bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Triplex and other DNA motifs show motif-specific associations with mitochondrial DNA deletions and species lifespan.

Authors

Kamil Pabis[1]

1. Georg August University of Göttingen, Göttingen, Germany.
Mail: Kamil.pabis@gmail.com

ABSTRACT

The "theory of resistant biomolecules" posits that long-lived species show resistance to molecular damage at the level of their biomolecules. Here, we test this hypothesis in the context of mitochondrial DNA (mtDNA) as it implies that predicted mutagenic DNA motifs should be inversely correlated with species maximum lifespan (MLS).

First, we confirmed that guanine-quadruplex and direct repeat (DR) motifs are mutagenic, as they associate with mtDNA deletions in the human major arc of mtDNA, while also adding mirror repeat (MR) and intramolecular triplex motifs to a growing list of potentially mutagenic features. What is more, triplex motifs showed disease-specific associations with deletions and an apparent interaction with guanine-quadruplex motifs.

Surprisingly, even though DR, MR and guanine-quadruplex motifs were associated with mtDNA deletions, their correlation with MLS was explained by the biased base composition of mtDNA. Only triplex motifs negatively correlated with MLS even after adjusting for body mass, phylogeny, mtDNA base composition and effective number of codons.

Taken together, our work highlights the importance of base composition for the comparative biogerontology of mtDNA and suggests that future research on mitochondrial triplex motifs is warranted.

ABBREVIATIONS

BPs, mtDNA deletion break points

DR, direct repeats

ER, everted repeats

GQ, guanine-quadruplexes

IR, inverted repeats

MLS, species maximum lifespan

MR, mirror repeats

nBMST, non-B DNA motif search tool

Nc, number of effective codons

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

PGLS, phylogenetic generalized least squares

SD, standard deviation

Trip, Triplex forming motif

XR, any repeat half-site or motif

mtDNA, mitochondrial DNA

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

INTRODUCTION

Macromolecular damage to lipids, proteins and DNA accumulates with aging (**Richardson and Schadt 2014, Gladyshev 2013**), whereas cells isolated from long-lived species are resistant to genotoxic and cytotoxic drugs, giving rise to the multistress resistance theory of aging (**Miller 2009, Hamilton and Miller 2016**). By extension of this idea, the "theory of resistant biomolecules" posits that lipids, proteins and DNA itself should be resilient in long-lived species (**Pamplona and Barja 2007**). In support of this theory, it was shown that long-lived species possess membranes that contain fewer lipids with reactive double bonds (**Valencak and Ruf 2007**) and perhaps a lower content of oxidation-prone cysteine and methionine in mitochondrially encoded proteins (see **Aledo et al. 2012** for a discussion).

Mitochondrial DNA (mtDNA) mutations constitute one type of macromolecular damage that accumulates over time. Point mutations accumulate in proliferative tissues like the colon and in some progeroid mice (**Kauppila et al. 2017**), while the accumulation of mtDNA deletions in postmitotic tissues may underpin certain age-related diseases like Parkinson's and sarcopenia (**Lawless et al. 2020, Bender et al. 2006**).

If the theory of resistant biomolecules can be generalized, the mtDNA of long-lived species should resist both point mutation and deletion formation. However, we will focus on deletions because they are more pathogenic than point mutations at the same level of heteroplasmy (**Gamamge et al. 2014**) and human tissues do not accumulate high levels of point mutations observed in progeroid mouse models (**Khrapko et al. 2006**).

Since deletion formation depends on the primary sequence of the mtDNA (sequence motifs) it is amenable to bioinformatic methods. Ever since a link between direct repeat (DR) motifs and deletion formation became known, variations of the theory of resistant biomolecules have been tested, although not necessarily under this name. It was reasoned that long-lived species evolved to resist deletion formation and mtDNA instability by reducing the number of mutagenic motifs in their mtDNA (**Khaidakov et al. 2006, Yang et al. 2013**).

We aim to extend these findings by re-evaluating and establishing new candidate motifs, which we then correlate with species maximum lifespan (MLS). Studying multiple motif classes at once also allows us to reveal relationships between potentially overlapping mtDNA motifs that may affect the data. We define candidate motifs as those that are associated with deletion formation inside the major arc of human mtDNA, because during asynchronous replication the major arc is single stranded for extended periods of time (**Persson et al. 2019**) which should favor the formation of secondary structures. Finally, we test if these motifs correlate with the MLS of mammals, birds and ray-finned fishes after correcting for potential biases, especially global mtDNA base composition which is an important confounder (**Aledo et al. 2012**) yet is neglected in some studies (**Yang et al. 2013**).

The choice of motifs to study is based on biological plausibility and published literature that will be briefly reviewed below. Mutagenic motifs include repeats as well as guanine-quadruplex (GQ)- and triplex-forming motifs. DR motifs can lead to DNA instability through strand-slippage if two DR motifs mispair during replication (**Persson et al. 2019**). Whereas inverted repeat (IR), G-quadruplex and triplex motifs destabilize progression of the replication fork through the formation of stable secondary structures. Some of the structures formed include hairpins for IR motifs (**Tremblay-Belzile et al. 2015**), triple stranded DNA for triplex motifs and bulky stacks of guanines for G-quadruplex motifs (**Bacolla et**

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

al. 2016; Fig. 1). Mirror repeat (MR) and everted repeat (ER) motifs, in contrast, do not allow stable Watson-Crick base pairing and are thus less likely to be mutagenic, although a subset of MR motifs may form triplex structures (**Kamat et al. 2016)**.

Thus, many motifs can be mutagenic in principle, but what is the evidence that these motifs are related to mtDNA instability, particularly deletions, and MLS?

Paradoxically, while DRs are the motif most consistently associated with mtDNA deletion breakpoints (BPs), despite preliminary reports (**Khaidakov et al. 2006, Lakshmanan et al. 2012**, **Yang et al. 2013**), no correlation with species MLS was seen in recent studies (**Lakshmanan et al. 2015**). In contrast, with the exception of one preprint (**Mikhailova et al. 2020**), IRs are not known to be associated with mtDNA deletions (**Dong et al. 2014**), although they do show a negative relationship with species MLS (**Yang et al. 2013**) and may contribute to inversions (**Tremblay-Belzile et al. 2015**). Whether age-related mtDNA inversions underlie any pathology, however, requires further study. Finally, G-quadruplex motifs are associated with both deletions (**Dong et al. 2014**) and point mutations (**Butler et al. 2020**), but no study tested if they correlate with MLS. Triplex motifs are poorly studied with one report finding no association between these motifs and deletions (**Oliveira et al. 2013**).

Based on these studies we decided to test the theory of resistant biomolecules by quantifying DR, MR, IR, ER, G-quadruplex- and triplex-forming motifs. We stipulate that if a motif class played a causal role in aging, it should be involved in deletion formation and its abundance should be negatively correlated with species MLS.

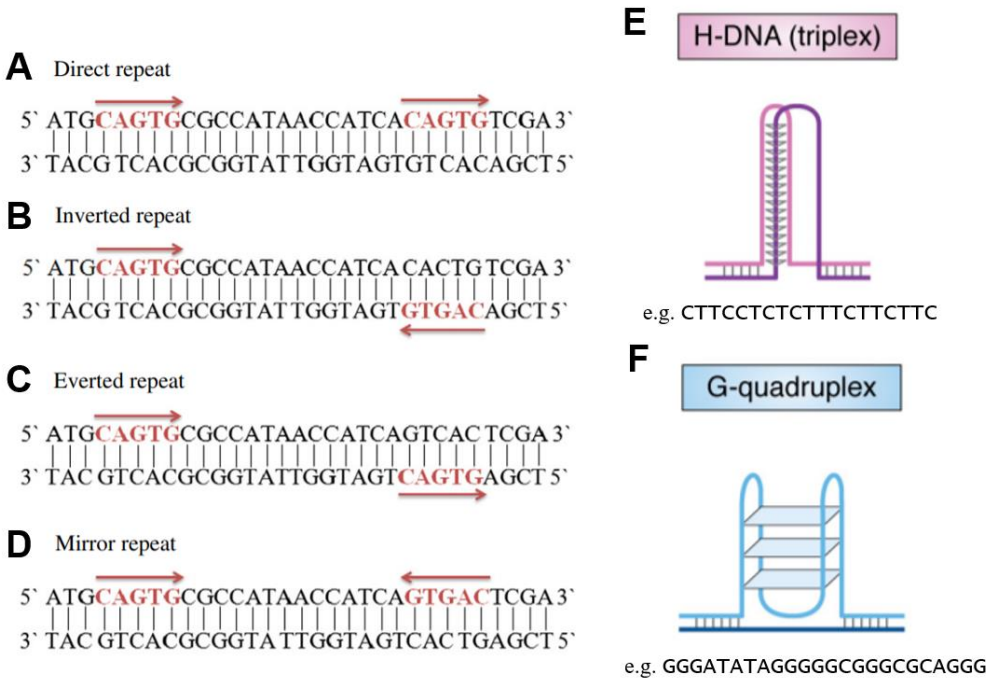


Figure 1

A. Direct repeat, both half-sites have the same orientation.

B. Inverted repeat, the half-sites are complementary and has mirror symmetry.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

C. Everted repeat, the half-sites are complementary.
D. Mirror repeat, the half-sites have mirror symmetry.

E. Triplex motifs can form a triple helical DNA structure also called H-DNA.
F. In a G-quadruplex multiple G-quartets (depicted as blue rectangles) stack on top of each other.

