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

Alina G. Mikhailova[1,2], Victor Shamansky[1], Alina A. Mikhailova[1], Kristina Ushakova[1], Evgenii Tretyakov[3], Sergey Oreshkov[1], Dmitry Knorre[4], Leonard Polishchuk[5], Dylan Lawless[6], Ilia Mazunin[7], Wolfram Kunz[8,9], Masashi Tanaka[10,11,12], Zoë Fleischmann[13], Dylan Aidlen[13], Vsevolod Makeev[2, 14], Alexander Kuptsov[15], Jacques Fellay[6], Konstantin Khrapko[13], Konstantin Gunbin[1,16], 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. Department of Molecular Neurosciences, Center for Brain Research, Medical University of Vienna, Vienna, Austria.
4. Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russian Federation.
5. Department of General Ecology and Hydrobiology, Biological Faculty, M.V. Lomonosov Moscow State University, Moscow, Russia.
6. Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland.
7. Center of Life Sciences, Skolkovo Institute of Science and Technology, Skolkovo, Russia.
8. Division of Neurochemistry, Department of Experimental Epileptology and Cognition Research, University Bonn, Bonn, Germany.
9. Department of Epileptology, University Hospital of Bonn, Bonn, Germany, Bonn, Germany.
10. Department for Health and Longevity Research, National Institutes of Biomedical Innovation, Health and Nutrition, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8636, Japan.
11. Department of Neurology, Juntendo University Graduate School of Medicine, 2-1-1 Hongo, Bunkyo-ku Tokyo 113-8421, Japan.
12. Department of Clinical Laboratory, IMS Miyoshi General Hospital, 974-3, Fujikubo, Miyoshi-machi, Iruma, Saitama Prefecture, 354-0041 Japan.
13. Northeastern University, Boston, MA, USA.
14. Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region.
15. Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Moscow, Russia.
16. Institute of Molecular and Cellular Biology SB RAS, Novosibirsk, Russia.

ABSTRACT

Mitochondrial mutational signature is very conserved and low deviations between species have been associated with longevity. By reconstructing species-specific mtDNA mutational spectrum for ray-finned fishes (Actinopterygii), we observed that temperature is a strong additional factor shaping the mtDNA mutational spectrum in ectotherms. The analysis of mammalian endotherms, with a special focus on species with temporarily or permanently low metabolic rates (hibernators, daily torpors, naked mole rat, etc.), confirmed the temperature effect, suggesting that two main factors shape between-species variation in mitochondrial mutational spectra: longevity and temperature.

MAIN

High mitochondrial mutation rate provides a rich source of variants, which are extensively used in tracing the history of species, populations, and more recently - cells in tissues (Ludwig et al. 2019). Here we propose an extension of utility for mtDNA polymorphic data in GenBank by deriving species-specific mutational spectra for many vertebrates. 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 different aspects of longevity, such as mammalian lifespan, human maternal age and tissue-specific turnover time (Mikhailova et al. 2019). Here, we hypothesise that mtDNA mutation spectrum can be sensitive to species-specific metabolic rate. Since the level of metabolism depends strongly on temperature (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.

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 neutral mtDNA polymorphisms we derived a 12-component mutational spectrum for 128 Actinopterygii species (Methods). The observed average mutational spectrum (Fig. 1a) 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, thus strongly resembling spectra of mammals (Mikhailova et al. 2019) and human cancers (Yuan et al. 2020; Ju et al. 2014). For each of the species we obtained mean annual water temperature (Methods) and analysed its associations with all 12 types of substitutions. We observed two significant correlations: positive for $A_H>G_H$ and negative for $T_H>C_H$ (Fig. 1b, Supplementary Mat.). The combined effect of both these substitutions was summarised as the ratio $A_H>G_H/T_H>C_H$ which demonstrates a strong positive correlation with temperature (Fig. 1b). This ratio is expected to be one if all mutations are caused by mtDNA polymerase, which symmetrically introduces mutations on both chains (Supplementary Mat.). An excess of $A_H>G_H$ over $T_H>C_H$ ($A_H>G_H/T_H>C_H > 1$) indicates an extra mutagen asymmetrically inducing A>G predominantly on the heavy chain ($A_H>G_H$). Thus, the increased ratio of $A_H>G_H/T_H>C_H$ in warm-water species (Fig. 1b) shows that this extra mutagen is temperature-sensitive.

