

RESEARCH ARTICLE

# Mitochondrial DNA sequence context in the penetrance of mitochondrial t-RNA mutations: A study across multiple lineages with diagnostic implications

Rachel A. Queen<sup>1</sup>, Jannetta S. Steyn<sup>1</sup>, Phillip Lord<sup>2</sup>, Joanna L. Elson<sup>1,3\*</sup>

**1** Institute of Genetic Medicine, Newcastle University, Newcastle-upon-Tyne, United Kingdom, **2** School of Computing Science, Newcastle University, Newcastle-upon-Tyne, United Kingdom, **3** Centre for Human Metabonomics, North-West University, Potchefstroom, South Africa

\* [joanna.elson@newcastle.ac.uk](mailto:joanna.elson@newcastle.ac.uk)



**OPEN ACCESS**

**Citation:** Queen RA, Steyn JS, Lord P, Elson JL (2017) Mitochondrial DNA sequence context in the penetrance of mitochondrial t-RNA mutations: A study across multiple lineages with diagnostic implications. PLoS ONE 12(11): e0187862. <https://doi.org/10.1371/journal.pone.0187862>

**Editor:** Dong Hoon Shin, Seoul National University College of Medicine, REPUBLIC OF KOREA

**Received:** March 2, 2017

**Accepted:** October 28, 2017

**Published:** November 21, 2017

**Copyright:** © 2017 Queen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Complete mitochondrial genome sequences were downloaded from NCBI in GenBank format. Please refer to the Material and Methods section of this paper for data access instructions and accession numbers.

**Funding:** Fund from a Newcastle University internal fund. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

Mitochondrial DNA (mtDNA) mutations are well recognized as an important cause of inherited disease. Diseases caused by mtDNA mutations exhibit a high degree of clinical heterogeneity with a complex genotype-phenotype relationship, with many such mutations exhibiting incomplete penetrance. There is evidence that the spectrum of mutations causing mitochondrial disease might differ between different mitochondrial lineages (haplogroups) seen in different global populations. This would point to the importance of sequence context in the expression of mutations. To explore this possibility, we looked for mutations which are known to cause disease in humans, in animals of other species unaffected by mtDNA disease. The mt-tRNA genes are the location of many pathogenic mutations, with the m.3243A>G mutation on the mt-tRNA-Leu(UUR) being the most frequently seen mutation in humans. This study looked for the presence of m.3243A>G in 2784 sequences from 33 species, as well as any of the other mutations reported in association with disease located on mt-tRNA-Leu(UUR). We report a number of disease associated variations found on mt-tRNA-Leu(UUR) in other chordates, as the major population variant, with m.3243A>G being seen in 6 species. In these, we also found a number of mutations which appear compensatory and which could prevent the pathogenicity associated with this change in humans. This work has important implications for the discovery and diagnosis of mtDNA mutations in non-European populations. In addition, it might provide a partial explanation for the conflicting results in the literature that examines the role of mtDNA variants in complex traits.

## Introduction

Mitochondria are double membrane-bound organelles found in eukaryote cells. They have a number of roles within the cells, including providing most of the cells energy through oxidative phosphorylation (OXPHOS). Additionally, they are involved in apoptotic cell death, the control of calcium concentration and other cellular pathways [1]. Mitochondria operate under

**Competing interests:** The authors have declared that no competing interests exist.

the dual control of the nuclear genome and mtDNA. The human mitochondrial genome or mitochondrial DNA (mtDNA) is a 16,569 bp circular strand of DNA. The mitochondrial chromosome contains 37 genes, which encode for 13 proteins, 22 tRNA molecules and 2 ribosomal RNA molecules. The 13 proteins form the core of four of the five enzymes that carry out OXPHOS. The mtDNA has several unusual features: mitochondria and mitochondrial DNA are inherited strictly through the maternal line [2], so the mtDNA does not undergo bi-parental recombination. MtDNA is found in high copy number ranging from several hundreds to thousands per cell. Within a single cell, mtDNA copies may all be same (*homoplasmic*) or different (*heteroplasmic*) [1]. In humans, pathogenic mutations are normally seen as heteroplasmic variants; that is there is a mixture of wild-type and mutant mtDNA [1, 3]. MtDNA mutations do not exhibit a clinical phenotype until they are present of the majority of the mtDNA molecules within the cells.

Mutations of mtDNA are an important cause of inherited disease; over 250 pathogenic mutations, deletions and re-arrangements have been identified in humans [1, 3]. Mitochondrial diseases predominantly affect tissues with a high energy demand; so, neurological and muscular symptoms are common. Mitochondrial diseases are the most common cause of inherited metabolic diseases within new-borns [4]. In the UK adult population, the minimum prevalence rate for mtDNA mutations is estimated to be 1 in 5,000 [5]. Despite excellent estimates of disease frequency in some, particularly European, populations [6, 7], the spectrum of mutations that cause disease is less well understood in others, such as Black Africans. This leads to less accurate assessment of the impact of mtDNA disease in these populations [8, 9].

With maternal inheritance, mtDNA evolution has resulted in the emergence of distinct lineages called haplogroups. While the high level of mtDNA variation seen within human populations is useful to study population histories [10], it has made the identification of disease causing mutations difficult [7, 11]. Mitochondrial variation in humans has been extensively studied and has been compiled in MITOMAP database (<http://www.mitomap.org>) [12], Phylotree (<http://www.phylotree.org/>) [13] and tRNA-MAMIT online resource [14]. tRNA-MAMIT provides an online resource comparing single representative mt-tRNA sequences from 150 mammals which have been aligned to the revised Cambridge Reference Sequence (rCRS) genome [15]. The specific mitochondrial background (or haplogroup) may influence the expression of a disease causing mutations [16–18]. In human populations, there is some evidence suggesting that additional polymorphisms can either repress or compensate for pathogenic mutations [19, 20]; this has been proposed as an explanation for how different human populations differ in both the spectrum of mtDNA mutations and the variable symptoms present in patients [16].

Many of the mtDNA point mutations reported to cause disease in humans are located in one of the 22 mt-tRNA's, which are encoded by mtDNA. Although mt-tRNA genes make up just 10% of the mitochondrial chromosome, ~60% of adult patients with mtDNA disease have a mutation within these genes [3]. Each mt-tRNA gene is approximately 70–75 base pairs in length. The secondary structure of tRNA molecule is described as clover leaf which is made of four domains, namely: the acceptor (acc) stem; dihydrouracil (D) stem/loop; the anticodon (ac) stem/loop; and the T Ψ C (T) stem/loop. The four stems in the tRNA molecule primarily exist as Watson-Crick base pairs, which maintain the clover leaf structure [21]. The clover leaf structure folds into a 3D L-shaped molecule which, in turn, is held together by long range tertiary interactions. These tertiary interactions between bases stabilise non Watson-Crick base pairings in the stem regions of the tRNA molecule. mt-tRNA is variable and significantly different to genomic tRNA both in terms of the size of the D and T domains and nucleotide sequence. Although there are some differences between the mitochondrial encoded and

nuclear encoded tRNA's, a detailed study of mammalian mt-tRNA has shown that many generalisations about the mt-tRNA molecules still apply [21].

The most common point mutation seen in Caucasian European patients with mitochondrial disease is called m.3243A>G. [22, 23]. Multiple phenotypes are associated with m.3243A>G, including Maternally Inherited Diabetes and Deafness (MIDD) and Mitochondrial Encephalomyopathy, Lactic acidosis, and Stroke-like episodes (MELAS). 80% of MELAS patients carry the m.3243A>G mutation [24]. The nucleotide at position 3243 within the mtDNA molecule, equates to nucleotide 14 within the mt-tRNA-Leu(UUR) molecule. This nucleotide is predicted to have a tertiary reaction with nucleotides at positions 8, 14 and 21 [21]. The mutation therefore disrupts the tertiary folding structure of the mt-tRNA-Leu which: reduces the capacity for amino-acylation [25], prevents the termination of translation by the 16S ribosomal subunit [26]; and, inhibits the methylation of the U in wobble position of the mt-tRNA molecule [1, 27]. Critically, this variant is rarely reported in Black African patients with mtDNA disease [28]

One way to explore the importance of haplogroup context on the expression of mtDNA disease is to look for mutations in other animals which are disease-causing in human in the absence of disease in the animal. Then to look for possible compensating mutations where disease is not evident in the non-human animal grouping. For this strategy to be valid it is important to demonstrate mitochondrial disease in other species. A deletion in the mt-tRNA-Tyr gene in position 5304, associated with sensory ataxic neuropathy, similar to human myopathies, has been observed in golden retrievers [29] this provides a clinical example that these variants can have an effect in other mammals. It would be expected that a non-domestic animal would have severely reduced fitness [30]. Additionally it is important to be able to examine a substantial number of sequences, some prior work in this area has used limited numbers of species only considering a single individual per species [31]. Perhaps most importantly clinically validated criteria need to be applied to variants reported as mutations to ensure that they are truly causative of primary mitochondrial disease [3, 7, 32], such clarity on the criteria for pathogenicity was not in place at the time of prior publications on this topic [30, 31].

In this investigation, we studied human disease-associated variants of mitochondrial mt-tRNA-Leu (UUR) gene in other chordates to further the understanding of sequence context in modulating disease expression. This study looked in detail at mt-tRNA-Leu (UUR) in all chordates, for which there are at least 30 complete mtDNA sequences listed in GenBank. We focused on the m.3243A>G mutation because this is both the most common point mutation known cause of mitochondrial disease and the best studied. We have found that this mutation does occur in quite a few animals and have identified several likely compensatory mutations. We also present a comprehensive survey of all the mt-tRNA's variants associated with disease in humans found to be fixed in other species, then consider those that are clinically proven as pathogenic [32]. We used publically available sequence resources that have grown massively over the last decade. This work supports the assertion that the mutational potential of mt-tRNA's is strongly contained by sequence context [30]. As such, this study helps us to better appreciate the importance of haplogroup contexts in the penetrance and expression on disease causing variants. In the last few years the importance of this type of understanding the variable penetrance of mtDNA mutations, and its impact the discovery and diagnosis of disease causing mutations, especially in populations that have been less well studied [8, 9] has become increasing clear.

## Material and methods

### Reference sequence

The revised Cambridge Reference Sequence (rCRS) [15], was used as a reference sequence used in this investigation (GenBank record NC\_012920.1). The position and order of mt-

tRNA genes within the mitochondrial chromosome varies between species therefore in this study, all genome positions refer to the equivalent nucleotide position within the rCRS rather than the actual nucleotide position for the individual species. Nucleotide positions within the mt-tRNA molecule itself are also included.

## Identification of SNPs associated with disease

A list of known or suspected pathogenic SNPs, a list of SNPs within mt-tRNA-Leu (UUR) gene, associated with disease, was compiled from the MITOMAP and tRNA-MAMIT databases.

