A mitochondrial mutational signature of temperature in ectothermic and endothermic vertebrates.

Alina G. Mikhailova[1,2*], Dmitrii Iliushchenko[1*], Victor Shamansky[1*], Alina A. Mikhailova[1, 3], Kristina Ushakova[1], Evgenii Tretyakov[4], Sergey Oreshkov[1], Dmitry Knorre[5], Leonard Polishchuk[6], Dylan Lawless[7], Aleksandr Kuzmin[1], Stepan Denisov[1], Ivan Kozenkov[1], Ilya Mazunin[8], Wolfram Kunz[9,10], Masashi Tanaka[11,12,13], Vsevolod Makeev[2, 14], Rita Castilho[15], Valerian Yurov[1], Alexander Kuptsov[16], Jacques Fellay[7], Konstantin Khrapko[17], Konstantin Gunbin[1], Konstantin Popadin[1,6].

1. Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal University, Kaliningrad, Russia.
2. Vavilov Institute of General Genetics RAS, Moscow, Russia.
3. Institute for Evolution and Biodiversity, University of Münster, Münster, Germany.
4. Department of Molecular Neurosciences, Center for Brain Research, Medical University of Vienna, Vienna, Austria.
5. Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russian Federation.
6. Department of General Ecology and Hydrobiology, Biological Faculty, M.V. Lomonosov Moscow State University, Moscow, Russia.
7. Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland.
8. Center of Life Sciences, Skolkovo Institute of Science and Technology, Skolkovo, Russia.
9. Division of Neurochemistry, Department of Experimental Epileptology and Cognition Research, University Bonn, Bonn, Germany.
10. Department of Epileptology, University Hospital of Bonn, Bonn, Germany, Bonn, Germany.
11. Department for Health and Longevity Research, National Institutes of Biomedical Innovation, Health and Nutrition, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8636, Japan.
12. Department of Neurology, Juntendo University Graduate School of Medicine, 2-1-1 Hongo, Bunkyo-ku Tokyo 113-8421, Japan.
13. Department of Clinical Laboratory, IMS Miyoshi General Hospital, 974-3, Fujikubo, Miyoshi-machi, Iruma, Saitama Prefecture, 354-0041 Japan.
14. Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region.
15. The Algarve Centre of Marine Sciences, CCMAR-Algarve, Portugal.
16. Severtsov Institute of Ecology and Evolution RAS, Moscow, Russia.
17. Northeastern University, Boston, MA, USA.

ABSTRACT

Mutational spectrum of the mitochondrial genome (mtDNA) fluctuates between species, however, factors responsible for this variation are mainly unknown. Recently, we demonstrated that the mammalian mtDNA mutational spectrum is associated with age-related oxidative damage by means of A>G substitutions on a heavy chain (hereafter Ah>Gh). Here we extend this logic hypothesizing that oxidative damage can also depend on a species-specific level of aerobic metabolism. Using body temperature of endotherms and the environmental temperature of ectotherms as a proxy for their levels of aerobic metabolism, we reconstructed and analyzed 1350 species-specific mtDNA mutational spectra of vertebrate species. First, within ray-finned fishes, we observed that temperature is associated with increased asymmetry of Ah>Gh substitution, estimated as a ratio Ah>Gh/Th>Ch. Second, comparing within-species geographically distinct clades of widely distributed European anchovy, we observed similarly increased Ah>Gh/Th>Ch in the tropical clade. Third, analysing nucleotide composition in the most neutral synonymous sites of fishes, we demonstrated that warm- versus cold-water fishes are expectedly more AC poor and GT rich. Fourth, within mammals we observed an increased Ah>Gh/Th>Ch in warmer species as compared to colder ones (hybernators, daily torpors, naked mole rat, etc.) Fifth, comparing mtDNA mutational spectra between five classes of vertebrates, we observed an increased Ah>Gh/Th>Ch in warm- (mammals and birds) versus cold- (actinopterygii, amphibia, reptilia) blooded classes. Altogether, we conclude that temperature, through the level of the metabolism and oxidative damage, shapes the properties of the mitochondrial genome.

INTRODUCTION

High mitochondrial mutation rate provides a rich source of effectively neutral variants which are extensively used in tracing the history of species, populations, organisms within populations and more recently - cells in tissues (Ludwig et al. 2019). Here we propose to extend the utility of mtDNA neutral polymorphic data by deriving species-specific mutational spectra and correlating it with life-history traits. We have shown recently that mammalian mtDNA mutational spectrum, and precisely A_H>G_H substitution (hereafter index $_H$ marks mtDNA heavy chain annotation, Methods), is associated with mammalian lifespan (Mikhaylova et al. 2021) most likely through the aging-related oxidative damage. Here, we hypothesize that the mtDNA mutational spectrum can be also sensitive to oxidative damage through the species-specific rate of aerobic metabolism. Since the level of metabolism depends strongly on temperature (J. F. Gillooly et al. 2001) and there is experimental evidence that some mtDNA mutations can be sensitive to temperature (Zheng et al. 2006) we decided to analyze mtDNA mutational spectrum of ectothermic and endothermic vertebrates. Effects of temperature on mutation rate have been intensively studied for decades (Timoféeff-Ressovsky 1936; Lindgren 1972; James F. Gillooly et al. 2005; Chu et al. 2018; Belfield et al. 2021; Matsuba et al. 2013) however temperature-specific mutational signature - i.e. the effect of temperature on changes in mutational spectrum became a focus of first studies just recently (Waldvogel and Pfenninger 2021; Chu et al. 2018) where mtDNA was not under the focus.

RESULTS

1. Species-specific mtDNA mutational spectrum of ray-finned fishes is associated with ambient water temperature by means of increased asymmetry of A_H>G_H

In order to test the potential effect of temperature on mtDNA mutagenesis we first focused on ray-finned fishes (Actinopterygii) - ectothermic animals spanning a wide range of ambient water temperatures. Using within-species synonymous mtDNA polymorphisms, polarised on a phylogenetic tree, we derived a 12-component mutational spectrum for 551 Actinopterygii species (Methods, Supplementary File 1). The average mutational spectrum demonstrates a pronounced excess of transitions among which C_H>T_H is the most common and A_H>G_H is the second most common one (Fig. 1A), thus strongly resembling spectra of mammals (Mikhaylova et al. 2021) and human cancers (Yuan et al. 2020; Ju et al. 2014). For 128 out of 551 species, we

obtained mean annual water temperature (Methods, Supplementary File 2) and analysed its associations with 12 types of substitutions from mtDNA mutational spectrum. We observed two types of substitutions, significantly associated with temperature: $A_H>G_H$, positively correlated with temperature, and $T_H>C_H$, negatively correlated with temperature (Fig. 1B, Supplementary Mat. 1a, 1b, 1c). It is important to note that these substitutions are complementary equivalents: a substitution annotated as A>G on a heavy chain ($A_H>G_H$) might be originated as T>C on the light chain ($T_L>C_L$) and opposite: substitution annotated as T>C on a heavy chain ($T_H>C_H$) can happen as A>G on a light chain ($A_L>G_L$). If the probability of A to mutate into G is identical on both heavy and light chains, we expect to see symmetry: rate of $A_H>G_H$ equals rate of $T_H>C_H$ (both rates are normalised by the frequency of ancestral nucleotides). However, the observed excess of $A_H>G_H$ over $T_H>C_H$ ($A_H>G_H/T_H>C_H > 1$) indicates a mutagen, inducing A>G predominantly on the heavy chain ($A_H>G_H$) (Supplementary Fig. 1.1). Thus the ratio $A_H>G_H/T_H>C_H$ can be interpreted as the level of asymmetry of $A_H>G_H$ substitutions, and interestingly, this ratio demonstrates strong positive correlation with temperature (Fig. 1B, Supplementary Mat. 1c). The increased ratio of $A_H>G_H/T_H>C_H$ in warm- versus cold- water species shows that this asymmetrical extra mutagen is temperature sensitive. This trend is robust to phylogenetic inertia (Supplementary Mat. 1d), remains qualitatively similar when we analyse family-specific data and when we split the whole dataset into several subgroups based on temperature (Supplementary Fig. 1.2),

In mammalian species, it has been shown that $A_H>G_H$ positively correlates with generation length (Mikhaylova et al. 2021). To test a similar relationship in fishes, we used an analogous metric - the time of maturation, estimated as age at first maturity when 50% of the cohort spawn for the first time. Multiple linear models describing the mutational spectrum ($A_H>G_H$, $T_H>C_H$ and $A_H>G_H/T_H>C_H$) as a function of both temperature and the time of maturation demonstrated the only significant effect of temperature (Supplementary Mat. 1e, 1f, 1g). Thus we concluded that the Actinopterygii mtDNA mutational spectrum is associated with temperature.

