Mitochondrion 68 (2023) 72–80

Contents lists available at ScienceDirect

Mitochondrion

journal homepage: www.elsevier.com/locate/mito

A novel, conserved and possibly functional motif "WHWGHTW" in mitochondrial transcription across Bilateria

Merve Nur Aydemir [a,*], Habeş Bilal Aydemir [a], Mahir Budak [b], Birsel Kızıltepe [c], Melissa Şafak Çelebi [b], Ertan Mahir Korkmaz [b], Hasan Hüseyin Başıbüyük [d]

[a] Tokat Gaziosmanpaşa University, Faculty of Science and Letters, Department of Molecular Biology and Genetics, 60250 Tokat, Turkey
[b] Sivas Cumhuriyet University, Faculty of Science, Department of Molecular Biology and Genetics, 58140 Sivas, Turkey
[c] Sivas Cumhuriyet University, Graduate School of Natural and Applied Sciences, Department of Bioinformatics, 58140 Sivas, Turkey
[d] Akdeniz University, Faculty of Health Sciences, Department of Gerontology, 07070 Antalya, Turkey

ARTICLE INFO

Keywords:
Mitochondrial transcription
Signal sequence
Conserved motif
Strand displacement

ABSTRACT

The animal mitogenomes which undergone a reductive evolution has an obvious loss of coding capacity compared to their known closest relatives, but it has not yet been fully investigated why and how the intergenic regions do not encode protein and have no known functions, are stably maintained, replicated, and transmitted by the genome. These relatively small intergenic regions may not be under neutral evolution and they may have functional and/or regulatory roles that have yet to be identified. Here, the distribution pattern, sequence content and location of a novel sequence motif of 'WWWGHTW' were bioinformatically investigated and characterised by constructing a sampling mitogenome dataset of 1889 species from 14 phyla representing the clade of Bilateria. This motif is reverse complementary of the previously described DmTTF binding sequence and found in the *nd4L*-(X) -trnT gene cluster. This cluster commonly exhibits a strand displacement region and an intergenic region among the bilaterian superphylums, particularly in Ecdysozoa. This motif may be accepted as a substrate providing binding sites for the specific interaction with transcription factors because of (i) its reverse complementarity of previously described DmTTF binding sequence, and (ii) the possession of G and T nucleotides in the fourth and sixth positions, (iii) the bias on T and G nucleotides instead of C and A in the degenerated positions. This suggestion is also supported by the presence of a strand displacement region in the *nd4L*- (X) -trnT gene cluster, particularly in Ecdysozoa consisting of the most rearranged mitogenomes among the bilaterian superphylums.

1. Introduction

Mitochondria have dense circular genomes that are able to replicate and transcribe using specialized systems, and these genomes typically encode 37 genes (22 tRNA, 13 PCG and 2 rRNA) involved in aerobic respiration processes in animals. The mitochondrial genome (mitogenome) also encompasses small intergenic regions in addition to a major noncoding regulatory region (Bernt et al., 2013). Despite the fact that the origins for their sequence content are mostly unidentified, these intergenic regions are generally considered to be pseudogene residues formed through gene duplications following random deletions (Akasaki et al., 2006; Boore, 2006; Serb, 2003). The percentage of intergenic regions in the mitogenome can reach substantial levels in some lineages (Aydemir and Korkmaz, 2020; Bandyopadhyay et al., 2008; Park et al., 2011; Wu et al., 2015b). Although the mitogenome which undergone a reductive evolution has an obvious loss of coding capacity compared to their known closest relatives (Andersson and Kurland, 1998), it has not yet been fully understood why and how these intergenic regions do not encode protein and have no known functions, are stably maintained, replicated, and transmitted by the genome. Differences in the presence or absence of some intergenic regions between or within closely related lineages probably suggest that the reductive evolution acting on these regions is an ongoing process (Wu et al., 2015a). However, the appearance of the sequence similarity in some of the intergenic regions across more distant lineages may indicate selective maintenance of these regions (Boore, 1999; Wu et al., 2015b). Consequently, the intergenic regions of mitogenomes may not be freely evolving and they may have functional and/or regulatory roles that have yet to be identified.

* Corresponding author.
E-mail address: ordekmervenur@hotmail.com (M.N. Aydemir).

https://doi.org/10.1016/j.mito.2022.11.004
Received 22 March 2022; Received in revised form 22 August 2022; Accepted 5 November 2022
Available online 16 November 2022

The results of several studies suggest that intergenic regions are either related with recombination (Andre et al., 1992) or they serve as substrates for mutations (Hahn et al., 2003; Lynch et al., 2005). Transcription of these non-coding residues has been confirmed in the *Drosophila* mitogenome (Stewart and Beckenbach, 2009). It was also proposed that multiple intergenic regions are found in mitogenomes because of their important role in mediating rearrangement events (Boore, 1999). However, a recent study reported that intergenic regions that can form stable secondary structures are relatively maintained among conserved genes (Breton et al., 2009). If these intergenic regions are greater than 20 bp and have a strand displacement position, which is involved in transcription polarity, they are potential candidates as control regions to regulate replication and transcription in the mitogenome (Boore, 2006). The functional role of these regions was empirically investigated by Roberti et al. (2003), and a conserved sequence motif comprising seven nucleotides both between the gene clusters of *nd1*-trnS2 and trnE-trnF was suggested as a recognition site for the mitochondrial transcription termination factor (DmTTF) in *Drosophila melanogaster* (Roberti et al., 2006, 2003). The localisation of this recognition site on the strand displacement position makes it an ideal sequence motif for the termination of mitochondrial transcription.

Our preliminary homology investigations on the previously reported hymenopteran (Arthropoda: Insecta) mitogenomes revealed the presence of a novel sequence motif in the conserved intergenic region between *nd4L* and trnT genes of the mitochondrial genome (data is not shown). This shared motif consists of a 'WWWGHTW' sequence in the hymenopteran species, located in the strand displacement position and is a reverse complementary of previously described DmTTF binding sequence (Roberti et al., 2006, 2003). Here, the presence or absence of this novel sequence motif as well as its sequence content and location were bioinformatically investigated by constructing a sampling mitogenome dataset of 1889 species from 14 phyla representing the clade of Bilateria. The consensus data of this novel motif was integrated in the hypothetic Bilaterian phylogeny and then investigated in terms of the nucleotide substitutions and conservation levels.

