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 (GQ) 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 GQ motifs.

Surprisingly, even though DR, MR and GQ 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 and these results remained stable 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

PGLS, phylogenetic generalized least squares

SD, standard deviation

Trip, Triplex forming motif

XR, any repeat half-site or motif

mtDNA, mitochondrial DNA

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 a lower content of oxidation-prone cysteine and methionine in mitochondrially encoded proteins (**Aledo et al. 2011**), although the interpretation of the latter is complicated due to mtDNA mutational and compositional biases (**Aledo et al. 2012**).

In the context of DNA, mutations constitute an irreversible type of damage that accumulates over time. Nuclear mutations, for example, are linked with aging and cancer (**Niedernhofer et al. 2018**). Similarly, mitochondrial DNA (mtDNA) point mutations accumulate with aging and in some progerias (**Kauppila et al. 2017**), while the accumulation of deletions 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 also applied to mtDNA, we should see fewer DNA motifs that form mutagenic structures in the mtDNA of long-lived species.

Potentially mutagenic motifs include repeats, leading to DNA mispairing, and motifs that promote the formation of non-B DNA conformations, like guanine-quadruplex (GQ)- and triplex-forming motifs (**Kamat et al. 2016, Bacolla et al. 2016; Fig. 1**). Among these, direct repeats (DR) can lead to strand-slippage if they mispair with a distant motif on the complementary strand (**Persson et al. 2019**) and inverted repeats (IR) can form intrastrand hairpins that destabilize progression of the replication fork (**Tremblay-Belzile et al. 2015**). In addition to DR and IR repeat motifs, there exist two related motif classes, mirror repeats (MR) and everted repeats (ER), that do not form stable Watson-Crick base pairs and have not been linked with mtDNA instability.

What is the evidence for the theory of resistant biomolecules in the context of mtDNA? 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 lifespan was seen in recent studies (**Lakshmanan et al. 2015**). In contrast, with the exception of one preprint (**Mikhailova et al. 2019**), IRs are not known to be associated with mtDNA deletions (**Dong et al. 2014**), although they do show a negative relationship with lifespan (**Yang et al. 2013**) and may contribute to inversions (**Yang et al. 2013, Tremblay-Belzile et al. 2015**). Whether age-related mtDNA inversions underlie any pathology, however, requires further study. Finally, G-quadruplex motifs are consistently associated with both deletions (**Dong et al. 2014**) and point mutations (**Butler et al. 2020**), but no study tested if they correlate with lifespan. Triplex motifs are poorly studied with one report finding no association between these motifs and deletions (**Oliveira et al. 2013**).

Here we 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 maximum lifespan (MLS).

