Molecular Biology and Evolution, 2025, **42**, 1–15
https://doi.org/10.1093/molbev/msae261
Advance access publication 4 February 2025
Discoveries

OXFORD MBE

Deciphering the Foundations of Mitochondrial Mutational Spectra: Replication-Driven and Damage-Induced Signatures Across Chordate Classes

Dmitrii Iliushchenko [1,†] Bogdan Efimenko [1,†] Alina G. Mikhailova [1,†] Victor Shamanskiy [1] Murat K. Saparbaev [2] Bakhyt T. Matkarimov [3,4] Ilya Mazunin [5,6] Alexandr Voronka [1] Dmitry Knorre [7] Wolfram S. Kunz [8] Philipp Kapranov [9] Stepan Denisov [10] Jacques Fellay [11] Konstantin Khrapko [12] Konstantin Gunbin [1,13] Konstantin Popadin [1,11,*]

[1]Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal University, Kaliningrad, Russian Federation
[2]Groupe "Mechanisms of DNA Repair and Carcinogenesis", CNRS UMR9019, Gustave Roussy Cancer Campus, Université Paris-Saclay, Villejuif, France
[3]National Laboratory Astana, Nazarbayev University, Astana, Kazakhstan
[4]Faculty of Information Technologies, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
[5]Department of Biology and Genetics, Petrovsky Medical University, Moscow, Russian Federation
[6]Research Centre for Medical Genetics, Moscow, Russian Federation
[7]Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russian Federation
[8]Department of Epileptology and Institute of Experimental Epileptology and Cognition Research, University Bonn Medical Center, Bonn, Germany
[9]School of Life Sciences, Xiamen University, Xiamen, China
[10]Faculty of Biology, Medicine and Health, School of Biological Sciences, The University of Manchester, Manchester, UK
[11]School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[12]Department of Biology, Northeastern University, Boston, MA, USA
[13]Institute of Molecular and Cellular Biology SB RAS, Novosibirsk, Russian Federation

[†]Equal contribution.
***Corresponding author:** E-mail: konstantinpopadin@gmail.com.
Associate editor: Weiwei Zhai

Abstract

Mitochondrial DNA (mtDNA) mutagenesis remains poorly understood despite its crucial role in disease, aging, and evolutionary tracing. In this study, we reconstructed a comprehensive 192-component mtDNA mutational spectrum for chordates by analyzing 118,397 synonymous mutations in the CytB gene across 1,697 species and five classes. This analysis revealed three primary forces shaping mtDNA mutagenesis: (i) symmetrical, replication-driven errors by mitochondrial polymerase (POLG), resulting in C > T and A > G mutations that are highly conserved across classes; (ii) asymmetrical, damage-driven C > T mutations on the single-stranded heavy strand with clock-like dynamics; and (iii) asymmetrical A > G mutations on the heavy strand, with dynamics suggesting sensitivity to oxidative damage. The third component, sensitive to oxidative damage, positions mtDNA mutagenesis as a promising marker for metabolic and physiological processes across various classes, species, organisms, tissues, and cells. The deconvolution of the mutational spectra into mutational signatures uncovered deficiencies in both base excision repair (BER) and mismatch repair (MMR) pathways. Further analysis of mutation hotspots, abasic sites, and mutational asymmetries underscores the critical role of single-stranded DNA damage (components ii and iii), which, uncorrected due to BER and MMR deficiencies, contributes roughly as many mutations as POLG-induced errors (component i).

Keywords: mitochondria, mutational spectrum, phylogenetics, neutral evolution

Introduction

DNA mutations can be a result of either replication errors or damage followed by incorrect repair (Chatterjee and Walker 2017), with a range of mutagens responsible for these changes (Chintalapati and Moorjani 2020). Reconstruction of a mutational spectrum, with 96 or 192 components, allows decomposition into distinct mutational signatures, helping us to trace the effects of various mutagens (Alexandrov et al. 2020). In the human nuclear genome, both germline and somatic spectra, including those from cancers, have been instrumental in reconstructing and deconvoluting these mutational signatures, leading to major breakthroughs in both fundamental (Seplyarskiy et al. 2019), and applied (Koh et al. 2021) research.

Despite significant advancements in comprehending the mutagenesis of the human nuclear genome, the mitochondrial genome is less well characterized, yet it plays a vital role in numerous human diseases (Taylor and Turnbull 2005) and aging (Kalous and Drahota 1996). The mutagenesis of the mitochondrial genome is mysterious: its mutation rate is a hundred times faster than in the nuclear genome, yet it remains poorly understood, since its spectra show none of the expected mutational

Received: April 5, 2024. **Revised:** November 8, 2024. **Accepted:** December 4, 2024

signatures of reactive oxygen species (ROS), UV light, nor tobacco smoke in associated cancer data (Yuan et al. 2020). Furthermore, the mechanisms of mitochondrial DNA (mtDNA) replication and repair have been the subject of ongoing debate, and recent detailed analysis of the mitochondrial mutational spectrum has offered important insights into these processes (Sanchez-Contreras et al. 2021). On an evolutionary scale, understanding mtDNA mutagenesis and the implementation of mtDNA spectrum-aware approaches are crucial for accurate phylogenetic inferences (Mikhailova et al. 2022), for uncovering variations in selection processes between species (Alexandrov et al. 2020; Boot et al. 2022), as well as for understanding the dynamics of deleterious human mtDNA variants.

The reconstruction of a detailed mutational spectrum of single-base substitutions (SBS) for each species demands extensive data. Traditionally, comparative-species analyses employ the transitions to transversions ratio as a basic characteristic of the mutational spectrum (Belle et al. 2005). The availability of more sequence data has allowed researchers to use more complex 6-component (focusing only on pyrimidines in Watson–Crick base pairs: C > A, C > G, C > T, T > A, T > C, T > G, under the assumption of symmetrical mutagenesis on complementary strands) (Dillon et al. 2017; Chu et al. 2018; Saclier et al. 2020) and 12-component (doubling the 6-component spectrum to account for asymmetrical mutagenesis on complementary strands) (Mikhailova et al. 2022) spectra in analyses for comparative studies. However, more comprehensive 96-component (expanding the 6-component spectrum with an inclusion of the nucleotide context) and 192-component (a doubled version of the 96-component spectrum, assuming asymmetry between complementary strands) spectra, which consider adjacent nucleotide context, have been limited to extensively studied species such as humans (Yuan et al. 2020) and SARS-CoV-2 (Yi et al. 2021). For rarely sequenced, nonmodel taxa, constructing these in-depth spectra is challenging.

Here, focusing on chordates, we overcame this limitation by integrating rare, species-specific mtDNA polymorphisms into a comprehensive spectrum that is representative of all *Chordata*. Reconstructing 118,397 synonymous polymorphic mutations from CytB sequences of 1697 chordate species, we compiled a 192-component mtDNA spectrum. This extensive spectrum enables the exploration of critical questions, related to damage, and replication-driven mechanisms of mtDNA mutagenesis, their etiologies, and dynamics.

Results

Integral 192-component mtDNA Mutational Spectrum of Chordates

The 192-component mutational spectrum for four mitochondrial protein-coding genes—CytB, ND2, CO1, and CO3—scaled and normalized to trinucleotide contexts, was derived by utilizing the NeMu pipeline (Efimenko et al. 2024), which processed tens of thousands of synonymous mutations reconstructed from polymorphic sequences of more than a thousand chordate species (supplementary table S1, Supplementary Material online). The resulting spectra of all four genes highlight a significant prevalence of asymmetric transitions, with $C_H > T_H$ and $A_H > G_H$ being the most frequent types (Fig. 1 and supplementary fig. S1, Supplementary Material online). This mutational pattern reflects findings observed in mammalian germline mutations (Mikhailova et al. 2022), somatic mutations in human cancers (Yuan et al. 2020), and healthy tissues (Sanchez-Contreras et al. 2023). It is important to note that although we refer to our mutational spectra as having 192 components, only 160 of them are effectively used in our approach because we focus on synonymous substitutions (see missing bins in Fig. 1c and explanation in supplementary fig. S2, Supplementary Material online).

The mutational spectra of all four genes appear very similar (see supplementary fig. S1, Supplementary Material online), confirmed by a high cosine similarity between them (around 0.9; see supplementary fig. S3 and Methods, Supplementary Material online). This suggests consistent mutational processes across the minor (ND2) and major (CO1, CO3, and CytB) mtDNA arcs, as well as along the different positions within the major arc (CO1 at the beginning, CO3 in the middle, and CytB at the end). Given this high similarity, our subsequent analyses will focus exclusively on CytB, a gene represented by a significantly higher number of mutations and species in our dataset: 118,397 synonymous mutations across 1,697 chordate species (see supplementary tables S1 and S2, Supplementary Material online). To further demonstrate the usability of CytB for analyzing comprehensive mtDNA mutational spectra across all chordates, we confirmed: (i) its highly stable location relative to the origin of replication on the light strand (oriL), which indicates minimal variation during the time spent single-stranded (TSSS) (see supplementary fig. S4, Supplementary Material online); (ii) a consistently high fraction of transitions in the mutational spectra across various species, exceeding 0.87 for each taxonomic group (see supplementary table S3, Supplementary Material online), serving as a hallmark of high-quality mtDNA mutational spectra (see Fig. 1c in Yuan et al. (2020) based on comprehensive analysis of human cancer mtDNA); (iii) a high level of similarity in the expected part of mutational spectra, which is used for normalization and is based on trinucleotides with synonymous mutations at the center (i.e. is influenced by the usage of codons and the colocalization of neighboring codons, see Materials and Methods and supplementary figs. S5 and S6 and table S4, Supplementary Material online). Altogether, we consider CytB as a widely sequenced gene with consistent key properties (stable gene location, high fraction of transitions as a quality control, similar expected mutational spectrum across classes), which allows it to serve as a hallmark of the mtDNA spectrum.

Variation in mtDNA Mutational Spectra-Driven by Life-History-Associated Damage Components

To assess variation in CytB-derived 192-component mutational spectra between chordate species, we calculated pairwise cosine similarities between all species and found no evidence of the expected clustering of species-specific spectra within classes (supplementary fig. S7a, Supplementary Material online). Technically speaking, this result may be explained by a high level of uncertainty in the 192-component spectrum at the species level due to the sparsity of the compared spectra (a high fraction of the bins in the spectrum are zeros, while many nonzero bins are based on a small number of observed events and thus exhibit high variance). Biologically speaking, the lack of similarities within class-specific spectra can indicate either extreme stability of mtDNA mutational spectra across classes or some evolutionary changes not linked to phylogeny (Chintalapati and Moorjani 2020). However, changes in mtDNA mutational spectra between different species have been documented before (Belle et al. 2005; Chintalapati and