This trend is robust to phylogenetic inertia and stays qualitatively similar in two subgroups: early- and -late maturing species (Methods, Supplementary Mat.).In mammalian species $A_H>G_H$ is shown to positively correlate with generation length (Mikhailova et al. 2019). To test this relationship in fishes, we used an analogous metric - the time of maturation (Methods). We performed multiple linear models with several signatures of the mutation spectrum ($A_H>G_H$, $T_H>C_H$ and $A_H>G_H/T_H>C_H$) being a function of both temperature and the time of maturation (Supplementary Mat.). We observed that all of them ($A_H>G_H$, $T_H>C_H$ and $A_H>G_H/T_H>C_H$) depend on temperature only (Supplementary Mat.). Thus, we can conclude that the Actinopterygii mtDNA mutational spectrum is associated with temperature.

The whole-genome nucleotide content may be affected by the mutational spectrum. For example, mutational bias towards 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, genome-wide in neutral sites of the genome. To test this prediction, we analysed synonymous four-fold degenerate nucleotide content in complete mitochondrial genomes of ray-finned fishes. Using pairwise rank

correlation we observed a decrease in A_H and C_H and an increase in G_H and T_H in warm-water species (Fig. 1c, Supplementary Mat.). We combined the sum of the fractions of T_H and G_H (S_{TG}) and the sum of the fractions of A_H and C_H (S_{AC}) into a single metric S_{TG}-S_{AC} (Methods) which involves all components of the temperature-sensitive signature of the mtDNA spectrum (Fig. 1c). Regressing it on temperature, we obtained an expected strong relationship (Fig. 1c, Supplementary Mat.). Keeping in mind the increased A_H>G_H in long- versus short-lived mammals (Mikhaylova et al. 2019) we tested a potential association of S_{TG}-S_{AC} with both temperature and the time of maturation of fishes. The result of the multiple linear regression is the following:

$$S_{TG}\text{-}S_{AC} = 0.422 + 0.046*\text{Temperature} + 0.027*\text{Time of maturation, all p-values} < 9.8e\text{-}05, N=65 \text{ (equation i)}$$

We see that both factors affect the S_{TG}-S_{AC} positively, but temperature has a higher impact on the S_{TG}-S_{AC} as compared to the time of maturation (all coefficients are standardized, Fig. 1d). These results are robust to phylogenetic inertia (Supplementary Mat.) and stay qualitatively similar if instead of S_{TG}-S_{AC} we analyze a fraction of A_H (Supplementary Mat.). Altogether, we conclude that the mtDNA nucleotide content is affected by two factors: a strong temperature-dependent mutagen and a weaker longevity-associated mutagen (Fig. 1c, Fig. 1d).

After describing the Actinopterygii mtDNA mutational spectrum, driven by temperature we decided to test if there is a temperature-associated signature in other vertebrates. Comparing mtDNA mutational spectrum between five classes of vertebrates, we observed that A_H>G_H (Fig. 2a) corresponds to the average body temperature in these classes (similar results with A_H>G_H/T_H>C_H, Supplementary Mat.). To investigate the potential effect of temperature in homeotherms in greater detail 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 uniform, demonstrates some variation (Supplementary Mat.) which might shape the mtDNA spectrum. Previously we demonstrated an increase in A_H>G_H in long-lived mammals which makes their genomes A_H poor and G_H rich (Mikhaylova et al. 2019). Here in order to deconvolute the effects of temperature and lifespan on mtDNA mutational spectrum we analysed mammalian species with known complete mitochondrial genomes, generation lengths and body temperatures (Methods). Running analyses analogues to equation i we observed qualitatively similar results:

$$S_{TG}\text{-}S_{AC} = 0.506 + 0.012*\text{Temperature} + 0.028*\text{Generation Length, all p-values} < 0.000225\text{, except Temperature } (p= 0.053679), N = 224 \text{ (equation ii)}$$

The effect of temperature was marginally significant and more than two times weaker than the effect of the generation length (all coefficients are standardised, Supplementary Mat.). Phylogenetically independent contrasts confirmed significance of generation length only (Supplementary Mat.). Taking into account the low number of species with known temperature we decided to split all mammalian species into two groups - “colder” and “warmer”, using not only temperature but also deep life-history annotation such as hibernation, daily torpor, etc. (Methods). We observed that S_{TG}-S_{AC} depends on both generation length and the temperature group (Fig. 2b, Supplementary Mat.). Results are qualitatively similar if instead of S_{TG}-S_{AC} skew we analyze fraction of A_H (Supplementary Mat.). Thus, mammalian nucleotide content is sensitive to both factors: temperature- and longevity- associated mutagens, where longevity component is stronger.

