differences in Drosophila development and physiology. *Current Opinion in Physiology*, *6*, 46–56. https://doi.org/10.1016/j.cophys.2018.04.002

Montooth, K. L., Meiklejohn, C. D., Abt, D. N., & Rand, D. M. (2010). Mitochondrial–nuclear epistasis affects fitness within species but does not contribute to fixed incompatibilities between species of Drosophila. *Evolution*, *64*(12), 3364–3379. https://doi.org/10.1111/j.1558-5646.2010.01077.x

Moore, F. L., & Reijo-Pera, R. A. (2000). Male Sperm Motility Dictated by Mother's mtDNA. *The American Journal of Human Genetics*, *67*(3), 543–548. https://doi.org/10.1086/303061

Mossman, J. A., Biancani, L. M., Zhu, C.-T. T., & Rand, D. M. (2016). Mitonuclear Epistasis for Development Time and Its Modification by Diet in Drosophila. *Genetics*, *203*(1), 463–484. https://doi.org/10.1534/genetics.116.187286

Mossman, J. A., Slate, J., Birkhead, T. R., Moore, H. D., & Pacey, A. A. (2012). Mitochondrial haplotype does not influence sperm motility in a UK population of men. *Human Reproduction*, *27*(3), 641–651. https://doi.org/10.1093/humrep/der438

Mukai, C., & Okuno, M. (2004). Glycolysis Plays a Major Role for Adenosine Triphosphate Supplementation in Mouse Sperm Flagellar Movement. *Biology of Reproduction*, *71*(2), 540–547. https://doi.org/10.1095/biolreprod.103.026054

Okamoto, N., Nakamori, R., Murai, T., Yamauchi, Y., Masuda, A., & Nishimura, T. (2013). A secreted decoy of InR antagonizes insulin/IGF signaling to restrict body growth in *Drosophila*. *Genes & Development*, *27*(1), 87–97. https://doi.org/10.1101/gad.204479.112

Patel, M. R., Miriyala, G. K., Littleton, A. J., Yang, H., Trinh, K., Young, J. M., Kennedy, S. R., Yamashita, Y. M., Pallanck, L. J., & Malik, H. S. (2016). A mitochondrial DNA hypomorph of cytochrome oxidase specifically impairs male fertility in Drosophila melanogaster. *ELife*, *5*, e16923. https://doi.org/10.7554/eLife.16923

Petavy, G., David, J. R., Gibert, P., & Moreteau, B. (2001). Viability and rate of development at different temperatures in Drosophila: A comparison of constant and alternating thermal regimes. *Journal of Thermal Biology*, *26*(1), 29–39. https://doi.org/10.1016/S0306-4565(00)00022-X

Pomatto, L. C. D., Carney, C., Shen, B., Wong, S., Halaszynski, K., Salomon, M. P., Davies, K. J. A., & Tower, J. (2017). The Mitochondrial Lon Protease Is Required for Age-Specific and Sex-Specific Adaptation to Oxidative Stress. *Current Biology*, *27*(1), 1–15. https://doi.org/10.1016/j.cub.2016.10.044

Rand, D. M. (2001). The Units of Selection on Mitochondrial DNA. *Annual Review of Ecology and Systematics*, *32*(1), 415–448. https://doi.org/10.1146/annurev.ecolsys.32.081501.114109

Rand, D. M., Clark, A. G., & Kann, L. M. (2001). Sexually antagonistic cytonuclear fitness interactions in Drosophila Melanogaster. *Genetics*, *159*(1), 173–187.

Reiff, T., Jacobson, J., Cognigni, P., Antonello, Z., Ballesta, E., Tan, K. J., Yew, J. Y., Dominguez, M., & Miguel-Aliaga, I. (2015). Endocrine remodelling of the adult intestine sustains reproduction in Drosophila. *ELife*, *4*, e06930. https://doi.org/10.7554/eLife.06930

Rideout, E. J., Narsaiya, M. S., & Grewal, S. S. (2015). The Sex Determination Gene transformer Regulates Male-Female Differences in Drosophila Body Size. *PLOS Genetics*, *11*(12), e1005683. https://doi.org/10.1371/journal.pgen.1005683

Ruiz-Pesini, E., Lapeña, A.-C., Díez-Sánchez, C., Pérez-Martos, A., Montoya, J., Alvarez, E., Díaz, M., Urriés, A., Montoro, L., López-Pérez, M. J., & Enríquez, J. A. (2000). Human mtDNA Haplogroups Associated with High or Reduced

Spermatozoa Motility. *The American Journal of Human Genetics*, *67*(3), 682–696. https://doi.org/10.1086/303040

Sawala, A., & Gould, A. P. (2017). The sex of specific neurons controls female body growth in Drosophila. *PLOS Biology*, *15*(10), e2002252. https://doi.org/10.1371/journal.pbio.2002252

Sgrò, C. M., Overgaard, J., Kristensen, T. N., Mitchell, K. A., Cockerell, F. E., & Hoffmann, A. A. (2010). A comprehensive assessment of geographic variation in heat tolerance and hardening capacity in populations of Drosophila melanogaster from eastern Australia: Variation in heat tolerance and hardening capacity. *Journal of Evolutionary Biology*, *23*(11), 2484–2493. https://doi.org/10.1111/j.1420-9101.2010.02110.x

Testa, N. D., Ghosh, S. M., & Shingleton, A. W. (2013). Sex-Specific Weight Loss Mediates Sexual Size Dimorphism in Drosophila melanogaster. *PLoS ONE*, *8*(3), e58936. https://doi.org/10.1371/journal.pone.0058936

Unckless, R. L., & Herren, J. K. (2009). Population genetics of sexually antagonistic mitochondrial mutants under inbreeding. *Journal of Theoretical Biology*, *260*(1), 132–136. https://doi.org/10.1016/j.jtbi.2009.06.004

Wade, M. J., & Brandvain, Y. (2009). Reversing mother's curse: Selection on male mitochondrial fitness effects. *Evolution*, *63*(4), 1084–1089. https://doi.org/10.1111/j.1558-5646.2009.00614.x

Werner, M., & Simmons, L. W. (2008). Insect Sperm Motility. *Biological Reviews*, *83*(2), 191–208. https://doi.org/10.1111/j.1469-185X.2008.00039.x

White-Cooper, H. (2010). Molecular mechanisms of gene regulation during Drosophila spermatogenesis. *REPRODUCTION*, *139*(1), 11–21. https://doi.org/10.1530/REP-09-0083

Willett, C. S. (2008). No evidence for faster male hybrid sterility in population crosses of an intertidal copepod (Tigriopus californicus). *Genetica*, *133*(2), 129–136. https://doi.org/10.1007/s10709-007-9191-0

Willett, C. S., & Burton, R. S. (2003). Environmental influences on epistatic interactions: Viabilities of cytochrome C genotypes in interpopulation crosses. *Evolution*, *57*(10), 2286–2292. https://doi.org/10.1111/j.0014-3820.2003.tb00240.x

Xu, H., DeLuca, S. Z., & O'Farrell, P. H. (2008). Manipulating the Metazoan Mitochondrial Genome with Targeted Restriction Enzymes. *Science*, *321*(5888), 575–577. https://doi.org/10.1126/science.1160226

Yang, H., & Yamashita, Y. M. (2015). The regulated elimination of transit-amplifying cells preserves tissue homeostasis during protein starvation in *Drosophila* testis. *Development*, *142*(10), 1756–1766. https://doi.org/10.1242/dev.122663

Yee, W. K. W., Sutton, K. L., & Dowling, D. K. (2013). In vivo male fertility is affected by naturally occurring mitochondrial haplotypes. *Current Biology*, *23*(2), R55–R56. https://doi.org/10.1016/J.CUB.2012.12.002

Zhang, C., Montooth, K. L., & Calvi, B. R. (2017). Incompatibility between mitochondrial and nuclear genomes during oogenesis results in ovarian failure and embryonic lethality. *Development*, *144*(13), 2490–2503. https://doi.org/10.1242/dev.151951

Zhu, C.-T., Ingelmo, P., & Rand, D. M. (2014). G×G×E for Lifespan in Drosophila: Mitochondrial, Nuclear, and Dietary Interactions that Modify Longevity. *PLoS Genetics*, *10*(5), e1004354. https://doi.org/10.1371/journal.pgen.1004354

CHAPTER 2

MITOCHONDRIAL OXPHOS GENES EXHIBIT HIGHER LEVELS OF MOLECULAR COMPENSATION OF HUMAN DISEASE-ASSOCIATED MUTATIONS RELATIVE TO NUCLEAR OXPHOS GENES IN MAMMALS

ABSTRACT

The epistatic effects of amino acid substitutions exert a strong influence on protein evolution trajectories. A substitution that is pathogenic in one genetic background may be neutral or even beneficial in the presence of other substitutions that compensate for its deleterious effects. Mutations in the mitochondrial genome are the cause of multisystemic disorders in humans, yet many human disease-associated residues (DARs) are observed as wild-type in non-human species. This phenomenon is commonly explained by compensatory epistasis wherein interacting residues present in the non-human species ameliorate the deleterious effects of the DAR. We mapped 293 and 419 DARs in human mitochondrial (mtDNA) and nuclear (nucDNA) oxidative phosphorylation (OXPHOS) proteins onto their orthologs in 1062 and ~120 mammalian species, respectively, to identify 202 instances of putatively compensated DARS in non-human species. We found different levels of compensation for the two genomes, with 49.48% of mtDNA DARs and 13.60% of nucDNA DARs present as wild-type in one or several non-human mammals. The significantly higher proportion of compensated DARs in the mtDNA cannot be completely explained by the relatively larger number of species in our mtDNA mammalian dataset. We also found that, on average, the non-disease-causing and the disease-causing amino acids at compensated DARs tend to be more

similar in physicochemical properties and have relatively smaller effects on protein structure than those at uncompensated DARs. The compensated DARs also had a more variable structural neighborhood and higher levels of covariation with their structural neighbors, suggesting that molecular compensation is largely the result of structurally local mutations with the likelihood of compensation depending on the amino acid residue's location in the protein.

INTRODUCTION

Comparative genomic studies have discovered that disease-associated alleles in humans are sometimes present as wild-type alleles in non-human species. This observation was first reported in 2002 (Kondrashov et al., 2002; Waterston et al., 2002), and since then multiple studies have reported that 2-18% of pathogenic variants are present as wild-type alleles in non-human species (Azevedo et al., 2009; Barešić et al., 2010; Ferrer-Costa et al., 2007; Jordan et al., 2015; Kulathinal et al., 2004). These variants are not enriched in associations with late-onset or mild human diseases, and their presence as wild-type in non-human species cannot be attributed to founder effects as might be hypothesized for laboratory mice (Gao & Zhang, 2003). Given the intricacies of the metabolic networks, the compensation of mitochondrial or metabolic disorders can happen through physiological (Rossignol et al., 2003; Zieliński et al., 2016) or genetic (Buglo et al., 2020; El-Brolosy & Stainier, 2017) modes. From a genetic compensation perspective, the presence of these disease-causing alleles in other species implies the existence of epistatic genetic interactions that alleviate the deleterious effects in the non-human species. For disease-causing mutations in protein-coding genes, compensation can

arise from a few correlated substitutions at structurally proximal interacting sites or from many delocalized mutations that cumulatively counter the deleterious effects of the disease-causing substitutions (Domingo et al., 2019). While the latter form of compensation is very difficult to observe, analyses of the structural neighborhoods of disease-causing mutations can help elucidate the prevalence of local structural compensation.

We modified the naming scheme used by Xu & Zhang (2014) and refer to the pathogenic mutations as Disease-Associated Residues (DARs) instead of the scheme employed by Kondrashov et al., 2002 (Compensated Pathogenic Deviations or CPDs). We classify DARs as either compensated-Disease Associated Residues (c-DARs) if the disease-associated allele in humans is present in the sequence of a non-human ortholog, or as uncompensated-Disease Associated Residues (u-DARs) if the disease-associated allele is never found in another species (Figure 1).

Two distinct evolutionary models have been proposed to explain the existence of c-DARs (Storz, 2016, 2018). In the first model, the deleterious allele might arise in a permissive genetic background where its deleterious effect is offset by one or more interacting sites (Gong et al., 2013; Harms & Thornton, 2014; Ortlund et al., 2007). Substitutions at these interacting sites in humans might abolish the stabilizing interactions and thereby render the DAR variant deleterious. In the alternative model, the deleterious mutation occurs first but is subsequently compensated by a second-site substitution in the non-human species(A. Poon & Chao, 2005; A. F. Y. Poon & Chao, 2006). The latter model of compensatory evolution posits some period of time in which the population experiences decreased fitness.

Given the central role of the mitochondria in generating much of the cell's energy through oxidative phosphorylation (OXPHOS) (Quirós et al., 2016), mediating apoptosis, and regulating signal transduction, it is unsurprising that mitochondrial dysfunction has a profound impact on organismal fitness. The prevalence of mitochondrial diseases associated with mutations either in the mitochondrial DNA (mtDNA) or nuclear DNA (nucDNA) is estimated at 5 in 100,000 for children (Skladal et al., 2003) and at 20 in 100,000 for adults (Gorman et al., 2015). As with all mitochondrial functions, the OXPHOS pathway is under dual-genomic control, as the five transmembrane complexes responsible for OXPHOS consist of 110 polypeptides in humans, 97 of which are encoded by the nuclear genome while the other 13 subunits are mitochondrially-encoded. Thus, mitochondrial respiratory chain disorders represent a dual-genome complex disease that is both genetically and clinically heterogeneous with a variable age of onset and varying levels of tissue-specificity (Wong, 2007). Point mutations in the mtDNA constitute a significant cause of mitochondrial diseases in humans, with an estimated population prevalence of 1 in 200 (Elliott et al., 2008). Disease-associated alleles have been reported in every mtDNA gene, including mt-tRNAs (Kern & Kondrashov, 2004), and are associated with clinical symptoms ranging from non-syndromic sensorineural deafness to syndromic neurologic conditions such as mitochondrial encephalopathy, lactic acidosis and stroke-like episodes (MELAS) (www.mitomap.org; Lott et al., 2013). Mitochondrial disorders due to mutations in the nucDNA can be a result of either isolated or multiple respiratory chain complex deficiencies and occur in genes encoding diverse mitochondrial functions, including the OXPHOS complexes and the mitochondrial

protein translational machinery (e.g., the mitochondrial aminoacyl-tRNA synthetases (mt-aaRSs)(Zhu et al., 2009).

mtDNA differs from nucDNA in nucleotide composition and genetic code, mutation rates, lack of recombination, and generally uniparental pattern of inheritance (Wolstenholme, 1992). Mitochondria also exhibit cellular polyploidy, with several hundreds or thousands of identical or different copies of mtDNA being present in each cell, the latter phenomenon being termed heteroplasmy(Stewart & Chinnery, 2015; Wallace & Chalkia, 2013). The mtDNA experiences higher substitution rates relative to nucDNA across many animal species, and it has been suggested that the mtDNA may experience less effective selection due to a reduced effective population size (Allio et al., 2017). Thus, a compensatory model is often invoked to describe a co-evolutionary dynamic between the mtDNA and nucDNA—deleterious mutations first arise in the mitochondrial genome and are subsequently compensated by structurally interacting residues in the nuclear genome (Rand et al., 2004). Disorders affecting oxidative phosphorylation are uniquely dual genomic in origin and provide a framework to test models of molecular compensation and coevolution.

Because of the dual-genomic control and high potential for intra- and inter-genomic epistasis underlying mitochondrial dysfunction, we sought to test whether c-DARs were more common in OXPHOS genes than in other datasets. By mapping human pathogenic mutations in both mt- and nuc-encoded OXPHOS proteins onto their corresponding mammalian orthologs, we found that specifically mt-encoded OXPHOS proteins exhibit a substantially higher proportion of c-DARs (%) than nuc-encoded OXPHOS proteins or any published data set. We then characterized c-DARs and u-DARs

in the context of protein structural models to assess whether molecular compensation of the deleterious effects stems from cis or trans-genomic structurally neighboring residues. Finally, we analyzed OXPHOS DARs in a phylogenetic framework to test models of compensatory mitochondrial-nuclear coevolution.

MATERIALS AND METHODS

Data sources

Protein-coding mutations in mitochondrial OXPHOS proteins that cause diseases in humans were obtained from the MITOMAP database (http://mitomap.org, accessed May 2019; (Lott et al., 2013)); disease-causing protein-coding mutations in nuclear OXPHOS proteins were obtained from the HGMD database (http://www.hgmd.cf.ac.uk/, accessed November 2019; (Stenson et al., 2017)). The mutations dataset was filtered to exclude 23 mitochondrial frameshift mutations and 23 mitochondrial and 85 nuclear nonsense mutations, resulting in a total of 293 and 419 disease-causing missense mutations in mitochondrial and nuclear OXPHOS proteins respectively. Mitochondrial sequence data was obtained from the RefSeq database in GenBank for all available mammalian species. The list of species included in the analysis are reported in Supplementary Table S1 and Figure 5. Sequence data for nuclear gene orthologs was obtained from GenBank RefSeq database (O'Leary et al., 2016), Ensembl database and UCSC genome database.

Identification of potentially compensated Disease Associated Residues (c-DARs)

Multiple sequence alignments for individual mitochondrial genes were computed using the MAFFT algorithm (v7.450) implemented in Geneious R11 (Geneious 2020.1.1) with BLOSUM62 as the scoring matrix and a gap open penalty of 1.53 and an offset value of 0.123. MAFFT has been found to be more efficient and accurate for large alignments (Pais et al., 2014). These alignments were further cleaned up and trimmed using the GBlocks webserver (Talavera et al., 2007) set on 'less stringent' parameters.

Nucleotide coordinates for disease annotations were transformed to amino acid coordinates for individual genes and disease annotations were then classified as compensated vs uncompensated using custom Python scripts. In brief, the classification algorithm was as follows, for each disease-causing mutation, the corresponding column in the alignment was scanned for the presence of the mutant amino acid (Figure 1). If the wild-type amino acid was conserved across the alignment and the mutant amino acid not observed, the mutation was classified as an uncompensated Disease Associated Residue (u-DAR). If the mutant amino acid was observed in a non-human species, the mutation was classified as a compensated Disease Associated Residue (c-DAR).

Effect of sample size on the identification of c-DARs

To test the effect of sample size on the detection of c-DARs in the mitochondrial genome and provide a direct comparison between the proportion of c-DARs in mtDNA and nucDNA, we subsetted the mtDNA alignment in two ways. First, we subsampled the species set to include species for which sequence data was available for both mtDNA and nucDNA. This subsetting resulted in a reduced dataset of 57 mammalian species. Second,

we randomly subsampled the mammalian alignment and pruned the mammalian tree to preserve phylogenetic relationships in increments of 50 species with 100 randomizations at each increment and estimated the proportion of c-DARs in each subsample.

Estimation of phylogenetic relationships

A concatenated alignment of amino acid sequences for all mitochondrial OXPHOS genes and nuclear OXPHOS genes was used to infer the phylogenetic relationships of the mammalian species set with *Gallus gallus, Danio rerio* and *Xenopus tropicalis* serving as outgroups. The best-fit model of amino acid substitution was selected using ModelTest-NG v0.1.5 (Darriba et al., 2019). Phylogenetic trees were then constructed using maximum likelihood calculations with IQ-TREE multicore v1.6.7 (Nguyen et al., 2015) with 1000 ultrafast bootstrap replicates (Hoang et al., 2018) and 1000 replicates for SH-aLRT. Ancestral state reconstruction for all internal nodes of the tree was also performed using the empirical Bayesian method in IQ-TREE multicore v1.6.7. The number of independent origins for mutations at each DAR were parsimoniously inferred using a custom Python script.

Construction of protein structural models

Homology-based structural modelling was used to construct three-dimensional protein models for the OXPHOS complexes using MODELLER v9.23 (Sali & Blundell, 1993). Three dimensional structures submitted in the Protein Data Bank (Berman et al., 2000) served as templates for the protein structural models (Table 4) where structural

information was unavailable for *Homo sapiens*. Amino acid sequences for the mitochondrial-encoded subunits were used from the revised Cambridge Reference Sequence (rCRS), GenBank Accession number: NC_012920 and the sequences for the nuclear-encoded subunits were retrieved from the Genome Reference Consortium Human Build 37, GRCh37/hg19 (GenBank Accession number: GCF_000001405.13).

10 structural models were constructed for each OXPHOS complex using the automodel class in MODELLER. These models were further screened and corrected for sub-optimal loop configurations using Molecular Dynamics (MD) simulations at the 'very slow refine' level implement in MODELLER, following which the sub-optimal rotamer conformations were corrected using the RepairPDB function implemented in FoldX 4.0 (Schymkowitz et al., 2005).

Quantifying the structural impact of DAR mutations

We estimated the effect of DAR mutations on the thermodynamic stability of proteins for the compensated and uncompensated DARs by quantifying the predicted change in Gibb's free energy upon mutation ($\Delta\Delta G = \Delta G_{mut} - \Delta G_{wt}$). For each DAR mutation, we constructed ten 3D models with the wild-type and the mutant allele each for individual OXPHOS subunits. We subsequently performed a 10-fold relaxation on the structural models using RepairPDB following which we calculated the thermodynamic stability using the Stability function from FoldX 4.0. Finally, we computed the predicted change in Gibb's free energy upon mutation as the difference between the average stability of the ten mutant models and the ten wild-type models ($\Delta\Delta G = \Delta G_{mut} - \Delta G_{wt}$).

We chose the above protocol as it has been demonstrated to quantify the change in thermodynamic stability upon mutation with the least bias (Usmanova et al., 2018).

An exchangeability measure (Ex_value) as described in Yampolsky & Stoltzfus, 2005 was used to quantify the physicochemical changes upon mutation for the different classes of DARs. Yampolsky & Stoltzfus, 2005 compiled data from ~ 10,000 engineered amino acid exchanges in 12 proteins to compute a matrix of 'experimental exchangeability' to serve as a proxy for the physicochemical similarities of amino acids.

The depth of an amino acid residue from the protein surface has been shown to correlate significantly with its energetic contribution to the stability of the protein (Chakravarty & Varadarajan, 1999). The extent to which DARs were buried within each OXPHOS complex was computed using the DEPTH webserver (Tan et al., 2013).

Statistical significance for differences in the distributions for the different DAR classes were tested using the Kruskal-Wallis test implemented in R (R Core Team, 2021).

Characterization of the local structural neighborhood for DARs

For testing differences in the local structural neighborhood of the different classes of Disease-Associated Residues (DARs), we mapped the aligned sequence data onto three-dimensional structural models and identified neighboring amino acid residues using custom Python scripts. Amino acid residues that had at least one atom within 5Å of within any atom of a DAR defined the structural neighborhood of the DAR and were labelled as 'interacting residues'.

Neighborhood variability for compensated (c-DAR) vs uncompensated (u-DAR) mutations was then estimated as mean Shannon's entropy of the interacting residues across all species present in the gene alignment (Equation 1). For each DAR classified as a c-DAR, we further divided the alignment into two subsets: one comprising of species that had the mutant (disease-causing) amino acid as wild type (c-DAR species set) and the other comprising of species that had any other amino acid present at that locus (non-c-DAR species set). Statistical significance for differences in the distributions of Shannon's entropy for c-DARs vs u-DARs and c-DAR species set, and non-c-DAR species set was assessed using the Krukal-Wallis test implemented in R.

Shannon's entropy was estimated as -

$$H(x) = \lambda_H \sum_{i}^{K} p_i \,.\, log(p_i)$$

where K, the alphabet size is 21 (20 amino acids plus one gap symbol), and p_i is the probability of observing the i th symbol type. λ_H scales the entropy for N sequences so it's bound to [0,1] and is defined as

$$\lambda_H = [log_2(min(N, K)]^{-1}$$

Covariation between the DARs and their local structural neighborhood was quantified using three different measures:

1. Observed Minus Expected Score ($OMES(i,j)$) – These scores are a variant of the chi-square tests that compare the expected probability of amino acid residue pairs to their observed probabilities. OMES scores are for a pair of columns i, j is calculated as

$$OMES(i,j) = \sum_{a,b} \frac{\left(p(a_i, b_j) - p(a_i).p(b_j)\right)^2}{p(a_i).p(b_j)}$$

where,

$p(a_i, b_j)$ = observed frequency of residue *a* at position *i* and residue *b* at position *j*

$p(a_i).p(b_j)$ = expected frequency of residue *a* at position *i* and residue *b* at position *j*

and the summation is over all 20 types of amino acid residues.