2. Material and methods

2.1. Sampling and dataset construction

The sampling dataset was constructed by retrieving the mitogenomes of 1889 species representing 14 phyla from the clade of Bilateria from NCBI annotated database (Table S1). We mainly considered the inclusion of the complete and/or partial RefSeq mitogenome data of one representative species from each family in the sampling dataset construction stage. If the data of involved species was not available as RefSeq, the relevant sequences from GenBank (INSDC) were included in the sampling dataset. Taxonomic information and sampling numbers of the species were presented in Table 1 (please also see Table S1 for the details of the sampling). The mitogenome dataset was imported into Geneious R9 (Kearse et al., 2012). Here, regardless of being intergenic, overlapping or sequential sequences, the region between *nd4L* and trnT genes only considering its location as being the strand displacement position were filtered with five nucleotides immediately their upstream and downstream using a custom script (Supplementary material S1). The filtered sequences were clustered on the basis of their class as a taxonomic range, and aligned to obtain a consensus alignment sequence. In order to determine whether this region positions on the transcription polarity changing region, the nucleotide compositions of *nd4L* and trnT genes were calculated, and strand asymmetry was verified by Chargaff' s second parity rule (PR2): AT- and GC- skew values of *nd4L* and trnT genes were calculated by (A-T)/(A + T) and (G-C)/(G + C) formula.

On the other hand, the possible function of this region has been estimated by combining the obtained data from Chip-Seq with the regions displaying protein interaction using the reference mitogenome of *D. melanogaster* accessed from the GTRD database (Gene Transcription Regulation Database) (Yevshin et al., 2019). This strand displacement region between *nd4L* and trnT is located from 9830 to 9843 bases in the reference mitogenome of *D. melanogaster*. The possible (or verified) functions of estimated proteins were investigated using KEGG (Kanehisa, 2000) and UniProt (Magrane and Consortium, 2011) with the organism priority of *D. melanogaster*.

2.2. Integration of consensus sequence into the phylogeny

The conservation pattern, nucleotide content and direction of substitution in the obtained sequence motif for each relevant terminal taxonomic group across the phylogenetic tree (please see Table S2 and Fig. 1) were investigated by integrating them into the Bilaterian phylogeny. The phylogenetic tree of Bilateria was hypothetically constructed by combining the trees from recently reported phylogenomics/phylogeny studies on the clade of Bilateria (Borner et al., 2014; Borner and Burmester, 2017; Halanych, 2004; Laumer et al., 2019; Park et al., 2006; Rehm et al., 2011; Telford et al., 2015).

2.3. Estimation of putative transcription pattern of this intergenic region in Apis model

In order to test whether the intergenic region between *nd4L*- (X) – trnT gene cluster is transcribed in *Apis* model, RNA-seq (*Apis mellifera:* SRX5798518) and mitogenome (*Apis mellifera sahariensis:* NC_035883) data of *Apis* were downloaded from NCBI database. RNA-seq data was filtered into reads containing at least 5 bp polyA sequences in any end of reads. Filtered data were directly mapped to the related mitogenome and mapped data were checked manually in this gene cluster.

3. Results and discussion

In this study, the investigation of the shared strand displacement positions in *nd4L*- (X) - trnT gene cluster across the mitogenomes of bilaterian animals indicated the presence of a novel and conserved motif "WHWGHTW" (Fig. 2, Table S2). In addition to the presence of degenerate nucleotides, its reverse complementarity to a previously described DmTTF binding sequence (Roberti et al., 2003), in conjunction with the similar binding feature with the eukaryotic transcription factors, which bind 6–10 bp recognition sites located near their target genes (Castellanos et al., 2020) may provide evidence supporting the potential functional of this motif. The presence of conserved G and T nucleotides in the fourth and sixth positions, respectively, might also reveal its possible relation to transcription factors. The occurrence of either as GC- (in 13.7 % of the sampling) or GT- (in 74.5 % of the sampling; Supplementary material S2) boxes of the nucleotide G in this motif can be related with the fact that DNA binding transcription factors bind to G-rich elements such as GC- or GT boxes and these elements have also similar roles to enhancers (Hamann et al., 1994). This motif is also biased to T and G nucleotides instead of C and A in degenerated positions. This bias may be due to the presence of selectional pressure toward G and T nucleotides to avoid the detrimental effects of replication mediated deamination mutations on the motif's likely function (Faith and Pollock, 2003). Another reason of the bias on G and T nucleotides (average number of nucleotide compositions are T %: 57,7 %, C %: 2,4, A %: 25,7 %, G %: 14,3 %) (Supplementary material S2), which display 23-fold lower frequency of mispairing with A and C nucleotides (Reyes et al., 1982), may be due to non-canonical Watson-Crick base pairing. On the other hand, the gene cluster which contain "WHWGHTW" motif between *nd4L* and trnT, was exhibited varying nucleotide contents and skewness based on the results of AT- GC- skew calculations (Figure S1). Hereby, the motif can act as a substrate providing binding sites for the specific interaction with transcription factors. Also, the strand displacement region in the *nd4L*- (X) -trnT gene cluster is observed in 45.89 % of all samples (Table 2). However, it has a heterogeneous distribution across the bilaterian superphylums: in 86.17 % of Ecdysozoa,

Table 1
Taxonomic information and representative numbers of the sampling species.

group								species
Bilateria	Ecdysozoa	Arthropoda	Pancrustacea	Hexapoda	Collembola			16
					Protura			6
					Diplura			12
					Insecta	Archaeognatha		2
						Thysanura		3
						Pterygota	Ephemeroptera	10
							Odonata	14
						Neoptera	Hemipteroid	Psocodea	11
								Thysanoptera	2
								Hemiptera	78
							Endopterygota	Megaloptera	2
								Raphidioptera	2
								Neuroptera	16
								Coleoptera	99
								Strepsiptera	2
								Diptera	53
								Mecoptera	6
								Siphonaptera	3
								Trichoptera	8
								Lepidoptera	47
								Hymenoptera	63
							Plecoptera		15
							Phasmida		7
							Orthoptera		30
							Mantophasmatodea		1
							Dictyoptera		22
							Dermaptera		3
							Grylloblattodea		1
				Crustacea	Branchiopoda			8
					Remipedia			1
					Cephalocarida			1
					Copepoda			7
					Ostracoda			1
					Branchiura			1
					Thecostraca			12
					Malacostraca			153
			Myriapoda	Pauropoda				1
				Diplopoda				10
				Chilopoda				7
				Symphyla				2
			Chelicerata	Arachnida				63
				Xiphosura				1
				Pycnogonida				4
		Onychophora						2
		Tardigrada						2
		Nematoda						41
		Priapulida						2
	Lophotrochozoa	Annelida						44
		Bryozoa						8
		Mollusca						196
		Nemertea						11
		Entoprocta						1
		Brachiopoda						5
	Deuterostomia	Echinodermata						37
		Hemichordata						4
		Chordata	Urochordata					9
			Craniata	Hyperotreti				2
				Vertebrata				735
								1905

