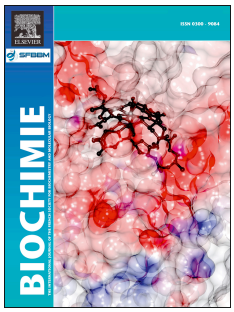


Conservation and over-representation of G-quadruplex sequences in regulatory regions of mitochondrial DNA across distinct taxonomic sub-groups

Natália Bohálová, Michaela Dobrovolná, Václav Brázda, Stefan Bidula

PII: S0300-9084(21)00294-7

DOI: https://doi.org/10.1016/j.biochi.2021.12.006

Reference: BIOCHI 6227

To appear in: *Biochimie*

Received Date: 8 October 2021

Revised Date: 22 November 2021

Accepted Date: 14 December 2021

Please cite this article as: Natá. Bohálová, M. Dobrovolná, Vá. Brázda, S. Bidula, Conservation and over-representation of G-quadruplex sequences in regulatory regions of mitochondrial DNA across distinct taxonomic sub-groups, *Biochimie* (2022), doi: https://doi.org/10.1016/j.biochi.2021.12.006.

Author contributions

Natália Bohálová: Software, Validation, Formal analysis, Investigation, Writing – Review and Editing, Visualization

Michaela Dobrovolná: Formal analysis, Investigation, Resources, Writing – Review and Editing

Václav Brázda: Conceptualization, Resources, Writing - Review and Editing, Supervision, Project Administration, Funding acquisition

Stefan Bidula: Conceptualization, Formal analysis, Investigation, Writing - Original Draft, Visualization, Supervision, Project Administration

Abstract

G-quadruplexes have important regulatory roles in the nuclear genome but their distribution and potential roles in mitochondrial DNA (mtDNA) are poorly understood. We analysed 11883 mtDNA sequences from 18 taxonomic sub-groups and identified their frequency and location within mtDNA. Large differences in both the frequency and number of putative quadruplex-forming sequences (PQS) were observed amongst all the organisms and PQS frequency was negatively correlated with an increase in evolutionary age. PQS were over-represented in the 3'UTRs, D-loops, replication origins, and stem loops, indicating regulatory roles for quadruplexes in mtDNA. Variations of the G-quadruplex-forming sequence in the conserved sequence block II (CSBII) region of the human D-loop were conserved amongst other mammals, amphibians, birds, reptiles, and fishes. This D-loop PQS was conserved in the duplicated control regions of some birds and reptiles, indicating its importance to mitochondrial function. The guanine tracts in these PQS also displayed significant length heterogeneity and the length of these guanine tracts were generally longest in bird mtDNA. This information provides further insights into how G4s may contribute to the regulation and function of mtDNA and acts as a database of information for future studies investigating mitochondrial G4s in organisms other than humans.

Key words: mitochondria, genome, G-quadruplex, evolution, D-loop

Conservation and over-representation of G-quadruplex sequences in regulatory regions of mitochondrial DNA across distinct taxonomic sub-groups

Running title – G-quadruplexes in mitochondria

Natália Bohálová,[a, b] Michaela Dobrovolná,[a, c] Václav Brázda,[a, c] Stefan Bidula[d]

[a] Institute of Biophysics of the Czech Academy of Sciences, Brno, Czech Republic

[b] Department of Experimental Biology, Faculty of Science, Masaryk University, Brno, Czech Republic

[c] Department of Food Chemistry and Biotechnology, Faculty of Chemistry, Brno University of Technology, Purkyňova 118, 61200 Brno, Czech Republic

[d] School of Biological Sciences, University of East Anglia, Norwich, UK

Corresponding author – Stefan Bidula (s.bidula@uea.ac.uk), University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ

Declarations of interest: None

Abbreviations

G4, G-quadruplex; PQS. Putative quadruplex-forming sequence; mitochondrial DNA, mtDNA; CBSII, conserved sequence block II; RHPS4, 3,11-Difluoro-6,8,13-trimethylquino[4,3,2-*kl*]acridinium methylsulfate; NCBI, National Center for Biotechnology and Information

Abstract

G-quadruplexes have important regulatory roles in the nuclear genome but their distribution and potential roles in mitochondrial DNA (mtDNA) are poorly understood. We analysed 11883 mtDNA sequences from 18 taxonomic sub-groups and identified their frequency and location within mtDNA. Large differences in both the frequency and number of putative quadruplex-forming sequences (PQS) were observed amongst all the organisms and PQS frequency was negatively correlated with an increase in evolutionary age. PQS were over-represented in the 3'UTRs, D-loops, replication origins, and stem loops, indicating regulatory roles for quadruplexes in mtDNA. Variations of the G-quadruplex-forming sequence in the conserved sequence block II (CSBII) region of the human D-loop were conserved amongst other mammals, amphibians, birds, reptiles, and fishes. This D-loop PQS was conserved in the duplicated control regions of some birds and reptiles, indicating its importance to mitochondrial function. The guanine tracts in these PQS also displayed significant length heterogeneity and the length of these guanine tracts were generally longest in bird mtDNA. This information provides further insights into how G4s may contribute to the regulation and function of mtDNA and acts as a database of information for future studies investigating mitochondrial G4s in organisms other than humans.

Key words: mitochondria, genome, G-quadruplex, evolution, D-loop

1. Introduction

Mitochondria are aptly referred to as the 'powerhouse' of the cell, whereby they generate energy for cells in the form of ATP [1]. However, they have now been demonstrated to participate in numerous important and diverse biological processes, including metabolic signalling, bioenergetics, calcium transport, production of reactive oxygen species, and regulation of cell death pathways [2]. Dysfunction of mitochondria, arising from either the acquisition of mutations in mitochondrial DNA (mtDNA) or nuclear DNA, can be catastrophic and can result in various diseases, such as Leigh syndrome, myoclonic epilepsy with ragged red fibres syndrome, and mtDNA depletion syndrome [3]. Thus, mitochondria require strict regulatory processes to ensure normal biological function. There is growing evidence that alternative DNA structures such as cruciforms, left-handed DNA (Z-DNA), R-loops, and quadruplexes play critical regulatory roles in fundamental biological functions, although our understanding of their roles in mtDNA are still in their infancy [4–7].

