10264–10277 *Nucleic Acids Research, 2022, Vol. 50, No. 18* *Published online 21 September 2022*
https://doi.org/10.1093/nar/gkac779

A mitochondria-specific mutational signature of aging: increased rate of A > G substitutions on the heavy strand

Alina G. Mikhailova[1,2,†], Alina A. Mikhailova[1,†], Kristina Ushakova[1,†], Evgeny O. Tretiakov[1,3,†], Dmitrii Iliushchenko[1], Victor Shamansky[1], Valeria Lobanova[1], Ivan Kozenkov[1], Bogdan Efimenko[1], Andrey A. Yurchenko[4], Elena Kozenkova[5], Evgeny M. Zdobnov[6,7], Vsevolod Makeev[2,8], Valerian Yurov[5], Masashi Tanaka[9], Irina Gostimskaya[10], Zoe Fleischmann[11], Sofia Annis[11], Melissa Franco[11], Kevin Wasko[11], Stepan Denisov[1,12], Wolfram S. Kunz[13], Dmitry Knorre[14], Ilya Mazunin[15,16,17], Sergey Nikolaev[4], Jacques Fellay[7,18,†], Alexandre Reymond[19,†], Konstantin Khrapko[11,†], Konstantin Gunbin[1,20,†] and Konstantin Popadin[1,7,18,*,†]

[1]Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal University, Kaliningrad, Russian Federation, [2]Vavilov Institute of General Genetics RAS, Moscow, Russia, [3]Department of Molecular Neurosciences, Center for Brain Research, Medical University of Vienna, Vienna, Austria, [4]INSERM U981, Gustave Roussy Cancer Campus, Université Paris Saclay, Villejuif, France, [5]Institute of Physics, Mathematics and Information Technology, Immanuel Kant Baltic Federal University, Kaliningrad, Russian Federation, [6]Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland, [7]Swiss Institute of Bioinformatics, Lausanne, Switzerland, [8]Moscow Institute of Physics and Technology, Moscow, Russian Federation, [9]Department of Neurology, Juntendo University Graduate School of Medicine, Tokyo, Japan, [10]Manchester Institute of Biotechnology, The University of Manchester, Manchester, United Kingdom, [11]Department of Biology, Northeastern University, Boston, MA, USA, [12]School of Biological Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom, [13]Department of Epileptology and Institute of Experimental Epileptology and Cognition Research, University Bonn, Bonn, Germany, [14]The A.N. Belozersky Institute Of Physico-Chemical Biology, Moscow State University, Moscow, Russian Federation, [15]Center for Molecular and Cellular Biology, Skolkovo Institute of Science and Technology (Skoltech), Skolkovo, Russian Federation, [16]Fomin Clinic, Moscow, Russian Federation, [17]Medical Genomics LLC, Moscow, Russian Federation, [18]School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, [19]Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland and [20]Institute of Molecular and Cellular Biology SB RAS, Novosibirsk, Russian Federation

Received April 28, 2022; Revised August 02, 2022; Editorial Decision August 25, 2022; Accepted September 07, 2022

ABSTRACT

The mutational spectrum of the mitochondrial DNA (mtDNA) does not resemble any of the known mutational signatures of the nuclear genome and variation in mtDNA mutational spectra between different organisms is still incomprehensible. Since mitochondria are responsible for aerobic respiration, it is expected that mtDNA mutational spectrum is affected by oxidative damage. Assuming that oxidative damage increases with age, we analyse mtDNA mutagenesis of different species in regards to their generation length. Analysing, (i) dozens of thousands of somatic mtDNA mutations in samples of different ages (ii) 70053 polymorphic synonymous mtDNA substitutions reconstructed in 424 mammalian species with different generation lengths and (iii) synonymous nucleotide content of 650 complete mitochondrial genomes of mammalian species we observed that the frequency of $A_H > G_H$ substitutions (H: heavy strand notation) is twice bigger in species with high versus low generation length making their mtDNA

*To whom correspondence should be addressed. Tel: +41 21 693 18 03; Email: konstantin.popadin@epfl.ch
†The authors wish it to be known that, in their opinion, the first four authors should be regarded as joint First Authors and the last five authors should be regarded as joint Last Authors.

more A_H poor and G_H rich. Considering that $A_H > G_H$ substitutions are also sensitive to the time spent single-stranded (TSSS) during asynchronous mtDNA replication we demonstrated that $A_H > G_H$ substitution rate is a function of both species-specific generation length and position-specific TSSS. We propose that $A_H > G_H$ is a mitochondria-specific signature of oxidative damage associated with both aging and TSSS.

INTRODUCTION

Molecular evolution is a function of both mutagenesis and selection. To uncover selection forces, it is crucial to reconstruct the mutational process. It has been suggested for example, that an excess of G nucleotides in the mitochondrial genome (mtDNA) of long-lived mammals is a result of selection, favouring more stable genomes in long-lived species (1). However, this conclusion could be premature prior to comparing mutational processes between short- and long-lived mammals. Indeed, significant changes in mtDNA mutational spectra between different species have been shown (2,3), however, no driving factors explaining this variation have been proposed till now.

Interestingly, a similar knowledge gap on mtDNA mutational spectra exists on the comparative-tissues level. Pan-cancer studies have shown that mtDNA has a unique mutational signature that differs from all known nuclear signatures (4,5). Moreover, well-known strong exogenous mutagens such as tobacco smoke in lung cancers of smokers or ultraviolet light in melanomas do not show expected effects on the mitochondrial mutational spectrum (4,5). Thus, the main mutagen of mtDNA as well as the causes of variation in mtDNA mutational spectra are unknown on both comparative-tissues and comparative-species levels.

The widely accepted expectation is that reactive oxygen species (ROS) produced by mitochondria can damage mtDNA (6). The well-documented ROS-induced mutational signature is the modification of the guanine (G) DNA base to 7,8-dihydro-8-oxo-20-deoxyguanosine (8-oxodG), which after mispairing with adenine, leads to G > T transversion mutations. Although G > T substitutions are considered to be the hallmark of oxidative damage in the nuclear DNA (COSMIC signature 18) (7–9), it is rather rare in mtDNA (4) and doesn't significantly increase with age in mtDNA (4,10,11). So, up to now, there is no well-established mutational signature of oxidative damage in mtDNA (12).

Taking into account recent progress in deciphering the variation in mutational spectra of the nuclear genome as a function of different cancer types (13), environmental agents (9), gene knockouts (14), human populations (15) and primate species (16), here we focus on mtDNA and perform a large-scale reconstruction of its mutational spectra across hundreds of mammalian species. Considering the tight association of the level of mtDNA metabolism (and thus potential mtDNA mutagens) with species-specific life-history traits, we aimed to uncover a correlation between the mtDNA mutational spectrum and life-history traits. Using collections of (i) somatic mtDNA mutations in mice and humans, (ii) polymorphic synonymous substitutions in hundreds of mammalian species and (iii) nucleotide content in whole mitochondrial genomes of mammals, we observed one universal trend: $A_H > G_H$ substitutions (H: heavy strand notation, see Materials and Methods) positively correlates with the generation length. Considering an additional association of the $A_H > G_H$ substitutions with the time spent single-stranded (TSSS) during asynchronous mtDNA replication and numerous literature data about A > G substitutions, we propose that the increased $A_H > G_H$ in mtDNA of long-lived mammals is a mutational signature of age-associated damage, specific to single-stranded DNA. Therefore, he described variation in the mtDNA mutational spectrum should be considered more widely in somatic, population and evolutionary mtDNA analyses.

MATERIALS AND METHODS

Heavy strand notation of the 12-component mutational spectrum of mtDNA

Although it is traditional to refer to mtDNA substitutions with respect to the light strand, which corresponds to the reference mtDNA sequences, here we will refer to them based on the complementary heavy strand as has been done previously by several other authors (17,18). The heavy strand was selected since it is more prone to acquiring mutations and thus the nature of most of its nucleotide substitutions would reflect the course of mutageniesis in a more meaningful way (4,10,19). Hereafter, to simplify the biological interpretability of the mtDNA mutational spectrum, we use a 12-component spectrum based on heavy strand notation.

Analysis of duplex sequencing mtDNA data

All data of somatic mtDNA mutations derived from the duplex sequencing approach were obtained from Sanchez-Contreras et al. (19). Control human mtDNA duplex sequencing data were used from obtained from two sources (20,21) with the reported age interval of 10-30 and 80 to 90 years, respectively.

Reconstruction of the species-specific mutational spectrum for mammalian species

Using all available intraspecies sequences (April 2016) of mitochondrial protein-coding genes we derived the mutational spectrum for each species. In brief, we collected all available mtDNA sequences of any protein-coding genes for any chordate species, reconstructed the intraspecies phylogeny using an outgroup sequence (closest species for analysed one), reconstructed ancestral states spectra in all positions at all inner tree nodes and finally got the list of single-nucleotide substitutions for each gene of each species. The pipeline is described in more detail in Supplementary Materials 1.1. Using species with at least 15 single-nucleotide synonymous substitutions at four-fold degenerate sites we estimated the mutational spectrum as the probability of each nucleotide mutating into any other nucleotide (vector of 12 types of substitutions with a sum equals one) for

more than a thousand Chordata species. Four-fold degenerate sites have little or no selective pressure and therefore can be considered the most neutral ones and reflect mutational bias (22,23). Moreover, a recent study revealed that synonymous four-fold degenerate sites are highly associated with mutational bias rather than selection (24). We normalised observed frequencies by nucleotide content in the third position of four-fold degenerative synonymous sites of a given gene. To eliminate the effect of nonuniform sampling of analysed genes between different species, most of our analyses were performed with the *Cytb* gene—the most common gene in our dataset.

Generation length in days (as the average age of parents of the current cohort, reflecting the turnover rate of breeding individuals in a population) for mammals was downloaded from the Dryad data base: https://doi.org/10.5061/dryad.gd0m3 (25).

Analyses of codon asymmetry

XXC_L codon asymmetry was defined as a median value of $XXC_L/(XXC_L + XXT_L)$ from each amino acid. Similarly, XXA_L asymmetry was defined as a median value of $XXA_L/(XXA_L + XXG_L)$.