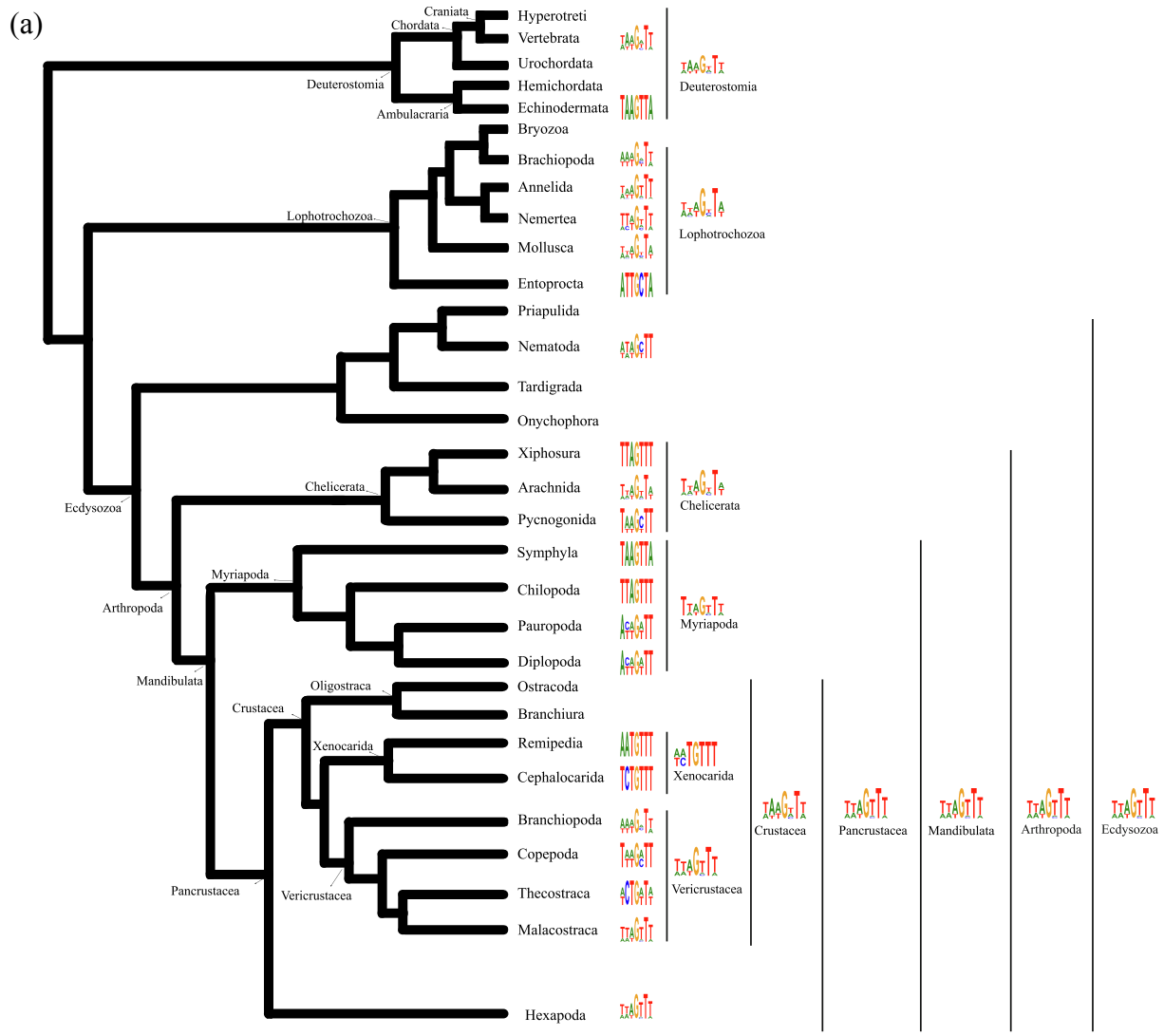

Fig. 1. The conservation pattern and nucleotide content of "WHWGHTW" motif for each terminal taxonomic group in the hypothetical Bilaterian phylogeny (in upper) and detailed representation of terminal Hexapoda group in the hypothetical Bilaterian phylogeny.

in 50.56 % of Lophotrochozoa and in 1.27 % Deuterostomia (Table 2). The common localisation of the motif WHWGHTW in the *nd4L*- trnT gene cluster indicates the consistent maintenance of this gene order. In detail, localisation of the strand displacement region was found between *nd4L* and trnT (*nd4L*-trnT) in 35.15 %, X-trnT in 9.05 % and *nd4L*-X in 1.69 % of all samples. Spatial organisation of the strand displacement region was mostly observed as intergenic (40.81 % of all sampling) as well as sequential and overlapping (2.43 % and 2.64 %, respectively) (Table 2 and Fig. 2). The variability in its spatial organisation in addition to the maintenance even at higher taxonomic ranks may support the argument that this region may provide opportunity for the binding of the related proteins.

The result of searches in the GTRD database suggested that "elav" and "pho (known as pleiohomeotic)" proteins speculatively bind to this region. The potential binding region of the elav protein (UniProt ID number, P16914) is between 9801 and 9846 positions, while the pho protein (UniProt ID number, Q8ST83) binds to the positions between 9838 and 9921. However, please note that neither of these two proteins has evidence of mitochondrial localization; therefore, *in vitro* analyses using mitochondrial targeting sequences or isoforms that have such as targeting signals should be conducted to make stronger inferences. Thus, whether these RNA binding proteins are imported to mitochondria or they have possible functions in mitochondria still remains unanswered and further studies are needed to support our findings.