G-quadruplexes (G4s) are four-stranded secondary structures in nucleic acids that form in guanine-rich regions. G4s form when four guanines associate through Hoogsteen hydrogen bonding to form a G-tetrad [8]. Several G-tetrads then stack on top of one another, linked by mixed-sequence nucleotides, to form the G4 structure itself. Given their localisation throughout the nuclear genome in promoters, untranslated regions, and telomeres, G4s and iMs have been identified to participate in critical regulatory processes, such as transcription, translation, and phase separation of RNA to name but a few [9,10]. However, their roles within mtDNA are poorly understood and practically nothing known about their roles in the mtDNA of organisms other than humans.

Therefore, it was important that we explored mtDNA for the presence of putative quadruplex-forming sequences (PQS) to enhance our understanding of where these structures were located in mtDNA and provide insight into the potential biological roles that G4s may play. To

this end, we have analysed the mtDNA of 11883 genomes from diverse species across 18 taxonomic sub-groups and highlighted the frequency and location of PQS, whilst also identifying conservation of a PQS in a critical regulatory region across taxonomic sub-groups.

2. Methods

2.1 mtDNA sequences

Complete and most recent mtDNA sequences were downloaded from the organelle genome database of the National Center for Biotechnology and Information (NCBI). The genomes were obtained for 11883 organisms from 18 different taxonomic sub-groups. Groups included amphibians (303 genomes), apicomplexans (47 genomes), ascomycetes (379 genomes), basidiomycetes (125 genomes), birds (977 genomes), fishes (3036 genomes), flatworms (153 genomes), green algae (92 genomes), insects (2534 genomes), land plants (259 genomes), mammals (1342 genomes), other (161 genomes), other animals (1760 genomes), other fungi (27 genomes), other plants (13 genomes), other protists (109 genomes), reptiles (382 genomes), and roundworms (184 genomes). Duplicated genomes were omitted from the analysis.

2.2 Data analysis

Genomes were analysed using G4Hunter (http://bioinformatics.ibp.cz) to identify PQS [11]. The parameters for analysis were set at a length of 25 nucleotides and a threshold of 1.2, as these settings have previously been shown to identify experimentally validated quadruplex structures [12]. The figures in the main manuscript body were all produced using these analysis parameters. However, we also analysed the genomes with thresholds between 1.2-1.4, 1.4-1.6, 1.6-1.8, 1.8-2.0, and more than 2.0. The raw data for all the genomes analysed can be found in **Supplementary Table S1** and provides information on the genome name,

NCBI identifier, length, GC genome content, and frequency and number of PQS. The tables of annotated genomic features were downloaded in tandem with the mitochondrial genomes, and we analysed the frequency of PQS within these annotated features (e.g., gene) and within ± 100 base pairs of the features for each genome (**Supplementary Table S2)**. The script used for analysis is publicly available at https://gitlab.com/PatrikKaura/DNA_analyser_IBP.

2.3 Statistical analysis

A cluster dendrogram was constructed in R, using the *pvclust* package [13]. The following values were used as input data: Mean PQS/kbp, Min PQS/kbp, Max PQS/kbp. and Cov % (% of genome covered by PQS). The 'ward.D2' clustering method was used with Euclidean distance and 10,000 bootstrap resampling. This cluster dendrogram can be found in **Supplementary Figure S2**. Correlation was determined by two-tailed Pearson's correlation coefficient. Normality of the data was determined via a Shapiro-Wilk test. Non-parametric Kruskal-Wallis tests with Dunn's multiple comparisons were used to determine significance. All figures and analysis were generated using GraphPad Prism (v 9.1.0).

3. Results and discussion

3.1 Large heterogeneity of PQS frequency in mtDNA

G4s and iMs have been shown to have important regulatory roles throughout nuclear genomes. However, an in-depth analysis of PQS in mitochondrial genomes had not been conducted to date. Using G4Hunter, we investigated the presence and frequency of PQS in 11883 mitochondrial sequences from highly varied species across 18 taxonomic sub-groups.

We utilised the default G4Hunter settings to identify PQS with a length of 25 nt using a threshold of 1.2. Data for all organisms can be found in **Supplementary Table S1** and

Supplementary Figure S1. The PQS frequencies between all groups were found to be significantly different (**Figure** 1). On average, bird mtDNA was found to have the highest PQS frequency (5.62 PQS/kbp), frequency of PQS relative to genome GC content (12.28 PQS/GC%), and highest genome GC content (45.63%; **Figure 1, Table 1**). Conversely, the mtDNA of land plants was found to have the greatest average total number of PQS (607.03 PQS; **Table 1**). Conversely, apicomplexan mtDNA contained the lowest PQS/kbp (0.16 PQS/kbp), PQS/GC% (0.49 PQS/GC%), and total PQS (0.94 PQS), whilst insect mtDNA contained the lowest genome GC content (23.78%; **Figure 1, Table 1**).