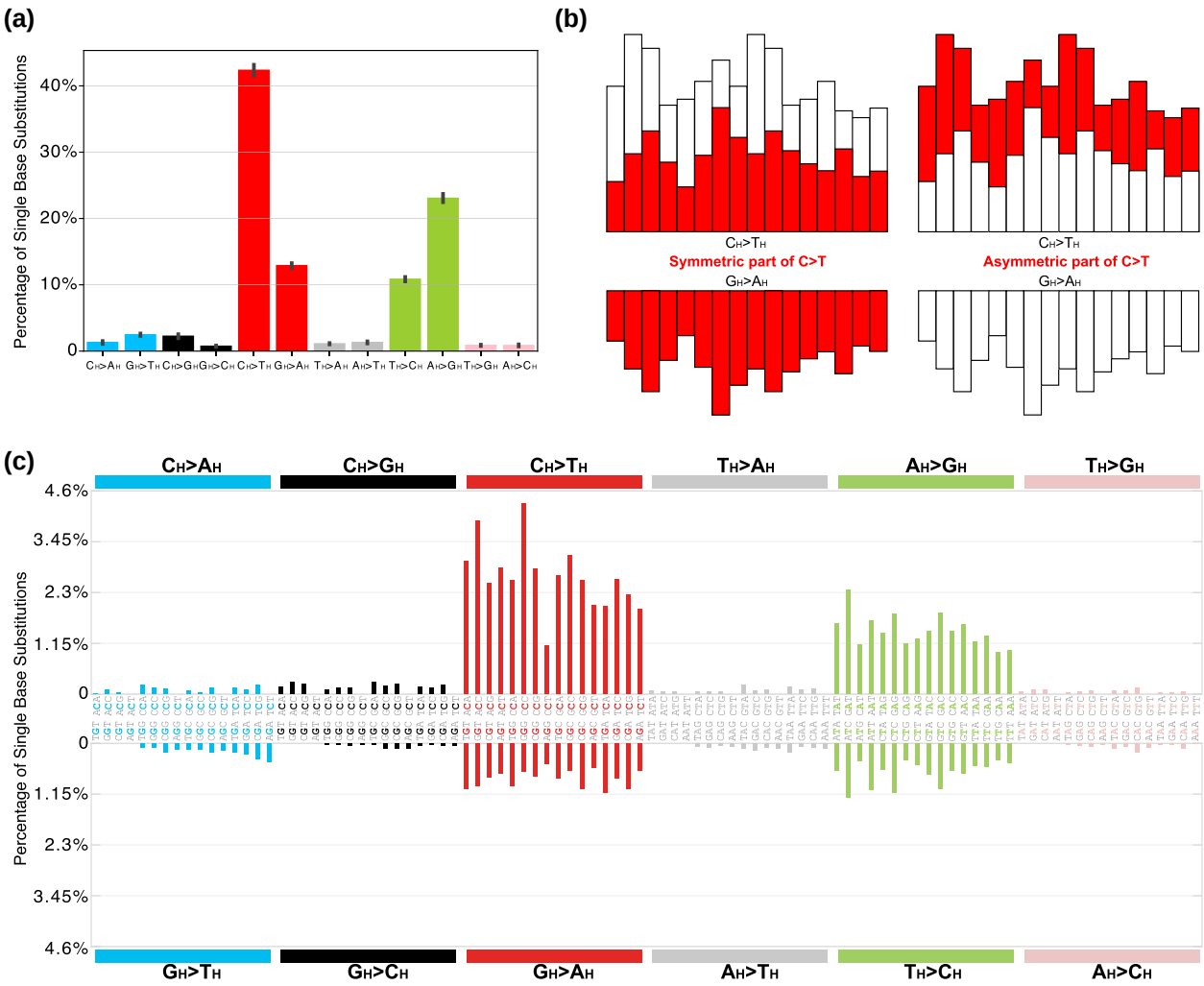

Fig. 1. Integral mtDNA mutational spectrum of chordate species based on CytB. a) An average 12-component mutational spectrum (N=1,697 species); b) A scheme, visualizing symmetrical and asymmetrical parts of a given mutation (e.g. C > T); c) integral 192-component mutational spectrum (N= 1,697 species). Missing 32 bins (4 different contexts in each of 8 transversions) are explained by the absence of possible synonymous substitutions in such contexts due to genetic code properties (see details in supplementary fig. S2, Supplementary Material online).

Moorjani 2020). Such changes can be explained either by variable replication/reparation-driven components (if POLG and basic reparation systems of mtDNA vary strongly between species) or by variable damage-driven components (if mtDNA damage is affected by the metabolism-related traits that can be highly variable even among close species). Taking into account the high evolutionary conservation of chordate POLG (Oliveira et al. 2015; Khan et al. 2020; Czernecki et al. 2023) and rather variable levels of aerobic metabolism in different chordates (Gavrilov et al. 2022), we propose that the damage effect may be the main reason for the observed variation.

Additionally, to rule out the technical scenario, we calculated pairwise cosine similarities across all chordate species using the 12-component CytB-derived mutational spectrum (supplementary fig. S7b, Supplementary Material online). With the 12-component spectra, we expect minimal impact from data sparsity, allowing us to disregard the technical scenario mentioned above. Despite utilizing enriched data to derive species-specific 12-component spectra, the lack of class-specific clustering observed supports a biological explanation. Contrasting replication-driven with damage-driven mechanisms at the level of 12-component spectra, we suggest the damage-driven one as the best working hypothesis, due to the expected conservative nature of replication and repair of mtDNA and previously demonstrated associations of life-history traits with 12-component mtDNA mutational spectra. Indeed, it has been shown before that $A_H > G_H$ mutation—one of the most asymmetrical and thus damage-driven mutations in mtDNA—correlates with life-history traits across species, factoring in phylogenetic inertia: longevity in mammals (Mikhailova et al. 2022), water temperature in fishes (Mikhailova et al. 2023), and average body temperature in five chordate classes (Mikhailova et al. 2023). Consequently, we suggest that both the 12- and 192-component mtDNA mutational spectra are significantly shaped by life-history-related damage factors in mtDNA.

To add more lines of evidence to the damage-effect hypothesis, we performed class-specific analyses, expecting that birds (Aves), a class characterized by the highest level of basal metabolic rate on average among chordates (Gavrilov et al. 2022), would show the most divergent patterns of the spectrum. We reconstructed 192-component mutational spectra for each of the five chordate classes (supplementary fig. S8 and table S5, Supplementary Material online), revealing notable similarities across them and to the overall mtDNA mutational pattern

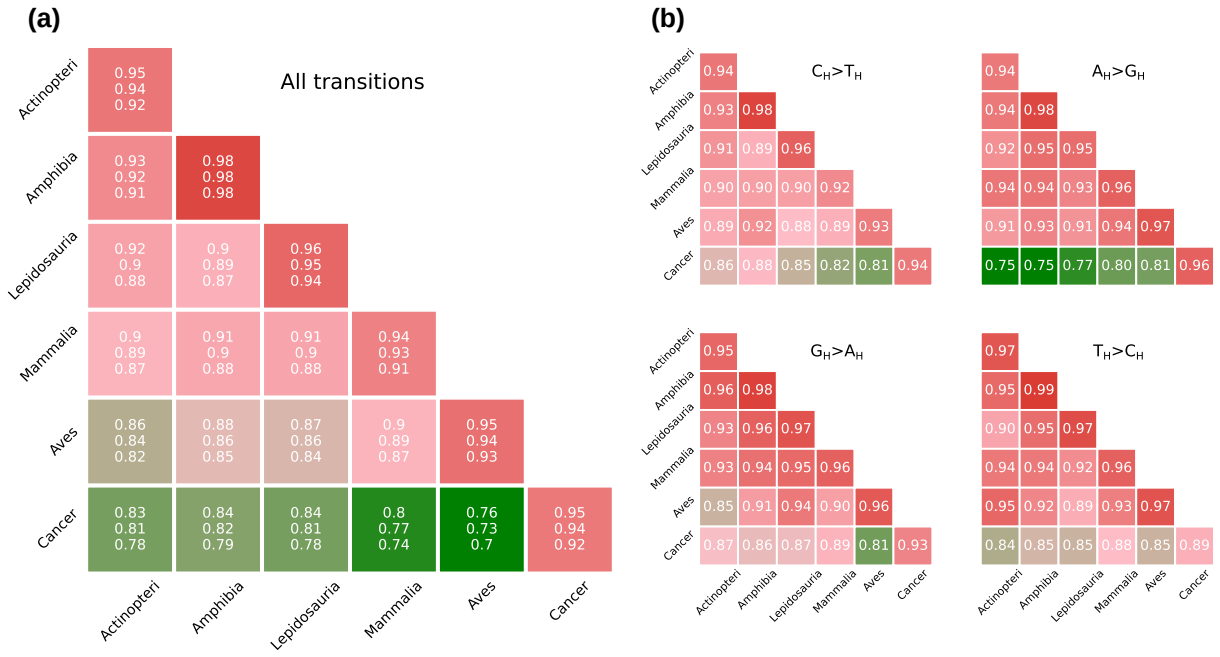

Fig. 2. Pairwise comparisons of mutational spectra of five chordate classes and human cancer samples. Cosine similarity of Jackknife samples underpinned the analysis (refer to Materials and Methods for details). a) The similarity based on mutational spectra of transitions, comprising 64 components. Each box contains the quartiles (Q1, median Q2, and Q3), representing the similarity distributions across samples. b) The similarity is grounded on distinct transition types within the mutational spectra, featuring 16 components for each transition. Numbers in boxes indicate medians. Notably, the $G_H > A_H$ spectra are more conserved than $C_H > T_H$ ones (see text for details). See supplementary fig. S9, Supplementary Material online, for the similar analyses based on all types of substitutions (192 components) and transversions only (128 components).

(Fig. 1c), suggesting a conserved mutational process within chordates. Further analysis, involving median pairwise cosine similarities between different classes, confirmed that birds (Aves) have the most distinct mtDNA mutational spectrum among all classes (Materials and Methods; Fig. 2a), with the lowest cosine similarities to all other chordate classes. This is in line with a hypothesis that the mtDNA mutational spectrum is shaped by metabolism-associated chemical damage (Almatarneh et al. 2006).

To further test the potential damage effect, we analyzed somatic mtDNA mutations from human cancers, which are often hypoxic (Bhandari et al. 2019) and thus may have less oxidative damage (Ericsson et al. 2015). Conducting pairwise comparisons between 192-component mutational spectra of human cancers (Yuan et al. 2020) and 192-component mutational spectra of five chordate classes, we indeed revealed the least cosine similarity between birds and cancers (Fig. 2a; supplementary fig. S9, Supplementary Material online). This indicates their strong deviations between them, presumably due to the lowest oxidative damage in cancers and the highest oxidative damage in birds (Trimarchi et al. 2000; Sugimura et al. 2012). Altogether, we propose that, due to the differences in the level of aerobic metabolism, all chordate species and classes, as well as human cancers, are characterized by different levels of mtDNA damage, which in turn shape the mtDNA mutational spectra.

Patterns Within mtDNA Spectrum: Conserved Symmetrical Replication-Driven Components, Variable Asymmetrical Damage-Driven Components, hot Mutational Motifs

To distinguish stable and variable patterns within the CytB-derived 192-component mutational spectra across classes, we analyzed transitions and transversions separately. Transitions showed consistently high similarities among chordate classes and human cancers (Fig. 2a), aligning with the full spectrum trends (supplementary fig. S9a, Supplementary Material online). Conversely, transversions showed lower similarities (supplementary fig. S9b, Supplementary Material online), likely due to their stochastic occurrence and rarity in our dataset (supplementary fig. S10, Supplementary Material online). Subsequently, we focused our analysis on each transition type separately.

Cosine similarities across classes were uniformly high for each of the four transition types (Fig. 2b), yet interesting variations were observed. Notably, $G_H > A_H$ exhibited higher similarity between classes than $C_H > T_H$, despite being complementary equivalents. This indicates that symmetrical C > T mutations on double-stranded DNA (dsDNA) (approximated by $G_H > A_H$, Fig. 1b) are more conserved compared with the asymmetrical part of C > T mutations in single-stranded DNA (ssDNA) (approximated by the difference between $C_H > T_H$ and $G_H > A_H$, Fig. 1b). We suggest that the conserved pattern of symmetrical C > T substitutions (equal to $G_H > A_H$) primarily results from internal replication errors, introduced by POLG. Conversely, the less conserved pattern of asymmetrical C > T mutations ($C_H > T_H$ minus $G_H > A_H$) on ssDNA, particularly divergent between cancer and chordate classes, may be influenced by chemical damage, including spontaneous deamination or oxidation.