3.1. Strand displacement region in nd4L-(X)-trnT gene cluster

Strand displacement region in *nd4L*-trnT is mainly found in the superphylum of Ecdysozoa, but there is no any representative in the Deuterostomia superphylum. The strand displacement region as a whole was counted together with five nucleotides from neighbouring genes and it was an average of 18.81 bp in size, ranging between 11 bp and 130 bp. The length of this region as the intergenic spatial organisation ranged from 9 to 128 bp, with an average of 16.86 bp. The overlapping organisation varies from 10 to 63 bp in length, with a 20.35 bp on average. However, the region having the sequential organisation was selected as 10 bp in length with only five nucleotides from each adjacent gene. For *nd4L*-X, this region was displayed similar pattern with the *nd4L*-trnT, occurring mainly in Ecdysozoa, but not in any deuterostomian representative (Table S2). In this gene cluster, the downstream gene (X in here) is highly variable with an unstable pattern. The length of this region as the intergenic organisation was between 11 and 198 bp, with an average of 60.46 bp. The average length was found to be increased in this gene cluster in comparison with *nd4L*-trnT as intergenic

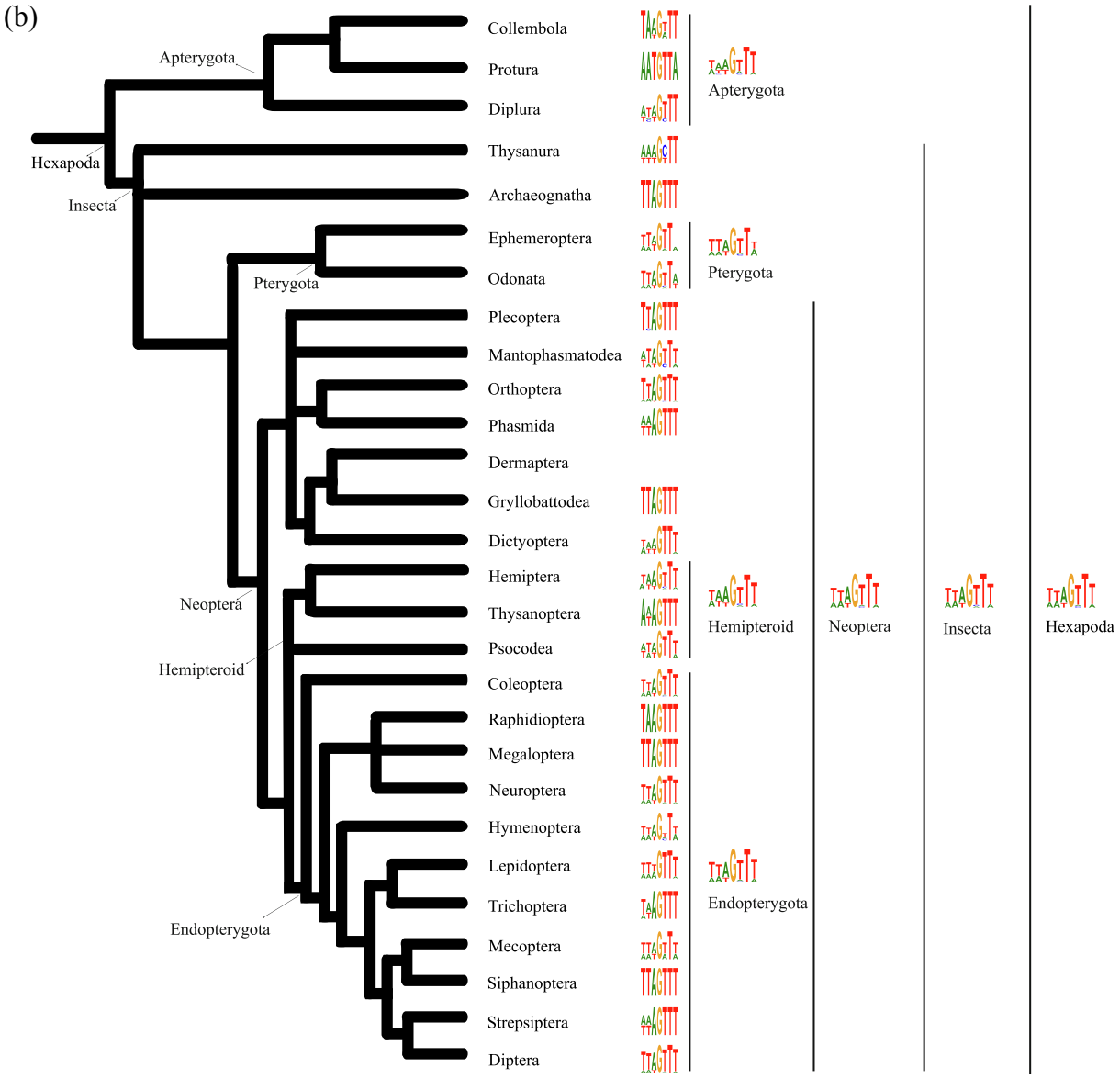

Fig. 1. (*continued*).

spatial organisation. In addition, the overlapping organisation ranged from 10 to 18 bp in length, with a 13.25 bp on average. Alterations in the gene order as *nd4L* - X or X – trnT have resulted in the remarkable size increments in this intergenic region (Table S2 and Fig. 3). This may be related with the rapid change and accumulation of nucleotides in consistent with the observed pattern (TDRL and rearrangement events) in the overlapping genes (Jühling et al., 2012). These findings also point to the necessity of the detailed examination of the intergenic regions in the larger taxonomic groups to explain why rearrangement events and conserved gene order are maintained. A different pattern was observed in the region of X-trnT, particularly in Lophotrochozoa (57.31 %), followed by Ecdysozoa (36.84 %) and Deuterostomia (5.85 %) (Table 2). In this gene cluster, the upstream gene (X in here) is more conserved than that of *nd4L*-X with a bias on being trnS2 (32.16 %), *nd4* (23.98 %) or trnP (26.90 %) genes (Table S2). The trnS2 and *nd4* genes displayed a bias since they are mostly found in the class of Mollusca, but the bias in trnP gene was remarkable. It is observed in almost all different taxonomic groups across samples. The length of this region as the intergenic organisation ranged between nine and 176 bp, with an average of 30.77 bp. Additionally, the overlapping organisation was from 10 to 147 bp in length, with a 28.90 bp on average. The most frequently encountered trnP – trnT order is commonly observed in the superphyla of Lophotrochozoa and Deuterostomia getting a more stable mitogenome gene order.

Another striking finding on the possible functionality of this motif is the neighbouring localisation of trnT gene (Table S2). The motif seems to be evolved with the trnT gene, because of the more conserved in the X-trnT gene cluster than that of the *nd4L*-X, and the most occurrence of the motif in X-trnT gene cluster in terminal groups such as Lophotrochozoa (Table S2). This hypothesis is also supported by the localisation of several nucleotides of the motif within the trnT gene. This trnT gene is found within the most conserved gene clusters across Insecta based on the PanGO (Babbucci et al., 2014), but it has a relatively high nucleotide variability in some of the previously studied groups according to INUC % calculations (Aydemir and Korkmaz, 2020; Zhang et al., 2016). The high variability in the trnT gene might points to the possible functionality of this region, because this tRNA gene may act as an intronic region that attracts mutations to maintain the intergenic region (Aydemir and Korkmaz, 2020). This assumption conflicts with the claim that intergenic regions are largely rearrangement residues and hence are not functional (Boore, 1999; Breton et al., 2009; Dowton et al., 2009).

