Mitochondrial mutational spectrum is associated with mammalian longevity: a novel signature of oxidative damage.

A. G. Mikhaylova[1*], A. A. Mikhailova[1,2*], K. Ushakova[1*], E.O. Tretiakov[1,3*], V. Shamansky[1], A. Iurchenko[4], M. Zazhytska[5], E. Kozenkova[6], E. Zdobnov[7,8], V. Makeev[9], V. Yurov[1], M. Tanaka[10], I. Gostimskaya[11], Z. Fleischmann[12], S. Annis[12], M. Franco[12], K. Wasko[12], W.S, Kunz[13], D.A. Knorre[14,15], I. Mazunin[16] S. Nikolaev[4], J. Fellay[8,17], A. Reymond[5]**, K. Khrapko[12]**, K. Gunbin[1]**, K. Popadin[1,5,8,17]**

1. Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal University, Kaliningrad, Russian Federation.
2. Institute for Evolution and Biodiversity, University of Münster, Münster, Germany.
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. Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland.
6. Institute of Physics, Mathematics and Information Technology, Immanuel Kant Baltic Federal University, Kaliningrad, Russian Federation.
7. Department of Genetic Medicine and Development, University of Geneva Medical School, 1211 Geneva, Switzerland.
8. Swiss Institute of Bioinformatics, Lausanne, Switzerland.
9. Vavilov Institute of General Genetics, Moscow, Russia.
10. Department for Health and Longevity Research, National Institutes of Biomedical Innovation, Health and Nutrition, Shinjuku-ku, Tokyo 162-8636, Japan.
11. The University of Manchester, Manchester Institute of Biotechnology, United Kingdom
12. Northeastern University, Massachusetts, USA.
13. Institute of Experimental Epileptology and Cognition Research, University Bonn, Sigmund-Freud-Str. 25, 53105 Bonn, Germany.
14. The A.N. Belozersky Institute Of Physico-Chemical Biology, MSU, Moscow, Russian Federation.
15. Institute of Molecular Medicine Sechenov First Moscow State Medical University Trubetskaya str. 8-2, 119991, Moscow.
16. Center of Life Sciences, Skolkovo Institute of Science and Technology, Skolkovo, Russia.
17. School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.

* equal contribution
** equal contribution

Corresponding author:
Konstantin Popadin
konstantin.popadin@epfl.ch
@KPopadin
Ecole polytechnique fédérale de Lausanne (EPFL), school of life science, SV GHI GR-FE, AAB 219 (Bâtiment AAB), Station 19, CH-1015 Lausanne.

A running title: mitochondrial Ah>Gh substitutions mark ageing-related oxidative damage

Keywords: mitochondrial genome, mutational spectrum, longevity

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 tissues and organisms is still incomprehensible. Since mitochondria is tightly involved in aerobic energy production, it is expected that mtDNA mutational spectra may be affected by the oxidative damage which is increasing with organismal aging. However, the well-documented mutational signature of the oxidative damage, G>T substitutions, is typical only for the nuclear genome while it is extremely rare in mtDNA. Thus it is still unclear if there is a mitochondria-specific mutational signature of the oxidative damage. Here, reconstructing mtDNA mutational spectra for 424 mammalian species with variable generation length which is a proxy for oocyte age, we observed that the frequency of A_H>G_H substitutions ($_H$ - heavy chain notation) is positively correlated with organismal longevity. This mutational bias from A_H to G_H significantly affected the nucleotide content of analyzed 650 complete mammalian mitochondrial genomes, where fourfold degenerative synonymous positions of long-lived species become more A_H poor and G_H rich. Because (i) A>G is a substitution, typical for mtDNA; (ii) it is characterized by very strong asymmetry: A>G is several-fold more frequent on a heavy chain as compared to the light one; (iii) it is sensitive to the time being single-stranded during mtDNA asynchronous replication; (iv) it is associated with oxidative damage of single-stranded DNA in recent experimental studies we propose that A>G is a novel mutational signature of age-associated oxidative damage of single-stranded mtDNA. The described association of the mtDNA mutational spectra with a species-specific life-history trait can significantly affect general patterns of molecular evolution of mtDNA.

Introduction

Evolution is a function of mutagenesis and selection. In order 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 mtDNA of long-lived mammals is a result of selection, favouring more stable genomes in long-lived species (Lehmann et al. 2006). However, this conclusion can be premature without a comparison of mutational processes between short- and long-lived mammals. Indeed, significant changes in mtDNA mutational spectra between different species have been shown (Belle et al. 2005; Montooth and Rand 2008), however, no driving factors explaining this variation has been proposed till now.

Interestingly, a similar gap of knowledge 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 (Yuan et al. 2017; Ju et al. 2014). 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 (Yuan et al. 2017; Ju et al. 2014). 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 scales.

The widely accepted expectation is that mitochondrially produced reactive oxygen species (ROS) can damage mtDNA (Alexeyev 2009). The well-documented ROS-induced modification of the guanine (G) DNA base is 7,8-Dihydro-8-oxo-20-deoxyguanosine (8-oxodG), which after mispairing with adenine, leads to G > T transversion mutations. Although G>T mutations are considered to be the hallmark of oxidative damage in the nuclear DNA (COSMIC signature 18) (Fraga et al. 1990; Ludmil B. Alexandrov et al. 2013; Kucab et al. 2019), it is rather rare in mtDNA (Yuan et al. 2020), doesn't significantly increase with age in mtDNA (Kennedy et al. 2013), and experimentally induced in mtDNA only at high concentrations of ROS (Guliaeva et al. 2006). Altogether, up to now, there is no well-established mtDNA mutational signature of oxidative damage (Zsurka et al. 2018).

Taking into account recent progress in deciphering the variation in mutational spectra of the nuclear genome between human populations (Harris and Pritchard 2017) and primate species (Moorjani et al.), here we perform a large-scale reconstruction of mtDNA mutational spectra across hundreds of mammalian species. Considering the tight association of the mtDNA metabolism with species-specific life-history traits, we aimed to uncover a correlation between the mtDNA mutational spectrum and life-history traits. Using a collection of polymorphic synonymous substitutions in most neutral sites from hundreds of mammalian species, as well as nucleotide content in whole mitochondrial genomes of mammals, we observed one universal trend: A_H>G_H substitution positively correlates with the species-specific generation length. Since the mtDNA of all mammals is inherited exclusively through the oocytes, and oocyte division is stopped after birth, the species-specific generation length correlates well with the lifespan of oocytes. Thus the increased A_H>G_H of long-lived mammals can be a mutational signature of mtDNA oxidative damage in aged oocytes. The described variation in the mtDNA mutational spectrum of mammals affects the whole genome nucleotide content, codon usage and thus should be taken into account in evolutionary analyses based on mtDNA.