Adapted from **Gurusaran et al. (2013)** and **Khristich and Mirkin (2020**) with permission. Half-sites shown in red.

METHODS

Detection of DNA motifs

Repeats were detected by a script written in R (vR-3.6.3). Briefly, to find all repeats with N basepairs (bps), the mtDNA light strand is truncated by 0 to N bps and each of the N truncated mtDNAs is then split every N bps. This generates every possible substring (and thus repeat) of length N. In the next step, duplicate strings are removed. Afterwards we can find DR (a substring with at least two matches in the mtDNA), MR (at least one match in the mtDNA and on its reverse), IR (at least one match in the mtDNA and on its reverse-complement) and ER motifs (at least one match in the mtDNA and on its complement). Overlapping and duplicate repeats were not counted for the correlation between repeats and MLS. The code for the analyses performed in this paper can be found on github (pabisk/aging_triplex2).

Unless stated otherwise, all analyses were performed in R. G-quadruplex motifs were detected by the pqsfinder package (v2.2.0, **Hon et al. 2017**). Intramolecular triplex-forming motifs were detected by the triplex package (v1.26.0, **Hon et al. 2013**) and duplicates were removed. We also compared the data with two other publicly available tools, Triplexator (**Buske et al. 2013**), and with the non-B DNA motif search tool (nBMST; **Cer et al. 2011**). Triplexator was run on a virtual machine in an Oracle VM VirtualBox (v6.1) in -ss mode on the human mitochondrial genome and its reverse complement, the results were combined and overlapping motifs from the output were removed. We used the web interface of nBMST to detect mirror repeats/triplexes (v1.0).

Association between motifs and major arc deletions

The major arc was defined as the region between position 5747 and 16500 of the human mtDNA (NC_012920.1). The following deletions and their breakpoints were located in this region and included: 1066 deletions from the MitoBreak database (**Damas et al. 2014, mtDNA Breakpoints.xlsx**), 1114 from **Persson et al. (2019)** and 1894 from **Hjelm et al. (2019)**.

Each deletion is defined by two breakpoints. A breakpoint pair was considered to associate with a motif if the motif fell within a defined window around one or both breakpoints, depending on the analysis. The window size was chosen in relation to the length of the studied motifs (30 bp for repeats and 50 bp for other motifs).

Three different motif orientations relative to the breakpoints were considered. Two orientations for motifs with half-sites (i.e. repeats), either both half-sites at any one breakpoint of a deletion, or one half-site per breakpoint of a deletion. Motifs with overlapping half-sites were not counted. In the third case, distinct G-quadruplex and triplex motifs could associate with one or both breakpoints of a deletion, but were at most counted once, since the latter case is sufficiently rare.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In order to exclude overlapping "hybrid" motifs, MR and DR motifs with the same sequence were removed whereas triplex and G-quadruplex motifs were removed if they were in proximity.

To generate controls, the mtDNA deletions as a whole were randomly redistributed inside the major arc which, because of the fixed deletion size, allowed us to approximate the original distribution of breakpoints (as suggested by **Oliveira et al. 2013**). Significance was determined via one-sample t-test in Prism (v7.04) by comparing actual breakpoints to 20 such randomized controls. Alternative controls were generated by shifting each breakpoint by 200 bp towards the midpoint of the major arc or as in **Fig. S9**.

Cancer associated breakpoints

We obtained all autosomal breakpoints available from the Catalogue Of Somatic Mutations In Cancer (COSMIC; release v92, 27th August 2020), which includes deletions, inversions, duplications and other abnormalities (n=587515 in total). After removing breakpoints whose sequences could not be retrieved (<1.7%), we quantified the number of predicted G-quadruplex and triplex motifs in a 500 bp window centered on the breakpoints using default settings for the detection of these motifs. Sequences of breakpoint regions were obtained from the GRCh38 build of the human genome using the BSgenome package (v1.3.1). Each breakpoint shifted by +3000 bps served as its own control.

Lifespan, base composition and life history traits

We included three phylogenetic classes in our analysis for which we had sufficient data (n>100), mammals, birds and ray-finned fishes (actinopterygii). MLS and body mass were determined from the AnAge database (**Tacutu et al. 2018**) and, for mammals, supplemented with data from **Pacifici et al. (2013)**. The mtDNA accessions were obtained from an updated version of MitoAge (unpublished; **Toren et al. 2016**). Species were excluded if body mass data was unavailable, if the sequence could not be obtained using the genbankr package (v1.14.0), or if the extracted cytochrome B DNA sequence did not allow for an alignment, precluding phylogenetic correction. The species data can be found in the supplementary (**Species Data.xlsx**).

We analyzed the full mtDNA sequence, heuristically defined as the mtDNA sequence between the first and last encoded tRNA, excluding the D-loop, which is rarely involved in repeat-mediated deletion formation (**Yang et al. 2013**). The effective number of codons was calculated using Wright's Nc (**Smith et al. 2019**). Base composition was calculated for the light-strand. GC skew was calculated as the fraction (G − C)/(G + C) and AT skew as (A − T)/(A + T). All correlations are Pearson's R. Partial correlations were performed using the ppcor package (v1.1).

Phylogenetic generalised least squares and phylogenetic correction

Observed correlations between traits and lifespan can be spurious due to shared species ancestry (**Speakman 2005**). To correct for this, we use phylogenetic generalised least squares (PGLS) implemented in the caper package (v1.0.1). Species phylogenetic trees were constructed via neighbor joining based on aligned cytochrome B DNA sequences using Clustal Omega from the msa package (v1.18.0) and in the resulting mammalian and bird tree, four branch edge lengths were equal to zero, which were set to the lowest non-zero value in the dataset.

RESULTS

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Direct repeats and mirror repeats are over-represented at mtDNA deletion breakpoints

In order to define candidate mtDNA motifs that could be linked with lifespan, we started by reanalyzing motifs that associate with mtDNA deletion breakpoints reported in the MitoBreak database (**Damas et al. 2014; Fig. S1; mtDNA Breakpoints.xlsx**). In the below analysis, we consider DR and IR motifs thought to be mutagenic, as well as MR and ER motifs, so far not known to be mutagenic and we pool all 6 to 15 bp long repeats, since the data is similar between different repeat lengths (**Fig. S2**).

As shown by others, we found that DR motifs often flank mtDNA deletions (**Fig. 2A**). In contrast, no strong association was seen for ER and IR motifs, even considering a larger window around the breakpoint to allow for the fact that IRs could bridge and destabilize mtDNA over long distances (**Persson et al. 2019; Fig. S3**).

Surprisingly, we also found MR motifs flanking deletion breakpoints more often than expected by chance (**Fig. 2A**). However, DR and MR motifs are known to correlate with each other (**Shamanskiy et al. 2019; Fig. 5B**) and indeed we noticed a large sequence overlap between MR and DR motifs (**Fig. 2B**), which could explain an apparent over-representation of MRs at breakpoints. Removal of overlapping MR-DR hybrid motifs confirmed this suspicion. After this correction, the degree of enrichment was strongly attenuated (**Fig. 2C**) and the total number of breakpoints flanked by MR motifs was reduced by >80%. Nevertheless, long MR motifs remained particularly over-represented around deletions (**Fig. S4**).

Since the prior analysis only considered motifs that flank both breakpoints, we next tested the idea that IR and other motifs could be mutagenic if both half-sites are found at any of the breakpoints. However, in this analysis no motif class showed enrichment around breakpoints (**Fig. 2D**).