2. Tropical clade of European anchovy shows an excess of $A_H>G_H$ and deficit of $T_H>C_H$ in mtDNA

Widely distributed species with genetic clinal variation over geographical ranges is a good model to test our findings on within-species scale. Intensive studies of the European anchovy (*Engraulis encrasicolus*) uncovered an existence of two mtDNA clades, clonally arranged along the eastern Atlantic Ocean: clade A is preferentially distributed in tropics, while clade B has an anti-tropical distribution, with frequencies decreasing towards the tropics (Silva et al. 2014). Analysing a fragment of cytochrome B for 285 organisms of clade A and 160 organisms of clade B (data generously provided by Dr. Rita Castilho) we reconstructed their mutational spectra: for each clade we reconstructed a consensus, and based on all synonymous deviations from consensus we derived the mutational spectra (Methods). Two types of substitutions demonstrated significant differences between the clades: $A_H>G_H$ was more common in clade A, while $T_H>C_H$ was less frequent in clade A (Fig. 2); frequencies of all other types of substitutions were similar. The observed pattern is completely in line with our previous findings, showing an excess of $A_H>G_H$ and deficit of $T_H>C_H$ in warm-water species (Fig. 1B).

The mutational spectrum with time can affect nucleotide composition, especially in the synonymous positions. Assuming that consensus sequences of clades A and B reflect ancestral state we compared their nucleotide content in synonymous positions. Totally we observed 21 synonymous differences between consensuses of clades A and B and, interestingly, nucleotide composition in these sites are completely in line with the expected mutational bias: tropical clade A versus B is more A poor (3 versus 10), G rich (9 versus 4), T rich (7 versus 1) and C poor (2 versus 6) (Supplementary Mat. 2a). All these changes might result from the clade-specific mutational spectra: an excess of $A_H>G_H$ and deficit of $T_H>C_H$ in tropical clade A (Fig 2). It is important to emphasize that observed divergence in the consensus sequences between clades A and B makes the difference in the mutational spectra even more pronounced and robust. For example, clade A demonstrates an excess of substitutions $A_H>G_H$ (deficit of $T_H>C_H$) despite the deficit of A_H (an excess of T_H) among synonymous sites of the consensus. Altogether we observed an increased asymmetry of $A_H>G_H$ in the tropical clade of European

anchovy as well as a line of evidence that the mutational bias affects the neutral nucleotide composition of the sequences.

3. The temperature sensitive mtDNA mutational spectrum of ray-finned fishes affects their neutral nucleotide composition: warm water species are more AC poor and GT rich

An anchovy example, described above shows that the nucleotide composition of recently diverged clades can be affected by the mutational spectrum. The same process can be visible also on a long-term evolutionary scale if mutational bias is stronger than selection and analyzed species are close to the nucleotide compositional equilibrium. For example, mutational spectrum characterized by increased A_H>G_H and decreased T_H>C_H in warm-water fishes, in the long-term perspective, may lead to a drop in A_H and rise in G_H as well as retention of relatively high fraction of T_H and low fraction of C_H.

In order to estimate how close the ray-finned fishes are to their compositional nucleotide equilibrium we compared expected and observed nucleotide compositions in the most neutral synonymous four-fold degenerate positions. First, we derived the typical mutational spectra for cold- and warm- water fishes taking the coldest (N=10, temp < 7.78℃) and the warmest (N= 11, temp > 27℃) deciles of the analysed species. As was anticipated, warm-water fishes show an excess of A_H>G_H and a deficit of T_H>C_H (Supplementary Mat. 3a). Second, based on these mutational spectra, using computer simulations and analytical solutions we derived the expected neutral nucleotide compositions for cold and warm-water fishes (both approaches converged to identical results, see Supplementary Fig. 3.1). By this, we addressed a question - what would be the nucleotide composition in the long-term perspective if the only factor affecting the molecular evolution is the mutagenesis (i.e. the specific mutational spectrum). Third, we obtained observed nucleotide composition, analysing synonymous four-fold degenerate nucleotide content in twelve (all except ND6) protein-coding genes of mtDNA from the coldest and warmest deciles, and compared it with the expected one (Fig. 3A). As a result we found that the observed nucleotide composition is rather similar with the expected one, which means that analyzed species are close enough to the compositional equilibrium. Moreover, we observed that warm-water species tend to be closer to the expected equilibrium (see the horizontal dotted lines on Fig. 3A) as compared to cold-water species, probably because warm water species have increased mutational rate and approach the equilibrium faster.

Taking into account that the mtDNA synonymous nucleotide composition of ray-finned fishes is close to the neutral equilibrium, and assuming that there is no strong selection on four-fold degenerate synonymous positions in mtDNA, we expect to observe a correlation between the nucleotide content and ambient temperature. Analysing synonymous four-fold degenerate nucleotide content in twelve (all except ND6) protein-coding genes of ray-finned fishes we observed all four expected trends: a decrease in A_H and C_H and an increase in G_H and T_H in warm- versus cold-water species (Fig. 3B, Supplementary Mat. 3b). It is important to note that all these changes reflect consequences of A_H>G_H, which is more asymmetrical in warm-water fishes (i.e. A_H>G_H is higher than T_H>C_H). To emphasize this we got a sum of the fractions of A_H and C_H (S_{AC}) and subtracted from it the sum of T_H and G_H (S_{TG}) thus combining all four nucleotide frequencies into a single metric S_{TG}-S_{AC} (Methods). Regressing S_{TG}-S_{AC} on temperature we obtained an expected strong relationship (Fig. 3C left panel, Supplementary Mat. 3c), which is robust to phylogenetic inertia (Supplementary Mat. 3d) and stays significant in two separate groups of fishes: short- and long-lived (Supplementary Fig. 3.2).

Keeping in mind the increased A_H>G_H in long- versus short-lived mammals (Mikhaylova et al. 2021) we tested a potential association of S_{TG}-S_{AC} with both temperature and the time of maturation among ray-finned fishes. The result of the multiple linear regression is the following:

$$S_{TG}\text{-}S_{AC} = 0.41 + 0.047*Temperature + 0.031*Time\ of\ maturation,\ all\ p\text{-}values < 9.8e\text{-}05,\ N=131\ (equation\ i)$$

We see that both factors affect the S_{TG}-S_{AC} positively, but the temperature has a higher impact on the S_{TG}-S_{AC} as compared to the time of maturation (all coefficients are standardized, Fig. 3C right panel, see also Supplementary Mat. 3e). Both temperature and the time of maturation marginally significantly affect S_{TG}-S_{AC} taking into account the phylogenetic inertia (Supplementary Mat. 3f). Of note, the effect of the time of maturation was not significant in our analysis of the 12-component mutational spectrum (chapter 1), probably due to the lower sample size (N = 65 instead of 131). Altogether, we conclude that the mtDNA nucleotide content of ray-finned fishes is affected by two factors: a strong temperature-dependent mutagen and a two-times weaker longevity-associated mutagen.

4. The mtDNA nucleotide composition of mammals is also partially shaped by temperature

To investigate the potential effect of temperature in homeotherms we focused on mammals - the class with the highest number of species with data available for both mutational spectrum and life-history traits. Mammalian body temperature, although more constant as compared to fishes, demonstrates some variation (Supplementary Fig. 4.1) which might affect the mtDNA mutational spectrum. Previously, we described a longevity mutational signature as a high frequency of A_H>G_H in long-lived mammals which makes their genomes A_H poor and G_H rich, increasing G_HA_H nucleotide skew $(G_H-A_H)/(G_H+A_H)$ (Mikhaylova et al. 2021). In this paper, analyzing Actinopterygii species, we observed a temperature mutational signature, realized in increased A_H>G_H and decreased T_H>C_H in warm-water species. The temperature mutational signature is similar to the longevity mutational signature (both of them have increased A_H>G_H), but they are not identical (only temperature signature is negatively associated with T_H>C_H). Thus, it is interesting to check if the temperature mutational signature can also differentiate mammals with different body temperature and generation length. First, we analysed 224 mammalian species with known complete mitochondrial genomes, generation lengths and body temperatures (Methods). A linear regression analysis of S_{TG}-S_{AC} as a function of generation length and temperature demonstrated a significant effect of the generation length but a marginally significant and weak effect of temperature (Supplementary Mat. 4a). Both these traits were not robust to phylogenetic inertia (Supplementary Mat. 4b). To overcome the limited number of mammalian species with known body temperature (N = 224) we split all mammalian species into two groups - “colder” and “warmer”, using an expert-curated annotation, which took into account not only the body temperature but also other life-history traits such as hibernation, daily torpor, etc (Methods). With the new metric we increased our sample size to 649 species and observed that S_{TG}-S_{AC} depends significantly on both generation length and the temperature group (Fig. 4A, Supplementary Mat. 4c), however, only the effect of the generation length was marginally significantly supported by the phylogenetically aware analysis (Supplementary Mat. 4d). Third, to test the effect of a temperature in a naive analysis, similar to the phylogenetically independent contrasts, we split all mammals into four quartiles based on the generation length and within each quartile compared S_{TG}-S_{AC} between ‘colder’ and ‘warmer’ subgroups. Interestingly, the expected trend, an increased S_{TG}-S_{AC} in warmer groups, was observed in all four quartiles, being significant in two of them and marginally significant in two others (Fig 4B). A minor modification of this analysis when we split mammals into ten groups based on deciles of the generation length and compared the S_{TG}-S_{AC} between ‘colder’ and ‘warmer’ subgroups within each decile gave similar results (Supplementary Mat. 4e). Thus, we suggest that the temperature-sensitive mutational signature S_{TG}-S_{AC} is weakly pronounced even in mammals.