Results

(1) Mutational spectrum, derived from the intra-species mtDNA polymorphisms of mammals, is shaped by generation length

The variation in mtDNA mutational spectra between different species (Montooth and Rand 2008; Belle et al. 2005) has no overarching explanation till now. However, recent experiments based on high-sensitive detection of de novo variants showed that mtDNA mutational spectrum, and particularly a fraction of mtDNA A_H>G_H transitions, is changing with oocyte age (Arbeithuber et al. 2020). Extending the logic, here we hypothesize that oocyte lifespan can be associated with mtDNA mutational spectra. Oocytes are the only lineage through which mtDNA is transmitted (with rare exceptions of paternal inheritance) from generation to generation in mammals (Sato and Sato 2017). Mammalian oocytes are arrested from birth to puberty, which takes weeks (mouse) or decades (humans) (Von Stetina and Orr-Weaver 2011) and allows us to use the species-specific generation length as a proxy for oocyte longevity in different mammalian species. The generation length is also associated with numerous ecological (body mass, litter size, effective population size) and physiological (basal metabolic rate) parameters of mammalian species (Ollason 1987; Damuth 1987) and is available for the vast majority of mammalian species (Pacifici et al. 2013; Tacutu et al. 2013).

Due to extensive mitochondrial sequences coming from ecological, evolutionary, and population genetics studies (Hebert et al. 2003) we were able to collect mtDNA polymorphisms and using our in-house pipeline (see Methods, Supplementary Materials 1.1) reconstruct the mutational spectrum for 611 mammalian species (Figure 1A, Supplementary Materials 1.2). Hereafter, to simplify the biological interpretability of the mtDNA mutational spectrum, we use a 12-component spectrum based on heavy chain notation. The choice of 12 components was based on the fact that the mitochondrial mutational spectrum is characterized by very strong asymmetry and thus complementary substitutions have different probabilities; the heavy chain was the chain of choice since it is more mutagenic and thus the direction of the majority of the nucleotide substitutions would reflect the mutagenic process. The average mutational spectrum for all mammalian species (Figure 1A) demonstrates an excess of C_H>T_H and A_H>G_H substitutions which have been shown in previous studies (Ju et al. 2014; Yuan et al. 2020).

To uncover the nature of mitochondrial mutagenesis shaping the species-specific variation in mutational spectra we focused on *MT-CYB* gene, which contained 56% of all analyzed mutations (39,112 out of 70,053 used to draw Figure 1A) and compared the *MT-CYB*-derived spectrum between species with variable life-history traits. As the first and simplest approximation of the mutational spectrum, we used Ts/Tv. For 424 mammalian species with known Ts/Tv and generation length, we observed a positive correlation between those two parameters (Figure S1B). To test the robustness of this correlation we performed several additional analyses splitting all species into several groups (by quartiles of the generation length, by median of the generation length or by families); results were also not affected by the phylogenetic inertia and the total number of polymorphisms observed in different species (Supplementary Materials 1.3: Figures S1B, S1C, S1D and S1E, Tables S1A and S1B).

To understand which substitution type predominantly shaped the observed correlation between Ts/Tv and generation length we performed twelve pairwise rank correlation analyses. We observed that only A_H>G_H frequency positively correlated with the generation length (Spearman's rho = 0.252, p-value = 1.188e-07) (Figure 1B), while several rare transversions showed a 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 p-values are < 0.0003). The inclusion of all five types of substitutions into a multiple linear model showed A_H>G_H being the strongest component associated with generation length (Supplementary Materials 1.4, equation 1) and the observed effect was robust to phylogenetic inertia and the number of analyzed polymorphisms (Supplementary Materials 1.4, Table S1C).

In order to derive mtDNA mutational signatures in an unsupervised way, we performed principal component analysis (PCA) of mutational spectra of 424 mammalian species. We observed that the first component is mainly driven by C_H>T_H substitutions, whereas the second is driven mainly by A_H>G_H substitutions (Figure 1C left panel). We assessed if the first ten principal components were correlated with generation length and observed that only the second component was significantly correlated (Figure 1C right panel).

Interestingly, the correlation of the second principal component with generation length was higher as compared to the sole effect of A_H>G_H frequency, suggesting that the second component could reflect a complex nature of longevity-associated mutational signature (Supplementary Materials 1.5). Additional analyses which take into account the effects of the total number of mutations (Supplementary Materials 1.5), potential sequencing errors (Supplementary Materials 1.6, Table S1D) and nucleotide content (Supplementary Materials 1.7, Table S1E) confirmed the robustness of our results. Of note, G>T transversions (G_H>T_H and C_H>A_H) despite the rareness and weak effect, demonstrate an opposite to A_H>G_H trend (Figure 1C 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.

Altogether, our analyses of within-species neutral polymorphisms in hundreds of mammalian species demonstrated a robust association of mtDNA mutational spectrum (mainly A_H>G_H) to the generation length.

(2) The long-term mutational bias affects nucleotide composition of complete mammalian mitochondrial genomes

Mutational bias, if stronger than selection, is expected to change genome-wide nucleotide content. We expect that increased A_H>G_H (Figure 1) would decrease frequencies of A_H and increase frequencies of G_H in effectively neutral positions of long-lived species. Using 650 complete mitochondrial genomes of mammalian species we estimated their nucleotide content in the most neutral - synonymous fourfold degenerate - positions: in 13 protein-coding genes without overlapped regions. Testing all four pairwise correlations between the species-specific generation length and the nucleotide content we observed two strongest correlations: negative with A_H and positive with G_H (Figure 2A; Supplementary Materials 2.1, Table S2A). Inclusion of all four types of nucleotide frequencies into 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 Materials 2.1). Thus we concluded that mtDNA of long- versus short-lived mammals is more A_H poor and G_H rich (Figure 2A), which is in line with the more intensive A_H>G_H mutagenesis in the former (Figure 1).

An excess of G_H and deficit of A_H in long-lived species determines the G_HA_H nucleotide skew which approximates the level of asymmetry in the distribution of these two nucleotides and is calculated as $(G_H-A_H)/(G_H+A_H)$. As expected we observed positive correlation between G_HA_H skew and generation length of mammalian species (phylogenetic independent contrasts: Spearman's rho = 0.09, p-value = 0.0164; phylogenetic generalized least squares: coefficient = 0.13, p-value = $2.9*10^{-4}$; see also Figure 2C). To visualize a contrast in G_HA_H skew between the shortest- and the longest-lived species we plotted A_H and G_H fractions along the major arc of mtDNA for opossum (generation length 341 days) and whale (generation length 18980 days) (Figure 2B). An excess of G_H and deficit of A_H were more pronounced in whale versus opossum.