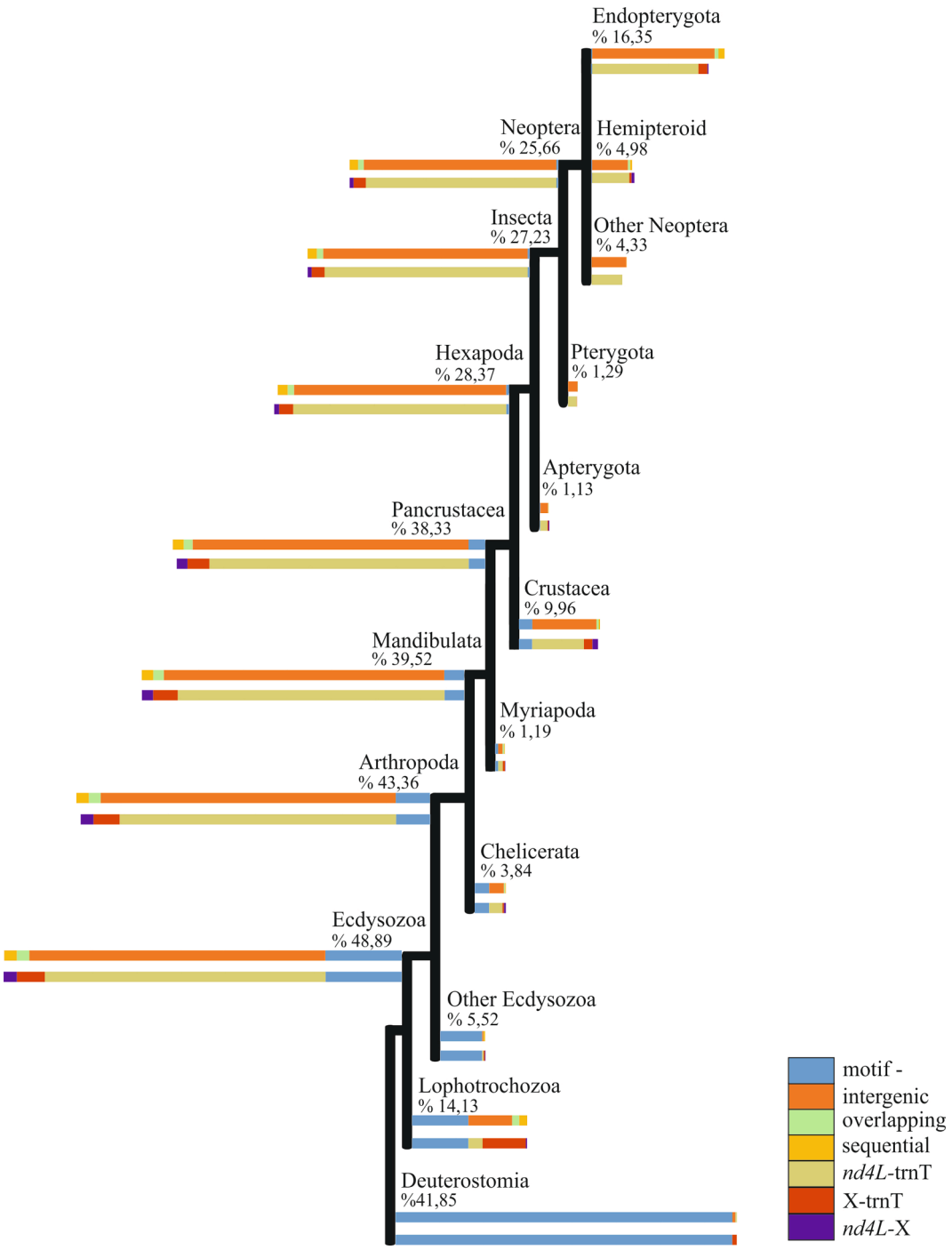

Fig. 2. Presence and general genomic properties of the "WHWGHTW" motif across Bilaterian phylogeny.

3.2. Conserved motif in the strand displacement region in nd4L-(X)-trnT gene cluster

The motif of WHWGHTW is reverse complement of the previously reported motif in *nd1*-trnS2 gene cluster from *D. melanogaster* and displayed both forward and reverse complement orientations. This motif was found almost in all representatives having the strand displacement region with a percentage of 88.66 % in *nd4L*-(X)-trnT gene cluster (Table 2, Fig. 2). 11.17 % of strand displacement regions do not contain this motif. The forward orientation was found in most of the strand displacement regions (79.13 %) in *nd4L*-trnT gene cluster, while the reverse complement orientation was detected in the 9.53 % (Table 2). Distribution pattern of the motif is heterogeneous among superphyla (Table S2). The motif is also not found in any of representatives of Deuterostomia, while it is rarely represented in Lophotrochozoa with a percentage of 12.45. The representatives of Ecdysozoa include the motif with 75.21 %. Furthermore, in the Ecdysozoa superphylum, the motif is represented especially in the class of Insecta (92.21 %) (Table 2). In the region of *nd4L*-X, the motif of WHWGHTW appeared in only 2.63 % of those containing this motif (Table 2), of which 69.57 % has the forward orientation and the remaining has the reverse-complement orientation. This observed pattern is also not found in Deuterostomia and is presented in only a few representatives of Lophotrochozoa (three of 23 species). The highest percentage is in the superphylum of Ecdysozoa, with 86.96 % (Table 2). For X-trnT, the motif is observed in 9.72 % of those containing this motif. Among these, the forward orientation is

Table 2
General pattern characteristics and species counts of the strand displacement region between *nd4L*-(X)-trnT gene cluster.