2. Mutual Information ($MI(i,j)$) – MI is an information-theory metric that measures the mutual dependence of two random variables. It quantifies what we know about column *j,* given column *i* and is calculated as –

$$MI(i,j) = H(i) + H(j) - H(i,j)$$

where,

$H(i), H(j)$ = Shannon's entropies of columns *i* and *j* respectively

$H(i,j)$ = Joint Entropy of columns *i* and *j* respectively and is given by -

$$H(i,j) = -\sum_{a}^{K} \sum_{b}^{K} p(a_i, b_j) .log\ log\ (p(a_i, b_j))$$

Put together,

$$MI(i,j) = \sum_{a,b} p(a_i, b_j).log\left(\frac{p(a_i, b_{j)}}{p(a_i).p(b_{j)}}\right)$$

Same definitions for observed frequencies apply as above

3. Corrected Mutual Information ($MI_p(i,j)$) - First introduced by Dunn et al., 2008, MI_p introduces an average product correction (APC) term to the formulation of MI to account for background noise due to phylogenetic and entropy correlations. MI_p has been shown to outperform other metrics in detecting covariation between amino acid residue pairs and is calculated as -

$$MI_p(i,j) = MI(i,j) - APC_{i,j}$$

where,

$APC_{i,j} = \frac{MI(i,\underline{x}) \,.\, MI(j,\underline{x})}{\underline{MI}}$

$MI(i,\underline{x})$ = Mean mutual information of column *i,* given by $\frac{1}{n-1} \sum \quad MI(a,x)$,

for an alignment with n columns

Similarly, for $MI(j,\underline{x})$ and

$\underline{MI}$ = Overall mean mutual information, given by $\frac{2}{n\,(n-1)} \sum \quad MI(x,y)$

We computed Z-scores to measure how different a given covariance value was relative to all other values in the dataset and used them to identify cases of molecular compensation between DARs and their neighborhoods. The Z-scores were computed as –

$$Z - score = \frac{x - \mu}{\sigma}$$

where,

x = observed value

μ = mean of the sample

σ = standard deviation of the sample

Identification of the compensating residues for c-DARs

The structural neighborhood of the c-DAR residue was screened in species that harbored the disease-causing variant as wild-type (c-DAR species) to identify potential compensating residues. For a neighboring residue to be considered potentially compensating, we imposed two restrictions: the amino acid variant should be different than the amino acid present in wild-type *Homo sapiens* and the amino acid variant should be present at a higher abundance in the c-DAR species alignment than the wild-type *Homo sapiens.*

RESULTS

Characterization of the DARs for mitochondrial and nuclear OXPHOS proteins

We used mutation data for human diseases from the MITOMAP database and the HGMD database for mitochondrial respiratory chain proteins encoded by the mitochondrial and the nuclear genome respectively to identify instances of compensatory epistatic evolution. We restricted our analysis to missense point mutations in the coding regions of OXPHOS proteins which resulted in 293 DARs in the 13 mtDNA-encoded OXPHOS proteins and 419 DARs in 44 nucDNA-encoded OXPHOS proteins. Although the number of amino acid sites in these 44 nucDNA-encoded OXPHOS proteins (15,361) is ~4-fold higher than the number of mtDNA sites (3776), the number of nucDNA DARs

is only 1.4-fold higher than the number of mtDNA DARs, indicating a significant enrichment of OXPHOS DARs in mtDNA-encoded proteins (Fisher's Exact Test P-value < 0.0001). Although Complex II comprises only four protein subunits and 4% of the nucDNA sites, 252 (60.9%) of the nucDNA DARs occur in Complex II, the only OXPHOS protein complex encoded entirely by nucDNA (FET P-value < 0.0001 for Complex II vs. other complexes). However, an overwhelming majority of the Complex II DARs (236 DARs or 93.6%) were associated with tumors and only a small proportion (18 DARs or 7%) were implicated in mitochondrial disorders. In contrast, all but 1 of the 162 nucDNA DARs in the remaining OXPHOS complexes (Complex I, III, IV and V) and all of the mtDNA DARs were associated with mitochondrial disorders. Excluding the DARs associated with cancer reduces the total number of nucDNA DARs to 183, or 1.8% of the 10,151 sites in the alignment.

We mapped human DARs onto multiple sequence alignments consisting of 1062 mammalian mtDNA sequences (1025 Eutherian, 34 Metatherian and 3 Prototherian) and 136 mammalian nucDNA proteins (131 Eutherian, 4 Metatherian and 1 Prototherian) to classify them as compensated (c-DARs which are found in at least one non-human species) or uncompensated (u-DARs). While only 57 out of 419 (13.60%) nucDNA DARs are a native residue in at least one non-human species, 145 out of 293 (49.48%) mtDNA DARs are the wild-type allele in another species and were annotated as c-DARs (Table 1). The proportion of c-DARs we detect in nucDNA OXPHOS proteins is comparable to estimates from previous studies, while the proportion of c-DARs in mtDNA OXPHOS proteins is significantly larger (Fisher's Exact test $p < 0.001$; Table 2).

Previous surveys of compensated DARs used smaller datasets (<60 species) (Barešić et al., 2010; Ferrer-Costa et al., 2007; Jordan et al., 2015; Kondrashov et al., 2002). To test whether the high proportion of c-DARs detected in our mtDNA alignment was simply a consequence of the large number of species in our dataset, we randomly subsampled the mtDNA alignment in increments of 50 species and estimated the proportion of c-DARs for each increment (Figure 2). For the smallest mtDNA subsamples of 50 species, on average ~27% of DARs were compensated, more than double the proportion observed in the full nuclear-encoded protein alignment. We further restricted our mtDNA dataset to match the species used in three previously published studies to enable direct comparison with previous c-DAR proportion estimates. In all three cases, the proportions of c-DARs were at least 7-fold higher in mtDNA than nucDNA: Jordan et al., 2015 (38.1% vs 5.3%), Azevedo et al., 2016 (37.4% vs 3.7%) and Yan et al., 2011 (17.1% vs 1.08%). We also calculated the proportion of c-DARs in a reduced dataset that includes only species for which sequence data was available for both mtDNA and nucDNA OXPHOS genes. In this subset of 57 species, 97 out of 289 mtDNA DARs (33.6%) were compensated while 28 out of 414 nucDARs (6.8%) were compensated. Together, these results suggest that, regardless of the number or identity of the mammalian species investigated, DARs in mtDNA are 3-4-fold more likely to be compensated than the DARs in nucDNA. This pattern is independent of the probability that a protein carries a DAR, since Complex II, which shows a much higher number of DARs than other nucDNA complexes, nonetheless has a similar proportion of DARS that are compensated (~14%) as other nucDNA complexes and a lower proportion than mtDNA complexes (~50%).

We compared the phylogenetic distributions of human disease residues found in non-human species across the mammalian tree. Among the 57 nucDNA c-DARs, we infer that 33 (58%) originated only once, 10 (18%) have two origins, the remaining 14 (25%) independently arose more than twice, up to a maximum of 9 independent origins across the 57 species. 47 nucDNA cDARs (82%) are taxonomically rare (found in <5% of species in the alignment) while five nucDNA c-DARs are common (>20% of species), and the most common is found in 56/136 species (41%). Among the 145 mtDNA c-DARs, 23 (16%) have a single origin, 14 (10%) have two origins, and the remaining 108 (74%) originated more than twice. 62 mtDNA c-DARs (43%) have ten or more independent origins. 87 mtDNA c-DARs (60%) are rare, while 16 mtDNA c-DARs are found in more than 20% of the species in the alignment, and the most common mtDNA c-DAR is found in 1009 (95%) of non-human species. While direct comparisons between the phylogenetic distributions of mtDNA vs nucDNA c-DARs are difficult given the disparate sizes of the two datasets, mtDNA c-DARs appear to be older, more taxonomically widespread, and independently arise more frequently than nucDNA cDARs.

Compensated DAR mutations tend to have a milder effect at the molecular level.

We estimated the effect of DAR substitutions on protein structure and function using metrics that reflect the likelihood that an amino acid substitution will disrupt protein function. We computed distributions of exchangeability values (Yampolsky & Stoltzfus, 2005) to quantify physico-chemical changes to proteins that result from the two different classes of DARs. c-DAR amino acid substitutions had higher exchangeability

values than u-DAR amino acid substitutions for both mtDNA (c-DAR: median value = 0.31 (IQR = 0.14), u-DAR: median value = 0.28 (IQR = 0.15); Wilcoxon ranked sum test, $p < 0.001$, effect size r = 0.22) and nucDNA OXPHOS proteins (c-DAR: median value = 0.29 (IQR = 0.10), u-DAR: median value = 0.26 (IQR = 0.12); Wilcoxon ranked sum test, $p = 0.017$, effect size r = 0.12). When compared between genomes, mtDNA c-DARs and u-DARs are more exchangeable than nucDNA c-DARs and u-DARs (Wilcoxon ranked sum test, c-DAR: $p = 0.032$, effect size r = 0.15, uDAR: $p = 0.007$, effect size r = 0.12) (Figure 5).

We also estimated the difference in thermodynamic stability between the wild-type and mutant alleles in human proteins ($\Delta\Delta G = \Delta G_{mut} - \Delta G_{wt}$). For each DAR, we calculated the mean change in Gibb's free energy ($\Delta\Delta G$) from 10 mutant structural models: more positive $\Delta\Delta G$ values correspond to more destabilizing mutations. The $\Delta\Delta G$ values indicate c-DAR mutations are less destabilizing (mtDNA: median $\Delta\Delta G$ = 0.037 (IQR = 4.28); nucDNA: median $\Delta\Delta G$ = -0.181 (IQR = 8.54)) than u-DAR mutations (mtDNA: median $\Delta\Delta G$ = 0.880 (IQR = 4.60); nucDNA: median $\Delta\Delta G$ = 0.688 (IQR = 5.72)) for both mtDNA and nucDNA DARs (Wilcoxon ranked sum test, mtDNA: $p = 0.049$, effect size r = 0.12; nucDNA: $p = 0.089$, effect size r = 0.09). mtDNA c-DARs and u-DARs were not significantly different than the corresponding nucDNA c-DARs and u-DARs (Wilcoxon ranked sum test, c-DAR: $p = 0.187$, effect size r = 0.10, u-DAR: $p = 0.300$, effect size r = 0.05). Together, these results suggest that the likelihood that a human DAR is the wild-type allele in a non-human mammal is in part determined by the extent to which the DAR allele disrupts protein structure and function, with milder perturbations being more likely to be compensated and observed in another species.

Specific substitutions between particular pairs of amino acids were also observed to be more likely to be compensated when they occur in mtDNA but not in nucDNA (Fisher's Exact test, mtDNA: $p < 0.001$, nucDNA: $p = 0.414$) (Figure 3 and 4). For mtDNA, three out of 78 amino acid substitutions were more likely to be a compensated c-DAR (Fisher's exact test; A->T ($p = 0.0003$), I->T ($p = 0.002$) and I->V ($p = 0.0008$)), while 2 mutations were more likely to be an uncompensated u-DAR (Fisher's exact test; L->P ($p = 0.0014$) and S->P ($p = 0.0016$)). In terms of raw numbers, 36 out of the 47 observed mutations to threonine in mtDNA appeared to be compensated in non-human species, whereas only 1 out of the 25 observed mutations to proline in mtDNA was classified as compensated. Overall, mutations to an aliphatic or a polar amino acid residue were more likely to be compensated than mutations to aromatic or charged amino acid residues in the mtDNA. For nucDNA, which has a higher proportion of uncompensated mutations, mutations to tyrosine and cysteine were largely uncompensated. Mutations from leucine to proline were more likely to be uncompensated in both mtDNA and nucDNA, with only 1 out of 12 L->P mutations being classified as c-DAR.

Three-dimensional localization of DARs in protein complexes.

Amino acids present on the surface of proteins exhibit higher rates of substitution and hence might be more frequently compensated than residues buried deep inside the protein (Halabi et al., 2009; Tseng & Liang, 2006). Furthermore, if the compensation of c-DARs is facilitated by intergenomic interactions, then we might predict that mtDNA c-DARs located at the interface between mitochondrially-encoded and nuclear-encoded

proteins are more likely to be compensated than mtDNA c-DARs on the interface between two mitochondrially-encoded proteins or at non-interface sites. To test these hypotheses, we computed residue depths for the different DAR classes for each OXPHOS complex and mapped them onto structural models to determine whether they sit at an interface between two proteins. For all OXPHOS complexes, on average, c-DARs were found to be more exposed than u-DARs ($\chi^2 = 8.33,\ df = 2, p - value = 0.015$). Previous studies looking at DARs in nuclear proteins also reported c-DARs to be biased towards more exposed locations for both mammalian and all-species datasets (Barešić et al., 2010; Ferrer-Costa et al., 2007).

We found the proportions of mtDNA c-DARs were similar across sites that contact a nuclear-encoded protein *versus* a mitochondrially-encoded protein *versus* do not contact residues in another protein (54%, 44%, and 49% respectively, Fisher's Exact Test $p > 0.3$ for all comparisons, Table 3). A similar result was observed within nucDNA c-DARs: 12%, 10%, and 15% of DARs that contact mitochondrially-encoded proteins, contact nuclear-encoded proteins, or do not contact another protein, are cDARs respectively (Fisher's Exact Test $p > 0.2$ for all comparisons). This suggests that the excess of compensation observed among mtDNA cDARs is not primarily driven by intergenomic co-evolution at interfaces between proteins.

Characterization of structural neighborhoods of DARs

To test the hypothesis that compensation is accomplished by amino acid substitutions in close physical proximity to the DAR, we used Shannon's entropy metric

to estimate amino acid variability at residues that physically interact with the DAR. Amino acid residues within 5 Å of a DAR were defined as interacting residues. The number of structural neighbors was higher for u-DARs than for c-DARs for both mtDNA (median number of neighbors: u-DAR: 13 (IQR = 4), c-DAR: 12 (IQR = 4); Wilcoxon ranked sum test, $p = 0.003$, effect size r = 0.17) and for nucDNA (median number of neighbors: u-DAR: 12 (IQR = 5), c-DAR: 8.50 (IQR = 6.25); Wilcoxon ranked sum test, $p < 0.001$, effect size r = 0.31). c-DARs had a more variable structural neighborhood than u-DARs for both mtDNA (c-DAR: median $H(x) = 0.10$ (IQR = 0.16), u-DAR: median $H(x) = 0.06$ (IQR = 0.07); Wilcoxon ranked sum test, $p < 0.001$, effect size r = 0.34) and nucDNA (c-DAR: median $H(x) = 0.11$ (IQR = 0.13), u-DAR: median $H(x) = 0.05$ (IQR = 0.07); Wilcoxon ranked sum test, $p < 0.001$, effect size r = 0.32). This pattern is consistent with local compensation by residues within a DAR's structural neighborhood. Further for each c-DAR, we defined two subsets of the species set: species that have the disease-causing amino acid at the DAR locus (c-DAR species) and the species that carried a different allele at the same locus (non-c-DAR species). We found that the structural variability around the DAR residue was more variable in non-c-DAR species than in c-DAR species for both the mtDNA (non c-DAR species: median $H(x) = 0.11$ (IQR = 0.17), c-DAR species: median $H(x) = 0.08$ (IQR = 0.14); Wilcoxon ranked sum test, $p = 0.003$, effect size r = 0.17) and nucDNA (non c-DAR species: median $H(x) = 0.10$ (IQR = 0.13), c-DAR species: median $H(x) = 0.00$ (IQR = 0.08); Wilcoxon ranked sum test, $p < 0.001$, effect size r = 0.56). The higher variability in the structural neighborhoods of c-DAR seems to suggest that molecular compensation of deleterious variants is more likely in structural pockets with high substitution rates. The

observation that c-DAR species (species with disease-causing amino acid at DAR locus) have a less variable structural neighborhood than non-c-DAR species (species that have a different amino acid at the DAR locus) suggests a functional constraint and evolutionary pressure to maintain specific interactions in the structural neighborhood that mask the deleterious effects of the disease-causing mutation.

We quantified the covariation between the DAR and its local structural neighborhood using three different measures of covariation: Observed Minus Expected Score (OMES), Mutual Information (MI) and Corrected Mutual Information (MI_p). For all three measures and across both genomes, c-DARs had higher covariation scores with their neighborhoods than u-DARs. Using OMES as a measure of covariation, for mtDNA, the median OMES values were, c-DARs: 0.01 (IQR = 1.25), u-DARs: 0.00 (IQR = 0.00); Wilcoxon ranked sum test, $p < 0.001$, effect size r = 0.48), while for nucDNA, the median OMES values were c-DARs: 0.02 (IQR = 0.10), u-DARs: 0.00 (IQR = 0.01); Wilcoxon ranked sum test, $p < 0.001$, effect size r = 0.27). Using MI as a measure of covariation, for mtDNA, the median MI values were, c-DARs: 0.00 (IQR = 0.02), u-DARs: 0.00 (IQR = 0.00); Wilcoxon ranked sum test, $p < 0.001$, effect size r = 0.47), while for nucDNA, the median MI values were c-DARs: 0.03 (IQR = 0.04), u-DARs: 0.01 (IQR = 0.03); Wilcoxon ranked sum test, $p < 0.001$, effect size r = 0.23). Using MI_p as a measure of covariation, for mtDNA, the median MI_p values were, c-DARs: 0.00 (IQR = 0.00), u-DARs: 0.00 (IQR = 0.00); Wilcoxon ranked sum test, $p = 0.027$, effect size r = 0.04), while for nucDNA, the median MI_p values were c-DARs: 0.00 (IQR = 0.00), u-DARs: 0.00 (IQR = 0.00); Wilcoxon ranked sum test, $p = 0.004$, effect size r = 0.04).

In line with the above observations, the higher levels of covariation between the c-DARs and their structural neighborhoods suggests that the molecular compensation of deleterious variants is largely a local phenomenon.

Nature of structural compensation

In order to identify the amino acid variants compensating for a DAR's deleterious effects, we screened the DAR's structural neighborhood with the following assumption: a compensating amino acid is one that is structurally adjacent to a DAR, is absent in the wild-type human sequence but present in sequences of species harboring the disease-causing variant of the DAR. We found potentially compensating partners for 56 out of 414 nucDNA c-DARs and 138 out of 289 mtDNA c-DARs, which respectively represent 13.52% and 47.75% of the nucDNA and mtDNA c-DARs identified in our study. We further restricted this categorization of compensating residues to amino acid variants that were present in greater abundance than the wild-type human variant. This reduced the number of c-DARs for which we could recover a compensating partner to 3 nucDNA c-DARs and 75 mtDNA c-DARs.

Intergenomic molecular compensation of a mtDNA DAR

Through structural analysis and *in-silico* mutagenesis screens we were able to identify a case of inter-genomic compensation between the mitochondrial and the nuclear genomes. The COX1:Met117Thr mutation has been reported in human patients suffering from prostate cancer (Brandon et al., 2006; Petros et al., 2005) The mutant amino acid

Threonine is present as a wild-type residue in the Sumatran Orangutan (*Pongo abelii)* and the Bornean Orangutan (*Pongo pygmaeus)*. Met117 also lies on a mitochondrial-nuclear interface and interacts with the COX7C protein. An alignment of the COX7C protein for all apes where sequence data was available reveals a single Val55Ile mutation that separates the *Pongo* clade from the rest of the apes. Structural analyses of these substitutions reveal that COX1:Met117 directly interacts with COX7C:Val55 and that this interaction is abolished in the presence of Thr117. A Val55Ile mutation in COX7C then compensates the deleterious effects of the Thr117 in the *Pongo* clade.

DISCUSSION

Pathogenic mutations that cause dysfunction in humans have been identified to exist in non-human species without apparent pathogenicity. These variants have been identified in protein-coding genes in the nuclear (Barešić & Martin, 2011; Ferrer-Costa et al., 2007; Kondrashov et al., 2002; Waterston et al., 2002) and mitochondrial genomes (Azevedo et al., 2009; De Magalhães, 2005; Tavares & Seuánez, 2017), as well as tRNA (Kern & Kondrashov, 2004; Meer et al., 2010) and rRNA genes (Emperador et al., 2015). Experimental investigations focusing on functional tests of such variants establish their pathogenicity and report subsequent rescue through the introduction of the non-human wild-type variant at an interacting site (Emperador et al., 2015; Jordan et al., 2015).

We mapped disease-associated residues associated with mitochondrial disorders in both mitochondrial (mtDNA) and nuclear (nucDNA) OXPHOS proteins and classified them as compensated (c-DAR) or uncompensated (u-DAR) based on whether the

deleterious amino acid variant was observed in a non-human mammalian species. While the proportion of DARs that are compensated (c-DARs) in nuclear proteins is similar to previous estimates, the proportion of c-DARs in mitochondrial proteins is significantly higher. Since our mtDNA sequence alignment covered a broader taxonomic range and included a larger number of species than previous studies as well as our nucDNA sequence alignment, we performed subsampling analyses to demonstrate that this result does not depend on the number, or the identity of the species used in the sequence alignment.

Compensatory mutations in the non-human species that alleviate the deleterious effects of the pathogenic variants represent a genetic compensation mechanism to explain their existence. This compensation can occur in two ways: the deleterious effects can be suppressed in a localized manner by a small number of discrete compensatory substitutions, typically occurring in close proximity to the deleterious variant, or, the suppression of the deleterious effects can be achieved by a global shift in the properties of the gene (such as thermodynamic stability) resulting from the cumulative action of a lot of substitutions that individually have small effects. It is likely that both these mechanisms are important in the maintenance of c-DARs.

Consistent with previous studies (Barešić et al., 2010; Ferrer-Costa et al., 2007; Marín et al., 2019), we find that c-DARs tend to have milder effects on protein structure and function than u-DARs for both mtDNA and nucDNA studies. c-DAR mutations tend to be physicochemically similar and cause smaller reductions in thermodynamic stability than u-DARs, suggesting that some mutations may be more compensatable than others. Structural analysis of the neighborhoods surrounding the DARs reveal that c-DARs

reside in neighborhoods with higher structural variation and exhibit stronger covariation with their neighborhoods relative to u-DARs. Within c-DARs, the structural neighborhoods are more conserved in species with the disease-causing allele than in species harboring a different allele. This suggests an evolutionary pressure on the residues in the structural neighborhoods in order to maintain favourable interactions with the DAR in species with the disease-causing variant. Once the disease-causing variant arises in the course of evolution, it exerts a purifying selective pressure on the subsequent substitutions, leading to a type of evolutionary conservatism, a phenomenon referred to as entrenchment (Shah et al., 2015).

Taken together, these results suggest that c-DARs are less likely to disrupt protein structure and function and exhibit stronger covariation with their structural neighbors than u-DARs. These patterns are consistent across both the genomes are hence unlikely to account for the differences we observe between genomes. The differences between the genomes can be partially explained by the differences in substitution rates between the genomes (Allio et al., 2017). Given the higher substitution rate in animal mitochondrial genomes, the background rate at which DARs arise in the mtDNA is expected to be higher than that for nucDNA. A higher proportion of these mtDNA DARs might find themselves in a permissive genetic background that would allow them to persist and eventually fix in the population.

While all the proteins analyzed in this study were part of the same biochemical pathway and hence expected to be subject to similar functional constraints, the mtDNA-encoded proteins in OXPHOS protein complexes often occupy central, core positions and perform crucial catalytic activities (Elurbe & Huynen, 2016). Many nuclear-encoded

OXPHOS proteins are considered ‘supernumerary’ and while not being directly involved in the catalytic activities, they play essential roles in the assembly, regulation, and stability of OXPHOS complexes. Mitochondrial OXPHOS proteins also experience higher levels of purifying selection than their corresponding nuclear counterparts in mammals (Popadin et al., 2013) suggesting mtDNA proteins may be subject to greater functional constraints than nucDNA proteins.

While the general hypothesis explaining the existence of these variants is the presence of compensatory mutations required for maintaining thermodynamic stability in the non-human species (DePristo et al., 2005; Wittenhagen & Kelley, 2003; Xu & Zhang, 2014), the deleterious effects of mutations that cause disease are often multi-faceted and hence genetic compensation through substitutions in the same or interacting genetic elements (Buglo et al., 2020; El-Brolosy & Stainier, 2017) represents one mode of compensation. Compensation of the deleterious effects can also arise through metabolic or physiological changes especially for disorders affecting mitochondrial function (Rossignol et al., 2003; Zieliński et al., 2016). The metabolic demands might also be significantly different in non-human species such that a mutation that has a deleterious effect in humans might not species with lower metabolic demands to the same extent. Hence, these mutations might be tolerated more readily in the non-human species than they are in humans.

FIGURES AND TABLES

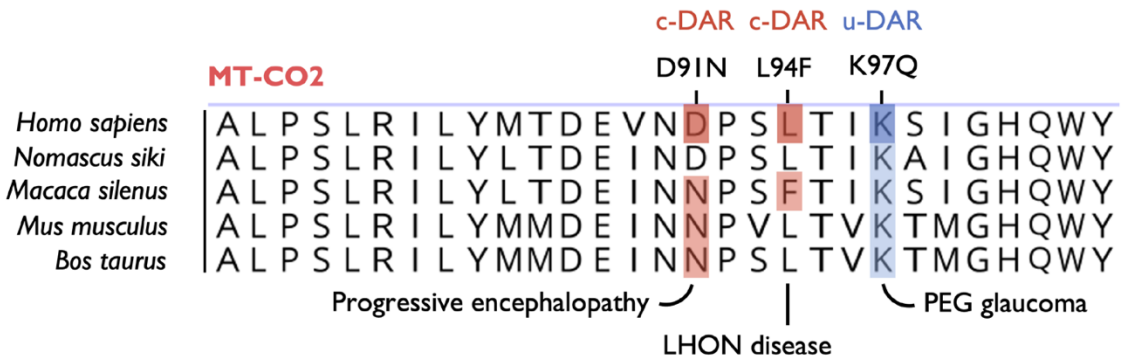

Figure 1. Disease Associated Residues (DARs) in the mitochondrially encoded Cytochrome oxidase 2 (MT-CO2) gene. The figure shows the alignment of the human MT-CO2 protein sequence with other non-human mammalian MT-CO2 protein sequences. Highlighted columns are amino acid positions associated with mitochondrial diseases; for example, a substitution from Aspartic acid (D) to Asparagine (N) at position 91 is associated with Progressive encephalopathy. Columns 91 and 94 represent compensated-Disease Associated Residues (c-DAR) as the mutant amino acid is present as a wild-type residue in non-human mammals whereas column 97 represents an uncompensated-Disease Associated Residue (u-DAR) as the disease-causing Q residue is never observed as the wild-type amino acid in another mammalian species.