Analyses of complete mitochondrial genomes

We downloaded whole mitochondrial genomes from GenBank using the following search query: 'Chordata [Organism] AND (complete genome [All Fields] AND mitochondrion [All Fields] AND mitochondrion [filter]'. We extracted non-overlapping regions of protein-coding genes, calculated codon usage and extracted fractions of A_H, T_H, G_H and C_H nucleotides in synonymous four-fold degenerate positions. We estimated the G_HA_H skew as $(G_H-A_H)/(G_H + A_H)$ using only synonymous four-fold degenerate sites for each protein-coding gene of each reference mitochondrial genome of mammalian species.

Analysis of the time spent single-stranded (TSSS)

Based on the asynchronous mode of mtDNA replication and assuming a constant rate of replication by DNA polymerase within major and minor arcs, we calculated relative TSSS for protein-coding genes, coded on a heavy strand of human mtDNA (all except ND6):

$$TSSS\ for\ the\ major\ arc = (GeneLocation - OL) * 2$$

$$TSSS\ for\ the\ minor\ arc = 16569 - (GeneLocation - OL) * 2$$

For all other mammalian species the rank of gene-specific TSSS was the same.

Additional methodological details are presented in Supplementary materials 1-6. All statistical analyses were performed in R.

RESULTS

Frequency of de novo $A_H > G_H$ mutations increases with age in the soma and germline

Mitochondrial genome is characterised by the strong strand asymmetry in the nucleotide content: the heavy strand (H-strand) is guanine rich (G_H) and cytosine poor (C_H), while the light strand (L-strand) is the opposite: cytosine rich (C_L) and guanine poor (G_L). The mutagenic explanation of this asymmetry is based on an assumption that mtDNA heavy strand, being single-stranded during asynchronous replication, is more susceptible to two of the most common mutations in mtDNA: $C_H > T_H$ and $A_H > G_H$ leading to a deficit of C_H and an excess of G_H. Analyses of complete mitochondrial genomes of mammals showed additionally that this nucleotide asymmetry forms a gradient along mtDNA (22,26,27): a global deficit of C_H over T_H and A_H over and G_H at the third codon positions is becoming more pronounced along the major arc from *COX1* to *CYTB* (Figure 1A). This gradient also supports the asynchronous mode of mtDNA replication and a mutagenic effect of the time spent single-stranded (TSSS): the two most common transitions $C_H > T_H$ and $A_H > G_H$ are more frequent in the region of *CYTB* which spent significantly more time in a single-stranded state as compared to *COX1* (17) (Figure 1A). Recently, a large collection of somatic mutations, generated by a highly sensitive duplex sequencing approach, allowed for precise reconstruction of the $C_H > T_H$ and $A_H > G_H$ gradients and unambiguously confirmed the mutagenic effect of TSSS during the asynchronous replication (Figure 1A) (19).

Recent confirmation of such mutational nature of the mtDNA gradients (19) provides a solid ground for further investigations of the mtDNA mutational spectra. It was noted, for example, that the positive G_H/A_H gradient significantly differs between primate species being higher in species with longer gestation time, while T_H/C_H gradient didn't show strong species-specific variations (17). This suggests that the $A_H > G_H$ mutations, shaping the G_H/A_H gradient, can be sensitive to some mutagens associated with gestation time or other life-history traits. Due to the existence of positive correlations between gestation time, body size, and longevity, which in turn are associated with generation length and the level of mitochondrial metabolism (28–30) we can expect differences in mtDNA mutagenesis between species with different life-history traits. To test this hypothesis, we compared datasets of somatic mtDNA mutations of mice and men.

To test the potential sensitivity of the two most common transitions to life-history traits, we reanalysed the recent dataset of somatic *de novo* mtDNA mutations, obtained by the duplex sequencing approach (19). Sanchez-Contreras et al. had three groups of samples: young mice (4–5 months), old mice (26 months) and humans (10–90 years), which allowed the authors to prove that the mutational gradient increases with age. Here we focus on the comparison of magnitudes of $C_H > T_H$ and $A_H > G_H$ gradients. First of all, for comparative purposes, we plotted $C_H > T_H$ and $A_H > G_H$ gradients on the same scale (Figure 1B) and re-ran the linear regression models describing the frequency of mutations

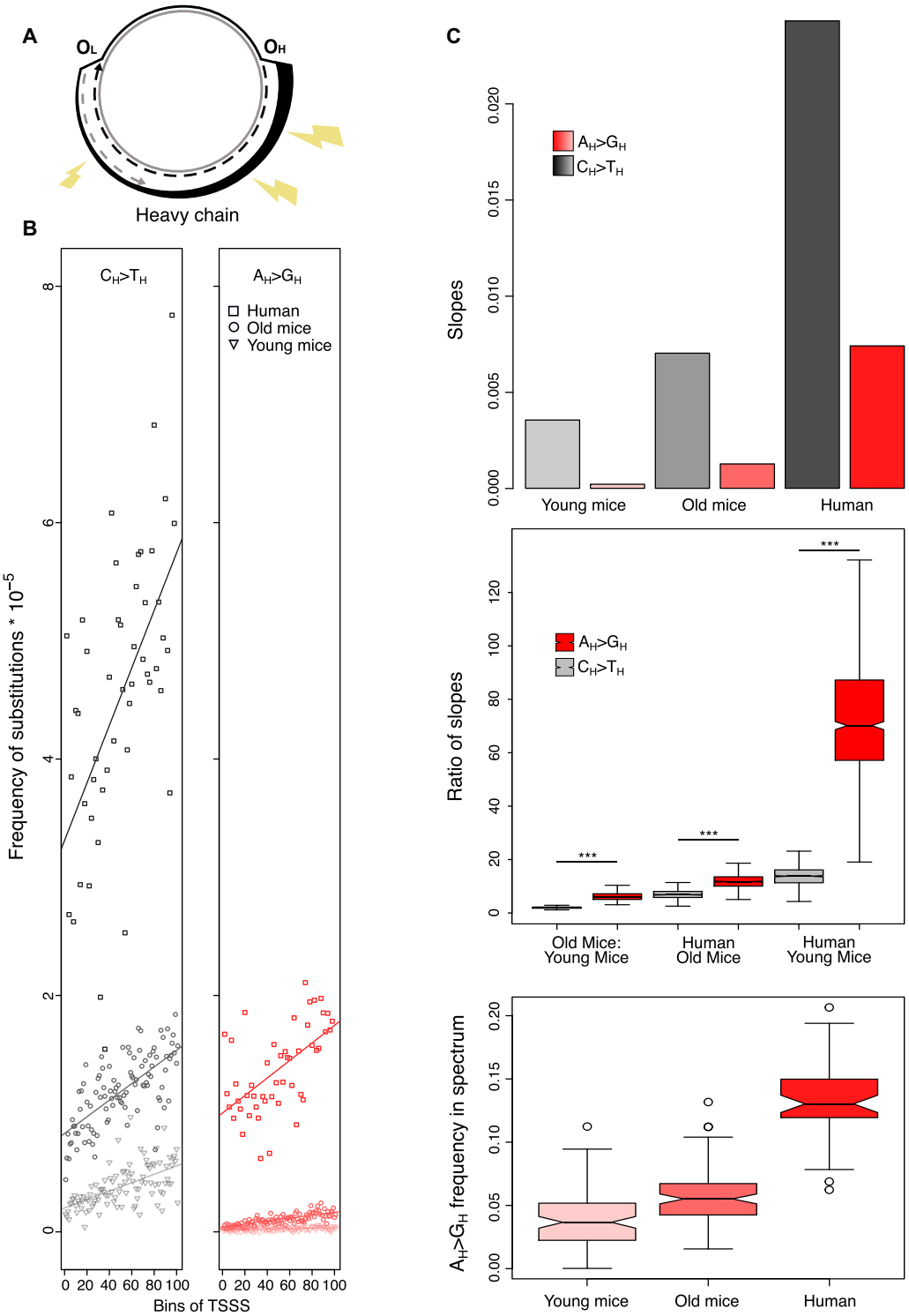

Figure 1. $A_H > G_H$ mtDNA mutational gradient is increasing with the sample age. (**A**) Asynchronous replication of mtDNA is associated with a long time spent single stranded (TSSS) by the parental heavy strand. TSSS in turn is associated with the high frequency of two the most common mtDNA transitions: $C_H > T_H$ and $A_H > G_H$ (daughter heavy strand: dashed black line; parental heavy strand: bold thickening black line reflecting the TSSS; daughter light strand: dashed gray line; parental light strand: solid gray line; OH: origin of replication of daughter heavy strand, OL: origin of replication of daughter light strand). (**B**) Gradients of $C_H > T_H$ and $A_H > G_H$ mutations along the major arc of mtDNA are more pronounced in humans versus old mice and in old mice versus young mice. Both intercepts and slopes are increasing with sample age. (**C**) $A_H > G_H$ substitution rate is increasing faster in aged samples. Upper panel: barplots visualize the slopes of the linear regressions between the mutation frequency and TSSS. Middle panel: $A_H > G_H$ slopes increase faster with age as compared to $C_H > T_H$ slopes. Boxplots are based on the ratio of slopes derived from 1000 bootstrapped samples. Bottom panel: frequency of $A_H > G_H$ in the total mutational spectrum is increasing with age (*P*-values from all three pairwise comparisons are less than 1.583e−08, Mann–Whitney *U* test). '***' marks *P* values < 0.001.

as a function of TSSS (Supplementary Table S1). As expected, and has already bee shown by Sanchez-Contreras et al. (19), both gradients demonstrate increased intercepts and slopes with age (Figure 1B). Second, we focused on the slopes of the linear regressions as a proxy of the sensitivity of mutations to TSSS. We compared the relative increase of slopes from young mice, to old mice, to humans (Figure 1C, upper panel) and observed that the rate of increase of slopes is higher for $A_H > G_H$ as compared to $C_H > T_H$: it is 6.05 (1.98 for $C_H > T_H$) fold higher in old versus young mice, 11.7 (6.9 for $C_H > T_H$) fold higher in humans versus old mice and 70.86 (13.7 for $C_H > T_H$) fold higher in humans versus young mice. One thousand bootstrap resamplings of windows with different TSSS with consequent recalculation of six slopes confirmed the robustness of this result: the increase in $A_H > G_H$ slope with age is higher than the increase in $C_H > T_H$ slope (Figure 1C middle panel, all *P*-values $< 10^{-16}$, Mann–Whitney *U* test). Additional deep resampling of individual molecules (mutated and nonmutated) in each window of each sample confirmed that slopes of $A_H > G_H$ increase with age faster as compared to $C_H > T_H$ (Supplementary Material S1.1, Supplementary Figures S1 and S2, Tables S2 and S3). $A_H > G_H$ mutations, due to the faster increase in slope with age, are expected to contribute proportionally more to the aged samples than $C_H > T_H$. We estimated the total mutational spectrum of young mice, old mice and humans and observed that the total fraction of $A_H > G_H$ is indeed increasing with age (Figure 1C, bottom panel), while the fraction of $C_H > T_H$ shows no monotonic increase with age (Supplementary Figure S3). Altogether, using the datasets of somatic mtDNA mutations, obtained by a highly sensitive duplex sequencing approach, we uncovered that $A_H > G_H$ is more sensitive to age as compared to $C_H > T_H$.