group										species	strand displacement							
											count	percentage	intergenic	overlapping	sequential	*nd4L*-trnT	*nd4L*-X	X-trnT
Bilateria	Ecdysozoa	Arthropoda	Mandibulata	Pancrustacea	Hexapoda	Apterygota			Collembola	12	11	91,66666667	9	1	1	9	2	
									Protura	3	2	66,66666667	2			1		1
									Diplura	6	6	100	6			5		1
						Insecta	Pterygota		Ephemeroptera	10	10	100	8	1	1	9	1	
									Odonata	14	14	100	14			14		
									Archaeognatha	2	2	100	2			2		
									Thysanura	3	3	100	2		1	3		
							Neoptera	Hemipteroid	Psocodea	11	9	81,81818182	8	1		7		2
									Thysanoptera	2	2	100	1		1	1	1	
									Hemiptera	78	78	100	71	4	3	71	4	3
								Endopterygota	Megaloptera	2	2	100	2			2		
									Raphidioptera	2	2	100	2			2		
									Neuroptera	16	16	100	16			16		
									Coleoptera	99	99	100	92	2	5	94	5	
									Strepsiptera	2	2	100	1	1		2		
									Diptera	53	52	98,11320755	49	2	1	48	2	2
									Mecoptera	6	6	100	6			6		
									Siphonaptera	3	3	100	3			3		
									Trichoptera	8	8	100	7		1	8		
									Lepidoptera	47	47	100	46	1		47		
									Hymenoptera	63	62	98,41269841	53	2	7	45	16	1
									Plecoptera	15	15	100	15			15		
									Phasmida	6	6	100	6			6		
									Orthoptera	30	30	100	29		1	30		
									Mantophasmatodea	1	1	100	1			1		
									Dictyoptera	22	21	95,45454545	21			21		
									Dermaptera	3	3	100	3			3		
									Grylloblattodea	1	1	100	1			1		
					Crustacea	Vericrustacea			Branchiopoda	8	5	62,5	5			5		
									Copepoda	7	3	42,85714286	2		1		1	2
									Thecostraca	12	8	66,66666667	7		1	1	6	1
									Malacostraca	153	131	85,62091503	125	5	1	109	12	10
						Xenocarida			Remipedia	1	1	100	1			1		
									Cephalocarida	1	1	100	1				1	
						Oligostraca			Ostracoda	1	1	100	1			1		
									Branchiura	1	0	0						
				Myriapoda	Pauropoda					1	1	100		1			1	
					Diplopoda					10	6	60	5		1	5	1	
					Chilopoda					7	6	85,71428571	3	3		4	1	1
					Symphyla					2	2	100	2				2	
			Chelicerata		Arachnida					63	33	52,38095238	28	3	2	25	4	4
					Xiphosura					1	1	100	1			1		
					Pycnogonida					4	4	100	4			4		
		Onychophora								2	0	0						
		Tardigrada								2	1	50			1	1		
		Nematoda								41	4	9,756097561	1	1	2	2	1	1
		Priapulida								2	2	100	2				2	
	Lophotrochozoa	Annelida								44	2	4,545454545	2				2	
		Bryozoa								8	1	12,5			1	1		
		Mollusca								196	117	59,69387755	87	16	14	28	86	3
		Nemertea								11	10	90,90909091	7		3		10	
		Entoprocta								1	1	100	1			1		
		Brachiopoda								5	3	60	3			3		
	Deuterostomia	Echinodermata								37	1	2,702702703	1				1	
		Hemichordata								4	0	0						
		Chordata		Craniata	Urochordata					9	0	0						
					Hyperotreti					2	0	0						
					Vertebrata					733	9	1,227830832	6	2	1		9	

found in 63.33 %, while the remaining (36.67 % %) has the reverse-complement orientation. In contrast to the other two gene clusters, two forward- and three reverse complement-oriented motifs were detected in Deuterostomia (Table 2). The motif in this gene cluster is also mostly represented with percent of 61.11 across Lophotrochozoa, while a relatively low representation was observed in Ecdysozoa (33.33 %).

3.3. The consensus sequences of the motif of WHWGHTW across the phylogenetic tree of Bilateria

The hypothetically constructed phylogenetic tree and consensus sequences of the novel motif for each relevant terminal taxonomic group were shown in Fig. 1. This motif is completely conserved in terms of the fourth and sixth positions as G and T nucleotides, respectively; whereas the second and fifth positions include the most degenerated nucleotides (A, T or C). The remaining positions are represented by A or T nucleotides (W), most probably based on the observed A + T bias in the animal mitogenomes. The sequence content of the motif is also relatively conserved at the subphylum down to the order level (Fig. 1). Considering the nucleotide bias of the motif throughout the subphyla, it exhibited a variable pattern as A or T. The nucleotides are mostly T-prone in all degenerated positions in Ecdysozoa, except for third position. However, in Lophotrochozoa, the second and last positions are T and A, respectively, *vice versa* in Deuterostomia (Fig. 1).

3.4. Estimation of putative transcription pattern of this intergenic region in Apis model

6090 of filtered reads were mapped to *nd4L* - trnT- trnP – *nd6* gene cluster. 1379 of which were located between *nd4L* and trnT regions, which covered WHWGHTW motif as “AATGATA” sequences. 95.70 % of sense trnP transcripts contain whole sequence of intergenic region (19 bp long); but, none of these cover any bases in the intergenic region according to the transcription direction of genes. Consequently, it can be suggested that this intergenic region between trnP and trnT genes is transcribed with trnP gene in a polycistronic unit.

On the other hand, 1928 of 6090 reads were located between trnT and trnP genes which covered WHWGHTW motif as “TAAGAAT” sequences with one based discorded. 98.65 % of antisense transcripts of trnT and 87.5 % of antisense transcripts of *nd4L* genes contain at least a few bases in the intergenic region (12 bp long); 56.31 % and 50 % of which transcripts were covered at least whole motif with trnT and *nd4L* genes, respectively.

As a consequence of all findings on the motif, it can be suggested that this motif is most likely responsible for the termination, pause and/or elongation of mitochondrial transcription, similarly to the reported motif in the intergenic region between *nd1*-trnS2 (Roberti et al., 2003; 2006).