Additionally, a literature search was performed to look for SNPs which had not been added to either of the databases. The PubMed search query "Mitochondrial tRNA" was used to look for any articles published between 2014 and 2015 which reported disease association with SNPs. The likelihood of a role in pathogenicity of each disease associated SNP was assessed the scoring algorithm described by Yarham et al was applied [32].

## Extraction of mt-tRNA sequences from GenBank files

Complete mitochondrial genome sequences were downloaded from NCBI in GenBank format [33]. Each record within the file contains the species name and positions of the individual mt-tRNA genes within the mitochondrial sequences, in the annotation section of the record. Custom python scripts were used to extract mt-tRNA sequences from GenBank files containing complete mtDNA sequences. The Biopython SeqIO module was used to extract the gene locations for each mt-tRNA genes within each individual record. A search for the term "product" within the record qualifiers object was used executed to capture the start and end locations of all mitochondrial genes. The mt-tRNA sequences were extracted from the file using the gene locations contained in the sequence annotations. Each mt-tRNA sequence, for each species studied, was saved in a separate FASTA file. The product names given to the sequences correspond to those used within the rCRS sequence. As there are two mitochondrial mt-tRNA-Leu and tRNA-Ser genes, the sequences were then compared to the rCRS using the Biopython pairwise module. The pairwise scores were used to determine which of the two genes were present and separate the genes into the correct file. When it was not possible to distinguish between the sequences based on pairwise score, the sequences were placed into a separate temporary file for visual inspection.

## Gene alignment

The mt-tRNA sequences from the different species were aligned to the complementary mt-tRNA sequence from the rCRS and these alignments were then used to look for nucleotide changes at specific locations associated with disease in humans.

The mt-tRNA gene sequences from the rCRS genome were added to the previously generated files. ClustalW was used to create the alignments of the sequences [34]. The Biopython AlignIO module was used to create the alignments for each file.

A Python script was written using the Biopython module AlignIO which was used to detect polymorphisms at specific positions within the alignments. The script allowed for gaps within the alignment by converting the nucleotide equivalent in the rCRS mt-tRNA molecule to the nucleotide position within the alignment.

Sequences were used to study 33 disease associated variants within tRNA-Leu (UUR) gene.

The mitoseq section of MITOMAP provides executable search terms to retrieve all available mitochondrial genome sequences in GenBank (<http://www.mitomap.org/bin/view.pl/MITOMAP/MitoSeqs>). To collect sequence data for study the search "Complete non-human mtDNA genomes" was executed, to create a file containing all non-human

mitochondrial genomes. The contents of this file was analysed using a script, which used the Biopython SeqIO module to generate a list of species to study. Species were selected if they met the following criteria:

1. 30 or more complete mitochondrial sequences were available within GenBank for the species. This threshold was set to allow the study of variation within species as well as between the species and humans to be conducted.
2. The species belonged to the phylum *Chordata*. The majority of mitochondrial diseases are neuromuscular therefore only species, possessing a central nervous system were chosen for study. Species which met these criteria are shown in [Table 1](#).

### Sequence quality control

GenBank contains both curated and uncurated sequences from a large number of sources, which are of differing quality. A number of tests were applied to the sequences to assess the

**Table 1. The 33 chordate species with greater than 30 complete mtDNA sequences used in the initial phase of the study.**

Species	Common Name	Number of Sequences before QC	Number of Sequences after QC
<i>Anguilla anguilla</i>	European Eel	55	54
<i>Anguilla rostrata</i>	American Eel	51	51
<i>Balaenoptera physalus</i>	Fin Whale	154	154
<i>Bison bison</i>	Bison	34	34
<i>Bos grunniens</i>	Yak	83	83
<i>Bos taurus</i>	Cow	275	275
<i>Canis lupus familiaris</i>	Dog	391	391
<i>Clupea harengus</i>	Atlantic Herring	100	100
<i>Coregonus lavaretus</i>	European whitefish	81	80
<i>Equus caballus</i>	Horse	254	245
<i>Gallus gallus</i>	Red Jungle Fowl	66	66
<i>Glyphis glyphis</i>	Speartooth Shark	94	94
<i>Hypophthalmichthys molitrix</i>	Silver carp	30	29
<i>Hypophthalmichthys nobilis</i>	Bighead carp	36	35
<i>Macaca fascicularis</i>	Crab-eating macaque	44	44
<i>Mus musculus</i>	mouse	53	50
<i>Mus musculus domesticus</i>	House mouse	59	59
<i>Myodes glareolus</i>	Bank vole	35	35
<i>Orcinus orca</i>	Killer Whale	87	87
<i>Ovis aries</i>	Sheep	94	94
<i>Pan paniscus</i>	Banobo	54	54
<i>Pan troglodytes schweinfurthii</i>	Eastern chimpanzee	33	33
<i>Pan troglodytes troglodytes</i>	Central chimpanzee	56	54
<i>Pan troglodytes verus</i>	Western chimpanzee	30	30
<i>Rattus norvegicus</i>	Brown Rat	66	66
<i>Sus scrofa</i>	Wild Boar	150	150
<i>Syncerus caffer</i>	African buffalo	45	45
<i>Tursiops truncatus</i>	Common bottlenose Dolphin	50	50
<i>Urocyon littoralis catalinae</i>	Island Fox	41	41
<i>Urocyon littoralis clementae</i>	Island Fox	33	33
<i>Urocyon littoralis santacruzae</i>	Island Fox	42	42
<i>Ursus arctos</i>	Brown Bear	74	74
<i>Ursus spelaeus</i>	Cave Bear (extinct)	34	20

<https://doi.org/10.1371/journal.pone.0187862.t001>

reliability of the data. (1) Sequence length of the mt-tRNA-Leu (UUR) genes extracted was compared to the length of the rCRS mt-tRNA-Leu (UUR) gene using the Biopython SeqIO module. Any sequences that were 5 nucleotides longer or shorter than the length of the rCRS mt-tRNA-Leu (UUR) gene were manually examined. (2) Unknown/unspecified bases (Ns), sequences with a high number of Ns are an indication of poor quality sequence data. The SeqIO module was used to determine the number of Ns with the mt-tRNA sequences. Any sequences which contained more than 5 Ns were removed from the study. (3) All of the mt-tRNA genes were aligned using ClustalW for each species and the statistics from these alignments was compiled. This file contained information about the number of sequences, the length of sequences within the file and a summary of the pairwise alignments scores. Pairwise scores are generated by ClustalW, before the full multiple sequence alignment, by comparing each pair of sequences to be aligned. They are a measure of the number of sequence identities divided by the sequence length. Mt-tRNA sequences from the same species should share a high degree of similarity and low similarity of sequences could be an indication of incorrect annotation, mislabelling of genes, or poor quality sequence data. The pairwise scores produced by ClustalW were used to highlight any sequences with a low degree of similarity to other sequences from the same species.

### tRNA secondary structure analysis

Two tools were used to study the secondary structures of tRNA genes. Alignments from this study were compared to the alignments in the tRNA-MAMIT [14] to determine the functional regions of the gene. “tRNAscan-SE Search Server” was used to predict the secondary structure of genes [35]. This software predicts the folding of tRNA molecules. The source was set to “Mito/ Chloroplast”.

### Phylogenetic analysis

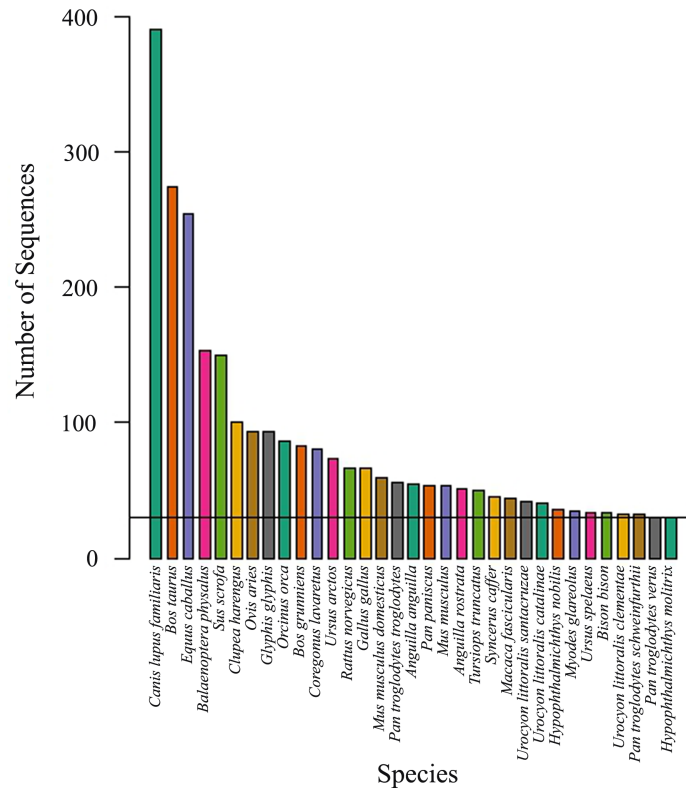
Phylogenetic analysis, NETWORK version 4.6.1.3 was used to study the phylogenetic relationship of mt-tRNA-Leu (UUR) sequences [36].

## Results

### Disease associated mitochondrial m-tRNA-Leu variants found in vertebrates

The first part of this study considered all disease-associated mutations located on mt-tRNA-Leu (UUR) using a panel of species. mt-tRNA-Leu(UUR) was chosen, as it contains a number of clinically significant mutations including the most common mtDNA point mutation m.3243A>G. Initially 16,992 complete eukaryote (non-human) mitochondrial DNA genomes were downloaded from the GenBank database, and then all chordate sequences were extracted. A total of 33 species were selected for use in the study, with the majority being vertebrates, each with 30 or more complete sequences available (Fig 1). Details of species selection (Table 1), quality control (QC) applied to the sequences used and pipeline validation are described in the methods, with the detailed results of the QC being described in the supplemental data.