Assuming a similarity of mtDNA mutagenesis in somatic and germ-line tissues we expect to observe also an excess of $A_H > G_H$ in aged germ-line tissues. Indeed, recent deep sequencing of *de novo* mtDNA mutations in aged versus young mice oocytes confirmed that the strongest hallmark of oocyte aging is an increased fraction of $A_H > G_H$ substitutions (31).

Additionally, we analysed the human de novo mtDNA mutations as a function of the female reproductive age, which is a proxy for oocyte age. It has been shown that the number of de novo mtDNA mutations in children increases with maternal age (32,33), however, no age-related changes in mtDNA mutational spectra have been documented yet due to the low sample size of the de novo mutations. We reanalysed all de novo germline mutations from two recent studies of mother-offspring pairs (32,33). Due to low available sample sizes, all the analyses of de novo mutations in mother-offspring pairs were only suggestive but consistently showed a trend of an increased fraction of $A_H > G_H$ with oocyte age (Supplementary Material S1.2). Thus, our results suggest the frequency of de novo $A_H > G_H$ mutations increases with age in both soma and germline of mammals.

Within-species comparisons (old versus young mice or old versus young humans) are expected to be more robust since mtDNA mutagenesis within the same species is most likely very similar. However, assuming that the basic mutagenesis of mtDNA is stable enough across all mammalian species, we extend the logic and perform comparative-species analyses as described in the next section.

$A_H > G_H$s are more prevalent in mammals with high generation length: evidence from polymorphism-derived neutral mutational spectra

The variation in mtDNA mutational spectra between different species (2,3) previously had no overarching explanation. Our findings (Figure 1) and literature data (31) suggest that this variation, and particularly a fraction of mtDNA $A_H > G_H$ transitions, can be associated with aging. Thus, we hypothesize that species-specific mtDNA mutational spectra depend on the generation length which is, in turn, a good proxy of oocyte age in mammals. Because mammalian oocytes are arrested from birth to puberty, which takes weeks (for mice) or decades (for humans) (34) we can use the species-specific generation length as a natural proxy for oocyte longevity in different mammalian species. Additionally, because oocytes are the only lineage through which mtDNA is transmitted (with rare exceptions of paternal inheritance) from generation to generation in mammals (35) we expect to observe a correlation between the species-specific mtDNA properties and the generation length (a proxy for oocyte longevity) in different mammalian species. The generation length is defined as 'the average age of parents of the current cohort' (25,36); it is available for the vast majority of mammalian species (25,36) and is also associated with numerous ecological (body mass, litter size, effective population size) and physiological (basal metabolic rate) parameters of mammalian species (37,38).

Numerous mitochondrial sequences from ecological, evolutionary, and population genetics studies of different species (39) provide a valuable source of mtDNA polymorphisms used in our analyses. Based on our in-house pipeline (Methods and Supplementary Material S2.1) we reconstructed the mutational spectrum of mammalian species. Briefly, we (i) downloaded all available nucleotide sequences of mitochondrial protein-coding genes of mammals, (ii) obtained multiple codon alignment for each gene of each species, (iii) rooted the mitochondrial within-species tree by the nearest neighbor sequence from another species, (iv) reconstructed the ancestral sequences in each inner node, (v) obtained a list of polarized single-nucleotide substitutions and (vi) normalized them by the frequency of ancestral nucleotides. Focusing on the most neutral 70,053 substitutions, located within the 4-fold degenerate synonymous sites, we reconstructed the neutral mutational spectrum for 611 mammalian species (Supplementary Material S2.2 and Supplementary Figure S4). The average mutational spectrum of all mammalian species (Figure 2A) demonstrates strong excess of $C_H > T_H$ and $A_H > G_H$ substitutions which have been shown in previous studies (4,5).

To focus on the species-specific variation in mutational spectra, we analysed in detail the CYTB gene, which was the most common in our database: it contained 56% of all extracted substitutions (39 112 out of 70 053 used to draw Figure 2A). Comparing the CYTB-derived spectrum between

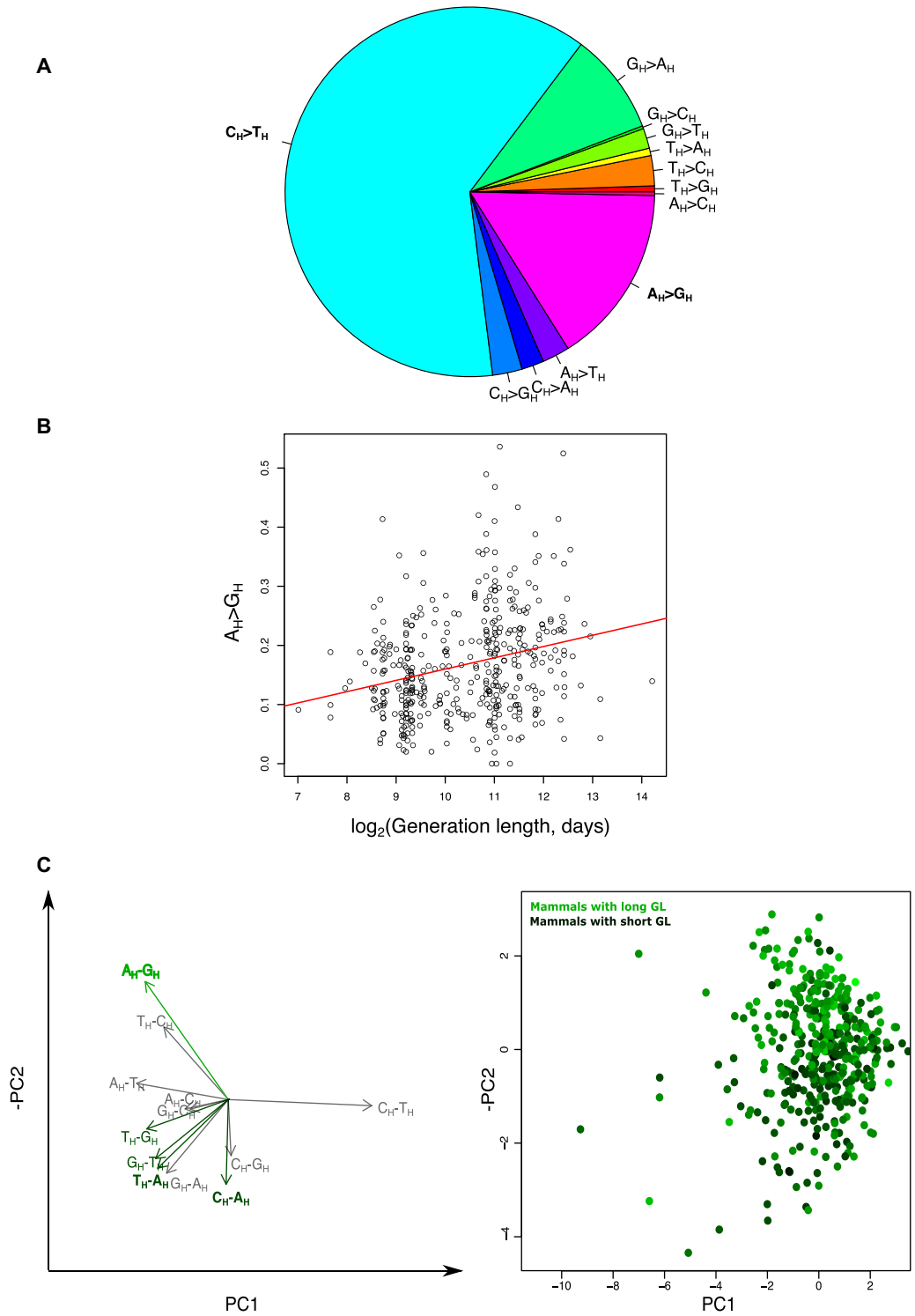

Figure 2. Variation in neutral mtDNA mutational spectrum of mammals is driven by the generation length. (**A**) An average mtDNA mutational spectrum of mammalian species ($N = 611$). Mutational spectrum is a probability of each nucleotide to mutate to each other based on the observed and normalized frequencies of twelve types of nucleotide substitutions in four-fold degenerate synonymous sites of all available within-species polymorphisms of mtDNA protein-coding genes. (**B**) Mutational spectra vary with species-specific generation length ($N = 424$). $A_H > G_H$ is the type of substitutions, frequency of which stronger correlated with the generation length. It shows approximately two-fold difference between the mammalian with very short and very long generation length. (**C**) The principal component analysis (PCA) of mtDNA mutational spectra of mammalian species ($N = 424$). Left panel: the biplot of the principal component analyses (first and the second components explains 16% and 12% of variation correspondingly). $C_H > T_H$ has the highest loading on the first principal component while $A_H > G_H$ has the highest loading on the second principal component. Note that we plotted negative PC2 to make it positively correlated with generation length. Right panel: The second principal component correlates with the generation length in mammals. Generation length is color-coded from dark green (the shortest generation length) to light green (the longest generation length).

species allowed us to eliminate the effect of the gradient (Figure 1) and focus on a potential effect of the life-history traits. As the simplest metric of the mutational spectrum, for each species, we calculated first the Transition/Transversion ratio (Ts/Tv) as the sum of frequencies of all transitions divided by the sum of frequencies of all transversions. For 424 mammalian species with reconstructed Ts/Tv and known generation length, we observed a positive correlation between them (Supplementary Material S2.3, Supplementary Figure S5). Several additional analyses proved the robustness of this correlation: (i) we repeated the same trend splitting all species into several groups by quartiles of the generation length (Supplementary Figure S6), by a median of the generation length (Supplementary Figure S7) and by families (Supplementary Figure S8, Table S4); (ii) we obtained the same trend with phylogenetic aware approach (Supplementary Material S2.4, Supplementary Table S5) and finally (iii) we confirmed the robustness of the results to the total number of polymorphisms used to calculate the mutational spectrum in different species.