Taking into account that asymmetrical $C_H > T_H$ (i.e. $C_H > T_H$ minus $G_H > A_H$) is most likely associated with chemical damage such as spontaneous deamination or oxidation at ssDNA (Sanchez-Contreras et al. 2021), we explored the profile of the $C_H > T_H$ substitutions in detail. A yeast experiment with ssDNA and ROS (oxygen peroxide and paraquat) demonstrated that cCc > cTc mutations are a primary target of

oxidative damage in ssDNA (Degtyareva et al. 2019). Notably, completely in line with these experimental results, we consistently identified cCc > cTc as the most common motif for $C_H > T_H$ substitutions in all chordate classes (supplementary fig. S11a, Supplementary Material online). Therefore, the cCc > cTc mutation arises not only from spontaneous deamination but also includes contributions from oxidative or other types of damage to ssDNA. Conversely, cancer data again demonstrate a distinct pattern: asymmetrical $C_H > T_H$ substitutions predominantly appear in the nCg context rather than cCc (supplementary fig. S11c, Supplementary Material online), suggesting the effects of potentially lower levels of oxidative damage (Ericson et al. 2012) or from cytosine methylation in mtDNA (Stoccoro and Coppedè 2021), resembling the well-known CpG > TpG substitutions in the nuclear genome.

Comparing $A_H > G_H$ and $T_H > C_H$, we observe a similar trend: $A_H > G_H$ (as opposed to $T_H > C_H$) is slightly more variable, particularly showing greater sensitivity to differences between cancers and Aves (more greenish on heatmaps; Fig. 2b). This pattern is expected if $A_H > G_H$ is more associated with basal metabolism or oxidative damage as compared with $T_H > C_H$.

Analyzing the local patterns of $A_H > G_H$ substitutions, we observed that nAt > nGt and nAg > nGg are the most common motifs across all chordates (supplementary fig. S11b, Supplementary Material online). Conversely, it alters in cancers, representing the diverse environments in different cancer cells compared with germline tissues (supplementary fig. S11d, Supplementary Material online). The $A_H > G_H$ mutation, recently linked to aging in mammals (Mikhailova et al. 2022) and body temperature in chordates (Mikhailova et al. 2023), suggests a signature of damage related to aerobic metabolism (Mikhailova et al. 2022), yet the exact process driving these substitutions is still unknown.

The potential mechanism of $A_H > G_H$ transitions can be explained at least partially by the mutagenic N6-methyldeoxyadenosine (6 mA) in mtDNA, being strongly enriched on the heavy strand (Koh et al. 2018) and associated with stressful hypoxic conditions. Although enrichment in 6 mA in mammalian mitochondria is still contradictory (Li et al. 2021; Kong et al. 2022) and a link between 6 mA and $A_H > G_H$ is rather suggestive, we observed that the 6 mA motifs described as (c/a)At and A(t/g) in previous studies (Koh et al. 2018; Hao et al. 2020) are similar to our $A_H > G_H$ motifs nAt > nGt and nAg > nGg (supplementary fig. S11b, Supplementary Material online). Future research should be directed toward elucidating the mechanisms underlying $A_H > G_H$ substitutions in mtDNA.

Altogether, we propose that the symmetrical and conservative part of mtDNA mutational spectra ($C_H > T_H$ equal to $G_H > A_H$; $A_H > G_H$ equal to $T_H > C_H$) can be shaped by replication-driven POLG-induced mutations, while the asymmetrical and variable part of mtDNA mutational spectra ($C_H > T_H$ minus $G_H > A_H$ and $A_H > G_H$ minus $T_H > C_H$) is most likely shaped by damage-driven mutations, induced by deamination, oxidation, or methylation.

mtDNA Mutations Through the Lens of Mutational Signatures: BER Deficiency, MMR Deficiency, Spontaneous Depurination

To deconvolute the CytB-derived 192-component mutational spectrum of mtDNA into its underlying mutational signatures, we utilized the COSMIC SBS database (https://cancer.sanger.ac.uk/signatures/sbs/). Since the COSMIC SBS database is built upon 96-component signatures, we divided our average class-specific 192-component mtDNA spectrum into three sets (Fig. 3a; Materials and Methods): a "high" spectrum for asymmetric transitions (substitutions with the highest rate out of two complementary substitutions: nCn > nTn and nAn > nGn), "low" spectrum for symmetric transitions (substitutions with the lowest rate out of two complementary substitutions: nGn > nAn and nTn > nCn), and a "diff" spectrum, typical for the heavy strand ssDNA, calculated by subtraction of "low" spectrum from the "high" (nCn > nTn minus nGn > nAn and nAn > nGn minus nTn > nCn). Given the rarity and variability of transversions in mtDNA, we conducted main analyses without transversions (Fig. 3); however, the inclusion of transversions into analyses did not significantly affect our findings (supplementary fig. S12, Supplementary Material online). Using two different methods: SigProfilerAssignment and mSigAct (Ng et al. 2017; Islam et al. 2022; Jiang et al. 2024), we observed several signatures, namely SBS30, SBS44, SBS26, SBS6, and SBS12, which are predominant in mtDNA mutations with varying contributions to high, low, and diff spectra (Fig. 3b; Materials and Methods). Two other signatures are not discussed in detail below: SBS5 exhibits a consistently uniform pattern that is more pronounced when including transversions and is traced only by SigProfilerAssignment. SBS23 is a rare signature with unknown etiology; however, it has a pattern similar to SBS30 and is primarily traced only by mSigAct (supplementary fig. S12, Supplementary Material online). Taken together, the results derived from the application of SigProfilerAssignment and mSigAct methodologies demonstrate consistency in the assignment analyses (supplementary tables S6 and S7, Supplementary Material online). Key signatures identified through these analyses are discussed below.

SBS30 is associated with deficient base excision repair (BER) in the nuclear genome and predominantly results in C > T mutations (Drost et al. 2017; Zou et al. 2021). In mitochondria, BER is the primary repair pathway for chemically damaged bases, including deaminated, oxidized, and alkylated bases (Rong et al. 2021). Our findings indicate a BER deficiency signature in mtDNA (Fig. 3b and c) (Ikeda et al. 1998). Another notable observation is the association of SBS30 in the nuclear genome with malfunctioning NTHL1 glycosylase, suggesting an inherent decreased efficiency of this enzyme in mtDNA. NTHL1 removes oxidized pyrimidine lesions in both nuclear DNA (nDNA) (Das et al. 2020) and mtDNA (Karahalil et al. 2003); significantly decreased NTHL1 activity in Friedreich's ataxia leads to the accumulation of mtDNA mutations dramatically, affecting mitochondrial functioning (Bhalla et al. 2016). These observations suggest that NTHL1 could serve as a strategic target for mitigating the mitochondrial mutation rate in clinical settings.

BER intermediates, such as abasic sites (AP sites), may provide additional insights into mtDNA mutagenesis, though they can also result from spontaneous base loss. When BER is ineffective on ssDNA, AP sites may form as an intermediate step following glycosylase activity. In such cases, we anticipate that these AP sites could display motifs similar to SBS30 patterns ($C_H > T_H$ mutations). Spontaneous base loss will show a distinct tendency: depurination (A and G loss) outpaces depyrimidination (C and T loss) by 20-fold, guanines are 1.5 times more prone to depurination than adenines, and depurination happens faster in ssDNA than in dsDNA (Thompson and Cortez 2020). Analyzing AP sites in mouse mtDNA,

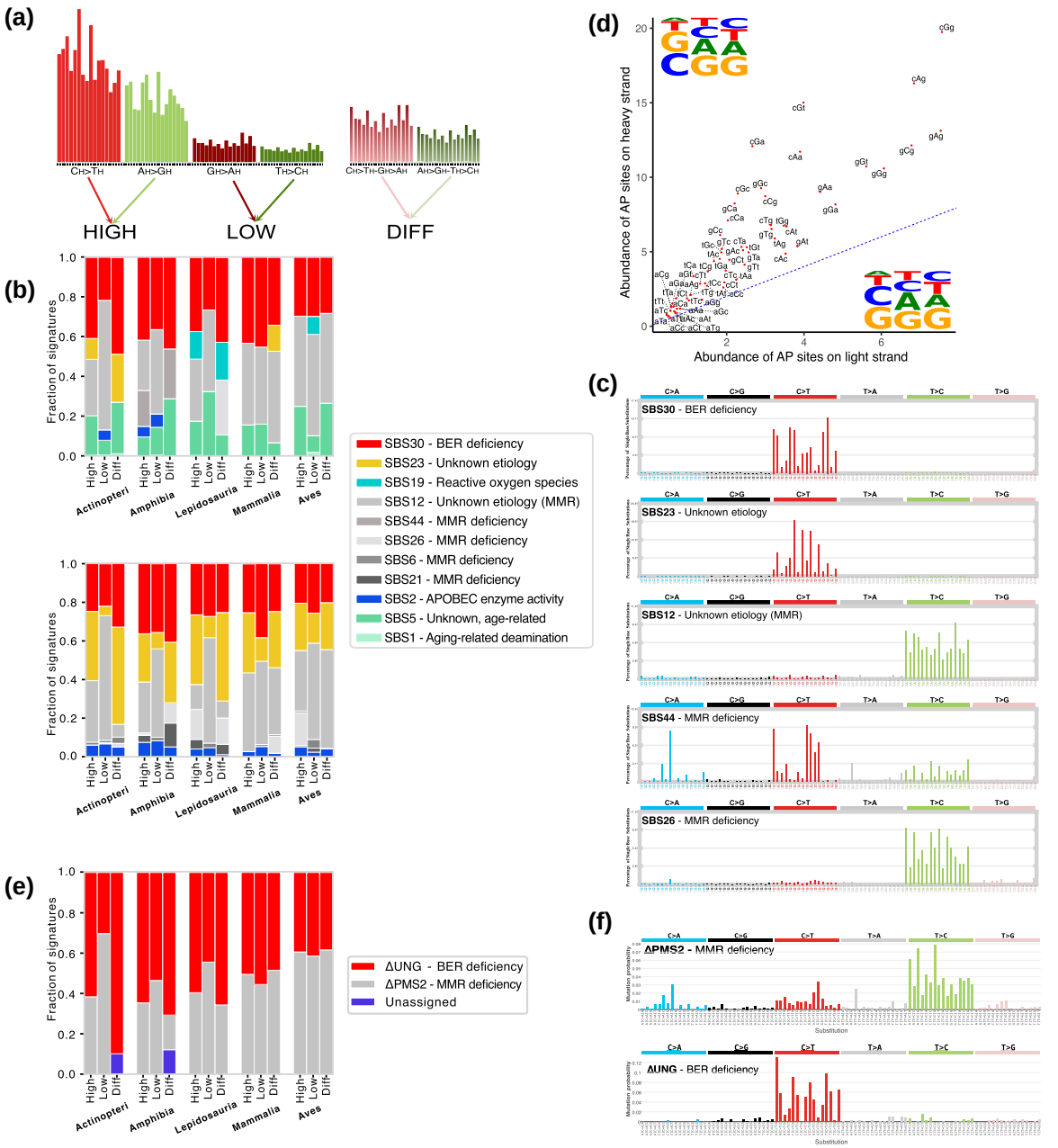

Fig. 3. Deconvolution of 192-component mutational spectra into COSMIC SBS signatures. a) The 192-component spectrum of the mitochondria is divided into three groups—"high", "low," and "diff"—for signature analysis (see details in the text). b) The signature assignment outcomes from the SigProfilerAssignment (top panel) and mSigAct (bottom panel). c) Main COSMIC signatures assigned to mtDNA mutational spectra of chordate classes (graphs generated by SigProfilerAssignment software, COSMIC version 3.3). Pattern of SBS30: BER deficiency mutations. Pattern of SBS44, SBS26, and SBS6: MMR deficiency mutations. The pattern of SBS12 may also be linked to MMR (see the text for details) d) Analysis of the trinucleotide pattern of AP sites within coding sequences of both the heavy and light strands reveals that trinucleotides on the heavy strand exhibit higher levels of damage, with G being the most frequently damaged nucleotide. e) The signature assignment outcomes using Signal Database. f) Mutational signatures from knockout models of key DNA repair genes PMS2 and UNG.

with precise single-nucleotide resolution on both strands (Cai et al. 2022), we observed that motifs with AP sites correlate well between the light and heavy strands, occurring about twice as often on the heavy strand, with Guanine being the most common AP site, followed by Adenine (Fig. 3d). This pattern is more consistent with spontaneous base loss, not BER. This process, upon replication by POLG that incorporates a dA residue opposite the abasic site (Pinz et al. 1995), can result in G > T and A > T transversions. The asymmetry observed for G > T and A > T transversions (Fig. 1a) goes in line with a predominance of AP sites on the heavy strand (Fig. 3d).