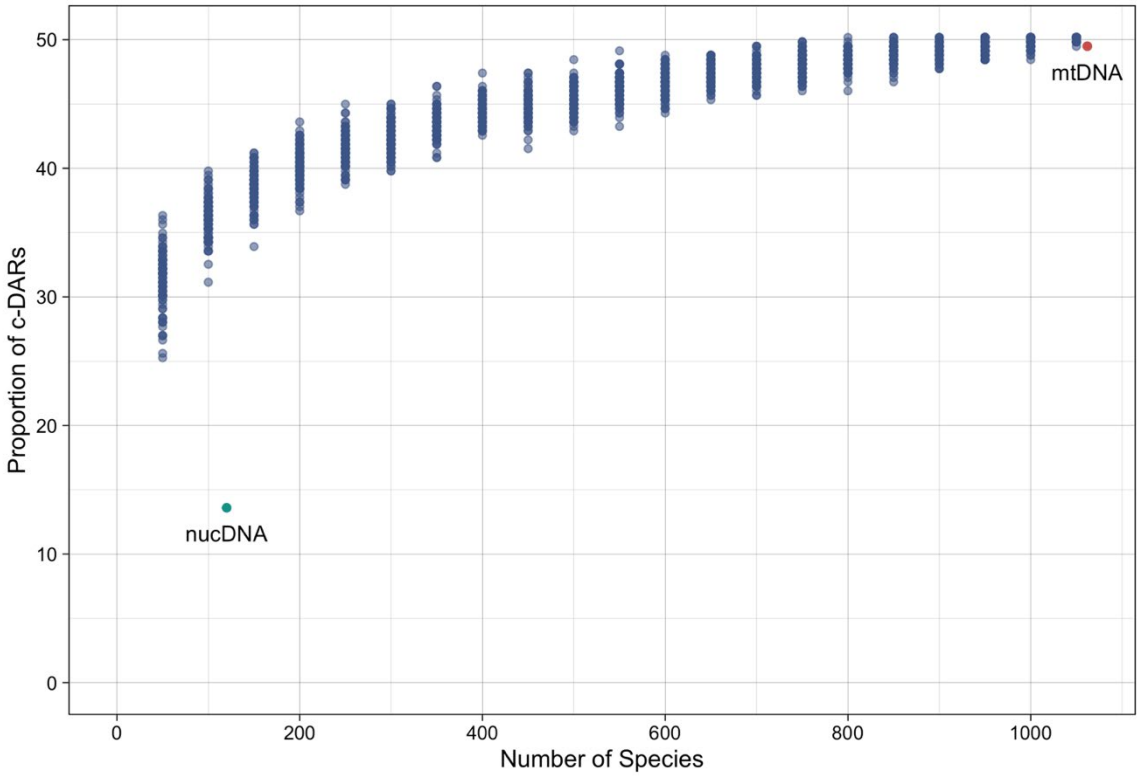

Figure 2. Subsampling analysis to test for the effect of sample size on the detection of c-DARs in mtDNA. The proportion of c-DARs is plotted against the number of species in the multiple sequence alignment. The red dot represents the proportion of c-DARs in the complete mtDNA dataset comprising 1063 species.

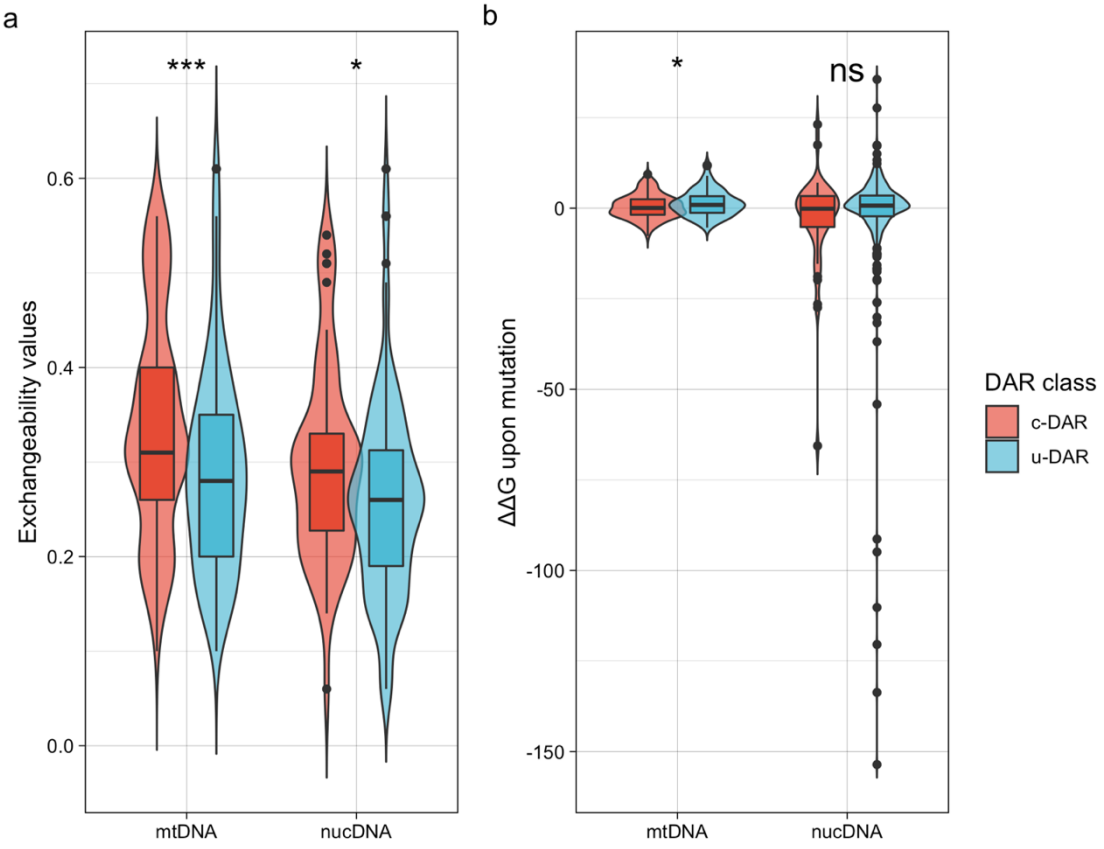

Figure 3. c-DAR mutations are more exchangeable and have milder effects on protein structure than u-DAR mutations for both mtDNA and nucDNA. Horizontal lines within the boxplot represent median values. Statistical significance levels for differences in median values correspond to *: p <= 0.05; ***: p <= 0.001.

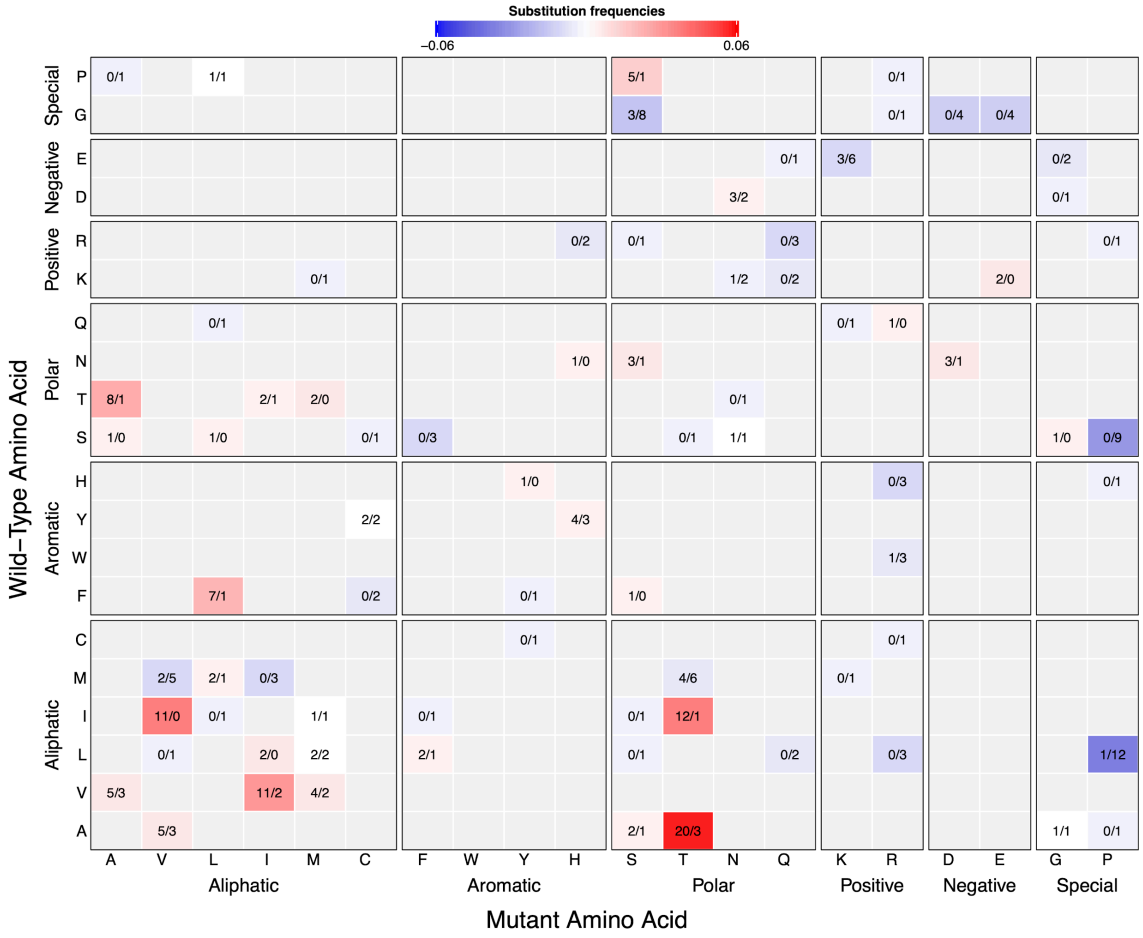

Figure 4. mtDNA DAR substitution differences. The redder a tile is, the higher the frequency of c-DARs than u-DARs for each particular type of mutation and vice versa for the bluer tiles. Numbers in the tiles represent the number of c-DARs divided by the number of u-DARs.

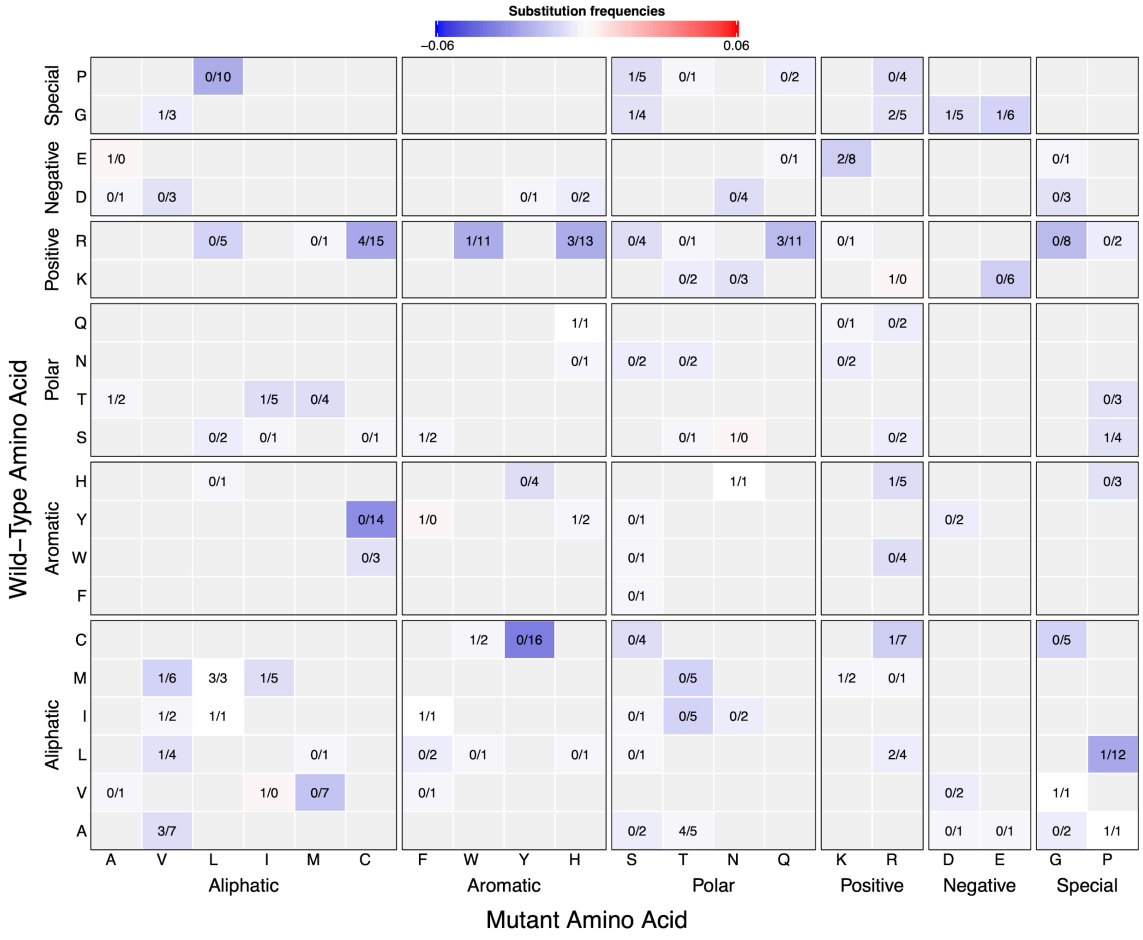

Figure 5. nucDNA DAR substitution differences: the redder a tile is, the higher the frequency of c-DARs than u-DARs for each particular type of mutation and vice versa for the bluer tiles. Numbers in the tiles represent the number of c-DARs divided by the number of u-DARs.

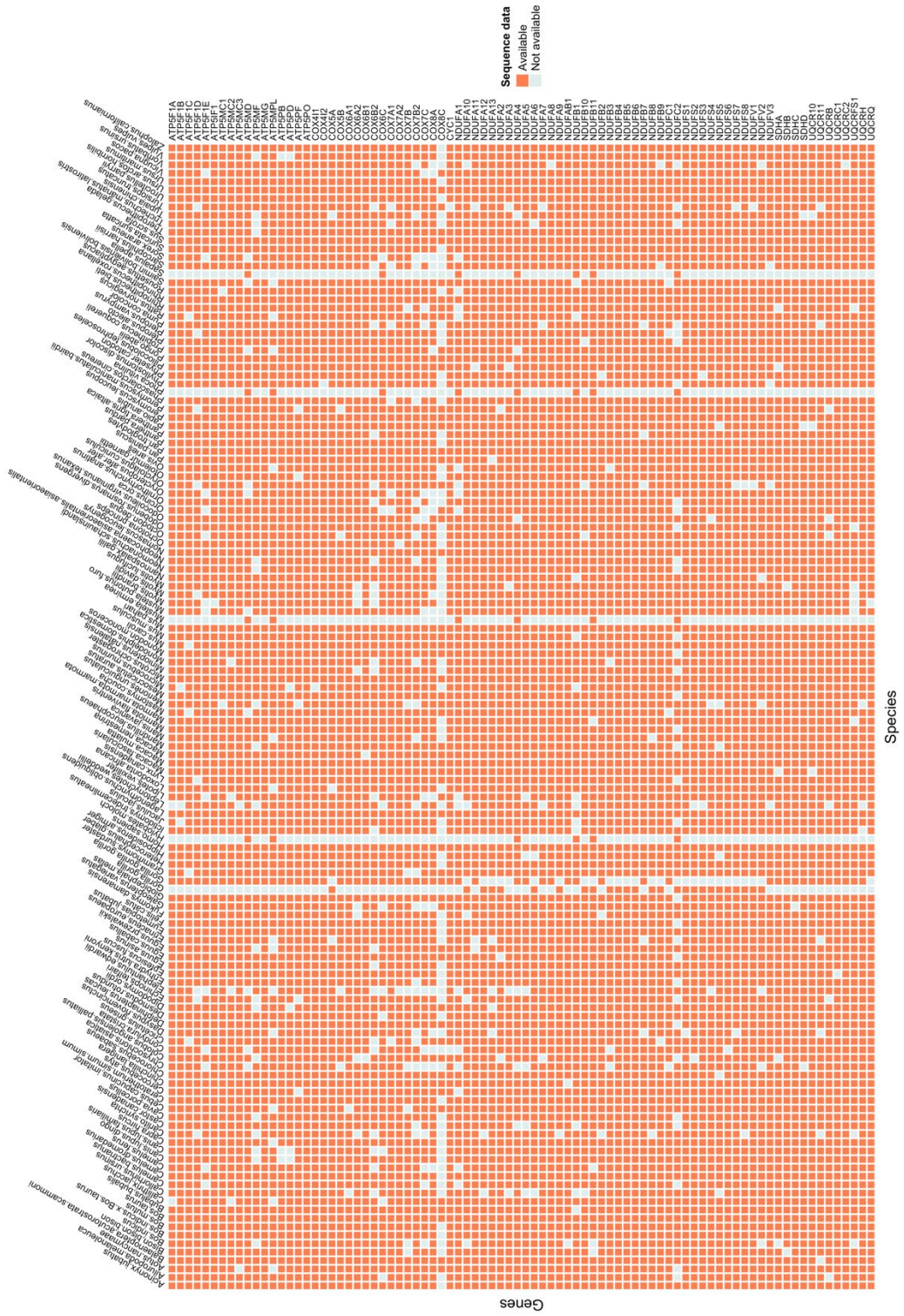

Figure 5. Species x Genes heatmap for nucDNA showing availability of sequence data.

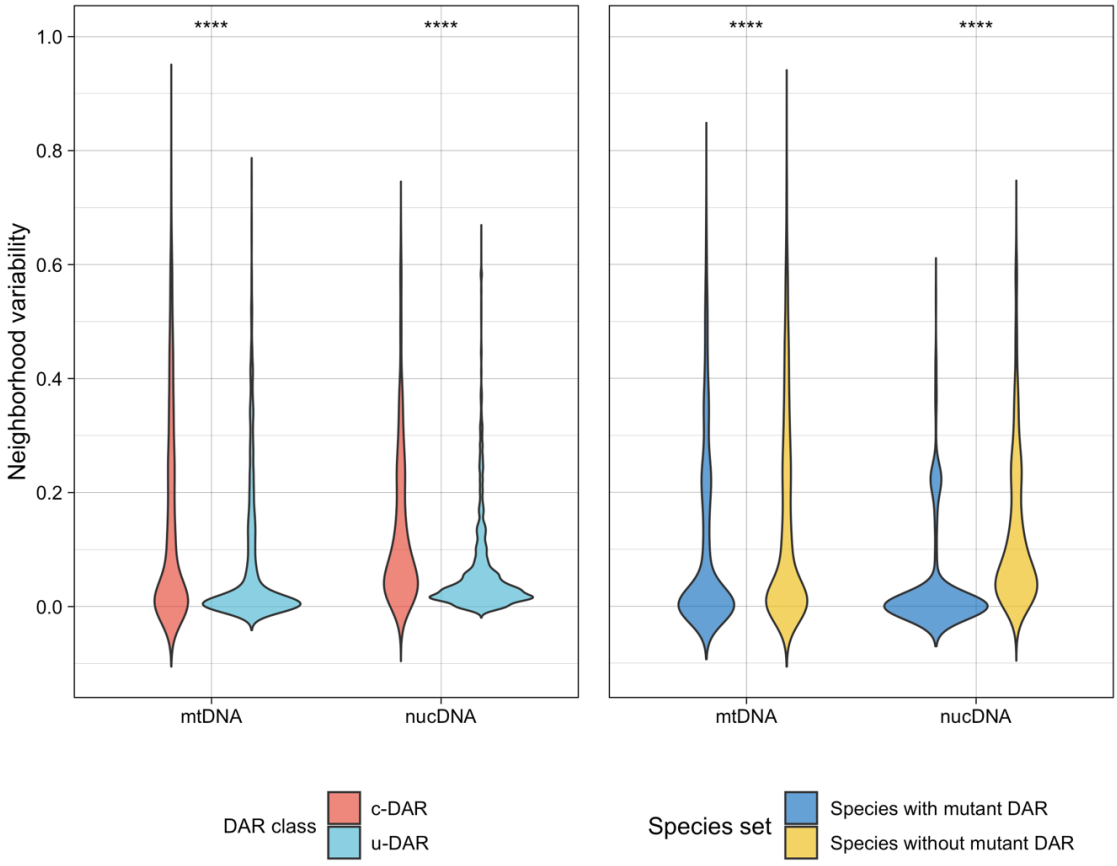

Figure 6. c-DARs have a more variable structural neighborhood than u-DARs as estimated using Shannon's entropy for both mtDNA and nucDNA. However, within c-DARs, the structural neighborhoods are more variable for species that do not contain the disease-causing amino acid as wild-type than for species that do. Statistical significance levels for differences in median values correspond to ****: p <= 0.0001.

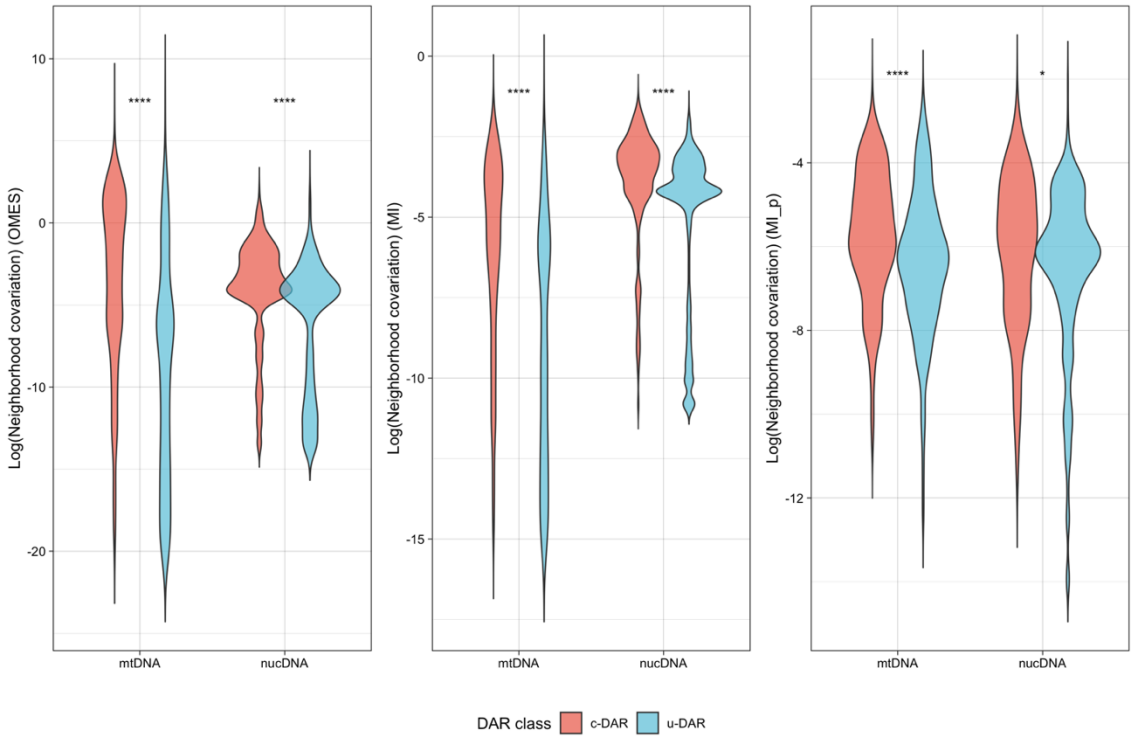

Figure 7. c-DARs show higher covariation scores with their structural neighborhoods than u-DARs for both mtDNA and nucDNA, using different metrics for quantifying covariation. Statistical significance levels for differences in median values correspond to *: p <= 0.05; ****: p <= 0.0001.