The restriction to species with 30 sequences allows us to assess the within-species variation critical to this study, as disease causing mutations will occur only at low frequencies. In total, 2784 sequences were available from the 33 species selected for use in the study. The sequences were derived from 406 independent studies. A total of 32 variants located on mt-tRNA-Leu (UUR) are reported to be associated with disease on either MITOMAP, tRNA-MAMIT or



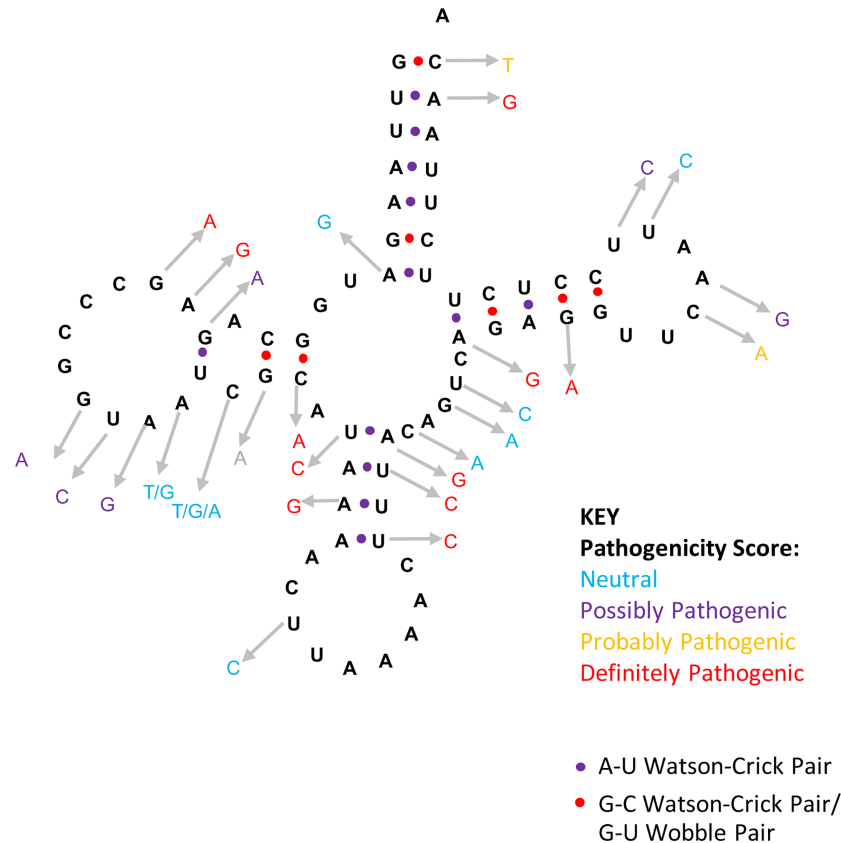
**Fig 1. Species from the phylum Chordata with over 30 complete mitochondrial sequences.**

<https://doi.org/10.1371/journal.pone.0187862.g001>

from a literature search. Therefore, we examined each of the 2784 sequences for any of these 32 variants.

One potential problem is that the disease associations in the various online databases may not all be well supported by clinical and laboratory evidence [37]. So, we conducted an up-to-date literature search to assess the variants reported as pathogenic in humans using a scoring system applied widely in the mitochondrial field [32]. This system considers factors such as the number of times a mutation has been reported and evolutionary conservation. However, critical to its reliability is the emphasis on laboratory investigation of the proposed mutation. In particular, experiments that link genotype and phenotype, such as single muscle fibre analysis. After completing the re-evaluation of the mutations reported in the literature for mt-tRNA-Leu (UUR), 12 of the 32 reported mutations were designated as neutral polymorphisms, 8 were scored possibly pathogenic, 1 was scored probably pathogenic, while 11 scored as definitely pathogenic. The distribution of the human mutations on the mt-tRNA structure is shown Fig 2. These results support previous analysis of the data in these public databases [7, 32, 37]. This assessment was used to guide our interpretation of the data that we found in other species. In other words, to ensure that any disease causing variants seen in other species in the absence of disease are bona fide disease causing mutations.

Of the 32 known or suspected pathogenic variants located on the human mt-tRNA-Leu (UUR), 12 out of 32 were *monomorphic* in another species. That is present in all sequences from one (or more) non-human species. In Table 2, the variants considered to be definitely or probably pathogenic, after the application of the Yarham criteria described above, are highlighted. Six of these variants were located within stem regions of the mt-tRNA molecule



**Fig 2. Classification of variants with reports of disease association within mt-tRNA-Leu (UUR) gene.**

<https://doi.org/10.1371/journal.pone.0187862.g002>

and six within the loop or the variable regions. Of the six present in the stem regions, all showed a corresponding change in the other arm of the stem, so maintaining the Watson-Crick type pairing; and also the secondary structure. This is important because variants which disrupt Watson-Crick pairs in mt-tRNA molecules are frequently classed as pathogenic mutations [37]. Interestingly, the m.3251A>G mutation associated with mitochondrial myopathy seen in the mt-tRNA d-loop [38, 39] was present in all 74 sequences from *Ursus arctos* (brown bear) and all 20 sequences from *Ursus spelaeus* (cave bear). This mutation corresponds to base 22 in mt-tRNA-Leu(UUR); it is involved in a long range tertiary interaction affecting the following triplet (13–22)–46 [21]. The *Ursus arctos* and *Ursus spelaeus* sequences also differ at base 46, with a change of a C to T changing the nature of the tertiary structure [21] and potentially acting as a compensatory change.

We observed a number of the variants associated with specific species or taxonomic groups such as the m.3264T>C, m.3271T>C, m.3273T>C, m.3275 C>A, m.3302A>G, which were all confined to species of fish. Other such observations include the occurrence of the m.3271T>C variant in fish and rodents; similarly, mutations are often seen in parallel branches of the human phylogeny, probably due to mtDNA’s high rate of mutation [40]. The m.3254C>G SNP was monomorphic in all species except *Pan paniscus* (bonobo), which had a C at this position, and the m.3226A>G SNP was monomorphic in all non-primate species.

We found an additional four disease associated variants in other species, but here there were two polymorphisms at the sites in sequences of the species in question (Table 3).



**Table 2. SNPs which are known or suspected to be pathogenic in humans which are seen in 100% of sequences from other species, positions shown equates to location within rCRS sequence.**

Position	Region	Variant	Status	Secondary Structure	Tertiary Structure	Species
3236 (Bosley <i>et al.</i> , 2008)	acc stem	A-G	Neutral			All non primate species
3250 (Goto <i>et al.</i> , 1992)	d-loop	T-C	Possibly Pathogenic	N/A	(8–14)-21	<i>Glyphis glyphis</i> , <i>Orcinus orca</i> , <i>Gallus gallus</i> , <i>Tursiops truncatus</i>
3251 (Sweeney <i>et al.</i> , 1993)	d-loop	A-G	Possibly Pathogenic	N/A	(13–22)-46G–A>G–C>T	<i>Ursus arctos</i> , <i>Ursus spelaeus</i>
3254 (Chen <i>et al.</i> , 2000)	d-stem	C-T	Neutral	A-C mismatch to A-T pair	(25–10)-45	All species except <i>Pan paniscus</i>
3264 (Suzuki <i>et al.</i> , 1997)	ac-loop	T-C	Neutral	N/A		<i>Glyphis glyphis</i> , <i>Coregonus lavaretus</i> , <i>Anguilla anguilla</i> , <i>Anguilla rostrata</i> , <i>Hypophthalmichthys nobilis</i> , <i>Hypophthalmichthys molitrix</i>
3271 (Goto <i>et al.</i> , 1991)	ac-stem	T-C	Definitely Pathogenic	A-T pair to C-G pair		<i>Glyphis glyphis</i> , <i>Coregonus lavaretus</i> , <i>Rattus norvegicus</i> , <i>Gallus gallus</i> , <i>Mus musculus domesticus</i> , <i>Anguilla anguilla</i> , <i>Anguilla rostrata</i> , <i>Mus musculus</i> , <i>Hypophthalmichthys nobilis</i> , <i>Hypophthalmichthys molitrix</i> , <i>Myodes glareolus</i>
3273 (Campos <i>et al.</i> , 2001)	ac-stem	T-C	Definitely Pathogenic	A-T pair to C-G pair	26-44 No Change	<i>Clupea harengus</i>
3275 (Garcia-Lozano <i>et al.</i> , 2000)	Variable Region	C-A	Neutral	N/A	(13–22)-46	<i>Hypophthalmichthys nobilis</i> , <i>Hypophthalmichthys molitrix</i>
3290 (Zhu <i>et al.</i> , 2009)	t-loop	T-C	Neutral	N/A		<i>Equus caballus</i> , <i>Sus scrofa</i> , <i>Anguilla anguilla</i> , <i>Anguilla anguilla</i> , <i>Anguilla rostrata</i> , <i>Macaca fascicularis</i> , <i>Urocyon littoralis santacruzae</i> , <i>Urocyon littoralis catalinae</i> , <i>Urocyon littoralis clementae</i>
3291 (Goto <i>et al.</i> , 202AD)	t-loop	T-C	Possibly Pathogenic	N/A		<i>Coregonus lavaretus</i>
3302 (Bindoff <i>et al.</i> , 1993)	acc stem	A-G	Definitely Pathogenic	A-T pair to C-G		<i>Clupea harengus</i> , <i>Coregonus lavaretus</i> , <i>Gallus gallus</i> , <i>Hypophthalmichthys nobilis</i> , <i>Hypophthalmichthys molitrix</i>
3303 (Silvestri <i>et al.</i> )	acc stem	C-T	Probably Pathogenic	A-T pair to C-G		<i>Ursus spelaeus</i> , <i>Sus scrofa</i> , <i>Rattus norvegicus</i> , <i>Myodes glareolus</i> , <i>Mus musculus domesticus</i> , <i>Mus musculus</i> , <i>Hypophthalmichthys nobilis</i> , <i>Hypophthalmichthys molitrix</i> , <i>Clupea harengus</i>

<https://doi.org/10.1371/journal.pone.0187862.t002>

- The m.3243A>G located in the d-loop of the mt-tRNA molecule, is the most common point mutation associated with mitochondrial disease in humans. The m.3243A>G mutation was seen in 57 out of 391 sequences from *Canis lupus familiaris* (dog) [29].
- The m.3244G>A mutation also located in the mt-tRNA's d-loop, was seen in 2 out of 72 sequences from *Ursus arctos* (HQ685945, HQ685940). Both sequences were obtained from a study of migration patterns among brown bears. The m.3244G>A mutation has been

**Table 3. The frequency of known/ suspected pathogenic SNPs which are polymorphic in other species (Position shown equates to location within rCRS genome sequence).**

Base	Region	Variant	Status	Species	No. Sequences
3243 (Goto <i>et al.</i> , 1990)	D-Loop	A>G	Definitely pathogenic	<i>Canis lupus familiaris</i>	57/391
3244 (Mimaki <i>et al.</i> , n 2009)	D-Loop	G>A	Definitely pathogenic	<i>Ursus arcto</i>	2/72
3249 (Seneca <i>et al.</i> , 2001)	D-Loop	A>G	Possibly Pathogenic	<i>Sus scrofa</i>	1/150
3290 (Zhu <i>et al.</i> , 2009)	T-Loop	T>C	Neutral	<i>Canis lupus familiaris</i>	2/391
3290 (Zhu <i>et al.</i> , 2009)	T-Loop	T>C	Neutral	<i>Pan paniscus</i>	1/54

<https://doi.org/10.1371/journal.pone.0187862.t003>

observed in a patient with MELAS [41] and using the pathogenicity scoring criteria [7] it is classed as definitely pathogenic in humans.