To understand further which substitution type(s) predominantly shaped the observed correlation between Ts/Tv and the generation length we performed twelve pairwise rank correlation analyses between each type of substitution and generation length. We observed that only $A_H > G_H$ frequency positively correlated with the generation length (Spearman's rho = 0.252, nominal P-value = 1.188e−07) (Figure 2B), while several rare transversions showed a weak and negative correlation ($T_H > A_H$, $T_H > G_H$, $C_H > A_H$ and $G_H > T_H$: all Spearman's rhos < −0.17, all nominal P-values are < 0.0003). Including all five types of these substitutions into a multiple linear model we showed $A_H > G_H$ being the strongest component associated with generation length (Supplementary Material S2.5). The observed effect was robust to phylogenetic inertia and the number of polymorphisms analyzed in each species (Supplementary Table S6). A wide range of thresholds in the number of polymorphisms used to reconstruct species-specific mutational spectra (from 36 to 156 mutations per species), robustly demonstrated a positive association between $A_H > G_H$ frequency and generation length, despite the significantly different sample size (Supplementary Material S2.6, Supplementary Tables S7 and S8).

To analyze the mtDNA mutational spectra of mammals in an unsupervised way, we performed a dimensionality reduction by reconstructing principal components of mutational spectra of the 424 mammalian species. We observed that the first component is mainly driven by the most common $C_H > T_H$ substitutions, whereas the second is driven mainly by $A_H > G_H$ substitutions (Figure 2C, left panel). We assessed if the first ten principal components were correlated with the generation length and observed that only the second one was significantly correlated with it (Figure 2C right panel, Supplementary Material S2.7). Interestingly, the correlation of the second principal component with generation length was stronger than the sole effect of $A_H > G_H$ frequency, suggesting that the second component could reflect a complex nature of a longevity-associated mutational signature (Supplementary Material S2.5). Additional analyses which take into account the effects of the total number of substitutions (Supplementary Material S2.5), potential sequencing errors (Supplementary Material S2.8, Supplementary Table S9) and nucleotide content (Supplementary Material S2.9, Supplementary Table S10) confirmed the robustness of our results. Altogether, our analyses of within-species four-fold degenerate synonymous polymorphisms in hundreds of mammalian species demonstrated a robust association of mtDNA mutational spectrum (mainly the frequency of $A_H > G_H$) with the species-specific generation length.

MtDNA of mammals with high generation length are more A_H poor and G_H rich due to intensive $A_H > G_H$ mutagenesis

Mutational bias, if stronger than selection, in the long-term perspective is expected to change the genome-wide nucleotide content. Below we test this assumption.

Firstly, to model the possible effect of the mutation bias on the nucleotide composition, we used a computational simulation that derived the expected neutral nucleotide composition based on an input 12-component mutational spectrum. We run this simulation separately for mammals with very short (less than the lower decile: 554 days, $N = 27$) and very long (higher than the upper decile: 5221 days, $N = 25$) generation lengths (Supplementary Table S11, Figure S9); an average mutational spectrum of mammals with very long generation length was characterized by almost two-fold increased frequency of $A_H > G_H$. The results of these simulations demonstrated that an expected nucleotide composition of mammals with a high generation length is characterized by the decreased frequency of A_H. The results of these simulations were confirmed by our analytic solution (Supplementary Material S3.2, Supplementary Material S6). Both approaches (simulations and analytic solution) also confirmed that expected neutral nucleotide composition at equilibrium depends exclusively on the mutational spectrum and doesn't depend on initial conditions (Supplementary Material S3.4; Supplementary Figures S11–S12; Supplementary Material S6). To estimate how close the mammals are to their compositional nucleotide equilibrium we compared the expected nucleotide composition with the observed ones, which was derived using synonymous four-fold degenerate nucleotide content of twelve (all except ND6) protein-coding genes from the same species with very short and very long generations length (Figure 3A) (see similar figure for Cytb gene: Supplementary Figure S10). We found that the observed nucleotide composition is rather similar to the expected one, which means that the analyzed species are close enough to the compositional equilibrium and continue to converge to the equilibrium (Supplementary Material S3.3). Moreover, we observed that species with short generation length tend to be closer to the expected equilibrium (see the horizontal dotted lines in Figure 3A) as compared to species with high generation length, probably because species with short generation length have an increased mutational rate (increased number of mtDNA replications per unit of time) and thus approach an equilibrium faster. Altogether, we observed that the synonymous fourfold degenerate nucleotide composition of mammals is close to their neutral equilibrium and thus we expect to observe an effect of the mutation bias on the nucleotide content of mammalian mtDNA.

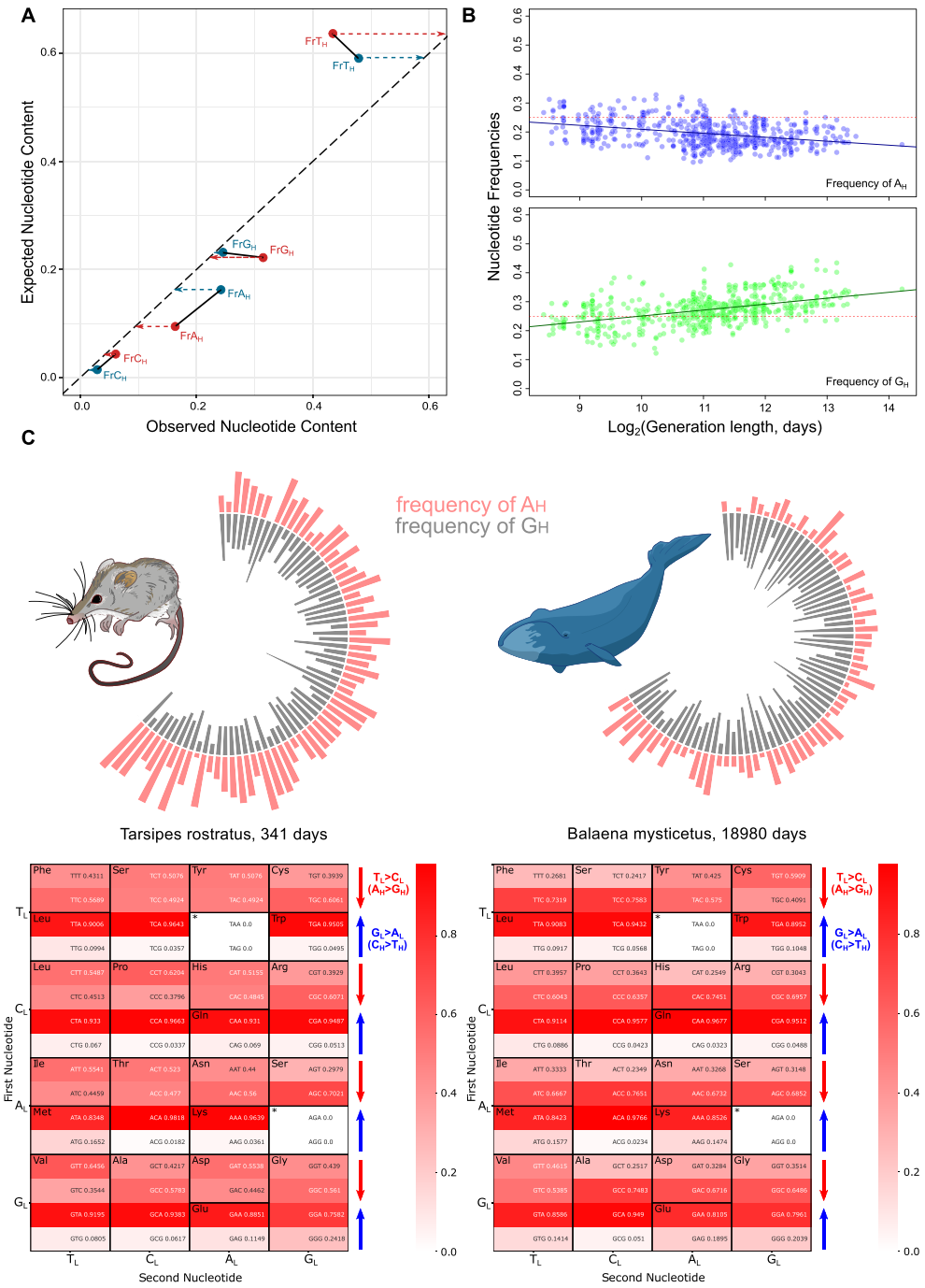

Figure 3. The long-term effect of the mutational bias: neutral nucleotide content in mammalian species. (**A**) A correlation between the expected (obtained in simulations) and observed neutral nucleotide content of mammals with very short and very long generation length. Due to an excess of $A_H > G_H$ substitutions in long-lived mammals (marked by red circles) they are more A_H poor and G_H rich for both expected and observed values. A location of all data points (red and blue circles) near the diagonal shows that mtDNA of mammals is close enough to a neutral equilibrium. However, short-lived species (marked by blue circles) are even closer to the diagonal (horizontal dotted blue line towards the diagonal are shorter than the red dotted lines), suggesting that they are evolving faster towards the neutral equilibrium. (**B**) Nucleotide frequencies in neutral sites of all 13 protein-coding genes as a function of generation length—fraction of A_H is decreasing while fraction of G_H is increasing ($N = 650$). (**C**) Structure of mtDNA of two mammalian species with extreme generation lengths: a honey possum and a whale. Upper panel: frequencies of A_H (red) and G_H (grey) nucleotides along the major arc of mtDNA of the most short-lived (honey possum) and the most long-lived (whale) mammalian species from our dataset. Each bar represents the nucleotide frequency in a 20-nucleotide window. In both mammals, A_H is decreasing and G_H is increasing along the major arc of mtDNA: from the bottom left (origin of replication of light strand) to the top right (origin of replication of heavy strand). However, additionally to the gradient, mtDNA of a whale has an integral, genome-wide, deficit of A_H and excess of G_H - a signature of an increased generation length. Bottom panel: heatmaps visualize asymmetry of the codon usage of 12 protein-coding genes (all except ND6). Whale is more contrast than honey possum in terms of an asymmetry driven by the age-related $T_L > C_L$ ($A_H > G_H$) substitutions. Heatmaps of both species are equally contrasted in terms of an asymmetry driven by $G_L > A_L$ ($C_H > T_H$) substitutions, which have high and similar (not age related) substitution rate in both species.