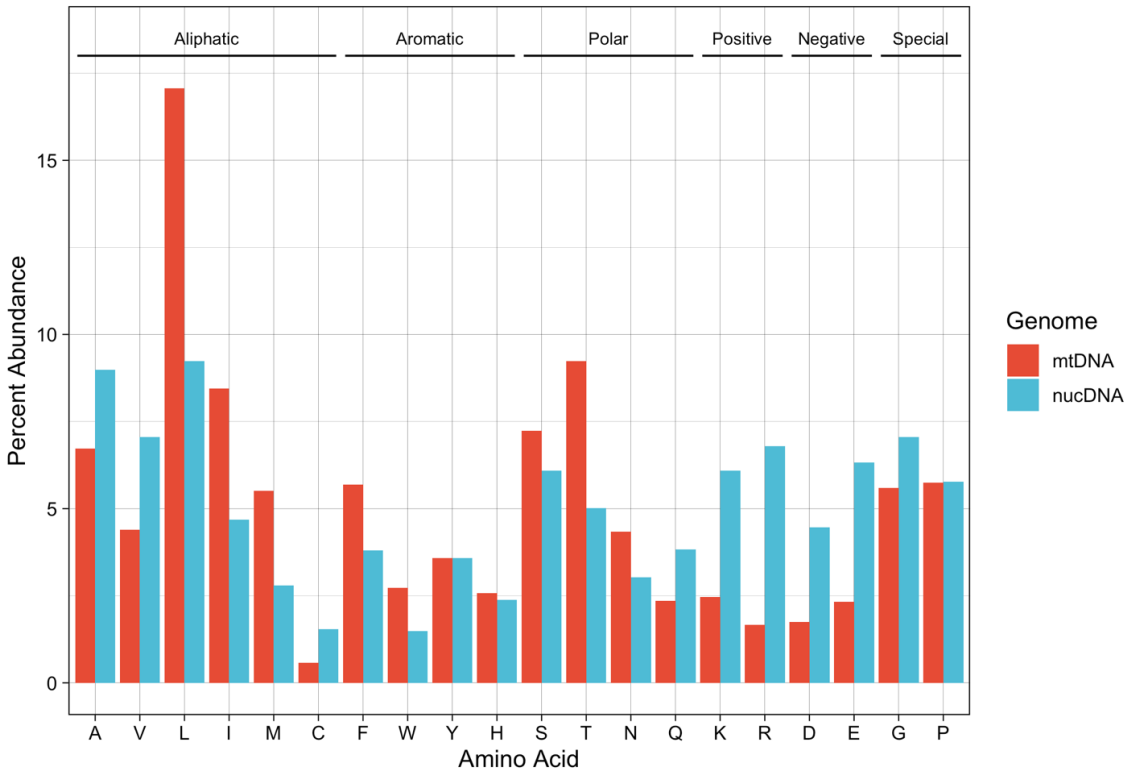

Figure 8. Amino-acid composition for the mtDNA and nucDNA OXPHOS genes in *Homo sapiens.*

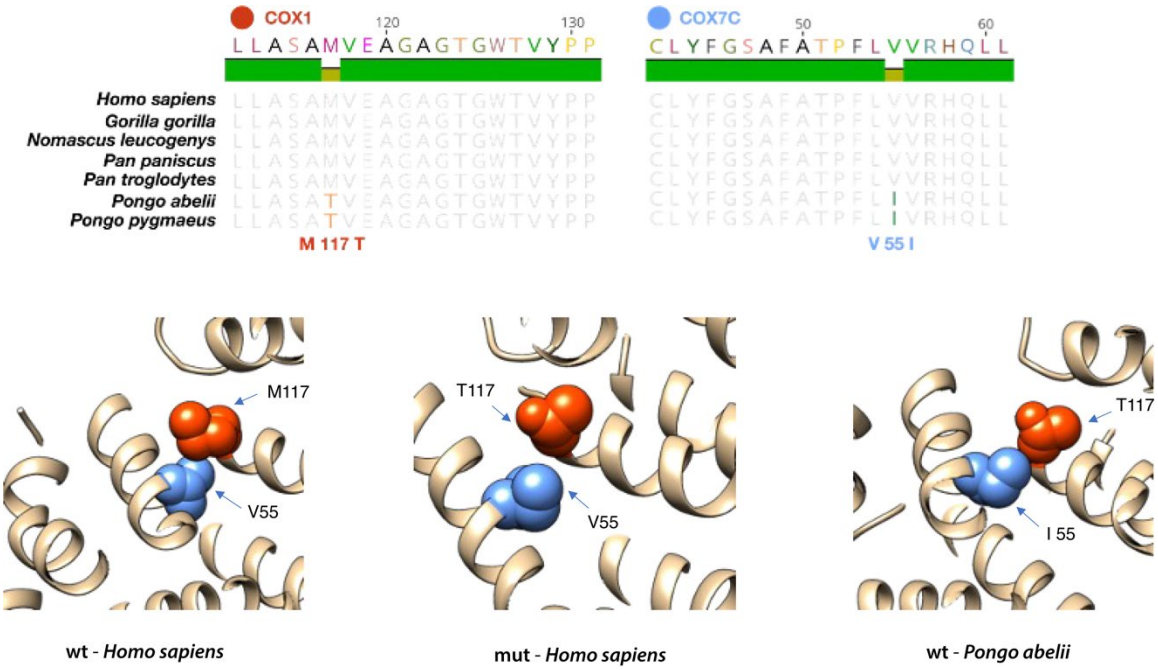

Figure 9. Compensation of a deleterious mutation associated with prostate cancer in humans: Methionine-117 in the mitochondrially-encoded MT-CO1 interacts with Valine-55 in the nuclear-encoded COX7C. Mutation of Methionine-117 to Threonine-117 abolishes the interaction with Valine-55 and presumably plays a role in inducing prostate cancer in humans. However, MT-CO1: Threonine-117 is present as a wild-type amino acid in *Pongo abelii* and *P. pygmaeus* but is also accompanied by a mutation from Valine-55 to Isoleucine-55 in their respective COX7C sequences. The mutation of Valine-55 to Isoleucine-55 re-establishes the non-bond contact with Threonine-117, thereby presumably leading to a non-disease phenotype.

Table 1. The proportion of DARs that appear as compensated in non-human mammals for nuclear (nucDNA) and mitochondrial (mtDNA) OXPHOS proteins.

Genome	# Amino Acids	# DARs	# c-DARs	% c-DARs
nucDNA	15361	419	57	13.60
mtDNA	3776	293	145	49.48

Table 2. Proportion of compensated Disease Associated Residues (c-DARs) identified in previous studies

Study	# Species	# DARs	# c-DARs	% c-DARs
Kondrashov et al., 2002	Not reported	608	4880	12.4
Ferrer-Costa et al., 2007*	Not reported	847	9426	9.8
Barešić et al., 2010	Not reported	453	2328	19.5
Yan et al., 2011*	4	550	50721	1.08
Xu & Zhang, 2014	Not reported	1077	9212	11.6
Jordan et al., 2015*	57	3705	69905	5.3
Azevedo et al., 2016*	39	1964	52765	3.7
Present study (nucDNA) *	~ 120	57	419	13.6
Present study (mtDNA) *	1062	145	293	49.4

* - Mammalian subsets reported

Table 3 Distribution of mtDNA (a) and nucDNA (b) DARs on protein-protein interfaces

a) For DARs in mtDNA -

DAR-class	Mito-nuc interface	Mito-mito interface	Non interface	Total
c-DAR	46	27	65	138
u-DAR	39	34	67	140
Total	85	61	132	278

b) For DARs in nucDNA -

DAR-class	Mito-nuc interface	Nuc-nuc interface	Non interface	Total
c-DAR	2	16	33	51
u-DAR	15	140	184	339
Total	17	156	217	390

Table 4 List of PDB IDs of three-dimensional protein complex structures used for structural mapping of DARs

OXPHOS Complex	**PDB ID**	**Organism**	**Reference**
Complex I	5XTD	*Homo sapiens*	Guo et al., 2017
Complex II	1ZOY	*Sus scrofa*	Sun et al., 2005
Complex III	5XTE	*Homo sapiens*	Guo et al., 2017
Complex IV	5Z62	*Homo sapiens*	Zong et al., 2018
Complex V	5ARA	*Bos taurus*	Zhou et al., 2015

Supplementary Table S1 RefSeq accession numbers for the mtDNA genomes for the species used in the study

Species name	RefSeq accession	Species name	RefSeq accession
Acinonyx jubatus	NC_005212.1	*Mammuthus primigenius*	NC_007596.2
Acomys cahirinus	NC_020758.1	*Mandrillus leucophaeus*	NC_028442.1
Addax nasomaculatus	NC_020674.1	*Mandrillus sphinx*	NC_021956.1
Aepyceros melampus	NC_020675.1	*Manis javanica*	NC_026781.1
Ailuropoda melanoleuca	NC_009492.1	*Manis pentadactyla*	NC_016008.1
Ailurus fulgens	NC_011124.1	*Manis temminckii*	NC_025769.1
Ailurus fulgens styani	NC_009691.1	*Manis tetradactyla*	NC_004027.1
Akodon montensis	NC_025746.1	*Manis tricuspis*	NC_026780.1
Alcelaphus buselaphus	NC_020676.1	*Marmota himalayana*	NC_018367.1
Alces alces	NC_020677.1	*Martes americana*	NC_020642.1
Allenopithecus nigroviridis	NC_023965.1	*Martes flavigula*	NC_012141.1
Alouatta caraya	NC_021938.1	*Martes foina*	NC_020643.1
Alouatta seniculus	NC_027825.1	*Martes martes*	NC_021749.1
Ammotragus lervia	NC_009510.1	*Martes melampus*	NC_009678.1
Anomalurus sp. GP-2005	NC_009056.1	*Martes pennanti*	NC_020664.1
Anoura caudifer	NC_022420.1	*Martes zibellina*	NC_011579.1
Anourosorex squamipes	NC_024563.1	*Mazama americana*	NC_020719.1
Antidorcas marsupialis	NC_020678.1	*Mazama gouazoupira*	NC_020720.1
Antilocapra americana	NC_020679.1	*Mazama nemorivaga*	NC_024812.1
Antilope cervicapra	NC_012098.1	*Mazama rufina*	NC_020721.1
Aotus azarai	NC_021939.1	*Megaladapis edwardsi*	NC_026088.1
Aotus azarai azarai	NC_018115.1	*Megaptera novaeangliae*	NC_006927.1
Aotus lemurinus	NC_019799.1	*Meles anakuma*	NC_009677.1
Aotus nancymaae	NC_018116.1	*Meles meles*	NC_011125.1
Apodemus agrarius	NC_016428.1	*Melogale moschata*	NC_020644.1
Apodemus chejuensis	NC_016662.1	*Melursus ursinus*	NC_009970.1
Apodemus chevrieri	NC_017599.1	*Mephitis mephitis*	NC_020648.1
Apodemus draco	NC_019584.1	*Meriones libycus*	NC_027683.1
Apodemus latronum	NC_019585.1	*Meriones meridianus*	NC_027684.1
Apodemus peninsulae	NC_016060.1	*Meriones tamariscinus*	NC_034314.1
Arctocephalus forsteri	NC_004023.1	*Meriones unguiculatus*	NC_023263.1
Arctocephalus pusillus	NC_008417.1	*Mesocricetus auratus*	NC_013276.1
Arctocephalus townsendi	NC_008420.1	*Mesoplodon densirostris*	NC_021974.1
Arctodus simus	NC_011116.1	*Mesoplodon europaeus*	NC_021434.1
Arctonyx collaris	NC_020645.1	*Mesoplodon ginkgodens*	NC_027593.1
Arctotherium sp.	NC_030174.1	*Mesoplodon grayi*	NC_023830.1
Artibeus jamaicensis	NC_002009.1	*Metachirus nudicaudatus*	NC_006516.1
Artibeus lituratus	NC_016871.1	*Microcebus murinus*	NC_028718.1
Ateles belzebuth	NC_019800.1	*Micromys minutus*	NC_027932.1
Avahi laniger	NC_021940.1	*Micronycteris megalotis*	NC_022419.1
Axis axis	NC_020680.1	*Microtus fortis fortis*	NC_015243.1
Axis porcinus	NC_020681.1	*Microtus kikuchii*	NC_003041.1
Balaena mysticetus	NC_005268.1	*Microtus levis*	NC_008064.1

Balaenoptera bonaerensis	NC_006926.1	*Microtus ochrogaster*	NC_027945.1
Balaenoptera borealis	NC_006929.1	*Mirounga leonina*	NC_008422.1
Balaenoptera brydei	NC_006928.1	*Mogera robusta*	NC_029836.1
Balaenoptera edeni	NC_007938.1	*Mogera wogura*	NC_005035.1
Balaenoptera musculus	NC_001601.1	*Monodelphis domestica*	NC_006299.1
Balaenoptera omurai	NC_007937.1	*Monodon monoceros*	NC_005279.1
Balaenoptera physalus	NC_001321.1	*Moschus anhuiensis*	NC_020017.1
Bandicota indica	NC_028335.1	*Moschus berezovskii*	NC_012694.1
Beatragus hunteri	NC_023542.1	*Moschus chrysogaster*	NC_020093.1
Berardius bairdii	NC_005274.1	*Moschus moschiferus*	NC_013753.1
Bison bonasus	NC_014044.1	*Mungotictis decemlineata*	NC_027828.1
Bison priscus	NC_027233.1	*Muntiacus crinifrons*	NC_004577.1
Bison schoetensacki	NC_033873.1	*Muntiacus muntjak*	NC_004563.1
Blarinella quadraticauda	NC_023950.1	*Muntiacus reevesi*	NC_004069.1
Blastocerus dichotomus	NC_020682.1	*Muntiacus reevesi micrurus*	NC_008491.1
Bootherium bombifrons	NC_032352.1	*Muntiacus vuquangensis*	NC_016920.1
Bos gaurus	NC_024818.1	*Murina leucogaster*	NC_025949.1
Bos indicus	NC_005971.1	*Murina ussuriensis*	NC_021119.1
Bos javanicus	NC_012706.1	*Mus caroli*	NC_025268.1
Bos mutus	NC_025563.1	*Mus cervicolor*	NC_025269.1
Bos primigenius	NC_013996.1	*Mus cookii*	NC_025270.1
Bos taurus	NC_006853.1	*Mus famulus*	NC_030342.1
Boselaphus tragocamelus	NC_020614.1	*Mus fragilicauda*	NC_025287.1
Brachyphylla cavernarum	NC_022421.1	*Mus musculus molossinus*	NC_006915.1
Bradypus pygmaeus	NC_028554.1	*Mus spretus*	NC_025952.1
Bradypus torquatus	NC_028555.1	*Mus terricolor*	NC_010650.1
Bradypus tridactylus	NC_006923.1	*Mustela altaica*	NC_021751.1
Bradypus variegatus	NC_028501.1	*Mustela erminea*	NC_025516.1
Bubalus bubalis	NC_006295.1	*Mustela eversmannii*	NC_028013.1
Bubalus depressicornis	NC_020615.1	*Mustela frenata*	NC_020640.1
Budorcas taxicolor	NC_013069.1	*Mustela itatsi*	NC_034330.1
Cabassous centralis	NC_028556.1	*Mustela kathiah*	NC_023210.1
Cabassous chacoensis	NC_028557.1	*Mustela nigripes*	NC_024942.1
Cabassous tatouay	NC_028558.1	*Mustela nivalis*	NC_020639.1
Cabassous unicinctus	NC_028559.1	*Mustela sibirica*	NC_020637.1
Cacajao calvus	NC_021967.1	*Myodes glareolus*	NC_024538.1
Caenolestes fuliginosus	NC_005828.1	*Myodes regulus*	NC_016427.1
Callimico goeldii	NC_024628.1	*Myodes rufocanus*	NC_029477.1
Callithrix geoffroyi	NC_021941.1	*Myospalax aspalax*	NC_026915.1
Callithrix jacchus	NC_025586.1	*Myospalax psilurus*	NC_026034.1
Callithrix kuhlii	NC_027658.1	*Myotis bechsteinii*	NC_034227.1
Callithrix penicillata	NC_030788.1	*Myotis bombinus*	NC_029342.1
Callithrix pygmaea	NC_021942.1	*Myotis brandtii*	NC_025308.1
Callorhinus ursinus	NC_008415.3	*Myotis davidii*	NC_025568.1
Callosciurus adamsi	NC_030071.1	*Myotis formosus*	NC_015828.1
Callosciurus erythraeus	NC_025550.1	*Myotis ikonnikovi*	NC_022698.1
Callospermophilus lateralis	NC_031210.1	*Myotis lucifugus*	NC_029849.1
Calyptophractus retusus	NC_028560.1	*Myotis macrodactylus*	NC_022694.1

Camelus bactrianus	NC_009628.2	*Myotis muricola*	NC_029422.1
Camelus dromedarius	NC_009849.1	*Myotis myotis*	NC_029346.1
Camelus ferus	NC_009629.2	*Myotis petax*	NC_027973.1
Canis anthus	NC_027956.1	*Myrmecobius fasciatus*	NC_011949.1
Canis latrans	NC_008093.1	*Myrmecophaga tridactyla*	NC_028572.1
Canis lupus	NC_008092.1	*Mystacina tuberculata*	NC_006925.1
Canis lupus chanco	NC_010340.2	*Naemorhedus baileyi*	NC_020722.1
Canis lupus familiaris	NC_002008.4	*Naemorhedus caudatus*	NC_013751.1
Canis lupus laniger	NC_011218.2	*Naemorhedus goral*	NC_021381.1
Canis lupus lupus	NC_009686.1	*Naemorhedus griseus*	NC_020723.1
Caperea marginata	NC_005269.1	*Nandinia binotata*	NC_024567.1
Capra aegagrus	NC_028161.1	*Nanger dama*	NC_020724.1
Capra caucasica	NC_020683.1	*Nanger granti*	NC_020725.1
Capra falconeri	NC_020622.1	*Nanger soemmerringii*	NC_020726.1
Capra hircus	NC_005044.2	*Nannospalax ehrenbergi*	NC_005315.1
Capra ibex	NC_020623.1	*Nannospalax galili*	NC_020754.1
Capra nubiana	NC_020624.1	*Nannospalax golani*	NC_020757.1
Capra pyrenaica	NC_020625.1	*Nannospalax judaei*	NC_020755.1
Capra sibirica	NC_020626.1	*Nasalis larvatus*	NC_008216.1
Capreolus capreolus	NC_020684.1	*Nasua nasua*	NC_020647.1
Capreolus pygargus	NC_025271.1	*Nectogale elegans*	NC_023351.1
Capricornis crispus	NC_012096.1	*Neodon irene*	NC_016055.1
Capricornis milneedwardsii	NC_023457.1	*Neofelis nebulosa*	NC_008450.1
Capricornis sp. YZ-2016	NC_030179.1	*Neomonachus schauinslandi*	NC_008421.1
Capricornis sumatraensis	NC_020629.1	*Neomys fodiens*	NC_025559.1
Capricornis swinhoei	NC_010640.1	*Neophoca cinerea*	NC_008419.1
Caracal caracal	NC_028306.1	*Neophocaena asiaeorientalis*	NC_026456.1
Carlito syrichta	NC_012774.1	*Neophocaena phocaenoides*	NC_021461.1
Carollia perspicillata	NC_022422.1	*Neotetracus sinensis*	NC_019626.1
Castor canadensis	NC_033912.1	*Neotoma fuscipes*	NC_033356.1
Castor fiber	NC_028625.1	*Neotragus batesi*	NC_020727.1
Catopuma badia	NC_028300.1	*Neotragus moschatus*	NC_020728.1
Catopuma temminckii	NC_027115.1	*Neovison vison*	NC_020641.1
Cavia porcellus	NC_000884.1	*Niviventer confucianus*	NC_023960.1
Cebus albifrons	NC_002763.1	*Niviventer excelsior*	NC_019617.1
Cephalopachus bancanus	NC_002811.1	*Niviventer fulvescens*	NC_028715.1
Cephalophus adersi	NC_020685.1	*Nomascus gabriellae*	NC_018753.1
Cephalophus callipygus	NC_020686.1	*Nomascus leucogenys*	NC_021957.1
Cephalophus dorsalis	NC_020687.1	*Nomascus siki*	NC_014051.1
Cephalophus jentinki	NC_020688.1	*Notoryctes typhlops*	NC_006522.1
Cephalophus leucogaster	NC_020689.1	*Nyctalus noctula*	NC_027237.1
Cephalophus natalensis	NC_020690.1	*Nyctereutes procyonoides*	NC_013700.1
Cephalophus nigrifrons	NC_020691.1	*Nycticebus bengalensis*	NC_021958.1
Cephalophus ogilbyi	NC_020692.1	*Nycticebus coucang*	NC_002765.1
Cephalophus rufilatus	NC_020693.1	*Nycticebus pygmaeus*	NC_033381.1
Cephalophus silvicultor	NC_020694.1	*Ochotona collaris*	NC_003033.1
Cephalophus spadix	NC_020695.1	*Ochotona curzoniae*	NC_011029.1
Cephalorhynchus heavisidii	NC_020696.1	*Ochotona princeps*	NC_005358.1

Cercocebus atys	NC_028592.1	*Octodon degus*	NC_020661.1
Cercocebus chrysogaster	NC_021943.1	*Odocoileus hemionus*	NC_020729.1
Cercocebus torquatus	NC_023964.1	*Okapia johnstoni*	NC_020730.1
Cercopithecus albogularis	NC_021944.1	*Onychomys leucogaster*	NC_029760.1
Cercopithecus diana	NC_023963.1	*Orcaella brevirostris*	NC_019590.1
Cercopithecus mitis	NC_023961.1	*Orcaella heinsohni*	NC_019591.1
Cervus canadensis songaricus	NC_014703.1	*Orcinus orca*	NC_023889.1
Cervus elaphus	NC_007704.2	*Oreamnos americanus*	NC_020630.1
Cervus nippon centralis	NC_006993.1	*Oreotragus oreotragus*	NC_020731.1
Cervus nippon hortulorum	NC_013834.1	*Ornithorhynchus anatinus*	NC_000891.1
Cervus nippon kopschi	NC_016178.1	*Oryctolagus cuniculus*	NC_001913.1
Cervus nippon sichuanicus	NC_018595.1	*Oryx beisa*	NC_020793.1
Cervus nippon taiouanus	NC_008462.1	*Oryx dammah*	NC_016421.1
Cervus nippon yakushimae	NC_007179.1	*Oryx gazella*	NC_016422.1
Cervus nippon yesoensis	NC_006973.1	*Oryx leucoryx*	NC_020732.1
Chaetophractus vellerosus	NC_028561.1	*Otocolobus manul*	NC_028323.1
Chaetophractus villosus	NC_028562.1	*Otolemur crassicaudatus*	NC_012762.1
Chalinolobus tuberculatus	NC_002626.1	*Ourebia ourebi*	NC_020733.1
Cheirogaleus medius	NC_021945.1	*Ovibos moschatus*	NC_020631.1
Cheracebus lugens	NC_024630.1	*Ovis ammon*	NC_020656.1
Chinchilla lanigera	NC_021386.1	*Ovis canadensis*	NC_015889.1
Chiropotes albinasus	NC_021946.1	*Ovis orientalis*	NC_026063.1
Chiropotes israelita	NC_024629.1	*Ovis vignei*	NC_026064.1
Chlamyphorus truncatus	NC_028563.1	*Ozotoceros bezoarticus*	NC_020766.1
Chlorocebus aethiops	NC_007009.1	*Paguma larvata*	NC_029403.1
Chlorocebus cynosuros	NC_024933.1	*Palaeopropithecus ingens*	NC_026090.1
Chlorocebus djamdjamensis	NC_034277.1	*Pan paniscus*	-
Chlorocebus pygerythrus	NC_009747.1	*Pan troglodytes*	NC_001643.1
Chlorocebus sabaeus	NC_008066.1	*Panthera leo*	NC_028302.1
Chlorocebus tantalus	NC_009748.1	*Panthera leo persica*	NC_018053.1
Choloepus didactylus	NC_006924.1	*Panthera onca*	NC_022842.1
Choloepus hoffmanni	NC_027964.1	*Panthera pardus*	NC_010641.1
Chrysochloris asiatica	NC_004920.1	*Panthera tigris amoyensis*	NC_014770.1
Chrysocyon brachyurus	NC_024172.1	*Panthera uncia*	NC_010638.1
Civettictis civetta	NC_033378.1	*Pantholops hodgsonii*	NC_007441.1
Coelodonta antiquitatis	NC_012681.1	*Papio anubis*	NC_020006.2
Coendou insidiosus	NC_021387.1	*Papio cynocephalus*	NC_020007.2
Colobus guereza	NC_006901.1	*Papio hamadryas*	NC_001992.1
Condylura cristata	NC_029762.1	*Papio kindae*	NC_020008.2
Connochaetes gnou	NC_020698.1	*Papio papio*	NC_020009.2
Connochaetes taurinus	NC_020699.1	*Papio ursinus*	NC_020010.2
Cricetulus griseus	NC_007936.1	*Pardofelis marmorata*	NC_028303.1
Cricetulus kamensis	NC_024592.1	*Pecari tajacu*	NC_012103.1
Cricetulus longicaudatus	NC_025330.1	*Pelea capreolus*	NC_020734.1
Cricetulus migratorius	NC_031802.1	*Peponocephala electra*	NC_019589.1
Crocidura attenuata	NC_026204.2	*Perameles gunnii*	NC_006521.1
Crocidura beatus	NC_027249.1	*Perodicticus potto*	NC_012764.1
Crocidura grayi	NC_027247.1	*Petaurista alborufus*	NC_023922.1

Crocidura lasiura	NC_029329.1	*Petaurista hainana*	NC_023089.1
Crocidura mindorus	NC_027248.1	*Petaurista yunanensis*	NC_033902.1
Crocidura negrina	NC_027245.1	*Petaurus breviceps*	NC_008135.1
Crocidura ninoyi	NC_027244.1	*Phacochoerus africanus*	NC_008830.1
Crocidura orientalis	NC_027242.1	*Phalanger vestitus*	NC_008137.1
Crocidura palawanensis	NC_027243.1	*Phascogale tapoatafa*	NC_006523.1
Crocidura panayensis	NC_027246.1	*Phascolarctos cinereus*	NC_008133.1
Crocidura russula	NC_006893.1	*Philantomba maxwellii*	NC_020735.1
Crocidura shantungensis	NC_021398.1	*Philantomba monticola*	NC_020736.1
Crocuta crocuta	NC_020670.1	*Phoca fasciata*	NC_008428.1
Ctenomys leucodon	NC_020659.1	*Phoca groenlandica*	NC_008429.1
Ctenomys sociabilis	NC_020658.1	*Phoca largha*	NC_008430.1
Cuon alpinus	NC_013445.1	*Phoca vitulina*	NC_001325.1
Cyclopes didactylus	NC_028564.1	*Phocarctos hookeri*	NC_008418.1
Cynomys leucurus	NC_026705.1	*Phocoena phocoena*	NC_005280.1
Cynomys ludovicianus	NC_026706.1	*Phodopus roborovskii*	NC_031809.1
Cynopterus brachyotis	NC_026465.1	*Physeter catodon*	NC_002503.2
Cystophora cristata	NC_008427.1	*Piliocolobus badius*	NC_008219.1
Dactylomys dactylinus	NC_029876.1	*Pipistrellus abramus*	NC_005436.1
Dactylopsila trivirgata	NC_008134.1	*Pipistrellus coromandra*	NC_029191.1
Dama dama	NC_020700.1	*Platanista minor*	NC_005275.1
Dama mesopotamica	NC_024819.1	*Plecotus auritus*	NC_015484.1
Damaliscus lunatus	NC_023543.1	*Plecotus macrobullaris*	NC_027977.1
Damaliscus pygargus	NC_020627.1	*Plecturocebus cupreus*	NC_021965.1
Dasypus hybridus	NC_028565.1	*Plecturocebus donacophilus*	NC_019801.1
Dasypus kappleri	NC_028566.1	*Pongo abelii*	NC_002083.1
Dasypus novemcinctus	NC_001821.1	*Pongo pygmaeus*	NC_001646.1
Dasypus pilosus	NC_028567.1	*Pontoporia blainvillei*	NC_005277.1
Dasypus sabanicola	NC_028568.1	*Potamochoerus porcus*	NC_020737.1
Dasypus septemcinctus	NC_028569.1	*Potorous tridactylus*	NC_006524.1
Dasypus yepesi	NC_028570.1	*Presbytis melalophos*	NC_008217.1
Dasyurus hallucatus	NC_007630.1	*Priodontes maximus*	NC_028573.1
Daubentonia madagascariensis	NC_010299.1	*Prionailurus bengalensis*	NC_028301.1
Delphinapterus leucas	NC_034236.1	*Prionailurus bengalensis euptilurus*	NC_016189.1
Delphinus capensis	NC_012061.1	*Prionailurus planiceps*	NC_028312.1
Dendrohyrax dorsalis	NC_010301.1	*Prionailurus rubiginosus*	NC_028304.1
Desmodus rotundus	NC_022423.1	*Prionailurus viverrinus*	NC_028305.1
Dicerorhinus sumatrensis	NC_012684.1	*Prionodon pardicolor*	NC_024569.1
Diceros bicornis	NC_012682.1	*Procapra gutturosa*	NC_020738.1
Dicrostonyx groenlandicus	NC_034313.1	*Procapra przewalskii*	NC_014875.1
Dicrostonyx hudsonius	NC_034307.1	*Procavia capensis*	NC_004919.1
Dicrostonyx torquatus	NC_034646.1	*Procolobus verus*	NC_020666.1
Didelphis virginiana	NC_001610.1	*Procyon lotor*	NC_009126.1
Dipus sagitta	NC_027499.1	*Proechimys longicaudatus*	NC_020657.1
Distoechurus pennatus	NC_008145.1	*Proedromys liangshanensis*	NC_013563.1
Dorcatragus megalotis	NC_020701.1	*Profelis aurata*	NC_028299.1
Dremomys rufigenis	NC_026442.1	*Prolemur simus*	NC_021959.1