Moreover, the frequencies of G_H and A_H are changing along the major arc (Figure 2B, 2C), reflecting the change in G_HA_H skew with gene position. It has been shown that the frequency of A_H>G_H substitutions depends on Time which heavy strand Spent in Single-Stranded (TSSS) condition during asynchronous mtDNA replication (Faith and Pollock 2003). 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. Thus, we expect that G_HA_H skew is a function of both: species-specific generation length and gene-specific TSSS. Plotting G_HA_H skew of short- and long-lived mammals for each gene we observed that indeed G_HA_H skew increases with both species-specific generation length and gene-specific TSSS (Figure 2C).

Increase in G_HA_H skew with TSSS is supported by a corresponding decrease in A_H and an increase in G_H in all species (Figure 2A, Supplementary Materials 2.2). Interestingly, comparing the rate of changes in the nucleotide frequencies with TSSS we observed that long-lived mammals demonstrate faster decrease in A_H and increase in G_H along the genome (Supplementary Materials 2.2, Figure S2, the same effect is visible on Figure 2B). Fast changes (stronger gradients) in A_H and G_H in long-lived mammals can be interpreted as an interaction between TSSS and longevity probably mediated by increased oxidative damage (see discussion) in genes with high TSSS in long-lived mammals.

Altogether, our results demonstrate that the neutral nucleotide content reflects the mutational bias from A_H to G_H: long-lived mammals are more A_H poor and G_H rich (Figure 2A), have stronger G_HA_H skew (Figure 2B, 2C) and

faster changes in nucleotide content (decrease in A_H and increase in G_H) along the genome (Supplementary Materials 2.2).

Discussion

We observed that the fraction of A_H>G_H substitutions increases with organismal longevity on different time scales: (i) from hundreds of thousands years which is the average time of segregation of neutral within-species mtDNA polymorphisms (Atkinson, Gray, and Drummond 2008) (Figure 1), to (ii) millions of years as in the case of a comparative species scale (Figure 2).

We propose that a process of mtDNA mutagenesis, rather than selection, is primarily responsible for all our findings. Due to little or no evidence of selection on synonymous four-fold degenerate sites in mtDNA of mammals (Faith and Pollock 2003; Uddin and Chakraborty 2017) we consider the vast majority of variants analysed in this study as effectively neutral and thus driven predominantly by mutagenesis.

We propose that mtDNA mutagenesis is sensitive to the mitochondrial microenvironment which in turn depends on organism-specific properties. We assume that oxidative damage - a main by-product of the aerobic metabolism - through the deamination of adenosine induces A_H>G_H substitutions on a single-stranded mtDNA heavy strand. Importantly, the well-documented ROS-induced 7,8-Dihydro-8-oxo-20-deoxyguanosine (8-oxodG), leading to G > T transversions, was expected to be one of the strongest mtDNA-markers of oxidative damage. However, mtDNA G > T is low frequency, age-independent and demonstrates weak if any association with oxidative damage (Yuan et al. 2020; Kennedy et al. 2013; Hoekstra et al. 2016; Zsurka et al. 2018). Here, instead, we describe a novel potential mtDNA-marker of oxidative damage: A_H>G_H substitution on a single-stranded mtDNA heavy strand and numerous recent papers, discussed below, support it.

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 hallmark of oocyte aging (Arbeithuber et al. 2020). This discovery also suggests that continuous mtDNA turnover within the dormant oocytes can be a universal trait of all mammalian species (Arbeithuber et al. 2020) - and thus variation in mtDNA mutational spectra between different species can reflect the level of oxidative damage in oocytes of different species. Extending the results of Arbeithuber et al (Arbeithuber et al. 2020) from young and old mice oocytes to oocytes of short- and long-lived mammalian species we observed the same trend with A_H>G_H marking more long-lived oocytes (Figure 1). These results suggest increased oxidative damage in mtDNA of oocytes of long-lived mammals.

Extending the trend of the species-specific mtDNA mutational spectra (Figure 1) to nucleotide content of mitochondrial genomes (Figure 2) confirms long-term effects of the described tendencies and goes in line with an excess of G_H in mtDNA of long-lived mammals which had been previously shown (Lehmann et al. 2006, Lehmann et al. 2008). Although Lehmann et al proposed a selective 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 result of mutagenesis, not selection.

Altogether, we propose that A_H>G_H substitutions are a marker of oxidative damage on the single-stranded DNA. It is important to mention that A_H>G_H transitions are the substitution type which is the most sensitive to the time being single-stranded during asymmetric replication of mammalian mtDNA (Faith and Pollock 2003; Tanaka and Ozawa 1994), which may be driven by the increased oxidative damage of single-stranded DNA. An excess of A:T>G:C has been observed also in aerobically versus anaerobically grown Escherichia coli and the signal was driven by the lagging strand, spending more time in single-stranded condition (Shewaramani et al. 2017). Moreover, A>G substitutions, associated with oxidative damage, can be a key process, explaining a long-standing puzzle of increased GC content of aerobic versus anaerobic bacteria (Naya et al. 2002; Romero et al. 2009; Aslam et al. 2019). Additional lines of evidence are coming from a mismatch repair pathway, which can prevent mutations caused by oxidative stress (Bridge, Rashid, and Martin 2014). Mismatch repair-deficient mouse cell line shows a 6-50 fold increase in the rate of the A:T>G:C substitutions, and this increase is washed out in the presence of antioxidants (Shin and Turker 2002). Also, A>G is the most asymmetric mutation in the nuclear genome, which is a hallmark of a mutagen, inducing strand-specific DNA damage (Seplyarskiy et al. 2019), which we propose can be oxidative damage. Also, it has been shown recently that the fraction of A_H>G_H in mtDNA positively correlates with the ambient temperature of Actinopterygii species and it is higher in non-hibernating versus hibernating mammals

(Mikhailova et al. 2020). Because higher temperature is associated with an increased level of aerobic metabolism (Martin and Palumbi 1993) we assume that A_H>G_H can be a marker of oxidative damage.