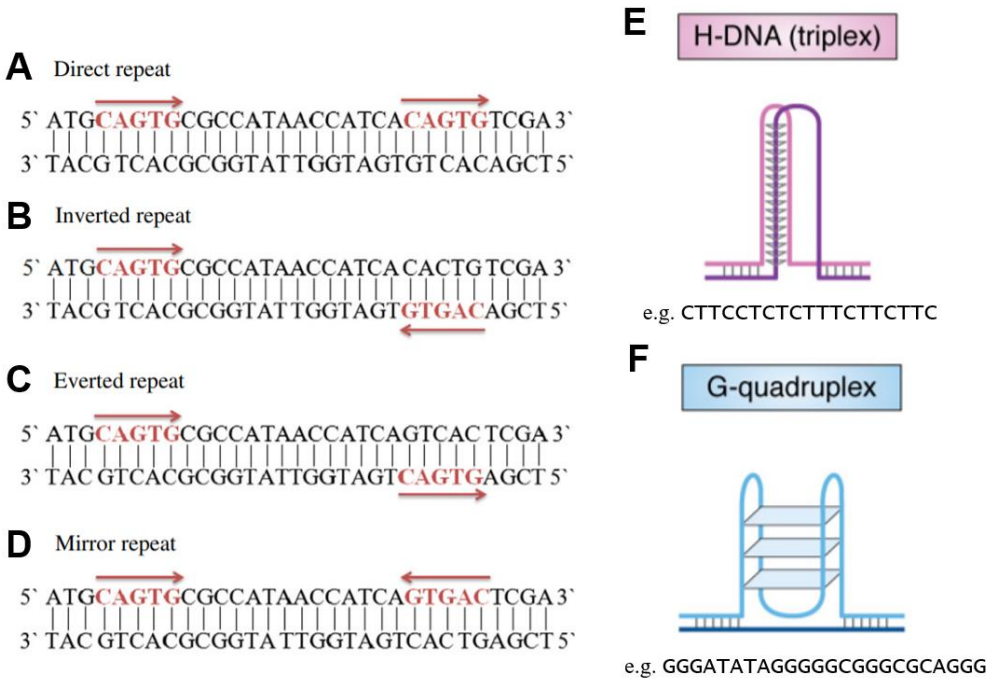


Figure 1

A. Direct repeat, second half-site has the same orientation.
B. Inverted repeat, second half-site is complementary and has mirror symmetry.
C. Everted repeat, second half-site is complementary.
D. Mirror repeat, second half-site has 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.

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

Our analysis included only major arc deletions, because during asynchronous mtDNA replication the major arc is single stranded for extended periods of time (**Persson et al. 2019**) which should favor the formation of triplexes and G-quadruplexes. The major arc was defined as the region between position 5747 and 16500 of the human mitochondrial genome (NC_012920.1), totaling 1066 deletions from the MitoBreak database (**Damas et al. 2014, mtDNA Breakpoints.xlsx**), 1114 deletions from **Persson et al. (2019)** and 1894 from **Hjelm et al. (2019)**.

A 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; see **Fig. S5A, B** for different window sizes).

Three different orientations relative to the breakpoints were considered. Two orientations for motifs with half-sites (i.e. repeats), either each of the two half-sites flanking one breakpoint of a deletion, or both half-sites flanking any one 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.

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 close proximity.

To generate controls, the mtDNA deletions were randomly redistributed inside the major arc which allowed us to approximate the original distribution and size 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. Several different shuffling methods produced similar results (data not shown).

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 a cleaning step, 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

We used Clustal Omega from the msa package (v1.18.0) to align cytochrome B DNA sequences and construct phylogenetic trees via neighbor joining. In the resulting mammalian and bird tree, four branch edge lengths were equal to zero and these were set to the lowest non-zero value in the dataset. Based on these trees we corrected the correlations between mtDNA motifs and lifespan for shared species ancestry using phylogenetic generalised least squares (PGLS) implemented in the caper package (v1.0.1).

RESULTS

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

In order to define the 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. Finally, in our analysis 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. 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 because they form a secondary structure at any of the breakpoints. However, in this analysis no motif class showed enrichment around breakpoints (**Fig. 2D**).