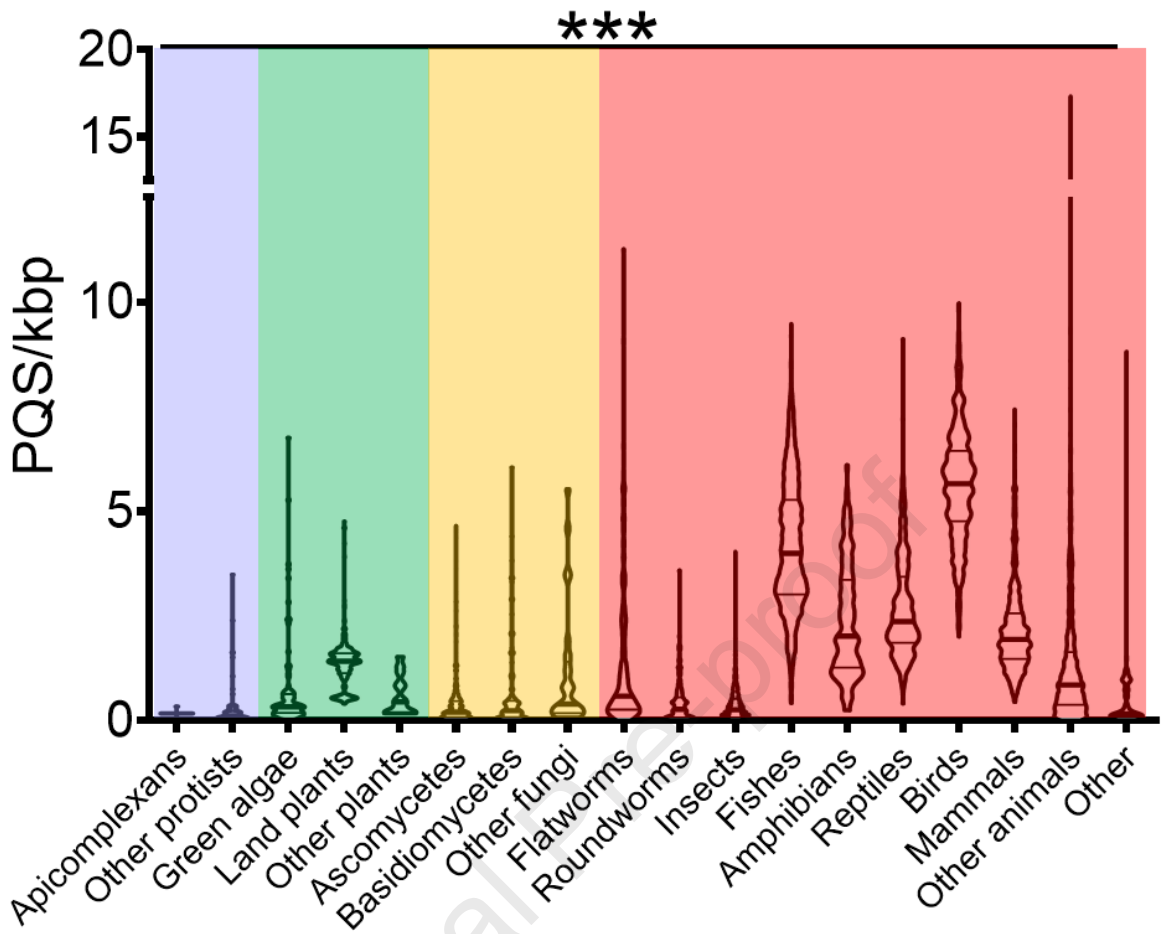
In general, mtDNA PQS frequencies were associated with evolutionary distance. More closely related organisms, such as birds, reptiles, mammals, amphibians, and fishes had the highest PQS frequencies and clustered together with statistical significance, whereas there appeared to be a loss of PQS with an increase in evolutionary distance (**Supplementary Figure S2**).

Interestingly, the frequency of PQS was found to be the inverse of what has previously been observed for inverted repeats in mtDNA [14]. Inverted repeats of six or more nucleotides are a pre-requisite for the formation of cruciforms; alternative DNA structures which have been demonstrated to regulate key biological processes [4,15]. Birds were found to have the lowest frequency of inverted repeats, whilst apicomplexans, fungi, and insects had the highest frequency [14]. This is almost a mirror-image of what we observed for PQS and could be indicative that inverted repeats are more important for the integrity or function of mitochondria in organisms such as apicomplexans, whereas inverted repeats may be detrimental to more complex organisms such as birds. Indeed, inverted repeat-related inversions tend to accumulate in tissues with high-energy metabolism in mammals, which can accelerate ageing and reduce longevity [16]. Birds possess fewer inverted repeats and have much longer lifespans comparative to similarly sized organisms [17]. Our understanding of the roles of

quadruplexes within mtDNA, particularly in organisms other than humans, are still in their infancy. Although, one could speculate that the roles of G4s in mtDNA could be organism specific and dependent upon the prevalence of PQS within the genome.

Table 1. An overview of the PQS frequencies, PQS number, and genome GC% of mitochondrial genomes in the study

Group name	Number of genomes	Average PQS/kbp	Lowest frequency (PQS/kbp)	Highest frequency (PQS/kbp)	Average genome GC%	Average PQS/GC%	Average total PQS
Apicomplexans	47	0.16	0.00	0.34	31.62	0.49	0.94
Other protists	109	0.24	0.00	3.49	27.17	0.74	22.65
Green algae	92	0.70	0.03	6.76	37.74	1.58	50.68
Land plants	259	1.38	0.40	4.76	44.10	3.08	607.03
Other plants	13	0.61	0.17	1.52	37.84	1.51	70.15
Ascomycetes	379	0.43	0.00	4.66	26.42	1.62	30.24
Basidiomycetes	125	0.59	0.00	6.05	28.09	1.87	55.35
Other fungi	27	1.14	0.07	5.53	33.97	2.78	87.22
Flatworms	153	1.37	0.00	11.27	31.73	3.76	20.05
Roundworms	184	0.41	0.00	3.59	25.39	1.46	5.89
Insects	2534	0.36	0.00	4.04	23.78	1.41	5.71
Fishes	3036	4.17	0.42	9.49	44.16	9.31	69.61
Amphibians	303	2.38	0.24	6.11	38.87	5.91	42.46
Reptiles	382	2.68	0.40	9.12	40.56	6.49	45.98
Birds	977	5.62	2.00	9.98	45.60	12.28	96.35
Mammals	1342	2.15	0.44	7.44	39.18	5.39	35.71
Other animals	1760	1.32	0.00	17.34	32.70	3.67	22.44
Other	161	0.44	0.00	8.82	31.35	1.24	14.63



159

160 ***Figure 1. There was large heterogeneity in the frequency of PQS in 11883 mtDNA***
161 ***sequences across the taxonomic sub-groups.*** *The frequency of PQS relative to genome*
162 *length expressed as PQS/kbp. The organisms are grouped by Kingdom into the Protista*
163 *(blue), Plantae (green), Fungi (yellow), and Animalia (red). The lines within the body of the*
164 *violin plots represent the median and upper and lower quartiles. The range and distribution of*
165 *sequences are also displayed. Genomes were obtained from the NCBI organelle genome*
166 *database. Frequencies across the groups were found to be significantly different to one*
167 *another (P<0.0001) as determined via a non-parametric Kruskal-Wallis test.*