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

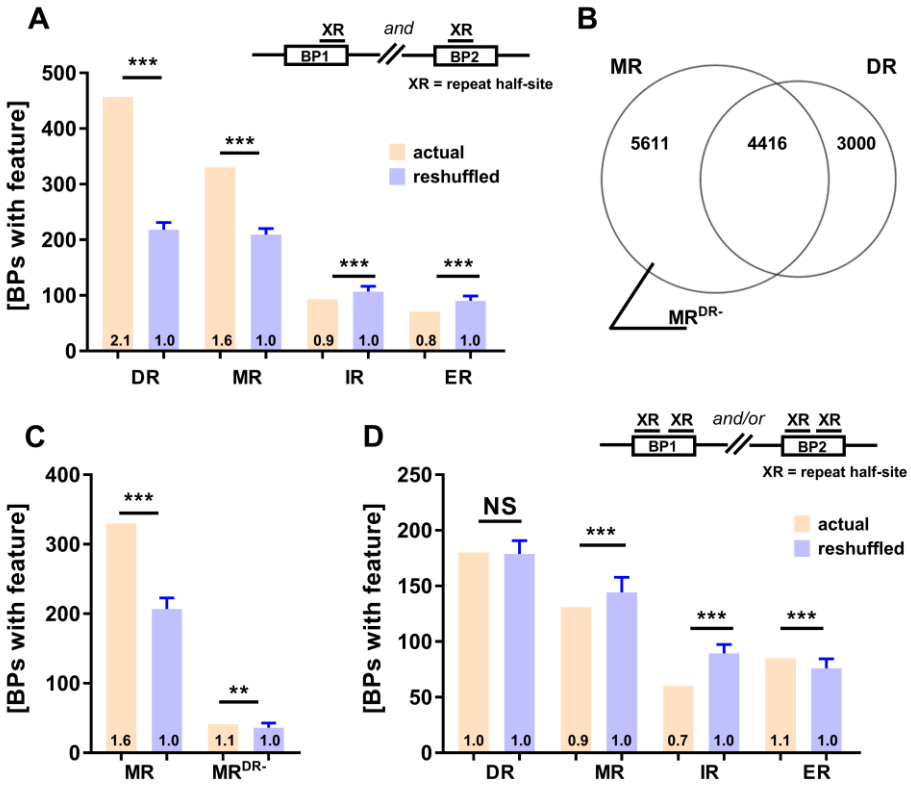


Figure 2

Direct repeat (DR) and mirror repeat (MR) motifs are significantly enriched around actual deletion breakpoints (BPs) compared to reshuffled BPs, but the same is not true for inverted repeat (IR) and everted repeat (ER) motifs (A, D). The surprising correlation between MR motifs and deletion BPs is attenuated when MRs that have the same sequence as DR motifs are removed (B, C). Controls were generated by reshuffling the deletion BPs while maintaining their distribution (n=20, mean ±SD shown). The schematic drawings above (A, D) depict the orientation of the repeat (XR) half-sites in relation to the BPs. *** $p < 0.001$; ** $p < 0.01$ by one sample t-test.

A) The number of deletions associated with DR, MR, IR or ER motifs at both BPs compared with reshuffled controls.

B) Venn diagram showing the number of MR, DR and hybrid MR-DR motifs that were identified within the major arc.

C) The number of deletions associated with MR motifs, before (MR) and after removal of hybrid MR-DR motifs (MR^{DR-}), compared with reshuffled controls.

D) The number of deletions associated with DR, MR, IR or ER motifs at either BP compared with reshuffled controls.

Predicted triplex-forming motifs are over-represented at mtDNA breakpoints

Given the association between MR motifs and breakpoints we decided to analyze triplex motifs, a special case of homopurine and homopyrimidine mirror repeats (**Khristich and Mirkin 2020, Bissler 2007**), and their association with deletion breakpoints in the MitoBreak database.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Here, we use the triplex package to predict intramolecular triplex motifs because it has several advantages compared to other software (**Hon et al. 2013**). For example, using the nBMST tool, as in a previous study of mtDNA instability (**Oliveira et al. 2013**), we only identified two potential triplex motifs within the major arc that did not overlap with the six motifs identified by the triplex package (**Table S1**). In contrast, using Triplexator (**Buske al. 2013**) we were able to detect four of the six triplex motifs and the motifs detected by Triplexator were also enriched at breakpoints (**Table S2**).

We noticed that predicted triplexes are G-rich and thus could be related to G-quadruplex motifs (**Doluca et al. 2013**). In a comparison of the two motif types, however, we found several differences (**Table S1, S3**). Triplex motifs were shorter and less abundant than predicted G-quadruplexes, associated with fewer breakpoints altogether (**Fig. 3**) and, in contrast to G-quadruplexes almost exclusive to the G-rich mtDNA heavy-strand, triplex motifs were also common on the light-strand.

The six triplex motifs detected by the triplex package were significantly enriched around deletion breakpoints and when we excluded triplex-G-quadruplex hybrid motifs the result was attenuated but remained significant (**Fig. 3A**). Given the higher risk of spurious findings with only six motifs, we repeated the analysis using a relaxed definition of triplex and the results were fundamentally unchanged (**Fig. 3B**). Furthermore, our results were not sensitive to reasonable changes in the size of the search window around breakpoints (**Fig. S5A, B**), motif quality scores (**Fig. S5C, D**) or inclusion of overlapping motifs (**Fig. S5E-G**).

Analogous to the situation with MR motifs we tested if overlapping triplex-DR hybrid motifs could bias our results. Given the rarity of triplex motifs and the many DRs in the mitochondrial genome we choose an alternative approach rather than excluding triplex motifs that overlapped any DR half-site. We compared the fraction of triplex and G-quadruplex positive deletions associated with DRs (GQ^+, DR^+ and $Trip^+$, DR^+) and not associated with DRs (GQ^+, DR^- and $Trip^+$, DR^-). We considered a deletion to be DR^+ if both breakpoints were flanked by the same DR sequence. In this case, only 44% of $Trip^+$ deletions associated with DRs whereas 66% of GQ^+ deletions did (**Table S4**).

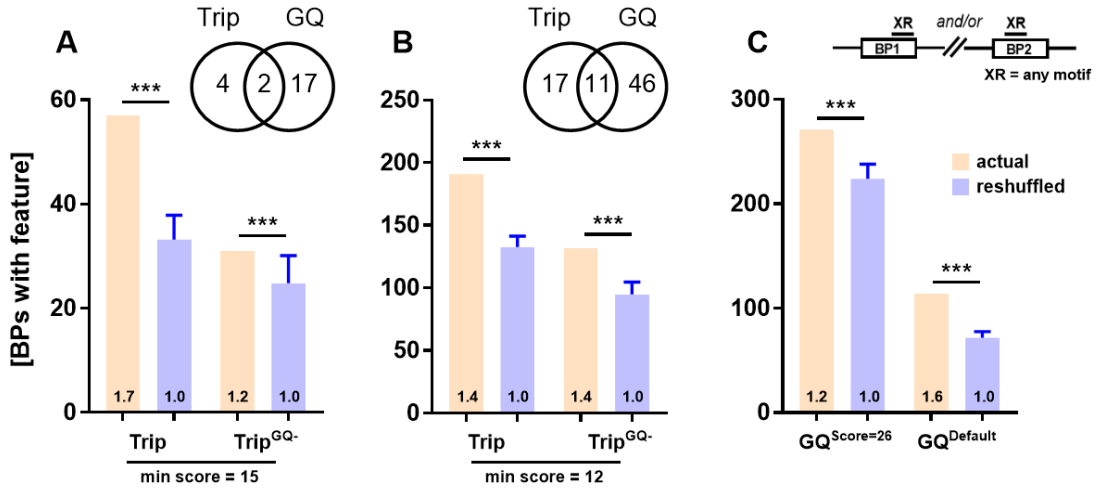


Figure 3

Triplex motifs are significantly enriched around actual breakpoints (BPs) compared to reshuffled BPs (A, B) even after removal of G-quadruplex (GQ)-triplex hybrid motifs ($Trip^{GQ-}$). The number of unique triplex

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

motifs, GQ motifs and of hybrid triplex-GQ motifs, within the mtDNA major arc, is shown in the Venn diagrams above (A, B). Enrichment of GQ motifs around BPs is shown for comparison in (C). Controls were generated by reshuffling the deletion BPs while maintaining their distribution (n=20, mean ±SD shown). The schematic drawing above (C) depicts the orientation of the GQ and triplex motifs (XR) in relation to the BPs. *** $p < 0.0001$ by one sample t-test.

A) The number of deletion BPs associated with triplex motifs compared with reshuffled controls. Analysis including (left side) or excluding triplex-GQ hybrid motifs (right side).
B) Same as (A) but with relaxed criteria for the detection of triplex motifs (min score=12) and GQ motifs (min score=26).
C) The number of deletion BPs associated with GQ motifs compared with reshuffled controls. Relaxed settings (left side, min score=26) and default settings (right side, min score=47).

Triplex forming motifs may be associated with mitochondrial disease breakpoints

Next, we sought to validate our findings on two recently published next generation sequencing datasets (**Hjelm et al. 2019, Persson et al. 2019**; **mtDNA Breakpoints.xlsx; Table S5**). We were able to confirm the enrichment of DR (**Fig. S6A, S7A**), MR (**Fig. S6A, S7A**) and G-quadruplex motifs (**Fig. 4A, B; S6C, D**) around deletion breakpoints. Additionally, we confirmed that hybrid MR-DR motifs are responsible in large part for the enrichment of MR motifs around breakpoints (**Fig. S6B, S7B**).

In contrast, we found that triplex motifs were not consistently enriched around breakpoints in the dataset of Hjelm et al. (**Fig. S6C, D**), which is based on post-mortem brain samples from patients without overt mitochondrial disease, whereas we saw enrichment in the dataset by Persson et al. (**Fig. 4A, B**), which is based on muscle biopsies from patients with mitochondrial disease. This unexpected discrepancy prompted us to take a second look at the MitoBreak data. In this dataset triplex motifs were significantly more enriched at breakpoints in the mtDNA single deletion subgroup compared to the healthy tissues subgroup (**Fig. S8**). In addition, we found more broadly that mitochondrial disease status might explain the heterogenous results across datasets we have seen (**Fig. 4C**).