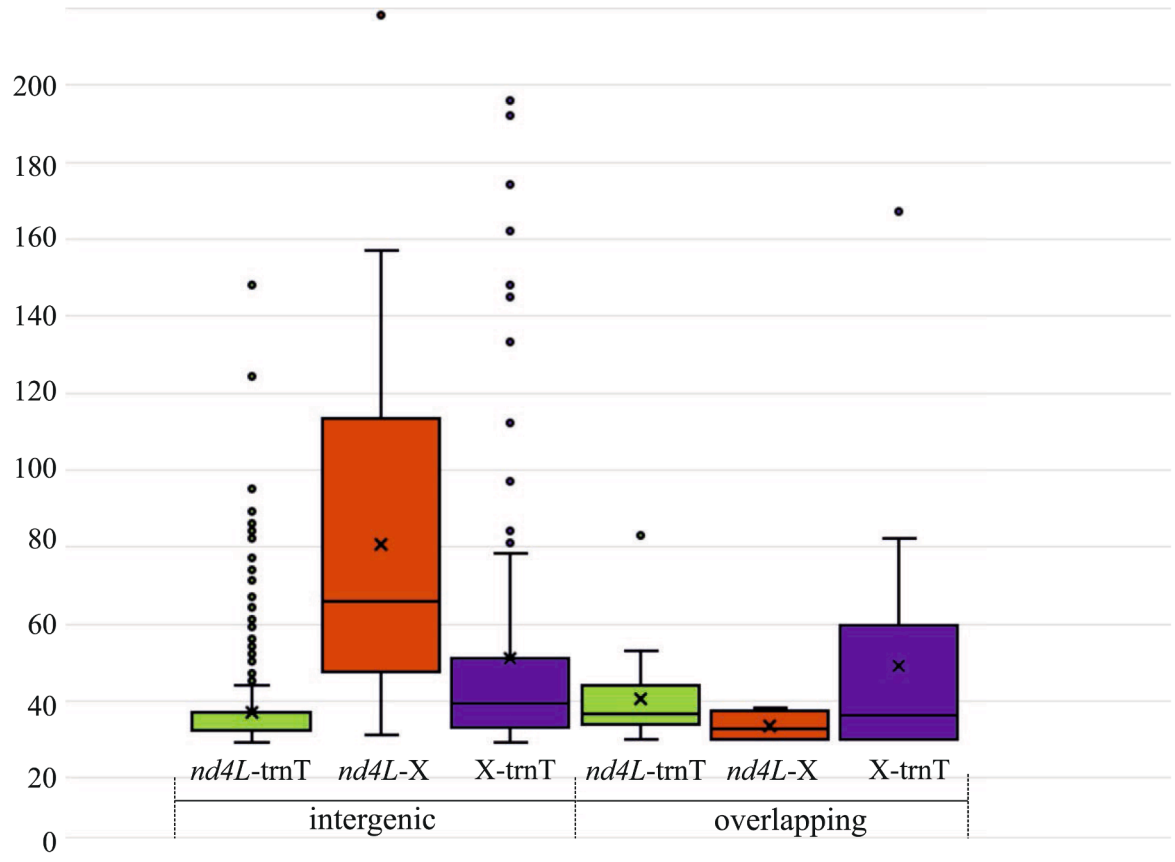

Fig. 3. Schematic representation of length variation in intergenic and overlapping regions for each gene cluster in *nd4L*-(X)-trnT order. Green represents *nd4L*-trnT; red represents *nd4L*-X and purple represents X-trnT gene order. "x" indicates average value of each cluster and dots represent excessive values.

4. Conclusion

Here, we bioinformatically investigated and characterised a novel motif of "WHWGHTW" in *nd4L*- (X) - trnT gene cluster across the mitogenomes of bilaterian animals. This motif can be suggested either as a recognition site of the transcription factors or a substrate providing binding sites for the specific interaction with transcription factors because of (i) its reverse complementarity to a previously described DmTTF binding sequence, and (ii) its possession of G and T nucleotides in the fourth and sixth positions, (iii) the bias on T and G nucleotides instead of C and A in the degenerated positions. This suggestion is also supported by the presence of a strand displacement region in the *nd4L*-(X) -trnT gene cluster, particularly in Ecdysozoa consisting of the most rearranged mitogenomes among the bilaterian superphylums. However, this assumption needs to be verified in further mitochondrial protein interaction studies to better understand its possible function(s) in the mitochondrial transcription.

Declaration of Competing Interest

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.

Acknowledgements

This work was supported by TÜBİTAK (The Scientific and Technological Research Council of Turkey) with a grant number of 117Z020). A part of this project was presented as a PhD thesis of the first author (M.N. A.).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mito.2022.11.004.

References

Akasaki, T., Nikaido, M., Tsuchiya, K., Segawa, S., Hasegawa, M., Okada, N., 2006. Extensive mitochondrial gene arrangements in coleoid Cephalopoda and their phylogenetic implications. Mol. Phylogenet. Evol. 38, 648–658. https://doi.org/10.1016/j.ympev.2005.10.018.

Andersson, S.G., Kurland, C.G., 1998. Reductive evolution of resident genomes. Trends Microbiol. 6, 263–268. https://doi.org/10.1016/S0966-842X(98)01312-2.

Andre, C., Levy, A., Walbot, V., 1992. Small repeated sequences and the structure of plant mitochondrial genomes. Trends Genet. 8, 128–132. https://doi.org/10.1016/0168-9525(92)90370-J.

Aydemir, M.N., Korkmaz, E.M., 2020. Comparative mitogenomics of Hymenoptera reveals evolutionary differences in structure and composition. Int. J. Biol. Macromol. https://doi.org/10.1016/j.ijbiomac.2019.12.135.

Babbucci, M., Basso, A., Scupola, A., Patarnello, T., Negrisolo, E., 2014. Is it an ant or a butterfly? Convergent evolution in the mitochondrial gene order of Hymenoptera and Lepidoptera. Genome Biol. Evol. 6, 3326–3343. https://doi.org/10.1093/gbe/evu265.

Bandyopadhyay, P.K., Stevenson, B.J., Ownby, J.-P., Cady, M.T., Watkins, M., Olivera, B. M., 2008. The mitochondrial genome of Conus textile, coxI–coxII intergenic sequences and Conoidean evolution. Mol. Phylogenet. Evol. 46, 215–223. https://doi.org/10.1016/j.ympev.2007.08.002.

Bernt, M., Braband, A., Schierwater, B., Stadler, P.F., 2013. Genetic aspects of mitochondrial genome evolution. Mol. Phylogenet. Evol. 69, 328–338. https://doi.org/10.1016/j.ympev.2012.10.020.

Boore, J.L., 2006. The complete sequence of the mitochondrial genome of Nautilus macromphalus (Mollusca: Cephalopoda). BMC Genomics 7, 182. https://doi.org/10.1186/1471-2164-7-182.

Boore, J.L., 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27, 1767–1780. https://doi.org/10.1093/nar/27.8.1767.

Borner, J., Burmester, T., 2017. Parasite infection of public databases: A data mining approach to identify apicomplexan contaminations in animal genome and transcriptome assemblies. BMC Genomics 18. https://doi.org/10.1186/s12864-017-3504-1.

Borner, J., Rehm, P., Schill, R.O., Ebersberger, I., Burmester, T., 2014. A transcriptome approach to ecdysozoan phylogeny. Mol. Phylogenet. Evol. 80, 79–87. https://doi.org/10.1016/j.ympev.2014.08.001.

Breton, S., Beaupré, H.D., Stewart, D.T., Piontkivska, H., Karmakar, M., Bogan, A.E., Blier, P.U., Hoeh, W.R., 2009. Comparative mitochondrial genomics of freshwater mussels (Bivalvia: Unionoida) with doubly uniparental inheritance of mtDNA: Gender-specific open reading frames and putative origins of replication. Genetics 183, 1575–1589. https://doi.org/10.1534/genetics.109.110700.

Castellanos, M., Mothi, N., Muñoz, V., 2020. Eukaryotic transcription factors can track and control their target genes using DNA antennas. Nat. Commun. 11, 540. https://doi.org/10.1038/s41467-019-14217-8.