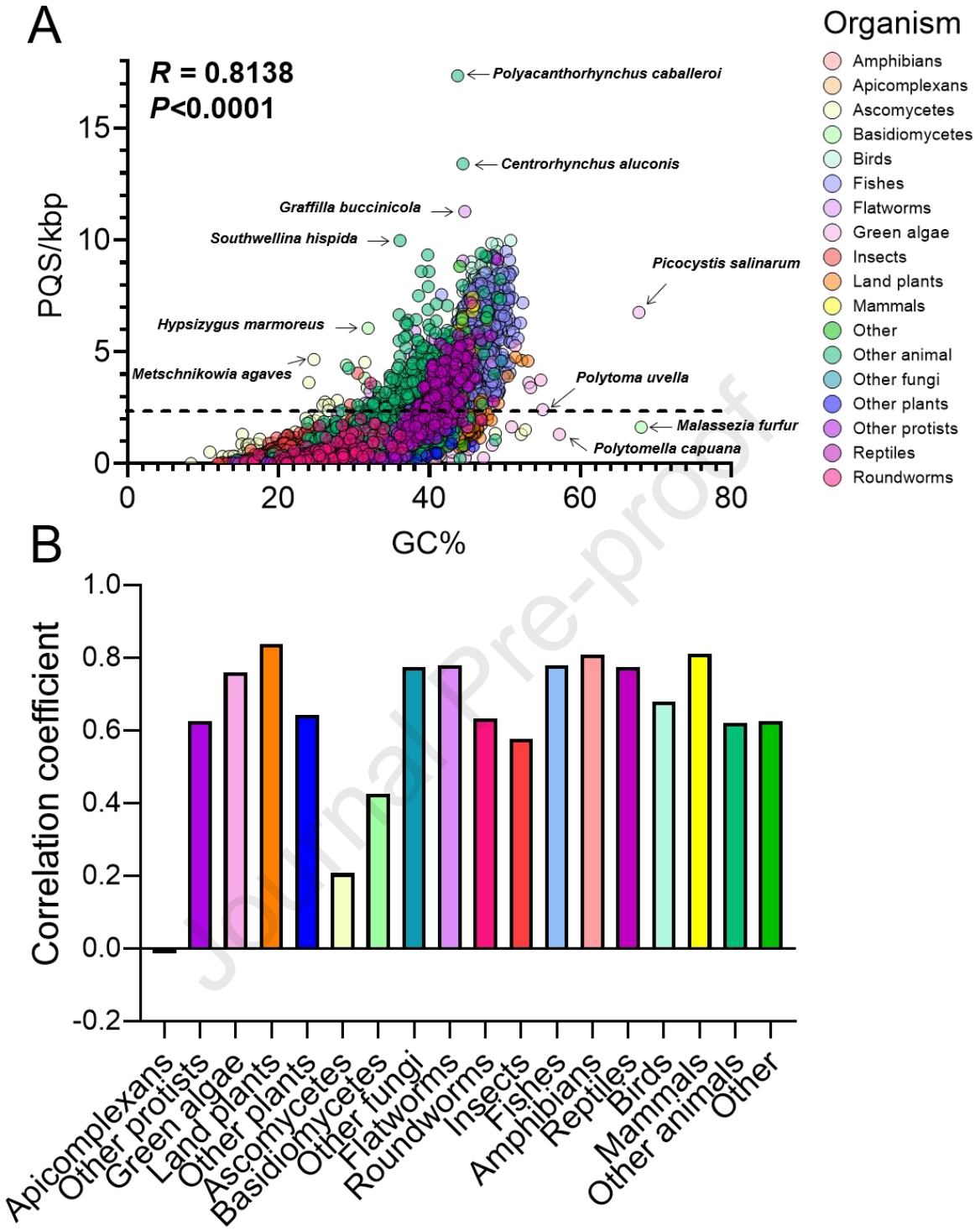
168

169 **3.2 PQS frequency is correlated with mtDNA GC content**

170 As quadruplexes can be found in GC-rich genomic regions, we next explored whether PQS
171 frequency was correlated with genome GC content in mtDNA. In general, PQS frequency was
172 found to be positively correlated with the genome GC content (*R*=0.8138; **Figure 2A**). This
173 positive correlation was observed in all sub-groups except for the apicomplexans (**Figure 2B**;
174 **Supplementary Figure S3**). However, some organisms had higher or lower PQS frequencies

175 relative to their GC content which deviated from the average. Some of these organisms with
176 high PQS frequencies included the parasites *Polyacanthorhynchus caballeroi, Centrorynchus*
177 *aluconis,* and *Southwellina hispida*, the fungus *Metschnikowia agaves*, and flatworm *Graffilla*
178 *buccinicola* (**Figure 2A**). Those with lower PQS frequencies included the fungi *Malassezia*
179 *furfur*, *Candida gigantensis,* and *Candida subhashii*, and the green algae *Polytoma uvella,*
180 *Picocystis salinarum,* and *Polytomella capuana* (**Figure 2A**).

181



182

183 ***Figure 2. The frequency of PQS is positively correlated with genome GC content.*** *(**A**)*
184 *There was a positive correlation between PQS frequency and genome GC content amongst*
185 *all genomes used in the study. (**B**) The correlation coefficients for each group. Some*
186 *organisms which had greater or fewer PQS than expected relative to their GC content are*
187 *highlighted. The dashed line represents the average PQS/kbp for all organisms used in the*
188 *study. Correlation was calculated via two-tailed Pearson's correlation coefficient (P<0.0001).*

189

3.3 Quadruplexes are found in critical regulatory regions in mtDNA

To garner some insight into the potential roles that quadruplexes may play in mtDNA, we endeavoured to identify where in the genome PQS were located. The genomic locations discussed here were categorised based upon the annotation definitions used in the NCBI database and we identified the PQS frequency in regions 100 bp before, within, and 100 bp after the annotated feature.

Examining all the mitochondrial genomes together highlighted that PQS were found at greater than average frequencies in key regulatory regions of mtDNA, including the 3'UTRs, D-loops, replication origins, and stem loops (**Figure 3**). Conversely, PQS were found to be depleted in the exons, introns, messenger RNA, and transfer RNA (**Figure 3**). However, when looking at each subgroup individually, the picture is more complex and the distribution of PQS in genomic features amongst the subgroups was highly varied (**Supplementary Table S2 and Supplementary Figure S3**). Other regions of interest where PQS could be found at a high frequency included the repeat regions (ascomycetes, basidiomycetes, flatworms, green algae, other animals, other fungi), and non-coding RNA (ascomycetes, basidiomycetes, land plants; **Supplementary Figure S3**).