Further strengthening our findings, triplex motifs were enriched in the MitoBreak and Persson et al. dataset regardless of the breakpoint shuffling method chosen and of our statistical assumptions (**Fig. S9**). What is more, triplex motifs were also enriched at breakpoints when we pooled all three datasets (**Fig. 4D**), although to a lesser extent.

Finally, G-quadruplex motifs close to triplex motifs were more strongly enriched at deletion breakpoints than solitary G-quadruplex motifs (**Fig. 4E; Fig. S10**), suggesting that triplex formation may further contribute to DNA instability.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

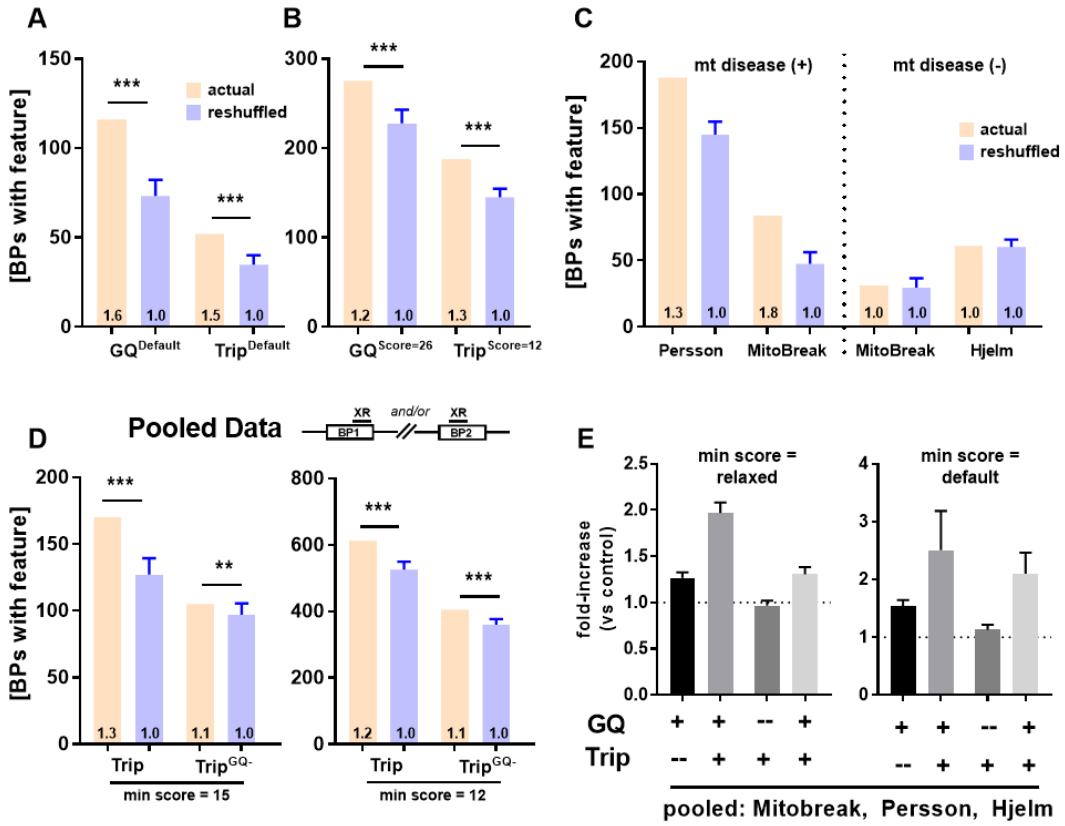


Figure 4

In the **Persson et al. (2019)** dataset, triplex and G-quadruplex (GQ) motifs are enriched around deletion breakpoints (BPs), using either default (A) or relaxed scoring criteria (B). Although triplex motifs predominate in mitochondrial disease datasets (C), we also find that triplex motifs are significantly enriched around BPs (D) after pooling the data from **MitoBreak, Persson et al. (2019) and Hjelm et al (2019).** Finally, GQ and triplex motifs show stronger enrichment around BPs than either of them in isolation (E). Controls were generated by reshuffling the deletion BPs while maintaining their distribution (n=20, mean ±SD shown). The schematic drawing above (D) depicts the orientation of the motifs (XR) in relation to the BPs. *** p<0.0001, **p<0.001 by one sample t-test.

A) The number of deletion BPs associated with GQ and triplex motifs compared with reshuffled controls (min score = default).

B) The number of deletion BPs associated with GQ and triplex motifs compared with reshuffled controls (min score = relaxed).

C) The number of deletion BPs associated with triplex motifs (relaxed settings, min score=12) stratified by mitochondrial disease status. MitoBreak data includes single and multiple mitochondrial deletion syndromes.

D) The number of deletion BPs associated with triplex motifs, or with triplex motifs excluding triplex-GQ hybrid motifs ($Trip^{GQ-}$), compared with reshuffled controls. Default settings (left side, min score=15) and relaxed settings (right side, min score=12).

E) The fold-enrichment of GQ and triplex motifs around deletion BPs is shown. Motifs were considered overlapping if their midpoints were within 50 bp.

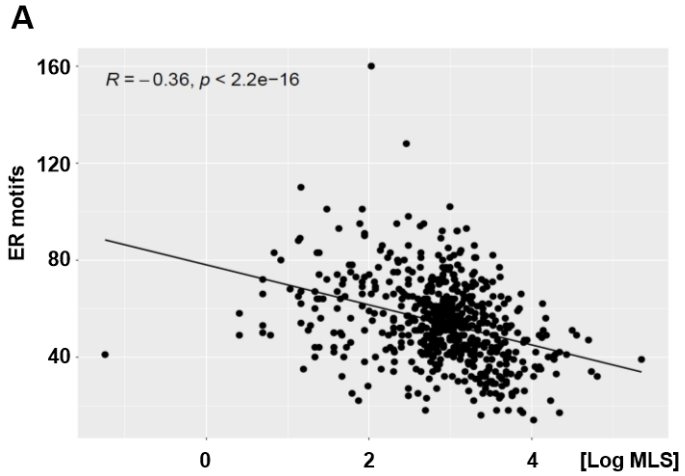
bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

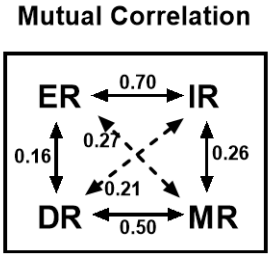
Repeats and lifespan: no support for the theory of resistant biomolecules

For our analysis, we focus on 11 bp long repeat motifs as short repeats are less likely to allow stable base pairing and longer repeats are rare (**Fig. S11**) and because results considering repeat motifs of different lengths usually agree with each other (**Table S6**; **Yang et al. 2013**). To allow comparability with other studies (**Lakshmanan et al. 2015**) we analyzed non D-loop motifs, but results for major arc motifs are numerically similar (**Table S7**).

First, consistent with **Yang et al. (2013)** we found that IR motifs show a negative correlation with the MLS of mammals in the unadjusted model. In addition, we identified ER motifs, a class of symmetrically related repeats, that show an even stronger inverse relationship with longevity (**Fig. 5A; Table 1**). However, these inverse correlations vanished after taking into account body mass, base composition and phylogeny in a PGLS model (**Table 1**). Second, in agreement with **Lakshmanan et al. (2015)** we found that DR motifs do not correlate with the MLS of mammals. The same was true for the symmetrically related MR motifs. Just as with IR motifs, modest inverse correlations vanished in the fully adjusted model (**Table 1**). We also found the same null results in two other vertebrate classes, birds and ray-finned fishes (**Table S6**). To gain hints as to causality, we finally tested if longer repeats, allowing more stable base pairing, show stronger correlations with MLS, but to our surprise we noticed the opposite (**Fig. S12A-D**).

Considering all four types of repeats together, we noticed that repeats with both half-sites on the same strand (DR and MR) or half-sites opposite strands (IR and ER) were correlated with each other (**Fig. 5B**) and with the same mtDNA compositional biases (**Fig. 5C**). Thus, for DR and MR motifs, an apparent relationship with MLS may be explained by their inverse relationship with GC content and for IR and ER motifs by an inverse relationship with GC content and a positive relationship with GC skew.





C

	GC	GC-skew	AT-skew	Nc
DR11	-0.342	-0.249	0.102	-0.555
MR11	-0.416	-0.138	0.044	-0.480
ER11	-0.791	0.604	-0.609	-0.066
IR11	-0.743	0.611	-0.628	-0.014

Figure 5

The number of everted repeat (ER) motifs is negatively correlated with species MLS in an unadjusted analysis (A). Repeats with a similar orientation correlate with each other (B). Direct repeat (DR) and mirror repeat (MR) motifs have a similar orientation since both half-sites are found on the same strand and in the case of ER and inverted repeat (IR) motifs the half-sites are on opposite strands. Finally, we show the major mtDNA compositional biases that co-vary with the four repeat classes (C) and may

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

explain an apparent correlation with MLS. Data is for 11 bp long repeats and Pearson's R is shown in (A-C).