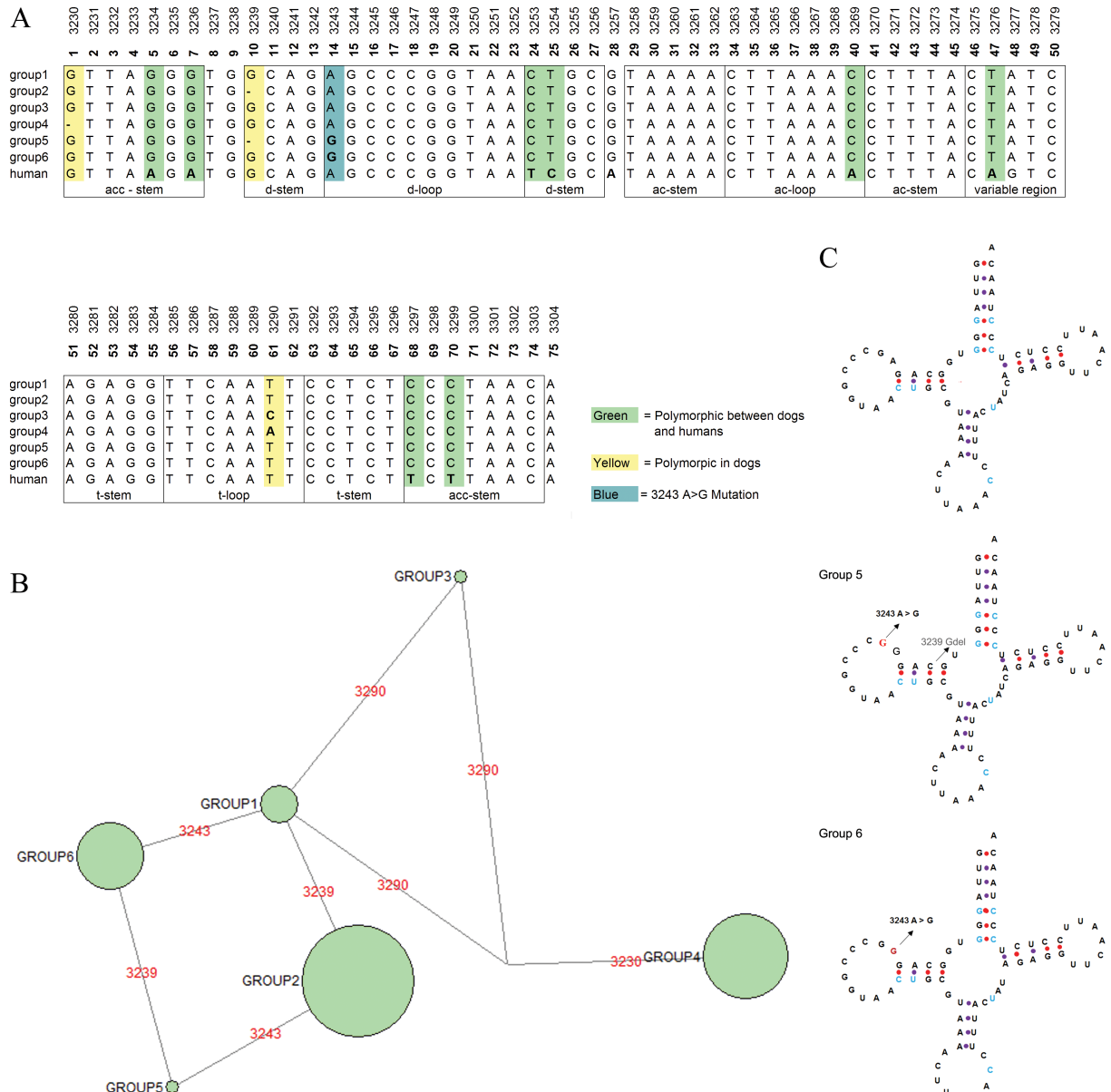
- The m.3249G>A mutation, which is located in the mt-tRNA's d-loop, was seen in 1 in 150 *Sus scrofa* (Wild boar) sequences (DQ268530). This mutation was first detected in 2001 in one patient with a clinical phenotype resembling Kearns-Sayre syndrome [42, 43]. However, it was only designated as possibly pathogenic using the scoring system of Yarham et al 2011 [32].
- The m.3290T>C mutation was seen in 2/391 sequences from *Canis lupus familiaris* and 1/54 *Pan paniscus* sequences. When this mutation was investigated, it was also found to be polymorphic in humans, being seen in multiple populations, and with limited evidence existed linking it with disease. It has, therefore, been classified as neutral [32].

### Differences in secondary structure of Mt-tRNA-Leu (UUR) with the m.3243A>G in dog (*Canis lupus familiaris*)

The largest number of sequences available for a single group was from the dog (*Canis lupus familiaris*) with a total of 391 sequences. All available sequences from the dog were aligned with the human mt-tRNA-Leu gene. The resulting alignment was 75bp in length, which corresponds to the entire length of the mt-tRNA-Leu(UUR) gene. The dog mt-tRNA-Leu(UUR) sequence is divergent from the human reference sequence at eight locations, shown in green on the alignment (Fig 3A). The m.3243A>G mutation was found in 57 out of 391 alignments of dog mt-tRNA-Leu(UUR). Prior work has suggested that this variant was fixed in dogs, this work was conducted when the available sequence data was more limited than today [30] the availability of clades with and without the m.3243A>G mutation makes for a more powerful model to investigate possible compensatory changes. In addition to the m.3243A>G mutation, there are three other positions within the mt-tRNA-Leu(UUR) which where the sequences were divergent from human mt-tRNA-Leu gene.

As mentioned, the presence of dog sequence with and without the well-studied m.3243A>G mutation has allowed a more detailed investigation into possible compensatory effects than was possible previously [30]. The polymorphisms were used to create a phylogenetic network that contained six haplogroups (Fig 3B), using NETWORK [36]. The haplotypes used are shown in (Table 4). The largest group was haplogroup 1 containing 328 sequences, which was 83% of the total. The other groups most likely evolved from haplogroup 1, as larger groups are believed to have a higher likelihood of being ancestral [36]. Two deletions were observed in position 3230 and 3239, seen in groups 4 and groups 2 and 5 respectively. The m.3290T>A variant was observed in group 2 and m.3290T T>C was observed in group 4. The m.3243A>G mutation was confined to two haplogroups (haplogroup 6 and haplogroup 5). We postulated alterations to the secondary structure of the mt-tRNA-Leu(UUR) suppress the pathogenicity of the m.3243A>G mutation. To address this, we analysed the secondary structure of sequences containing the m.3243A>G polymorphism. Two tools were used to study the secondary structures of the gene: we compared alignments to those from tRNA-MAMIT [14] to determine the functional regions of the gene; and "tRNAscan-SE Search Server, was used to predict the secondary structure of genes [35]. The predicted secondary structures of the three mt-tRNA-Leu (UUR) dog haplotypes are shown in (Fig 3C).

Two positions, divergent from the human sequence, and within the D-Stem of the dog sequences are particularly interesting. First, m.3253T>C, as all of the sequences have a C at position 24 of the mt-tRNA-Leu molecule, whereas the human sequence has T at the same position. This corresponds to the final base in the D- Stem and changes a wobble pair into a Watson-Crick pair. This polymorphism was only seen in 8 other sequences included in this



**Fig 3.** (A) Alignment of 6 unique mt-tRNA-Leu (UUR) sequences from *Canis lupus familiaris* with the rCRS. Reticulation indicates one of the variants observed at positions 3239, and 3243 has occurred more than once in *Canis lupus familiaris*. (B) A phylogenetic network, created using *Canis lupus familiaris* mt-tRNA-Leu (UUR) sequences. The first structure shows the positions of polymorphisms between all of the sequences studied from *Canis lupus familiaris* and humans. The second two structures show the tRNA molecule where the 3243 A>G polymorphism was observed in *Canis lupus familiaris*. (C) The secondary structure of the mt-tRNA-Leu (UUR) from *Canis lupus familiaris*.

<https://doi.org/10.1371/journal.pone.0187862.g003>

study, all from *Macaca fascicularis* (crab-eating macaque). Second m.3254 C>T, as there is T at position 25 within the molecule instead of a C creating an extra Watson-Crick pair in the d-stem of the dog mt-tRNA-Leu molecule. Both of these variants change the secondary structure of the molecule and we hypothesise that they could suppress the pathogenicity of the m.3243A>G mutation.

The sequences from dog are divergent from the human sequence in six other locations. Four of these changes are in the acc-stem maintaining Watson-Crick base pairings but

**Table 4. The frequency of key SNPs on the *Canis lupus familiaris* phylogeny (Position shown equates to location within rCRS, human mtDNA reference sequence).**

Sequence	Character				Frequency
	3230	3239	3243	3290	
Group 1	G	G	A	T	328
Group 2	G	-	A	T	3
Group 3	G	G	A	C	2
Group 4	-	G	A	A	1
Group 5	G	-	G	T	2
Group 6	G	G	G	T	55

<https://doi.org/10.1371/journal.pone.0187862.t004>

changing the pairs from A-T in humans to G-C in *Canis lupus familiaris*. A further difference is located in the variable region, as well as the T-loop. These six variants maintain and do not alter the secondary structure of the molecule and so were not studied further.

### The distribution of predicted compensatory variants, identified in *Canis lupus familiaris*, in other vertebrates

One possible reason for the presence of mutations that are human disease-associated, in species other than humans is that these mutations are compensated for; therefore, we conducted an extended study of the variants m.3254C>T and m.3253T>C, which may compensate for the m.3243A>G mutation. The investigation was extended by studying m.3254C>T, m.3253T>C where it co-occurred with the m.3243A>G in 10,426 GenBank records from vertebrates. The m.3254C>T and m.3253T>C variants are interesting as they change the secondary structure of the mt-tRNA molecule and are present in all sequences containing the m.3243A>G in *dog*. The deletion m.3239G>- was not studied further because it was only observed in 3.5% of the sequences, belonging to phylogenetic group 5 (see Fig 3C), which contain the m.3243A>G mutation and the network analysis suggested that it was acquired after the m.3243A>G. The m.3254C>T polymorphism was within 97% of the vertebrate species sequences analysed with an even distribution amongst taxonomic groups. The m.3254C appears to be the most common genotype in the majority of vertebrate species studied.

Considering next the m.3253T>C polymorphism within other vertebrate, it was observed within 15 out of 139 orders/suborders/superorders studied (Table 5). It was present with the highest frequency within the taxonomic order *Proboscidea*, where it was found in all 37 sequences analysed. Additionally, it was observed with a frequency of 0.436 within the order *Carnivora*, which contains canid species. Within the family *Canidae*, m.3253T>C is present in 421 out of 631 sequences. It is present in 100% of sequences from sub species of *Canis lupus* and in 1/5 *Canis latrans* (coyote) sequences (Table 6).