Secondly, we tested if the increased $A_H > G_H$ (Figure 2) in species with high generation length would decrease A_H's frequencies and increase G_H's frequencies in the corresponding reference sequences. Since generation length correlates with the strength of the $A_H > G_H$ (Figure 2) we expect that generation length should demonstrate a positive correlation with G_H and a negative one with A_H. Testing all four pairwise correlations between the species-specific generation length and the nucleotide content (A_H, T_H, G_H, C_H) we observed two strongest correlations: negative with A_H and positive with G_H (Figure 3B; Supplementary Table S12). Including all four types of nucleotide frequencies in the multiple linear model confirmed the importance of A_H and G_H only, the effect of which was also robust to the phylogenetic inertia (Supplementary Table S13). Thus, we concluded that mtDNA of mammals with long- versus short-generation length is more A_H poor and G_H rich (Figure 3B), which is in line with the more intensive $A_H > G_H$ mutagenesis in the former (Figure 2).

Third, we tested whether an excess of G_H and deficit of A_H in long-lived species determines the positive G_HA_H nucleotide skew. The G_HA_H nucleotide skew approximates the level of asymmetry in the distribution of these two nucleotides and is calculated as $(G_H - A_H)/(G_H + A_H)$. Based on four-fold degenerate synonymous positions of 12 genes (all except ND6) we estimated the G_HA_H skew for each mammalian species and correlated it with the generation length. As expected, we obtained a positive correlation (phylogenetic generalized least squares: coefficient = 0.13, *P*-value = 2.9×10^{-4}; see also Figure 3C). To visualize a contrast in G_HA_H skew between the shortest- and the longest-lived species in our dataset we plotted A_H and G_H fractions along the major arc of mtDNA for honey possum (generation length 341 days) and whale (generation length 18980 days) (Figure 3C). It is evident that on average honey possum mtDNA has an excess of A_H (red color in Figure 3C) while whale has an excess of G_H (gray color in Figure 3C).

Fourth, we analyzed how $A_H > G_H$ mutagenesis affects asymmetry in codon usage. If $A_H > G_H$ is a strong and uniform mutational force, we expect that most amino acids would demonstrate a deficit of XXA_H and excess of XXG_H codons. Because the light strand of mtDNA is equivalent to mRNA of 12 protein-coding genes (all except ND6) we analyze codon usage in terms of the light strand notation and thus, $A_H > G_H$ substitutions are expected to decrease the frequency of XXT_L codons (because T_L is complementary to A_H) and increase the frequency of XXC_L codons (because C_L is complementary to G_H). Similarly, $C_H > T_H$ mutations are expected to increase frequencies of XXA_L and decrease frequencies of XXG_L codons. To visualize this effect we plotted heatmaps of the codon usage of honey possum and whale (based on 12 genes, all except ND6) (Figure 3C, bottom panels). We observed that XXC_L asymmetry (see Methods) is indeed stronger in a whale (0.70) as compared to a honey possum (0.52). Interestingly, XXA_L asymmetry (see Methods) is similarly high in both these species (whale: 0.93, honey possum: 0.95). Quantifying the XXC_L asymmetry and XXA_L asymmetry across all mammalian species with complete mitochondrial genomes, we observed that (i) both XXC_L and XXA_L are significantly higher than the null expectation of 0.5 (Supplementary Figure S13); (ii) XXA_L asymmetry is much stronger as compared to XXC_L asymmetry (Supplementary Table S14), supporting that the $C_H > T_H$ substitution rate is much higher as compared to $A_H > G_H$ (see also Figure 2A) and (iii) XXC_L asymmetry correlates strongly and positively with the generation length of mammals, while XXA_L demonstrates weak negative correlation, which is not supported by the phylogeny aware statistics (Supplementary Table S15), supporting that only $A_H > G_H$ substitutions (those which affect XXC_L asymmetry) are associated with the generation length.

Altogether we demonstrated that $A_H > G_H$ mutagenesis, which is more pronounced in long-lived species, strongly shapes its reference sequences: nucleotide content (low A_H and high G_H frequency), nucleotide skew (strong positive G_HA_H skew) and the codon usage (positive XXC_L asymmetry).

G_HA_H nucleotide skew is a function of both time spent single-stranded (TSSS) and the generation length

It has been shown that the frequency of $A_H > G_H$ substitutions depends on how much Time the parental heavy strand Spent in a Single-Stranded (TSSS) condition during asynchronous mtDNA replication. Genes, located close to the origin of light strand replication (O_L), such as *COX1*, spend minimal time being single-stranded and demonstrate the low frequency of $A_H > G_H$, while genes located far away from O_L spend more time being single-stranded and demonstrate correspondingly higher frequencies of $A_H > G_H$ (Figure 1). Thus, we expect that the effectively neutral nucleotide composition of mtDNA is a function of both: gene-specific TSSS and species-specific generation length. To test this, we derived for each gene of each species the G_HA_H skew and split all mammalian species into species with short and long generation length according to the median (median = 2190 days, *N* short = 325, *N* long = 319). Next, we plotted G_HA_H skew of mammals with short and long generation lengths for each gene, ranking them along the major arc from *COX1* (rank equals 1) to *CYTB* (rank equals 10), corresponding to the increasing TSSS. As expected, we observed that G_HA_H skew increases with both gene-specific TSSS and species-specific generation length (Figure 4A). Performing multiple linear models, where G_HA_H skew is a function of both TSSS and generation length, we confirmed that both factors affected the skew, moreover, to a very similar degree (Supplementary Table S16). Genes, located close to the origin of light strand replication (O_L), such as *COX1*, spend minimal time being single-stranded and demonstrate the low frequency of $A_H > G_H$, while genes located far away from O_L spend more time being single-stranded and demonstrate correspondingly higher frequencies of $A_H > G_H$ (Figure 1) (19,22). Thus, we expect that effectively neutral nucleotide composition of mtDNA is a function of both: gene-specific TSSS and species-specific generation length. To test this, we derived for each gene of each species the G_HA_H skew and split all mammalian species into species with short and long generation length according to the median (median = 2190 days, *N* short = 325, *N* long = 319). Next, we plotted G_HA_H

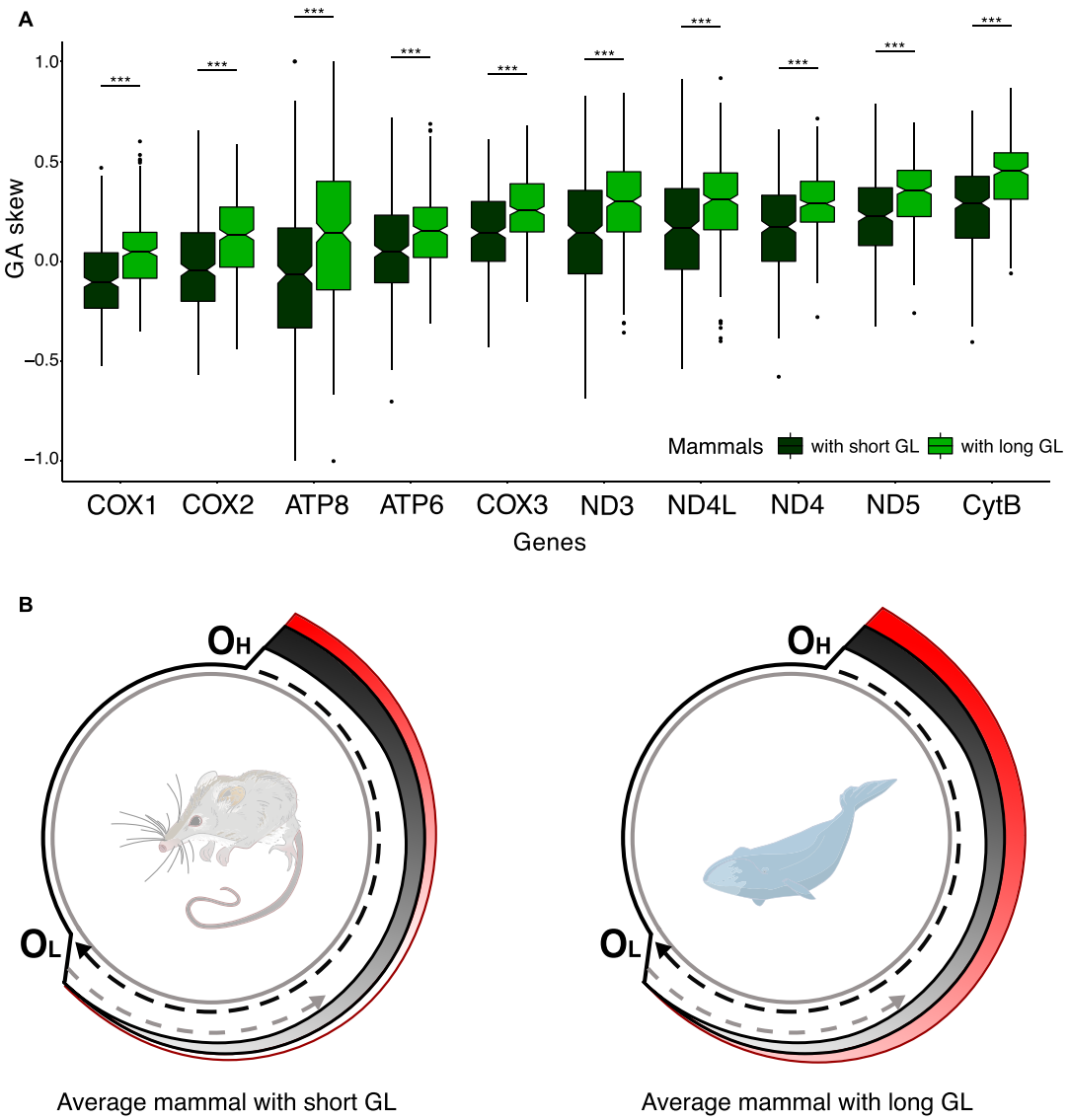

Figure 4. (**A**) Changes in nucleotide content along mtDNA of short- and long-lived mammals ($N = 650$). All genes (except for ND6) located in the major arc are ranked according to the time spent single stranded: from COX1 to CYTB. Pairs of boxplots for each gene represent G_HA_H skew for short- and long-lived mammals splitted by the median generation length. G_HA_H is increasing with both gene-specific TSSS and the species-specific generation length. (**B**) A visual summary of the main finding: $A_H > G_H$ substitution rate (marked as red gradient) is increasing with both gene-specific TSSS and the species-specific generation length. The effect size of the G_L is comparable with the effect size of TSSS. $C_H > T_H$ substitution rate (marked as grey gradient) is sensitive to TSSS only.

skew of mammals with short and long generation length for each gene, ranking them along the major arc from *COX1* (rank equals 1) to *CYTB* (rank equals 10), corresponding to the increasing TSSS. As expected, we observed that G_HA_H skew increases with both gene-specific TSSS and species-specific generation length (Figure 4A). Performing multiple linear models, where G_HA_H skew is a function of both TSSS and generation length, we confirmed that both factors affected the skew, moreover, the effect sizes of TSSS and generation length were very similar (Supplementary Table S17).