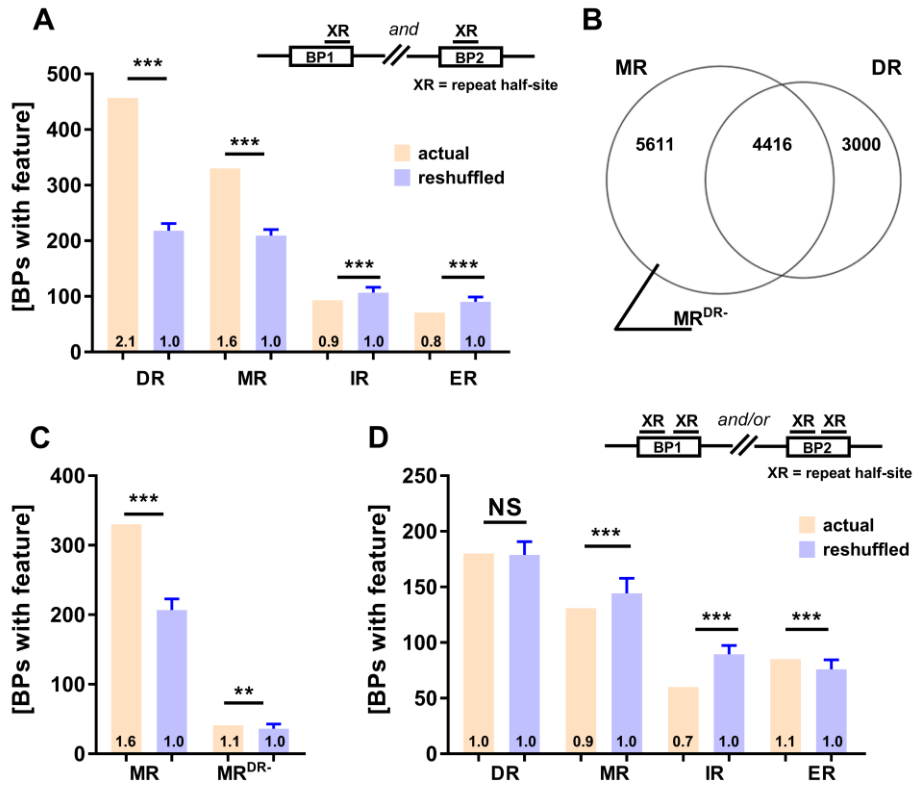


Figure 2

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 adjusted for MRs that are equivalent to a DR motif (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**).

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. 2012**) 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. Triplex motifs were shorter and less abundant than predicted G-quadruplexes (**Table S1, S3**) and associated with fewer breakpoints altogether (**Fig. 3**). In contrast to G-quadruplexes, almost exclusive to the G-rich mtDNA heavy-strand, triplex motifs were also common on the light-strand and more often associated with the 3'-breakpoints compared to G-quadruplex motifs (**Table S4**).

The six triplex motifs detected by the triplex package were significantly enriched around deletion breakpoints 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 triplex 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 DRs. For this, 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 S5**).

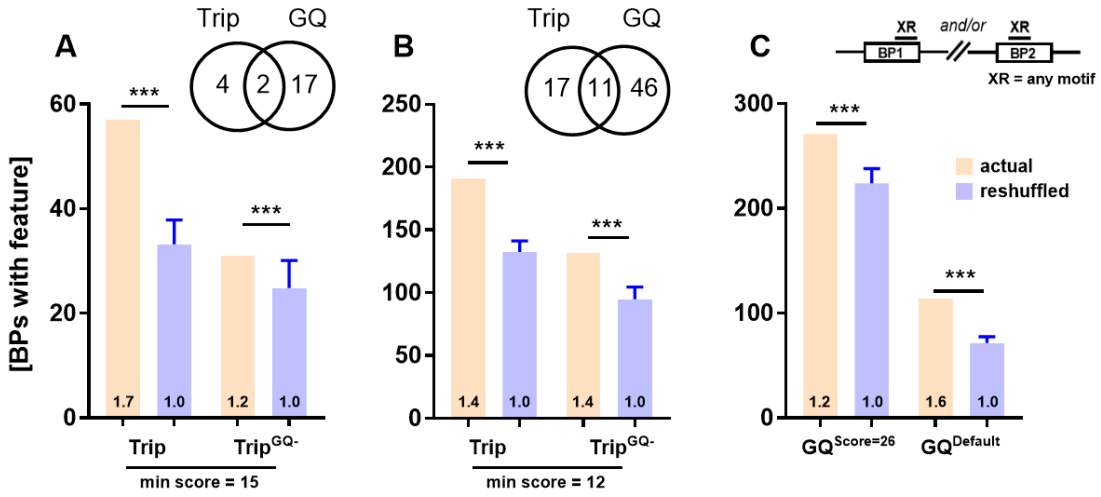


Figure 3

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 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 S6**). 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**). Nevertheless, triplex motifs were also enriched at breakpoints in the pooled analysis (**Fig. 4D**), although to a lesser extent.

Despite differences between these datasets we found that often the same individual motifs were associated with breakpoints across all three datasets (**Table S7**). Using the pooled dataset we also confirmed that triplex motifs are more often associated with 3'-breakpoints compared to G-quadruplex motifs (**Table S8**) and that G-quadruplex motifs close to triplex motifs are more strongly enriched at deletion breakpoints than solitary G-quadruplex motifs (**Fig. 4E; Fig. S9**), suggesting that triplex formation may further contribute to DNA instability.