We have shown that mtDNA mutational spectrum within Chordata species is shaped by two main factors: temperature- and longevity-associated mutagens, temperature being the strongest among ectotherms and longevity among endotherms (Fig. 2c). Interestingly, in the integral model with both Actinopterygii and Mammals (Fig. 2c) we observe that S_{TG}-S_{AC} depends on temperature and longevity, but not on the class (Supplementary Mat.). This confirms that mtDNA mutagenesis of different vertebrate classes follows very similar universal rules, defined stronger by temperature and longevity than by belonging to a specific phylogenetic group.

In this study we assumed that the vast majority of fourfold degenerate synonymous substitutions is effectively neutral, but we cannot rule out some yet unknown selection forces. To our best knowledge there is no evidence

of selection acting on synonymous codons in mtDNA of vertebrates, in turn there is strong mutational effect shaping synonymous codon usage of the human mtDNA (Ju et al. 2014). Until now, the biggest collection of mitochondrial mutations was derived from human cancers, which showed high similarity in spectra (Yuan et al. 2020; Ju et al. 2014). An evolutionary approach allows us to compare mutational spectra between species with contrasting life-history and physiological traits. The existence of between-species variation in mtDNA mutational spectrum has been shown before, but left without explanation (Belle et al. 2005); our large-scale study uncovered relationships, which can shed light on mtDNA mutagens.

Polymerase-introduced mutations are expected to be enriched in $A_H>G_H$ and $T_H>C_H$ and symmetrical (Lee and Johnson 2006). Asymmetric nature of mutations observed in our study ($A_H>G_H/T_H>C_H > 1$) and its relationship with temperature and longevity requires a special explanation. The temperature effect might be explained by the increased level of aerobic metabolism, which in turn leads to increased level of molecular oxygen. High sensitivity of the adenine deamination to the level of molecular oxygen ((Shin and Turker 2002)) may result in higher deamination ($A_H>G_H$) in species with increased temperature due to their higher level of oxidative metabolism. The longevity effect demonstrates relatively lower deamination (decreased $A_H>G_H$ and $A_H>G_H/T_H>C_H$) in short-lived species. Taking into account the generation time and the metabolic rate effects (Martin and Palumbi 1993) we assume that short-lived species per unit of time have an excess of replication-driven mutations, which are expected to be symmetrical in case of mtDNA ($A_H>G_H \sim T_H>C_H$, Supplementary Mat.). This replication-driven input increases the overall symmetry of spectra diminishing the relative effect of asymmetrical chemically-induced mutations (making $A_H>G_H/T_H>C_H$ lower in short-lived species). In other words, relatively increased asymmetrical component of spectra in long-lived species can be driven mainly by the decreased fraction of their symmetrical replication-based component, rather than by increased asymmetrical chemically-induced component (Supplementary Mat.).

A sensitivity of mtDNA mutational spectrum to environmental and life-history traits opens a possibility to use this metric in ecological, evolutionary and genomic studies.

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 (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.
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).

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.

REFERENCES:

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.

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, 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.

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

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.

Mikhaylova, A. G., A. A. Mikhailova, K. Ushakova, E. Tretiakov, A. Yurchenko, M. Zazhytska, D. A. Knorre, et al. 2019. "Mitochondrial Mutational Spectrum Provides an Universal Marker of Cellular and Organismal Longevity." *bioRxiv*. https://doi.org/10.1101/589168.

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, 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.

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.

ACKNOWLEDGMENTS:

We thank the laboratory of Evolutionary Genomics of the Moscow State University for providing the computational cluster "Makarych" resources. 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. 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].

FIGURES

Fig. 1 mtDNA mutational spectrum of Actinopterygii (mtDNA heavy chain notation). (a) average mutational spectrum of all Actinopterygii; (b) mutational spectrum is associated with temperature; (c) whole-genome neutral nucleotide content is associated with ambient temperature; (d) S_{TG}-S_{AC} is sensitive to both temperature and the time of maturation, temperature being the strongest factor.

Fig. 2 mtDNA mutational spectrum of all Vertebrates and mammals (mtDNA heavy chain notation). (a) Between-class comparison shows an excess of A_H>G_H in homeothermic versus poikilothermic animals. (b) S_{TG}-S_{AC} among mammalian species is sensitive to both the generation length and temperature, generation length being the strongest factor. (d) S_{TG}-S_{AC} is sensitive to both temperature and lifespan in ectotherms and endotherms.