The m.3253T>C variant is also present in the families *Phocidae*, *Felidae* and *Urisdae* with frequencies of 25%, 21% and 0.6%. The m.3243A>G mutation was found along with m.3253T>C mutation in 2 sequences from the species *Leptonychotes weddellii* (weddell seal), which belongs to a taxonomic family of earless seals and *Phocidae* within the order *Carnivora* (Fig 4). Only two sequences were available for the species *Leptonychotes weddellii*. A total of 32 sequences were analysed from the family *Phocidae*, with the m.3253T>C mutation present in 8 of these sequences. All of species within *Phocidae* contained the m.3254C>T, which provides an additional Watson-Crick pairing in the d-stem of the mt-tRNA-Leu(UUR) molecule that is not seen in humans. An additional polymorphism was present at position m.3256C>T in 2 sequences from *Hydrurga leptonyx* (leopard seal) and 2 sequences from *Lobodon carcinophaga* (crabeater seal); this decreases the number of Watson-Crick pairs in the mt-tRNA-Leu(UUR)

**Table 5. The number of sequences containing the m.3253T>C polymorphism and the frequency of the occurrence within vertebrata groupings.**

Order	Present	Total Sequences	Frequency
Proboscidea	37	37	1.000
Carnivora	438	1005	0.436
Pholidota	4	11	0.364
Primates	50	604	0.237
Neoteleostei	320	1886	0.170
Acipenseriformes	2	33	0.061
Perissodactyla	9	299	0.030
Chiroptera	2	87	0.023
Cryptodira	3	150	0.020
Squamata	7	379	0.018
Caudata	2	186	0.011
Anguilliformes	2	217	0.009
Ostariophysi	9	1261	0.007
Rodentia	2	454	0.004
Cetartiodactyla	2	1599	0.001

<https://doi.org/10.1371/journal.pone.0187862.t005>

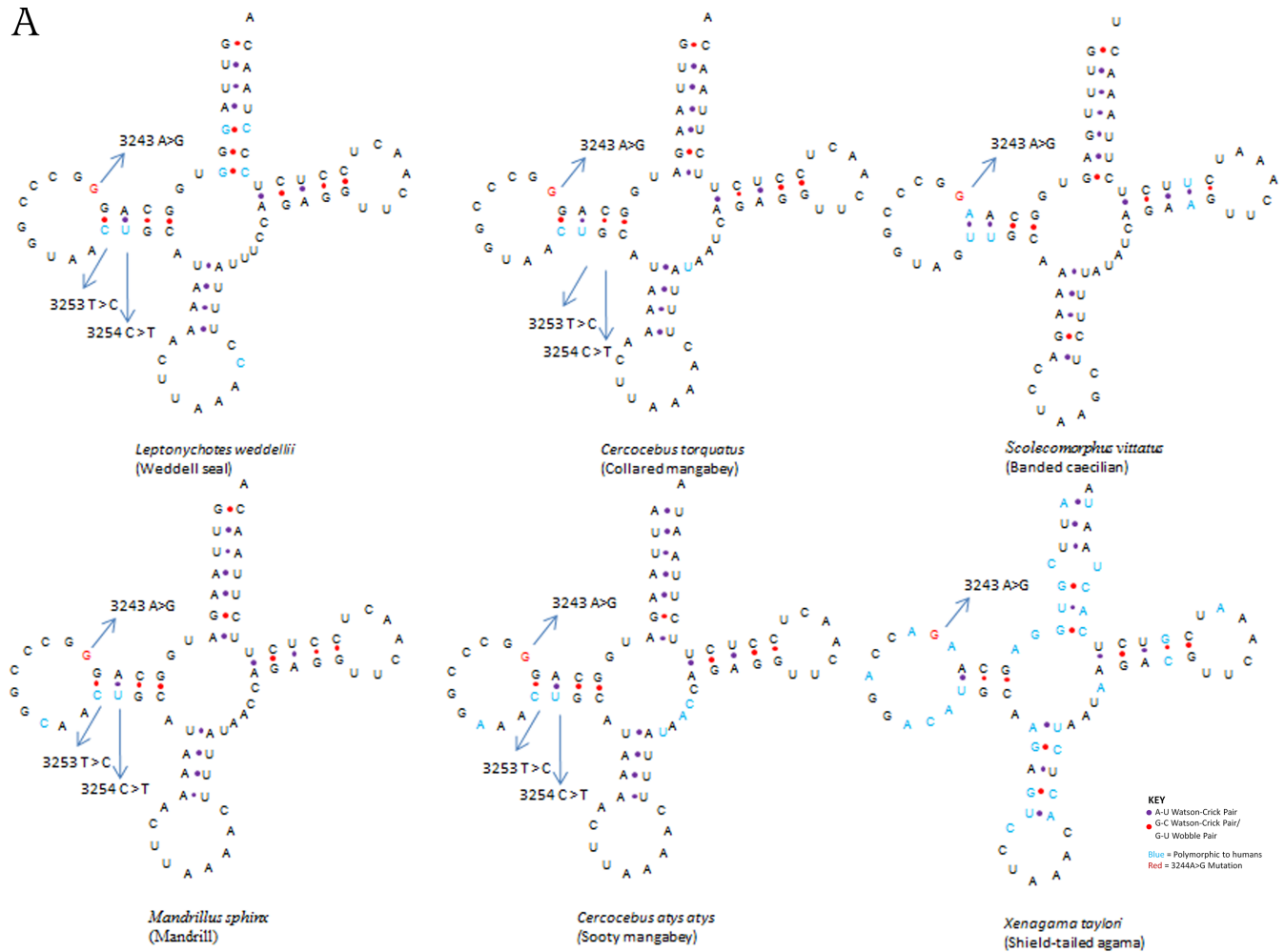
d-stem in these species. Additionally, a total of 604 mt-tRNA-Leu(UUR) sequences from primate species were analysed. The m.3243A>G mutation along with m.3253T>C variant were found in three species of old world monkeys within the family *Cercopithecidae*. In the species *Mandrillus sphinx*, these SNPs were present in all 3 of the sequences available for the species (KJ434963.1, NC\_021956.1, KC757403.1). Both SNPs were also observed in the only sequence from (KP090062.1) and in the only 2 sequences available for *Cercocebus torquatus* (white collared mangabey) (NC\_023964.1, KJ434959.1).

The m.3243A>G mutation was found in 2 sequences from *Scolecophorus vittatus* (AY456253.1 and NC\_006304.1), a legless amphibian. The sequences came from different sources but are identical. The ClustalW pairwise score between the human and *Scolecophorus vittatus* sequences within the alignment is 73.33 indicating a low degree of similarity between the gene sequences. The pairwise score is a measurement of the number of identities in a pair

**Table 6. Percentage of sequences carrying the 3243A>G and 3253T>C SNP within the order carnivora.**

Family	Species	3243A>G	3253A>C
Canidae	<i>Canis lupus familiaris</i>	15	100
Canidae	<i>Canis lupus</i>	0	100
Canidae	<i>Canis lupus campestris</i>	0	100
Canidae	<i>Canis lupus desertorum</i>	0	100
Canidae	<i>Canis lupus lupus</i>	0	100
Canidae	<i>Canis lupus chanco</i>	0	100
Canidae	<i>Canis lupus laniger</i>	0	100
Canidae	<i>Canis latrans</i>	0	20
Phocidae	<i>Lobodon carcinophaga</i>	100	100
Phocidae	<i>Hydrurga leptonyx</i>	0	100
Phocidae	<i>Leptonychotes weddellii</i>	100	100
Phocidae	<i>Mirounga leonina</i>	0	100
Ursidae	<i>Ursus thibetanus</i>	0	8
Felidae	<i>Puma concolor</i>	0	100
Felidae	<i>Felis catus</i>	0	100
Felidae	<i>Acinonyx jubatus</i>	0	100

<https://doi.org/10.1371/journal.pone.0187862.t006>



KEY  
 ● A-U Watson-Crick Pair  
 ● G-C Watson-Crick Pair/  
 ● G-U Wobble Pair  
 Blue = Polymorphic to humans  
 Red = 3244A>G Mutation

B

Carnivora - Phocidae			Primates - Cercopithecidae		
D - Stem	D- Loop	D-Stem	D - Stem	D- Loop	D-Stem
<i>Pusa sibirica</i>	G C A G A G C C C G G C A A T T G C	G C A G A G C C C G G C A A T T G C	<i>Mandrillus sphinx</i>	G C A G G G C C C G G C A A C T G C	G C A G G G C C C G G C A A C T G C
<i>Phoca caspica</i>	G C A G A G C C C G G C A A T T G C	G C A G A G C C C G G C A A T T G C	<i>Cercocebus atys aty</i>	G C A G G G C C C G G A A A C T G C	G C A G G G C C C G G A A A C T G C
<i>Phoca largha</i>	G C A G A G C C - G G C A A T T G C	G C A G A G C C - G G C A A T T G C	<i>Cercocebus torquatus</i>	G C A G G G C C C G G T A A C T G C	G C A G G G C C C G G T A A C T G C
<i>Halichoerus grypus</i>	G C A G A G C C - G G C A A T T G C	G C A G A G C C - G G C A A T T G C	<i>Cercocebus chrysogaster</i>	G C A G A G C C C G G T A A C T G C	G C A G A G C C C G G T A A C T G C
<i>Phoca vitulina</i>	G C A G A G C C - G G C A A T T G C	G C A G A G C C - G G C A A T T G C	<i>Cercopithecus albogularis</i>	G C A G A G C C C G G C A A T T G C	G C A G A G C C C G G C A A T T G C
<i>lchoerus grypus</i>	G C A G A G C C - G G C A A T T G C	G C A G A G C C - G G C A A T T G C	<i>Homo sapiens</i>	G C A G A G C C C G G T A A T C G C	G C A G A G C C C G G T A A T C G C
<i>Phoca groenlandica</i>	G C A G A G C C C G G C A A T T G C	G C A G A G C C C G G C A A T T G C			
<i>Pusa hispida</i>	- C A G A G C C C G G C A A T T G C	- C A G A G C C C G G C A A T T G C			
<i>Phoca caspica</i>	G C A G A G C C C G G C A A T T G C	G C A G A G C C C G G C A A T T G C			
<i>Phoca groenlandica</i>	G C A G A G C C C G G C A A T T G C	G C A G A G C C C G G C A A T T G C			
<i>Cystophora cristata</i>	G C A G A G C C C G G C A A T T G C	G C A G A G C C C G G C A A T T G C			
<i>Phoca fasciata</i>	G C A G A G C C C G G C A A T T G C	G C A G A G C C C G G C A A T T G C			
<i>Monachus schauinslandi</i>	G C A G A G C C C G G C A A T T G C	G C A G A G C C C G G C A A T T G C			
<i>Erigonathus barbatu</i>	G C A G A G C C C G G C A A T T G C	G C A G A G C C C G G C A A T T G C			
<i>Hydrurga leptonyx</i>	G C A G A G C C C G G T A A C T G T	G C A G A G C C C G G T A A C T G T			
<i>Lobodon carcinophaga</i>	G C A G A G C C C G G T A A C T G T	G C A G A G C C C G G T A A C T G T			
<i>Leptonychotes weddellii</i>	G C A G G G C C C G G T A A C T G C	G C A G G G C C C G G T A A C T G C			
<i>Mirounga leonina</i>	G C A G A G C C C G G T A A C T G C	G C A G A G C C C G G T A A C T G C			
<i>Homo sapiens</i>	G C A G A G C C C G G T A A T C G C	G C A G A G C C C G G T A A T C G C			

Key  
 Green = Polymorphic from rCRS  
 Yellow = Polymorphic between species  
 Blue = 3243 A>G mutation

Fig 4. (A) The secondary structure of mt-tRNA-Leu molecules carrying the 3243 A>G mutation. (B) Alignments of tRNA's D-Loop of from carnivore and primates carrying the 3243 A>G mutation.