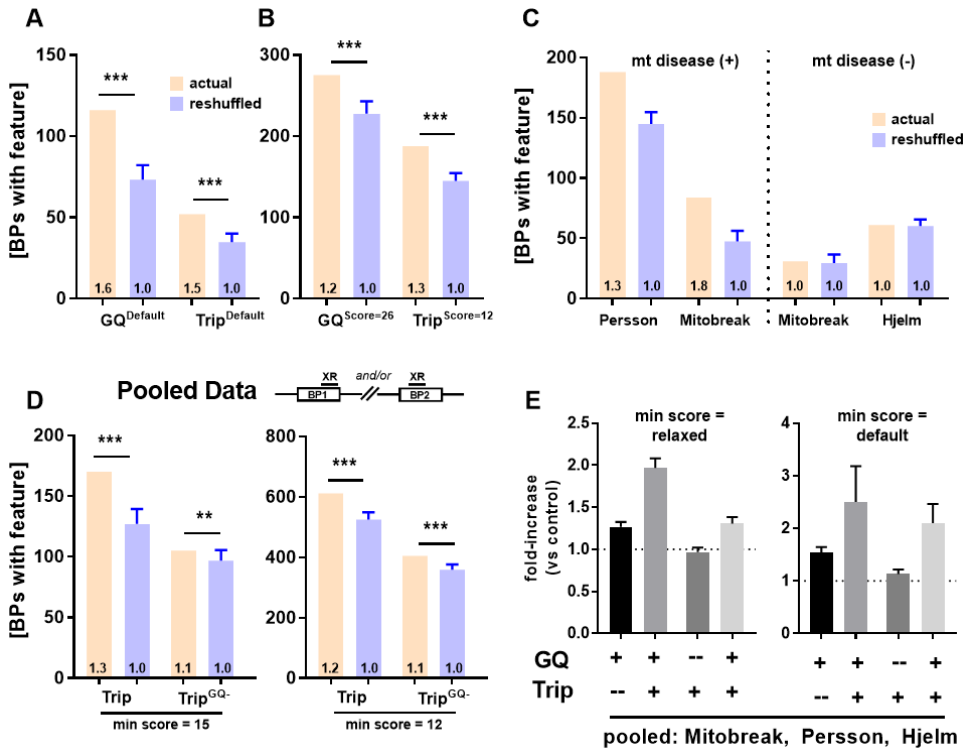


Figure 4

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 repeat (XR) half-sites and motifs 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 positive vs negative. 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 plotted. Motifs were considered overlapping if their midpoints were within 50 bp.

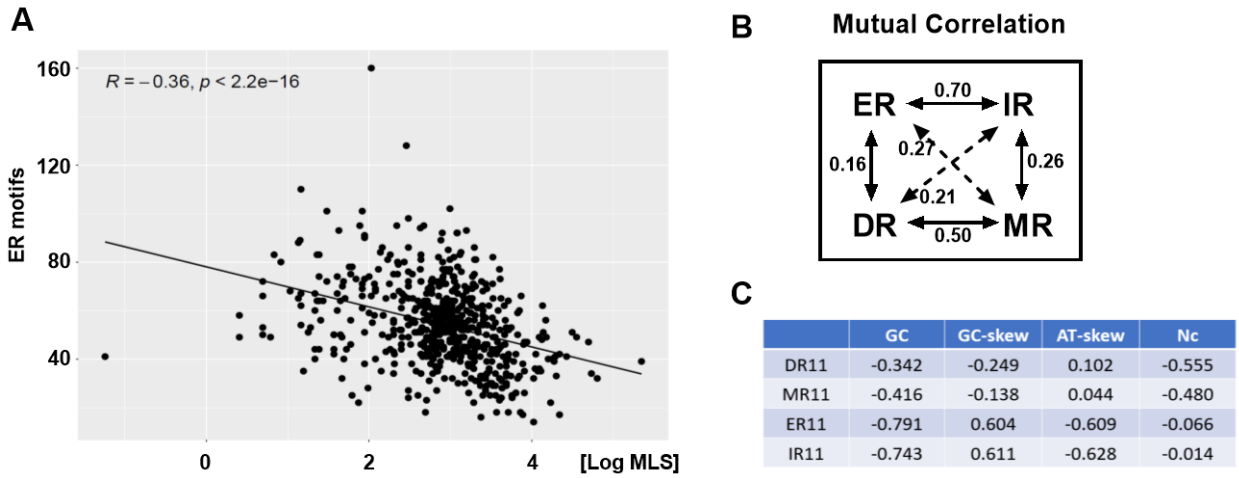
Repeats and lifespan: no support for the theory of resistant biomolecules