The results of our study can significantly expand the usability of mtDNA mutational spectra. Low-heteroplasmy somatic mtDNA mutations from a merely neutral marker used to trace cellular lineages (Ludwig et al. 2019) can be transformed to a metric, associated with cell-specific aging state - a metric which can be especially important in highly heterogeneous tissues such as cancers. Low-heteroplasmy de novo germ-line mtDNA mutations (Arbeithuber et al. 2020) can predict the biological age of human oocytes, which can be used in *in vitro* fertilization techniques. MtDNA mutational spectra, reconstructed for non-model vertebrate species, may help to approximate their average generation length - a metric that is not always easier to estimate empirically.

Direct experimental investigation of the effect of different mutagens (Kucab et al. 2019) and especially an effect of oxidative damage ((Degtyareva et al. 2019) and chemical modifications of nucleotides (Koh et al. 2018, Hao et al. 2020) on single-stranded mtDNA will significantly improve our understanding of mtDNA mutagenesis.

Figures

FIGURE 1: variation in mammalian mutational spectrum is driven by generation length

1A. Average mtDNA mutational spectrum of mammalian species (N = 611). Mutational spectrum as 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.

1B. Mutational spectra vary with species-specific generation length (N = 424). Frequency of A_H>G_H is the best type of substitutions, correlated with generation length.

1C. The principal component analysis 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 reverted PC2 to make it positively correlated with generation length. Right panel: The second principal component correlates with generation length in mammals. Generation length is color-coded from dark green (the shortest generation length) to light green (the longest generation length).

FIGURE_2: the long-term effect of the mutational bias: neutral nucleotide content in mammalian species

2A. 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).

2B. Frequencies of A_H (red) and G_H (grey) nucleotides along the major arc of mtDNA of the most short-lived and most long-lived 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 mtDNA: from bottom left (origin of replication of light strand) to top right (origin of replication of heavy strand). Also, mtDNA of a long-lived mammal has a deficit of A_H and excess of G_H genome-wide - a signature of increased longevity.

2C. 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.

Methods

Reconstruction of the species-specific mutational spectrum for mammalian species

Using all available intraspecies sequences (at April 2016) of mitochondrial protein-coding genes we derived the mutational spectrum for each species. Briefly, 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 analyzed 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 details in Supplementary Materials 1.1. Using species with at least 15 single-nucleotide synonymous mutations at four-fold degenerate sites we estimated mutational spectrum as a probability of each nucleotide to mutate into any other nucleotide (vector of 12 types of substitutions with sum equals one) for more than a thousand of chordata species. We normalized 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 analyzed genes between different species the vast majority of our analyses were performed with *MT-CYB* 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 Pacifici et al. (Pacifici et al. 2013).

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 fourfold degenerate positions. For each protein-coding gene of each reference mitochondrial genome of mammalian species we estimated G_HA_H skew as $(G_H-A_H)/(G_H+A_H)$ using only synonymous fourfold degenerate sites.

All statistical analyses were performed in R.

Acknowledgments

We thank Athanasios Kousathanas for statistical comments, the whole laboratory of Alexandre Reymond, Mikhail Gelfand and Vladimir Katanaev for discussions and valuable suggestions. K.P. was supported by the 5 Top 100 Russian Academic Excellence Project at the Immanuel Kant Baltic Federal University. E.O.T. is supported by a scholarship from the Austrian Science Fund (FWF, DOC 33-B27). This work was also supported by Russian Foundation of Basic Research [No. 19-29-04101 to I.M]. This work was also supported by Russian Science Foundation [No. 21-75-20143 to A.G., K.P, K.G. & V.Y.]

Literature

Alexandrov, L. B., J. Kim, N. J. Haradhvala, M. N. Huang, Tian Ng Aw, Y. Wu, A. Boot, et al. 2020. "The Repertoire of Mutational Signatures in Human Cancer." *Nature* 578 (7793). https://doi.org/10.1038/s41586-020-1943-3.

Alexandrov, Ludmil B., Serena Nik-Zainal, David C. Wedge, Samuel A J, Sam Behjati, Andrew V. Biankin, Graham R. Bignell, et al. 2013. "Signatures of Mutational Processes in Human Cancer." *Nature* 500 (7463): 415–21.

Arbeithuber, Barbara, James Hester, Marzia A. Cremona, Nicholas Stoler, Arslan Zaidi, Bonnie Higgins, Kate Anthony, Francesca Chiaromonte, Francisco J. Diaz, and Kateryna D. Makova. 2020. "Age-Related Accumulation of de Novo Mitochondrial Mutations in Mammalian Oocytes and Somatic Tissues." *PLoS Biology* 18 (7): e3000745.

Aslam, Sidra, Xin-Ran Lan, Bo-Wen Zhang, Zheng-Lin Chen, Li Wang, and Deng-Ke Niu. 2019. "Aerobic Prokaryotes Do Not Have Higher GC Contents than Anaerobic Prokaryotes, but Obligate Aerobic Prokaryotes Have." *BMC Evolutionary Biology* 19 (1): 35.

Atkinson, Q. D., R. D. Gray, and A. J. Drummond. 2008. "mtDNA Variation Predicts Population Size in Humans and Reveals a Major Southern Asian Chapter in Human Prehistory." *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msm277.

Belle, Elise M. S., Gwenael Piganeau, Mike Gardner, and Adam Eyre-Walker. 2005. "An Investigation of the Variation in the Transition Bias among Various Animal Mitochondrial DNA." *Gene* 355 (August): 58–66.

Bhandari, Vinayak, Constance H. Li, Robert G. Bristow, Paul C. Boutros, and PCAWG Consortium. 2020. "Divergent Mutational Processes Distinguish Hypoxic and Normoxic Tumours." *Nature Communications* 11 (1): 737.

Bridge, Gemma, Sukaina Rashid, and Sarah A. Martin. 2014. "DNA Mismatch Repair and Oxidative DNA Damage: Implications for Cancer Biology and Treatment." *Cancers* 6 (3): 1597–1614.

Damuth, John. 1987. "Interspecific Allometry of Population Density in Mammals and Other Animals: The Independence of Body Mass and Population Energy-Use." *Biological Journal of the Linnean Society*. https://doi.org/10.1111/j.1095-8312.1987.tb01990.x.

Degtyareva, Natalya P., Natalie Saini, Joan F. Sterling, Victoria C. Placentra, Leszek J. Klimczak, Dmitry A. Gordenin, and Paul W. Doetsch. 2019. "Mutational Signatures of Redox Stress in Yeast Single-Strand DNA and of Aging in Human Mitochondrial DNA Share a Common Feature." *PLOS Biology*. https://doi.org/10.1371/journal.pbio.3000263.

Ericson, Nolan G., Mariola Kulawiec, Marc Vermulst, Kieran Sheahan, Jacintha O'Sullivan, Jesse J. Salk, and Jason H. Bielas. 2012. "Decreased Mitochondrial DNA Mutagenesis in Human Colorectal Cancer." *PLoS Genetics* 8 (6): e1002689.