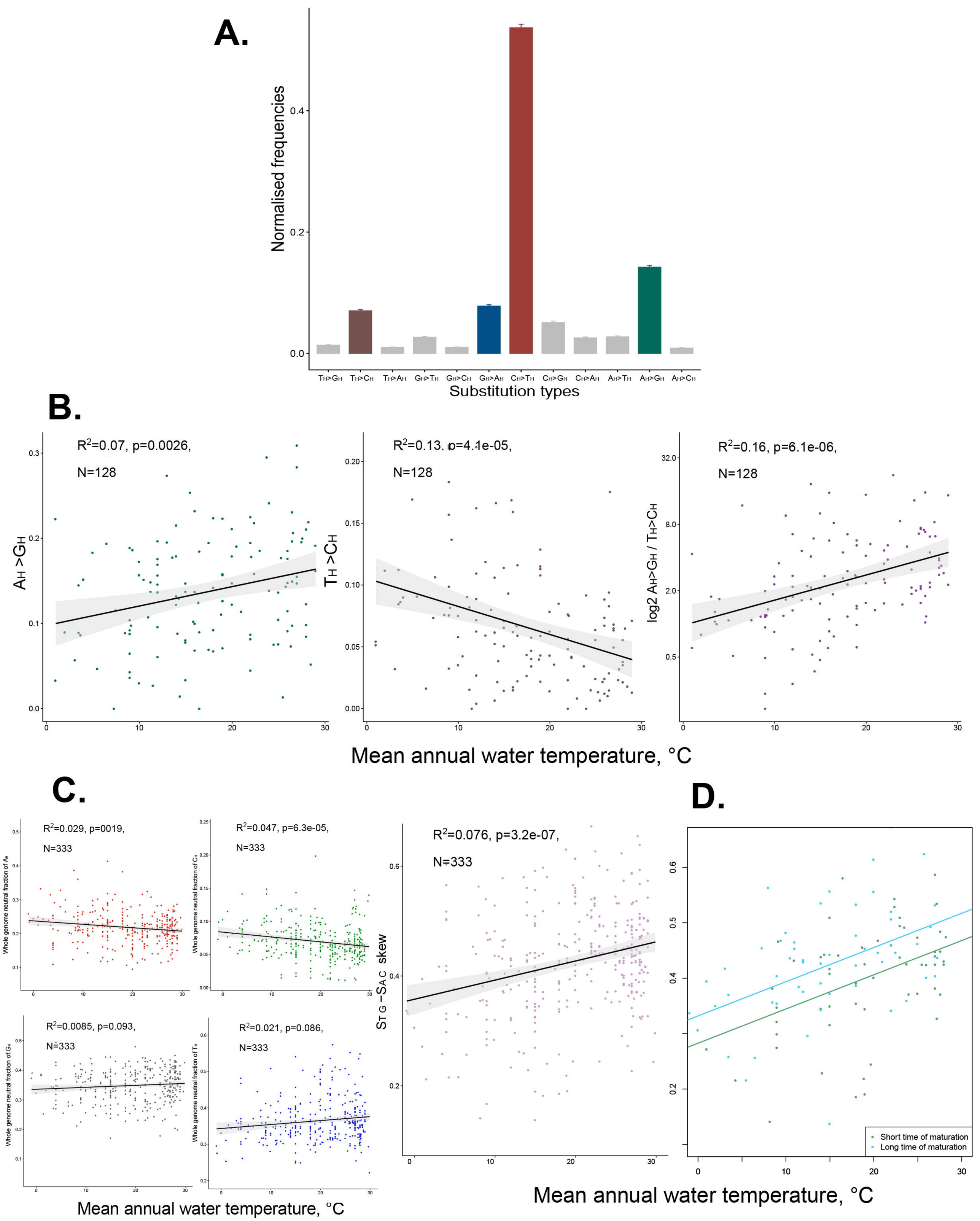
A.
Normalised frequencies
0.0
0.2
0.4
TH>GH TH>CH TH>AH GH>TH GH>CH GH>AH CH>TH CH>GH CH>AH AH>TH AH>GH AH>CH
Substitution types
B.
R2=0.07, p=0.0026,
N=128
AH >GH
R2=0.13, p=4.1e-05,
N=128
TH >CH
R2=0.16, p=6.1e-06,
N=128
log2 AH>GH / TH>CH
Mean annual water temperature, °C
C.
R2=0.029, p=0019,
N=333
Whole genome neutral fraction of AH
R2=0.047, p=6.3e-05,
N=333
Whole genome neutral fraction of CH
R2=0.0085, p=0.093,
N=333
Whole genome neutral fraction of GH
R2=0.021, p=0.086,
N=333
Whole genome neutral fraction of TH
Mean annual water temperature, °C
D.
R2=0.076, p=3.2e-07,
N=333
ST G –SA C skew
Short time of maturation
Long time of maturation
Mean annual water temperature, °C