Dromiciops gliroides	NC_005826.1	*Propithecus coquereli*	NC_011053.1
Dugong dugon	NC_003314.1	*Propithecus diadema*	NC_026084.1
Echinops telfairi	NC_002631.2	*Propithecus edwardsi*	NC_026086.1
Echinosorex gymnura	NC_002808.1	*Propithecus tattersalli*	NC_027740.1
Echymipera rufescens australis	NC_007632.1	*Propithecus verreauxi*	NC_028210.1
Elaphodus cephalophus	NC_008749.1	*Przewalskium albirostris*	NC_016707.1
Elaphurus davidianus	NC_018358.1	*Pseudocheirus peregrinus*	NC_006519.1
Elephantulus sp. VB001	NC_004921.1	*Pseudois nayaur*	NC_020632.1
Elephas maximus	NC_005129.2	*Pseudois schaeferi*	NC_016689.1
Enhydra lutris	NC_009692.1	*Pseudomys chapmani*	NC_014698.1
Eospalax fontanierii baileyi	NC_018098.1	*Pseudorca crassidens*	NC_019577.1
Eospalax fontanierii cansus	NC_021129.1	*Pseudoryx nghetinhensis*	NC_020616.1
Eospalax rothschildi	NC_018535.1	*Pteromys volans*	NC_019612.1
Eothenomys chinensis	NC_013571.1	*Pteronotus parnellii*	NC_023368.1
Eothenomys melanogaster	NC_027418.1	*Pteronotus personatus*	NC_033353.1
Eothenomys miletus	NC_030330.1	*Pteronotus rubiginosus*	NC_022425.1
Eozapus setchuanus	NC_027578.1	*Pteropus alecto*	NC_023122.1
Episoriculus caudatus	NC_026131.1	*Pteropus dasymallus*	NC_002612.1
Episoriculus fumidus	NC_003040.1	*Pteropus scapulatus*	NC_002619.1
Episoriculus macrurus	NC_029840.1	*Pteropus vampyrus*	NC_026542.1
Epomophorus gambianus	NC_029375.1	*Pudu mephistophiles*	NC_020739.1
Eptesicus serotinus	NC_022474.1	*Pudu puda*	NC_020740.1
Equus asinus	NC_001788.1	*Puma concolor*	NC_016470.1
Equus caballus	NC_001640.1	*Puma yagouaroundi*	NC_028311.1
Equus grevyi	NC_020432.2	*Pusa caspica*	NC_008431.1
Equus hemionus	NC_016061.1	*Pusa hispida*	NC_008433.1
Equus hemionus kulan	NC_018782.1	*Pusa sibirica*	NC_008432.1
Equus kiang	NC_020433.1	*Pygathrix cinerea 1 RL-2012*	NC_018062.1
Equus ovodovi	NC_018783.1	*Pygathrix cinerea 2 RL-2012*	NC_018063.1
Equus przewalskii	NC_024030.1	*Pygathrix nemaeus*	NC_008220.1
Equus zebra	NC_020476.1	*Pygathrix nigripes*	NC_018061.1
Equus zebra hartmannae	NC_018780.1	*Rangifer tarandus*	NC_007703.1
Eremitalpa granti	NC_010304.1	*Raphicerus campestris*	NC_020741.1
Erignathus barbatus	NC_008426.1	*Rattus exulans*	NC_012389.1
Erinaceus europaeus	NC_002080.2	*Rattus fuscipes*	NC_014867.1
Erythrocebus patas	NC_021947.1	*Rattus leucopus*	NC_014855.1
Eschrichtius robustus	NC_005270.1	*Rattus lutreolus*	NC_014858.1
Eubalaena australis	NC_006930.1	*Rattus niobe*	NC_023347.1
Eubalaena japonica	NC_006931.1	*Rattus nitidus*	NC_032286.1
Euchoreutes naso	NC_027500.1	*Rattus praetor*	NC_012461.1
Eudorcas rufifrons	NC_020702.1	*Rattus rattus*	NC_012374.1
Eulemur fulvus fulvus	NC_012766.1	*Rattus sordidus*	NC_014871.1
Eulemur fulvus mayottensis	NC_012769.1	*Rattus tanezumi*	NC_011638.1
Eulemur macaco macaco	NC_012771.1	*Rattus tiomanicus*	NC_029888.1
Eulemur mongoz	NC_010300.1	*Rattus tunneyi*	NC_014861.1
Eulemur rubriventer	NC_026098.1	*Rattus villosissimus*	NC_014864.1
Eulemur rufus	NC_021948.1	*Ratufa bicolor*	NC_023780.1
Eumetopias jubatus	NC_004030.2	*Redunca arundinum*	NC_020794.1

Euphractus sexcinctus	NC_028571.1	*Redunca fulvorufula*	NC_020742.1
Exilisciurus exilis	NC_030072.1	*Rhinoceros sondaicus*	NC_012683.1
Felis catus	NC_001700.1	*Rhinoceros unicornis*	NC_001779.1
Felis chaus	NC_028307.1	*Rhinolophus ferrumequinum korai*	NC_016191.1
Felis margarita	NC_028308.1	*Rhinolophus ferrumequinum quelpartis*	NC_020326.1
Felis nigripes	NC_028309.1	*Rhinolophus formosae*	NC_011304.1
Felis silvestris	NC_028310.1	*Rhinolophus luctus*	NC_018539.1
Feresa attenuata	NC_019588.1	*Rhinolophus macrotis*	NC_026460.1
Fukomys damarensis	NC_027742.1	*Rhinolophus monoceros*	NC_005433.1
Galago moholi	NC_021949.1	*Rhinolophus pumilus*	NC_005434.1
Galago senegalensis	NC_012761.1	*Rhinolophus rex*	NC_028536.1
Galemys pyrenaicus	NC_008156.1	*Rhinolophus thomasi*	NC_034306.1
Galeopterus variegatus	NC_004031.1	*Rhinophylla pumilio*	NC_022426.1
Gazella bennettii	NC_020703.1	*Rhinopithecus avunculus*	NC_015485.1
Gazella cuvieri	NC_020704.1	*Rhinopithecus bieti*	NC_015486.1
Gazella dorcas	NC_020705.1	*Rhinopithecus bieti 1 RL-2012*	NC_018058.1
Gazella erlangeri	NC_020706.1	*Rhinopithecus bieti 2 RL-2012*	NC_018060.1
Gazella gazella	NC_020707.1	*Rhinopithecus brelichi*	NC_018057.1
Gazella leptoceros	NC_020708.1	*Rhinopithecus roxellana*	NC_008218.1
Gazella spekei	NC_020709.1	*Rhinopithecus strykeri*	NC_018059.1
Gazella subgutturosa	NC_020710.1	*Rhizomys pruinosus*	NC_021478.1
Genetta servalina	NC_024568.1	*Rhizomys sinensis*	NC_026124.1
Giraffa camelopardalis	NC_024820.1	*Rhyncholestes raphanurus*	NC_005829.1
Glis glis	NC_001892.1	*Rousettus aegyptiacus*	NC_007393.1
Globicephala macrorhynchus	NC_019578.2	*Rucervus duvaucelii*	NC_020743.1
Globicephala melas	NC_019441.1	*Rucervus eldii*	NC_014701.1
Gorilla gorilla	NC_001645.1	*Rupicapra pyrenaica*	NC_020789.1
Gorilla gorilla gorilla	NC_011120.1	*Rupicapra rupicapra*	NC_020633.1
Grampus griseus	NC_012062.1	*Rusa alfredi*	NC_020744.1
Gulo gulo	NC_009685.1	*Rusa timorensis*	NC_020745.1
Halichoerus grypus	NC_001602.1	*Rusa unicolor*	NC_031835.1
Hapalemur griseus	NC_021950.1	*Rusa unicolor swinhoei*	NC_008414.3
Helarctos malayanus	NC_009968.1	*Saguinus oedipus*	NC_021960.1
Hemiechinus auritus	NC_005033.1	*Saiga tatarica*	NC_020746.1
Hemitragus jemlahicus	NC_020628.1	*Saimiri boliviensis*	NC_021966.1
Hexaprotodon liberiensis	NC_020697.1	*Saimiri boliviensis boliviensis*	NC_018096.1
Hippocamelus antisensis	NC_020711.1	*Saimiri oerstedii citrinellus*	NC_023211.1
Hippopotamus amphibius	NC_000889.1	*Saimiri sciureus*	NC_012775.1
Hipposideros armiger	NC_018540.1	*Sapajus xanthosternos*	NC_021961.1
Hippotragus equinus	NC_020712.1	*Sarcophilus harrisii*	NC_018788.1
Hippotragus niger	NC_020713.1	*Scapanulus oweni*	NC_025777.1
Homo heidelbergensis	NC_023100.1	*Sciurus vulgaris*	NC_002369.1
Homo sapiens	NC_012920.1	*Semnopithecus entellus*	NC_008215.1
Homo sapiens neanderthalensis	NC_011137.1	*Sicista concolor*	NC_027579.1

Hoolock hoolock	NC_033885.1	*Simias concolor*	NC_020667.1
Hoolock leuconedys	NC_033882.1	*Sminthopsis crassicaudata*	NC_007631.1
Hoolock tianxing	NC_033884.1	*Sminthopsis douglasi*	NC_006517.1
Hyaena hyaena	NC_020669.1	*Sorex araneus*	NC_027963.1
Hydropotes inermis	NC_011821.1	*Sorex cylindricauda*	NC_025278.1
Hydropotes inermis argyropus	NC_018032.1	*Sorex roboratus*	NC_034808.1
Hydrurga leptonyx	NC_008425.1	*Sorex tundrensis*	NC_025327.1
Hyemoschus aquaticus	NC_020714.1	*Sorex unguiculatus*	NC_005435.1
Hylobates agilis	NC_014042.1	*Sousa chinensis*	NC_012057.1
Hylobates lar	NC_002082.1	*Spalacopus cyanus*	NC_020660.1
Hylobates pileatus	NC_014045.1	*Spalax carmeli*	NC_020756.1
Hylomys suillus	NC_010298.1	*Spermophilus dauricus*	NC_027283.1
Hylopetes alboniger	NC_031847.1	*Spilogale putorius*	NC_010497.1
Hylopetes phayrei	NC_026443.1	*Stenella attenuata*	NC_012051.1
Hyperoodon ampullatus	NC_005273.1	*Stenella coeruleoalba*	NC_012053.1
Hypsugo alaschanicus	NC_029939.1	*Stenella longirostris*	NC_032301.1
Ictidomys tridecemlineatus	NC_027278.1	*Sturnira tildae*	NC_022427.1
Indopacetus pacificus	NC_034348.1	*Stylodipus telum*	NC_027692.1
Indri indri	NC_026095.1	*Suncus murinus*	NC_024604.1
Inia geoffrensis	NC_005276.1	*Sus barbatus*	NC_026992.1
Isoodon macrourus	NC_002746.1	*Sus cebifrons*	NC_023541.1
Jaculus jaculus	NC_005314.1	*Sus celebensis*	NC_024860.1
Kobus ellipsiprymnus	NC_020715.1	*Sus scrofa*	NC_000845.1
Kobus leche	NC_018603.1	*Sus scrofa domesticus*	NC_012095.1
Kogia breviceps	NC_005272.1	*Sus scrofa taivanus*	NC_014692.1
Lagenorhynchus albirostris	NC_005278.1	*Sus verrucosus*	NC_023536.1
Lagorchestes hirsutus	NC_008136.1	*Sylvicapra grimmia*	NC_020747.1
Lagostrophus fasciatus	NC_008447.1	*Symphalangus syndactylus*	NC_014047.1
Lagothrix lagotricha	NC_021951.1	*Syncerus caffer*	NC_020617.1
Lama glama	NC_012102.1	*Tachyglossus aculeatus*	NC_003321.1
Lama guanicoe	NC_011822.1	*Talpa europaea*	NC_002391.1
Laonastes aenigmamus	NC_030184.1	*Tamandua mexicana*	NC_028574.1
Lariscus insignis	NC_030070.1	*Tamandua tetradactyla*	NC_004032.1
Lasiopodomys mandarinus	NC_025283.1	*Tamias canipes*	NC_032372.1
Lasiurus borealis	NC_016873.1	*Tamias cinereicollis*	NC_032374.1
Leggadina lakedownensis	NC_014696.1	*Tamias dorsalis*	NC_032373.1
Lemur catta	NC_004025.1	*Tamias quadrivittatus*	NC_032370.1
Leontopithecus rosalia	NC_021952.1	*Tamias rufus*	NC_032371.1
Leopardus colocolo	NC_028314.1	*Tamias sibiricus*	NC_025277.1
Leopardus geoffroyi	NC_028320.1	*Tamias striatus*	NC_032375.1
Leopardus guigna	NC_028321.1	*Tamias umbrinus*	NC_032376.1
Leopardus jacobita	NC_028322.1	*Tamiops maritimus*	NC_029325.1
Leopardus pardalis	NC_028315.1	*Tamiops swinhoei*	NC_026875.1
Leopardus tigrinus	NC_028317.1	*Tapirus indicus*	NC_023838.1
Leopardus wiedii	NC_028318.1	*Tarsipes rostratus*	NC_006518.1
Leopoldamys edwardsi	NC_025670.1	*Tarsius dentatus*	NC_024052.1
Lepilemur aeeclis	NC_034730.1	*Tarsius lariang*	NC_024051.1
Lepilemur ahmansoni	NC_034733.1	*Tarsius wallacei*	NC_024053.1

Lepilemur ankaranensis	NC_034717.1	*Taurotragus derbianus*	NC_020618.1
Lepilemur betsileo	NC_034726.1	*Taxidea taxus*	NC_020646.1
Lepilemur dorsalis	NC_034721.1	*Tetracerus quadricornis*	NC_020788.1
Lepilemur edwardsi	NC_034720.1	*Theropithecus gelada*	NC_019802.1
Lepilemur fleuretae	NC_034718.1	*Thryonomys swinderianus*	NC_002658.1
Lepilemur grewcocki	NC_034729.1	*Thylacinus cynocephalus*	NC_011944.1
Lepilemur hollandorum	NC_034738.1	*Thylamys elegans*	NC_005825.1
Lepilemur hubbardorum	NC_014453.1	*Tlacuatzin canescens*	NC_029381.1
Lepilemur jamesi	NC_034736.1	*Tolypeutes matacus*	NC_028575.1
Lepilemur leucopus	NC_034719.1	*Tolypeutes tricinctus*	NC_028576.1
Lepilemur microdon	NC_034739.1	*Tonatia saurophila*	NC_022428.1
Lepilemur milanoii	NC_034725.1	*Trachypithecus cristatus*	NC_023971.1
Lepilemur mittermeieri	NC_034728.1	*Trachypithecus francoisi*	NC_023970.1
Lepilemur otto	NC_034724.1	*Trachypithecus germaini*	NC_019580.1
Lepilemur petteri	NC_034723.1	*Trachypithecus hatinhensis*	NC_019579.1
Lepilemur randrianasoloi	NC_034722.1	*Trachypithecus johnii*	NC_019583.1
Lepilemur ruficaudatus	NC_021953.1	*Trachypithecus obscurus*	NC_006900.1
Lepilemur sahamalazensis	NC_034734.1	*Trachypithecus pileatus*	NC_024529.1
Lepilemur scottorum	NC_034737.1	*Trachypithecus poliocephalus*	NC_034795.1
Lepilemur seali	NC_034731.1	*Trachypithecus shortridgei*	NC_019581.1
Lepilemur septentrionalis	NC_034727.1	*Trachypithecus vetulus*	NC_019582.1
Lepilemur tymerlachsoni	NC_034735.1	*Tragelaphus angasii*	NC_020748.1
Lepilemur wrighti	NC_034732.1	*Tragelaphus eurycerus*	NC_020749.1
Leptailurus serval	NC_028316.1	*Tragelaphus imberbis*	NC_020619.1
Leptonychotes weddellii	NC_008424.1	*Tragelaphus oryx*	NC_020750.1
Lepus americanus	NC_024043.1	*Tragelaphus scriptus*	NC_020751.1
Lepus capensis	NC_015841.1	*Tragelaphus spekii*	NC_020620.1
Lepus coreanus	NC_024259.1	*Tragelaphus strepsiceros*	NC_020752.1
Lepus europaeus	NC_004028.1	*Tragulus kanchil*	NC_020753.1
Lepus granatensis	NC_024042.1	*Tremarctos ornatus*	NC_009969.1
Lepus hainanus	NC_025902.1	*Trichosurus vulpecula*	NC_003039.1
Lepus sinensis	NC_025316.1	*Tscherskia triton*	NC_013068.1
Lepus timidus	NC_024040.1	*Tupaia belangeri*	NC_002521.1
Lepus tolai	NC_025748.1	*Tursiops aduncus*	NC_012058.1
Lepus townsendii	NC_024041.1	*Tursiops australis*	NC_022805.1
Lipotes vexillifer	NC_007629.1	*Tursiops truncatus*	NC_012059.1
Litocranius walleri	NC_020716.1	*Tympanoctomys barrerae*	NC_020792.1
Lobodon carcinophaga	NC_008423.1	*Typhlomys cinereus*	NC_033382.1
Lophocebus aterrimus	NC_021954.1	*Urocitellus richardsonii*	NC_031209.1
Lophostoma silvicolum	NC_022424.1	*Urocyon cinereoargenteus*	NC_026723.1
Loris lydekkerianus	NC_021955.1	*Uropsilus gracilis*	NC_031330.1
Loris tardigradus	NC_012763.1	*Uropsilus soricipes*	NC_023244.1
Loxodonta africana	NC_000934.1	*Uropsilus sp. 4 FT-2015*	NC_028150.1
Loxodonta cyclotis	NC_020759.1	*Urotrichus talpoides*	NC_005034.1
Lutra lutra	NC_011358.1	*Ursus americanus*	NC_003426.1
Lycaon pictus	-	*Ursus arctos*	NC_003427.1
Lynx canadensis	NC_028313.1	*Ursus maritimus*	NC_003428.1
Lynx lynx	NC_027083.1	*Ursus spelaeus*	NC_011112.1

Lynx pardinus	NC_028319.1
Lynx rufus	NC_014456.1
Macaca arctoides	NC_025201.1
Macaca assamensis	NC_023795.1
Macaca cyclopis	NC_027449.1
Macaca fascicularis	NC_012670.1
Macaca fuscata	NC_025513.1
Macaca leonina	NC_027604.1
Macaca leucogenys	NC_031156.1
Macaca mulatta	NC_005943.1
Macaca nemestrina	NC_026976.1
Macaca nigra	NC_026120.1
Macaca silenus	NC_025221.1
Macaca sylvanus	NC_002764.1
Macaca thibetana	NC_011519.1
Macaca tonkeana	NC_025222.1
Macropus giganteus	NC_027424.1
Macroscelides proboscideus	NC_004026.1
Macrotis lagotis	NC_006520.1
Madoqua kirkii	NC_020717.1
Madoqua saltiana	NC_020718.1
Mammut americanum	NC_009574.1
Mammuthus columbi	NC_015529.1
Ursus thibetanus	NC_009971.1
Ursus thibetanus formosanus	NC_009331.1
Ursus thibetanus mupinensis	NC_008753.1
Ursus thibetanus thibetanus	NC_011118.1
Ursus thibetanus ussuricus	NC_011117.1
Vampyrum spectrum	NC_022429.1
Varecia variegata	NC_026085.1
Varecia variegata variegata	NC_012773.1
Vespertilio murinus	NC_033347.1
Vespertilio sinensis	NC_024558.1
Vicugna pacos	NC_002504.1
Vicugna vicugna	NC_013558.1
Viverricula indica	NC_025296.1
Vombatus ursinus	NC_003322.1
Vulpes corsac	NC_023958.1
Vulpes ferrilata	NC_027935.1
Vulpes lagopus	NC_026529.1
Vulpes vulpes	NC_008434.1
Vulpes zerda	NC_023459.1
Wiedomys cerradensis	NC_025747.1
Zaedyus pichiy	NC_028577.1
Zaglossus bruijni	NC_006364.1
Zalophus californianus	NC_008416.1
Ziphius cavirostris	NC_021435.1

REFERENCES

Allio, R., Donega, S., Galtier, N., & Nabholz, B. (2017). Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: Implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Molecular Biology and Evolution*, *34*(March), 2762–2772. https://doi.org/10.1093/molbev/msx197

Azevedo, L., Carneiro, J., van Asch, B., Moleirinho, A., Pereira, F., & Amorim, A. (2009). Epistatic interactions modulate the evolution of mammalian mitochondrial respiratory complex components. *BMC Genomics*, *10*(1), 266. https://doi.org/10.1186/1471-2164-10-266

Azevedo, L., Mort, M., Costa, A. C., Silva, R. M., Quelhas, D., Amorim, A., & Cooper, D. N. (2016). Improving the in silico assessment of pathogenicity for compensated variants. *European Journal of Human Genetics*, *25*(1), 2–7. https://doi.org/10.1038/ejhg.2016.129

Barešić, A., Hopcroft, L. E. M., Rogers, H. H., Hurst, J. M., & Martin, A. C. R. (2010). Compensated Pathogenic Deviations: Analysis of Structural Effects. *Journal of Molecular Biology*, *396*(1), 19–30. https://doi.org/10.1016/j.jmb.2009.11.002

Barešić, A., & Martin, A. C. R. (2011). Compensated pathogenic deviations. *Biomolecular Concepts*, *2*(4), 281–292. https://doi.org/10.1515/bmc.2011.025

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235–242. https://doi.org/10.1093/nar/28.1.235

Brandon, M., Baldi, P., & Wallace, D. C. (2006). Mitochondrial mutations in cancer. *Oncogene*, *25*(34), 4647–4662. https://doi.org/10.1038/sj.onc.1209607

Buglo, E., Sarmiento, E., Martuscelli, N. B., Sant, D. W., Danzi, M. C., Abrams, A. J., Dallman, J. E., & Züchner, S. (2020). Genetic compensation in a stable slc25a46 mutant zebrafish: A case for using F0 CRISPR mutagenesis to study phenotypes caused by inherited disease. *PLOS ONE*, *15*(3), e0230566. https://doi.org/10.1371/journal.pone.0230566

Chakravarty, S., & Varadarajan, R. (1999). Residue depth: A novel parameter for the analysis of protein structure and stability. *Structure (London, England : 1993)*, *7*(7), 723–732. https://doi.org/10.1016/s0969-2126(99)80097-5

Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., & Flouri, T. (2019). ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msz189

De Magalhães, J. P. (2005). Human disease-associated mitochondrial mutations fixed in nonhuman primates. *Journal of Molecular Evolution*, *61*(4), 491–497. https://doi.org/10.1007/s00239-004-0258-6

DePristo, M. A., Weinreich, D. M., & Hartl, D. L. (2005). Missense meanderings in sequence space: A biophysical view of protein evolution. *Nature Reviews Genetics*, *6*(9), 678–687. https://doi.org/10.1038/nrg1672

Domingo, J., Baeza-Centurion, P., & Lehner, B. (2019). The Causes and Consequences of Genetic Interactions (Epistasis). *Annual Review of Genomics and Human Genetics*, *20*, 433–460. https://doi.org/10.1146/annurev-genom-083118-014857

Dunn, S. D., Wahl, L. M., & Gloor, G. B. (2008). Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics*, *24*(3), 333–340. https://doi.org/10.1093/bioinformatics/btm604

El-Brolosy, M. A., & Stainier, D. Y. R. (2017). Genetic compensation: A phenomenon in search of mechanisms. *PLOS Genetics*, *13*(7), e1006780. https://doi.org/10.1371/journal.pgen.1006780

Elliott, H. R., Samuels, D. C., Eden, J. A., Relton, C. L., & Chinnery, P. F. (2008). Pathogenic Mitochondrial DNA Mutations Are Common in the General Population. *The American Journal of Human Genetics*, *83*(2), 254–260. https://doi.org/10.1016/j.ajhg.2008.07.004

Elurbe, D. M., & Huynen, M. A. (2016). The origin of the supernumerary subunits and assembly factors of complex I: A treasure trove of pathway evolution. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1857*(7), 971–979. https://doi.org/10.1016/j.bbabio.2016.03.027

Emperador, S., Pacheu-Grau, D., Bayona-Bafaluy, M. P., Garrido-Pérez, N., Martín-Navarro, A., López-Pérez, M. J., Montoya, J., & Ruiz-Pesini, E. (2015). An MRPS12 mutation modifies aminoglycoside sensitivity caused by 12S rRNA mutations. *Frontiers in Genetics*, *5*. https://www.frontiersin.org/articles/10.3389/fgene.2014.00469