In the linear models, we did not observe a significant interaction between TSSS and the generation length, suggesting that either these factors affect nucleotide composition independently of each other (Supplementary Material S4.1) or the interaction signal is too weak to be significant with our sample size. Our specific analyses, which focused on uncovering a potential interaction between TSSS and generation length indeed showed a positive trend, suggesting that mammals with high generation length demonstrate a faster decrease in A_H and increase in G_H along the genome (Supplementary Figure S14, the same effect is visible on Figure 3B). Faster changes (stronger gradients) in A_H and G_H along the major arc of mtDNA in long-lived mammals can be interpreted as an interaction between TSSS and generation length as if the substitution rate $A_H > G_H$ in-

creases faster as a function of TSSS in case of high generation length. Altogether, our results demonstrate that the nucleotide content, shaped by the mutational bias from A_H to G_H, positively and strongly depends on both TSSS and the generation length.

DISCUSSION

We observed that $A_H > G_H$ substitution rate and its consequences such as nucleotide frequencies, G_HA_H skew, and codon usage asymmetry increase with organismal longevity. This finding was repeated on three rather different time scales: (i) months and years for somatic and de novo germline mutations (Figure 1); (ii) dozens or hundreds of thousands of years for neutral within-species mtDNA polymorphisms (which is the average time of segregation of neutral within-species mtDNA polymorphisms (40)) (Figure 2); (iii) millions of years for neutral mtDNA substitutions fixed between species (Figures 3 and 4). This trend, universal in time and across all mammalian species, requires a special explanation.

We consider that a process of mtDNA mutagenesis, rather than selection, is primarily responsible for our findings. Firstly, we expect no effect of selection in case of extremely rare, with variant allele frequency less than 1%, mtDNA mutations, called in duplex sequencing approach (19) (Figure 1). Secondly, due to little or no evidence of selection on synonymous four-fold degenerate sites in mtDNA of mammals (22,23) we consider the polymorphic variants (Figure 2) as well as the variants fixed between mammalian species (Figures 3 and 4) as effectively neutral.

The mitochondrial mutational spectrum, and particularly $A_H > G_H$ substitutions, can be affected by both errors of DNA polymerase gamma and damage associated with the mitochondrial microenvironment. A recent elegant experiment with mice homozygous for exonuclease deficient gamma DNA polymerase showed that $A_H > G_H$ and $C_H > T_H$ gradients are mainly shaped by an exogenous mutagen associated with asynchronous replication of DNA rather than by DNA polymerase mistakes (19). Considering that reactive oxygen species are the main deleterious byproduct of aerobic metabolism, we hypothesize that our key findings can be attributed to the effects of oxidative damage.

Our hypothesis is based on the high sensitivity of A > G to TSSS: single-stranded DNA is more vulnerable to DNA damaging agents (41). DNA exists in single-stranded form during the process of replication, transcription, and DNA repair (41). A mutational gradient along the major and minor arcs of mtDNA (19) suggests that TSSS during replication is more mutagenic in the case of mtDNA than the TSSS during transcription and reparation. However, non-zero intercepts observed for both common transitions when replication-driven TSSS equals zero (19) (see also Figure 1) suggest that background mutagenesis, probably associated with transcription or reparation-driven TSSS can play some role.

Aging is associated with multiple changes in cellular metabolism. For instance, tissues of aged mice produce less ATP and have lower NAD+ levels than the corresponding young mice tissues (42). Given that mitochondria is a central hub of energy metabolism, its functioning also changes with age (43). However, it is still unclear if there is a universal pattern of metabolic changes during aging. Here we showed that mtDNA mutagenesis depends on organism age and species-specific generation length implying some common changes in the mitochondrial microenvironment. We suggest that oxidative damage can be a universal factor that aggravates with age. Indeed, the protein carbonylation level is higher in the fibroblasts of old than young people (44), and lipid peroxidation is associated with many age-related diseases (45). Moreover, it is assumed that the mitochondria is a major source of age-related oxidative damage: expression of mitochondria-targeted antioxidant enzyme—catalase—increases murine lifespan (46).

It is important to emphasize, that our discovered substitution differs from the well-documented signature of the reactive oxygen species (ROS): G > T transversions which are results of ROS-induced guanosine modification 7,8-dihydro-8-oxo-20-deoxyguanosine (8-oxodG). It has been shown before that this expected hallmark of oxidative damage, namely G > T substitutions, does not occur in mtDNA: it is very rare, age-independent and demonstrates weak if any association with oxidative damage (4,10,12,21). Interestingly, G > T transversions ($G_H > T_H$ and $C_H > A_H$) in our work, despite the rareness and weak effect, demonstrate an opposite to $A_H > G_H$ trend (Figure 2C, left panel), which may reflect the classical oxidative damage signature (ROS-induced 8-oxodG) more pronounced in short-lived species due to their higher relative basal metabolic rate. However, its effect indeed is weak.

Here, instead of G > T substitutions, we propose that the deamination of adenosine, which is the main source of $A_H > G_H$ substitutions, is associated with oxidative damage of a single-stranded mtDNA. We propose that A > G is a novel mtDNA marker of oxidative damage typical for single-stranded DNA. Our results (Figures 1–4) and earlier publications (discussed below) support our hypothesis:

1. Recent deep sequencing of de novo mtDNA mutations in aged versus young mice oocytes confirmed that an increased fraction of $A_H > G_H$ substitutions is the best mtDNA hallmark of oocyte aging (31). This discovery also suggests that continuous mtDNA turnover within the dormant oocytes can be a universal trait of all mammalian species (31)—and thus variation in mtDNA mutational spectra between different species can reflect the age-associated processes in different mammalian species.

2. An excess of A:T > G:C substitutions has been observed recently in aerobically versus anaerobically grown *Escherichia coli*. It is important to emphasize that the effect was driven by the lagging strand, spending more time in single-stranded conditions (47). These experiment results are compatible with our hypothesis that A > G transitions are associated with oxidative damage of single-stranded DNA.

3. Evidence that oxidative stress can cause A:T > G:C mutations comes from eukaryotes, i.e. yeast (48) and mice (49). In both studies apparent reduction of oxidative stress (by imposing aerobic conditions or use of antioxidants, respectively) resulted in a specific reduction of the A:T > G:C rates, implying that a portion of these mutations results from oxidative damage. Interestingly, both studies used mismatch repair deficient systems, which resulted in the marked

increase of A:T > G:C mutations compared to MMR-proficient ones. A large proportion of this MMR-dependent increase should be oxidative-stress dependent. This potentially implies MMR, in addition to its traditional role in repairing replication errors, in repairing oxidative lesions. This is consistent with a growing body of evidence from other studies (reviewed in (50)). This intriguing area remains controversial and requires more research.

4. It has been shown that A > G is the most asymmetric substitution in the human nuclear genome, which is a hallmark of a mutagen acting via strand-specific DNA damage (51). Even though the key mutagen still has not been determined, we suggest that it can be associated with oxidative damage in the case of mtDNA.

5. It has been shown that the fraction of $A_H > G_H$ substitutions in mtDNA positively correlates with the ambient temperature of *Actinopterygii* species (52). Because higher temperature is associated with increased aerobic metabolism (53), we assume that $A_H > G_H$ is a marker of oxidative damage.

6. A > G substitutions, associated with oxidative damage, can be a key process, explaining a long-standing evolutionary puzzle of the increased GC content of aerobic versus anaerobic bacteria (54–56). According to the conventional knowledge, G > T is expected to be higher in aerobic versus anaerobic bacteria, making genomes of the former more GC poor and AT rich. However, numerous pieces of empirical evidence show the opposite: aerobic bacteria have an increased GC content as compared to the anaerobic ones (54–56). Assuming that oxidative damage has similar mutagenic effects on both mitochondrial and bacterial genomes, we hypothesize that an extensive A > G substitution rate in aerobic bacteria can make their genomes more G-rich.

An excess of G_H in mtDNA of long-lived mammals has been previously shown by Lehmann et al.(1). They proposed a selection-based explanation of this observation assuming increased stability of the G_H-rich genomes which may confer an advantage to long-lived mammals. Our findings demonstrate that an excess of G_H in long-lived mammals may be a neutral consequence of $A_H > G_H$ mutagenesis, not a result of selection-driven mechanism (Figures 3 and 4). Additionally, the low effective population size of long-lived mammals increases the strength of the random genetic drift's strength and the fixation rate of slightly-deleterious variants in their mtDNA (57–59) making the explanation based on selection even less probable. However, to evaluate the selection-based argument additionally, we analysed the correlations of the nucleotide content (A, T, G, C) with generation length for each of three nucleotide positions in codons separately. We observed that A_H and G_H nucleotide frequencies at the third - the most neutral position of codons show the strongest correlations (negative for A_H and positive for G_H) with generation length (Supplementary Material S5.1, Supplementary Table S18). Although we cannot rule out the selection-based argument (if selection acts on overall nucleotide composition to increase GC content, we can expect the strongest effect at the 3rd codon positions being the least constrained and further research is needed to deconvolute the mutagenesis and selection, our result suggests that mutagenesis can also explain the observed excess of G_H in long-lived mammals.