SBS44, SBS26, and SBS6 are three of seven known signatures linked to defective DNA mismatch repair (MMR), crucial for correcting mismatches during DNA replication (Li 2008). Although the presence of MMR in mtDNA has been debated, it is now widely accepted that MMR in mtDNA is deficient compared with the nuclear genome (Copeland and Longley 2014; Rong et al. 2021), which is completely in line with our results. Interestingly, MSH1 and mtMutS, the mitochondrial MMR proteins, were detected exclusively in the mitochondrial genomes of several nonbilateral animals (Muthye and Lavrov 2021). This suggests that their

loss in bilaterians may influence the mtDNA mutational spectrum. A detailed comparative analysis of the mtDNA mutational spectra of species with and without MMR proteins coded within mtDNA would be of great interest for future research.

SBS12, despite its unknown origins in the nuclear genome, might be an interesting signature because it is associated with A > G (T > C in COSMIC notation, Fig. 3c), which is the second most common transition in mtDNA and is sensitive to age and temperature (Mikhailova et al. 2022, 2023). Notably, SBS12 in the nuclear genome shows an increase with replication timing (see the "Replication timing" section in COSMIC). It also exhibits transcriptional strand asymmetry with more A > G mutations occurring on the nontranscribed strand (more T > C mutations on the transcribed strand in COSMIC notation). Additionally, SBS12 displays replication strand asymmetry, with A > G mutations occurring on the lagging strand (T > C on the leading strand in COSMIC notation). All these properties resemble the asymmetric behavior of $A_H > G_H$ substitutions in mtDNA (Mikhailova et al. 2022, 2023), and thus may shed light on the similar mutagenesis mechanisms. SBS12 is frequently observed in liver hepatocellular tumors and Biliary Tract Adenocarcinoma (Alexandrov et al. 2020). Although the etiology of SBS12 remains unknown, it has been noted (Mas-Ponte et al. 2022) that its mutational profile resembles that of PMS2 knockout models, which are associated with defects in the MMR pathway (Zou et al. 2021). Further evidence linking SBS12 to MMR comes from a study where organwise independent extractions of mutational signatures were performed (Degasperi et al. 2020). The analysis revealed that MMR reference signature 2 (RefSig MMR2) closely resembles SBS12, further supporting the association with MMR deficiency.

When comparing the "diff", "high," and "low" sets of mutations, we observed a trend where BER-related deficiencies (e.g. SBS30 and potentially SBS23, based on their pattern similarity to SBS30) are more pronounced in the "diff" and "high" categories (Fig. 3b; supplementary fig. S12, Supplementary Material online). This could be attributed to the inefficiency of BER in repairing ssDNA. In contrast, MMR-related deficiencies (SBS12, SBS44, SBS26, and SBS6) are more commonly observed in the "low" category (Fig. 3b; supplementary fig. S12, Supplementary Material online), likely reflecting the more symmetrical nature of these mutations. To further investigate this tendency, we utilized experimentally derived mutational signatures from knockout models of key DNA repair genes (Zou et al. 2021), using data from the Signal database (Degasperi et al. 2020). Based on our earlier results (Fig. 3b), which showed that most mtDNA mutations are driven by BER and MMR deficiencies, we deconvoluted our mtDNA spectra using signatures derived exclusively from experimentally obtained knockouts of nine genes involved in various repair pathways (see Materials and Methods). We confirmed that BER-associated deficiencies, such as UNG (which resembles SBS30 due to NTHL1 deficiency), are more prominent in the "diff" and "high" categories compared with "low" (Fig. 3c). In contrast, MMR-related signatures (e.g. ΔPMS2, resembling SBS12) show the opposite trend, being predominant in the "low" category (Fig. 3e). Notably, this pattern is strong and consistent across all three cold-blooded classes (Actinopteri, Amphibia, and Lepidosauria), while the contributions of BER to the "high", "low," and "diff" categories remain relatively stable and low in warm-blooded classes (Mammalia and Aves) (Fig. 3b and e). This stability may be due to elevated levels of metabolism-associated damage, which is visible as uncorrected by MMR signature of ΔPMS2, characterized by A > G (Fig. 3f, upper panel), which is known to be higher in warm-blooded animals (Mikhailova et al. 2023).

Altogether, the deconvolution analysis identified two primary mutational sources in mtDNA: (i) inefficient BER, predominantly associated with asymmetrical C > T mutations (NTHL1-driven SBS30 in COSMIC, which is similar to ΔUNG pattern in Signal (Fig. 3f, lower panel)); and (ii) inefficient MMR, mainly associated with A > G mutations (SBS12 in COSMIC, which is similar to the ΔPMS2 pattern in Signal). The BER deficiency is more pronounced in the ssDNA of cold-blooded species; however, this effect diminishes in warm-blooded species. This change is likely due to an increasing prevalence of A > G mutations uncorrected by MMR, associated with higher levels of oxidative damage observed in our data (Fig. 3e—note the stabilization of red bins and the increase in gray from cold-blooded to warm-blooded animals) and reported in the literature (Bridge et al. 2014). Expectedly, the approach used in this chapter cannot recover the POLG signature, which does not exist in the corresponding databases and remains poorly characterized. This highlights a broader potential drawback: the use of nDNA-derived mutational signatures in our analysis may introduce bias due to the transition-rich nature of mtDNA mutagenesis. This raises the concern that the identified signatures could be inherently skewed toward transitions, thereby limiting the ability to detect mtDNA-specific processes, such as those driven by POLG. Future studies utilizing de novo decomposition of signatures will be essential to independently validate and refine our findings. Considering that POLG-related mutations are expected to be symmetrical, i.e. contribute mainly to the "low" spectrum, we may assume that the POLG signature is hidden within the MMR-related one. More studies are needed to deconvolute the mutational spectrum of mtDNA into more detailed signatures. The deficiency of both BER and MMR pathways makes mtDNA especially intriguing, as DNA damage at least partially driven by life-history traits and physiological processes can be incorporated as mutations without undergoing repair. This makes mtDNA mutagenesis a highly sensitive marker for tracking various physiological processes and environmental influences across different biological contexts.

Strong Asymmetry in mtDNA Mutagenesis is Shaped by Single-strand DNA Damage During Asynchronous Replication

Throughout our paper, we have repeatedly demonstrated the pronounced asymmetry of mtDNA mutagenesis: (i) the most common transitions C > T and A > G occur several times more often on the heavy strand; (ii) transversions G > T and C > G also demonstrate increased frequencies on the heavy strand in our (Fig. 1a) and other studies (Yuan et al. 2020). Estimation of the total level of asymmetry in mtDNA (Materials and Methods) shows that ~42% of all mtDNA mutations are asymmetrical, i.e. occur exclusively on a heavy strand, while the rest of the mutations occur symmetrically on both heavy and light strands. Here, we analyze deeper the phenomenon of asymmetry to identify its primary causes.

In the nuclear genome, two types of mutational asymmetry have been described T-asymmetry (transcription asymmetry,

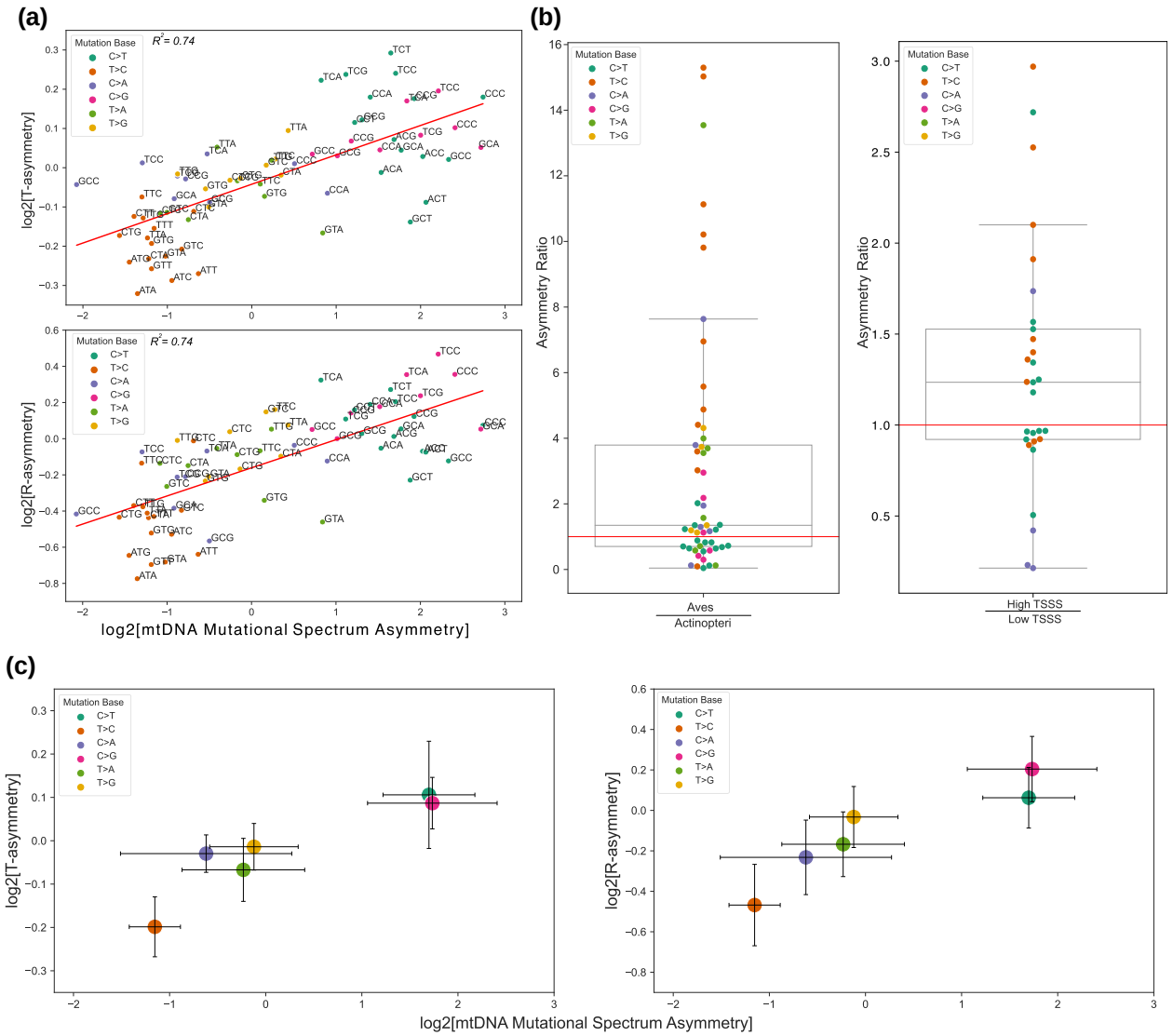

Fig. 4. Comparison of the mitochondrial asymmetry based on the global mitochondrial mutational spectrum with the T and R asymmetries in nDNA (each dot represents substitution type with a context). a) The analysis of the mitochondrial asymmetry reveals a notable positive correlation with both T (upper panel, Spearman's Rho = 0.45, $P = 2.8 \times e{-}05$, $N = 80$) and R (bottom panel, Spearman's Rho = 0.38, $P = 0.0005$, $N = 80$) nuclear asymmetries. The elimination of zero-rated substitutions (rare transversions, never observed in chordates) from the mito-asymmetry significantly improved the positive associations with both T- (Spearman's Rho = 0.74, $P = 9.1 \times e{-}13$, $N = 68$) and R-asymmetry (Spearman's Rho = 0.74, $P = 8.1 \times e{-}13$, $N = 68$). b) Direct comparison of mitochondrial asymmetries between warm- (birds) and cold- (fishes) blooded species shows an increased strength of the asymmetry in warm-blooded (the ratio is higher than one) (Wilcoxon test, $P = 3.09e{-}05$, $N = 60$). A similar trend is observed for high vs. low TSSS regions (Wilcoxon test, $P = 0.06$, $N = 29$). c) Among the six primary substitution types, a robust association is observed between mito-asymmetry and T (left panel) and R (right panel) asymmetries, superseding the influence of nucleotide content.