5. The mtDNA mutational spectrum is temperature sensitive in all classes of vertebrates

The temperature-sensitivity of mtDNA mutational spectra within ray-finned fishes (Fig 1, 2, 3) and more weak but still a similar trend in mammals (Fig 4), opens a possibility that this phenomenon is rather universal. In order to test it we compared mtDNA mutational spectra between five classes of vertebrates, grouping them into cold-blooded (Actinopterygii, Amphibia, Reptilia) and warm-blooded (Mammalia, Aves) ones. Comparing mtDNA mutational spectrum (A_H>G_H, T_H>C_H , A_H>G_H/ T_H>C_H) between these groups, we observed all the

expected trends: excess of $A_H>G_H$, deficit of $T_H>C_H$ and increased $A_H>G_H$/ $T_H>C_H$ in warm-blooded classes (Fig 5), suggesting that mtDNA mutational spectrum is temperature sensitive in all vertebrates.

DISCUSSION

Till now the biggest collections of mitochondrial mutations were derived from human cancers (Yuan et al. 2020; Ju et al. 2014) and mice pedigrees (Arbeithuber et al. 2020), both of which showed high similarity in spectra with each other and our data. An evolutionary approach, used in our manuscript, allowed us to reconstruct the highest number of mtDNA mutations - dozens of thousands of substitutions in mtDNA of vertebrates. This dataset gave us a possibility to compare mutational spectra between species with contrasting life-history and physiological traits. Although the existence of between-species variation in the mtDNA mutational spectrum has been shown before (Belle et al. 2005), it was left without explanation. In this study, we assumed that the vast majority of fourfold degenerate synonymous substitutions is effectively neutral and thus different mutational spectra would reflect different mutagens rather than selection. To our best knowledge, there is no evidence of selection acting on synonymous codons in mtDNA of vertebrates; on the contrary, it has been shown that the mutational bias shapes the synonymous codon usage of the human mtDNA (Ju et al. 2014).

Polymerase-induced mutations are expected to be symmetrical and enriched in $A_H>G_H$ and $T_H>C_H$ (Lee and Johnson 2006). Asymmetric nature of mtDNA mutations observed in our and other studies (Arbeithuber et al. 2020), (Yuan et al. 2020; Ju et al. 2014) and positive relationship of $A_H>G_H$ with temperature and longevity requires a special explanation, involving an extra mutagen affecting predominantly the single-stranded heavy chain during mtDNA replication (Fig S1.1). The temperature sensitivity of the mutational spectrum, analyzed in our paper (Fig 1 - 5) might be explained by the increased oxidative damage in warm, highly-metabolic species (Martin and Palumbi 1993). Indeed, it has been shown experimentally on mouse cells (Shin and Turker 2002) and bacteria (Shewaramani et al. 2017) that oxidative damage predominantly increases adenine deamination especially on single-stranded DNA. Thus, higher adenine deamination ($A_H>G_H$) in species with increased temperature can be explained by their higher oxidative damage due to intensive aerobic metabolism. To test the effect of oxidative damage more directly we extracted the experimentally derived oxygen consumption (mg oxygen per kilogram fish per hour at 20°C at standard activity level) for several Actinopterygii species with known mutational spectrum. Despite the extremely low sample size (N = 10), we observed a positive trend between the oxygen consumption and $A_H>G_H$ (Supplementary Mat. 7).

The primary result of our paper - temperature effect on the mutational spectrum - can be explained through several straightforward logical steps: (i) temperature is associated with the level of aerobic metabolism, (ii) oxidative damage is a normal by-product of the aerobic metabolism, (iii) asymmetry of $A_H>G_H$ is increasing due to the oxidative damage. The increased fraction of transitions under high temperature conditions has been demonstrated in recent mutational accumulation experiments with Chironomus riparius and E. coli (Waldvogel and Pfenninger 2021; Chu et al. 2018), however an uncovering of more detailed mutational signatures of temperature requires future experimental studies.

The secondary result of our paper - longevity-associated changes in the mtDNA mutational spectrum is in line with our recent paper (Mikhaylova et al. 2021). Longevity effect, i.e. relatively higher adenine deamination (increased $A_H>G_H$) in long- versus short- lived species, can be based on three non-mutually exclusive scenarios: (i) increased oxidative damage in oocytes of long-lived species, (ii) increased impact of mtDNA replication-mediated mutations in short-lived species and (iii) increased fidelity of mtDNA polymerase of long-lived species. The first scenario is supported by the recent experimental discovery that aged mouse oocytes are characterized by significantly increased $A_H>G_H$ as compared to young mouse oocytes (Arbeithuber et al. 2020). An extension of the results from young and old mouse oocytes to short- and long- lived species will bring us to the observed correlations (Fig 4, 6) due to age-associated oxidative damage. The second scenario assumes that short-lived species per unit of time have an excess of replication-driven mtDNA mutations, which are expected to be symmetrical ($A_H>G_H \sim T_H>C_H$, Supplementary Fig 1.1). This replication-driven input increases

the overall symmetry of spectra diminishing the relative effect of asymmetrical oxidative damage-induced mutations in short-lived species. The third scenario is similar with the second one since it is associated with an excess of symmetrical replication-driven mutations in short-lived species, however it is based on higher fidelity of mtDNA polymerase in long-lived species (Nabholz, Glémin, and Galtier 2008), which could evolve to support decelerated somatic mutagenesis of mtDNA in long-lived large-bodied species.

Taking into account that all our observations (effects of temperature and longevity) might be explained through the same mediator - oxidative damage, which facilitates A>G substitutions on a single stranded heavy chain of mtDNA, we consider the oxidative damage explanation as the most parsimonious one. In other words, we propose that oxidative mutational signature in mtDNA is a function of both: metabolism-associated damage (approximated by temperature) and aging-associated damage (approximated by longevity) (Fig 6B). In order to derive one integral model we merged datasets of Actinopterygii and Mammals, having for each species S_{TG}-S_{AC}, temperature and a longevity proxy (time of maturation for fishes and generation length for mammals). We observe that S_{TG}-S_{AC} depends on temperature and longevity (Fig 6A), but not on the class of species (Supplementary Mat. 6). This suggests that mtDNA mutagenesis of different vertebrates follows very similar universal rules, shaped by temperature and longevity, but not by belonging to a specific phylogenetic group. Thus from the point of view of the mtDNA mutational spectrum, mammals mutate as warm fishes. Assuming a limited amount of oxidative damage (approximated by S_{TG}-S_{AC}) per lifespan of all species, it is expected that there is a trade-off between temperature and longevity - a correlation which is well known on within (Waldvogel and Pfenninger 2021) and between species scales (Keil, Cummings, and de Magalhães 2015). Using S_{TG}-S_{AC} as a marker of the oxidative damage we compared effect of the temperature and effect of the longevity on the oxidative damage and obtained roughly that an increase in temperature in 1 degree of celsius increases oxidative damage in the comparable level as an increase in longevity in one year (Supplementary materials 6). Thus, we propose that the mtDNA mutational spectrum, associated with both temperature and longevity, can help to better understand the mechanisms underlying the trade-off between temperature and longevity (Keil, Cummings, and de Magalhães 2015).

An oxidative damage is a normal by-product of aerobic metabolism and the higher the metabolism, the higher the damage is expected. However, numerous antioxidant protective mechanisms may evolve in parallel with the increased level of the metabolism, partially or completely compensating the increasing oxidative damage, and thus decreasing the mutational rate. Observed in our paper correlations between several components of the mutational spectrum and temperature confirms the direct logic: the higher the metabolism, the higher the damage, demonstrating that the antioxidant protective mechanisms even if they compensate the oxidative damage, do it partially. This suggests that on the comparative-species scale, the fluctuations in the oxidative damage can be an effectively neutral trait not significantly affecting fitness of the analysed species.