Ferrer-Costa, C., Orozco, M., & Cruz, X. de la. (2007). Characterization of Compensated Mutations in Terms of Structural and Physico-Chemical Properties. *Journal of Molecular Biology*, *365*(1), 249–256. https://doi.org/10.1016/j.jmb.2006.09.053

Gao, L., & Zhang, J. (2003). Why are some human disease-associated mutations fixed in mice? *Trends in Genetics*, *19*(12), 674–678. https://doi.org/10.1016/j.tig.2003.10.006

Gong, L. I., Suchard, M. A., & Bloom, J. D. (2013). Stability-mediated epistasis constrains the evolution of an influenza protein. *ELife*, *2*, e00631. https://doi.org/10.7554/eLife.00631

Gorman, G. S., Schaefer, A. M., Ng, Y., Gomez, N., Blakely, E. L., Alston, C. L., Feeney, C., Horvath, R., Yu-Wai-Man, P., Chinnery, P. F., Taylor, R. W., Turnbull, D. M., & McFarland, R. (2015). Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease. *Annals of Neurology*, *77*(5), 753–759. https://doi.org/10.1002/ana.24362

Guo, R., Zong, S., Wu, M., Gu, J., & Yang, M. (2017). Architecture of Human Mitochondrial Respiratory Megacomplex I2III2IV2. *Cell*, *170*(6), 1247-1257.e12. https://doi.org/10.1016/j.cell.2017.07.050

Halabi, N., Rivoire, O., Leibler, S., & Ranganathan, R. (2009). Protein Sectors: Evolutionary Units of Three-Dimensional Structure. *Cell*, *138*(4), 774–786. https://doi.org/10.1016/j.cell.2009.07.038

Harms, M. J., & Thornton, J. W. (2014). Historical contingency and its biophysical basis in glucocorticoid receptor evolution. *Nature*, *512*(7513), 203–207. https://doi.org/10.1038/nature13410

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Jordan, D. M., Frangakis, S. G., Golzio, C., Cassa, C. A., Kurtzberg, J., Task Force for Neonatal Genomics, T. F. for N., Davis, E. E., Sunyaev, S. R., & Katsanis, N. (2015). Identification of cis-suppression of human disease mutations by comparative genomics. *Nature*, *524*(7564), 225–229. https://doi.org/10.1038/nature14497

Kern, A. D., & Kondrashov, F. A. (2004). Mechanisms and convergence of compensatory evolution in mammalian mitochondrial tRNAs. *Nature Genetics*, *36*(11), 1207–1212. https://doi.org/10.1038/ng1451

Kondrashov, A. S., Sunyaev, S., & Kondrashov, F. A. (2002). Dobzhansky-Muller incompatibilities in protein evolution. *Proc Natl Acad Sci U S A*, *99*(23), 14878–14883. https://doi.org/10.1073/pnas.232565499

Kulathinal, R. J., Bettencourt, B. R., & Hartl, D. L. (2004). Compensated deleterious mutations in insect ganomes. *Science*, *306*(5701), 1553–1554. https://doi.org/10.1126/science.1100522

Lott, M. T., Leipzig, J. N., Derbeneva, O., Michael Xie, H., Chalkia, D., Sarmady, M., Procaccio, V., & Wallace, D. C. (2013). MtDNA variation and analysis using Mitomap and Mitomaster. *Current Protocols in Bioinformatics*, *SUPPL.44*. https://doi.org/10.1002/0471250953.bi0123s44

Marín, Ò., Aguirre, J., & de la Cruz, X. (2019). Compensated pathogenic variants in coagulation factors VIII and IX present complex mapping between molecular impact and hemophilia severity. *Scientific Reports*, *9*(1), 9538. https://doi.org/10.1038/s41598-019-45916-3

Meer, M. V., Kondrashov, A. S., Artzy-Randrup, Y., & Kondrashov, F. A. (2010). Compensatory evolution in mitochondrial tRNAs navigates valleys of low fitness. *Nature*, *464*(7286), 279–282. https://doi.org/10.1038/nature08691

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., … Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, *44*(D1), D733-745. https://doi.org/10.1093/nar/gkv1189

Ortlund, E. A., Bridgham, J. T., Redinbo, M. R., & Thornton, J. W. (2007). Crystal structure of an ancient protein. *Science (New York, N.Y.)*, *317*(5844), 1544–1548. https://doi.org/10.1126/science.1142819

Pais, F. S. M., Ruy, P. de C., Oliveira, G., & Coimbra, R. S. (2014). Assessing the efficiency of multiple sequence alignment programs. *Algorithms for Molecular Biology*, *9*(1). https://doi.org/10.1186/1748-7188-9-4

Petros, J. A., Baumann, A. K., Ruiz-Pesini, E., Amin, M. B., Sun, C. Q., Hall, J., Lim, S., Issa, M. M., Flanders, W. D., Hosseini, S. H., Marshall, F. F., & Wallace, D. C. (2005). MtDNA mutations increase tumorigenicity in prostate cancer.

Proceedings of the National Academy of Sciences, *102*(3), 719–724. https://doi.org/10.1073/pnas.0408894102

Poon, A., & Chao, L. (2005). The Rate of Compensatory Mutation in the DNA Bacteriophage φX174. *Genetics*, *170*(3), 989–999. https://doi.org/10.1534/genetics.104.039438

Poon, A. F. Y., & Chao, L. (2006). Functional Origins of Fitness Effect-Sizes of Compensatory Mutations in the Dna Bacteriophage Øx174. *Evolution*, *60*(10), 2032–2043. https://doi.org/10.1111/j.0014-3820.2006.tb01841.x

Popadin, K. Y., Nikolaev, S. I., Junier, T., Baranova, M., & Antonarakis, S. E. (2013). Purifying Selection in Mammalian Mitochondrial Protein-Coding Genes Is Highly Effective and Congruent with Evolution of Nuclear Genes. *Molecular Biology and Evolution*, *30*(2), 347–355. https://doi.org/10.1093/molbev/mss219

Quirós, P. M., Mottis, A., & Auwerx, J. (2016). Mitonuclear communication in homeostasis and stress. *Nature Reviews Molecular Cell Biology*, *17*(4), 213–226. https://doi.org/10.1038/nrm.2016.23

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rand, D. M., Haney, R. a, & Fry, A. J. (2004). Cytonuclear coevolution: The genomics of cooperation. *Trends in Ecology & Evolution*, *19*(12), 645–653. https://doi.org/10.1016/j.tree.2004.10.003

Rossignol, R., Faustin, B., Rocher, C., Malgat, M., Mazat, J.-P., & Letellier, T. (2003). Mitochondrial threshold effects. *Biochemical Journal*, *370*(3), 751–762. https://doi.org/10.1042/bj20021594

Sali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, *234*, 779–815. https://doi.org/10.1006/jmbi.1993.1626

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., & Serrano, L. (2005). The FoldX web server: An online force field. *Nucleic Acids Research*, *33*(Web Server), W382–W388. https://doi.org/10.1093/nar/gki387

Shah, P., McCandlish, D. M., & Plotkin, J. B. (2015). Contingency and entrenchment in protein evolution under purifying selection. *Proceedings of the National Academy of Sciences*, *112*(25), E3226–E3235. https://doi.org/10.1073/pnas.1412933112

Skladal, D., Halliday, J., & Thorburn, D. R. (2003). Minimum birth prevalence of mitochondrial respiratory chain disorders in children. *Brain*, *126*(8), 1905–1912. https://doi.org/10.1093/brain/awg170

Stenson, P. D., Mort, M., Ball, E. V., Evans, K., Hayden, M., Heywood, S., Hussain, M., Phillips, A. D., & Cooper, D. N. (2017). The Human Gene Mutation Database: Towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Human Genetics*, *136*(6), 665–677. https://doi.org/10.1007/s00439-017-1779-6

Stewart, J. B., & Chinnery, P. F. (2015). The dynamics of mitochondrial DNA heteroplasmy: Implications for human health and disease. *Nature Reviews Genetics*, *16*(9), 530–542. https://doi.org/10.1038/nrg3966

Storz, J. F. (2016). Causes of molecular convergence and parallelism in protein evolution. *Nature Reviews Genetics*, *17*(4), 239–250. https://doi.org/10.1038/nrg.2016.11

Storz, J. F. (2018). Compensatory mutations and epistasis for protein function. *Current Opinion in Structural Biology*, *50*, 18–25. https://doi.org/10.1016/j.sbi.2017.10.009

Sun, F., Huo, X., Zhai, Y., Wang, A., Xu, J., Su, D., Bartlam, M., & Rao, Z. (2005). Crystal structure of mitochondrial respiratory membrane protein Complex II. *Cell*, *121*(7), 1043–1057. https://doi.org/10.1016/j.cell.2005.05.025

Talavera, G., Castresana, J., Kjer, K., Page, R., & Sullivan, J. (2007). Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Systematic Biology*, *56*(4), 564–577. https://doi.org/10.1080/10635150701472164

Tan, K. P., Nguyen, T. B., Patel, S., Varadarajan, R., & Madhusudhan, M. S. (2013). Depth: A web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. *Nucleic Acids Research*, *41*(Web Server issue), W314–W321. https://doi.org/10.1093/nar/gkt503

Tavares, W. C., & Seuánez, H. N. (2017). Disease-associated mitochondrial mutations and the evolution of primate mitogenomes. *PLoS ONE*, *12*(5), 1–23. https://doi.org/10.1371/journal.pone.0177403

Tseng, Y. Y., & Liang, J. (2006). Estimation of Amino Acid Residue Substitution Rates at Local Spatial Regions and Application in Protein Function Inference: A Bayesian Monte Carlo Approach. *Molecular Biology and Evolution*, *23*(2), 421–436. https://doi.org/10.1093/molbev/msj048

Usmanova, D. R., Bogatyreva, N. S., Bernad, J. A., Eremina, A. A., Gorshkova, A. A., Kanevskiy, G. M., Lonishin, L. R., Meister, A. V., Yakupova, A. G., Kondrashov, F. A., & Ivankov, D. N. (2018). Self-consistency test reveals systematic bias in programs for prediction change of stability upon mutation. *Bioinformatics*, *34*(21), 3653–3658. https://doi.org/10.1093/bioinformatics/bty340

Wallace, D. C., & Chalkia, D. (2013). Mitochondrial DNA Genetics and the Heteroplasmy Conundrum in Evolution and Disease. *Cold Spring Harbor Perspectives in Biology*, *5*(11), a021220–a021220. https://doi.org/10.1101/cshperspect.a021220

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., … Lander, E. S. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, *420*(6915), 520–562. https://doi.org/10.1038/nature01262

Wittenhagen, L. M., & Kelley, S. O. (2003). Impact of disease-related mitochondrial mutations on tRNA structure and function. *Trends in Biochemical Sciences*, *28*(11), 605–611. https://doi.org/10.1016/j.tibs.2003.09.006

Wolstenholme, D. R. (1992). Animal Mitochondrial DNA: Structure and Evolution. *International Review of Cytology*, *141*(C), 173–216. https://doi.org/10.1016/S0074-7696(08)62066-5

Wong, L.-J. C. (2007). Pathogenic mitochondrial DNA mutations in protein-coding genes. *Muscle & Nerve*, *36*(3), 279–293. https://doi.org/10.1002/mus.20807

Xu, J., & Zhang, J. (2014). Why human disease-associated residues appear as the wild-type in other species: Genome-scale structural evidence for the compensation hypothesis. *Molecular Biology and Evolution*, *31*(7), 1787–1792. https://doi.org/10.1093/molbev/msu130

Yampolsky, L. Y., & Stoltzfus, A. (2005). The exchangeability of amino acids in proteins. *Genetics*, *170*(4), 1459–1472. https://doi.org/10.1534/genetics.104.039107

Yan, G., Zhang, G., Fang, X., Zhang, Y., Li, C., Ling, F., Cooper, D. N., Li, Q., Li, Y., Van Gool, A. J., Du, H., Chen, J., Chen, R., Zhang, P., Huang, Z., Thompson, J. R., Meng, Y., Bai, Y., Wang, J., … Wang, J. (2011). Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. *Nature Biotechnology*, *29*(11), 1019–1023. https://doi.org/10.1038/nbt.1992

Zhou, A., Rohou, A., Schep, D. G., Bason, J. V., Montgomery, M. G., Walker, J. E., Grigorieff, N., & Rubinstein, J. L. (2015). Structure and conformational states of the bovine mitochondrial ATP synthase by cryo-EM. *ELife*, *4*, e10180. https://doi.org/10.7554/eLife.10180

Zhu, X., Peng, X., Guan, M.-X., & Yan, Q. (2009). Pathogenic mutations of nuclear genes associated with mitochondrial disorders. *Acta Biochimica et Biophysica Sinica*, *41*(3), 179–187. https://doi.org/10.1093/abbs/gmn021

Zong, S., Wu, M., Gu, J., Liu, T., Guo, R., & Yang, M. (2018). Structure of the intact 14-subunit human cytochrome c oxidase. *Cell Research*, *28*(10), 1026–1034. https://doi.org/10.1038/s41422-018-0071-1

Zieliński, Ł. P., Smith, A. C., Smith, A. G., & Robinson, A. J. (2016). Metabolic flexibility of mitochondrial respiratory chain disorders predicted by computer modelling. *Mitochondrion*, *31*, 45–55. https://doi.org/10.1016/j.mito.2016.09.003

CHAPTER 3

EVIDENCE OF MITONUCLEAR MOLECULAR COEVOLUTION IN MAMMALIAN GENOMES

ABSTRACT

The primary cellular function of the mitochondria - aerobic respiration, is wholly dependent on intimate interactions between the products of both the mitochondrial (mtDNA) and nuclear (nucDNA) genomes. Upwards of a thousand nuclear-encoded proteins are required for mitochondrial DNA's replication and maintenance. This strong interdependence between mtDNA and nucDNA creates pressure for the coevolution of the two genomes in order to maintain organismal fitness. We investigate patterns of coevolution between mtDNA and nucDNA by comparing the evolutionary rates of the mitochondrial-encoded proteins to three nuclear-encoded protein sets with varying levels of interactions with mitochondria in 29 mammalian species. We find that evolutionary rate correlations between the mtDNA-encoded and nucDNA-encoded proteins are the strongest in nucDNA-encoded proteins that physically interact with mtDNA-encoded products, followed by the proteins that are encoded in the nucDNA and function in the mitochondria and finally proteins from the glycolysis pathway which do not interact with the mitochondria showed the weakest correlations. These evolutionary rate correlations are the strongest for branches with higher mtDNA substitution rates suggesting mtDNA evolutionary rate plays an influential role in shaping the coevolutionary dynamics between the mitochondrial and the nuclear genomes.

INTRODUCTION

Aerobic metabolism in eukaryotic cells is the result of extensive interactions between gene products encoded by two distinct genomes: the nuclear (nucDNA) and the mitochondrial (mtDNA) genomes. While the mitochondria serve as the primary sites for ATP synthesis, mitochondrial maintenance and function is intricately tied to nuclear-encoded products. Specifically, both nuclear and mitochondrial-encoded proteins physically interact to form the multi-subunit oxidative phosphorylation (OXPHOS) enzyme complexes. In addition to the OXPHOS proteins, mtDNA also codes for ribosomal RNA and tRNAs, which in concert with the nucDNA-encoded rRNAs and tRNA-synthetases perform mitochondrial protein synthesis in the mitochondrial matrix, independent of the cytoplasmic ribosomes. The nuclear-encoded DNA polymerase γ along with a number of accessory proteins and transcription factors is responsible for the replication and repair of mitochondrial DNA (Graziewicz et al., 2006). Further, more than 1000 nuclear encoded proteins are essential for proper mitochondrial function and are imported into the mitochondria (Calvo & Mootha, 2010). Despite the strong interdependence between the two genomes, they are replicated and inherited in fundamentally different ways, resulting in fascinating coevolutionary dynamics across the eukaryotic tree of life.

Although mitochondria descended from a single endosymbiotic merger between an Archaea-like prokaryote and an α-proteobacterium (reviewed in Martin et al., 2015), eukaryotes have quite divergent mitochondrial substitution rates, even in some closely related species. The substitution rates for bilaterian animal mitochondrial genomes are ≈ 100 times higher than plants and there is considerable variation in the rates within plants

and animals. However, substitution rates in the nuclear genome are similar between plants and animals. Mitochondrial genomes in bilaterian animals also predominantly lack recombination and are uniparentally inherited resulting in a four-fold smaller effective population size than the nuclear genomes. The smaller effective population size results in a lower efficacy of natural selection on mtDNA predisposing it to accumulate mildly deleterious mutations by genetic drift, which might further reach fixation by hitchhiking during positive selection sweeps in the non-recombining mitochondrial genomes (Meiklejohn et al., 2007).

Protein abundance is another factor that affects evolutionary rates; proteins that are highly abundant may be under more functional constraints and thereby exhibit slower rates of evolution as protein misfolding and misinteractions can be more costly for abundant proteins, resulting in stronger purifying selection (Zhang & Yang, 2015). A comparative analysis of expression levels and evolutionary rates for mtDNA and nucDNA OXPHOS proteins found that mutational rate differences between mtDNA and nucDNA explains 84% of the variance observed in mitochondrial vs nuclear substitution rates (Havird & Sloan, 2016). In addition, mitochondrial mutation rates in animals tend to be higher than nuclear mutation rates with vertebrates typically exhibiting higher μ_{mito}/μ_{nuc} ratios than invertebrates (Allio et al., 2017; Denver et al., 2000; Duda, 2021; Haag-Liautard et al., 2008; Xu et al., 2012). Although the elevated mutation rates in bilateral animal mitochondrial genomes were originally ascribed to the low fidelity of DNA polymerase γ (DeBalsi et al., 2017), recent experiments report that mtDNA mutation rates are higher due to the indirect impact of the highly oxidative mitochondrial

environment which reduces the fidelity of DNA polymerase γ by 10-100 fold (Anderson et al., 2020).

The elevated mutation rates and reduced efficacy of selection that result in faster substitution rates for mtDNA in contrast to nucDNA present a clear challenge to the maintenance of mitochondrial function, thereby necessitating mitochondrial-nuclear coevolution. As the rapid evolution of mitochondrial genome likely results in the accumulation of deleterious mutations, adaptive compensatory mutations are likely to arise in the nuclear genome, accelerated by higher recombination and relatively larger population to alleviate the deleterious mutations in mtDNA (Rand et al., 2004). However, compensatory mutations are also likely to arise in the mitochondrial genome and since it lacks recombination, an adaptive sweep that leads the mtDNA haplotype harboring the compensatory mutations to fixation is also likely to fix mildly deleterious mutations by genetic draft (Gillespie, 2000). This process is then expected to result in a positive feedback loop, strongly influenced by the mitochondrial mutation rate, a model termed as the "Compensation-Draft Feedback" model (Oliveira et al., 2008).

Consistent with this model is the hypothesis that mitochondria-associated nuclear-encoded (mt-nucDNA) products are under selective pressure to "keep up" with fast mitochondrial evolution and hence would exhibit similar fast rates of evolution. Indeed, evidence for this hypothesis in the mitochondrial translation machinery comes from the observation that mitochondrial ribosomal proteins, and aminoacyl tRNA synthetases show elevated evolutionary rates in contrast to their cytosolic counterparts (Adrion et al., 2016; Barreto & Burton, 2013). Similar patterns in the OXPHOS system have also been reported in angiosperms (Williams et al., 2019), insects (Yan et al., 2019) and bivalves

(Piccinini et al., 2021). In plants, mitochondria-associated nuclear-encoded (mt-nucDNA) proteins exhibit higher substitution rates in lineages with faster mtDNA evolution, while no such correlation was observed for the other nucDNA proteins (Havird et al., 2015; Sloan et al., 2014). Since among vertebrates, mammals and in particular primates exhibit some of the largest differences in μ_{mito}/μ_{nuc} (Allio et al., 2017; Osada & Akashi, 2012), we focused on testing these hypotheses in mammals.

Here, we estimate evolutionary rate correlations in mammalian species between mitochondrial-encoded proteins (mtDNA-OXPHOS) and three categories of nuclear-encoded proteins: nuclear-encoded OXPHOS proteins (nucDNA-OXPHOS), mitochondria-associated nuclear-encoded proteins (mt-nucDNA) and nuclear-encoded glycolysis proteins that serve as our non-mitochondria interacting, essential protein set. We show that evolutionary rate correlations are the strongest between mtDNA-OXPHOS and nucDNA-OXPHOS proteins, followed by those between mtDNA-OXPHOS and mt-nucDNA and finally the correlations between mtDNA-OXPHOS and glycolysis which were the weakest.

MATERIALS AND METHODS

Gene sets, compilation, and download

To test our hypothesis of correlated evolutionary rates between the mitochondrial and the nuclear genomes, we defined four gene categories: the mitochondrial-encoded and the nuclear-encoded genes from the Oxidative Phosphorylation pathway (mtDNA-OXPHOS and nucDNA-OXPHOS gene sets

respectively), genes encoded by the nuclear genome but associated with mitochondrial function (mt-nucDNA gene set) and genes from the glycolysis pathway, which function in the cytoplasm and are encoded solely by the nuclear genome (glycolysis gene set).

For the OXPHOS gene sets, we used HUGO Gene Nomenclature Committee to search for genes in the gene group "Mitochondrial respiratory chain complexes" (HGNC Group ID: 639) and retrieved 13 mtDNA-OXPHOS and 84 nucDNA-OXPHOS genes. For the mitochondrial-associated nuclear-encoded (mt-nucDNA) gene set, we used MitoCarta 3.0 (Rath et al., 2021) to download gene sets for human genes with high confidence of mitochondrial localization based on an integrated proteomics, computation, and microscopy approach. We retrieved 1025 genes in the mt-nucDNA from the MitoCarta dataset.

For the glycolysis gene set, we used the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto, 2000) annotation database to identify genes participating in the glycolysis pathway (KEGG pathway ID: ko00010). We used humans as our focal species identifying the genes across the gene sets and retrieved 62 human glycolysis genes using the KEGG pathway ID: hsa00010.

Orthologs for nuclear genes were downloaded using the NCBI Datasets command-line tool (https://www.ncbi.nlm.nih.gov/datasets/). Mitochondrial genomes were downloaded from GenBank's Organelle Genome Resources (Wolfsberg et al., 2001). Amino acid sequences were extracted from NCBI data packages using custom Python scripts**.**

We subsetted the above gene sets to include ≈ 50 genes in each gene set in order to maximize species × genes coverage based on sequence data availability and further

restricted it to one species per genus. This subsetting resulted in a final gene set with amino acid sequence data for 26 species (Table 2) and with 13 mtDNA-OXPHOS genes, 45 nucDNA-OXPHOS genes, 50 mt-nucDNA genes and 49 glycolysis genes (Table 3).

Sequence alignment and curation

Amino acid sequences for each gene were individually aligned using MAFFT v7.407 set to automatically select the appropriate alignment strategy using the '--auto' option (Katoh & Standley, 2013). Genewise amino acid alignments were curated to retain high-confidence homology clusters using divvier with the 'full divvying' option and the 'mincol' parameter set to 4 (Ali et al., 2019). Sequence alignments were then concatenated for each gene set using AMAS v1.0 (Borowiec, 2016) implemented via a custom Python script. Codon-aware alignments were generated for each gene by aligning nucleotide sequences to protein sequences using PAL2NAL v14.1 (Suyama et al., 2006) with columns with gaps and inframe stop codons removed with the '-nogap' option.

Phylogenetic tree download and manipulation

A set of 1000 sample trees were downloaded from the VertLife phylogeny subsets (http://vertlife.org/) using the "Mammals birth-death tip-dated DNA-only trees (4098 species, set of 10k trees)" selection for the 26 species in our dataset. The VertLife trees were built using Bayesian inference on a 31-gene supermatrix that included 22 nuclear-encoded exons, 5 non-coding regions and 4 mtDNA genes (Upham et al., 2019).

Topological differences between the sample of 1000 trees were quantified and summarized using the Triples metric, Robinson-Fould Clustering distance and the

Matching Clusters metric using TreeCmp (Bogdanowicz et al., 2012). A 95% Majority Rule Consensus (MRC) tree was constructed from the sample of 1000 trees using SumTrees from the DendroPy v4.5.2 Python package (Sukumaran & Holder, 2010). Multichotomies in the 95% MRC were resolved using mutli2di and the resulting tree was made ultrametric using the ape v5.6.2 package (Paradis & Schliep, 2019) in R v4.1.2 (R Core Team, 2021).

Selecting best-fit models of evolution and estimating evolutionary rates

The best-fit models of amino acid substitutions were inferred for each concatenated gene set by specifying a partitioned model with each gene making up a partition using ModelTest-NG v0.1 (Darriba et al., 2020). The amino acid substitution models for each partition were then chosen using the corrected Akaike Information Criterion (AICc). For each gene set, the number of amino acid substitutions per site were estimated by re-optimizing branch length on the 95% MRC tree using the concatenated protein alignments using RAxML-NG v1.1 (Kozlov et al., 2019). Branch lengths were then extracted using Newick Utilities (Junier & Zdobnov, 2010).

For each gene set, we estimated the substitution rates at synonymous and nonsynonymous sites with concatenated backtranslated alignments using Codeml (PAML v4.9 (Yang, 2007) by specifying a branch model that estimates a different dN/dS value for each branch in the phylogeny (branch model = 1, NSsites = 0).

Computing Evolutionary Rate Correlations (ERC)

Evolutionary rate correlations (ERC) were estimated using the 95% majority rule consensus phylogenetic tree (described above). Using the same tree topology enables us to use branch lengths as measures of relative evolutionary rates and permits comparisons of branch lengths across the gene sets. Pairwise evolutionary rate correlations between the gene sets were then computed for each node-to-node distance across the phylogeny to account for phylogenetic dependence in the evolutionary rates. A non-parametric measure of correlation, Spearman's ρ was computed in R v4.1.2 and correlation plots were made using the ggpubr (Kassambara, 2020).