Faith, Jeremiah J., and David D. Pollock. 2003. "Likelihood Analysis of Asymmetrical Mutation Bias Gradients in Vertebrate Mitochondrial Genomes." *Genetics* 165 (2): 735–45.

Fraga, C. G., M. K. Shigenaga, J. W. Park, P. Degan, and B. N. Ames. 1990. "Oxidative Damage to DNA during Aging: 8-Hydroxy-2'-Deoxyguanosine in Rat Organ DNA and Urine." *Proceedings of the National Academy of Sciences of the United States of America* 87 (12): 4533–37.

Hao, Ziyang, Tong Wu, Xiaolong Cui, Pingping Zhu, Caiping Tan, Xiaoyang Dou, Kai-Wen Hsu, et al. 2020. "N-Deoxyadenosine Methylation in Mammalian Mitochondrial DNA." *Molecular Cell* 78 (3): 382–95.e8.

Hebert, Paul D. N., Alina Cywinska, Shelley L. Ball, and Jeremy R. deWaard. 2003. "Biological Identifications through DNA Barcodes." *Proceedings. Biological Sciences / The Royal Society* 270 (1512): 313–21.

Hoekstra, Jake G., Michael J. Hipp, Thomas J. Montine, and Scott R. Kennedy. 2016. "Mitochondrial DNA Mutations Increase in Early Stage Alzheimer Disease and Are Inconsistent with Oxidative Damage." *Annals of Neurology* 80 (2): 301–6.

Ju, Young Seok, Ludmil B. Alexandrov, Moritz Gerstung, Inigo Martincorena, Serena Nik-Zainal, Manasa Ramakrishna, Helen R. Davies, et al. 2014. "Origins and Functional Consequences of Somatic Mitochondrial DNA Mutations in Human Cancer." *eLife* 3 (October). https://doi.org/10.7554/eLife.02935.

Kennedy, Scott R., Jesse J. Salk, Michael W. Schmitt, and Lawrence A. Loeb. 2013. "Ultra-Sensitive Sequencing Reveals an Age-Related Increase in Somatic Mitochondrial Mutations That Are Inconsistent with Oxidative Damage." *PLoS Genetics* 9 (9): e1003794.

Koh, Casslynn W. Q., Yeek Teck Goh, Joel D. W. Toh, Suat Peng Neo, Sarah B. Ng, Jayantha Gunaratne, Yong-Gui Gao, Stephen R. Quake, William F. Burkholder, and Wee Siong S. Goh. 2018. "Single-Nucleotide-Resolution Sequencing of Human N6-Methyldeoxyadenosine Reveals Strand-Asymmetric Clusters Associated with SSBP1 on the Mitochondrial Genome." *Nucleic Acids Research* 46 (22): 11659–70.

Kucab, Jill E., Xueqing Zou, Sandro Morganella, Madeleine Joel, A. Scott Nanda, Eszter Nagy, Celine Gomez, et al. 2019. "A Compendium of Mutational Signatures of Environmental Agents." *Cell* 177 (4): 821–36.e16.

Lehmann, Gilad, Arie Budovsky, Khachik K. Muradian, and Vadim E. Fraifeld. 2006. "Mitochondrial Genome Anatomy and Species-Specific Lifespan." *Rejuvenation Research* 9 (2): 223–26.

Ludwig, Leif S., Caleb A. Lareau, Jacob C. Ulirsch, Elena Christian, Christoph Muus, Lauren H. Li, Karin Pelka, et al. 2019. "Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics." *Cell* 176 (6): 1325–39.e22.

Martin, A. P., and S. R. Palumbi. 1993. "Body Size, Metabolic Rate, Generation Time, and the Molecular Clock." *Proceedings of the National Academy of Sciences of the United States of America* 90 (9): 4087–91.

Mikhailova, Alina G., Victor Shamansky, Alina A. Mikhailova, Kristina Ushakova, Evgenii Tretyakov, Sergey Oreshkov, Dmitry Knorre, et al. 2020. "A Mitochondrial Mutational Signature of Temperature and Longevity in Ectothermic and Endothermic Vertebrates." *Cold Spring Harbor Laboratory*. https://doi.org/10.1101/2020.07.25.221184.

Montooth, Kristi L., and David M. Rand. 2008. "The Spectrum of Mitochondrial Mutation Differs across Species." *PLoS Biology* 6 (8): e213.

Naya, Hugo, Héctor Romero, Alejandro Zavala, Beatriz Alvarez, and Héctor Musto. 2002. "Aerobiosis Increases the Genomic Guanine plus Cytosine Content (GC%) in Prokaryotes." *Journal of Molecular Evolution* 55 (3): 260–64.

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)." *Journal of Tropical Ecology*. https://doi.org/10.1017/s0266467400002224.

Pacifici, Michela, Luca Santini, Moreno Di Marco, Daniele Baisero, Lucilla Francucci, G. Grottolo Marasini, Piero Visconti, and Carlo Rondinini. 2013. "Generation Length for Mammals." *Nature Conservation* 5: 87–94.

Rebolledo-Jaramillo, B., M. S. -W. Su, N. Stoler, J. A. McElhoe, B. Dickins, D. Blankenberg, T. S. Korneliussen, et al. 2014. "Maternal Age Effect and Severe Germ-Line Bottleneck in the Inheritance of Human Mitochondrial DNA." *Proceedings of the National Academy of Sciences* 111 (43): 15474–79.

Romero, Héctor, Emiliano Pereira, Hugo Naya, and Héctor Musto. 2009. "Oxygen and Guanine-Cytosine Profiles in Marine Environments." *Journal of Molecular Evolution* 69 (2): 203–6.

Rosario, S. R., M. D. Long, H. C. Affronti, A. M. Rowsam, K. H. Eng, and D. J. Smiraglia. 2018. "Pan-Cancer Analysis of Transcriptional Metabolic Dysregulation Using The Cancer Genome Atlas." *Nature Communications* 9 (1): 5330.

Sato, Ken, and Miyuki Sato. 2017. "Multiple Ways to Prevent Transmission of Paternal Mitochondrial DNA for Maternal Inheritance in Animals." *Journal of Biochemistry* 162 (4): 247–53.

Seplyarskiy, Vladimir B., Evgeny E. Akkuratov, Natalia Akkuratova, Maria A. Andrianova, Sergey I. Nikolaev, Georgii A. Bazykin, Igor Adameyko, and Shamil R. Sunyaev. 2019. "Error-Prone Bypass of DNA Lesions during Lagging-Strand Replication Is a Common Source of Germline and Cancer Mutations." *Nature Genetics* 51 (1): 36–41.