<https://doi.org/10.1371/journal.pone.0187862.g004>

of sequences divided by the length of the sequence, and is expressed as a percentage. Within the alignment the *Scolecormorphus vittatus* sequence differs from the human reference genome in 20 out of 75 nucleotides, with 13 of these polymorphisms occurring in the stem regions of the mt-tRNA molecule. The m.3253A>C variant is not present in this sequence. However, when the secondary structure was examined, the tRNA d-Stem contained four Watson-Crick pairs as opposed to the two that are found in human mt-tRNA-Leu(UUR). Two sequences from *Xenagama taylori* (NC\_008065.1 and DQ008215.1) the shield-tailed agama contained the 3243A>G. The two sequences were 73 bp in length and were identical. They were the only available sequences for this species. The sequences were highly divergent from the rCRS with a pairwise score of 56.16%. There were a number of changes in the secondary structure: the d-stem was only 3bp and the ac-stem contained an additional nucleotide and was 5bp in length.

The evidence gathered from current databases shows there are variants that have the potential to compensate for mutations known to cause disease in humans, and that the spread of these variants is such that there is unlikely to be a universal list of disease causing variants across different species, and even within a species.

### The distribution of predicted compensatory variants in other humans

Finally, we have investigated directly whether compensatory mutations are present in human mtDNA present in the public databases. The m.3243A>G, m.3253T>C and m.3254C>T variants were also studied within human mitochondrial sequences present on the public databases. Out of a total of 30,524 complete human mitochondrial sequences downloaded from GenBank, the m.3243A>G mutation was observed in 8 sequences. Unsurprisingly, 7 of these sequences were derived from sequences described as patient data. The 3243A>G mutation was also present in another human sequence KJ185483.1, which was derived from a population study, and there is no evidence to confirm whether or not this individual showed any disease symptoms. No potential compensatory mutations were observed in this sequence.

Distribution of the putative m.3253 T>C and m.3254C>T putative compensatory variants in humans was investigated, using the MITOMAP database of mitochondrial sequence variants derived from 29,867 complete human mitochondrial sequences present in GenBank in December 2014. Within these sequences, the m.3253T>C occurs in 7 sequences with a frequency of 0.02% and the m.3254C>T variant is present in 9 sequences with a frequency of 0.03%. The sequences that contained the m.3253T>C mutations were derived from 7 independent studies. Of these sequences, 3 belonged to the haplogroup M10a1, 3 belonged to haplogroup L2 and one belonged to haplogroup U6a3 (Table 7). Using information about haplogroup frequencies contained within MITOMAP the m.3253T>C variant was found to occur at the highest frequency within the L2 haplogroup. Although the numbers of human sequences bearing the m.3253 T>C and m.3254C>T are relatively low, their presence supports the need to consider sequence context when making decisions about the pathogenicity of variants in the human mt-tRNA's. It also suggests that, where known disease causing variant are seen in novel phylogenetic contexts with unexpected presentations, it would be of interest to conduct a number of the gold standard laboratory analysis [32] used to link genotype-phenotype to ensure the known disease causing mutation is causing disease in this context.

### Disease associated variants from all mitochondrial m-tRNA studied in vertebrates

We concluded the study by determining whether the results we had seen in mt-tRNA-Leu (UUR) were specific to this tRNA molecule or whether the trend would be replicated in other mt-RNAs. We selected a total of 246 disease associated variants in the remaining 21 m-tRNA

**Table 7. Sequences that carry the 3253 point mutation.**

Sequence ID	Haplogroup	Percentage within haplogroup
KP702293.1	U6a3	0.04
KF451676.1	M10a	0.11
KJ446421.1	M10a	0.11
JQ045037.1	L2b	0.35
JQ044890.1	L2b	0.35
JN857060.1	M10a	0.11
DQ112702.2	L2b	0.35

<https://doi.org/10.1371/journal.pone.0187862.t007>

molecules and looked for their presence or absence in the original panel of 33 species. Only 4% of these variants were not observed in any species and the remaining 235 mutations were observed in at least one species. This panel of mutations was compared to data published by Yarham et al., 62 of these mutations had been previously classified as definitely pathogenic in humans were studied further [32]. Of the definitely pathogenic mutations 47 were seen to be monomorphic for at least one species (Fig 5). These numbers differ to those in a prior report [30] as we applied a clinically validated scoring system [3, 7, 32] to ensure that we look at variants defiantly associated with clinically manifesting disease this algorithm was not available prior to the first report being published.

Most of these definitely pathogenic mutations were only monomorphic in less than 10 of the species studied. However, but there were a number of mutations which were found in the majority of species studied. Notably, G5703A, G14710A, A7472C, and G14724A, were observed in more than 75% of the species studied.

## Discussion

There have been many discussions about the causes of the variable presentation of mitochondrial disease. One possible much discussed explanation for this is the wide sequence context. The unique inheritance pattern of mtDNA results in the emergence of distinct maternal lineages or haplogroups. This has the effect that once a *compensatory* mutation, has occurred a subsequent normally pathogenic mutation can occur on the lineage and spread. This leads to the possibility that some lineages might be more robust than others to the consequences of mutation. Public databases now contain a wealth of sequence information from individuals, both human and other species. Prior studies that have considered this question have been limited by availability of sequence data, with some studies only using a single sequence form each of the species considered [31]. We have used a greatly expanded sequence database to investigate the prevalence and penetrance of variants located in the gene for mt-tRNA-Leu (UUR). This is a known location of disease-causing mutations in humans. Supporting the idea outlined above, we have found a number of widely recognized human mutations in other species at high frequencies, and have been able to suggest compensatory mutations.

### Observed secondary structure variation in tRNA-Leu(UUR) molecule in species with 3243A>G mutation

In this study, we show that m.3243A>G mutation occurs at high frequency within sequences from 6 species *Canis lupus familiaris*, *Mandrillus sphinx*, *Cercocebus atysatys*, *Cercocebus torquatus*, *Leptonychotes weddellii*, *Scolecormorphus vittatus* and *Xenagama taylori*. This suggests that the mutation is not pathogenic within these species. This is especially likely with samples taken from wild animals, which are under intense selection. In *Canis lupus familiaris*, the



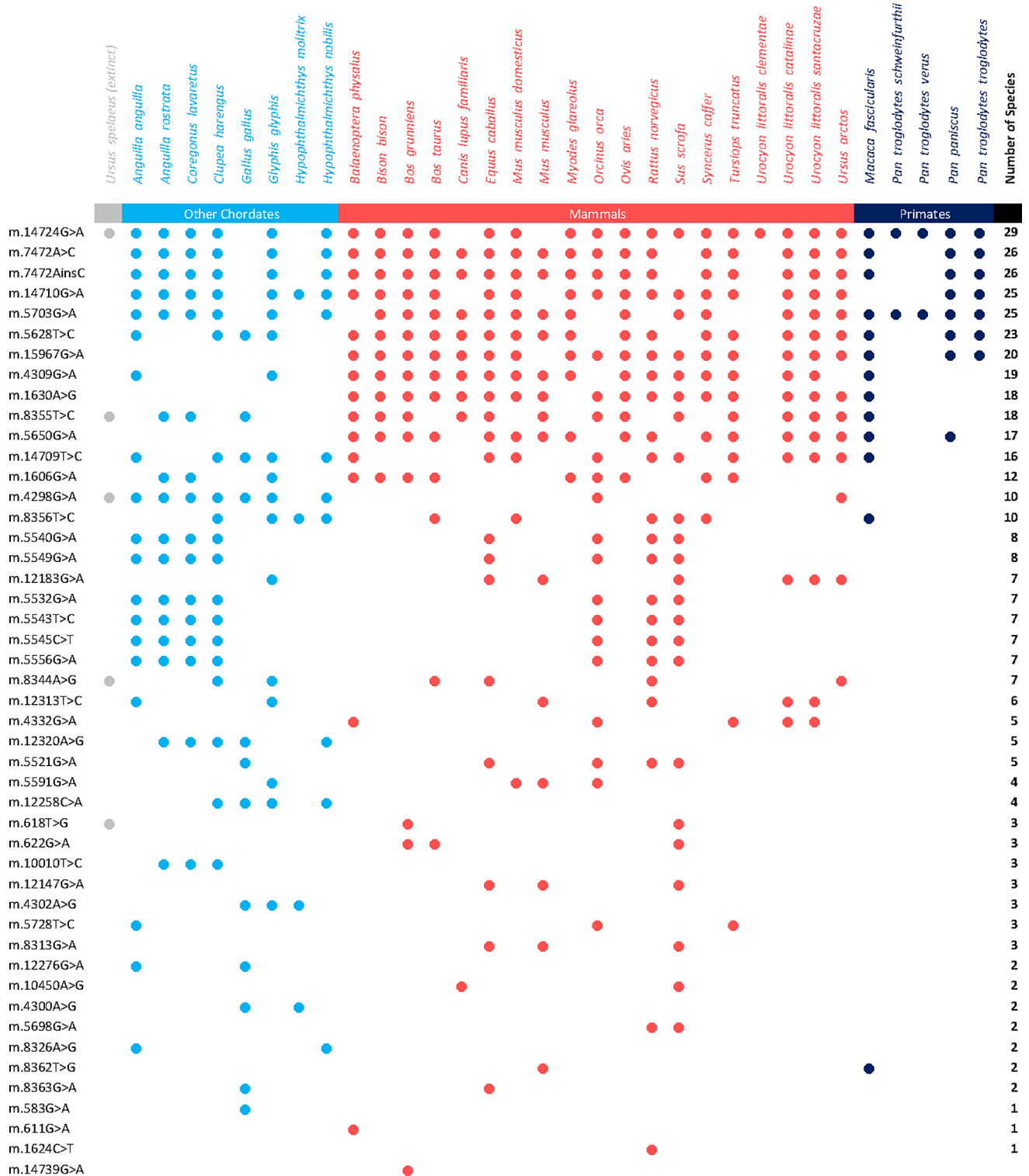


Fig 5. Mutations which are classified as definitely pathogenic in humans that are found to be fixed in other species.

<https://doi.org/10.1371/journal.pone.0187862.g005>

m.3243A>G mutation is present in approximately 15% of sequences. While only limited data is available for *Mandrillus sphinx*, *Cercocebus atys*, *Cercocebus torquatus*, *Leptonychotes*

*weddellii*, *Scolecormorphus vittatus* and *Xenagama taylori*, the variant is present in 100% of the available sequences. Although the m.3243A>G nucleotide is not present within a stem region of the mt-tRNA-Leu(UUR) molecule, it is involved in a tertiary interaction between bases 8, 14 and 21 of the mt-tRNA-Leu [21]. Disruption to the 3D structure of the molecule is believed to contribute to the pathogenic effect of the mutation [25].