RESULTS

To investigate the coevolution between mitochondrial-encoded and nuclear-encoded products, we estimated evolutionary rate correlations (ERC) between three nuclear-encoded protein datasets and one mitochondrial-encoded protein dataset in mammals. The 13 mitochondrial-encoded subunits of the oxidative phosphorylation (OXPHOS) pathway made up the mitochondrial protein set (mtDNA-OXPHOS). We defined the nuclear-encoded protein datasets at three different levels of functional interactions with the mitochondrial-encoded proteins to test for coevolutionary hypotheses. The nucDNA-OXPHOS dataset comprised of 45 nuclear-encoded OXPHOS proteins and represented the set of nuclear proteins that physically interact with the mitochondrial proteins and together constitute the respiratory chain complexes. We defined the set of mitochondrial-associated nuclear-encoded proteins (mt-nucDNA) that co-localize in the mitochondria based on annotations from the MitoCarta database.

Proteins involved in cellular glycolysis serve essential, housekeeping roles in the cell and were chosen as our nuclear-encoded control set for the analysis.

We employed a nonparametric Spearman correlation approach to estimate correlations in evolutionary rates between mitochondrial and nuclear proteins and tested the hypothesis that ERC would be strongest between proteins that form close, physical contact (mtDNA OXPHOS - nucDNA OXPHOS), followed by between proteins co-localizing in the mitochondria (mtDNA OXPHOS - mt-nucDNA) and lastly between proteins sharing essential cellular functions (mtDNA OXPHOS - glycolysis). For this, we compiled amino acid sequence data for 26 mammalian species and used phylogenetic tree topologies from VertLife. We concatenated protein sequence alignments for each protein dataset and re-estimated evolutionary rates as branch lengths of the 95% majority rule consensus (MRC) tree compiled from a sample of 1000 trees.

Phylogenetic relationships for the mammalian species

We downloaded a sample of 1000 birth-death tip-dated trees for the 26 species in our dataset from VertLife. Molecular data from both the mtDNA and nucDNA was used in the phylogenetic inferences by the VertLife group (Figure 1; see Methods). We recovered only two tree topologies from our sample of 1000 trees, of which one was represented only 17 times. The two topologies differed in the placement of the relative positions of the Carnivora, Artiodactyla and Chiroptera orders with respect to *Condylura cristata* (Figure 2). The quantitative differences between the trees are given in Table 1.

The summarization of the phylogenetic information into a majority rule consensus tree introduced a polytomy at the node where the two topologies diverged. We resolved the polytomy and made the tree ultrametric before optimizing the branch lengths.

Evolutionary rates for the protein datasets

mtDNA-OXPHOS proteins had higher amino acid substitutions per site (μ = 0.312, s.d. = 0.214) than the nuclear-encoded proteins (Figure 3). The evolutionary rates for the nucDNA-OXPHOS proteins (μ = 0.094, s.d. = 0.048) were not significantly different than those for mt-nucDNA proteins (μ = 0.064, s.d. = 0.086). The glycolysis proteins had the slowest evolutionary rates (μ = 0.058, s.d. = 0.024) which were significantly lower than the mtDNA-OXPHOS and the other nuclear-encoded proteins. For all the proteins, Rodentia and Chiroptera have consistently higher rates of evolution (Figure 5).

The ratio of the nonsynonymous to synonymous substitution rates (dN/dS ratios) are useful for representing the rates of protein evolution that is normalized by the background substitution rate. mtDNA-OXPHOS genes had the lowest dN/dS ratios (μ = 0.058, s.d. = 0.030) while nucDNA-OXPHOS genes had the highest dN/dS ratios (μ = 0.271, s.d. = 0.143) among the four gene sets (Figure 4A). For the other two nuclear gene sets, the dN/dS ratios were: mt-nucDNA (μ = 0.222, s.d. = 0.132) and glycolysis (μ = 0.137, s.d. = 0.078). The rates of nonsynonymous variation followed patterns similar to that of synonymous variation with mtDNA-OXPHOS genes having higher rates than nuclear-encoded genes (μ = 0.027, s.d. = 0.021) (Figure 4B). The rate of nonsynonymous

substitutions per nonsynonymous site for the nuclear genes were: nucDNA-OXPHOS (μ = 0.009, s.d. = 0.009), mt-nucDNA (μ = 0.008, s.d. = 0.009), and glycolysis (μ = 0.006, s.d. = 0.006). mtDNA-OXPHOS genes had the highest rates of synonymous substitutions per synonymous site (μ = 0.534, s.d. = 0.571) than the nuclear-encoded genes (Figure 4C). The synonymous rates of substitutions for the nuclear-encoded genes were: nucDNA-OXPHOS (μ = 0.037, s.d. = 0.04), mt-nucDNA (μ = 0.038, s.d. = 0.041), and glycolysis (μ = 0.042, s.d. = 0.045).

Evolutionary rate correlations between the protein datasets

The pattern of ERC between the protein datasets was consistent with our hypothesis. The correlations between the mtDNA-OXPHOS and nucDNA-OXPHOS proteins was the strongest (ρ = 0.96, $p < 0.001$), followed by the correlations between mtDNA-OXPHOS and mt-nucDNA proteins (ρ = 0.9, $p < 0.001$) and finally between mtDNA-OXPHOS and glycolysis proteins (ρ = 0.86, $p < 0.001$). While a linear curve fit the mtDNA-OXPHOS:nucDNA-OXPHOS regression well, the mtDNA-OXPHOS:mt-nucDNA and mtDNA-OXPHOS:glycolysis correlations were better represented by quadratic curves. The scatterplots between the branch lengths of the protein sets suggest that branches with slower mtDNA evolution exhibit a distinct trend than the branches with faster mtDNA evolution (Figure 6).

To further study this, we estimated Spearman's correlation coefficients for two subsets of the dataset based on the rate of mtDNA evolution: "fast branches" where the rates for mtDNA-OXPHOS proteins was > 0.2 and "slow branches" with mtDNA-

OXPHOS protein rates < 0.2. For the "slow branches" subset, all three pairwise correlations were similar (mtDNA-OXPHOS:nucDNA-OXPHOS ρ = 0.87, mtDNA-OXPHOS:mt-nucDNA ρ = 0.84, mtDNA-OXPHOS:glycolysis ρ = 0.88). For the "fast branches" subset, mtDNA-OXPHOS:nucDNA-OXPHOS correlation was still strong (ρ = 0.81, $p < 0.001$), while mtDNA-OXPHOS:mt-nucDNA (ρ = 0.5, $p = 0.009$) and mtDNA-OXPHOS:glycolysis (ρ = 0.2, $p = 0.32$) were not significant. These results suggest that mtDNA mutational rate might be driving the correlation patterns observed in this dataset.

Since any individual branch length from our amino acid sequence analysis would be a product of substitution rate × divergence time, a short branch in our phylogeny could have a high rate but a very small amount of elapsed time, or a very low rate but a longer amount of time. The rate of synonymous substitutions at synonymous sites is representative of the background rate of accumulation of mutations in the DNA as per the neutral theory of evolution. Hence, the ratio of nonsynonymous to synonymous substitution rates (dN/dS ratio) is expected to capture the rate of protein evolution that is normalized by divergence time.

We then estimated Spearman's correlation coefficients for synonymous and nonsynonymous substitutions and the ratio of nonsynonymous to synonymous rates across the gene sets. The correlations for nonsynonymous-to-synonymous substitution rates between the mtDNA-OXPHOS genes and the nuclear-encoded genes recapitulated trends similar to those observed for amino acid substitutions per site between the gene sets (Figure 7). The correlations were strongest between mtDNA-OXPHOS:nucDNA-OXPHOS (ρ = 0.7, $p < 0.001$), followed by the correlations between mtDNA-

OXPHOS:mt-nucDNA ($\rho = 0.51$, $p < 0.001$). The correlations between mtDNA-OXPHOS:glycolysis ($\rho = 0.3$, $p = 0.036$) were the weakest. The nonsynonymous substitution rates were correlated for all the pairwise comparisons (Figure 8): mtDNA-OXPHOS:nucDNA-OXPHOS ($\rho = 0.71$, $p < 0.001$), mtDNA-OXPHOS:mt-nucDNA ($\rho = 0.58$, $p < 0.001$) and mtDNA-OXPHOS:glycolysis ($\rho = 0.71$, $p < 0.001$). As expected, the rates of synonymous variation are correlated across the gene sets (Figure 9). The pairwise correlations for synonymous substitution rates between the mtDNA-OXPHOS genes and the nuclear genes were: mtDNA-OXPHOS:nucDNA-OXPHOS ($\rho = 0.72$, $p < 0.001$), mtDNA-OXPHOS:mt-nucDNA ($\rho = 0.64$, $p < 0.001$) and mtDNA-OXPHOS:glycolysis ($\rho = 0.66$, $p < 0.001$).

DISCUSSION

We estimated evolutionary rate correlations between mtDNA-encoded proteins and three categories of nucDNA-encoded proteins to test hypotheses of concerted evolution between the mitochondrial and mitochondria-associated gene products in mammalian species. Correlating evolutionary rates across branches requires all proteins to share identical phylogenetic relationships. A well-known phenomenon in mito-nuclear genetics is the phylogenetic discordance between the mitochondrial and nuclear genetic elements i.e. mitochondrial genes recover a phylogenetic relationship that is different than the relationship recovered using nuclear genes (Toews & Brelsford, 2012). One possible explanation for discordance among gene trees is incomplete lineage sorting, in which ancestral polymorphisms in young species splits are initially maintained in

multiple lineages (Degnan & Rosenberg, 2009). While this is a likely scenario for cases in which there are differences between gene tree topologies and the overall species tree, incomplete lineage sorting is unlikely to cause mitochondrial-nuclear discordance. As a consequence of the smaller effective population size in mitochondrial genomes, mtDNA polymorphisms are expected to sort out rapidly and hence are less likely to differ from the nucDNA tree topologies (Good et al., 2015).

A more likely cause of mitochondrial-nuclear discordance appears to be introgression, which through a myriad of genetic and demographic routes can differentially affect nuclear vs mitochondrial loci (Sloan et al., 2016). Indeed, numerous biogeographic studies report high levels of mitochondrial introgression in the absence of any detectable levels of nuclear gene flow (Bernatchez et al., 1995; Good et al., 2015; Milá et al., 2010; Zieliński et al., 2013). Our taxonomic dataset comprises of a single mammalian species per genus thereby sampling deeper phylogenetic relationships and reducing the potential of both incomplete lineage sorting and asymmetric introgression leading to mitonuclear discordance in our dataset (Table 1). The phylogenetic topology used to estimate the mtDNA and nucDNA evolutionary rates for our species set is a summarization of 1000 topologies subsetted for our 26 species from the posterior distribution of a tree building process that estimates the phylogenetic relationships between ≈ 4100 mammals through bayesian inference. The 31-gene supermatrix used in the construction of the phylogenetic tree consisted of both mtDNA, nucDNA and non-coding genetic elements (Upham et al., 2019). The majority rule consensus (MRC) tree used in our analysis was supported by 98.3% of the 1000 topologies providing robust support for the topology represented by the MRC tree.

Consistent with our hypothesis, the evolutionary rate correlations between mtDNA-OXPHOS and nucDNA-OXPHOS proteins were represented by an almost perfect positive correlation ($\rho = 0.96$). The evolutionary rate correlations between mtDNA-OXPHOS and mt-nucDNA proteins ($\rho = 0.90$) as well as mtDNA-OXPHOS and glycolyis proteins ($\rho = 0.86$) were better represented by quadratic relationships and had weaker correlations than the mtDNA-OXPHOS:nucDNA-OXPHOS comparison. The evolutionary rate correlations we estimate from our mammalian dataset are almost identical to similar correlations in bivalves ($\rho = 0.967$) (Piccinini et al., 2021) and are only slightly different than the correlations in insects ($\rho = 0.77$) (Yan et al., 2019). However, while our mtDNA and control set comparison exhibits strong correlations (mtDNA-OXPHOS:Glycolysis $\rho= 0.86$), similar comparisons in bivalves ($\rho = 0.43$) and insects ($\rho = 0.47$) recovered weaker relationships, this difference is likely due to the choice of the control nuclear set used in the comparisons. Yan et al., 2019 chose nuclear cell-cycle proteins as their control set, while Piccinini et al., 2021 chose random nuclear orthologs with no mitochondrial interactions.

Subsetting the branches into two categories based on mtDNA substitution rates uncovered distinct patterns for the correlations between branches with faster mtDNA evolution and those with slower mtDNA evolution. The correlations for all mtDNA-OXPHOS:protein set comparisons were similar for branches with slower mtDNA rates, which was consistent with the patterns we observed for the synonymous site variation in our gene sets. However, for the subset of branches with fast mtDNA evolution, the difference in correlations was stark with only the mtDNA-OXPHOS:nucDNA-OXPHOS showing a significant positive relationship in their evolutionary rates. Consistent with the

Compensatory-Draft Feedback model, these patterns suggest that mtDNA evolutionary rates strongly influences and drives the faster evolution of nuclear-encoded proteins that interact with the mitochondria.

FIGURES AND TABLES

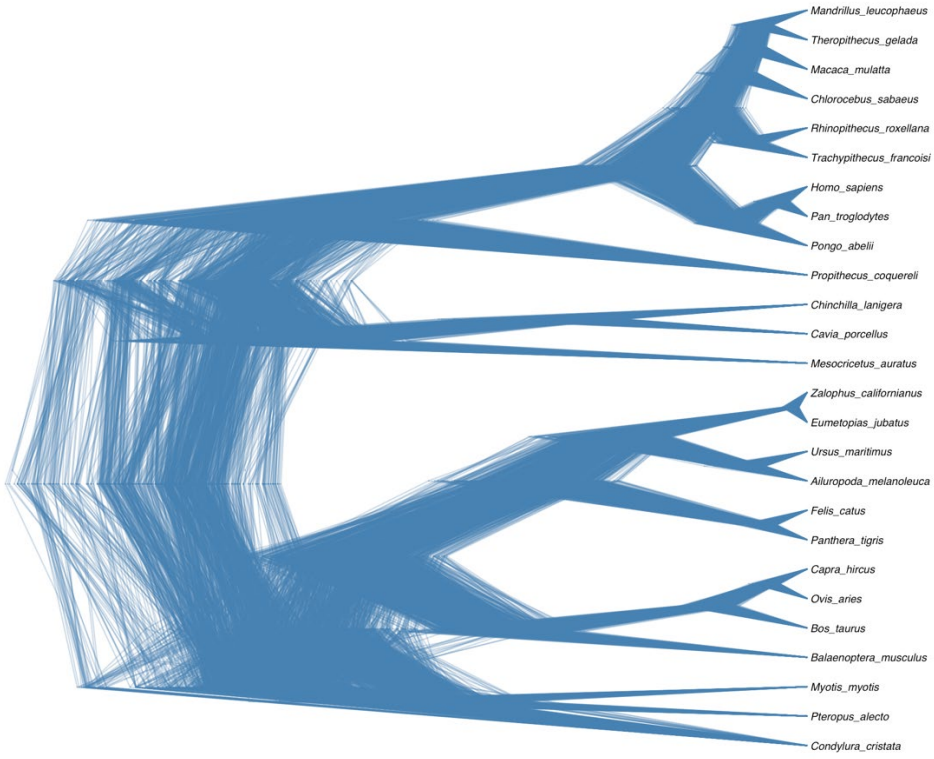

Figure 1. DensiTree plot of the 1000 mammalian phylogenetic trees downloaded from VertLife.

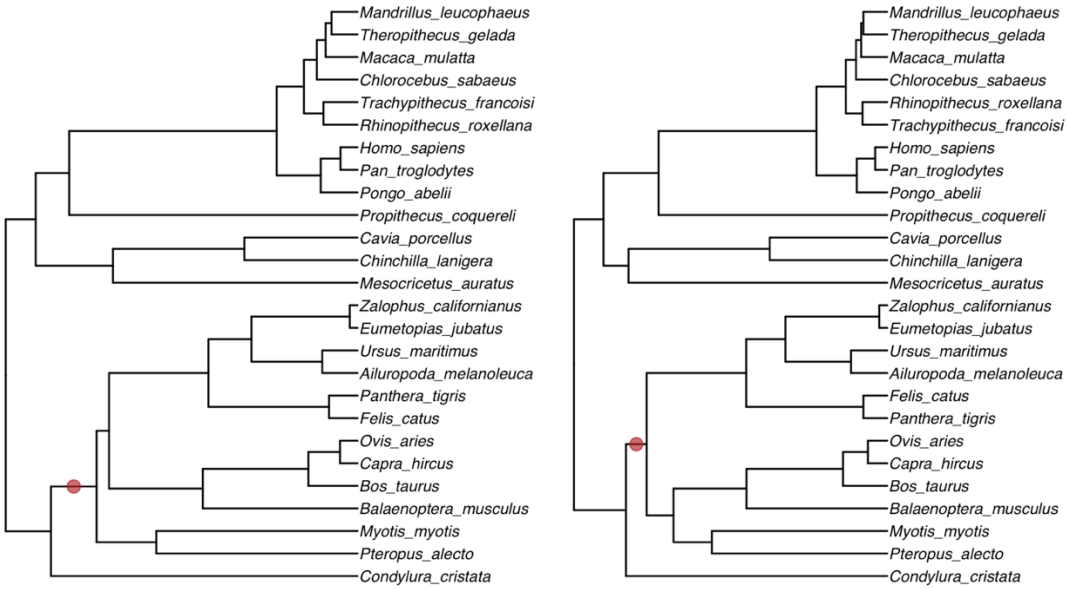

Figure 2. The two distinct tree topologies recovered from the VertLife 1000 tree sample. The trees differ in the placement of the node connecting *Condylura cristata* to the basal branch of Artiodactyla, Chiroptera and Carnivora (represented by the red dot).

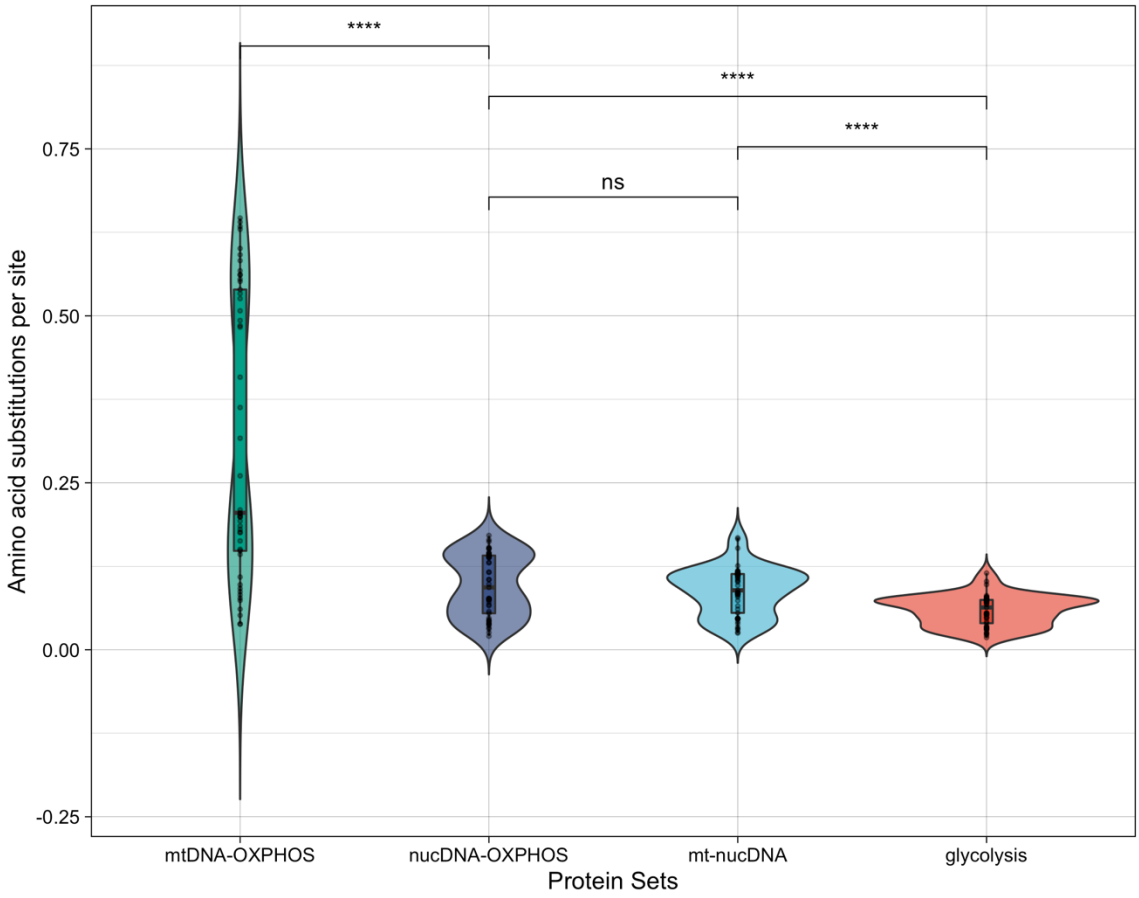

Figure 3. Branch lengths representing amino acid substitutions per site for the four protein sets. mtDNA-OXPHOS has the highest evolutionary rate, followed by nucDNA-OXPHOS, mt-nucDNA and glycolysis proteins. The difference in means of all pairwise comparisons except for the nucDNA-OXPHOS:mt-nucDNA comparison.

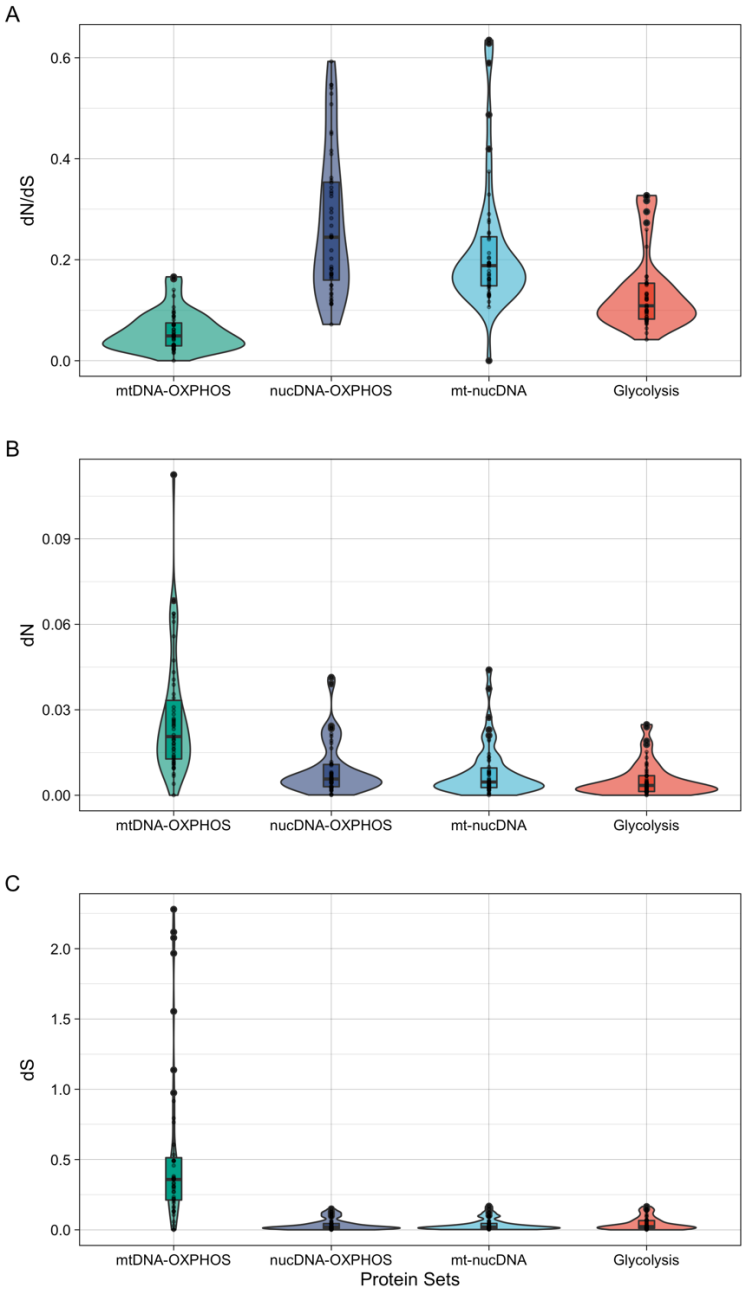

Figure 4. The rates of synonymous and nonsynonymous variation for the four protein sets. mtDNA-OXPHOS genes have the highest rate of synonymous substitution as well as the lowest values for dN/dS ratios.

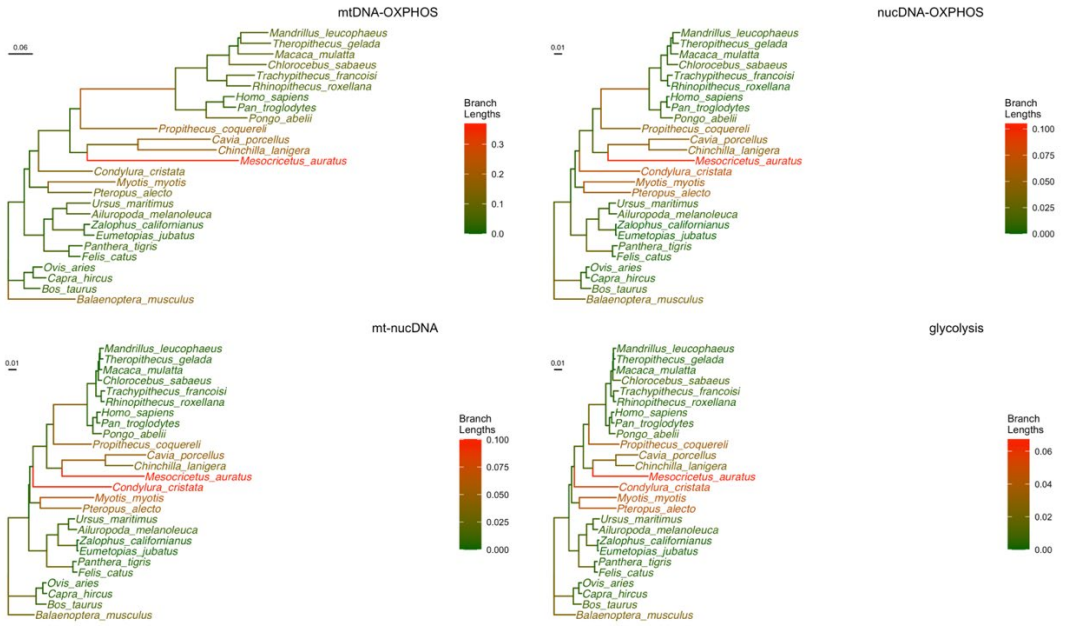

Figure 5. Number of amino acid substitutions per site mapped onto the 95% MRC tree for the four protein sets. Rodentia and Chiroptera appear to have overall higher rates of evolution across all the protein sets.