For our analysis, we focus on 11 bp long repeat motifs because short repeats are unlikely to allow stable base pairing and longer repeats are very rare (**Fig. S10**). What is more, results considering repeat motifs of different lengths usually agree with each other (**Table S9**; **Yang et al. 2013**). To allow comparability with early studies (**Lakshmanan et al. 2015**) we also focus on non D-loop motifs here.

First of all, 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**). Similarly, we confirmed prior data from **Lakshmanan et al. (2015)** finding that DR motifs do not correlate with the MLS of mammals. This also applies to the symmetrically related MR motifs. Just as before, 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 S9**).

Considering all four types of repeats together, we noticed that repeats with both half-sites on the same strand (DR and MR) or both half-sites on the opposite strand (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.

To gain hints as to causality, we also tested if longer repeats, allowing more stable base pairing, show stronger correlations with MLS, but to our surprise, we noticed the opposite (**Figure S11A-D**). Finally, it has been suggested that GC rich repeats are selected against (**Shamanskiy et al. 2019**). However, even restricting our analysis to repeats that have a GC content of >45% we did not find an inverse correlation with MLS (data not shown).



	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 both 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 explain an apparent correlation with MLS. Data in (A-C) is for 11 bp long repeats. Pearson's R 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 GC content, GC skew, AT skew and number of effective codons. Significant correlations from the raw and adjusted model are bolded and 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 non D-loop G-quadruplex motifs produce

similar results (**Lombardi and Londoño-Vallejo 2020; Fig. S12**). 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 (as discussed before), we found that the two tools made very different predictions regarding triplex counts across species (**Fig. S13**).

The limited agreement between the 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 S10**) and that breakpoints were preferably found in highly unstable regions with multiple such motifs (**Fig. S14**).

Next, we turned again to mtDNA to test whether mutagenic motifs are negatively correlated with lifespan as predicted by the theory of resistant biomolecules. To the contrary, we found that G-quadruplex motifs were positively correlated with MLS (**Fig. S15**), 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 relationship between G-quadruplex motifs and MLS (**Table 1**).

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. S11E**).

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. S16B, D**). Phylogenetic orders with higher numbers of triplex motifs in 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 S11**).

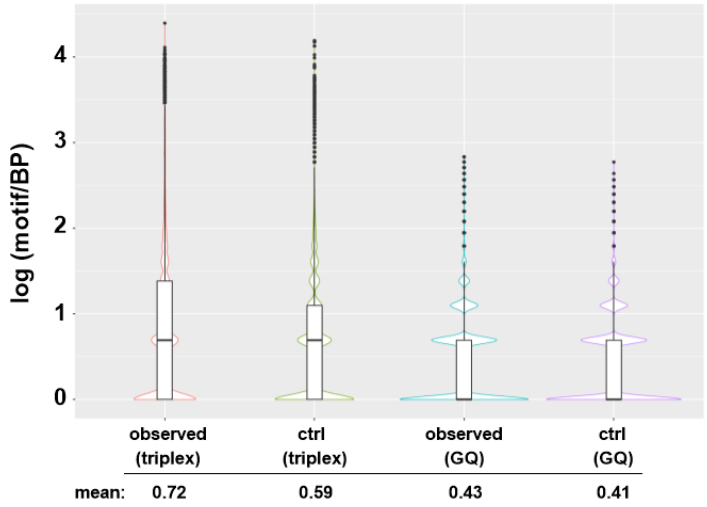


Figure 6

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.