MtDNA G4s have been found with increased prevalence within cancer cells, to be associated with DNA damage, and found in regions linked with the formation of deletion breakpoints in patients with mitochondrial disorders (Falabella *et al.*, 2019; Butler *et al.*, 2020; Dahal *et al.*, 2021). However, growing evidence supports a role for G4s in biologically relevant mitochondrial functions. Falabella and colleagues found that the G4-stabilising compound 3,11-Difluoro-6,8,13-trimethylquino[4,3,2-*kl*]acridinium methylsulfate (RHPS4) could preferentially localise to mitochondria in non-cancerous cells to bind to and stabilise mitochondrial G4s (Falabella *et al.*, 2019). Depending upon the concentration used and the

amount of G4 stabilisation, this compound could either induce DNA damage, modulate replication, or limit transcription. Our observations that PQS are found within regulatory regions such as the replication origins and stem loops, in addition to their localisation immediately before and within the genes themselves, is indicative of a potential regulatory role within mtDNA and is supportive of these experimental observations. Notably, G4-stabilisation by high concentrations of RHPS4 resulted in decreased strand-specific RNA abundance at the D-loop (Falabella et al., 2019). The D-loop is the site of first strand replication and stabilisation of G4s in this region could significantly modulate mitochondrial function. Thus, we explored the mtDNA genomes further to identify potential G4s in the D-loops that might indicate conserved mitochondrial functions in non-human organisms.

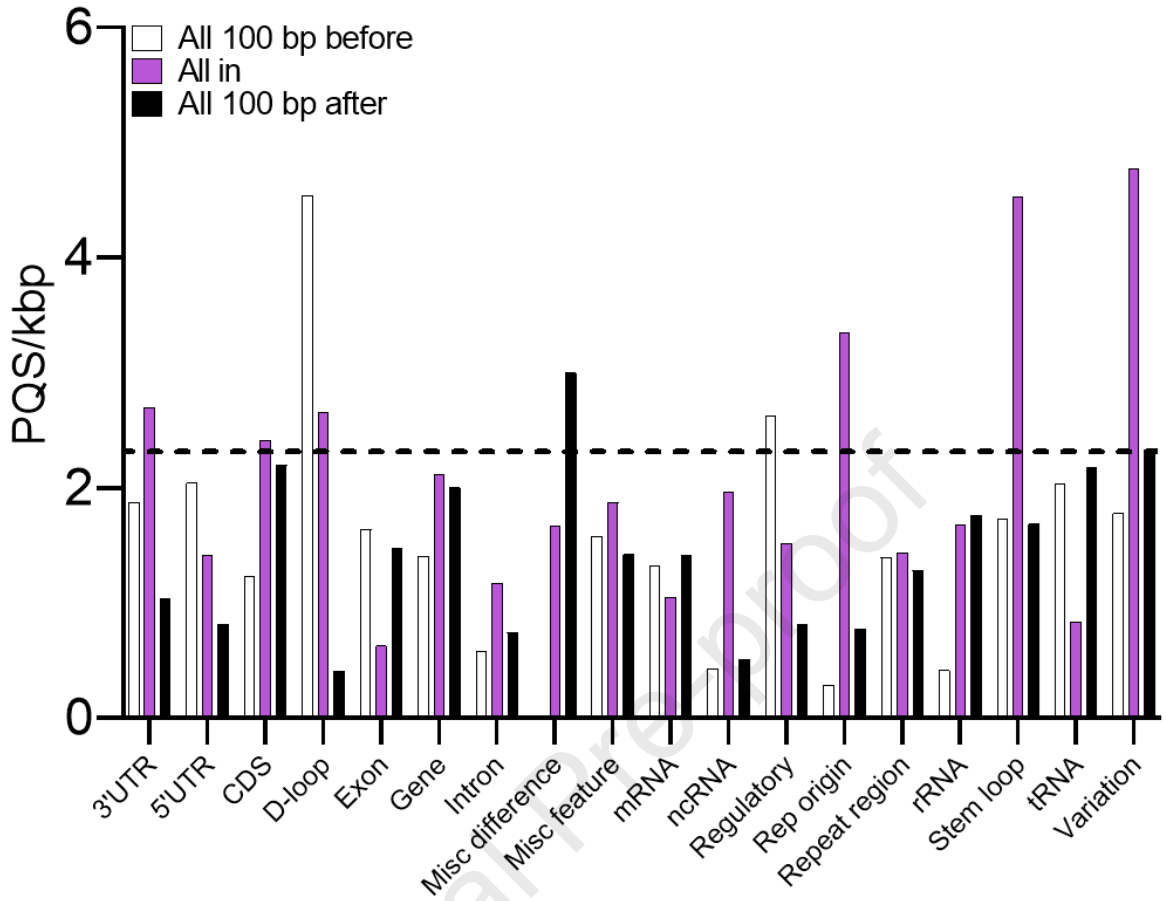


Figure 3. PQS are highly represented in key regulatory regions of mtDNA in all the genomes analysed. *The frequency of PQS (PQS/kbp) within an annotated genomic feature, or within ±100 bp of the feature was quantified. The dashed line represents the average PQS/kbp of all the genomes. Features which were found fewer than 10 times were omitted for clarity. Annotations were obtained from the NCBI database and data represents the average for all 11883 genomes analysed.*

3.4 The D-loop/control region quadruplex sequence displays significant length heterogeneity and variations are conserved in birds, fishes, reptiles, amphibians, and mammals

Several quadruplexes have recently been identified to form in human mtDNA but the best studied quadruplex-forming sequence in human mtDNA (GCGGGGGAGGGGGGGTTTG) falls within the CSBII region of the D-loop [22,23; Figure 4A].

Although this sequence has previously been well characterised in human mtDNA and identified in mice, similar sequences have not yet been fully explored in other organisms. In support of its potential importance in mitochondrial biology, we found that this sequence (or variations of this sequence) could be found highly conserved in D-loops and control regions amongst the mammal (83.7%), amphibian (98.0%), bird (97.0%), reptile (95.8%), and fishes (96.2%) mtDNA sequences analysed (**Supplementary Table S3**). This sequence was noted if it fulfilled two criteria. First, the sequence had to be located within the D-loop region, and second, it had to be formed of a contiguous run of guanines or two guanine tracts separated by a short loop sequence composed of adenine and thymine. These sequences could not be found in all organisms, but this may have been a limitation of genome quality, rather than that these organisms lacked these sequences. However, equivalent sequences seemed to be practically absent in all other subgroups.