Table 1. Correlation between potentially mutagenic motifs and species lifespan

Motif	Type	Raw	Adjusted
DR11	11bp	**-0.113**	0.055
MR11	11bp	**-0.155**	-0.002
IR11	11bp	**-0.336**	**0.105**
ER11	11bp	**-0.356**	-0.047
triplex	default	**-0.296**	**-0.211****
triplex	relaxed	**-0.190**	**-0.127^**
GQ	default	**0.264**	0.068
GQ	relaxed	**0.283**	-0.097**

The adjusted model takes into account body mass, GC content, GC skew, AT skew and number of effective codons. Significant correlations in the raw or adjusted model are bolded/underlined ($p<0.05$). The PGLS model additionally considers phylogeny. ^denotes p-values of $0.05<p<0.10$ in the PGLS model and ** p-values of $p<0.05$. The table shows Pearson's R.

Triplex motifs, not G-quadruplexes, show an inverse relationship with species lifespan

So far, no survey of G-quadruplex and triplex motifs has been conducted across species, although it is known that human mtDNA contains more G-quadruplexes than mouse, rat or monkey mtDNA (**Bharti et al. 2014**). First, we confirmed that different tools to detect G-quadruplex motifs agree with each other (**Lombardi and Londoño-Vallejo 2020; Fig. S13**). The results of different triplex detection tools, however, were inconsistent. While we were able to detect some overlap between the motifs found with Triplexator and the triplex package in human mtDNA (**Table S1**), we found that the two tools made very different predictions regarding triplex counts across species (**Fig. S14**).

The limited agreement between publicly available triplex detection tools raised the question whether our preferred tool, the triplex package in R, detects the same class of mutagenic triplex motifs as reported in earlier studies using in-house scripts (**Bacolla et al. 2016**). To answer this question, we reanalyzed a large dataset of over 500000 cancer associated breakpoints. In line with this earlier study, we found that both triplex and G-quadruplex motifs were significantly enriched around actual breakpoints compared to control breakpoints (**Fig. 6; Table S8**) and that breakpoints were preferably found in highly unstable regions with multiple such motifs (**Fig. S15**).

Next, we turned again to mtDNA to test whether mutagenic motifs are negatively correlated with MLS as predicted by the theory of resistant biomolecules. To the contrary, we found that G-quadruplex motifs were positively correlated with MLS (**Fig. S16**), although this may be secondary to their strong correlation with GC content. However, even after taking into account base composition and phylogeny using PGLS there was no convincing relationship between G-quadruplex motifs and MLS (**Table 1**).

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In contrast, we found a moderate, negative correlation between intramolecular triplex motifs and the MLS of mammals (**Fig. 7A**) that was significant after correcting for body mass and base composition and, for triplex motifs using default scoring, also after correction for phylogeny using PGLS (**Table 1**). This result remained stable when we varied the score-cutoff and in fact triplexes were the only motif class for which we found that higher confidence motifs, i.e. motifs predicted to be more stable, showed a stronger relationship with MLS (**Fig. S12E**).

Although we found no significant relationship between triplex motifs and MLS of birds or ray-finned fishes, we noticed that this finding was modulated by the number of triplex motifs in the mtDNA of birds (**Fig. 7B**) and mammals (**Fig. S17B, D**). Phylogenetic orders with higher numbers of triplex motifs in their mtDNA showed a stronger inverse relationship between MLS and triplex counts than orders with few such motifs. Thus, when we split the bird and mammal data into orders with mitochondrial triplex levels above or below the median, we found a significant inverse relationship between MLS and triplex motifs in the high group for both birds and mammals (**Table S9**).

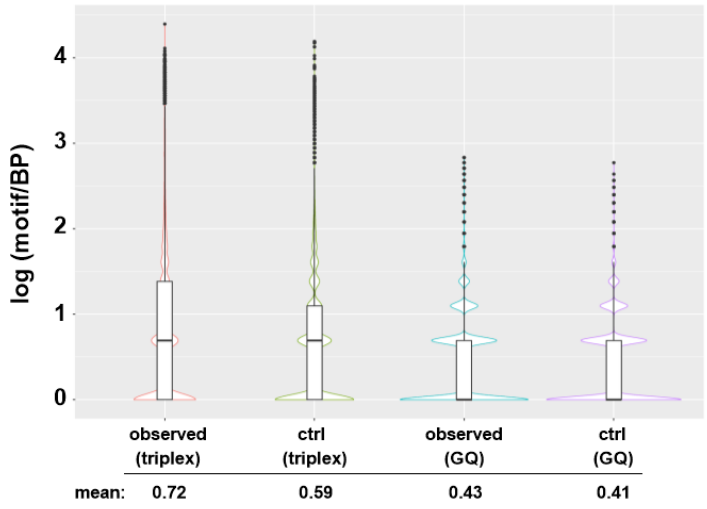


Figure 6

Actual cancer-related breakpoints (BPs) are associated with more triplex and G-quadruplex (GQ) motifs compared to control BPs ($p < 2.2e\text{-}16$; Wilcoxon rank sum test). To allow better visualization of the data, the number of motifs for each BP was log-transformed and log(0) values were excluded. Box whisker plots show the median, interquartile range and outliers while the underlying violin plot shows the actual distribution.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

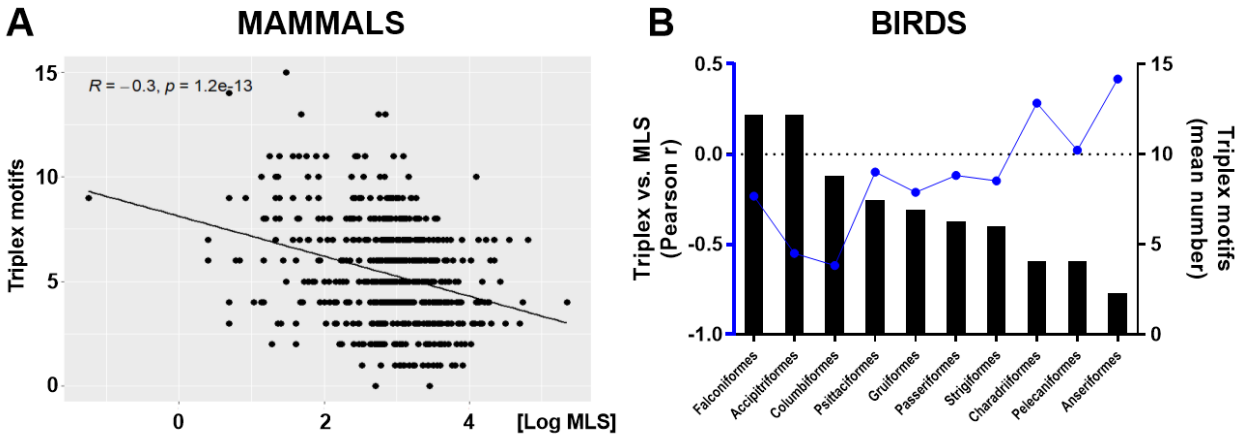


Figure 7

Across mammals, mitochondrial triplex motifs are inversely correlated with maximum lifespan (MLS) in an unadjusted analysis (A). Although, birds do not show a correlation between triplex motifs and MLS in the whole dataset (**Table S6**), the correlations between triplex motifs and MLS are stronger in bird orders with higher mean triplex levels (B).

A) Mammals with a higher number of triplex motifs in their mtDNA (default settings) are on average shorter-lived.

B) The higher the mean number of triplex motifs in a bird order (bar graphs) the stronger the inverse correlation between triplex motifs and MLS in the same order (blue line). All bird orders with more than 5 species in our dataset are included in this graph.

DISCUSSION

Repeat motifs and mtDNA deletions

Our goal was to define motifs associated with mtDNA deletions that would be then expected to show an inverse correlation with MLS under the theory of resistant biomolecules. We found that DR, MR, G-quadruplex and triplex motifs would be plausible candidates.

Specifically, our results support the consensus that DR and G-quadruplex motifs are major drivers of deletion mutagenesis (**Persson et al. 2019**) but not IR motifs (**Dong et al. 2014;** but see also: **Damas et al. 2012**, **Mikhailova et al. 2020**). In addition, we found a novel association between MR motifs and mtDNA deletions (**Fig. 2A**). This was unexpected since MR motifs do not allow for canonical base-pairing. Nonetheless, some studies have found MR motifs to be associated with indels (**Georgakopoulos-Soares et al. 2018**), perhaps due to their ability to form triplex DNA (**Kamat et al. 2016**).

Triplex forming motifs are mutagenic in bacteria (**Holder et al. 2015**), associated with deletions and translocations in cancer (**Bacolla et al. 2016**) and with mutations underlying various inherited diseases (**Kamat et al. 2016**), but little was known about their role in mitochondria. This is the first report that mitochondrial triplex motifs are indeed associated with mtDNA deletions (**Fig. 3**) and MLS (**Fig. 7**).

Some mtDNA motifs colocalize and may interact

One key finding that emerged during our analysis was that motif classes correlate with each other, e.g.,

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

if both classes are enriched in DNA regions with a biased composition. Mitochondrial DR motifs are often close to MR and G-quadruplex motifs (**Fig. S18, Table S4**), IR motifs to ER motifs (**Table S10**) and triplex motifs to G-quadruplex motifs (**Fig. 3**). Our work is a first attempt to disentangle these interactions both in relation to deletion formation and to lifespan.