A sensitivity of mtDNA mutational spectrum to environmental and life-history traits opens a possibility to use this metric in ecological and evolutionary studies. The reconstruction of the mtDNA mutational spectrum from neutral segregating polymorphisms of each species is rather straightforward and simple (our software is under preparation) which can shed light on species-specific traits. For example, we ranked all analyzed Actinopterygii species by A_H>G_H and compared species from the bottom and top deciles. Expectedly we observed that top decile (with high A_H>G_H) consists of many large-bodied freshwater species inhabiting predominantly tropic lakes and rivers: *Labeo gonius*, which spend a lot of time in Tropical climate zone; *Brachyhypopomus occidentalis* frequently found in shallow and calm streams; *Petrochromis trewavasae*, who lives in african *Lake Tanganyika*; *Girardinus metallicus* who lives in ponds, lakes and streams of Cuba in Central America; reef fish, such as *Mugil curema*, *Amblyeleotris wheeleri* and *Beaufortia kweichowensis*, living in shallow waters with oxygen-saturated water; *Lepisosteus oculatus* which is capable to compensate for the lack of oxygen by gulping air bubbles into a primitive lung called a gas bladder. Contrary, species from the bottom decile (with low A_H>G_H) are mainly small-sized sea or deepwater lake fishes such as: *Oncorhynchus mykiss*, which inhabit clear, cold headwaters, creeks, small to large rivers, lakes, and intertidal areas, usually not stocked in water reaching

summer temperatures above 25°C or ponds with very low oxygen concentrations; *Hippoglossoides robustus*, that lives on bottoms at depths of up to 425 metres where temperature reach up to 8℃, *Comephorus dybowskii* from Baikal lake found beyond 1,000 m depth. Also, a similar trend can be observed comparing unusual mammals such as a naked mole rat *Heterocephalus glaber* and its relative *Cryptomys hottentotus*. *Heterocephalus glaber,* being a cold-blooded mammal (32.1 ºC) with extended lifespan (average generation length is 2023 days) has lower frequencies of $A_H>G_H$ (0.1070143) as compared to it's more warm-blooded (34.4 ºC) and more short-lived (1606 days) relative *Cryptomys hottentotus* ($A_H>G_H$ 0.1464897). This suggests, that despite the existence of a global trade-off between the temperature and longevity (visualized in scheme Fig 6B), the level of oxidative damage (approximated through $A_H>G_H$) is decreased in naked mole rate, which is in line with the previous conclusions that not only decreased temperature but some additional factors (probably affecting the decreased oxidative damage) contribute to the extended lifespan of the naked mole rat (Keil, Cummings, and de Magalhães 2015).

FIGURE CAPTIONS

FIG1 mtDNA mutational spectrum of Actinopterygii species (mtDNA heavy chain notation).

(A) average mutational spectrum of all analyzed Actinopterygii species;
(B) species-specific mutational spectrum is associated with water temperature. Left panel: positive correlation of A_H>G_H with temperature (Spearman's rho = -0.36, p-value = 3.3e-05, N = 128); middle panel: negative correlation of T_H>C_H with temperature (Spearman's rho = 0.26, p-value = 0.0025, N = 128); right panel: positive correlation of log2(Ah>Gh/Th>Ch) with temperature (Spearman's rho = 0.44, p-value = 3.1e-07, N = 123)

FIG2 mtDNA mutational spectrum of European anchovy clades.

(A) Synonymous substitutions, reconstructed for clades A and B were split into five groups: four types of transitions and one group of transversions (all transversions were merged together due to their rarity). More tropically distributed clade A shows an excess of A_H>G_H (Fisher's odds ratio = 1.81, p-value = 0.049) and a deficit of T_H>C_H (Fisher's odds ratio = 0.48, p-value = 0.0025). Frequencies of all other types of substitutions were similar (all p-values > 0.5).
(B) Analysis of only two fluctuating types of transitions: A_H>G_H versus T_H>C_H between clades A and B emphasizes an excess of A_H>G_H as compared to T_H>C_H in tropical clade A (Fisher's odds ratio = 2.6, p = 0.005). One asterisk denotes p value < 0.05, two asterisks mark p value < 0.01. Two asterisks mark p value < 0.01

FIG3 mtDNA mutational spectrum of Actinopterygii (mtDNA heavy chain notation).

(A) observed and expected nucleotide compositional equilibrium of warm- and cold- water fishes. Expected nucleotide composition has been estimated based on computational simulations and analytical solutions (Supplementary Figure 3.1)
(B) whole-genome neutral nucleotide content is associated with ambient temperature (A_H: Spearman's rho = 0.117, p-value = 0.03301; T_H: Spearman's rho = -0.153, p-value = 0.00537; G_H: Spearman's rho = -0.249, p-value = 4.711e-06; C_H: Spearman's rho = 0.132, p-value = 0.01655, N = 123).
(C) S_{TG}-S_{AC} is sensitive to both temperature and the time of maturation, temperature being the strongest factor. Left panel shows a positive relationship between S_{TG}-S_{AC} and temperature for 333 species (statistical details are in supplementary Mat. 3d and 3c). Right panel demonstrates similar positive slopes between S_{TG}-S_{AC} and temperature for early and late maturing fishes (N = 131), where late-maturing ones have an increased intercept (statistical details are in supplementary Mat. 3e and 3f).

FIG4

(A) S_{TG}-S_{AC} among mammalian species is sensitive to both the generation length and temperature, generation length being the strongest factor (statistical details are in Supplementary Mat. 4c and 4d).
(B) S_{TG}-S_{AC} of mammalian species divided into 8 groups. S_{TG}-S_{AC} tends to be higher in warmer versus colder mammals in each category of species grouped into quartiles of the generation length. 1-st quartile: 47 colder versus 116 warmer species, p-value = 3.548e-05; 2-nd quartile: 47 colder versus 117 warmer species, p-value = 0.07805; 3-rd quartile: 23 colder versus 155 warmer species, p-value = 0.01811; 4-th quartile: 23 colder versus 121 warmer species, p-value = 0.07046. P-values between colder and warmer subgroups in each quartile were obtained using one-sided Mann-Whitney U test with a-priori hypothesis that warmer species have an increased S_{TG}-S_{AC}.

FIG5 mtDNA mutational spectrum of all Vertebrates as a function of the average class-specific body temperature (the bottom panel). Between-class comparison shows an excess of A_H>G_H (the first panel), deficit of T_H>C_H (the second panel), an excess of A_H>G_H/T_H>C_H (the third panel) in warm-versus cold-blooded species. All p-values < $2*10^{-13}$, Mann-Whitney U test.

FIG6

(A) S_{TG}-S_{AC} among Actinopterygii and Mammals as a function of temperature and generation length (statistical details are in Supplementary Mat. 6).
(B) A scheme visualizing that S_{TG}-S_{AC}, as a proxy of oxidative damage, can help to explain a trade-off between temperature and generation length.

SUPPLEMENTARY FILES

Supplementary file 1 A complete collection of 12-component mtDNA mutational spectra for 551 Actinopterygii species, used in the current study. For different analyses in the paper different subsets of this list were used depending on availability of additional parameters such as ambient temperature, time of maturation etc.

Supplementary file 2 A subset of the supplementary file 1 with 169 Actinopterygii species with known ambient temperature AND time of maturation.

METHODS

The widely accepted annotation of mitochondrial genomes is based on the light chain, however it is known that the majority of mitochondrial mutations occur on a heavy chain (Faith and Pollock 2003). In order to emphasize the chemical nature of observed mutations, we presented all substitutions with heavy-chain notation.

The mutational spectrum reconstruction has been described in detail in our accompanying paper (Mikhaylova et al. 2021). Here, we describe the logic of the analysis briefly. The mutational spectrum (a probability of one nucleotide to mutate into any other nucleotide) for each Vertebrate species was derived from all available intraspecies sequences (as of April 2018) of mitochondrial protein-coding genes. Using this database with intraspecies polymorphisms and developed pipeline we 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. Using species with at least 15 single-nucleotide synonymous mutations at four-fold degenerate sites we estimated the mutational spectrum for more than a thousand chordata species. We normalized observed frequencies by nucleotide content in the third position of four-fold degenerative synonymous sites of a given gene. Details of the pipeline are described in our recent paper (Mikhaylova et al. 2021) and the separate methodological paper is under preparation.

All statistical analyses were performed in R using Spearman rank correlations and Multiple models (for mammalian analyses dummy variables for each group were used). PGLS method (package “caper”, version 1.0.1) was used for the analysis of phylogenetic inertia.

The annual mean environmental (water) temperature in Celsius and time of maturation in years (mean or median age at first maturity, at which 50% of a cohort spawn for the first time) for fishes were downloaded from https://www.fishbase.se/ (at September 2019).

Mutational spectra of the European anchovy clades were estimated using a simplified algorithm, as compared to the between-species analyses described above. First, for each clade (285 organisms of clade A and 160 organisms of clade B, provided by Dr. Rita Castilho) we reconstructed a consensus of the region of MT-CYB, assuming that consensus reflects the ancestral state. Second, within each clase we counted all synonymous deviations from consensus (irrespectively of the frequency of the minor allele - from singletons to alleles with MAF < 50%), splitting these SNPs into five groups: four types of transitions and a group of all transversions (as

in Fig 2) which were used in the analyses. We didn't normalize the observed substitutions by the nucleotide content taking into account that compared sequences and clades are very similar.

S_{TG}-S_{AC} was calculated as a difference between sums of pairs of relative nucleotide frequencies: (sum(FrA+FrC) - sum(FrT+FrG))/(sum(FrA+FrC) + sum(FrT+FrG)), where (sum(FrA+FrC) + sum(FrT+FrG) = 1.