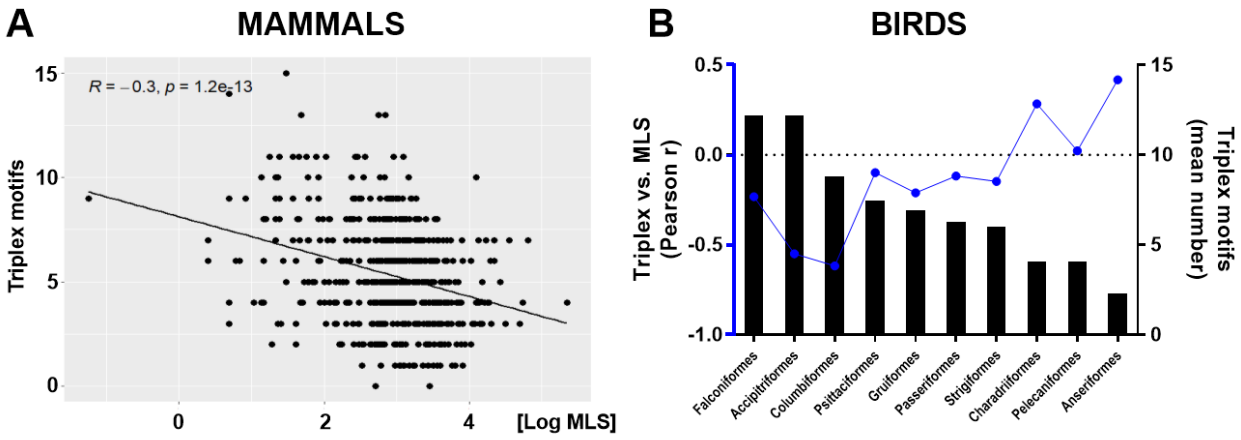


Figure 7

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 S9**), 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

Motifs and mtDNA instability

Our results were in line with the consensus that DR motifs are a major driver of deletion mutagenesis (**Persson et al. 2019**). However, we also 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**). Regardless of the postulated mechanisms, we could not establish if the association is causal because DR and MR motifs correlate with each other (**Shamanskiy et al. 2019; Fig. 5B**) and with the base composition of the mtDNA (**Fig. 5C**). After we removed MR motifs also found in the set of DR motifs, only few deletion breakpoints were associated with MR motifs (**Fig. 2A, B; Fig. S6B, S7B**).

Our findings (**Fig. 2A, D**) also agree with earlier studies that showed no relationship between breakpoints and IR motifs (**Dong et al. 2014**). Nevertheless, we cannot exclude that IRs contribute to the formation of some deletion types (**Damas et al. 2012**, **Mikhailova et al. 2019**) and we also did not address their influence on inversions (**Yang et al. 2013, Tremblay-Belzile et al. 2015**). ER motifs were not associated with breakpoints and behaved like IR motifs in our testing (**Fig. 2A, D; Fig. 5B**).

In addition, we identified novel roles for triplex motifs and their interaction with G-quadruplex motifs. 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.

We found that triplex motifs are indeed associated with mtDNA deletions (**Fig. 3A, B**). One limitation of our data is that many triplex motifs we identified were close to, or overlapped, G-quadruplex motifs. However, even excluding overlapping G-quadruplexes, triplex motifs were enriched around breakpoints (**Fig. 3A, B; Fig. 4D**). What is more, this exclusion may be overly conservative. A high G-content is known to both promote G-quadruplex formation and to stabilize triplex DNA (**Buske et al. 2011**, **Kaufmann et al. 2019**). Although G-quadruplexes may compete with the formation of triplex DNA, this is not always so (**Sole et al. 2017, del Mundo et al. 2017**), especially in mtDNA where overlapping triplex and G-quadruplex motifs can be located on opposite strands. Consistent with this idea, we found that G-quadruplex and triplex motifs in proximity to each other showed a stronger association with deletion breakpoints than isolated motifs, suggesting that their effects are additive (**Fig. S9**).

Furthermore, 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 could allow more time for the formation of triplex DNA during replication, thereby explaining this finding. Hence, the moderate enrichment after pooling all our datasets (n=4074 deletions; **Fig. 4D**) may be a conservative estimate for some subgroups.

Why did an earlier study (**Oliveira et al. 2013**) fail to uncover an association between triplex motifs and mtDNA deletions? One reason may be that the nBMST tool used in that study, is not very sensitive to detect bona fide triplexes (**Hon et al. 2013**), whereas the triplex package used here, not only detected

more mitochondrial triplex motifs than nBMST, but also produced data that agreed with prior studies of nuclear triplex motifs (**Table S10**).