Shewaramani, Sonal, Thomas J. Finn, Sinead C. Leahy, Rees Kassen, Paul B. Rainey, and Christina D. Moon. 2017. "Anaerobically Grown Escherichia Coli Has an Enhanced Mutation Rate and Distinct Mutational Spectra." *PLoS Genetics* 13 (1): e1006570.

Shin, Chi Y., and Mitchell S. Turker. 2002. "A:T → G:C Base Pair Substitutions Occur at a Higher Rate than Other Substitution Events in Pms2 Deficient Mouse Cells." *DNA Repair*. https://doi.org/10.1016/s1568-7864(02)00149-0.

Tacutu, Robi, Thomas Craig, Arie Budovsky, Daniel Wuttke, Gilad Lehmann, Dmitri Taranukha, Joana Costa, Vadim E. Fraifeld, and João Pedro de Magalhães. 2013. "Human Ageing Genomic Resources: Integrated Databases and Tools for the Biology and Genetics of Ageing." *Nucleic Acids Research* 41 (Database issue): D1027–33.

Tanaka, M., and T. Ozawa. 1994. "Strand Asymmetry in Human Mitochondrial DNA Mutations." *Genomics* 22 (2): 327–35.

Tomasetti, Cristian, Rick Durrett, Marek Kimmel, Amaury Lambert, Giovanni Parmigiani, Ann Zauber, and Bert Vogelstein. 2017. "Role of Stem-Cell Divisions in Cancer Risk." *Nature* 548 (7666): E13–14.

Tomasetti, Cristian, and Bert Vogelstein. 2015. "Cancer Etiology. Variation in Cancer Risk among Tissues Can Be Explained by the Number of Stem Cell Divisions." *Science* 347 (6217): 78–81.

Uddin, Arif, and Supriyo Chakraborty. 2017. "Synonymous Codon Usage Pattern in Mitochondrial CYB Gene in Pisces, Aves, and Mammals." *Mitochondrial DNA. Part A, DNA Mapping, Sequencing, and Analysis* 28 (2): 187–96.

Von Stetina, Jessica R., and Terry L. Orr-Weaver. 2011. "Developmental Control of Oocyte Maturation and Egg Activation in Metazoan Models." *Cold Spring Harbor Perspectives in Biology* 3 (10): a005553.

Wallace, Douglas C. 2005. "A Mitochondrial Paradigm of Metabolic and Degenerative Diseases, Aging, and Cancer: A Dawn for Evolutionary Medicine." *Annual Review of Genetics* 39: 359.

Wei, Wei, Salih Tuna, Michael J. Keogh, Katherine R. Smith, Timothy J. Aitman, Phil L. Beales, David L. Bennett, et al. 2019. "Germline Selection Shapes Human Mitochondrial DNA Diversity." *Science* 364 (6442). https://doi.org/10.1126/science.aau6520.

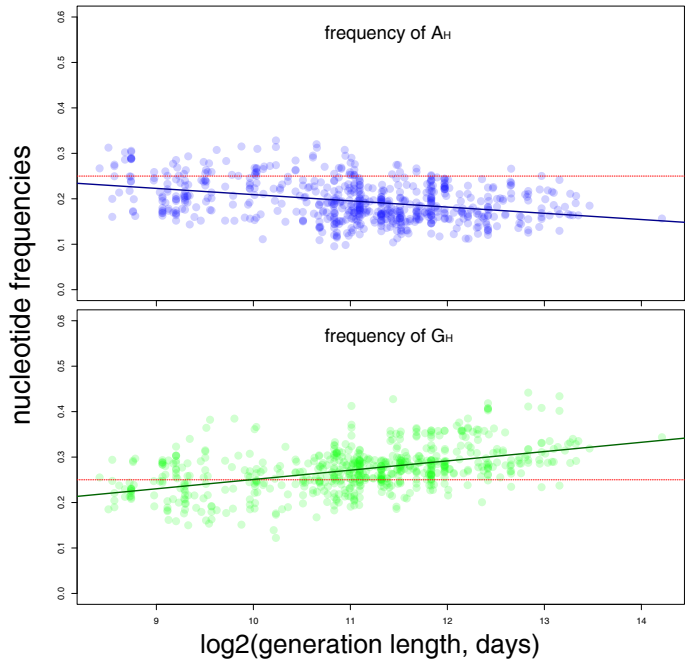
Yuan, Yuan, Young Seok Ju, Youngwook Kim, Jun Li, Yumeng Wang, Yang Yang, Inigo Martincorena, et al. 2017. "Comprehensive Molecular Characterization of Mitochondrial Genomes in Human Cancers." *bioRxiv*. https://doi.org/10.1101/161356.

Yuan, Yuan Yuan, PCAWG Consortium, Young Seok Ju, Youngwook Kim, Jun Li, Yumeng Wang, et al. 2020. "Comprehensive Molecular Characterization of Mitochondrial Genomes in Human Cancers." *Nature Genetics*. https://doi.org/10.1038/s41588-019-0557-x.