This G4-forming sequence was also found in the duplicated control regions of some birds and reptiles, which suggests this G4 may have been evolutionary favoured and likely to play important roles in mtDNA (**Supplementary Table S3**). Duplicated control regions, particularly in parrots, have been shown to provide a selective advantage due to more efficient initiation of replication or transcription and more replicating genomes per organelle ([24]. Consequently, birds with replicated genomes live longer, have larger body masses, and are predisposed to a more active flight. Again, when considering the potential detrimental effects that G4 stabilisation may have to mitochondria and that this region is prone to mutation, the exact reason why sequences with potential to form these structures are so strongly conserved, or even favoured, is unknown [25]. A potential explanation may be linked with the interaction between the G4 in CSBII and RNA which forms an R-loop stabilising hybrid G4 [23]. The mitochondrial R-loop plays critical roles in the replication, organisation, and expression of mtDNA and compromising R-loop formation can result in mtDNA aggregation and disease [6,26]. Therefore, G4s may also be a key participant in mtDNA and segregation through the

stabilisation of R-loops. The presence of G4s has been shown to result in premature transcription of POLRMT but has been suggested to form R-loop structures which provide free 3' ends to prime subsequent DNA synthesis [27–29]. Recent evidence also indicates that the G4-stabilised R-loop leads to increased transcription through a mechanism involving successive rounds of R-loop formation [30].

Significant heterogeneity in the number of guanines in both the first and second G tracts of this D-loop sequence were observed amongst all organisms and we noted at least 106 different G tract length combinations (**Supplementary Table S4**). In general, the second G tract was found to be longer than the first, as seen previously, and the most common combination observed throughout all organisms was 6 guanines in each tract. However, the length of the first G-tract was not always shorter, as has been observed for humans [31]. Moreover, there was significant length heterogeneity in this region with G tract lengths ranging from 2 to 21 and from 3 to 22 guanine residues in the first and second G tracts, respectively (**Supplementary Table S4**). There were also large differences found in the linking sequence, but the most frequently observed linking sequences between the G tracts were A, TA, or TTA (**Supplementary Table S3**).

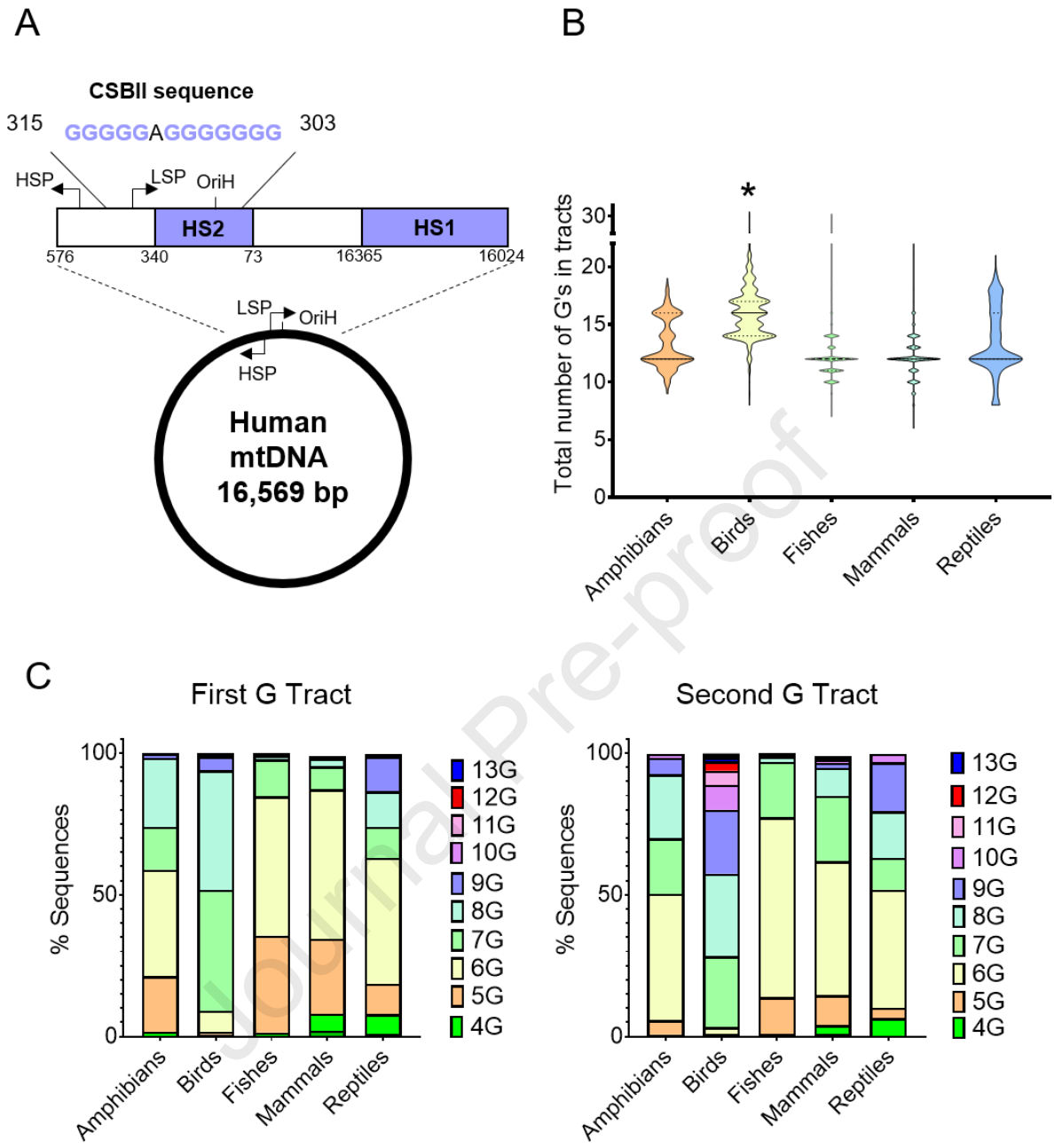


Figure 4. A D-loop quadruplex sequence in humans is conserved in amphibians, birds, fishes, reptiles, and other mammals. *(**A**) The G4-forming sequence between nucleotides 315 and 303 in the CSBII region of the D-loop in human mtDNA. The control region is indicated by dashed lines and the locations of the light and heavy strand promoters, the origin of H-strand replication, region of hypervariable sequence 1 and 2 (HS1 and HS2) are indicated. Adapted from Tan et al. 2006. (**B**) The total number of guanines in the first and second G-tracts combined for each group. The lines within the violin body represent the median (solid) and upper and lower quartiles (dashed). Birds had significantly more guanines in these tracts compared to the other groups (P<0.0001) as determined by a Kruskal-Wallis test with Dunn's multiple comparisons. (**C**) The percentage of organisms in each group with first or second G-tract lengths of between 4 and 13 guanines. G-tract lengths outside of these ranges were less common and these were omitted to focus on the predominant groups.*