Motifs and lifespan: little support for the theory of resistant biomolecules

Lakshmanan et al. (2015) showed that DR motifs are correlated with mtDNA compositional features rather than MLS and we extend this finding to all four classes of perfect repeats, i.e. DR, MR, IR and ER motifs. At the same time, our data is consistent with **Yang et al. (2013)** who found an inverse relationship between IR motifs and MLS without correction for mtDNA composition. We found a similar correlation which 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. S11C**). Secondly, IR motifs with half-sites close together failed to show a stronger relationship with MLS, although these should form secondary structures more easily.

When considering all four motif classes together (**Fig. 5B, C**) the importance of mtDNA compositional bias becomes even more obvious. 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.

Similarly, for G-quadruplex motifs we showed that mtDNA composition also accounts for the correlation with lifespan (**Table 1**). Only in the case of triplex motifs, were we unable to explain the findings by base composition effects. Indeed, we showed that intramolecular triplex motifs detected by the triplex package are inversely correlated with mammalian lifespan after correcting for body mass, composition and phylogeny (**Table 1**).

Whenever data was available, we also tested if the theory of resistant biomolecules holds up in non-mammalian species. While it is not clear a priori which motifs can be expected to be mutagenic in species with different mtDNA biology, we did find preliminary evidence that triplex motifs may be inversely correlated with MLS in bird orders with high numbers of triplex motifs in their mtDNA (**Fig. 7B**). Since we observed the same trend in mammals, this suggests there may be a threshold effect. 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 seen in the nuclear context, **Fig. S14**).

Finally, our data also highlights the importance of intramolecular (intrastrand) triplex forming sequences, because intermolecular triplex motifs detected by Triplexator were not correlated with lifespan nor with intramolecular motifs found by the triplex package across species (**Fig. S13**).

Summary

Despite promising data from earlier comparative studies (**Lakshmanan et al. 2012, Yang et al. 2013**), we found only limited evidence for the theory of resistant biomolecules which implies an association between mutagenic motifs and lifespan. Perhaps the mitochondrial DNA is too streamlined to allow for such profound changes in sequence. Given thousands of short, potentially mutagenic motifs (**Fig. S10**) that would need to change, 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 summary, intramolecular triplex motifs, which we found to be associated with both mitochondrial and nuclear DNA instability, were the only motif class that was independently associated with MLS, especially in orders with high levels of triplex motifs in their mtDNA like rodents. However, we cannot rule out that some unmeasured compositional bias or feature of mtDNA is responsible for this correlation.

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., Li, Y., Magalhães, J. P. de, Ruíz-Camacho, M., & Pérez-Claros, J. A. (2011). Mitochondrially encoded methionine is inversely related to longevity in mammals. *Aging Cell*, *10*(2), 198–207. https://doi.org/10.1111/j.1474-9726.2010.00657.x

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 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

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

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

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

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

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

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

Hjelm, B. E., Rollins, B., Morgan, L., Sequeira, A., Mamdani, F., Pereira, F., Damas, J., Webb, M. G., Weber, M. D., Schatzberg, A. F., Barchas, J. D., Lee, F. S., Akil, H., Watson, S. J., Myers, R. M., Chao, E. C., Kimonis, V., Thompson, P. M., Bunney, W. E., & Vawter, M. P. (2019). Splice-Break: Exploiting an RNA-seq splice

junction algorithm to discover mitochondrial DNA deletion breakpoints and analyses of psychiatric disorders. *Nucleic Acids Research*, *47*(10), e59–e59. https://doi.org/10.1093/nar/gkz164

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

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

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

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

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

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

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

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

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

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

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

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

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

Niedernhofer, L. J., Gurkar, A. U., Wang, Y., Vijg, J., Hoeijmakers, J. H. J., & Robbins, P. D. (2018). Nuclear Genomic Instability and Aging. *Annual Review of Biochemistry*, *87*(1), 295–322. https://doi.org/10.1146/annurev-biochem-062917-012239

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

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

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

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

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

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