Overall, our results offer a valuable possibility to expand the usability of mtDNA mutational spectra in a considerable way. For example, low-heteroplasmy somatic mtDNA mutations from a merely neutral marker used to trace cellular lineages (60) can be transformed to a metric, associated with a cell-specific aging state which can be especially important in highly heterogeneous tissues such as cancers. Low-heteroplasmy de novo germline mtDNA mutations (31) can predict the biological age of human oocytes, which can be used in IVF (In Vitro Fertilization) techniques. MtDNA mutational spectra, reconstructed for non-model chordate species, may help to approximate their average generation length - a metric that is not always straightforward to estimate empirically.

In the current work we cannot prove unambiguously that $A_H > G_H$ substitutions are driven by oxidative damage (instead it can be another age-related mutagen). By putting together several observations that collectively corroborate our assertions, we are proposing the working hypothesis, which is testable and can be evaluated by the scientific community in the future. There is no well established experimental link between oxidative damage and A→G yet, so we would like to offer a plausible speculation. Interestingly, the product of oxidative (or hydrolytic) ssDNA-specific deamination of adenosine is inosine, which preferentially pairs with C and therefore induces A > G/T > C mutations (61). Steady state levels of inosine in DNA were reported on the order of 1 per 1 000 000 nucleotides (62). Intriguingly, this level is not incompatible with the observed A > G/T > C mutational fractions in aged tissues, which is reported to be about 1 mutation per 100 000 nucleotides (10). Of note, with 10^{-6} steady state level of inosine, we are expecting mutational rate of 10^{-6} per DNA duplication (i.e. 10^{-5} per 10 duplications), which is more than enough to account for the observed mutant fractions, as cell lineages in aged tissues go through >40 duplications. In reality, mutation rates are affected by many unknowns such as levels of damage specifically in the replicating pool of mtDNAs. Also the above levels of inosine were reported for nuclear, not mitochondrial, DNA. Nevertheless, these rough estimates imply that inosine may be considered a promising candidate for the link between oxidative stress and asymmetric A > G/T > C mutations. We hope that our study will prompt further research in this field. Direct experimental investigation of the effect of different mutagens (9) and especially an effect of oxidative damage (63) and chemical modifications of nucleotides (64,65) on single-stranded mtDNA will significantly improve our understanding of mtDNA mutagenesis.

Altogether, we demonstrated that A > G substitutions depend on both TSSS and generation length and this relationship can be mediated by the sensitivity of this type of substitution to the oxidative damage of the single-stranded DNA (Figure 4B).

DATA AVAILABILITY

All data and scripts are available in our public GitHub repository: https://github.com/polarsong/mtDNA_mutspectrum/.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Athanasios Kousathanas for comments on statistical tools, the whole laboratory of Alexandre Reymond, Mikhail Gelfand and Vladimir Katanaev for valuable discussions; Alexander Voronka and Nikita Van Leiden for helpful suggestions.

FUNDING

The design of the study by K.P. and data processing by V.L, D.I, I.K. were supported by the federal academic leadership program Priority 2030 at the Immanuel Kant Baltic Federal University. E.O.T. was supported by a scholarship from the Austrian Science Fund (FWF, DOC 33-B27). Population and evolutionary genetics analysis by A.G.M. and V.Y. were supported by the Russian Science Foundation grant No. 21-75-20143 and analysis of somatic and germline mtDNA mutagenesis by I.M. was supported by the Russian Science Foundation No. 21-75-10081. Analysis of the de novo mtDNA mutations as a function of the female reproductive age by K.G. was supported by the Russian Science Foundation grant No. 21-75-20145; K.K., Z.F., S.A. and M.F. were supported in part by the National Institutes of Health [R01-HD091439]. Funding for open access charge: the federal academic leadership program Priority 2030 at the Immanuel Kant Baltic Federal University.
Conflict of interest statement. None declared.

REFERENCES

1. Lehmann,G., Budovsky,A., Muradian,K.K. and Fraifeld,V.E. (2006) Mitochondrial genome anatomy and species-specific lifespan. *Rejuvenation Res.*, **9**, 223–226.
2. Belle,E.M.S., Piganeau,G., Gardner,M. and Eyre-Walker,A. (2005) An investigation of the variation in the transition bias among various animal mitochondrial DNA. *Gene*, **355**, 58–66.
3. Montooth,K.L. and Rand,D.M. (2008) The spectrum of mitochondrial mutation differs across species. *PLoS Biol.*, **6**, e213.
4. Yuan,Y. and PCAWG ConsortiumPCAWG Consortium, Ju,Y.S., Kim,Y., Li,J., Wang,Y., Yoon,C.J., Yang,Y., Martincorena,I. *et al.* (2020) Comprehensive molecular characterization of mitochondrial genomes in human cancers. *Nat. Genet.*, **52**, 342–352.
5. Ju,Y.S., Alexandrov,L.B., Gerstung,M., Martincorena,I., Nik-Zainal,S., Ramakrishna,M., Davies,H.R., Papaemmanuil,E., Gundem,G., Shlien,A. *et al.* (2014) Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer. *Elife*, **3**, e24284.
6. Alexeyev,M.F. (2009) Is there more to aging than mitochondrial DNA and reactive oxygen species?*FEBS J.*, **276**, 5768–5787.
7. Fraga,C.G., Shigenaga,M.K., Park,J.W., Degan,P. and Ames,B.N. (1990) Oxidative damage to DNA during aging: 8-hydroxy-2′-deoxyguanosine in rat organ DNA and urine. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 4533–4537.
8. Alexandrov,L.B., Nik-Zainal,S., Wedge,D.C., Aparicio,S.A.J.R., Behjati,S., Biankin,A.V., Bignell,G.R., Bolli,N., Borg,A., Børresen-Dale,A.-L. *et al.* (2013) Signatures of mutational processes in human cancer. *Nature*, **500**, 415–421.
9. Kucab,J.E., Zou,X., Morganella,S., Joel,M., Nanda,A.S., Nagy,E., Gomez,C., Degasperi,A., Harris,R., Jackson,S.P. *et al.* (2019) A compendium of mutational signatures of environmental agents. *Cell*, **177**, 821–836.
10. Kennedy,S.R., Salk,J.J., Schmitt,M.W. and Loeb,L.A. (2013) Ultra-sensitive sequencing reveals an age-related increase in somatic mitochondrial mutations that are inconsistent with oxidative damage. *PLos Genet.*, **9**, e1003794.
11. Gouliaeva,N.A., Kuznetsova,E.A. and Gaziev,A.I. (2006) Proteins associated with mitochondrial DNA protect it against X-rays and hydrogen peroxide. *Biophysics*, **51**, 620–623.
12. Zsurka,G., Peeva,V., Kotlyar,A. and Kunz,W.S. (2018) Is there still any role for oxidative stress in mitochondrial DNA-Dependent aging?*Genes*, **9**, 175.
13. Koh,G., Degasperi,A., Zou,X., Momen,S. and Nik-Zainal,S. (2021) Mutational signatures: emerging concepts, caveats and clinical applications. *Nat. Rev. Cancer*, **21**, 619–637.
14. Zou,X., Koh,G.C.C., Nanda,A.S., Degasperi,A., Urgo,K., Roumeliotis,T.I., Agu,C.A., Badja,C., Momen,S., Young,J. *et al.* (2021) A systematic CRISPR screen defines mutational mechanisms underpinning signatures caused by replication errors and endogenous DNA damage. *Nat Cancer*, **2**, 643–657.
15. Harris,K. and Pritchard,J.K. (2017) Rapid evolution of the human mutation spectrum. *Elife*, **6**, e24284.
16. Moorjani,P., Amorim,C.E.G., Arndt,P.F. and Przeworski,M. (2016) Variation in the molecular clock of primates. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 10607–10612.
17. Raina,S.Z., Faith,J.J., Disotell,T.R., Seligmann,H., Stewart,C.-B. and Pollock,D.D. (2005) Evolution of base-substitution gradients in primate mitochondrial genomes. *Genome Res.*, **15**, 665–673.
18. Krishnan,N.M., Seligmann,H., Raina,S.Z. and Pollock,D.D. (2004) Detecting gradients of asymmetry in site-specific substitutions in mitochondrial genomes. *DNA Cell Biol.*, **23**, 707–714.
19. Sanchez-Contreras,M., Sweetwyne,M.T., Kohrn,B.F., Tsantilas,K.A., Hipp,M.J., Schmidt,E.K., Fredrickson,J., Whitson,J.A., Campbell,M.D., Rabinovitch,P.S. *et al.* (2021) A replication-linked mutational gradient drives somatic mutation accumulation and influences germline polymorphisms and genome composition in mitochondrial DNA. *Nucleic Acids Res.*, **49**, 11103–11118.
20. Baker,K.T., Nachmanson,D., Kumar,S., Emond,M.J., Ussakli,C., Brentnall,T.A., Kennedy,S.R. and Risques,R.A. (2019) Mitochondrial DNA Mutations are Associated with Ulcerative Colitis Preneoplasia but Tend to be Negatively Selected in Cancer. *Mol. Cancer*, **17**, 488–498.
21. Hoekstra,J.G., Hipp,M.J., Montine,T.J. and Kennedy,S.R. (2016) Mitochondrial DNA mutations increase in early stage Alzheimer disease and are inconsistent with oxidative damage. *Ann. Neurol.*, **80**, 301–306.
22. Faith,J.J. and Pollock,D.D. (2003) Likelihood analysis of asymmetrical mutation bias gradients in vertebrate mitochondrial genomes. *Genetics*, **165**, 735–745.
23. Uddin,A. and Chakraborty,S. (2017) Synonymous codon usage pattern in mitochondrial CYB gene in pisces, aves, and mammals. *Mitochondrial DNA*, **28**, 187–196.
24. Dunn,C.D. (2021) The population frequency of human mitochondrial DNA variants is highly dependent upon mutational bias. *Biol. Open*, **10**, bio059072.
25. Pacifici,M., Santini,L., Di Marco,M., Baisero,D., Francucci,L., Marasini,G.G., Visconti,P. and Rondinini,C. (2013) Generation length for mammals. *Nat. Conserv.*, **5**, 87–94.
26. Reyes,A., Gissi,C., Pesole,G. and Saccone,C. (1998) Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. *Mol. Biol. Evol.*, **15**, 957–966.
27. Tanaka,M. and Ozawa,T. (1994) Strand asymmetry in human mitochondrial DNA mutations. *Genomics*, **22**, 327–335.
28. Polishchuk,L.V. and Tseitlin,V.B. (1999) Scaling of population density on body mass and a number-size trade-off. *Oikos*, **86**, 544–556.
29. Damuth,J. (1981) Population density and body size in mammals. *Nature*, **290**, 699–700.
30. White,C.R. and Seymour,R.S. (2005) Allometric scaling of mammalian metabolism. *J. Exp. Biol.*, **208**, 1611–1619.
31. Arbeithuber,B., Hester,J., Cremona,M.A., Stoler,N., Zaidi,A., Higgins,B., Anthony,K., Chiaromonte,F., Diaz,F.J. and Makova,K.D. (2020) Age-related accumulation of de novo mitochondrial mutations in mammalian oocytes and somatic tissues. *PLoS Biol.*, **18**, e3000745.
32. Rebolledo-Jaramillo,B., Su,M.S.-W., Stoler,N., McElhoe,J.A., Dickins,B., Blankenberg,D., Korneliussen,T.S., Chiaromonte,F.,