Birds not only have the greatest frequency of PQS within their mtDNA, on average they also have significantly more guanines in these D loop G tracts combined (**Figure 4B**). When comparing the G tract lengths in detail, the most frequent G tract length observed in both the first and second G tracts of mammals, fishes, amphibians, and reptiles was six (**Figure 4C**). However, in birds, the most frequent G tract lengths were longer and were found to be 7 and 8 for the first and second G tracts, respectively (**Figure 4C**). Length heterogeneity of the G-tracts in the D-loop/control region has been found to be associated with the amount of transcription termination, with longer G-tracts resulting in increased termination [31]. Interestingly, increased length of these G-tracts is found favoured in cells with elevated growth characteristics and it may be that increased transcription termination is associated with higher levels of mtDNA replication [31].

4. Conclusions

Taken together, the conservation and prevalence of quadruplex-forming sequences in mtDNA, and the D-loop is indicative of potential key regulatory roles for quadruplexes within mtDNA. This study provides an in-depth overview of the similarities and differences between mtDNA in highly diverse organisms and an insight into the importance of quadruplexes in their genomes. Longer G-tract lengths and presence of G4-forming sequences in duplicated control regions could be evolutionarily favoured in organisms which require higher levels of mtDNA replication, such as birds which require increased replication of mitochondria to fulfil the energy requirements necessary for flight, for example. However, we still have much to discover, and the roles of quadruplexes in mtDNA require much greater examination and experimental validation.

Acknowledgements

This work was supported by The Czech Science Foundation (18-15548S).

Conflict of interest

The authors disclose no conflicts of interest.

Author contributions

S.B. and V. conceived the study, N.B., M.D., and S.B. collected the data. N.B., M.D., and S.B. analysed the data. N.B., V.B., and S.B. wrote the manuscript, V.B., acquired the funding. All authors have approved the final article.

References

[1] J.R. Friedman, J. Nunnari, Mitochondrial form and function., Nature. 505 (2014) 335–343. https://doi.org/10.1038/nature12985.

[2] E. Murphy, H. Ardehali, R.S. Balaban, F. DiLisa, G.W. 2nd Dorn, R.N. Kitsis, K. Otsu, P. Ping, R. Rizzuto, M.N. Sack, D. Wallace, R.J. Youle, Mitochondrial Function, Biology, and Role in Disease: A Scientific Statement From the American Heart Association., Circ. Res. 118 (2016) 1960–1991. https://doi.org/10.1161/RES.0000000000000104.

[3] G.S. Gorman, P.F. Chinnery, S. DiMauro, M. Hirano, Y. Koga, R. McFarland, A. Suomalainen, D.R. Thorburn, M. Zeviani, D.M. Turnbull, Mitochondrial diseases., Nat. Rev. Dis. Prim. 2 (2016) 16080. https://doi.org/10.1038/nrdp.2016.80.

[4] V. Brázda, R.C. Laister, E.B. Jagelská, C. Arrowsmith, Cruciform structures are a common DNA feature important for regulating biological processes., BMC Mol. Biol. 12 (2011) 33. https://doi.org/10.1186/1471-2199-12-33.

[5] A. Herbert, Z-DNA and Z-RNA in human disease., Commun. Biol. 2 (2019) 7. https://doi.org/10.1038/s42003-018-0237-x.

[6] J.M. Santos-Pereira, A. Aguilera, R loops: new modulators of genome dynamics and function., Nat. Rev. Genet. 16 (2015) 583–597. https://doi.org/10.1038/nrg3961.

[7] J. Robinson, F. Raguseo, S.P. Nuccio, D. Liano, M. Di Antonio, DNA G-quadruplex structures: more than simple roadblocks to transcription?, Nucleic Acids Res. (2021). https://doi.org/10.1093/nar/gkab609.

[8] S. Burge, G.N. Parkinson, P. Hazel, A.K. Todd, S. Neidle, Quadruplex DNA: sequence, topology and structure., Nucleic Acids Res. 34 (2006) 5402–5415. https://doi.org/10.1093/nar/gkl655.

[9] S. Balasubramanian, L.H. Hurley, S. Neidle, Targeting G-quadruplexes in gene

promoters: a novel anticancer strategy?, Nat. Rev. Drug Discov. 10 (2011) 261–275. https://doi.org/10.1038/nrd3428.

[10] Y. Zhang, M. Yang, S. Duncan, X. Yang, M.A.S. Abdelhamid, L. Huang, H. Zhang, P.N. Benfey, Z.A.E. Waller, Y. Ding, G-quadruplex structures trigger RNA phase separation., Nucleic Acids Res. 47 (2019) 11746–11754. https://doi.org/10.1093/nar/gkz978.

[11] V. Brázda, J. Kolomazník, J. Lýsek, M. Bartas, M. Fojta, J. Šťastný, J.-L. Mergny, G4Hunter web application: a web server for G-quadruplex prediction., Bioinformatics. 35 (2019) 3493–3495. https://doi.org/10.1093/bioinformatics/btz087.

[12] A. Bedrat, L. Lacroix, J.-L. Mergny, Re-evaluation of G-quadruplex propensity with G4Hunter., Nucleic Acids Res. 44 (2016) 1746–1759. https://doi.org/10.1093/nar/gkw006.

[13] R. Suzuki, H. Shimodaira, Pvclust: an R package for assessing the uncertainty in hierarchical clustering., Bioinformatics. 22 (2006) 1540–1542. https://doi.org/10.1093/bioinformatics/btl117.

[14] J. Cechová, J. Lýsek, M. Bartas, V. Brázda, Complex analyses of inverted repeats in mitochondrial genomes revealed their importance and variability., Bioinformatics. 34 (2018) 1081–1085. https://doi.org/10.1093/bioinformatics/btx729.