which originates from mutations on the nontranscribed strand) and R-asymmetry (replication asymmetry, which predominantly occurs on the lagging strand) have been described and a tight correlation between both of them has been shown (Seplyarskiy et al. 2019). We calculated mito-asymmetry from our 192-component spectra (supplementary table S7, Supplementary Material online; Materials and Methods) and observed a significant correlation between mito-asymmetry and both R- and T-asymmetries of the nDNA (Fig. 4a). Our findings show that mutations prevalent in the mtDNA heavy strand are also common in the lagging and nontranscribed strands of the nDNA. Further analysis revealed that the correlations (Fig. 4a) are primarily driven by 6 base-specific asymmetries (Fig. 4c), not by their context (supplementary figs. S13 to S14, Supplementary Material online). This indicates that, for example, C > T mutations, regardless of context, occur more frequently than G > A across all three areas: (i) mtDNA heavy strand, (ii) nDNA lagging strand, and (iii) nDNA nontranscribed strand. Categorizing the mutation pairs by the degree of asymmetry, from highest to lowest, with the first substitution in each pair occurring more frequently, we got the next ranking: C > T ≫ G > A, C > G ≫ G > C, A > G ≫ T > C, G > T ≫ C > A (Fig. 4c). Interestingly, C > T and A > G are the most common substitutions in the integral mtDNA mutational spectrum (Fig. 1a), thus shaping an overly strong asymmetry of the whole mtDNA mutational spectrum.

What is the plausible mechanism for the asymmetry's origin? The shared characteristic of the three areas described above is their single-stranded nature, suggesting a uniform mutational process influenced by ssDNA damage.

If single-stranded specific damage (Tanaka and Ozawa 1994; Faith and Pollock 2003; Sanchez-Contreras et al. 2021) is a viable hypothesis, we anticipate asymmetry to grow with (i) increased TSSS and (ii) increased total damage level. Assuming that TSSS is linearly increasing during asynchronous replication of mtDNA along the major arc (Sanchez-Contreras et al. 2021), we analyzed human cancer mtDNA data (Yuan et al. 2020). Dividing the major arc into low and high TSSS zones (see Materials and Methods), we revealed that asymmetry increases in high TSSS areas, with the high-to-low TSSS asymmetry ratio exceeding one (Fig. 4b, right panel). Testing our hypothesis of different sources of mDNA damage, we leverage the convention that mitochondrial damage is linked to aerobic metabolic rates (Ericson et al. 2012; Mikhailova et al. 2023): thus, mito-asymmetry in warm-blooded species should surpass that in cold-blooded ones. Comparing mito-asymmetry within the same gene (CytB) between the "warmest" (birds) and the "coldest" (fishes) chordate species (Mikhailova et al. 2023) in our dataset, we indeed observed an asymmetry ratio of "warm to cold" species greater than one (Fig. 4b, left panel, P value = 3.09e−05).

While an alternative explanation for asymmetry, proposed for nDNA, involves low-fidelity translesion DNA synthesis (TLS) polymerases capable of error-prone bypassing of DNA lesions (Seplyarskiy et al. 2019), we find it less likely for mito-asymmetry. First, PrimPol, an error-prone TLS polymerase observed in mammalian mitochondria (Rudd et al. 2014), is rarely recruited and has a strong preference for generating base insertions and deletions (Guilliam et al. 2015), which are rarely observed in mtDNA (Yuan et al. 2020). Second, the deamination of C and A on the heavy strand of mtDNA, resulting in the most common and asymmetrical mutations, C > T and A > G, are not expected to be helix-distorting changes that stall replication forks and necessitate PrimPol recruitment (see Zheng et al. 2006). In fact, C > T substitutions are the most common among POLG-mediated errors in in vitro experiments (Zheng et al. 2006), suggesting that POLG can make C > T transitions via cytosine deamination without issues (supplementary fig. S15, Supplementary Material online). In summary, we suggest that ssDNA damage is a key factor contributing to mutational asymmetry in mtDNA and potentially has some influence on nDNA as well.

Discussion

In this study, by integrating species-specific mtDNA mutational spectra from various chordates, we have reconstructed a comprehensive 192-component mutational spectrum. Our analyses suggest the possibility of deconvoluting this spectrum into three primary sources of mitochondrial mutations: replication-driven symmetrical mutations and two asymmetrical damage-driven categories of mutations: $C_H > T_H$ and $A_H > G_H$ characterized by distinct etiologies and dynamics.

The first component, primarily driven by symmetrical POLG replication errors (however, a possible source of symmetrical damage-driven mutations is also possible), comprises about 58% of all de novo mutations and includes C > T symmetrical mutations (26%), A > G symmetrical mutations (21%), and a symmetrical part of the majority of transversions (11%) (ds-DNA specific mutations, Fig. 5). The most prevalent type within this component is the symmetrical C > T substitutions, which also shows the most conservative pattern across chordate classes (see $G_H > A_H$ in Fig. 2b) and is known as the most common mutation for all polymerases, including POLG, due to the mispairing of T opposite to G, leading to C > T mutations (Lee and Johnson 2006; Zou et al. 2021). If the frequency of symmetrical C > T ($G_H > A_H$) mutations and most transversions increases with each mtDNA replication round, we expect a positive correlation among mutation types in the component. Notably, exactly this correlation has been confirmed recently by a principal component analysis in our comparative-species mammalian study, which shows collinearity between the majority of transversions and $G_H > A_H$ substitutions (Fig. 2c in Mikhailova et al. 2022).

The second and third components involve asymmetrical mutations, likely driven by damage, though replication-driven sources could also contribute (see Matkarimov and Saparbaev 2020). The precise chemical mechanisms behind these mutations in mtDNA require additional experimental validation. However, a plausible and parsimonious explanation is that these mutations stem from deamination events on ssDNA: the deamination of cytosine leading to C > T transitions, and the deamination of adenine resulting in A > G transitions. The distinct dynamics observed between C > T and A > G mutations may stem from differing balances between oxidative and hydrolytic deamination pathways. While adenine's deamination is relatively slow under standard physiological conditions, oxidative stress speeds up its oxidative conversion to hypoxanthine, particularly when ssDNA is exposed to high ROS levels. Cytosine deamination, by contrast, is primarily a hydrolytic process rather than oxidative, meaning that its rate is less sensitive to oxidative stress compared with adenine. This difference accounts for the observed sensitivity of A > G mutations to aerobic metabolism, while C > T mutations remain largely unaffected. Deficiencies in both BER and MMR pathways allow these deamination-based substitutions to persist uncorrected.

The second component, representing roughly 30% of mutations (ss-DNA specific C > T, Fig. 5), comprises asymmetrical C > T mutations due to damage. It is likely shaped by ssDNA damage caused by spontaneous deamination. Although the observed peak of cCc > cTc mutations (supplementary fig. S11a and c, Supplementary Material online) indicates the oxidative damage on ssDNA, the oxidative component of C > T mutations is expected to be rather low since these mutations maintain a rather constant rate across species, showing no sensitivity to metabolic or life-history changes (Mikhailova et al. 2022, 2023). This component's insensitivity to metabolic or life-history traits and replication, highlighted by its distinct position relative to both the first component ($G_H > A_H$ mutations and the majority of transversions) and the third component ($A_H > G_H$) in our comparative-species PCA plot (Fig. 2c in Mikhailova et al. 2022), suggests its potential as a molecular clock in mtDNA, similar to SBS1 in nDNA, which is characterized by C > T mutations in a CpG context due to deamination of 5-methylcytosine (SBS1). By analogy with SBS1, we can assume that the methylation of cytosine in mtDNA may explain some fraction of C > T substitutions—although the methylation of cytosine in mtDNA remains uncertain and probably low (Stoccoro and Coppedè 2021)—Somatic mtDNA mutations C > T in cancers clearly show a CpG context, suggesting that formation of 5-methylcytosine and subsequent deamination may occur (supplementary fig. S11c, Supplementary Material online, see Yuan et al. 2020), leading to clock-like dynamics.

The third component, representing around 12% of the mutation spectrum, consists of asymmetrical $A_H > G_H$ mutations

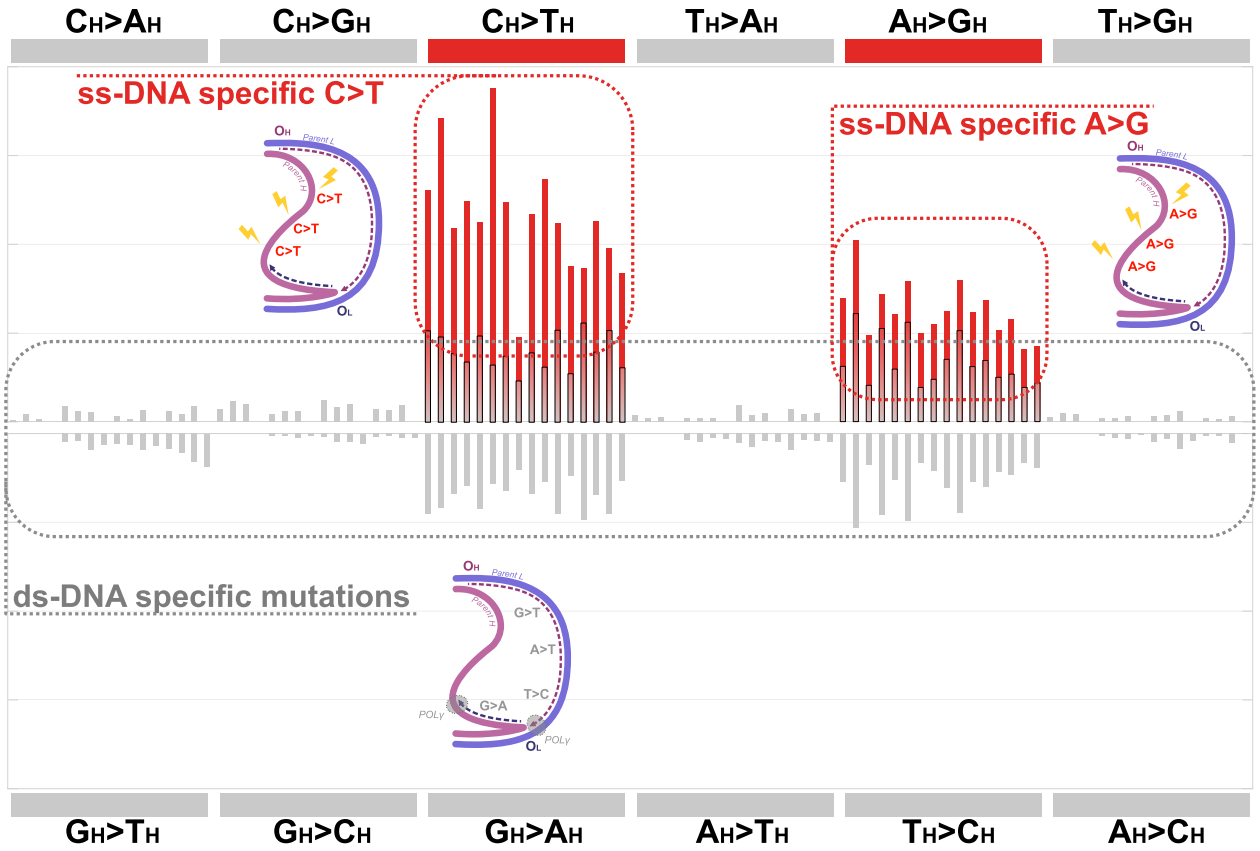

Fig. 5. Graphical visualization of the main mutational signatures in mtDNA: (i) symmetrical mutations, predominantly C > T, A > G, and rare transversions (ds-DNA specific mutations), linked to POLG's replication errors; (ii) asymmetrical C > T mutations (ss-DNA specific C > T), indicative of single-stranded DNA damage with clock-like dynamics; (iii) asymmetrical A > G mutations (ss-DNA specific A > G), resulting from single-stranded oxidative-associated DNA damage.