In principle one can correct for this by excluding hybrid motifs, i.e. two motifs located close to each other, from the analysis, as we have done in **Fig. 2-4**, but it is not always clear whether this is biologically sensible or overly conservative.

In the case of MR repeats we are inclined to conclude that they do not contribute to mtDNA instability, since there is little experimental data to implicate non-triplex-forming MRs in DNA instability and the enrichment of MR motifs at deletion breakpoints is attenuated after we exclude MR-DR hybrid motifs from the analysis (**Fig. 2C; Fig. S6, S7**).

In contrast, both triplex and G-quadruplex motifs are strongly implicated in DNA instability. A high G-content is known to promote G-quadruplex formation and to stabilize triplex DNA (**Buske et al. 2011**, **Kaufmann et al. 2019**) thus explaining their colocalization**.** Although G-quadruplexes may compete with the formation of triplex DNA, this is not always so (**Solé et al. 2017, del Mundo et al. 2017**). Furthermore, in mtDNA, where overlapping triplex and G-quadruplex motifs can be located on opposite strands, a secondary structure on one strand could potentially promote formation of a secondary structure on the other strand by preventing reannealing of the two strands, as has been suggested for R-loops that promote the formation of G-quadruplexes (**De Magis et al. 2019**).

In support of the idea that motifs can interact, we found that G-quadruplex and triplex motifs in proximity to each other showed stronger associations with deletion breakpoints than isolated motifs, suggesting that their effects are additive (**Fig. S10**).

Similarly, it is not clear if the enrichment of G-quadruplex motifs at mtDNA deletion breakpoints needs to be corrected for their proximity to DR motifs (**Table S4**) found in regions of high GC-skew (**Fig. S18**). In a preliminary analysis we found that removal of DR-G-quadruplex hybrid sequences can greatly attenuate the over-representation of G-quadruplex motifs around breakpoints (data not shown). A thorough re-analysis of this phenomenon is beyond the scope of this work, but it does suggest that colocalization of different motifs is not a problem unique to our triplex dataset.

mtDNA motifs in health and disease

We found that triplex motifs were particularly enriched around breakpoints in datasets from patients with mitochondrial disease (**Fig. 4C**). Conceivably, polymerase stalling associated with mitochondrial disease (**Wanrooij and Falkenberg 2010**) could allow more time for the formation of triplex DNA during replication, thereby explaining this finding. Consistent with this idea, a preprint by **Lakshmanan et al. (2018)** showed that mtDNA deletions in patients with mitochondrial disease due to POLG mutations are explained by predominantly short repeats and this bias towards short repeats was recently confirmed by a larger analysis that compared POLG patients to healthy controls (**Lujan et al. 2020**). These findings are intriguing because medium and long repeats are preferably associated with strand-slippage mechanisms whereas short repeats are consistent with strand breaks (**Nissanka et al. 2019, Lakshmanan et al. 2018**) which can be caused by ROS or secondary structures like G-quadruplex and triplex motifs.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Analysis of mutagenic motifs and their relationship with lifespan casts doubt on the theory of resistant biomolecules

The theory of resistant biomolecules can be seen to imply a correlation between mutagenic motifs and MLS, but despite promising data from earlier studies (**Lakshmanan et al. 2012, Yang et al. 2013**), we found that neither G-quadruplex, DR, MR, IR nor ER motifs were associated with MLS after correcting for phylogeny and base composition (**Table 1**).

Perhaps the mtDNA is too streamlined to allow for the necessary changes in sequence. Given thousands of short, potentially mutagenic motifs (**Fig. S11**) that would need to be lost, we speculate it is more likely that nuclear proteins involved in mtDNA maintenance and metabolism account for the different rates of age-related deletion accumulation across species (**Guo et al. 2010**). In fact, repeats enable error prone DNA repair (**Tadi et al. 2016**), which might be advantageous under genotoxic stress to prevent wholescale mtDNA degradation and only detrimental late in life due to deletion accumulation, as would be predicted by the theory of antagonistic pleiotropy (**Campisi and Vijg 2009**).

The literature appears largely consistent with the null hypothesis. The work by **Yang et al. (2013)** may be considered the strongest case in favor of resistant biomolecules. They found an inverse relationship between IR motifs and MLS after correcting for phylogeny only. We found a similar correlation which, however, is fully accounted for by base composition. Two other findings of their paper are also consistent with the null hypothesis. Longer IR motifs, presumably allowing more stable base pairing, did not show a stronger inverse correlation with MLS, which we confirmed (**Fig. S12C**). Secondly, IR motifs with half-sites close together failed to show a stronger relationship with MLS, although these should form secondary structures more easily.

The importance of mtDNA base composition in motif-lifespan correlations

It is striking that most of the findings were attenuated when we controlled for the base composition of mtDNA. Although the importance of such an adjustment was already highlighted by **Lakshmanan et al. (2015)** for DR motifs, we are the first to apply these corrections in the study of multiple DNA motifs.

We find it particularly informative to consider related motif classes together (**Fig. 5B, C**). ER motifs, for example, just like IR motifs show a strong correlation with GC content and GC skew, which in turn correlate with MLS. Importantly, ER motifs were not associated with mtDNA deletion breakpoints (**Fig. 2A, D**), are not known to be mutagenic, do not permit canonical base pairing and yet show a stronger inverse correlation with MLS than do IR motifs, which is fully attenuated after adjustment for base composition (**Table 1**). Thus, we conclude that most earlier results were driven by compositional bias.

Can triplex motifs save the theory of resistant biomolecules?

We showed that intramolecular triplex motifs detected by the triplex package are inversely correlated with mammalian MLS after correcting for body mass, base composition and phylogeny (**Table 1**). Although this moderate inverse correlation with MLS provides some support for this theory, it is not clear why major arc triplex motifs show a weaker than expected correlation with MLS (**Table S7**) nor if the triplex data can be generalized to non-mammalian species.

While we do not know a priori which motifs might be mutagenic in species with different mtDNA biology, we did find preliminary evidence that triplex motifs are inversely correlated with MLS in bird orders with high numbers of triplex motifs in their mtDNA (**Fig. 7B**) but no such correlation in ray-finned fishes. Since we observed the same trend in mammals, this suggests a potential threshold effect.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Perhaps a small number of triplex motifs in mtDNA is well tolerated, whereas larger numbers are progressively more destabilizing, especially if they occur in clusters as demonstrated for nuclear triplex motifs (**Fig. S15**).

Going forward, finding the right motifs?

Why did an earlier study (**Oliveira et al. 2013**) fail to uncover an association between triplex motifs and mtDNA deletions? According to **Hon et al. (2013)** the nBMST tool used in that study only detects triplex-like mirror repeats and we also found it to be not very sensitive for the detection of putative triplex sequences. Given these shortcomings we also tested two other tools. The first one was Triplexator, which detects triplex target and triplex forming sites that could form both intra- and intermolecular triplexes, and the second one was the triplex package that detects intramolecular triplexes. We reasoned that intramolecular triplexes would be the most stable thus settling to use the latter tool. Using the triplex package we were able to confirm the well-established finding that nuclear triplex motifs are found near cancer-related breakpoints, supporting the validity of this tool (**Table S8**).

Although some of the sequences detected with Triplexator and the triplex package overlapped, many did not and in a comparison of 600 mammalian species the numbers of detected motifs did not correlate between the two tools (**Fig. S14**). This highlights a broader issue about the selection of motif detection tools. While multiple mature G-quadruplex detection tools exist that tend to agree with each other (**Fig. S13; Lombardi and Londoño-Vallejo 2020**), the choices for triplex motifs are more limited and there is little in vivo validation in mammalian genomes (**Kaufmann et al. 2019**).

Given these uncertainties, perhaps going forward we can refine the search parameters to find motifs that show a stronger correlation with lifespan and mtDNA deletions. To this end, our analysis of mitochondrial and nuclear breakpoints raises the possibility that hybrid and clustered motifs are particularly harmful (**Fig. 4E; Fig. S15**), but we have not yet analyzed these across genomes. Another issue that we did not address in detail is the type and orientation of triplex motifs involved in mtDNA deletions. Most importantly, future studies should explore to what extent these predicted intramolecular triplexes form in mitochondria.

ACKNOWLEDGEMENTS

We thank Doug Turnbull, Amy Vincent, David Meyer and Teresa Valencak for helpful comments and/or support. Reginald Smith for providing the python code to calculate Wright's Nc, Mitya Toren for sharing updated MitoAge data and John A. Tainer for sharing triplex-related data.