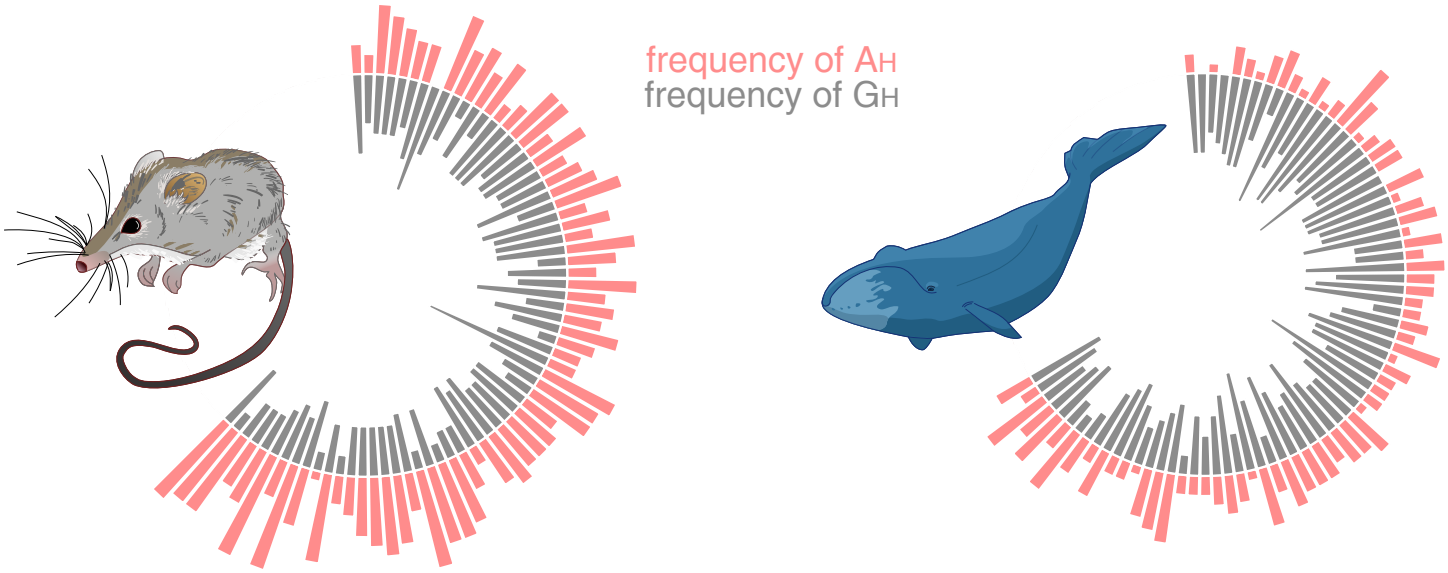
Zaidi, Arslan A., Peter R. Wilton, Marcia Shu-Wei Su, Ian M. Paul, Barbara Arbeithuber, Kate Anthony, Anton Nekrutenko, Rasmus Nielsen, and Kateryna D. Makova. 2019. "Bottleneck and Selection in the Germline and Maternal Age Influence Transmission of Mitochondrial DNA in Human Pedigrees." *Proceedings of the National Academy of Sciences of the United States of America* 116 (50): 25172–78.

Zsurka, Gábor, Viktoriya Peeva, Alexander Kotlyar, and Wolfram S. Kunz. 2018. "Is There Still Any Role for Oxidative Stress in Mitochondrial DNA-Dependent Aging?" *Genes* 9 (4). https://doi.org/10.3390/genes9040175.

A



B



Tarsipes rostratus, 341 days

Balaena mysticetus, 18980 days

C

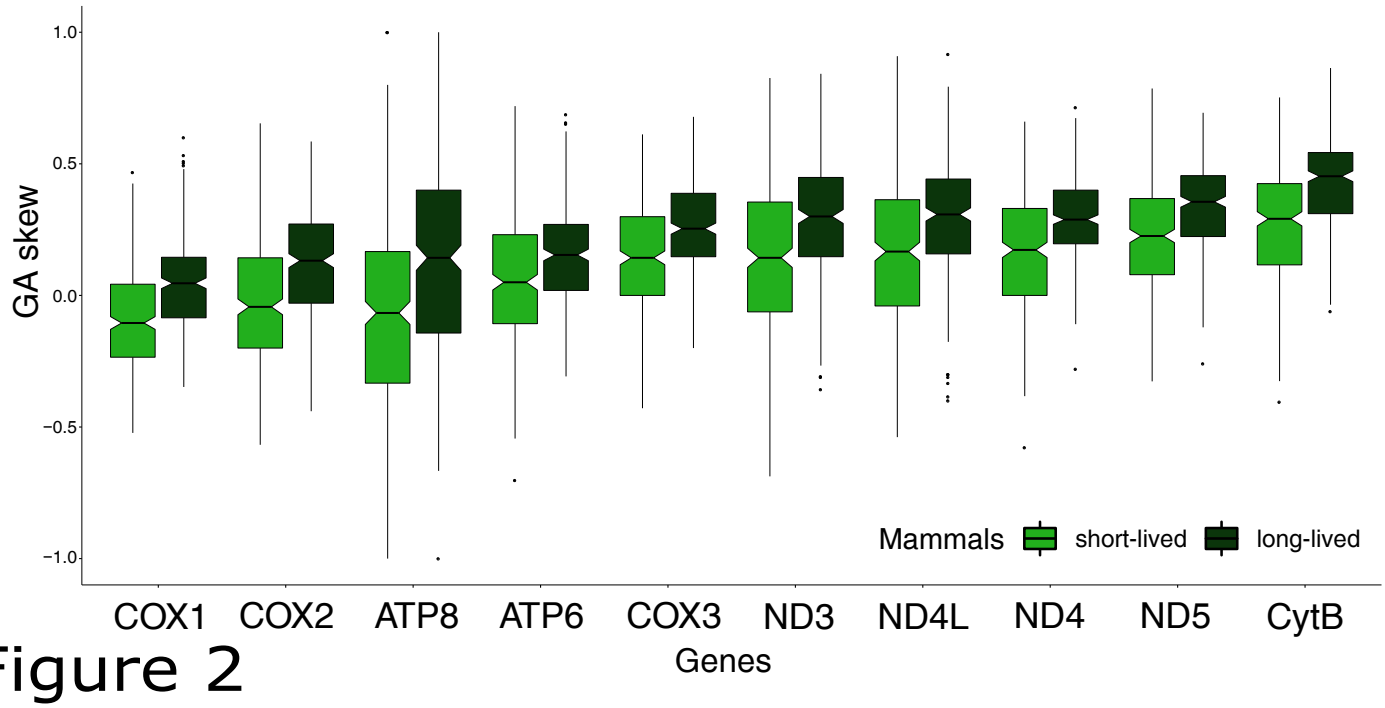
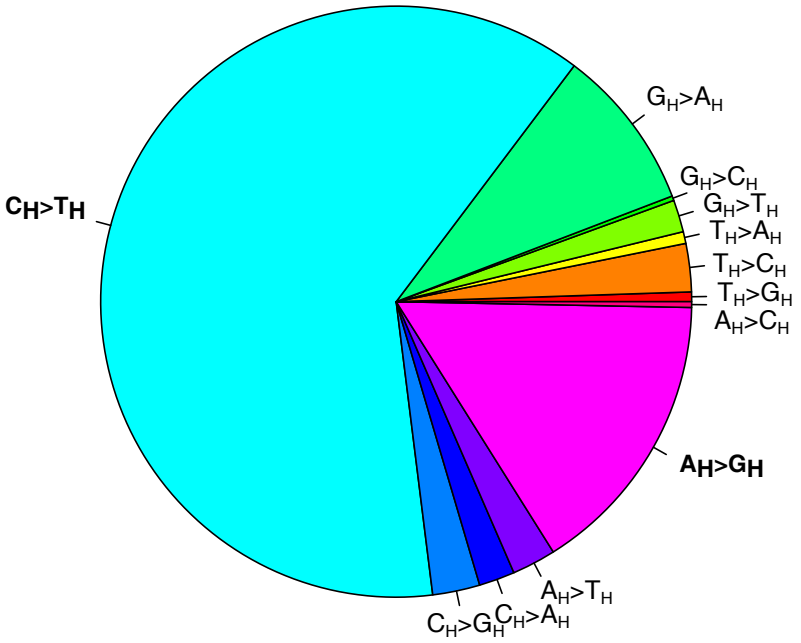
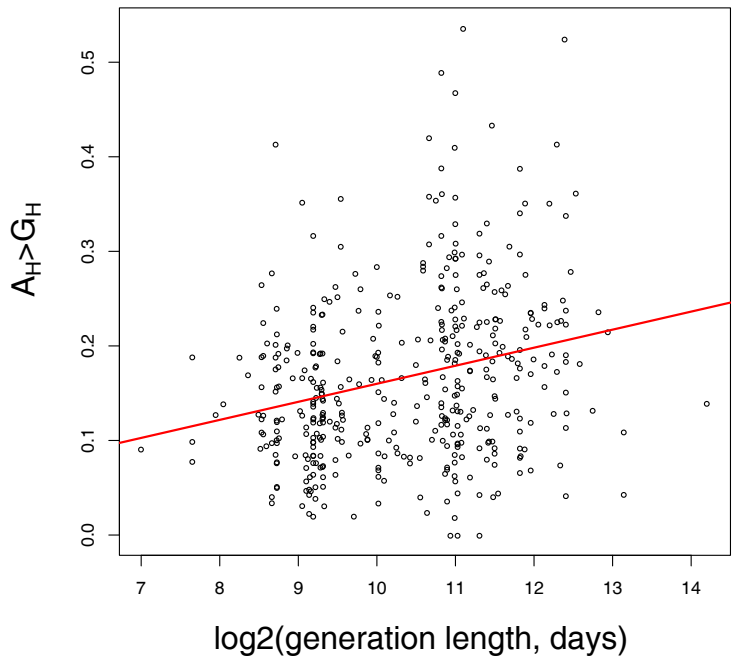