Nielsen,R., Holland,M.M. *et al.* (2014) Maternal age effect and severe germ-line bottleneck in the inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 15474–15479.
33. Wei,W., Tuna,S., Keogh,M.J., Smith,K.R., Aitman,T.J., Beales,P.L., Bennett,D.L., Gale,D.P., Bitner-Glindzicz,M.A.K., Black,G.C. *et al.* (2019) Germline selection shapes human mitochondrial DNA diversity. *Science*, **364**, eaau6520.
34. Von Stetina,J.R. and Orr-Weaver,T.L. (2011) Developmental control of oocyte maturation and egg activation in metazoan models. *Cold Spring Harb. Perspect. Biol.*, **3**, a005553.
35. Sato,K. and Sato,M. (2017) Multiple ways to prevent transmission of paternal mitochondrial DNA for maternal inheritance in animals. *J. Biochem.*, **162**, 247–253.
36. Tacutu,R., Craig,T., Budovsky,A., Wuttke,D., Lehmann,G., Taranukha,D., Costa,J., Fraifeld,V.E. and de Magalhães,J.P. (2013) Human ageing genomic resources: integrated databases and tools for the biology and genetics of ageing. *Nucleic Acids Res.*, **41**, D1027–D33.
37. Ollason,J.G. (1987) R. H. Peters 1986. The Ecological Implications of Body Size. Cambridge University Press, Cambridge. 329 pages. ISBN 0-521-2886-x. Price: £12.50, US$16.95 (paperback). *J. Trop. Ecol.*, **3**, 286–287.
38. Damuth,J. (1987) Interspecific allometry of population density in mammals and other animals: the independence of body mass and population energy-use. *Biol. J. Linn. Soc.*, **31**, 193–246.
39. Hebert,P.D.N., Cywinska,A., Ball,S.L. and deWaard,J.R. (2003) Biological identifications through DNA barcodes. *Proc. Biol. Sci.*, **270**, 313–321.
40. Atkinson,Q.D., Gray,R.D. and Drummond,A.J. (2008) mtDNA variation predicts population size in humans and reveals a major southern asian chapter in human prehistory. *Mol. Biol. Evol.*, **25**, 468–474.
41. Saini,N. and Gordenin,D.A. (2020) Hypermutation in single-stranded DNA. *DNA Repair (Amst.)*, **91-92**, 102868.
42. Gomes,A.P., Price,N.L., Ling,A.J.Y., Moslehi,J.J., Montgomery,M.K., Rajman,L., White,J.P., Teodoro,J.S., Wrann,C.D., Hubbard,B.P. *et al.* (2013) Declining NAD(+) induces a pseudohypoxic state disrupting nuclear-mitochondrial communication during aging. *Cell*, **155**, 1624–1638.
43. Bellanti,F., Romano,A.D., Giudetti,A.M., Rollo,T., Blonda,M., Tamborra,R., Vendemiale,G. and Serviddio,G. (2013) Many faces of mitochondrial uncoupling during age: damage or defense?*J. Gerontol. A Biol. Sci. Med. Sci.*, **68**, 892–902.
44. Stadtman,E.R. (2006) Protein oxidation and aging. *Free Radic. Res.*, **40**, 1250–1258.
45. Ademowo,O.S., Dias,H.K.I., Burton,D.G.A. and Griffiths,H.R. (2017) Lipid (per) oxidation in mitochondria: an emerging target in the ageing process?*Biogerontology*, **18**, 859–879.
46. Schriner,S.E., Linford,N.J., Martin,G.M., Treuting,P., Ogburn,C.E., Emond,M., Coskun,P.E., Ladiges,W., Wolf,N., Van Remmen,H. *et al.* (2005) Extension of murine life span by overexpression of catalase targeted to mitochondria. *Science*, **308**, 1909–1911.
47. Shewaramani,S., Finn,T.J., Leahy,S.C., Kassen,R., Rainey,P.B. and Moon,C.D. (2017) Anaerobically grown Escherichia coli has an enhanced mutation rate and distinct mutational spectra. *PLoS Genet.*, **13**, e1006570.
48. Earley,M.C. and Crouse,G.F. (1998) The role of mismatch repair in the prevention of base pair mutations in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 15487–15491.
49. Shin,C.Y. and Turker,M.S. (2002) A:T → G:C base pair substitutions occur at a higher rate than other substitution events in Pms2 deficient mouse cells. *DNA Repair (Amst.)*, **1**, 995–1001.
50. Bridge,G., Rashid,S. and Martin,S.A. (2014) DNA mismatch repair and oxidative DNA damage: implications for cancer biology and treatment. *Cancers*, **6**, 1597–1614.
51. Seplyarskiy,V.B., Akkuratov,E.E., Akkuratova,N., Andrianova,M.A., Nikolaev,S.I., Bazykin,G.A., Adameyko,I. and Sunyaev,S.R. (2019) Error-prone bypass of DNA lesions during lagging-strand replication is a common source of germline and cancer mutations. *Nat. Genet.*, **51**, 36–41.
52. Mikhailova,A.G., Shamansky,V., Mikhailova,A.A., Ushakova,K., Tretyakov,E., Oreshkov,S., Knorre,D., Polishchuk,L., Lawless,D., Mazunin,I. *et al.* (2020) A mitochondrial mutational signature of temperature and longevity in ectothermic and endothermic vertebrates. bioRxiv doi: https://doi.org/10.1101/2020.07.25.221184, 01 November 2021, preprint: not peer reviewed .
53. Martin,A.P. and Palumbi,S.R. (1993) Body size, metabolic rate, generation time, and the molecular clock. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 4087–4091.
54. Naya,H., Romero,H., Zavala,A., Alvarez,B. and Musto,H. (2002) Aerobiosis increases the genomic guanine plus cytosine content (GC%) in prokaryotes. *J. Mol. Evol.*, **55**, 260–264.
55. Romero,H., Pereira,E., Naya,H. and Musto,H. (2009) Oxygen and guanine-cytosine profiles in marine environments. *J. Mol. Evol.*, **69**, 203–206.
56. Aslam,S., Lan,X.-R., Zhang,B.-W., Chen,Z.-L., Wang,L. and Niu,D.-K. (2019) Aerobic prokaryotes do not have higher GC contents than anaerobic prokaryotes, but obligate aerobic prokaryotes have. *BMC Evol. Biol.*, **19**, 35.
57. Popadin,K., Polishchuk,L.V., Mamirova,L., Knorre,D. and Gunbin,K. (2007) Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 13390–13395.
58. Nikolaev,S.I., Montoya-Burgos,J.I., Popadin,K., Parand,L., Margulies,E.H., Antonarakis,S.E and of Health Intramural Sequencing Center Comparative Sequencing Program,N.I. and Othersof Health Intramural Sequencing Center Comparative Sequencing Program,N.I. and Others (2007) Life-history traits drive the evolutionary rates of mammalian coding and noncoding genomic elements. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 20443–20448.
59. Popadin,K.Y., Nikolaev,S.I., Junier,T., Baranova,M. and Antonarakis,S.E. (2013) Purifying selection in mammalian mitochondrial protein-coding genes is highly effective and congruent with evolution of nuclear genes. *Mol. Biol. Evol.*, **30**, 347–355.
60. Ludwig,L.S., Lareau,C.A., Ulirsch,J.C., Christian,E., Muus,C., Li,L.H., Pelka,K., Ge,W., Oren,Y., Brack,A. *et al.* (2019) Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell*, **176**, 1325–1339.
61. Alseth,I., Dalhus,B. and Bjørås,M. (2014) Inosine in DNA and RNA. *Curr. Opin. Genet. Dev.*, **26**, 116–123.
62. Pang,B., Zhou,X., Yu,H., Dong,M., Taghizadeh,K., Wishnok,J.S., Tannenbaum,S.R. and Dedon,P.C. (2007) Lipid peroxidation dominates the chemistry of DNA adduct formation in a mouse model of inflammation. *Carcinogenesis*, **28**, 1807–1813.
63. Degtyareva,N.P., Saini,N., Sterling,J.F., Placentra,V.C., Klimczak,L.J., Gordenin,D.A. and Doetsch,P.W. (2019) Mutational signatures of redox stress in yeast single-strand DNA and of aging in human mitochondrial DNA share a common feature. *PLoS Biol.*, **17**, e3000263.
64. Koh,C.W.Q., Goh,Y.T., Toh,J.D.W., Neo,S.P., Ng,S.B., Gunaratne,J., Gao,Y.-G., Quake,S.R., Burkholder,W.F. and Goh,W.S.S. (2018) Single-nucleotide-resolution sequencing of human N6-methyldeoxyadenosine reveals strand-asymmetric clusters associated with SSBP1 on the mitochondrial genome. *Nucleic Acids Res.*, **46**, 11659–11670.
65. 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.* (2020) N-Deoxyadenosine methylation in mammalian mitochondrial DNA. *Mol. Cell*, **78**, 382–395.