REFERENCES

Aledo, J. C., Valverde, H., & de Magalhães, J. P. (2012). Mutational Bias Plays an Important Role in Shaping Longevity-Related Amino Acid Content in Mammalian mtDNA-Encoded Proteins. Journal of Molecular Evolution, 74(5–6), 332–341. https://doi.org/10.1007/s00239-012-9510-7

Bacolla, A., Tainer, J. A., Vasquez, K. M., & Cooper, D. N. (2016). Translocation and deletion breakpoints in cancer genomes are associated with potential non-B DNA-forming sequences. Nucleic Acids Research, 44(12), 5673–5688. https://doi.org/10.1093/nar/gkw261

Bender, A., Krishnan, K. J., Morris, C. M., Taylor, G. A., Reeve, A. K., Perry, R. H., Jaros, E., Hersheson, J. S., Betts, J., Klopstock, T., Taylor, R. W., & Turnbull, D. M. (2006). High levels of mitochondrial DNA deletions

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

in substantia nigra neurons in aging and Parkinson disease. Nature Genetics, 38(5), 515–517. https://doi.org/10.1038/ng1769

Bharti, S. K., Sommers, J. A., Zhou, J., Kaplan, D. L., Spelbrink, J. N., Mergny, J.-L., & Brosh, R. M. (2014). DNA Sequences Proximal to Human Mitochondrial DNA Deletion Breakpoints Prevalent in Human Disease Form G-quadruplexes, a Class of DNA Structures Inefficiently Unwound by the Mitochondrial Replicative Twinkle Helicase. Journal of Biological Chemistry, 289(43), 29975–29993. https://doi.org/10.1074/jbc.M114.567073

Bissler, J. J. (2007). Triplex DNA and human disease. Frontiers in Bioscience: A Journal and Virtual Library, 12, 4536–4546. https://doi.org/10.2741/2408

Buske, F. A., Bauer, D. C., Mattick, J. S., & Bailey, T. L. (2013). Triplex-Inspector: An analysis tool for triplex-mediated targeting of genomic loci. Bioinformatics, 29(15), 1895–1897. https://doi.org/10.1093/bioinformatics/btt315

Buske, F. A., Mattick, J. S., & Bailey, T. L. (2011). Potential in vivo roles of nucleic acid triple-helices. RNA Biology, 8(3), 427–439. https://doi.org/10.4161/rna.8.3.14999

Butler, T. J., Estep, K. N., Sommers, J. A., Maul, R. W., Moore, A. Z., Bandinelli, S., Cucca, F., Tuke, M. A., Wood, A. R., Bharti, S. K., Bogenhagen, D. F., Yakubovskaya, E., Garcia-Diaz, M., Guilliam, T. A., Byrd, A. K., Raney, K. D., Doherty, A. J., Ferrucci, L., Schlessinger, D., … Brosh, R. M. (2020). Mitochondrial genetic variation is enriched in G-quadruplex regions that stall DNA synthesis in vitro. Human Molecular Genetics, 29(8), 1292–1309. https://doi.org/10.1093/hmg/ddaa043

Campisi, J., & Vijg, J. (2009). Does Damage to DNA and Other Macromolecules Play a Role in Aging? If So, How? The Journals of Gerontology: Series A, 64A(2), 175–178. https://doi.org/10.1093/gerona/gln065

Cer, R. Z., Bruce, K. H., Donohue, D. E., Temiz, A. N., Bacolla, A., Mudunuri, U. S., Yi, M., Volfovsky, N., Luke, B. T., Collins, J. R., & Stephens, R. M. (2011). Introducing the non-B DNA Motif Search Tool (nBMST). Genome Biology, 12(1), P34. https://doi.org/10.1186/1465-6906-12-S1-P34

Damas, J., Carneiro, J., Amorim, A., & Pereira, F. (2014). MitoBreak: The mitochondrial DNA breakpoints database. Nucleic Acids Research, 42(D1), D1261–D1268. https://doi.org/10.1093/nar/gkt982

Damas, J., Carneiro, J., Gonçalves, J., Stewart, J. B., Samuels, D. C., Amorim, A., & Pereira, F. (2012). Mitochondrial DNA deletions are associated with non-B DNA conformations. Nucleic Acids Research, 40(16), 7606–7621. https://doi.org/10.1093/nar/gks500

del Mundo, I. M. A., Zewail-Foote, M., Kerwin, S. M., & Vasquez, K. M. (2017). Alternative DNA structure formation in the mutagenic human c-MYC promoter. Nucleic Acids Research, 45(8), 4929–4943. https://doi.org/10.1093/nar/gkx100

Doluca, O., Withers, J. M., & Filichev, V. V. (2013). Molecular Engineering of Guanine-Rich Sequences: Z-DNA, DNA Triplexes, and G-Quadruplexes. Chemical Reviews, 113(5), 3044–3083. https://doi.org/10.1021/cr300225q

Dong, D. W., Pereira, F., Barrett, S. P., Kolesar, J. E., Cao, K., Damas, J., Yatsunyk, L. A., Johnson, F. B., & Kaufman, B. A. (2014). Association of G-quadruplex forming sequences with human mtDNA deletion breakpoints. BMC Genomics, 15(1), 677. https://doi.org/10.1186/1471-2164-15-677

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

583 Gammage, P. A., Rorbach, J., Vincent, A. I., Rebar, E. J., & Minczuk, M. (2014). Mitochondrially targeted
584 ZFNs for selective degradation of pathogenic mitochondrial genomes bearing large-scale deletions or
585 point mutations. EMBO Molecular Medicine, 6(4), 458–466. https://doi.org/10.1002/emmm.201303672

586 Georgakopoulos-Soares, I., Morganella, S., Jain, N., Hemberg, M., & Nik-Zainal, S. (2018). Noncanonical
587 secondary structures arising from non-B DNA motifs are determinants of mutagenesis. Genome Research,
588 28(9), 1264–1271. https://doi.org/10.1101/gr.231688.117

589 Gladyshev, V. N. (2013). The Free Radical Theory of Aging Is Dead. Long Live the Damage Theory!
590 Antioxidants & Redox Signaling, 20(4), 727–731. https://doi.org/10.1089/ars.2013.5228

591 Guo, X., Kudryavtseva, E., Bodyak, N., Nicholas, A., Dombrovsky, I., Yang, D., Kraytsberg, Y., Simon, D. K., &
592 Khrapko, K. (2010). Mitochondrial DNA deletions in mice in men: Substantia nigra is much less affected in
593 the mouse. Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1797(6), 1159–1162.
594 https://doi.org/10.1016/j.bbabio.2010.04.005

595 Gurusaran, M., Ravella, D., & Sekar, K. (2013). RepEx: Repeat extractor for biological sequences.
596 Genomics, 102(4), 403–408. https://doi.org/10.1016/j.ygeno.2013.07.005

597 Hamilton, K. L., & Miller, B. F. (2016). What is the evidence for stress resistance and slowed aging?
598 Experimental Gerontology, 82, 67–72. https://doi.org/10.1016/j.exger.2016.06.001

599 Hjelm, B. E., Rollins, B., Morgan, L., Sequeira, A., Mamdani, F., Pereira, F., Damas, J., Webb, M. G., Weber,
600 M. D., Schatzberg, A. F., Barchas, J. D., Lee, F. S., Akil, H., Watson, S. J., Myers, R. M., Chao, E. C., Kimonis,
601 V., Thompson, P. M., Bunney, W. E., & Vawter, M. P. (2019). Splice-Break: Exploiting an RNA-seq splice
602 junction algorithm to discover mitochondrial DNA deletion breakpoints and analyses of psychiatric
603 disorders. Nucleic Acids Research, 47(10), e59–e59. https://doi.org/10.1093/nar/gkz164

604 Holder, I. T., Wagner, S., Xiong, P., Sinn, M., Frickey, T., Meyer, A., & Hartig, J. S. (2015). Intrastrand triplex
605 DNA repeats in bacteria: A source of genomic instability. Nucleic Acids Research, 43(21), 10126–10142.
606 https://doi.org/10.1093/nar/gkv1017

607 Hon, J., Martínek, T., Rajdl, K., & Lexa, M. (2013). Triplex: An R/Bioconductor package for identification
608 and visualization of potential intramolecular triplex patterns in DNA sequences. Bioinformatics, 29(15),
609 1900–1901. https://doi.org/10.1093/bioinformatics/btt299

610 Hon, J., Martínek, T., Zendulka, J., & Lexa, M. (2017). pqsfinder: An exhaustive and imperfection-tolerant
611 search tool for potential quadruplex-forming sequences in R. Bioinformatics, 33(21), 3373–3379.
612 https://doi.org/10.1093/bioinformatics/btx413

613 Kamat, M. A., Bacolla, A., Cooper, D. N., & Chuzhanova, N. (2016). A Role for Non-B DNA Forming
614 Sequences in Mediating Microlesions Causing Human Inherited Disease. Human Mutation, 37(1), 65–73.
615 https://doi.org/10.1002/humu.22917

616 Kaufmann, B., Willinger, O., Eden, N., Kermas, L., Anavy, L., Solomon, O., Atar, O., Yakhini, Z., Goldberg, S.,
617 & Amit, R. (2019). Identifying triplex binding rules in vitro leads to creation of a new synthetic regulatory
618 tool in vivo. BioRxiv, 2019.12.25.888362. https://doi.org/10.1101/2019.12.25.888362