Figure 2

A


$G_H>A_H$
$G_H>C_H$
$G_H>T_H$
$T_H>A_H$
$T_H>C_H$
$T_H>G_H$
$A_H>C_H$
$C_H>T_H$
$A_H>G_H$
$A_H>T_H$
$C_H>A_H$
$C_H>G_H$

B


$A_H>G_H$
0.0
0.1
0.2
0.3
0.4
0.5
7
8
9
10
11
12
13
14
log2(generation length, days)

C

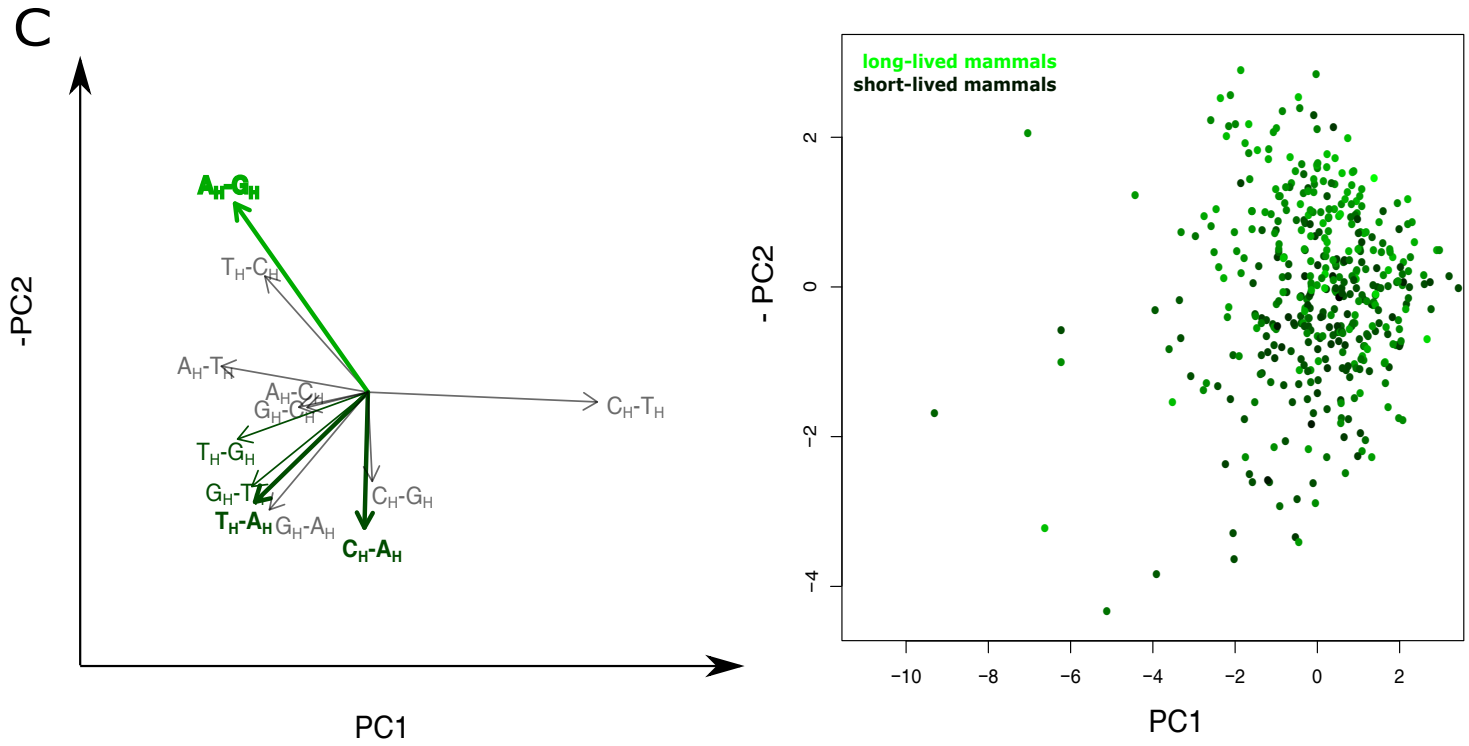

-PC2
PC1
A_H-G_H
T_H-C_H
A_H-T_H
A_H-C_H
G_H-C_H
T_H-G_H
G_H-T_H
T_H-A_H
G_H-A_H
C_H-A_H
C_H-G_H
C_H-T_H
long-lived mammals
short-lived mammals
- PC2
PC1
2
0
−2
−4
−10
−8
−6
−4
−2
0
2

Figure 1