In the current study, we found possible compensatory mutations, which alter the secondary structure of the mt-tRNA-Leu(UUR) molecule. These mutations may affect the way the mt-tRNA molecule folds and negate the negative effects of m.3243A>G. The compensatory mutations, observed in the d-stem of the mt-tRNA-Leu(UUR) molecule, were the m.3253A>C and m.3254C>T SNPs. They were present along with the m.3243A>G mutation in *Canis lupus familiaris*, *Mandrillus sphinx*, *Cercocebus atys*, *Cercocebus torquatus* and *Leptonychotes weddellii*. The d-stem of the human mitochondrial mt-tRNA-Leu(UUR) gene is 4 nucleotides long. It contains 2 Watson-Crick pairs, two nucleotides which are unpaired and a G-U wobble base pair in the position adjacent to the d-loop. The two compensatory mutations create a Watson-Crick base pair in the d-Stem of the mt-tRNA molecule in the position of the Wobble pair, and an extra Watson-Crick pair in the place of the two unpaired nucleotides. Changes to the secondary structure may suppress the pathogenic effect of the m.3243A>G mutation by altering the 3D shape of the mt-tRNA molecule in two ways: (1) Wobble pairs have been shown to change the 3D structure of mt-tRNA molecules due to the fact that G-U pairs form different glycosidic pairings to Watson-Crick pairs. This alters the angle of the bond with respect to the backbone of the molecule resulting in changes to the 3D shape of the molecule; (2) G-U pairs display conformational flexibility. These pairings react more sharply to sequence context than Watson-Crick pairs. The twisting in the molecule is influenced by the identity of the base pairs immediately adjacent to the wobble pair. As the m.3243A>G is directly adjacent to the wobble base pair, it possibly causes the molecule to twist in a way that prevents the correct 3D structure forming [44]. The tRNA-Leu(UUR) molecule of *Xenagama taylori* is significantly different from the rCRS and other mt-tRNA-Leu(UUR) molecules studied here. The reduced length position in d-stem means that the base identified as m.3243A>G is actually the second nucleotide in the d-loop rather than the first. Therefore, it is difficult to make direct comparisons about the structural importance of this base change.

The m.3254T>C mutation appears to be the most common allele and is prolific throughout the species studied, whereas the m.3253T>C mutation is confined to specific taxonomic groups. Within carnivores the m.3253T>C is seen in related species of wolf, dog and dog-like species, but no other closely related species. This indicates that the mutation was acquired after divergence of the canids. Variants which occur in regions that are highly conserved across species are most likely to be pathogenic and disease is most likely to be associated with rare variants. The wide distribution of these variants indicates that they are neutral mutations.

## Evidence that haplogroup background may suppress expression of disease

It has previously been reported that mitochondrial diseases from patients belonging to the African haplogroup L do not show the same phenotypic expression of disease as patients with European haplogroups [8, 9, 16]. One instance of the m.3243A>G mutation was detected in human from a population study where no link to disease was reported [45]. The m.3253T>C variant, which appears to be a compensatory mutation, is seen with the highest frequency in sequences from the African haplogroup L, and the Asian haplogroup M, was seen on 3 sequences from each haplogroup. Whereas the m.3254T>C is seen with highest frequency in sequences from the European haplogroup J, and was seen in five sequences from the sub-

haplogroup J1b. Secondary mutations which are specific to certain haplogroups have been shown to play a role in the phenotypic expression of mitochondrial disorders [17, 46]. It may be that some population variants are neutral on their own, but, when combined with a second mutation they can increase the severity of a disease [47].

The evidence presented here strongly indicates that changes to secondary structure of the mt-tRNA-Leu(UUR) molecule prevent pathogenic effects of 3243A>G. This supports the idea that sequence context (haplogroup background) is one of the important factors in the expression on mtDNA diseases. However, it must be remembered the phenotypes resulting from the m.3243A>G mutation are varied, and debate still exists as to its mechanism of action. The 3243A>G may impair methylation of mt-tRNA in position 10 of the molecule [21]. The presence of the mutation decreases the methylation of a uracil molecule in the first wobble position of the anticodon. This leads to a deficiency in the molecule in decoding UUG codons. Studying the methylation patterns of mt-tRNA molecules with additional mutations could provide information about how the mutation affects functionality. Neutral mutations elsewhere in other mitochondrial genes may suppress the 3243 A>G mutation [20]. In humans, the tRNA-Leu(UUR) gene is adjacent to the 16S ribosomal RNA gene. The m.3243A>G mutation within the tRNA-Leu(UUR) gene may interfere with a transcription and termination site for the 16S RNA molecule [20], leading to an accumulation of unprocessed RNA. In the *Canis lupus familiaris* and *Leptonychotes weddellii* sequences, including those possessing the m.3243A>G mutation, the locations of mt-tRNA-Leu(UUR) gene and the 16S ribosomal RNA gene are the same as the human mitochondrial genome and the transcription termination site is unchanged.

Previous research suggests Watson-Crick pairings are important in correct functioning of tRNA molecules and SNPs which break these bonds are more likely to be pathogenic [37, 48]. Evidence obtained in this study supports this hypothesis. There were three disease causing variants, with good evidence of pathogenicity within the stem regions of the tRNA molecule that were observed in 100% of sequences from non-humans, which strongly indicated that these mutations are not pathogenic within the species in question. In all instances, these mutations were accompanied by a compensatory change on the other arm of the stem of the tRNA molecule. This observation supports the previous hypothesis that it is not the SNP itself that is linked to disease but, rather, the disruption to the Watson-Crick pairs in these regions.

We identified a further 47 variants from other tRNA molecules which are classified as “definitely pathogenic” in 100% of sequences from other species. Again this provides strong evidence that variations within the molecules of these species prevent the pathogenic effects that have been seen in humans.

To conclude, evidence does exist to support the hypothesis that animals can be affected by mitochondrial disease, and as such they represent a valid system for considering the penetrance of mtDNA mutations and the importance of lineage context [29]. Notably, the results here strongly suggest that the m.3243A>G not seen to be present at high levels in humans in the absence of disease, is present as a non-pathogenic variant in other species, and that mtDNA sequence context is key to the modulation of the impact of this mutation. The importance of sequence context has previously been considered in the context of mutations causing Leber’s hereditary optic neuropathy (LHON) [17, 18]. As we sequence more, and investigate disease in more lineages [9], we are likely to find out more about the importance of sequence context in the expression of mtDNA variants. This knowledge will impact on the study of the role of mtDNA variants in clinical disease, and how we investigate any role of mtDNA variation in common complex diseases [49].

## Acknowledgments

Tom May for careful and thoughtful proof reading of the manuscript.

## Author Contributions

**Conceptualization:** Joanna L. Elson.

**Data curation:** Rachel A. Queen.

**Formal analysis:** Rachel A. Queen, Jannetta S. Steyn.

**Methodology:** Rachel A. Queen, Jannetta S. Steyn, Phillip Lord, Joanna L. Elson.

**Project administration:** Phillip Lord.

**Software:** Jannetta S. Steyn, Phillip Lord.

**Supervision:** Phillip Lord, Joanna L. Elson.

**Writing – original draft:** Rachel A. Queen.

**Writing – review & editing:** Phillip Lord, Joanna L. Elson.

## References

1. Tuppen HA, Blakely EL, Turnbull DM, Taylor RW. Mitochondrial DNA mutations and human disease. *Biochim Biophys Acta*. 2010; 1797(2):113–28. <https://doi.org/10.1016/j.bbabi.2009.09.005> Epub Sep 15. PMID: 19761752
2. Elson JL, Andrews RM, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Analysis of European mtDNAs for recombination. *Am J Hum Genet*. 2001; 68(1):145–53. Epub 2000 Dec 11. <https://doi.org/10.1086/316938> PMID: 11115380
3. Yarham JW, Elson JL, Blakely EL, McFarland R, Taylor RW. Mitochondrial tRNA mutations and disease. *Wiley Interdiscip Rev RNA*. 2010; 1(2):304–24. <https://doi.org/10.1002/wrna.27> Epub 2010 Jul 21. PMID: 21935892
4. Skladal DH, J. Thorburn D.R. Minimum birth prevalence of mitochondrial respiratory chain disorders in children. *Brain*. 2003; 126(Pt 8):1905–12. Epub 2003 May 21. <https://doi.org/10.1093/brain/awg170> PMID: 12805096
5. Gorman GS, Schaefer AM, Ng Y, Gomez N, Blakely EL, Alston CL, et al. Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease. *Ann Neurol*. 2015; 77(5):753–9. <https://doi.org/10.1002/ana.24362> Epub 2015 Mar 28. PMID: 25652200
6. Martikainen MH, Rönnemaa T, Majamaa K. Prevalence of mitochondrial diabetes in southwestern Finland: a molecular epidemiological study. *Acta Diabetol*. 2013; 50(5):737–41. <https://doi.org/10.1007/s00592-012-0393-2> Epub 2012 Apr 11. PMID: 22492248
7. Yarham JW, McFarland R, Taylor RW, Elson JL. A proposed consensus panel of organisms for determining evolutionary conservation of mt-tRNA point mutations. *Mitochondrion*. 2012; 12(5):533–8. <https://doi.org/10.1016/j.mito.2012.06.009> Epub Jul 7. PMID: 22781547
8. van der Westhuizen FH, Sinxadi PZ, Dandara C, Smuts I, Riordan G, Meldau S, et al. Understanding the Implications of Mitochondrial DNA Variation in the Health of Black Southern African Populations: The 2014 Workshop. *Hum Mutat*. 2015; 36(5):569–71. <https://doi.org/10.1002/humu.22789> PMID: 25764011
9. van der Walt EM, Smuts I, Taylor RW, Elson JL, Turnbull DM, Louw R, et al. Characterization of mtDNA variation in a cohort of South African paediatric patients with mitochondrial disease. *Eur J Hum Genet*. 2012; 20(6):650–6. <https://doi.org/10.1038/ejhg.2011.262> Epub 2 Jan 18. PMID: 22258525
10. Howell N, Elson JL, C. H, Turnbull DM. Relative rates of evolution in the coding and control regions of African mtDNAs. *Mol Biol Evol*. 2007; 24(10):2213–21. Epub 007 Jul 21. <https://doi.org/10.1093/molbev/msm147> PMID: 17642471
11. Smith PM, Elson JL, Greaves LC, Wortmann SB, Rodenburg RJ, Lightowlers RN, et al. The role of the mitochondrial ribosome in human disease: searching for mutations in 12S mitochondrial rRNA with high disruptive potential. *Hum Mol Genet*. 2014; 23(4):949–67. <https://doi.org/10.1093/hmg/ddt490> Epub 2013 Oct 2. PMID: 24092330

12. Lott MT, Leipzig JN, Derbeneva O, Xie HM, Chalkia D, Sarmady M, et al. mtDNA Variation and Analysis Using MITOMAP and MITOMASTER. *Curr Protoc Bioinformatics*. 2013; 1(123):1.23.1–1.6.
13. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat*. 2009; 30(2):E386–94. <https://doi.org/10.1002/humu.20921> PMID: 18853457
14. Pütz J, Dupuis B, Sissler M, Florentz C. Mamit-tRNA, a database of mammalian mitochondrial tRNA primary and secondary structures. *RNA*. 2007; 13(8):1184–90. Epub 2007 Jun 21. <https://doi.org/10.1261/rna.588407> PMID: 17585048
15. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet*. 1999; 23(2):147. <https://doi.org/10.1038/13779> PMID: 10508508
16. Smuts I, Louw R, du Toit H, Klopper B, Mienie LJ, van der Westhuizen FH. An overview of a cohort of South African patients with mitochondrial disorders. *J Inherit Metab Dis*. 2010; 33(Suppl 3):S95–104. <https://doi.org/10.1007/s10545-009-9031-8> Epub 2010 Feb 5. PMID: 20135231
17. Hudson G, Carelli V, Spruijt L, Gerards M, Mowbray C, Achilli A, et al. Clinical expression of Leber hereditary optic neuropathy is affected by the mitochondrial DNA-haplogroup background. *Am J Hum Genet*. 2007; 81(2):228–33. Epub 2007 Jun 4. <https://doi.org/10.1086/519394> PMID: 17668373
18. Cai W, Fu Q, Zhou X, Qu J, Tong Y, Guan MX. Mitochondrial variants may influence the phenotypic manifestation of Leber's hereditary optic neuropathy-associated ND4 G11778A mutation. *J Genet Genomics*. 2008; 35(11):649–55. [https://doi.org/10.1016/S1673-8527\(08\)60086-7](https://doi.org/10.1016/S1673-8527(08)60086-7) PMID: 19022198
19. Elson JL, Swalwell H, Blakely EL, McFarland R, Taylor RW, Turnbull DM. Pathogenic mitochondrial tRNA mutations—which mutations are inherited and why? *Hum Mutat*. 2009; 30(11):E984–92. <https://doi.org/10.1002/humu.21113> PMID: 19718780
20. El Meziane A, Lehtinen SK, Hance N, Nijtmans LG, Dunbar D, Holt IJ, et al. A tRNA suppressor mutation in human mitochondria. *Nat Genet*. 1998; 18(4):350–3. <https://doi.org/10.1038/ng0498-350> PMID: 9537417
21. Helm M, Brulé H, Friede D, Giegé R, Pütz D, Florentz C. Search for characteristic structural features of mammalian mitochondrial tRNAs. *RNA*. 2000; 6(10):1356–79. PMID: 11073213
22. Uusimaa J, Moilanen JS, Vainionpää L, Tapanainen P, Lindholm P, Nuutinen M, et al. Prevalence, segregation, and phenotype of the mitochondrial DNA 3243A>G mutation in children. *Ann Neurol*. 2007; 62(3):278–87. <https://doi.org/10.1002/ana.21196> PMID: 17823937
23. Manwaring N, Jones MM, Wang JJ, Rochtchina E, Howard C, Mitchell P, et al. Population prevalence of the MELAS A3243G mutation. *Mitochondrion*. 2007; 7(3):230–3. Epub 2007 Jan 8. <https://doi.org/10.1016/j.mito.2006.12.004> PMID: 17300999
24. Goto Y, Nonaka I, Horai S. A mutation in the tRNA(Leu)(UUR) gene associated with the MELAS subgroup of mitochondrial encephalomyopathies. *Nature*. 1990; 348(6302):651–3. <https://doi.org/10.1038/348651a0> PMID: 2102678
25. Roy MD, Wittenhagen LM, Kelley SO. Structural probing of a pathogenic tRNA dimer. *RNA*. 2005; 11(3):254–60. <https://doi.org/10.1261/rna.7143305> PMID: 15701731
26. Hess JF, Parisi MA, Bennett JL, Clayton DA. Impairment of mitochondrial transcription termination by a point mutation associated with the MELAS subgroup of mitochondrial encephalomyopathies. *Nature*. 1991; 351(6323):236–9. <https://doi.org/10.1038/351236a0> PMID: 1755869
27. Kirino Y, Yasukawa T, Ohta S, Akira S, Ishihara K, Watanabe K, et al. Codon-specific translational defect caused by a wobble modification deficiency in mutant tRNA from a human mitochondrial disease. *Proc Natl Acad Sci U S A*. 2004; 101(42):15070–5. Epub 2004 Oct 11. <https://doi.org/10.1073/pnas.0405173101> PMID: 15477592
28. van der Westhuizen FH, Sinxadi PZ, Dandara C, Smuts I, Riordan G, Meldau S, et al. Understanding the Implications of Mitochondrial DNA Variation in the Health of Black Southern African Populations: The 2014 Workshop. *Hum Mutat*. 2015; 36(5):569–71. <https://doi.org/10.1002/humu.22789> PMID: 25764011
29. Baranowska I, Jäderlund KH, Nennesmo I, Holmqvist E, Heidrich N, Larsson NG, et al. Sensory ataxic neuropathy in golden retriever dogs is caused by a deletion in the mitochondrial tRNA<sup>Tyr</sup> gene. *PLoS Genet*. 2009; 5(5):e1000499. <https://doi.org/10.1371/journal.pgen.1000499> Epub 2009 May 29. PMID: 19492087
30. Kern AD, Kondrashov FA. Mechanisms and convergence of compensatory evolution in mammalian mitochondrial tRNAs. *Nat Genet*. 2004; 36(11):1207–12. Epub 2004 Oct 24. <https://doi.org/10.1038/ng1451> PMID: 15502829
31. de Magalhaes JP. Human disease-associated mitochondrial mutations fixed in nonhuman primates. *J Mol Evol*. 2005; 61(4):491–7. Epub 2005 Aug 25. <https://doi.org/10.1007/s00239-004-0258-6> PMID: 16132471

32. Yarham JW, Al-Dosary M, Blakely EL, Alston CL, Taylor RW, Elson JL, et al. A comparative analysis approach to determining the pathogenicity of mitochondrial tRNA mutations. *Hum Mutat.* 2011; 32(11):1319–25. <https://doi.org/10.1002/humu.21575> Epub 2011 Sep 19. PMID: 21882289
33. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics.* 2009; 25(11):1422–3. Epub 2009 Mar 20. <https://doi.org/10.1093/bioinformatics/btp163> PMID: 19304878
34. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007; 23(21):2947–8. Epub 007 Sep 10. <https://doi.org/10.1093/bioinformatics/btm404> PMID: 17846036
35. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 1997; 25(5):955–64. PMID: 9023104
36. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol.* 1999; 16(1):37–48. PMID: 10331250
37. McFarland R, Elson JL, Taylor RW, Howell N, Turnbull DM. Assigning pathogenicity to mitochondrial tRNA mutations: when "definitely maybe" is not good enough. *Trends Genet.* 2004; 20(12):591–6. <https://doi.org/10.1016/j.tig.2004.09.014> PMID: 15522452
38. Sweeney MG, Bunday S, Brockington M, Poulton KR, Winer JB, Harding AE. Mitochondrial myopathy associated with sudden death in young adults and a novel mutation in the mitochondrial DNA leucine transfer RNA(UUR) gene. *Q J Med.* 1993; 86(11):709–13. PMID: 8265770
39. Houshmand M, Larsson NG, Oldfors A, Tulinius M, Holme E. Fatal mitochondrial myopathy, lactic acidosis, and complex I deficiency associated with a heteroplasmic A→G mutation at position 3251 in the mitochondrial tRNA<sup>Leu</sup>(UUR) gene. *Hum Genet.* 1996; 97(3):269–73. PMID: 8786060
40. Herrstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, et al. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet.* 2002; 70(5):1152–71. Epub 2002 Apr 5. <https://doi.org/10.1086/339933> PMID: 11938495
41. Mimaki M, Hatakeyama H, Ichiyama T, Isumi H, Furukawa S, Akasaka M, et al. Different effects of novel mtDNA G3242A and G3244A base changes adjacent to a common A3243G mutation in patients with mitochondrial disorders. *Mitochondrion.* 2009; 9(2):115–22. <https://doi.org/10.1016/j.mito.2009.01.005> Epub Jan 21. PMID: 19460299
42. Seneca S, Verhelst H, De Meirleir L, Meire F, Ceuterick-De Groote C, Lissens W, et al. A new mitochondrial point mutation in the transfer RNA(Leu) gene in a patient with a clinical phenotype resembling Kearns-Sayre syndrome. *Arch Neurol.* 2001; 58(7):1113–8. PMID: 11448301
43. Yakubovskaya MG, Belyakova AA, Gasanova VK, Belitsky GA, Dolinnaya NG. Comparative reactivity of mismatched and unpaired bases in relation to their type and surroundings. Chemical cleavage of DNA mismatches in mutation detection analysis. *Biochimie.* 2010; 92(7):762–71. <https://doi.org/10.1016/j.biochi.2010.02.016> Epub Feb 18. PMID: 20171258
44. Varani G, McClain WH. The G x U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep.* 2000; 1(1):18–23. <https://doi.org/10.1093/embo-reports/kvd001> PMID: 11256617
45. Barbieri C, Güldemann T, Naumann C, Gerlach L, Berthold F, Nakagawa H, et al. Unraveling the complex maternal history of Southern African Khoisan populations. *Am J Phys Anthropol.* 2014; 153(3):435–48. <https://doi.org/10.1002/ajpa.22441> Epub 2013 Dec 9. PMID: 24323467
46. Man PY, Howell N, Mackey DA, Nørby S, Rosenberg T, Turnbull DM, et al. Mitochondrial DNA haplogroup distribution within Leber hereditary optic neuropathy pedigrees. *J Med Genet.* 2004; 41(4):e41. <https://doi.org/10.1136/jmg.2003.011247> PMID: 15060117
47. Chinnery PF, Howell N, Andrews RM, Turnbull DM. Mitochondrial DNA analysis: polymorphisms and pathogenicity. *J Med Genet.* 1999; 36(7):505–10. PMID: 10424809
48. McFarland R, Taylor RW, Elson JL, Lightowlers RN, Turnbull DM, Howell N. Proving pathogenicity: when evolution is not enough. *Am J Med Genet A.* 2004; 131(1):107–8; author reply 9–10. <https://doi.org/10.1002/ajmg.a.30318> PMID: 15384096
49. Salas A, Elson JL. Mitochondrial DNA as a risk factor for false positives in case-control association studies. *J Genet Genomics.* 2015; 42(4):169–72. <https://doi.org/10.1016/j.jgg.2015.03.002> Epub Mar 17. PMID: 25953355