619 Kauppila, T. E. S., Kauppila, J. H. K., & Larsson, N.-G. (2017). Mammalian Mitochondria and Aging: An
620 Update. Cell Metabolism, 25(1), 57–71. https://doi.org/10.1016/j.cmet.2016.09.017

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

621 Khaidakov, M., Siegel, E. R., & Shmookler Reis, R. J. (2006). Direct repeats in mitochondrial DNA and
622 mammalian lifespan. Mechanisms of Ageing and Development, 127(10), 808–812.
623 https://doi.org/10.1016/j.mad.2006.07.008

624 Khrapko, K., Kraytsberg, Y., Grey, A. D. D., Vijg, J., & Schon, E. A. (2006). Does premature aging of the
625 mtDNA mutator mouse prove that mtDNA mutations are involved in natural aging? Aging Cell, 5(3), 279–
626 282. https://doi.org/10.1111/j.1474-9726.2006.00209.x

627 Khristich, A. N., & Mirkin, S. M. (2020). On the wrong DNA track: Molecular mechanisms of repeat-
628 mediated genome instability. Journal of Biological Chemistry, 295(13), 4134–4170.
629 https://doi.org/10.1074/jbc.REV119.007678

630 Lakshmanan, L. N., Gruber, J., Halliwell, B., & Gunawan, R. (2012). Role of Direct Repeat and Stem-Loop
631 Motifs in mtDNA Deletions: Cause or Coincidence? PLoS ONE, 7(4).
632 https://doi.org/10.1371/journal.pone.0035271

633 Lakshmanan, L. N., Gruber, J., Halliwell, B., & Gunawan, R. (2015). Are mutagenic non D-loop direct repeat
634 motifs in mitochondrial DNA under a negative selection pressure? Nucleic Acids Research, 43(8), 4098–
635 4108. https://doi.org/10.1093/nar/gkv299

636 Lakshmanan, L. N., Yee, Z., Gruber, J., Halliwell, B., & Gunawan, R. (2018). Thermodynamic analysis of
637 mitochondrial DNA breakpoints reveals mechanistic details of deletion mutagenesis [Preprint]. Systems
638 Biology. https://doi.org/10.1101/254631

639 Lawless, C., Greaves, L., Reeve, A. K., Turnbull, D. M., & Vincent, A. E. (2020). The rise and rise of
640 mitochondrial DNA mutations. Open Biology, 10(5), 200061. https://doi.org/10.1098/rsob.200061

641 Lujan, S. A., Longley, M. J., Humble, M. H., Lavender, C. A., Burkholder, A., Blakely, E. L., Alston, C. L.,
642 Gorman, G. S., Turnbull, D. M., McFarland, R., Taylor, R. W., Kunkel, T. A., & Copeland, W. C. (2020).
643 Ultrasensitive deletion detection links mitochondrial DNA replication, disease, and aging. Genome
644 Biology, 21(1), 248. https://doi.org/10.1186/s13059-020-02138-5

645 Magis, A. D., Manzo, S. G., Russo, M., Marinello, J., Morigi, R., Sordet, O., & Capranico, G. (2019). DNA
646 damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells.
647 Proceedings of the National Academy of Sciences, 116(3), 816–825.
648 https://doi.org/10.1073/pnas.1810409116

649 Mikhailova, A. A., Shamanskiy, V., Ushakova, K., Mikhailova, A. G., Oreshkov, S., Knorre, D., Tretiakov, E.
650 O., Zazhytska, M., Lukowski, S. W., Liou, C.-W., Lin, T.-K., Kunz, W. S., Reymond, A., Mazunin, I., Bazykin, G.
651 A., Gunbin, K., Fellay, J., Tanaka, M., Khrapko, K., & Popadin, K. (2020). Risk of mitochondrial deletions is
652 affected by the global secondary structure of the mitochondrial genome. BioRxiv, 603282.
653 https://doi.org/10.1101/603282

654 Miller, R. A. (2009). Cell Stress and Aging: New Emphasis on Multiplex Resistance Mechanisms. The
655 Journals of Gerontology: Series A, 64A(2), 179–182. https://doi.org/10.1093/gerona/gln072

656 Nissanka, N., Minczuk, M., & Moraes, C. T. (2019). Mechanisms of Mitochondrial DNA Deletion
657 Formation. Trends in Genetics, 35(3), 235–244. https://doi.org/10.1016/j.tig.2019.01.001

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Oliveira, P. H., Silva, C. L. da, & Cabral, J. M. S. (2013). An Appraisal of Human Mitochondrial DNA Instability: New Insights into the Role of Non-Canonical DNA Structures and Sequence Motifs. PLOS ONE, 8(3), e59907. https://doi.org/10.1371/journal.pone.0059907

Pacifici, M., Santini, L., Marco, M. D., Baisero, D., Francucci, L., Marasini, G. G., Visconti, P., & Rondinini, C. (2013). Generation length for mammals. Nature Conservation, 5, 89–94. https://doi.org/10.3897/natureconservation.5.5734

Pamplona, R., & Barja, G. (2007). Highly resistant macromolecular components and low rate of generation of endogenous damage: Two key traits of longevity. Ageing Research Reviews, 6(3), 189–210. https://doi.org/10.1016/j.arr.2007.06.002

Persson, Ö., Muthukumar, Y., Basu, S., Jenninger, L., Uhler, J. P., Berglund, A.-K., McFarland, R., Taylor, R. W., Gustafsson, C. M., Larsson, E., & Falkenberg, M. (2019). Copy-choice recombination during mitochondrial L-strand synthesis causes DNA deletions. Nature Communications, 10(1), 759. https://doi.org/10.1038/s41467-019-08673-5

Puig Lombardi, E., & Londoño-Vallejo, A. (2020). A guide to computational methods for G-quadruplex prediction. Nucleic Acids Research, 48(1), 1–15. https://doi.org/10.1093/nar/gkz1097

Richardson, A. G., & Schadt, E. E. (2014). The Role of Macromolecular Damage in Aging and Age-related Disease. The Journals of Gerontology: Series A, 69(Suppl_1), S28–S32. https://doi.org/10.1093/gerona/glu056

Shamanskiy, V. N., Timonina, V. N., Popadin, K. Yu., & Gunbin, K. V. (2019). ImtRDB: A database and software for mitochondrial imperfect interspersed repeats annotation. BMC Genomics, 20(3), 295. https://doi.org/10.1186/s12864-019-5536-1

Smith, R. (2019). Enhanced effective codon numbers to understand codon usage bias. BioRxiv, 644609. https://doi.org/10.1101/644609

Solé, A., Delagoutte, E., Ciudad, C. J., Noé, V., & Alberti, P. (2017). Polypurine reverse-Hoogsteen (PPRH) oligonucleotides can form triplexes with their target sequences even under conditions where they fold into G-quadruplexes. Scientific Reports, 7(1), 39898. https://doi.org/10.1038/srep39898

Speakman, J. R. (2005). Correlations between physiology and lifespan – two widely ignored problems with comparative studies. Aging Cell, 4(4), 167–175. https://doi.org/10.1111/j.1474-9726.2005.00162.x

Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., Diana, E., Lehmann, G., Toren, D., Wang, J., Fraifeld, V. E., & de Magalhães, J. P. (2018). Human Ageing Genomic Resources: New and updated databases. Nucleic Acids Research, 46(D1), D1083–D1090. https://doi.org/10.1093/nar/gkx1042

Tadi, S. K., Sebastian, R., Dahal, S., Babu, R. K., Choudhary, B., & Raghavan, S. C. (2015). Microhomology-mediated end joining is the principal mediator of double-strand break repair during mitochondrial DNA lesions. Molecular Biology of the Cell, 27(2), 223–235. https://doi.org/10.1091/mbc.e15-05-0260

Toren, D., Barzilay, T., Tacutu, R., Lehmann, G., Muradian, K. K., & Fraifeld, V. E. (2016). MitoAge: A database for comparative analysis of mitochondrial DNA, with a special focus on animal longevity. Nucleic Acids Research, 44(Database issue), D1262–D1265. https://doi.org/10.1093/nar/gkv1187

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.13.381475; this version posted January 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

695 Tremblay-Belzile, S., Lepage, É., Zampini, É., & Brisson, N. (2015). Short-range inversions: Rethinking
696 organelle genome stability. BioEssays, 37(10), 1086–1094. https://doi.org/10.1002/bies.201500064

697 Valencak, T. G., & Ruf, T. (2007). N−3 polyunsaturated fatty acids impair lifespan but have no role for
698 metabolism. Aging Cell, 6(1), 15–25. https://doi.org/10.1111/j.1474-9726.2006.00257.x

699 Wanrooij, S., & Falkenberg, M. (2010). The human mitochondrial replication fork in health and disease.
700 Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1797(8), 1378–1388.
701 https://doi.org/10.1016/j.bbabio.2010.04.015

702 Yang, J.-N., Seluanov, A., & Gorbunova, V. (2013). Mitochondrial Inverted Repeats Strongly Correlate with
703 Lifespan: MtDNA Inversions and Aging. PLOS ONE, 8(9), e73318.
704 https://doi.org/10.1371/journal.pone.0073318