(ss-DNA specific A > G, Fig. 5), associated with ssDNA damage, primarily attributed to adenosine deamination. This mutation type, in contrast to the second component, exhibits notable correlations with eco-physiological traits in mammals (Mikhailova et al. 2022) and across chordates (Mikhailova et al. 2023). Its distinctiveness may be linked to its sensitivity to temperatures (Karran and Lindahl 1980), alkaline conditions (Wang and Hu 2016) and potentially oxidative conditions (i.e. the prevalence of oxidative over hydrolytic pathways of deamination on ssDNA in a mitochondria environment). The temperature dependency, pH dependency, and oxidative deamination could account for the increased $A_H >$ G_H mutations in warm-blooded compared with cold-blooded chordates (Mikhailova et al. 2023), and long-lived mammals (Mikhailova et al. 2022). All these suggestions highlight important avenues for further research.

The categorization of the mtDNA mutational spectrum into replication-driven, ssDNA damage-driven molecular clock-like $C_H > T_H$ and ssDNA damage-driven oxidative-associated $A_H >$ G_H components can enrich our understanding of species' molecular evolution. It may reveal new aspects of species' origins, history, and potential genetic factors influencing mutation rates, as it was shown in comparative studies on nDNA (Harris and Pritchard 2017; Chintalapati and Moorjani 2020; Sasani et al. 2022; Gao et al. 2023; Jayakodi et al. 2023). Additionally, these components can offer insights into the dynamics of somatic mtDNA mutation patterns in different cancer types and healthy tissues (Mikhaylova et al. 2021).

Understanding the primary mutagens in mtDNA mutational spectrum can guide strategies to reduce mtDNA mutation load, causative for many human diseases and aging. For instance, evidence that damage contributes almost half (around 42%) of de novo mutations suggests approaches to mitigate damage-induced mutations, either by reducing damage (for example, by inducing hypoxia (Ericsson et al. 2015)) or by enhancing repair processes, like boosting mitochondrial BER or MMR. Further research is required to elucidate the damage processes impacting mtDNA.

Materials and Methods

Mutational Spectra Dataset Preparation for Mitochondrial Genes of Chortate Species

In this paper, we use heavy strand notation for all mtDNA mutations and spectra. In the mutational spectra reconstruction for mitochondrial genes, we made the assumption that the majority of mtDNA synonymous polymorphic variants within chordate species are effectively neutral. Although recent suggestions indicate that certain synonymous sites in mtDNA may not be entirely neutral (Lareau et al. 2023), we evaluated the effect of removing highly constrained synonymous sites on our data. The results suggested that omitting these conservative sites did not markedly change the results (supplementary fig. S16, Supplementary Material online).

To reconstruct the species- and gene-specific mutational spectra, we used the NeMu pipeline (Efimenko et al. 2024) developed in our lab for automatic neutral spectrum derivation for different genes. First, we extracted all available mitochondrial protein sequences for all chordates available in the MIDORI2 database (reference database of mtDNA sequences

of *Eukaryota*) (version GenBank259_2023-12-17) (Leray et al. 2022) with at least 10 entries and derived 18,937 sequences of 13 protein-coding genes for 11,887 species. Second, we executed the NeMu pipeline to reconstruct mutational spectra based on synonymous substitutions. During processing, NeMu (i) collects nucleotide orthologous sequences using tblastn on the same MIDORI2 database, (ii) filters out duplicates, (iii) aligns input sequences, (iv) builds a phylogenetic tree with reconstructed ancestor sequences, (v) reconstructs mutations from all tree branches, and (vi) calculates mutational spectra using all mutations and adjusting substitution counts by the context frequencies. After the NeMu execution, we obtained 9,134 raw mutational spectra (supplementary tables S9 and S10, Supplementary Material online). Most of the input proteins currently have no 10 unique sequences in the MIDORI2 database required for nearly adequate spectrum reconstruction and were excluded in the initial stages of NeMu processing.

We used only five chordate classes *Mammalia, Actinopteri, Aves, Lepidosauria,* and *Amphibia*, that have a lot of species in the Genbank and therefore a lot of spectra in our dataset, and only 4 most sequenced mitochondrial genes at different locations of mtDNA Cytochrome b, NADH dehydrogenase subunit 2, Cytochrome c oxidase subunit I, Cytochrome c oxidase subunit III (CytB, ND2, CO1, and CO3). These spectra were carefully verified and filtered: (i) we excluded spectra of incomplete genes to make comparisons of different species homogeneous and comparable (supplementary fig. S17, Supplementary Material online); (ii) we also filtered out artifact substitutions from long tree branches, carrying more than six mutations, observed due to automatic low-accurate input data preparation for the NeMu (supplementary fig. S18a, Supplementary Material online); (iii) besides, we excluded mutational spectra that, based on <20 reconstructed mutations, contain <16 out of 64 possible transitions or have transitions fraction <65% (supplementary fig. S18b, c, and d, Supplementary Material online). The final spectra dataset contains 2,866 species- and gene-specific spectra for five Chordata classes and is described in the supplementary table S1, Supplementary Material online.

Reconstruction of the Integral 192-component Mutational Spectrum for Mitochondrial Genes of Chordate Classes

To compute the integral gene-specific and taxa-specific 192-component mutational spectrum, we averaged species-specific mutational spectra vectors. Such a procedure was performed on the level of all chordates and five chordate classes for selected mitochondrial genes (Cytb, ND2, CO1, and CO3) separately.

Comparison of the Mutational Spectra

We primarily used cosine similarity to estimate the closeness of the compared spectra. Comparison has been performed on two levels: between species and between classes. First, we conducted pairwise comparisons of mutational spectra at the species-specific level within and between classes, considering all possible combinations. Second, we estimated differences between classes using jackknife resampling of species spectra. In this process, we randomly selected 50 species from each pair of classes, calculated the 192-component mutational spectrum for both classes as described in the previous paragraph of methods, and computed the cosine similarity of either the overall mutational spectrum or its parts (transitions and transversions). In the case of human cancer spectra, we sampled 50% of patients instead of 50 species due to the low number of mutations in each patient (supplementary figs. S19 and S20, Supplementary Material online). We repeated this process in 1,000 iterations for every conceivable class/cancer combination. Through the use of jackknife resampling with fixed sample size, we effectively adjusted the influence of varying species numbers in different classes.

Calculation of the mtDNA Mutational Spectrum Asymmetry

The assessment of mtDNA asymmetry involved the transformation of a 192-component mutational spectrum into a 96-component spectrum. This was achieved by selecting frequencies of 96 SBS of pyrimidines only (C > A, C > G, C > T, T > A, T > C, and T > G) from the mtDNA mutational spectrum (see COSMIC signatures, e.g. Fig. 3c) and dividing them by complementary substitutions frequencies.

The total mitochondrial asymmetry was determined by summing the differences between mutations presented in the 96-component mitochondrial mutational spectrum and their complementary mutations.

Annotation of Mitochondrial Mutational Signatures

To decompose class-specific mtDNA mutational spectra into COSMIC signatures, we used the SigProfilerAssignment tool v0.1.6 (Islam et al. 2022). Since main COSMIC signatures are symmetrical and do not account for strand-specific mutagenesis, we split the asymmetrical mitochondrial 192-component spectra into two ("low" and "high") 96-component spectra based on the abundance of specific transitions. The "high" spectra include more frequent $C_H > T_H$ and $A_H > G_H$ transitions, while the "low" spectra include $G_H > A_H$ and $T_H > C_H$ transitions, where H signifies heavy strand notation of substitution. Also, to analyze the asymmetrical component of mutagenesis, we derived "diff" spectra by subtracting "low" from "high" spectra in a context-dependent manner (Fig. 3a). All complementary pairs of transversion rates (for example, $A_H > C_H$ and $T_H > G_H$) were averaged and equally added to the "low", "diff," and "high" spectra; otherwise, to test the impact of noisy and rare transversions, they were zeroed in these spectra. In addition, SigProfilerAssignment, like any other SigProfiler tool, is capable of processing distinct substitution counts on the reference human nuclear genome. To ensure correspondence of input spectra to the human nuclear genome, we rescaled spectra of selected chordate classes multiplying them by the trinucleotide frequencies of the human nuclear genome following the rescaling approach described in Beichman et al. (2023), i.e. we emulated mtDNA mutational processes of chordate classes by human nuclear genome mutation counts. We applied SigProfilerAssignment using the cosmic_fit function, with the following parameters: genome_build='GRCh37', nnls_add_penalty = 0.01 (reduce the number of derived noisy signatures that explain a low number of mutations), and cosmic_version = 3.3. Furthermore, to reduce the effect of the unexpected signatures in the mtDNA during decomposition of average chordate class spectra, we excluded the following subgroups of signatures from the analysis: immunosuppressants, treatment, colibactin, lymphoid, and artifact.

To enhance the robustness of our decomposition results, we conducted a parallel assignment procedure utilizing the

mSigAct package v3.0.1 (Ng et al. 2017). We analyzed the same renormalized mutational spectra of chordate classes as input spectra. The MAPAssignActivity function within the mSigAct package was employed to analyze the "low", "diff," and "high" spectra, while concurrently excluding the same signature subgroups as in the SigProfilerAssignment approach. Since the proportions of mitochondrial mutational spectra signatures were unknown, we assigned a uniform proportion of "1" for all signatures as required by the function's input parameter "sigs.presence.prop".

Additionally, we performed mutational signature assignments of our "high," "low," and "diff" spectra using the online platform Signal (Degasperi et al. 2022). The "Analyze" option was selected with default parameters to compare our spectra against experimental signatures associated with knockouts of DNA repair genes. The knockout signatures included were ΔEXO1, ΔMLH1, ΔMSH2, ΔMSH6, ΔOGG1, ΔPMS1, ΔPMS2, ΔRNF168, and ΔUNG.

Abasic sites Patterns in mtDNA

The distribution of abasic sites (AP) at single-nucleotide and single-strand (separately for light and heavy) resolution in the mouse mitochondrial genome (mm10) was obtained from Cai et al. (2022). We quantified the AP sites occurrences within all 64 trinucleotide sequences of each strand of mouse mtDNA, excluding the control region. These values were then adjusted by the counts of trinucleotide motifs in a strand-specific manner.

We created trinucleotide logos using the Python package logomaker (Tareen and Kinney 2020). These logos used the normalized AP sites count for each trinucleotide as the nucleotide weights. These weights were averaged across all 64 trinucleotides, and final nucleotide counts at each trinucleotide position were normalized and converted to frequencies.

Analysis of Cytb Gene Location in mtDNA

The Cytb loci were obtained from the GenBank RefSeq mtDNA genomes as of 2024 May 13, spanning a total of 7,832 chordate species. For a subset of 6,363 genomes within this group, the oriL positions were predicted using the MITOS 2.1.9 Python package (Bernt et al. 2013). Of the species represented in the Cytb mutational spectra dataset, 749 out of 1,697 were analyzed in this analysis. Other species from our dataset probably don't have a full mtDNA genome. To ascertain the relative position of Cytb across different species, the oriL start coordinates were subtracted from the Cytb start position, and the resulting distribution was visualized.

mtDNA Mutational Spectrum of Human Cancers

Somatic mutations in the full human mtDNA were derived from a comprehensive analysis of the human mitochondrial genome by (Yuan et al. 2020). The 192-component mutational spectrum was constructed by aggregating SBS within a trinucleotide context. This was meticulously adjusted to account for all conceivable synonymous substitutions expected in the human mtDNA reference sequence NC_012920.1 (CRS). While it is possible to classify all mtDNA substitutions in human cancers as neutral, as observed in Fig. 3a of Yuan et al. (2020), our analysis was deliberately confined to the predominantly neutral component of the mutation sample—synonymous mutations.