Dowton, M., Cameron, S.L., Dowavic, J.I., Austin, A.D., Whiting, M.F., 2009. Characterization of 67 mitochondrial tRNA gene rearrangements in the Hymenoptera suggests that mitochondrial tRNA gene position is selectively neutral. Mol. Biol. Evol. 26, 1607–1617. https://doi.org/10.1093/molbev/msp072.

Faith, J.J., Pollock, D.D., 2003. Likelihood Analysis of Asymmetrical Mutation Bias Gradients in Vertebrate Mitochondrial Genomes. Genetics 165, 735–745. https://doi.org/10.1093/genetics/165.2.735.

Hahn, M.W., Stajich, J.E., Wray, G.A., 2003. The effects of selection against spurious transcription factor binding sites. Mol. Biol. Evol. 20, 901–906. https://doi.org/10.1093/molbev/msg096.

Halanych, K.M., 2004. The New View of Animal Phylogeny. Annu. Rev. Ecol. Evol. Syst. 35, 229–256. https://doi.org/10.1146/annurev.ecolsys.35.112202.130124.

Hamann, L., Bayer, K.U., Jensen, K., Harbers, K., 1994. Interaction of several related GC-box- and GT-box-binding proteins with the intronic enhancer is required for differential expression of the gb110 gene in embryonal carcinoma cells. Mol. Cell. Biol. 14, 5786–5793. https://doi.org/10.1128/mcb.14.9.5786.

Jühling, F., Pütz, J., Bernt, M., Donath, A., Middendorf, M., Florentz, C., Stadler, P.F., 2012. Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. Nucleic Acids Res. 40, 2833–2845. https://doi.org/10.1093/nar/gkr1131.

Kanehisa, M., 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27–30. https://doi.org/10.1093/nar/28.1.27.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649. https://doi.org/10.1093/bioinformatics/bts199.

Laumer, C.E., Fernández, R., Lemer, S., Combosch, D., Kocot, K.M., Riesgo, A., Andrade, S.C.S., Sterrer, W., Sørensen, M.V., Giribet, G., 2019. Revisiting metazoan phylogeny with genomic sampling of all phyla. Proc. R. Soc. B Biol. Sci. 286, 20190831. https://doi.org/10.1098/rspb.2019.0831.

Lynch, M., Scofield, D.G., Hong, X., 2005. The Evolution of Transcription-Initiation Sites. Mol. Biol. Evol. 22, 1137–1146. https://doi.org/10.1093/molbev/msi100.

Magrane, M., Consortium, U., 2011. UniProt Knowledgebase: a hub of integrated protein data. Database 2011, bar009-bar009. https://doi.org/10.1093/database/bar009.

Park, E., Song, J.-I., Won, Y.-J., 2011. The complete mitochondrial genome of Calicogorgia granulosa (Anthozoa: Octocorallia): Potential gene novelty in unidentified ORFs formed by repeat expansion and segmental duplication. Gene 486, 81–87. https://doi.org/10.1016/j.gene.2011.07.003.

Park, J.-K., Rho, H.S., Kristensen, R.M., Kim, W., Giribet, G., 2006. First Molecular Data on the Phylum Loricifera – An Investigation into the Phylogeny of Ecdysozoa with Emphasis on the Positions of Loricifera and Priapulida. Zoolog. Sci. 23, 943–954. https://doi.org/10.2108/zsj.23.943.

Rehm, P., Borner, J., Meusemann, K., von Reumont, B.M., Simon, S., Hadrys, H., Misof, B., Burmester, T., 2011. Dating the arthropod tree based on large-scale transcriptome data. Mol. Phylogenet. Evol. 61, 880–887. https://doi.org/10.1016/j.ympev.2011.09.003.

Reyes, A., Gissi, C., Pesole, G., Saccone, C., 1982. Asymmetrical Directional Mutation Pressure in the Mitochondrial Genome of Mammals 957–966.

Roberti, M., Bruni, F., Polosa, P.L., Gadaleta, M.N., Cantatore, P., 2006. The Drosophila termination factor DmTTF regulates in vivo mitochondrial transcription. Nucleic Acids Res. 34, 2109–2116. https://doi.org/10.1093/nar/gkl181.

Roberti, M., Loguercio Polosa, P., Bruni, F., Musicco, C., Gadaleta, M.N., Cantatore, P., 2003. DmTTF, a novel mitochondrial transcription termination factor that recognises two sequences of Drosophila melanogaster mitochondrial DNA. Nucleic Acids Res. 31, 1597–1604. https://doi.org/10.1093/nar/gkg272.

Serb, J.M., 2003. Complete mtDNA Sequence of the North American Freshwater Mussel, Lampsilis ornata (Unionidae): An Examination of the Evolution and Phylogenetic Utility of Mitochondrial Genome Organization in Bivalvia (Mollusca). Mol. Biol. Evol. 20, 1854–1866. https://doi.org/10.1093/molbev/msg218.

Stewart, J.B., Beckenbach, A.T., 2009. Characterization of mature mitochondrial transcripts in Drosophila, and the implications for the tRNA punctuation model in arthropods. Gene 445, 49–57. https://doi.org/10.1016/j.gene.2009.06.006.

Telford, M.J., Budd, G.E., Philippe, H., 2015. Phylogenomic insights into animal evolution. Curr. Biol. https://doi.org/10.1016/j.cub.2015.07.060.

Wu, Z., Cuthbert, J.M., Taylor, D.R., Sloan, D.B., 2015a. The massive mitochondrial genome of the angiosperm Silene noctiflora is evolving by gain or loss of entire chromosomes. Proc. Natl. Acad. Sci. 112, 10185–10191. https://doi.org/10.1073/pnas.1421397112.

Wu, Z., Stone, J.D., Štorchová, H., Sloan, D.B., 2015b. High transcript abundance, RNA editing, and small RNAs in intergenic regions within the massive mitochondrial genome of the angiosperm Silene noctiflora. BMC Genomics 16, 938. https://doi.org/10.1186/s12864-015-2155-3.

Yevshin, I., Sharipov, R., Kolmykov, S., Kondrakhin, Y., Kolpakov, F., 2019. GTRD: a database on gene transcription regulation—2019 update. Nucleic Acids Res. 47, D100–D105. https://doi.org/10.1093/nar/gky1128.

Zhang, H.L., Liu, B.B., Wang, X.Y., Han, Z.P., Zhang, D.X., Su, C.N., 2016. Comparative mitogenomic analysis of species representing six subfamilies in the family tenebrionidae. Int. J. Mol. Sci. https://doi.org/10.3390/ijms17060841.