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 https://datadryad.org/stash/dataset/doi:10.5061/dryad.gd0m3 (at April 2018). Mammalian body temperatures and life-history traits, associated with the level of metabolism, were collected from AnAge (Tacutu et al. 2018) with manual annotation of missing data.

ACKNOWLEDGMENTS

K.P., A.G.M, V.Sh. were 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. 18-29-13055 to K.P, A.G.M, No. 18-04-01143 to K.P., A.A.M, L.P. and No. 19-29-04101 to I.M].

REFERENCES:

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.

Belfield, Eric J., Carly Brown, Zhong Jie Ding, Lottie Chapman, Mengqian Luo, Eleanor Hinde, Sam W. van Es, et al. 2021. "Thermal Stress Accelerates Mutation Rate." *Genome Research* 31 (1): 40–50.

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.

Chu, Xiao-Lin, Bo-Wen Zhang, Quan-Guo Zhang, Bi-Ru Zhu, Kui Lin, and Da-Yong Zhang. 2018. "Temperature Responses of Mutation Rate and Mutational Spectrum in an Escherichia Coli Strain and the Correlation with Metabolic Rate." *BMC Evolutionary Biology*. https://doi.org/10.1186/s12862-018-1252-8.

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

Gillooly, James F., Andrew P. Allen, Geoffrey B. West, and James H. Brown. 2005. "The Rate of DNA Evolution: Effects of Body Size and Temperature on the Molecular Clock." *Proceedings of the National Academy of Sciences of the United States of America* 102 (1): 140–45.

Gillooly, J. F., J. H. Brown, G. B. West, V. M. Savage, and E. L. Charnov. 2001. "Effects of Size and Temperature on Metabolic Rate." *Science* 293 (5538): 2248–51.

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.

Keil, Gerald, Elizabeth Cummings, and João Pedro de Magalhães. 2015. "Being Cool: How Body Temperature Influences Ageing and Longevity." *Biogerontology* 16 (4): 383–97.

Lee, Harold R., and Kenneth A. Johnson. 2006. "Fidelity of the Human Mitochondrial DNA Polymerase." *The Journal of Biological Chemistry* 281 (47): 36236–40.

Lindgren, D. 1972. "The Temperature Influence on the Spontaneous Mutation Rate. I. Literature Review." *Hereditas* 70 (2): 165–78.

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.

Matsuba, Chikako, Dejerianne G. Ostrow, Matthew P. Salomon, Amit Tolani, and Charles F. Baer. 2013. "Temperature, Stress and Spontaneous Mutation in Caenorhabditis Briggsae and Caenorhabditis Elegans." *Biology Letters* 9 (1): 20120334.

Mikhaylova, A. G., A. A. Mikhailova, K. Ushakova, E. O. Tretiakov, V. Shamansky, A. Yurchenko, M. Zazhytska, et al. 2021. "Mammalian Mitochondrial Mutational Spectrum as a Hallmark of Cellular and Organismal Aging." *bioRxiv*. https://doi.org/10.1101/589168.

Nabholz, Benoit, Sylvain Glémin, and Nicolas Galtier. 2008. "Strong Variations of Mitochondrial Mutation Rate across Mammals--the Longevity Hypothesis." *Molecular Biology and Evolution* 25 (1): 120–30.

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.

Silva, Gonçalo, Fernando P. Lima, Paulo Martel, and Rita Castilho. 2014. "Thermal Adaptation and Clinal Mitochondrial DNA Variation of European Anchovy." *Proceedings. Biological Sciences / The Royal Society* 281 (1792). https://doi.org/10.1098/rspb.2014.1093.

Tacutu, Robi, Daniel Thornton, Emily Johnson, Arie Budovsky, Diogo Barardo, Thomas Craig, Eugene Diana, et al. 2018. "Human Ageing Genomic Resources: New and Updated Databases." *Nucleic Acids Research* 46 (D1): D1083–90.

Timoféeff-Ressovsky, N. W. 1936. "Qualitativer Vergleich Der Mutabilität vonDrosophila Funebris

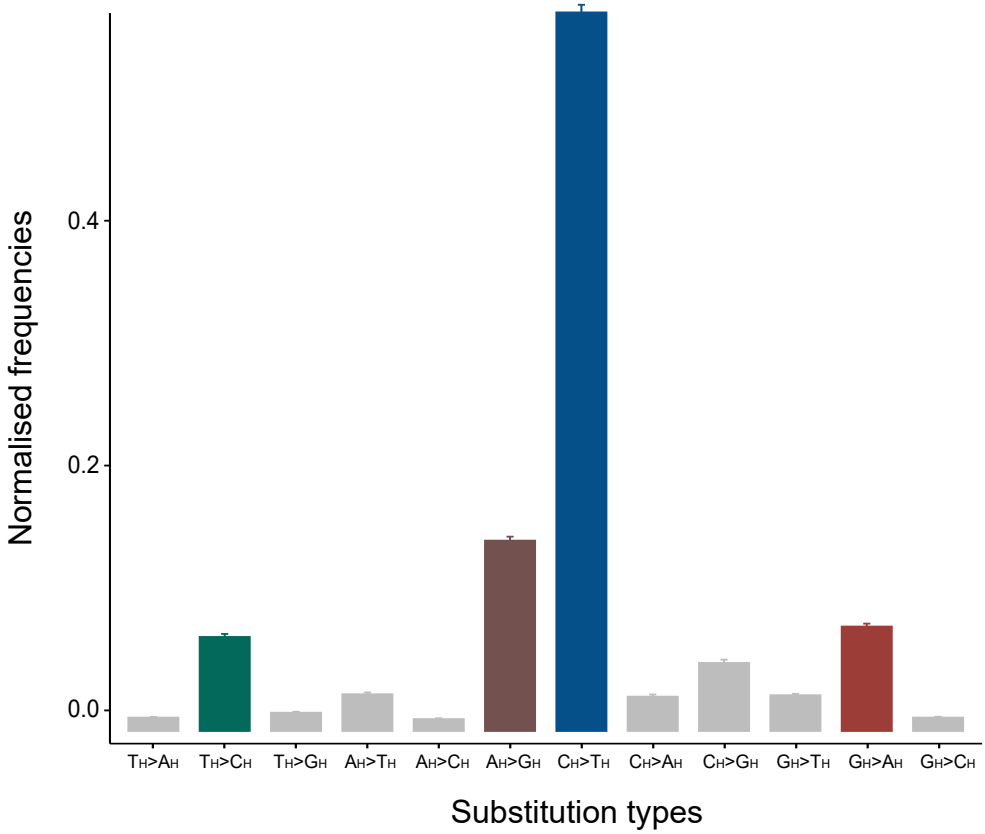
undDrosophila Melanogaster." *Zeitschrift Für Induktive Abstammungs- Und Vererbungslehre*. https://doi.org/10.1007/bf01848865.

Waldvogel, Ann-Marie, and Markus Pfenninger. 2021. "Temperature Dependence of Spontaneous Mutation Rates." *Genome Research* 31 (9): 1582–89.

Yuan, Yuan, Young Seok Ju, Youngwook Kim, Jun Li, Yumeng Wang, Christopher J. Yoon, Yang Yang, et al. 2020. "Comprehensive Molecular Characterization of Mitochondrial Genomes in Human Cancers." *Nature Genetics* 52 (3): 342–52.

Zheng, Weiming, Konstantin Khrapko, Hilary A. Coller, William G. Thilly, and William C. Copeland. 2006. "Origins of Human Mitochondrial Point Mutations as DNA Polymerase Gamma-Mediated Errors." *Mutation Research* 599 (1-2): 11–20.

A.



B.

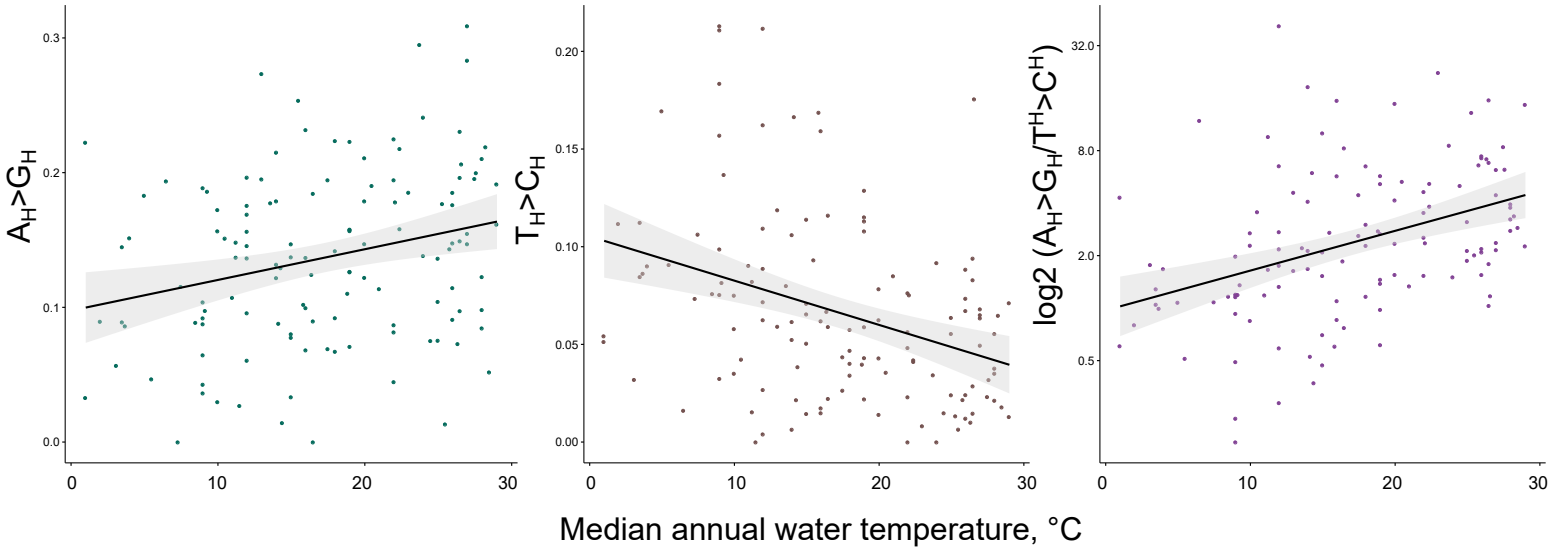


Figure 1

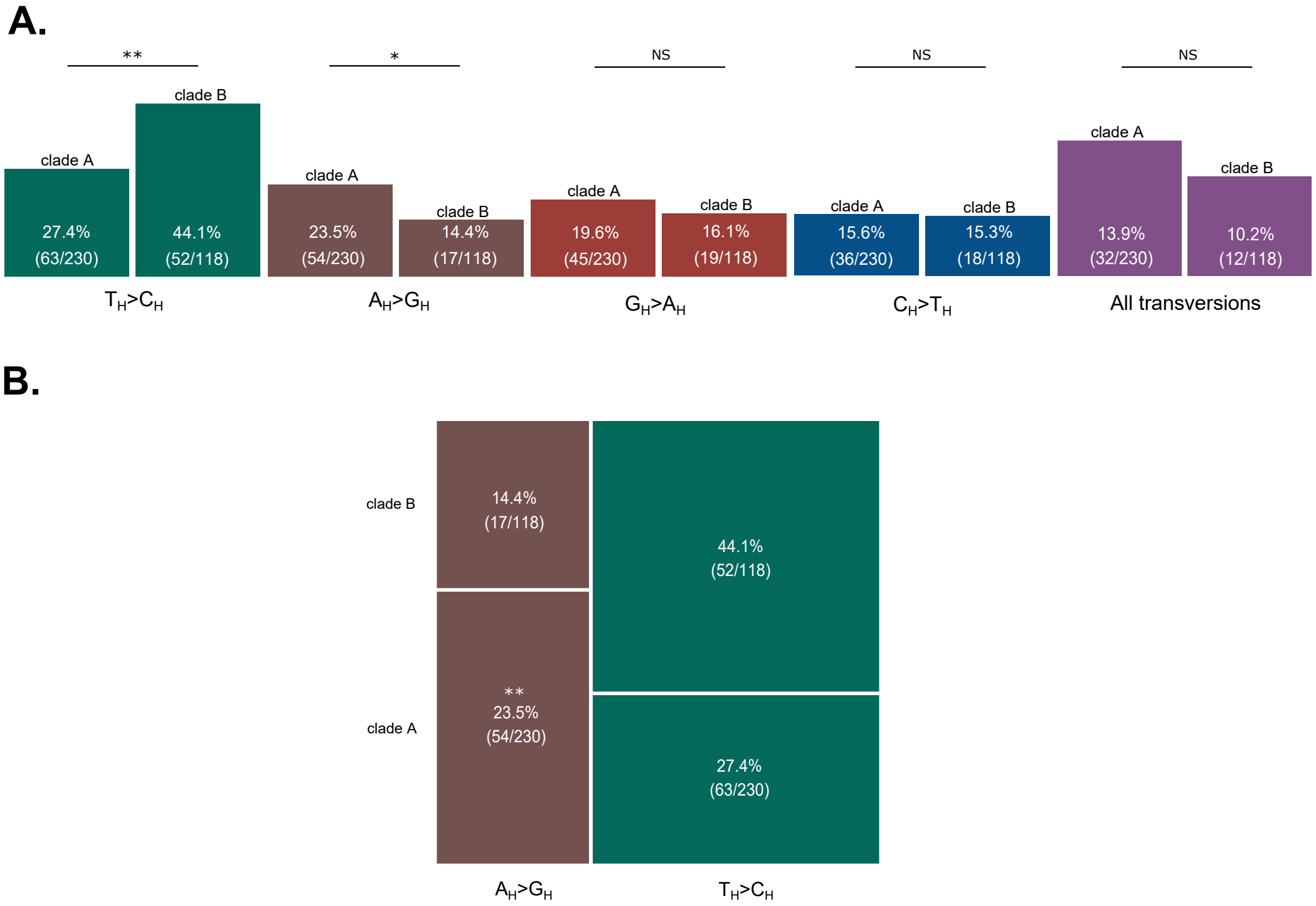
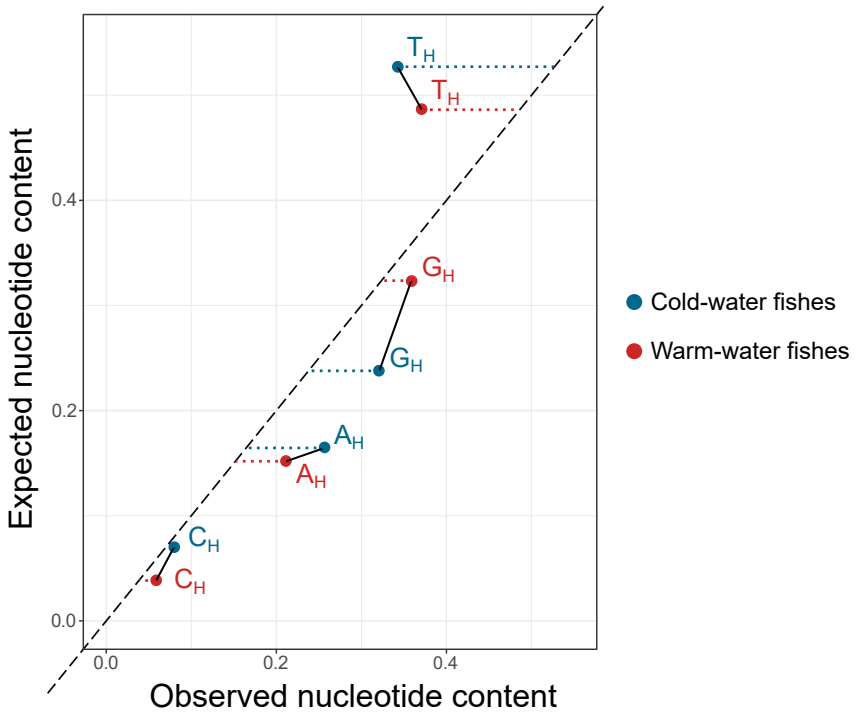
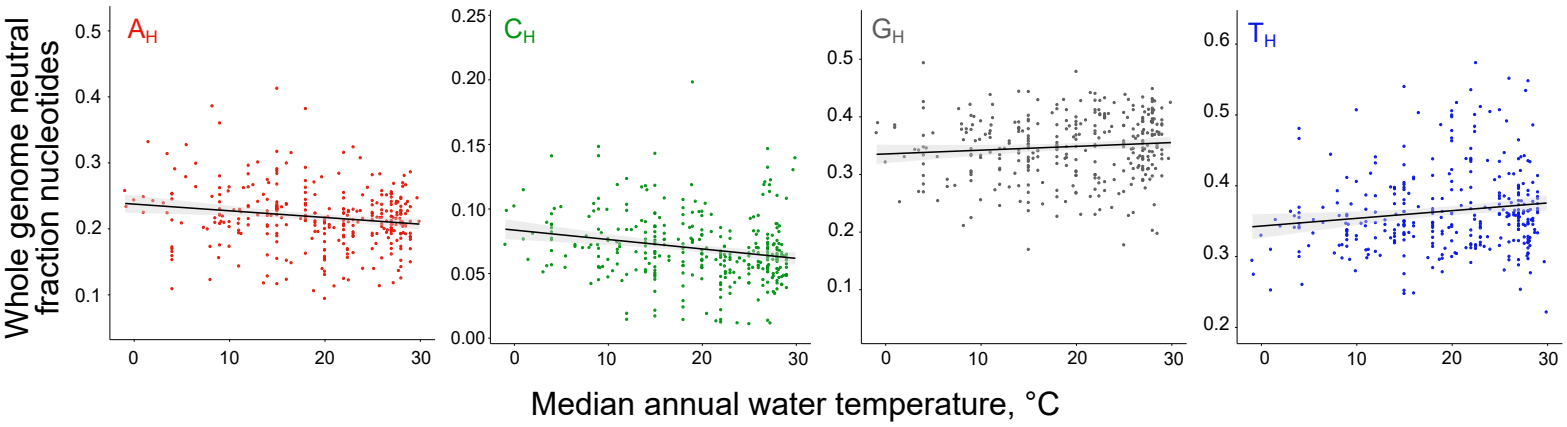

A.
**
*
NS
NS
NS
clade A
clade B
27.4%
(63/230)
44.1%
(52/118)
23.5%
(54/230)
14.4%
(17/118)
19.6%
(45/230)
16.1%
(19/118)
15.6%
(36/230)
15.3%
(18/118)
13.9%
(32/230)
10.2%
(12/118)
T_H>C_H
A_H>G_H
G_H>A_H
C_H>T_H
All transversions
B.
clade B
clade A
14.4%
(17/118)
44.1%
(52/118)
**
23.5%
(54/230)
27.4%
(63/230)
A_H>G_H
T_H>C_H

Figure 2

A.



B.



C.

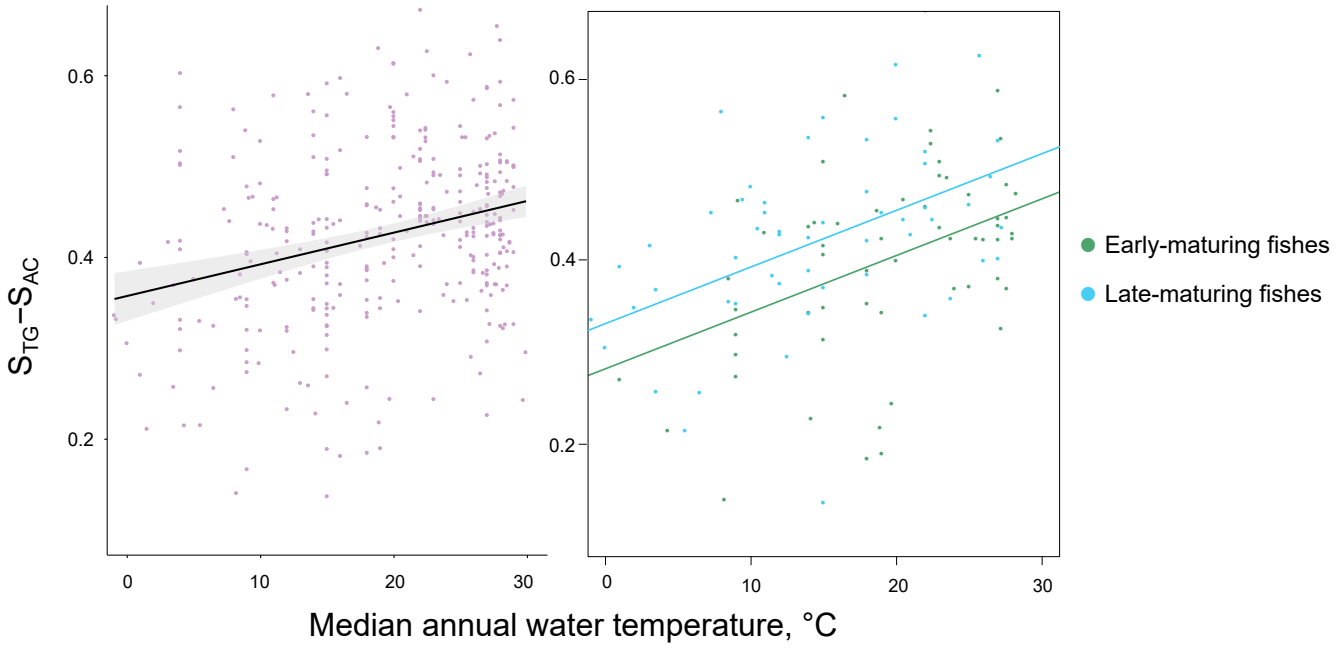
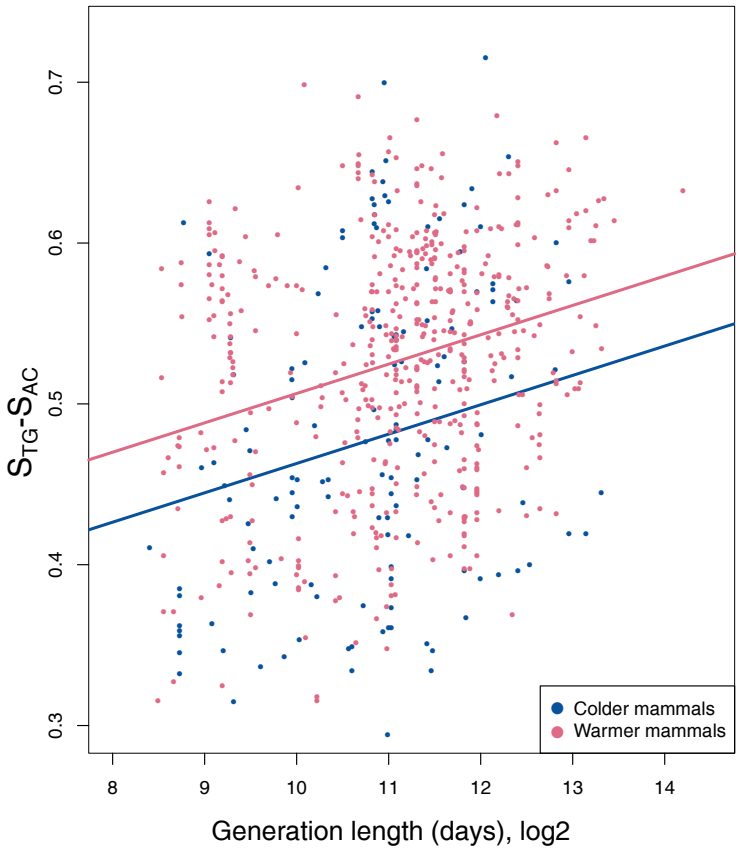


Figure 3

A.



B.

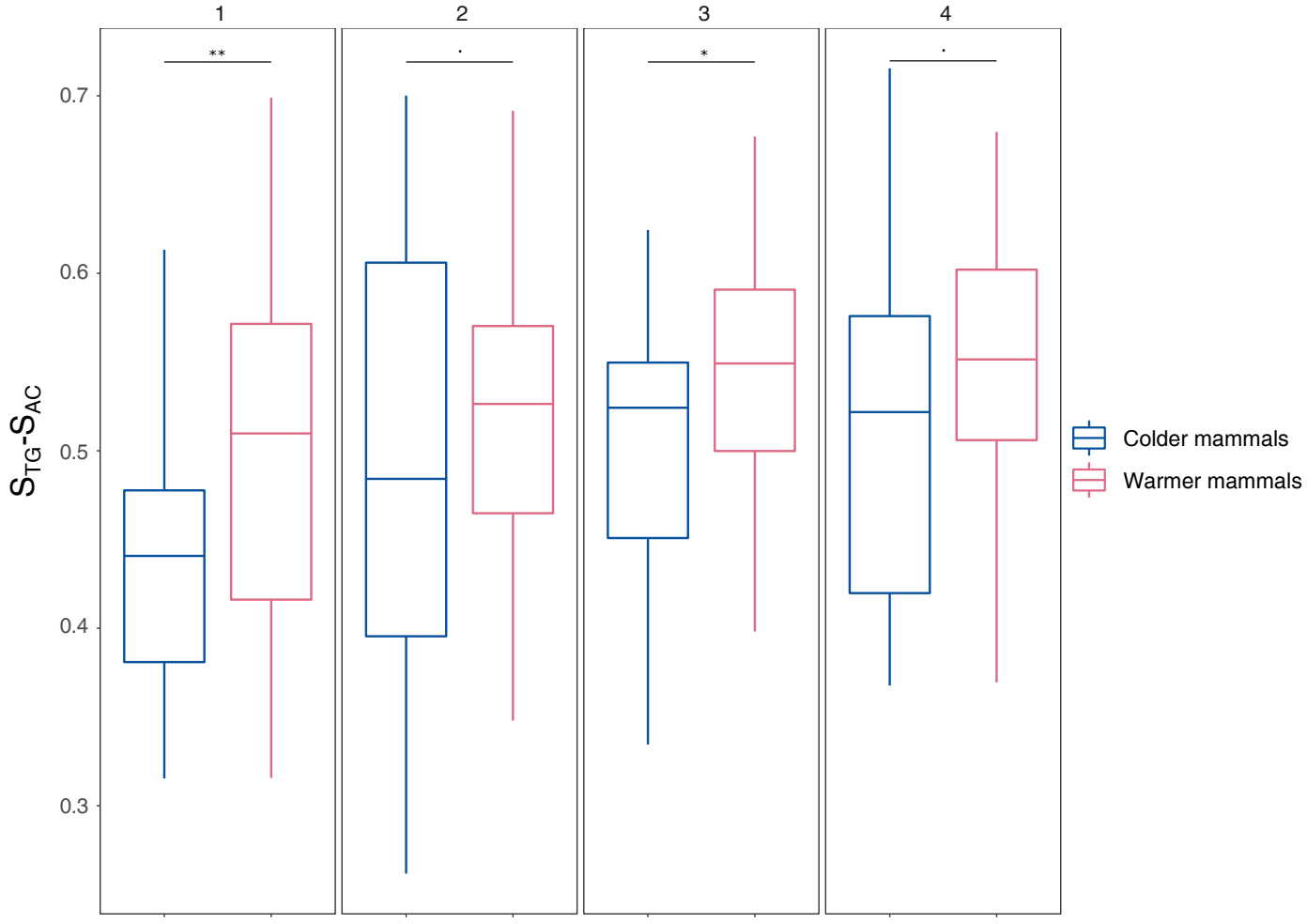


Figure 4

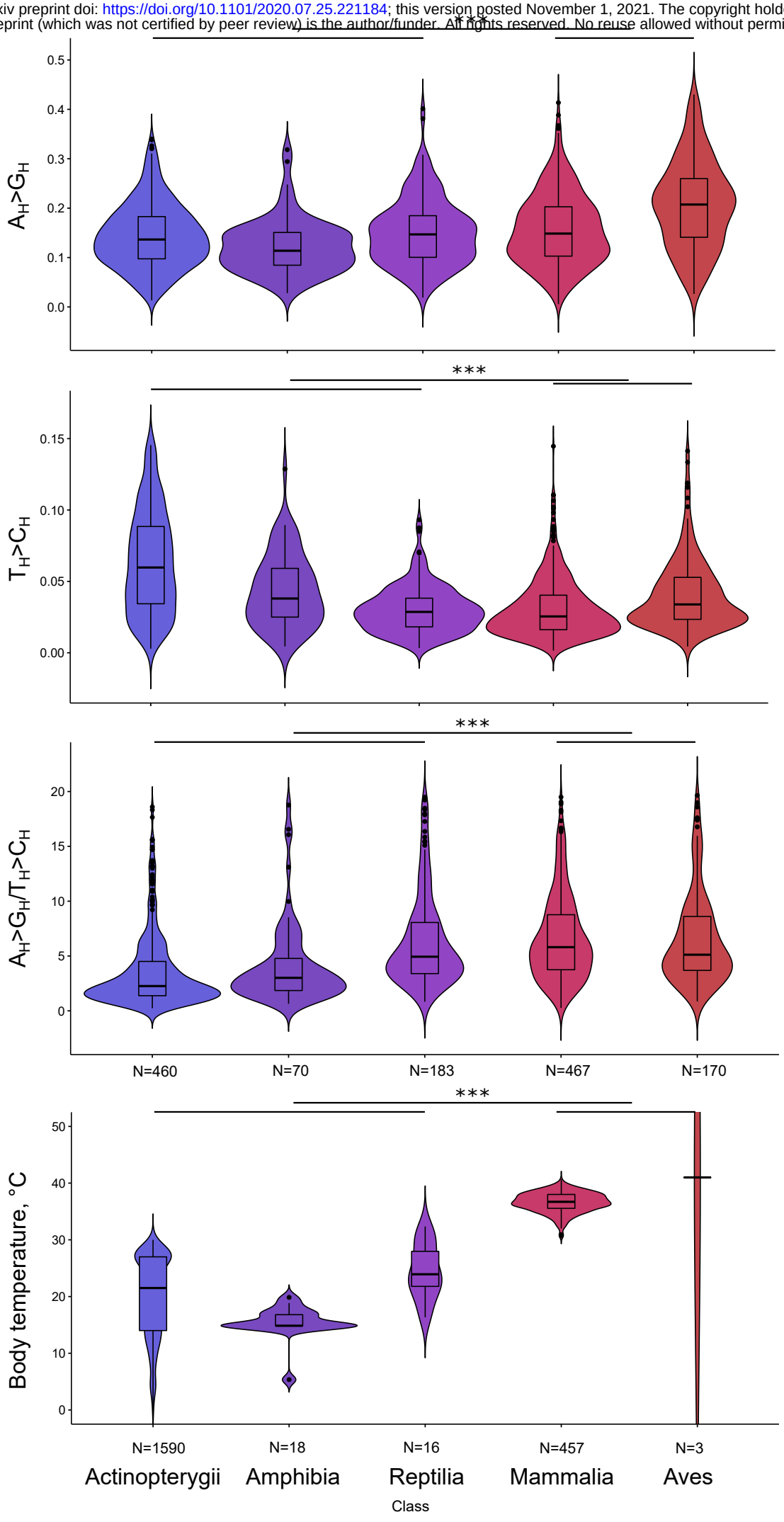


Figure 5

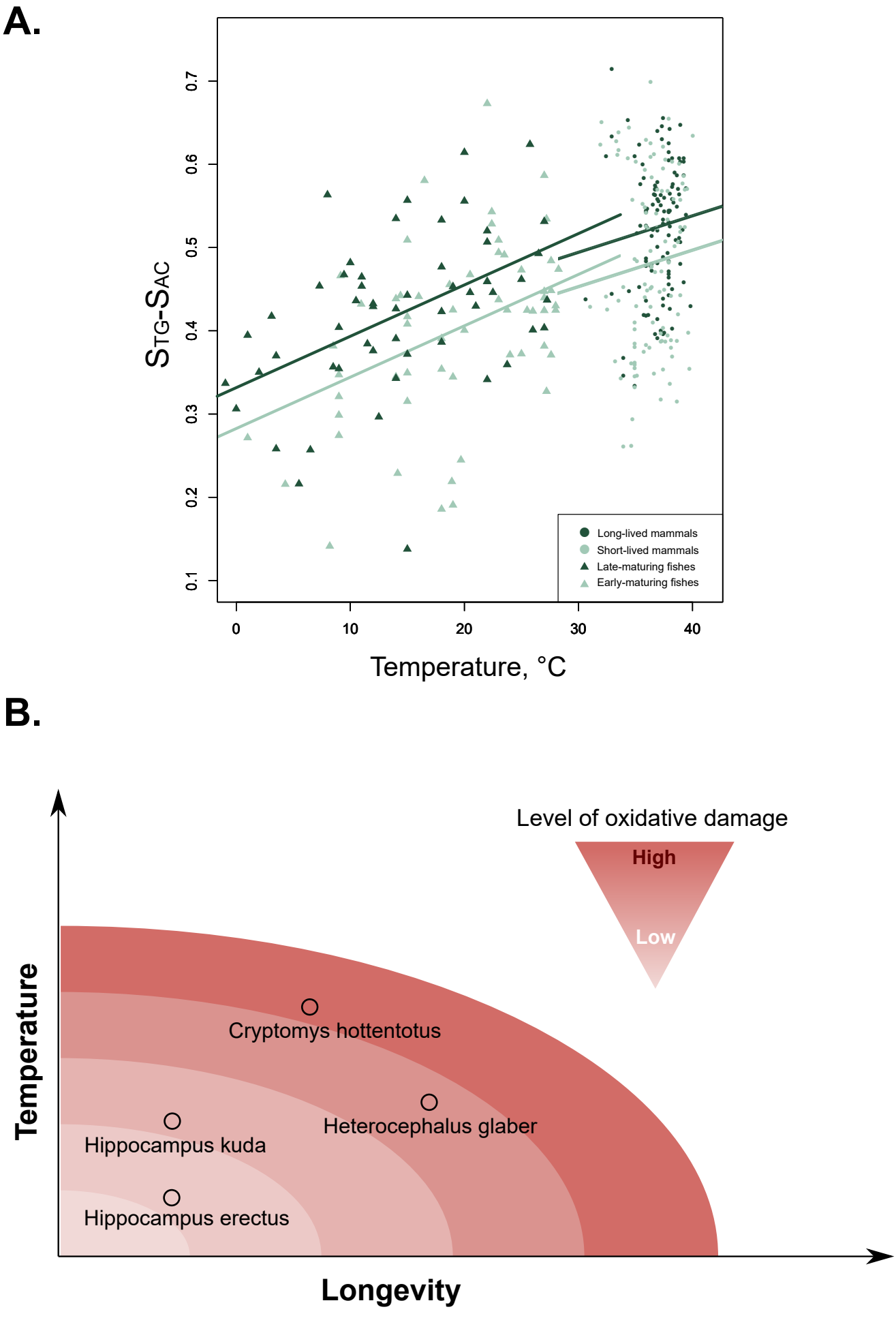


Figure 6