We also calculated spectra for different parts of human mtDNA. To compare the mutational spectra of genome regions that have low and high TSSS due to asynchronous replication, we separately calculated spectra for the region with low TSSS (first quartile of the major arc, 5,800–8,300) and the region with high TSSS (last quartile of the major arc, 11,000–13,500).

Furthermore, to enhance the accuracy of our comparisons with the mutational spectra of different chordate classes, we incorporated a patient-specific sampling in the calculation of mutation spectra (supplementary fig. S20, Supplementary Material online). Each replicate was constructed using mutations from a randomly selected half of the patient cohort, thereby accounting for the influence of patient-specific mutagenesis patterns.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Author Contributions

The design of the study developed by K.P. Data mining and processing performed by D.I., B.E., and K.G. Manuscript prepared by K.P., D.I., B.E., and A.G.M. All authors discussed in depth the manuscript and the rationale behind the project. All authors read and approved the final manuscript. The authors express their appreciation to Vladimir Seplyarskiy for his valuable contributions and insightful discussions regarding asymmetry and repair mechanisms. We also thank the high-performance computing platform at the Immanuel Kant Baltic Federal University.

Funding

This work was supported by the Federal Academic Leadership Program Priority 2030 at the Immanuel Kant Baltic Federal University (to D.I. and K.P.). A.G.M. and B.E. is supported by the Russian Science Foundation grant no. 21-75-20143. K.G. is supported by the Russian Science Foundation grant No. 21-75-20145. V.S. is supported by the Ministry of Science and Higher Education of the Russian Federation (agreement no. 075-15-2021-1084). W.S.K. is supported by Deutsche Forschungsgemeinschaft (grant KU 911/22-1). M.K.S. and B.T.M. are supported by the Ministry of Science and Higher Education of the Republic of Kazakhstan (grant nos AP23485899, BR24992841, and BR24993023).

Conflict of Interest

None declared.

Data Availability

All analyses we performed in Python and R. Scripts and data are available on GitHub: https://github.com/mitoclub/mtdna-192component-mutspec-chordata.

References

Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, Boot A, Covington KR, Gordenin DA, Bergstrom EN, *et al.* The repertoire of mutational signatures in human cancer. *Nature*. 2020:578(7793):94–101. https://doi.org/10.1038/s41586-020-1943-3.

Almatarneh MH, Flinn CG, Poirier RA, Sokalski WA. Computational study of the deamination reaction of cytosine with H_2O and OH-. *J Phys Chem A*. 2006:110(26):8227–8234. https://doi.org/10.1021/jp062300u.

Beichman AC, Robinson J, Lin M, Moreno-Estrada A, Nigenda-Morales S, Harris K. Evolution of the mutation spectrum across a mammalian phylogeny. *Mol Biol Evol*. 2023:40(10):msad213. https://doi.org/10.1093/molbev/msad213.

Belle EMS, Piganeau G, Gardner M, Eyre-Walker A. An investigation of the variation in the transition bias among various animal mitochondrial DNA. *Gene*. 2005:355:58–66. https://doi.org/10.1016/j.gene.2005.05.019.

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol*. 2013:69(2):313–319. https://doi.org/10.1016/j.ympev.2012.08.023.

Bhalla AD, Khodadadi-Jamayran A, Li Y, Lynch DR, Napierala M. Deep sequencing of mitochondrial genomes reveals increased mutation load in Friedreich's ataxia. *Ann Clin Transl Neurol*. 2016:3(7):523–536. https://doi.org/10.1002/acn3.322.

Bhandari V, Hoey C, Liu LY, Lalonde E, Ray J, Livingstone J, Lesurf R, Shiah Y-J, Vujcic T, Huang X, *et al*. Molecular landmarks of tumor hypoxia across cancer types. *Nat Genet*. 2019:51(2):308–318. https://doi.org/10.1038/s41588-018-0318-2.

Boot A, Liu M, Stantial N, Shah V, Yu W, Nitiss KC, Nitiss JL, Jinks-Robertson S, Rozen SG. Recurrent mutations in topoisomerase IIα cause a previously undescribed mutator phenotype in human cancers. *Proc Natl Acad Sci U S A*. 2022:119(4):e2114024119. https://doi.org/10.1073/pnas.2114024119.

Bridge G, Rashid S, Martin SA. DNA mismatch repair and oxidative DNA damage: implications for cancer biology and treatment. *Cancers (Basel)*. 2014:6(3):1597–1614. https://doi.org/10.3390/cancers6031597.

Cai Y, Cao H, Wang F, Zhang Y, Kapranov P. Complex genomic patterns of abasic sites in mammalian DNA revealed by a high-resolution SSiNGLe-AP method. *Nat Commun*. 2022:13(1):5868. https://doi.org/10.1038/s41467-022-33594-1.

Chatterjee N, Walker GC. Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen*. 2017:58(5):235–263. https://doi.org/10.1002/em.22087.

Chintalapati M, Moorjani P. Evolution of the mutation rate across primates. *Curr Opin Genet Dev*. 2020:62:58–64. https://doi.org/10.1016/j.gde.2020.05.028.

Chu X-L, Zhang B-W, Zhang Q-G, Zhu B-R, Lin K, Zhang D-Y. Temperature responses of mutation rate and mutational spectrum in an *Escherichia coli* strain and the correlation with metabolic rate. *BMC Evol Biol*. 2018:18(1):126. https://doi.org/10.1186/s12862-018-1252-8.

Copeland WC, Longley MJ. Mitochondrial genome maintenance in health and disease. *DNA Repair (Amst)*. 2014:19:190–198. https://doi.org/10.1016/j.dnarep.2014.03.010.

Czernecki D, Nourisson A, Legrand P, Delarue M. Reclassification of family A DNA polymerases reveals novel functional subfamilies and distinctive structural features. *Nucleic Acids Res*. 2023:51(9):4488–4507. https://doi.org/10.1093/nar/gkad242.

Das L, Quintana VG, Sweasy JB. NTHL1 in genomic integrity, aging and cancer. *DNA Repair (Amst)*. 2020:93:102920. https://doi.org/10.1016/j.dnarep.2020.102920.

Degasperi A, Amarante TD, Czarnecki J, Shooter S, Zou X, Glodzik D, Morganella S, Nanda AS, Badja C, Koh G, *et al*. A practical framework and online tool for mutational signature analyses show inter-tissue variation and driver dependencies. *Nat Cancer*. 2020:1(2):249–263. https://doi.org/10.1038/s43018-020-0027-5.

Degasperi A, Zou X, Amarante TD, Martinez-Martinez A, Koh GCC, Dias JML, Heskin L, Chmelova L, Rinaldi G, Wang VYW, *et al*. Substitution mutational signatures in whole-genome-sequenced cancers in the UK population. *Science*. 2022:376(6591):science.abl9283. https://doi.org/10.1126/science.abl9283.

Degtyareva NP, Saini N, Sterling JF, Placentra VC, Klimczak LJ, Gordenin DA, Doetsch PW. Mutational signatures of redox stress in yeast single-strand DNA and of aging in human mitochondrial DNA share a common feature. *PLoS Biol*. 2019:17(5):e3000263. https://doi.org/10.1371/journal.pbio.3000263.

Dillon MM, Sung W, Sebra R, Lynch M, Cooper VS. Genome-wide biases in the rate and molecular spectrum of spontaneous mutations in *Vibrio cholerae* and *Vibrio fischeri*. *Mol Biol Evol*. 2017:34(1):93–109. https://doi.org/10.1093/molbev/msw224.

Drost J, van Boxtel R, Blokzijl F, Mizutani T, Sasaki N, Sasselli V, de Ligt J, Behjati S, Grolleman JE, van Wezel T, *et al*. Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science*. 2017:358(6360):234–238. https://doi.org/10.1126/science.aao3130.

Efimenko B, Popadin K, Gunbin K. NeMu: a comprehensive pipeline for accurate reconstruction of neutral mutation spectra from evolutionary data. *Nucleic Acids Res*. 2024:52(W1):W108–W115. https://doi.org/10.1093/nar/gkae438.

Ericson NG, Kulawiec M, Vermulst M, Sheahan K, O'Sullivan J, Salk JJ, Bielas JH. Decreased mitochondrial DNA mutagenesis in human colorectal cancer. *PLoS Genet*. 2012:8(6):e1002689. https://doi.org/10.1371/journal.pgen.1002689.

Ericsson AC, Akter S, Hanson MM, Busi SB, Parker TW, Schehr RJ, Hankins MA, Ahner CE, Davis JW, Franklin CL, *et al*. Differential susceptibility to colorectal cancer due to naturally occurring gut microbiota. *Oncotarget*. 2015:6(32):33689–33704. https://doi.org/10.18632/oncotarget.5604.

Faith JJ, Pollock DD. Likelihood analysis of asymmetrical mutation bias gradients in vertebrate mitochondrial genomes. *Genetics*. 2003:165(2):735–745. https://doi.org/10.1093/genetics/165.2.735.

Gao Z, Zhang Y, Cramer N, Przeworski M, Moorjani P. 2023. Limited role of generation time changes in driving the evolution of the mutation spectrum in humans. *Elife*. 2023;12:e81188. https://doi.org/10.7554/eLife.81188.

Gavrilov VM, Golubeva TB, Bushuev AV. Evolution of metabolic scaling among the tetrapod: effect of phylogeny, the geologic time of class formation, and uniformity of species within a class. *Integr Zool*. 2022:17(5):904–917. https://doi.org/10.1111/1749-4877.12611.

Guilliam TA, Jozwiakowski SK, Ehlinger A, Barnes RP, Rudd SG, Bailey LJ, Skehel JM, Eckert KA, Chazin WJ, Doherty AJ. Human PrimPol is a highly error-prone polymerase regulated by single-stranded DNA binding proteins. *Nucleic Acids Res*. 2015:43(2):1056–1068. https://doi.org/10.1093/nar/gku1321.

Hao Z, Wu T, Cui X, Zhu P, Tan C, Dou X, Hsu K-W, Lin Y-T, Peng P-H, Zhang L-S, *et al*. N6-deoxyadenosine methylation in mammalian mitochondrial DNA. *Mol Cell*. 2020:78(3):382–395.e8. https://doi.org/10.1016/j.molcel.2020.02.018.

Harris K, Pritchard JK. Rapid evolution of the human mutation spectrum. *Elife*. 2017:6:e24284. https://doi.org/10.7554/eLife.24284.

Ikeda S, Biswas T, Roy R, Izumi T, Boldogh I, Kurosky A, Sarker AH, Seki S, Mitra S. Purification and characterization of human NTH1, a homolog of *Escherichia coli* endonuclease III. Direct identification of Lys-212 as the active nucleophilic residue. *J Biol Chem*. 1998:273(34):21585–21593. https://doi.org/10.1074/jbc.273.34.21585.

Islam SMA, Díaz-Gay M, Wu Y, Barnes M, Vangara R, Bergstrom EN, He Y, Vella M, Wang J, Teague JW, *et al*. Uncovering novel mutational signatures by extraction with SigProfilerExtractor. *Cell Genom*. 2022:2(11):None. https://doi.org/10.1016/j.xgen.2022.100179.

Jayakodi M, Golicz AA, Kreplak J, Fechete LI, Angra D, Bednář P, Bornhofen E, Zhang H, Boussageon R, Kaur S, *et al*. The giant diploid faba genome unlocks variation in a global protein crop. *Nature*. 2023:615(7953):652–659. https://doi.org/10.1038/s41586-023-05791-5.