Figure 1

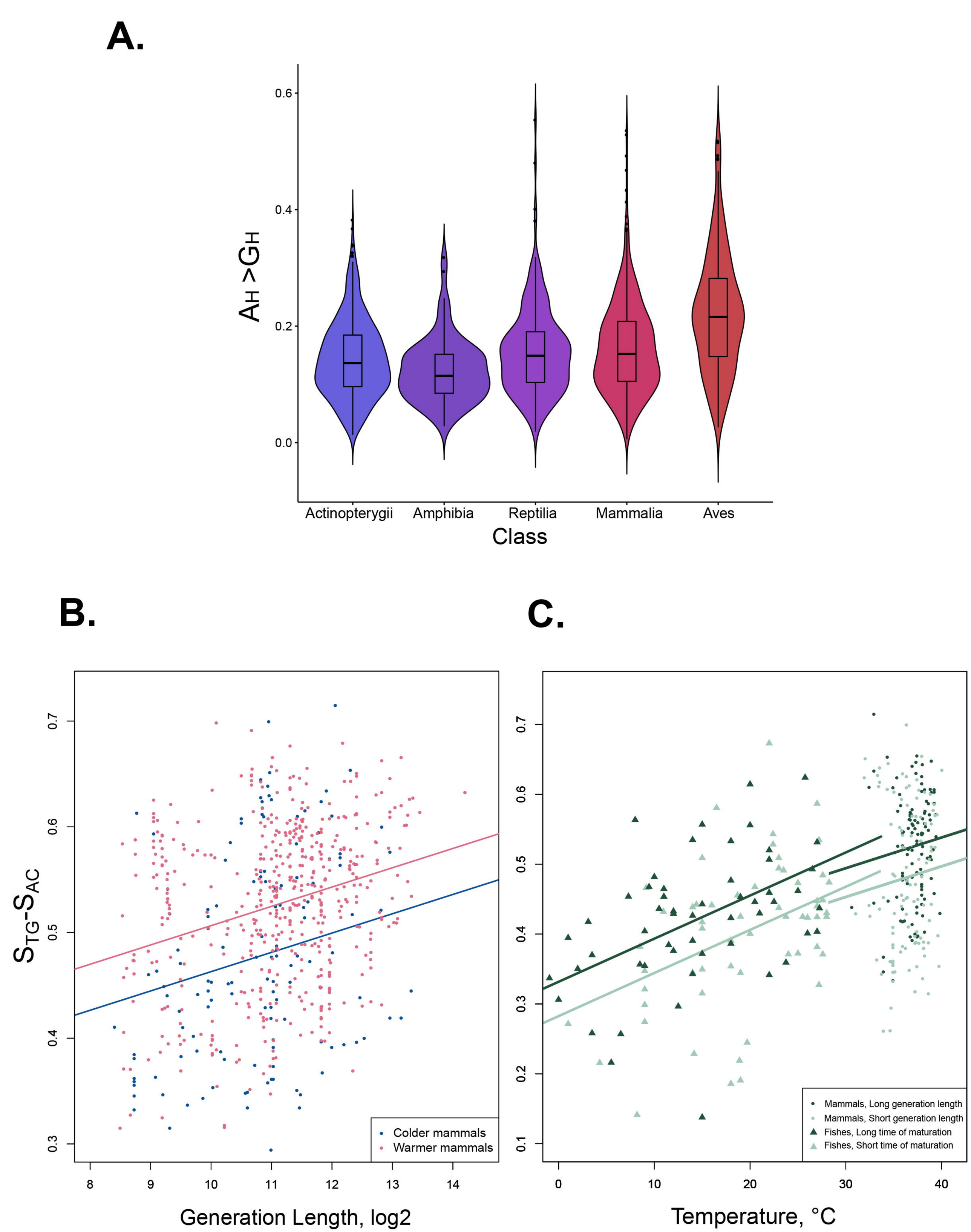
A.
$A_H > G_H$
0.0
0.2
0.4
0.6
Actinopterygii
Amphibia
Reptilia
Mammalia
Aves
Class
B.
S_{TG}-S_{AC}
0.3
0.4
0.5
0.6
0.7
8
9
10
11
12
13
14
Generation Length, log2
Colder mammals
Warmer mammals
C.
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
10
20
30
40
Temperature, °C
Mammals, Long generation length
Mammals, Short generation length
Fishes, Long time of maturation
Fishes, Short time of maturation

Figure 2