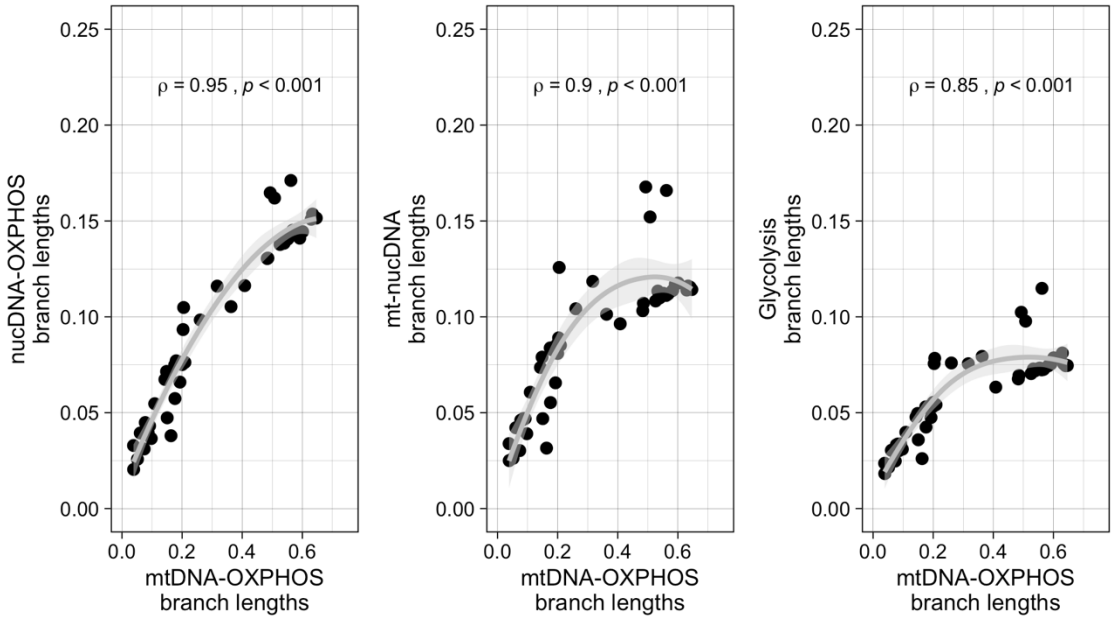

Figure 6. Evolutionary rate correlations using all branches between the mtDNA-OXPHOS proteins and nuclear-encoded proteins. A local regression curve was fitted to the data with "loess" smoothing.

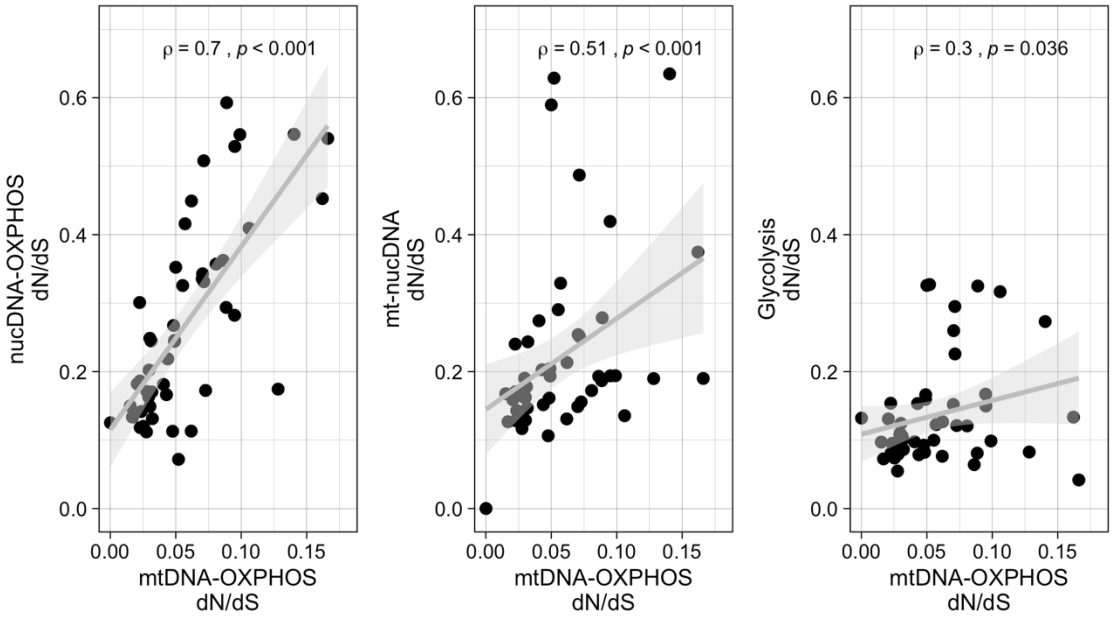

Figure 7. Evolutionary rate correlations for dN/dS ratios using all branches. The dN/dS ratios represent the rate of protein evolution normalized by the background rate of substitutions. The correlations between mtDNA-OXPHOS:nucDNA-OXPHOS are the strongest followed by mtDNA-OXPHOS:mt-nucDNA and then mtDNA-OXPHOS-Glycolysis. A local regression curve was fitted to the data.

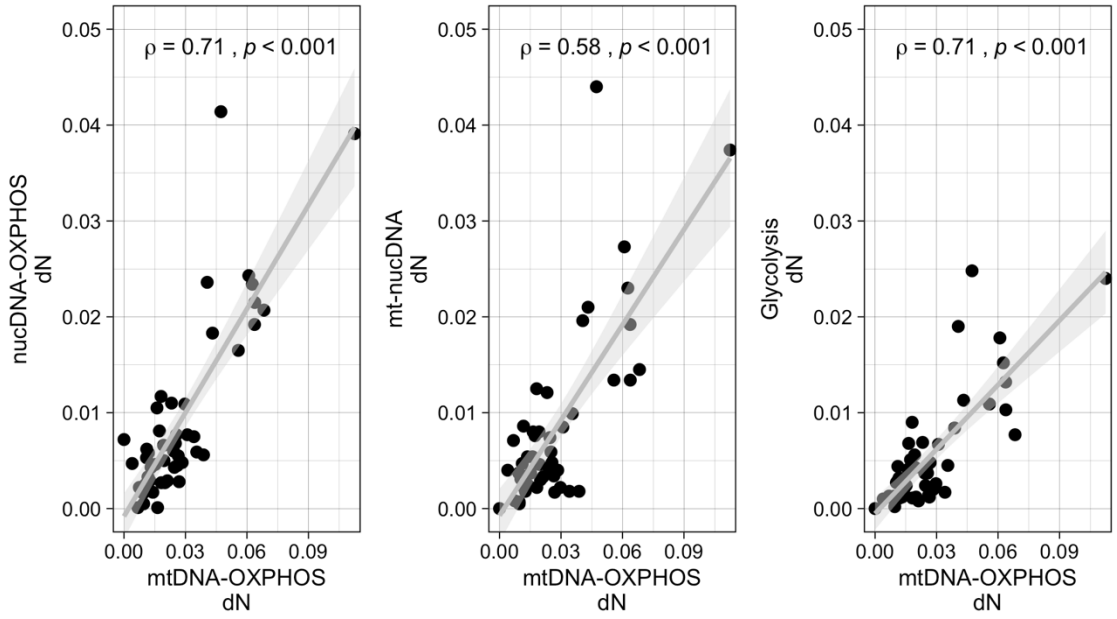

Figure 8. Evolutionary rate correlations for dN ratios using all branches. The dN rates represent the rate of accumulation of nonsynonymous substitutions in the protein sets. The correlations are similar across all the pairwise comparisons. A local regression curve was fitted to the data.

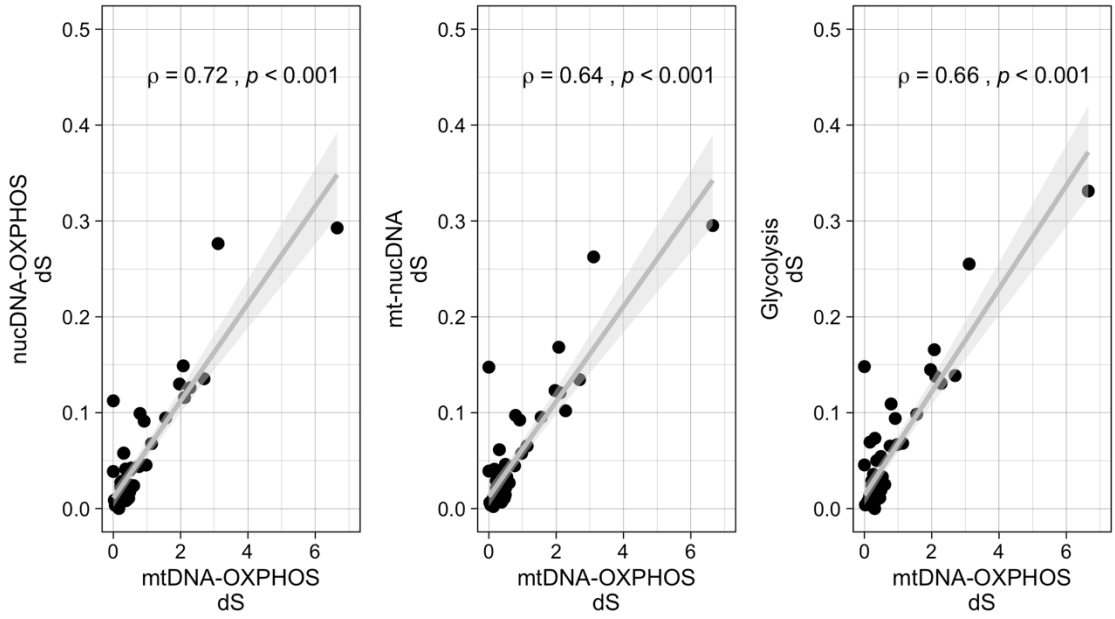

Figure 9. Evolutionary rate correlations for dS ratios using all branches. The dS rates represent the rate of accumulation of synonymous substitutions in the protein sets. The correlations are similar across all the pairwise comparisons. A local regression curve was fitted to the data.

Table 1. List of the 26 mammalian species used for the estimation of evolutionary rate correlations

Scientific name	**Order**
Balaenoptera musculus	Artiodactyla
Bos taurus	Artiodactyla
Capra hircus	Artiodactyla
Ovis aries	Artiodactyla
Ailuropoda melanoleuca	Carnivora
Eumetopias jubatus	Carnivora
Felis catus	Carnivora
Panthera tigris	Carnivora
Ursus maritimus	Carnivora
Zalophus californianus	Carnivora
Myotis myotis	Chiroptera
Pteropus alecto	Chiroptera
Condylura cristata	Eulipotyphla
Chlorocebus sabaeus	Primates
Homo sapiens	Primates
Macaca mulatta	Primates
Mandrillus leucophaeus	Primates
Pan troglodytes	Primates
Pongo abelii	Primates
Propithecus coquereli	Primates
Rhinopithecus roxellana	Primates
Theropithecus gelada	Primates
Trachypithecus francoisi	Primates
Cavia porcellus	Rodentia
Chinchilla lanigera	Rodentia
Mesocricetus auratus	Rodentia

Table 2. List of proteins in different categories for the estimation of evolutionary rate correlations

mtDNA OXPHOS	nucDNA OXPHOS	mt-nucDNA	Glycolysis
ATP6	ATP5F1A	ABCB7	ACSS1
ATP8	ATP5F1B	ACOD1	ACSS2
COX1	ATP5F1D	ADCK2	AKR1A1
COX2	ATP5MC1	AIFM2	ALDH1A3
COX3	ATP5MC3	ALAS2	ALDH1B1
CYTB	ATP5MG	ALDH5A1	ALDH2
ND1	ATP5PF	BCKDHA	ALDH3A1
ND2	ATP5PO	BCL2	ALDH3A2
ND3	COX4I1	CLPP	ALDH3B1
ND4	COX4I2	COQ8A	ALDH7A1
ND4L	COX5A	CPT1B	ALDH9A1
ND5	COX5B	DMAC2L	ALDOA
ND6	COX7A1	DNAJC15	ALDOB
	COX7A2	ECHDC3	ALDOC
	NDUFA10	FASTKD5	BPGM
	NDUFA11	FLAD1	DLAT
	NDUFA12	GFM1	DLD
	NDUFA13	GLYCTK	ENO1
	NDUFA5	GOT2	ENO2
	NDUFA6	LYRM4	ENO3
	NDUFA8	MARS2	FBP1
	NDUFA9	MCU	FBP2
	NDUFB11	METTL15	G6PC
	NDUFB2	MIGA2	G6PC2
	NDUFB4	MPC2	GALM
	NDUFB5	MRPL20	GAPDH
	NDUFB6	MRPL22	GCK
	NDUFB8	MRPS22	GPI
	NDUFB9	NFS1	HK1
	NDUFC1	NGRN	HK2
	NDUFS1	NIT1	HK3
	NDUFS3	NTHL1	LDHB
	NDUFS4	OGDHL	PCK1
	NDUFS7	PDK1	PCK2
	NDUFS8	PHB	PDHA1

NDUFV1	PNPLA8	PDHA2
NDUFV2	POLG	PDHB
NDUFV3	PPA2	PFKL
SDHB	PRXL2A	PFKM
SDHC	PTRH2	PFKP
SDHD	PUS1	PGAM1
UQCR11	RCC1L	PGAM2
UQCRC1	RSAD1	PGK1
UQCRC2	SAMM50	PGK2
UQCRH	SLC25A14	PGM1
	SLC25A48	PGM2
	TOMM34	PKLR
	TRIT1	PKM
	UQCC2	TPI1
	YRDC	

Table 3. Quantitative measures of tree discordance for the 1000 tree samples downloaded from VertLife computed using TreeCmp.

Metric	Average	S.D.	Minimum	Maximum	Count
Triples	1.60	8.63	0.0	48.0	499500
RFCluster (0.5)	0.03	0.17	0.0	1.0	499500
Matching Pair	0.40	2.15	0.0	12.0	499500
Matching Cluster	0.26	1.43	0.0	8.0	499500

REFERENCES

Adrion, J. R., White, P. S., & Montooth, K. L. (2016). The roles of compensatory evolution and constraint in aminoacyl trna synthetase evolution. *Molecular Biology and Evolution*, *33*(1), 152–161. https://doi.org/10.1093/molbev/msv206

Ali, R. H., Bogusz, M., & Whelan, S. (2019). Identifying Clusters of High Confidence Homologies in Multiple Sequence Alignments. *Molecular Biology and Evolution*, *36*(10), 2340–2351. https://doi.org/10.1093/molbev/msz142

Allio, R., Donega, S., Galtier, N., & Nabholz, B. (2017). Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: Implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Molecular Biology and Evolution*, *34*(March), 2762–2772. https://doi.org/10.1093/molbev/msx197

Anderson, A. P., Luo, X., Russell, W., & Yin, Y. W. (2020). Oxidative damage diminishes mitochondrial DNA polymerase replication fidelity. *Nucleic Acids Research*, *48*(2), 817–829. https://doi.org/10.1093/nar/gkz1018

Barreto, F. S., & Burton, R. S. (2013). Evidence for compensatory evolution of ribosomal proteins in response to rapid divergence of mitochondrial rRNA. *Molecular Biology and Evolution*, *30*(2), 310–314. https://doi.org/10.1093/molbev/mss228

Bernatchez, L., Glémet, H., Wilson, C. C., & Danzmann, R. G. (1995). Introgression and fixation of Arctic char (Salvelinus alpinus) mitochondrial genome in an allopatric population of brook trout (Salvelinus fontinalis). *Canadian Journal of Fisheries and Aquatic Sciences*, *52*(1), 179–185. https://doi.org/10.1139/f95-018

Bogdanowicz, D., Giaro, K., & Wróbel, B. (2012). TreeCmp: Comparison of trees in polynomial time. *Evolutionary Bioinformatics*, *2012*(8), 475–487. https://doi.org/10.4137/EBO.S9657

Borowiec, M. L. (2016). AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ*, *4*. https://doi.org/10.7717/peerj.1660

Calvo, S. E., & Mootha, V. K. (2010). The Mitochondrial Proteome and Human Disease. *Annual Review of Genomics and Human Genetics*, *11*(1), 25–44. https://doi.org/10.1146/annurev-genom-082509-141720

Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., & Flouri, T. (2020). ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Molecular Biology and Evolution*, *37*(1), 291–294. https://doi.org/10.1093/molbev/msz189

DeBalsi, K. L., Hoff, K. E., & Copeland, W. C. (2017). Role of the mitochondrial DNA replication machinery in mitochondrial DNA mutagenesis, aging and age-related diseases. *Ageing Research Reviews*, *33*, 89–104. https://doi.org/10.1016/j.arr.2016.04.006

Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, *24*(6), 332–340. https://doi.org/10.1016/j.tree.2009.01.009

Denver, D. R., Morris, K., Lynch, M., Vassilieva, L. L., & Thomas, W. K. (2000). High Direct Estimate of the Mutation Rate in the Mitochondrial Genome of *Caenorhabditis elegans*. *Science*, *289*(5488), 2342–2344. https://doi.org/10.1126/science.289.5488.2342

Duda, T. F. (2021). Patterns of variation of mutation rates of mitochondrial and nuclear genes of gastropods. *BMC Ecology and Evolution*, *21*(1), 13. https://doi.org/10.1186/s12862-021-01748-2

Gillespie, J. H. (2000). Genetic Drift in an Infinite Population: The Pseudohitchhiking Model. *Genetics*, *155*(2), 909–919. https://doi.org/10.1093/genetics/155.2.909

Good, J. M., Vanderpool, D., Keeble, S., & Bi, K. (2015). Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. *Evolution*, *69*(8), 1961–1972. https://doi.org/10.1111/evo.12712

Graziewicz, M. A., Longley, M. J., & Copeland, W. C. (2006). DNA Polymerase γ in Mitochondrial DNA Replication and Repair. *Chemical Reviews*, *106*(2), 383–405. https://doi.org/10.1021/cr040463d

Haag-Liautard, C., Coffey, N., Houle, D., Lynch, M., Charlesworth, B., & Keightley, P. D. (2008). Direct Estimation of the Mitochondrial DNA Mutation Rate in *Drosophila melanogaster*. *PLoS Biology*, *6*(8), e204. https://doi.org/10.1371/journal.pbio.0060204

Havird, J. C., & Sloan, D. B. (2016). The Roles of Mutation, Selection, and Expression in Determining Relative Rates of Evolution in Mitochondrial versus Nuclear Genomes. *Molecular Biology and Evolution*, *33*(12), 3042–3053. https://doi.org/10.1093/molbev/msw185

Havird, J. C., Whitehill, N. S., Snow, C. D., & Sloan, D. B. (2015). Conservative and compensatory evolution in oxidative phosphorylation complexes of angiosperms with highly divergent rates of mitochondrial genome evolution. *Evolution*, *69*(12), 3069–3081. https://doi.org/10.1111/evo.12808

Junier, T., & Zdobnov, E. M. (2010). The Newick utilities: High-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics (Oxford, England)*, *26*(13), 1669–1670. https://doi.org/10.1093/bioinformatics/btq243

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Kassambara, A. (2020). *ggpubr: "ggplot2" based publication ready plots* [Manual]. https://CRAN.R-project.org/package=ggpubr

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, *35*(21), 4453–4455. https://doi.org/10.1093/bioinformatics/btz305

Martin, W. F., Garg, S., & Zimorski, V. (2015). Endosymbiotic theories for eukaryote origin. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1678), 20140330. https://doi.org/10.1098/rstb.2014.0330

Meiklejohn, C. D., Montooth, K. L., & Rand, D. M. (2007). Positive and negative selection on the mitochondrial genome. *Trends in Genetics : TIG*, *23*(6), 259–263. https://doi.org/10.1016/j.tig.2007.03.008

Milá, B., Carranza, S., Guillaume, O., & Clobert, J. (2010). Marked genetic structuring and extreme dispersal limitation in the Pyrenean brook newt Calotriton asper (Amphibia: Salamandridae) revealed by genome-wide AFLP but not mtDNA.

Molecular Ecology, *19*(1), 108–120. https://doi.org/10.1111/j.1365-294X.2009.04441.x

Oliveira, D. C. S. G., Raychoudhury, R., Lavrov, D. V., & Werren, J. H. (2008). Rapidly Evolving Mitochondrial Genome and Directional Selection in Mitochondrial Genes in the Parasitic Wasp Nasonia (Hymenoptera: Pteromalidae). *Molecular Biology and Evolution*, *25*(10), 2167–2180. https://doi.org/10.1093/molbev/msn159

Osada, N., & Akashi, H. (2012). Mitochondrial-nuclear interactions and accelerated compensatory evolution: Evidence from the primate cytochrome C oxidase complex. *Molecular Biology and Evolution*, *29*(1), 337–346. https://doi.org/10.1093/molbev/msr211

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Piccinini, G., Iannello, M., Puccio, G., Plazzi, F., Havird, J. C., & Ghiselli, F. (2021). Mitonuclear Coevolution, but not Nuclear Compensation, Drives Evolution of OXPHOS Complexes in Bivalves. *Molecular Biology and Evolution*, *38*(6), 2597–2614. https://doi.org/10.1093/molbev/msab054

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rand, D. M., Haney, R. a, & Fry, A. J. (2004). Cytonuclear coevolution: The genomics of cooperation. *Trends in Ecology & Evolution*, *19*(12), 645–653. https://doi.org/10.1016/j.tree.2004.10.003

Rath, S., Sharma, R., Gupta, R., Ast, T., Chan, C., Durham, T. J., Goodman, R. P., Grabarek, Z., Haas, M. E., Hung, W. H. W., Joshi, P. R., Jourdain, A. A., Kim, S. H., Kotrys, A. V., Lam, S. S., McCoy, J. G., Meisel, J. D., Miranda, M., Panda, A., … Mootha, V. K. (2021). MitoCarta3.0: An updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Research*, *49*(D1), D1541–D1547. https://doi.org/10.1093/nar/gkaa1011

Sloan, D. B., Havird, J. C., & Sharbrough, J. (2016). The On-Again, Off-Again Relationship between Mitochondrial Genomes and Species Boundaries. *Molecular Ecology*, *38*(1), 42–49. https://doi.org/10.1111/ijlh.12426

Sloan, D. B., Triant, D. A., Wu, M., & Taylor, D. R. (2014). Cytonuclear Interactions and Relaxed Selection Accelerate Sequence Evolution in Organelle Ribosomes. *Molecular Biology and Evolution*, *31*(3), 673–682. https://doi.org/10.1093/molbev/mst259

Sukumaran, J., & Holder, M. T. (2010). DendroPy: A Python library for phylogenetic computing. *Bioinformatics (Oxford, England)*, *26*(12), 1569–1571. https://doi.org/10.1093/bioinformatics/btq228

Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, *34*(Web Server), W609–W612. https://doi.org/10.1093/nar/gkl315

Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, *21*(16), 3907–3930. https://doi.org/10.1111/j.1365-294X.2012.05664.x

Upham, N. S., Esselstyn, J. A., & Jetz, W. (2019). Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLOS Biology*, *17*(12), e3000494. https://doi.org/10.1371/journal.pbio.3000494

Williams, A. M., Friso, G., van Wijk, K. J., & Sloan, D. B. (2019). Extreme variation in rates of evolution in the plastid Clp protease complex. *The Plant Journal*, *98*(2), 243–259. https://doi.org/10.1111/tpj.14208

Wolfsberg, T. G., Schafer, S., Tatusov, R. L., & Tatusova, T. A. (2001). Organelle genome resources at NCBI. *Trends in Biochemical Sciences*, *26*(3), 199–203. https://doi.org/10.1016/S0968-0004(00)01773-4

Xu, S., Schaack, S., Seyfert, A., Choi, E., Lynch, M., & Cristescu, M. E. (2012). High Mutation Rates in the Mitochondrial Genomes of Daphnia pulex. *Molecular Biology and Evolution*, *29*(2), 763–769. https://doi.org/10.1093/molbev/msr243

Yan, Z., Ye, G., & Werren, J. H. (2019). Evolutionary Rate Correlation between Mitochondrial-Encoded and Mitochondria-Associated Nuclear-Encoded Proteins in Insects. *Molecular Biology and Evolution*, *36*(5), 1022–1036. https://doi.org/10.1093/molbev/msz036

Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591. https://doi.org/10.1093/molbev/msm088

Zhang, J., & Yang, J.-R. (2015). Determinants of the rate of protein sequence evolution. *Nature Reviews Genetics*, *16*(7), 409–420. https://doi.org/10.1038/nrg3950

Zieliński, P., Nadachowska-Brzyska, K., Wielstra, B., Szkotak, R., Covaciu-Marcov, S. D., Cogălniceanu, D., & Babik, W. (2013). No evidence for nuclear introgression

despite complete mtDNA replacement in the Carpathian newt (Lissotriton montandoni). *Molecular Ecology*, *22*(7), 1884–1903. https://doi.org/10.1111/mec.12225