Jiang N, Wu Y, Rozen GS. A new approach to the challenging problem of mutational signature attribution. *bioRxiv 594967*. 2024. https://doi.org/10.1101/2024.05.20.594967v1.

Kalous M, Drahota Z. The role of mitochondria in aging. *Physiol Res*. 1996:45:351–359. https://pubmed.ncbi.nlm.nih.gov/9085362/.

Karahalil B, de Souza-Pinto NC, Parsons JL, Elder RH, Bohr VA. Compromised incision of oxidized pyrimidines in liver mitochondria of mice deficient in NTH1 and OGG1 glycosylases. *J Biol Chem*. 2003:278(36):33701–33707. https://doi.org/10.1074/jbc.M301617200.

Karran P, Lindahl T. Hypoxanthine in deoxyribonucleic acid: generation by heat-induced hydrolysis of adenine residues and release in free form by a deoxyribonucleic acid glycosylase from calf thymus. *Biochemistry*. 1980:19(26):6005–6011. https://doi.org/10.1021/bi00567a010.

Khan YA, Jungreis I, Wright JC, Mudge JM, Choudhary JS, Firth AE, Kellis M. Evidence for a novel overlapping coding sequence in POLG initiated at a CUG start codon. *BMC Genet*. 2020:21(1):25. https://doi.org/10.1186/s12863-020-0828-7.

Koh CWQ, Goh YT, Toh JDW, Neo SP, Ng SB, Gunaratne J, Gao Y-G, Quake SR, Burkholder WF, Goh WSS. Single-nucleotide-resolution sequencing of human N6-methyldeoxyadenosine reveals strand-asymmetric clusters associated with SSBP1 on the mitochondrial genome. *Nucleic Acids Res*. 2018:46(22):11659–11670. https://doi.org/10.1093/nar/gky1104.

Koh G, Degasperi A, Zou X, Momen S, Nik-Zainal S. Mutational signatures: emerging concepts, caveats and clinical applications. *Nat Rev Cancer*. 2021:21(10):619–637. https://doi.org/10.1038/s41568-021-00377-7.

Kong Y, Cao L, Deikus G, Fan Y, Mead EA, Lai W, Zhang Y, Yong R, Sebra R, Wang H, *et al.* Critical assessment of DNA adenine methylation in eukaryotes using quantitative deconvolution. *Science*. 2022:375(6580):515–522. https://doi.org/10.1126/science.abe7489.

Lareau CA, Yin Y, Gutierrez JC, Dhindsa RS, Gribling-Burrer A-S, Hsieh Y-H, Nitsch L, Buquicchio FA, Abay T, Zielinski S, *et al.* Codon affinity in mitochondrial DNA shapes evolutionary and somatic fitness. *bioRxiv 537997*. 2023. https://doi.org/10.1101/2023.04.23.537997v1.

Lee HR, Johnson KA. Fidelity of the human mitochondrial DNA polymerase. *J Biol Chem*. 2006:281(47):36236–36240. https://doi.org/10.1074/jbc.M607964200.

Leray M, Knowlton N, Machida RJ. MIDORI2: a collection of quality controlled, preformatted, and regularly updated reference databases for taxonomic assignment of eukaryotic mitochondrial sequences. *Environ DNA*. 2022:4(4):894–907. https://doi.org/10.1002/edn3.303.

Li G-M. Mechanisms and functions of DNA mismatch repair. *Cell Res*. 2008:18(1):85–98. https://doi.org/10.1038/cr.2007.115.

Li X, Zhang Z, Luo X, Schrier J, Yang AD, Wu TP. The exploration of N6-deoxyadenosine methylation in mammalian genomes. *Protein Cell*. 2021:12(10):756–768. https://doi.org/10.1007/s13238-021-00866-3.

Mas-Ponte D, McCullough M, Supek F. Spectrum of DNA mismatch repair failures viewed through the lens of cancer genomics and implications for therapy. *Clin Sci (Lond)*. 2022:136(5):383–404. https://doi.org/10.1042/CS20210682.

Matkarimov BT, Saparbaev MK. DNA repair and mutagenesis in vertebrate mitochondria: evidence for asymmetric DNA strand inheritance. *Adv Exp Med Biol*. 2020:1241:77–100. https://doi.org/10.1007/978-3-030-41283-8_6.

Mikhailova AG, Iliushchenko D, Shamansky V, Mikhailova AA, Ushakova K, Tretyakov E, Oreshkov S, Knorre D, Polishchuk L, Lawless D, *et al.* A > G substitutions on a heavy chain of mitochondrial genome marks an increased level of aerobic metabolism in warm versus cold vertebrates. *bioRxiv 221184*. 2023. https://doi.org/10.1101/2020.07.25.221184v3.

Mikhailova AG, Mikhailova AA, Ushakova K, Tretiakov EO, Iliushchenko D, Shamansky V, Lobanova V, Kozenkov I, Efimenko B, Yurchenko AA, *et al.* A mitochondria-specific mutational signature of aging: increased rate of A > G substitutions on the heavy strand. *Nucleic Acids Res*. 2022:50(18):10264–10277. https://doi.org/10.1093/nar/gkac779.

Mikhaylova AG, Mikhailova AA, Ushakova K, Tretiakov EO, Shamansky V, Yurchenko A, Zazhytska M, Zdobnov E, Makeev V, Yurov V, *et al.* Mammalian mitochondrial mutational spectrum as a hallmark of cellular and organismal aging. *bioRxiv 589168*. 2021. https://doi.org/10.1101/589168v3.full.

Muthye V, Lavrov DV. Multiple losses of MSH1, gain of mtMutS, and other changes in the MutS family of DNA repair proteins in animals. *Genome Biol Evol*. 2021:13(9):evab191. https://doi.org/10.1093/gbe/evab191.

Ng AWT, Poon SL, Huang MN, Lim JQ, Boot A, Yu W, Suzuki Y, Thangaraju S, Ng CCY, Tan P, *et al.* Aristolochic acids and their derivatives are widely implicated in liver cancers in Taiwan and throughout Asia. *Sci Transl Med*. 2017:9(412):eaan6446. https://doi.org/10.1126/scitranslmed.aan6446.

Oliveira MT, Haukka J, Kaguni LS. Evolution of the metazoan mitochondrial replicase. *Genome Biol Evol*. 2015:7(4):943–959. https://doi.org/10.1093/gbe/evv042.

Pinz KG, Shibutani S, Bogenhagen DF. Action of mitochondrial DNA polymerase gamma at sites of base loss or oxidative damage. *J Biol Chem*. 1995:270(16):9202–9206. https://doi.org/10.1074/jbc.270.16.9202.

Rong Z, Tu P, Xu P, Sun Y, Yu F, Tu N, Guo L, Yang Y. The mitochondrial response to DNA damage. *Front Cell Dev Biol*. 2021:9:669379. https://doi.org/10.3389/fcell.2021.669379.

Rudd SG, Bianchi J, Doherty AJ. PrimPol-A new polymerase on the block. *Mol Cell Oncol*. 2014:1(2):e960754. https://doi.org/10.4161/23723548.2014.960754.

Saclier N, Chardon P, Malard F, Konecny-Dupré L, Eme D, Bellec A, Breton V, Duret L, Lefebure T, Douady CJ. Bedrock radioactivity influences the rate and spectrum of mutation. *Elife*. 2020:9:e56830. https://doi.org/10.7554/eLife.56830.

Sanchez-Contreras M, Sweetwyne MT, Kohrn BF, Tsantilas KA, Hipp MJ, Schmidt EK, Fredrickson J, Whitson JA, Campbell MD, Rabinovitch PS, *et al.* A replication-linked mutational gradient drives somatic mutation accumulation and influences germline polymorphisms and genome composition in mitochondrial DNA. *Nucleic Acids Res*. 2021:49(19):11103–11118. https://doi.org/10.1093/nar/gkab901.

Sanchez-Contreras M, Sweetwyne MT, Tsantilas KA, Whitson JA, Campbell MD, Kohrn BF, Kim HJ, Hipp MJ, Fredrickson J, Nguyen MM, *et al.* The multi-tissue landscape of somatic mtDNA mutations indicates tissue-specific accumulation and removal in aging. *Elife*. 2023:12:e83395. https://doi.org/10.7554/eLife.83395.

Sasani TA, Ashbrook DG, Beichman AC, Lu L, Palmer AA, Williams RW, Pritchard JK, Harris K. A natural mutator allele shapes mutation spectrum variation in mice. *Nature*. 2022:605(7910):497–502. https://doi.org/10.1038/s41586-022-04701-5.

Seplyarskiy VB, Akkuratov EE, Akkuratova N, Andrianova MA, Nikolaev SI, Bazykin GA, Adameyko I, Sunyaev SR. Error-prone bypass of DNA lesions during lagging-strand replication is a common source of germline and cancer mutations. *Nat Genet*. 2019:51(1):36–41. https://doi.org/10.1038/s41588-018-0285-7.

Stoccoro A, Coppedè F. Mitochondrial DNA methylation and human diseases. *Int J Mol Sci*. 2021:22(9):4594. https://doi.org/10.3390/ijms22094594.

Sugimura S, Matoba S, Hashiyada Y, Aikawa Y, Ohtake M, Matsuda H, Kobayashi S, Konishi K, Imai K. Oxidative phosphorylation-linked respiration in individual bovine oocytes. *J Reprod Dev*. 2012:58(6):636–641. https://doi.org/10.1262/jrd.2012-082.

Tanaka M, Ozawa T. Strand asymmetry in human mitochondrial DNA mutations. *Genomics*. 1994:22(2):327–335. https://doi.org/10.1006/geno.1994.1391.

Tareen A, Kinney JB. Logomaker: beautiful sequence logos in python. *Bioinformatics*. 2020:36(7):2272–2274. https://doi.org/10.1093/bioinformatics/btz921.

Taylor RW, Turnbull DM. Mitochondrial DNA mutations in human disease. *Nat Rev Genet*. 2005:6(5):389–402. https://doi.org/10.1038/nrg1606.

Thompson PS, Cortez D. New insights into abasic site repair and tolerance. *DNA Repair (Amst)*. 2020:90:102866. https://doi.org/10.1016/j.dnarep.2020.102866.

Trimarchi JR, Liu L, Porterfield DM, Smith PJ, Keefe DL. Oxidative phosphorylation-dependent and -independent oxygen consumption

by individual preimplantation mouse embryos. *Biol Reprod*. 2000:62(6):1866–1874. https://doi.org/10.1095/biolreprod62.6.1866.

Wang S, Hu A. Comparative study of spontaneous deamination of adenine and cytosine in unbuffered aqueous solution at room temperature. *Chem Phys Lett*. 2016:653:207–211. https://doi.org/10.1016/j.cplett.2016.05.001.

Yi K, Kim SY, Bleazard T, Kim T, Youk J, Ju YS. Mutational spectrum of SARS-CoV-2 during the global pandemic. *Exp Mol Med*. 2021:53(8):1229–1237. https://doi.org/10.1038/s12276-021-00658-z.

Yuan Y, Ju YS, Kim Y, Li J, Wang Y, Yoon CJ, Yang Y, Martincorena I, Creighton CJ, Weinstein JN, *et al*. Comprehensive molecular characterization of mitochondrial genomes in human cancers. *Nat Genet*. 2020:52(3):342–352. https://doi.org/10.1038/s41588-019-0557-x.

Zheng W, Khrapko K, Coller HA, Thilly WG, Copeland WC. Origins of human mitochondrial point mutations as DNA polymerase gamma-mediated errors. *Mutat Res*. 2006:599(1-2):11–20. https://doi.org/10.1016/j.mrfmmm.2005.12.012.

Zou X, Koh GCC, Nanda AS, Degasperi A, Urgo K, Roumeliotis TI, Agu CA, Badja C, Momen S, Young J, *et al*. A systematic CRISPR screen defines mutational mechanisms underpinning signatures caused by replication errors and endogenous DNA damage. *Nat Cancer*. 2021:2(6):643–657. https://doi.org/10.1038/s43018-021-00200-0.