[15] A.L. Mikheikin, A.Y. Lushnikov, Y.L. Lyubchenko, Effect of DNA supercoiling on the geometry of holliday junctions., Biochemistry. 45 (2006) 12998–13006. https://doi.org/10.1021/bi061002k.

[16] J.-N. Yang, A. Seluanov, V. Gorbunova, Mitochondrial inverted repeats strongly correlate with lifespan: mtDNA inversions and aging., PLoS One. 8 (2013) e73318. https://doi.org/10.1371/journal.pone.0073318.

[17] I. Skujina, R. McMahon, V.P. Lenis, G. V Gkoutos, M. Hegarty, Duplication of the

399 mitochondrial control region is associated with increased longevity in birds., Aging
400 (Albany. NY). 8 (2016) 1781–1789. https://doi.org/10.18632/aging.101012.

401 [18] M. Falabella, R.J. Fernandez, F.B. Johnson, B.A. Kaufman, Potential Roles for G-
402 Quadruplexes in Mitochondria., Curr. Med. Chem. 26 (2019) 2918–2932.
403 https://doi.org/10.2174/0929867325666180228165527.

404 [19] T.J. Butler, K.N. Estep, J.A. Sommers, R.W. Maul, A.Z. Moore, S. Bandinelli, F.
405 Cucca, M.A. Tuke, A.R. Wood, S.K. Bharti, D.F. Bogenhagen, E. Yakubovskaya, M.
406 Garcia-Diaz, T.A. Guilliam, A.K. Byrd, K.D. Raney, A.J. Doherty, L. Ferrucci, D.
407 Schlessinger, J. Ding, R.M. Brosh, Mitochondrial genetic variation is enriched in G-
408 quadruplex regions that stall DNA synthesis in vitro., Hum. Mol. Genet. 29 (2020)
409 1292–1309. https://doi.org/10.1093/hmg/ddaa043.

410 [20] S. Dahal, H. Siddiqua, V.K. Katapadi, D. Iyer, S.C. Raghavan, Characterization of G4
411 DNA formation in mitochondrial DNA and their potential role in mitochondrial genome
412 instability., FEBS J. (2021). https://doi.org/10.1111/febs.16113.

413 [21] M. Falabella, J.E. Kolesar, C. Wallace, D. de Jesus, L. Sun, Y. V Taguchi, C. Wang,
414 T. Wang, I.M. Xiang, J.K. Alder, R. Maheshan, W. Horne, J. Turek-Herman, P.J.
415 Pagano, C.M. St Croix, N. Sondheimer, L.A. Yatsunyk, F.B. Johnson, B.A. Kaufman,
416 G-quadruplex dynamics contribute to regulation of mitochondrial gene expression.,
417 Sci. Rep. 9 (2019) 5605. https://doi.org/10.1038/s41598-019-41464-y.

418 [22] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell,
419 Reanalysis and revision of the Cambridge reference sequence for human
420 mitochondrial DNA., Nat. Genet. 23 (1999) 147. https://doi.org/10.1038/13779.

421 [23] P.H. Wanrooij, J.P. Uhler, Y. Shi, F. Westerlund, M. Falkenberg, C.M. Gustafsson, A
422 hybrid G-quadruplex structure formed between RNA and DNA explains the
423 extraordinary stability of the mitochondrial R-loop., Nucleic Acids Res. 40 (2012)
424 10334–10344. https://doi.org/10.1093/nar/gks802.

[24] A.D. Urantówka, A. Kroczak, T. Silva, R.Z. Padrón, N.F. Gallardo, J. Blanch, B. Blanch, P. Mackiewicz, New Insight into Parrots' Mitogenomes Indicates That Their Ancestor Contained a Duplicated Region., Mol. Biol. Evol. 35 (2018) 2989–3009. https://doi.org/10.1093/molbev/msy189.

[25] M. Stoneking, Hypervariable sites in the mtDNA control region are mutational hotspots., Am. J. Hum. Genet. 67 (2000) 1029–1032. https://doi.org/10.1086/303092.

[26] G. Akman, R. Desai, L.J. Bailey, T. Yasukawa, I. Dalla Rosa, R. Durigon, J.B. Holmes, C.F. Moss, M. Mennuni, H. Houlden, R.J. Crouch, M.G. Hanna, R.D.S. Pitceathly, A. Spinazzola, I.J. Holt, Pathological ribonuclease H1 causes R-loop depletion and aberrant DNA segregation in mitochondria., Proc. Natl. Acad. Sci. U. S. A. 113 (2016) E4276-85. https://doi.org/10.1073/pnas.1600537113.

[27] T.A. Brown, A.N. Tkachuk, D.A. Clayton, Native R-loops persist throughout the mouse mitochondrial DNA genome., J. Biol. Chem. 283 (2008) 36743–36751. https://doi.org/10.1074/jbc.M806174200.

[28] B. Xu, D.A. Clayton, RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA-DNA hybrids serving as primers., EMBO J. 15 (1996) 3135–3143.

[29] D. Kang, K. Miyako, Y. Kai, T. Irie, K. Takeshige, In vivo determination of replication origins of human mitochondrial DNA by ligation-mediated polymerase chain reaction., J. Biol. Chem. 272 (1997) 15275–15279. https://doi.org/10.1074/jbc.272.24.15275.

[30] C.-Y. Lee, C. McNerney, K. Ma, W. Zhao, A. Wang, S. Myong, R-loop induced G-quadruplex in non-template promotes transcription by successive R-loop formation., Nat. Commun. 11 (2020) 3392. https://doi.org/10.1038/s41467-020-17176-7.

[31] B.G. Tan, F.C. Wellesley, N.J. Savery, M.D. Szczelkun, Length heterogeneity at conserved sequence block 2 in human mitochondrial DNA acts as a rheostat for RNA

450 polymerase POLRMT activity., Nucleic Acids Res. 44 (2016) 7817–7829.

451 https://doi.org/10.1093/nar/gkw648.

452

Highlights

- PQS frequency decreases with an increase in evolutionary distance
- PQS are over-represented in the 3'UTR, D-loops, replication origins, and stem loops
- Variation of G4 sequence in the D-loop is conserved across taxonomic sub-groups
- D-loop sequence is conserved in duplicated control regions of birds and reptiles
- Significant length heterogeneity in guanine tracts of the conserved D-loop sequence

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: