Heliyon 9 (2023) e17070

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Full-length transcriptome profiling of *Aphidius gifuensis* mitochondrial genome with gene rearrangement and control region duplication

Xinjie Zhao [a,1], Shiwen Xu [a,1], Jingrui Li [b], Hailin Yang [c], Li Tian [a], Fan Song [a], Wanzhi Cai [a], Zhonglong Lin [d,**], Hu Li [a,*]

[a] *Department of Entomology and MOA Key Lab of Pest Monitoring and Green Management, College of Plant Protection, China Agricultural University, Beijing, 100193, China*
[b] *State Key Laboratory of Plant Physiology and Biochemistry, College of Biological Sciences, China Agricultural University, Beijing, 100193, China*
[c] *Yuxi Branch of Yunnan Tobacco Company, Yuxi, 653100, China*
[d] *Yunnan Tobacco Company of China National Tobacco Corporation, Kunming, 650011, China*

ARTICLE INFO

Keywords:
Aphidius gifuensis
full-length transcriptome
Mitochondrial gene transcription
Post-transcriptional cleavage process

ABSTRACT

Although mitochondrial gene rearrangement has been observed in many insect lineages, little is known about how it affects mitochondrial gene transcription. To address this question, we first constructed a quantitative transcription map for *Aphidius gifuensis*, a species of parasitoid wasp known to have a highly rearranged mitochondrial genome (mitogenome) and two potential control regions (CRs). Based on this transcription map, we assessed the models of the mitochondrial transcription and post-transcription cleavage. We found that the J and N strand of this mitogenome differ significantly in transcriptional regulation. On the J strand, we found two transcription initiation sites (TISs), five transcription termination sites (TTSs), and six polycistronic primary transcripts whereas only one TIS, one TTS and one polycistronic primary transcript can be found on the N strand. Most of the non-coding regions of both strands were transcribed into primary transcripts and cleaved after transcription. The proposed mode of transcription of *A. gifuensis* was similar to that of *Drosophila*, a model organism with no gene rearrangement. And two rearranged gene clusters (*trnI-CR1-trnM-CR2-trnQ* and *trnW-trnY-trnC*) seemed to have little effects on the mode of transcription. In addition, our results revealed the presence of TISs in CR1 and CR2, implying that both CRs maybe required for transcriptional regulation. Analysis of the post-transcriptional cleavage process showed that there were both "forward cleavage" and "reverse cleavage" models in *A. gifuensis*, and more than one way of cleavages were found in three regions. The incomplete transcripts suggested that the direction of mitochondrial RNA degradation was from 5′ to 3′ end and supported the view of polyadenylation-dependent RNA degradation. Our study provides insights into the transcriptional and post-transcriptional regulation processes of highly rearranged insect mitogenomes.

* Corresponding author.
** Corresponding author.
E-mail addresses: sdlzl1983@163.com (Z. Lin), lih@cau.edu.cn (H. Li).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2023.e17070
Received 16 May 2022; Received in revised form 17 September 2022; Accepted 6 June 2023
Available online 8 June 2023

1. Introduction

Insect mitochondrial genome (mitogenome) is a 15–18 kb double-strand circular DNA typically comprising 37 coding genes, including 13 protein-coding genes (PCGs), 22 transfer RNA genes (tRNAs) and two ribosomal RNA genes (rRNAs) [1]. In addition to the coding genes, there are non-coding regions of various length, with the longest ones containing initiation sites for both replication and transcription and commonly known as the control region (CR) [1,2]. The arrangement of the mitochondrial genes is generally conserved across most inset orders [2,3]. With the rapid advance of DNA sequencing technology, an increasing number of insect mitogenomes has been characterized, yet few studies have been focused on transcription and post-transcriptional regulation of insect mitochondrial genes. Among insects, *Drosophila* is perhaps the most extensively studied group in regards to mitochondrial gene transcription. Since the report of the first *Drosophila* mitogenome, numbers of studies have been done to assess its transcriptional process. For instance, Berthier et al. [4] characterized two transcripts together with previously reported transcripts, and inferred that five TISs (transcription initiation sites) formed five transcriptional cassettes in *Drosophila* and most antisense genes were not transcribed. The result was confirmed by a subsequent study by Stewart et al. [5] using RACE and RT-PCR methods. However, Roberti et al. [6] studied the role of *Drosophila* mitochondrial transcription termination factor (DmTTF) in vivo and proposed two models of transcription containing two and four TISs respectively, in which most antisense genes were transcribed. The difference between these two models was whether two antisense gene clusters (*AstrnS2/AsCytb/AsND6* on the J strand and *AstrnF/AsND5/AstrnH/AsND4/AsND4L* on the N strand) were transcribed. The post-transcriptional cleavage process was thought to follow the "tRNA punctuation" model [7]. In this model, tRNAs were removed from the primary transcripts, leaving messenger RNA (mRNA) and rRNA transcripts. In addition, according to tRNAs removed by 5′ to 3′ cleavage or 3′ to 5′ cleavage, two models of post-transcriptional cleavage have been proposed, including the "reverse cleavage" model in *Drosophila* [5] and "forward cleavage" model in *Erthesina fullo* [8]. These studies have provided critical insights into the mechanisms of the mitochondrial gene transcription, RNA processing and mRNA maturation.

Nevertheless, most knowledge about mitochondrial transcription has come from a few model insect species with ancestral and conserved mitochondrial structure [4–6,8]. Little is known about how mitochondrial transcription is regulated in insect groups with highly rearranged mitogenomes [9–13]. Hymenopteran insects (bees, wasps and ants) are known to display very high rates of mitochondrial gene rearrangements, with at least one translocated tRNA being found in every sequenced mitogenome of hymenopteran species [13]. Previous study found that the mitogenome of *Aphidius gifuensis,* a parastoid wasp species, have underwent intense gene order modification, with tRNA gene rearrangements and two CRs found in two gene clusters (*trnI-CR1-trnM-CR2-trnQ* and *trnW-trnY-trnC*) [14]. These position this species an ideal model to study the process of mitochondrial gene transcription in a highly rearranged insect mitogenome.

In this study, we constructed a quantitative transcription map for *A. gifuensis* using the high quality, full-length transcripts generated from single molecule real-time (SMRT) sequencing, a RNA sequencing technique that have been proved to be particularly good at dealing with complex RNA structures (e.g., transcript isoforms, long non-coding RNAs (LncRNAs), splice sites and etc [8, 15–18]). Based on this transcriptome, we accessed the mode of mitochondrial transcription and post-transcriptional cleavage in this species. Our study revealed some unique features of RNA processing and mRNA maturation of mitochondrial genes of this species, and provided new insights into the pattern of gene transcription and posttranscriptional regulation of highly rearranged mitochondrial genomes.

2. Materials and methods

2.1. Sample collection, RNA extraction, library preparation and sequencing

Total RNA was extracted from 75 adult males of *A. gifuensis* from the laboratory in Kunming, Yunnan, China using TRIzol Universal (Tiangen, Beijing, China). The cDNA was synthesized using the Clontech SMARTer PCR cDNA Synthesis Kit (Clontech Laboratories, Inc, CA, USA) with 3′ SMART CDS Primer II A (5′-AAGCAGTGGTATCAACGCAGAGTAC-T(30)-3′) and SMARTer II A Oligonucleotide (5′-AAGCAGTGGTATCAACGCAGAGTACATGGG-3′). After size-selection, the cDNA was amplified and the library with the insert size between 0.5 and 6 kb was constructed according to the PacBio Iso-Seq protocol. Finally, one SMRT cell was performed on Pacbio Sequel platform with circular consensus sequencing (CCS) mode at GrandOmics Biosciences Company (Beijing, China).

2.2. Data assessment and quality control

The PacBio SMART Analysis v10.2 (http://www.pacb.com/devnet/) was performed on the raw reads data. The CCS v5.0.0 with parameters (Minimum Full Passes = 1, Minimum Predicted Accuracy = 0.9) was used to process the sequenced reads into the high-quality circular consensus sequencing (CCS) reads, then the lima v2.0.0 was used to produce full-length non chimera (FLNC) reads as known as draft transcripts by removing the full-length chimera, 5′ and 3′ ends cDNA primers and 3′ ployA sequences. The samtools v1.11 (http://www.htslib.org/) was used to convert BAM file to FASTA file.

2.3. Bioinformatic analysis

The annotated complete mitogenome sequence of *A. gifuensis* was collected from previous study [14], with accession number MT264907 in GenBank. The draft transcripts in FASTA format were mapped to complete mitogenome reference separately using

Geneious v10.2.6 (http://www.geneious.com/) [19] with parameters (Maximum gap per read = 10%, Maximum gap size = 20, Minimum overlap identity = 85%, maximum mismatches per read = 10%, Maximum ambiguity = 16). Statistics and plotting were performed to identify the 5′ and 3′ ends of mature transcripts, polycistronic transcripts and antisense transcripts.

3. Results

3.1. The quantitative transcription map of A. gifuensis mitogenome

A total of 18.63 Gb raw data with 684,943 raw reads was obtained. The mean polymerase read length is 27,194 bp and the polymerase read N50 is 51,560 bp. After reads was split into subreads, a total of 18.21 Gb data with 9,774,531 subreads were obtained. The mean subread length is 1863 bp, subread N50 is 2360 bp and longest subread length is 122,470 bp. The raw data of the full-length transcriptome was deposited in the NCBI SRA database with accession number SRR11568632. Then, the raw reads generated 386,940 high-quality CCS (accuracy≥0.9) transcripts, and produced 306,247 draft transcripts. To build a quantitative transcription map (Fig. 1), draft transcripts were mapped to the complete mitogenome of *A. gifuensis* generated from previous study [14]. The mean coverage of mitogenome of *A. gifuensis* is 435.7 bases. The relative expression levels of mitochondrial genes, if ranked from highest to lowest, were *16S rRNA*, *COI*, *COIII*, *Cytb*, *ATP8/ATP6*, *COII*, *ND4/ND4L*, *ND6*, *ND1*, *ND3*, *ND5*, *ND2* and *12S rRNA*. As can be seen from the trapezoid formed by the long to short transcripts in the quantitative transcription map, the number of transcripts degraded at 3′ end was much lower than that of 5′ end in a gene, which may imply that the orientation of mitochondrial RNA degradation was from 5′ to 3′ end.

In addition, we re-annotated mitochondrial genes of *A. gifuensis* precisely according to the full-length transcripts (Table 1). Compared to previous annotations using DNA sequencing [14], annotations of six PCGs (*ND2*, *ATP6*, *COIII*, *ND3*, *Cytb* and *ND1*) and *12S rRNA* were modified, resulting in shortened gene overlap regions and a compact mitogenome. The annotation of 3′ ends of five genes (*ND2*, *ATP6*, *ND3*, *Cytb* and *ND1*) was altered, leaving a single T residue as incomplete stop codon. The annotation of 5′ ends of four genes (*ND2*, *COIII*, *ND3* and *12S rRNA*) was altered, resulting in the change of start codon in PCGs (Fig. S1) and a non-coding region of 39 bp between *12S rRNA* and *trnI*. It is worth mentioning that the overlaps between *ND1* and *trnS2*, *trnQ* and *ND2* were replaced by non-coding regions after re-annotation.

3.2. RNA precursors and RNA degradation

Two mature transcripts (*ATP8/ATP6* and *ND4/ND4L*) were polycistronic mRNAs, as reported in other species [5,20,21], while the other mature transcripts were monocistronic mRNAs or rRNAs. In addition, there were some polycistronic transcripts and antisense transcripts, which were considered to be the primary transcripts or the products of post-transcriptional process. A total of 74 polycistronic transcripts (except for *ATP8/ATP6* and *ND4/ND4L*) and antisense transcripts were mapped to mitogenome, including 28

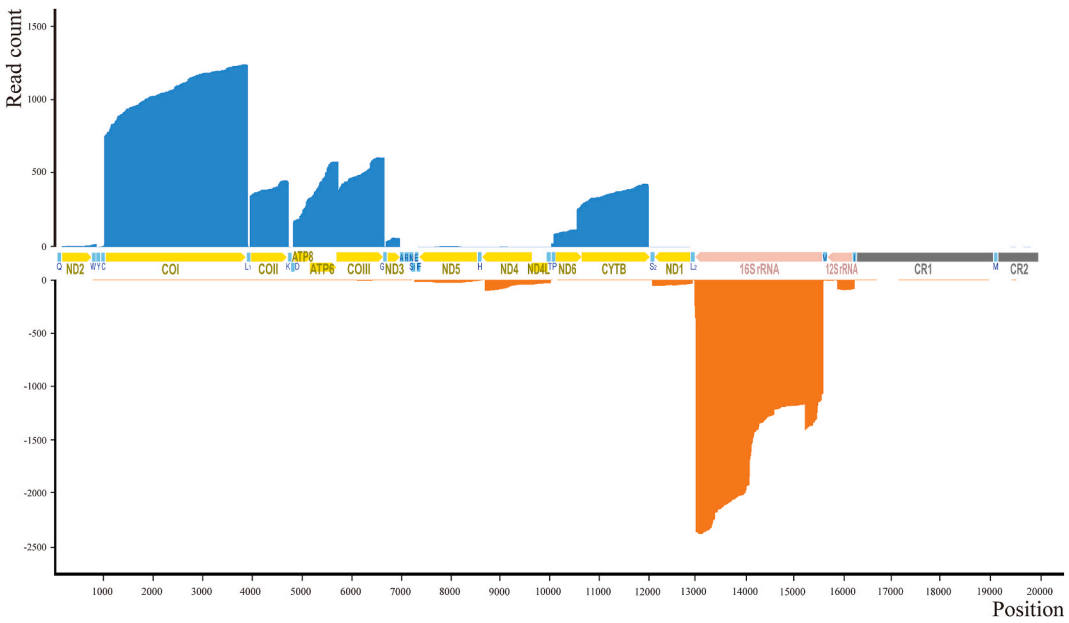

Fig. 1. The quantitative transcription map of *Aphidius gifuensis* mitogenome. The mitogenome was arranged in the J strand orientation. Alignments of transcripts encoded by J strand were piled along the positive y-axis in blue color. Alignments of transcripts encoded by N strand were piled along the negative y-axis in orange color. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1
The precise annotation of *Aphidius gifuensis* mitogenome based on the full-length transcripts.

Gene	Coding strand	Annotations using DNA sequencing				Annotations according to transcripts			
		Position (J strand as reference)	Length/bp	Start codon	Stop codon	Position (J strand as reference)	Length/bp	Start codon	Stop codon
trnQ	N	69–1	69			69–1	69		
ND2	J	67-1074	1008	AUA	UAA	82-1072	991	AUU	U
trnW	J	1073–1139	67			1073–1139	67		
trnY	N	1208–1142	67			1208–1142	67		
trnC	N	1275–1212	64			1275–1212	64		
COI	J	1276–2814	1539	AUG	UAA	1276–2814	1539	AUG	UAA
trnL$_1$	J	2821–2886	66			2821–2886	66		
COII	J	2887–3570	684	AUU	UAA	2887–3570	684	AUU	UAA
trnK	J	3575–3644	70			3575–3644	70		
trnD	J	3644–3710	67			3644–3710	67		
ATP8	J	3711–3872	162	AUU	UAA	3711–3872	162	AUU	UAA
ATP6	J	3866–4543	678	AUA	UAA	3866–4529	664	AUA	U
COIII	J	4527–5313	787	AUU	U	4530–5313	784	AUG	U
trnG	J	5314–5376	63			5314–5376	63		
ND3	J	5374–5730	357	AUA	UAG	5377–5728	352	AUU	U
trnA	J	5729–5792	64			5729–5792	64		
trnR	J	5793–5859	67			5793–5859	67		
trnN	J	5861–5926	66			5861–5926	66		
trnS$_1$	J	5923–5992	70			5923–5992	70		
trnE	J	5995–6063	69			5995–6063	69		
trnF	N	6122–6056	67			6122–6056	67		
ND5	N	7797–6124	1674	AUA	UAA	7797–6124	1674	AUA	UAA
trnH	N	7861–7798	64			7861–7798	64		
ND4	N	9208–7862	1347	AUG	UAA	9208–7862	1347	AUG	UAA
ND4L	N	9486–9202	285	AUU	UAA	9486–9202	285	AUU	UAA
trnT	J	9488–9549	62			9488–9549	62		
trnP	N	9611–9549	63			9611–9549	63		
ND6	J	9613–10,159	547	AUU	U	9613–10,159	547	AUU	U
CytB	J	10,160–11,302	1143	AUG	UAG	10,160–11,300	1141	AUG	U
trnS$_2$	J	11,301–11,366	66			11,301–11,366	66		
ND1	N	12,336–11,365	972	AUU	UAA	12,336–11,373	964	AUU	U
trnL$_2$	N	12,402–12,338	65			12,402–12,338	65		
16S rRNA	N	13,683–12,403	1281			13,683–12,403	1281		
trnV	N	13,747–13,684	64			13,747–13,684	64		
12S rRNA	N	14,527–13,748	780			14,488–13,748	741		
trnI	N	14,592–14,528	65			14,592–14,528	65		
CR1	J	14,593–18,408	3816			14,593–18,408	3816		
trnM	N	18,471–18,409	63			18,471–18,409	63		
CR2	J	18,472–20,134	1663			18,472–20,134	1663		

encoded by N strand and 46 encoded by J strand (Table S1; Fig. 2). These 74 transcripts covered 24 genes on the coding DNA strand: 12 PCGs (except for *ND5*), two rRNAs, and 10 tRNAs (*trnY*, *trnC*, *trnL*$_2$, *trnV*, *trnI*, *trnL*$_1$, *trnK*, *trnG*, *trnT* and *trnS*$_2$), while covered 29 gene regions on the non-coding DNA strand: regions of 13 PCGs and 16 tRNAs (*AstrnW*, *AstrnL*$_1$, *AstrnD*, *AstrnK*, *AstrnG*, *AstrnA*, *AstrnR*, *AstrnN*, *AstrnS*$_1$, *AstrnE*, *AstrnT*, *AstrnS*$_2$, *AstrnY*, *AstrnC*, *AstrnH* and *AstrnP*). This was different from the *Drosophila* polycistron, which covers 37 genes [5]. The gene regions on the non-coding DNA strand have relatively higher coverage, probably because some transcripts were cleaved from polycistrons shortly after transcription. In addition, we observed eight LncRNAs transcribed from the CRs. Four LncRNAs of CR1 were encoded by N strand, three LncRNAs of CR2 were encoded by J strand, and only one LncRNA of CR2 was encoded by N strand.

Furthermore, we observed 29 incomplete polycistrons (Fig. 2). The ends of 15 polycistrons aligned to AsRNAs. That is, once the attached sense RNA was cleaved, AsRNA begun to degrade at the cleavage site, regardless of whether it was still attached to other sense RNA. It was also observed in human and may indicate processing and degradation of the mitochondrial polycistrons occur simultaneously [22]. The incomplete monocistronics were mainly degraded at 5′ end. In addition to a bunch of transcripts indicating TTSs in *12S rRNA* and *16S rRNA*, 3′ ends of some monocistronics aligned to the intragenic regions of *COI*, *16S rRNA*, *Cytb*, *COII*, *ND5*, *COIII*, *ND6*, *12S rRNA*, *ATP8/ATP6*, *ND4/ND4L*, and *ND1* (Table S2). The 3′ ends were also observed within *COII* and *COIII* open reading frames in human mitochondrial RNA and considered as decay intermediates [22].

3.3. The proposed model of mitochondrial transcription

The CR of mitogenome plays essential role in mitochondrial transcription and replication [23]. Since LncRNAs of CR were defined as long transcription initial RNAs (ltiRNAs) in human mitogenome, which were considered to be the precursors of the transcription initial RNAs (tiRNAs) [24]. And tiRNAs are produced at sites associated with stalled RNA polymerase as well as at TISs and splice sites [25]. So, the location of TIS in CR can be assessed based on the presence of LncRNAs in these regions (Fig. 2). For the J strand, we found four of five LncRNAs were concentrated in CR1 (see Nos. 69599317, 67240121, 51184039 and 60162668 in Table S1). In addition, the polycistrons *CR1/trnI/12S rRNA/trnV/16S rRNA* (ChrM: 15,081–13,491) indicated the presence of primary transcripts covering CR1 to *16S rRNA*. Therefore, we proposed that CR1 contains TIS of the J strand. And referring to the explanation of the high expression of *12S rRNA* in human mitochondrial transcription [26], the presence of the TIS upstream of *16S rRNA* may explain much higher level of expression of *16S rRNA* than *12S rRNA*. For the N strand, three LncRNAs (see Nos. 50266705, 9896047 and 73925251 in Table S1) were all present in the CR2, and two of them initiated at position 19,023, so we proposed that the TIS of N strand was located in the CR2.

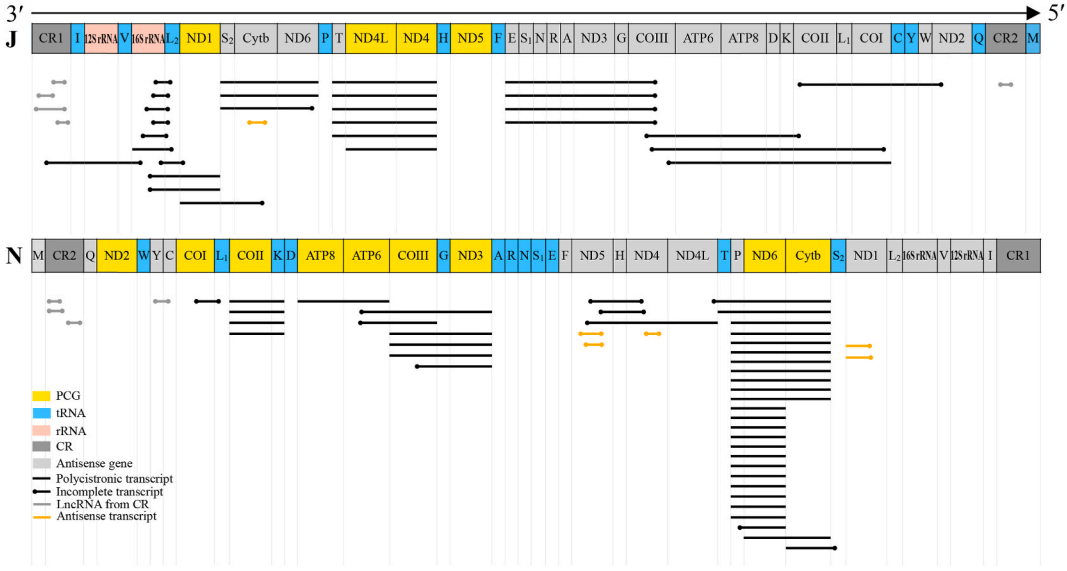

Fig. 2. The polycistronic transcripts, antisense transcripts and LncRNAs from control regions transcribed from J strand (above) and N strand (below). The black, orange and gray strips represented polycistronic transcripts, antisense transcripts and LncRNAs from CRs, respectively. The dot of the strips indicated incomplete transcripts. Two mature transcripts, *ATP8/ATP6* and *ND4/ND4L*, were represented by one strip respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

The TTSs region of mitochondrial transcription can be located according to the positions of the ends of transcripts. For the J strand, 75 sequences (ChrM: 14,488–14,055) were aligned from the 5′ end of *12S rRNA* but only to position 14,055 inside it, and no 5′ end of transcripts located in positions 14,054–13,748. The 75 sequences were unable to be translated into proteins, indicating the presence of TTS around position 14,055. We also observed 238 sequences (ChrM: 13,683–13,491) were aligned from the 5′ end of *16S rRNA* but only to position 13,491 inside it, and a polycistron *CR1/trnI/12S rRNA/trnV/16S rRNA* (ChrM: 15,081–13,491) terminated at position 13,491. Therefore, we proposed that there was another TTS near position 13,491. By analogy, we inferred two other TTSs, one was inside $trnL_2$ because six polycistrons *16S rRNA/*$trnL_2$ terminated around position 12,367 (Table S1), and another one was inside *COIII* because four polycistrons $AstrnE/AstrnS_1/AstrnN/AstrnR/AstrnA/AsND3/AstrnG/AsCOIII$ (ChrM: 6055–4825) terminated at position 4825. In addition, we introduced a TTS downstream of *trnM* to achieve transcription of all genes. Mitochondrial genes of N strand were transcribed except for a few antisense genes. We observed that two transcripts of *AsND1* terminated at positions 12,280 and 12,282, respectively, and there was no transcript downstream of position 12,282. Therefore, we believed that there was a TTS near position 12,282.

Based on the above analysis, we proposed a transcriptional model of mitogenome in *A. gifuensis* (Fig. 3). There were at least five TTSs and six polycistronic primary transcripts on the J strand, and at least one TTS and one polycistronic primary transcript on the N strand with antisense genes downstream of antisense *ND1* not transcribed.

3.4. The post-transcriptional cleavage process

We found four polycistronic transcripts of *COII/trnK* (ChrM: 2887–3643) (Fig. 2) which were contrary to the putative "reverse cleavage" model (Fig. 4A) [5] and "forward cleavage" models (Fig. 4B) [8]. In other words, these transcripts had the ability to remove *trnD* by 3′ to 5′ cleavage and $trnL_1$ by 5′ to 3′ cleavage (Fig. 4C). In addition, some transcripts indicated the presence of more than one way of cleavage in three regions (Fig. 4D–I). The transcripts of *16S rRNA/*$trnL_2$*/ND1* (ChrM: 13,199–11,373) and $AstrnS_2/AsCyt\text{-}b/AsND6$ (ChrM: 11,372–9619) showed the cleavage preferentially occurred between the AsRNA ($AstrnS_2/AsCytb/AsND6$) and mRNA (*ND1*) (Fig. 4D), while the *ND1/*$AstrnS_2$*/AsCytb* (ChrM: 12,337–10,440) showed the opposite situation, with the cleavage occurring preferentially between rRNA (*16S rRNA*) and tRNA (*trnL2*) (Fig. 4E). And a similar pattern was found between AsRNA and mRNA when the transcript *AstrnP/ND6/Cytb* (ChrM: 9550–11,300) was cleaved (Fig. 4F–G). The transcript *COIII/trnG/ND3* (ChrM: 4530–5728) showed preferent cleavage between mRNAs (Fig. 4H), while the transcript *ATP8/ATP6/COIII* (ChrM: 4068–5313) showed preferent cleavage between tRNA and mRNA (Fig. 4I). After cleaved from polycistrons, most monocistrons of *COI*, *COII* and *ND5* still carried the untranslated regions (UTRs) of 6 bp, 4 bp and 1 bp encoded by the intergenic regions. The UTRs were only found at the 3′ end of mRNAs, meaning that the UTRs would be carried behind the stop codon but removed before the start codon.

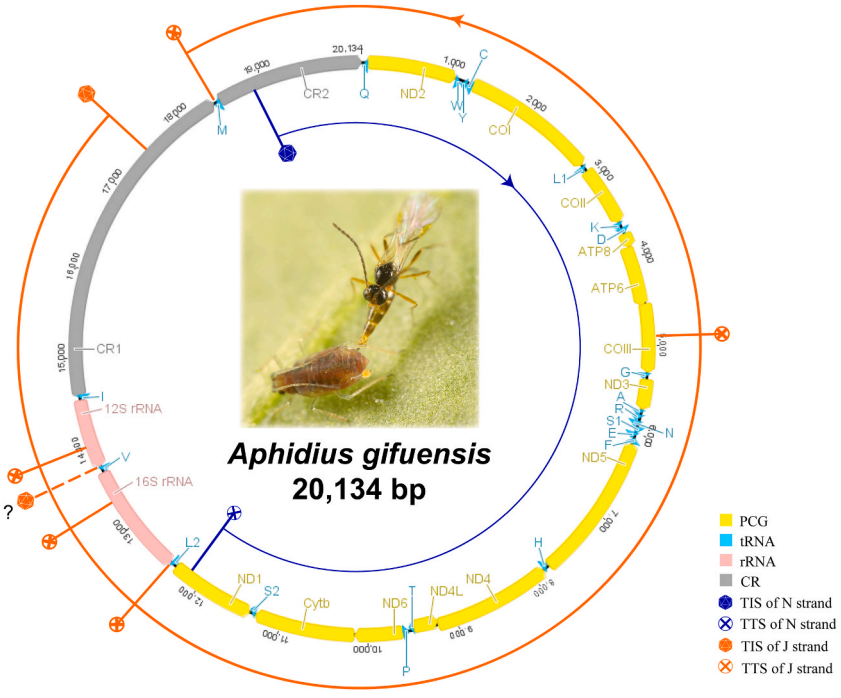

Fig. 3. The proposed models of mitochondrial transcription in *Aphidius gifuensis*. The blue and orange curves represented the polycistronic primary transcripts of N and J strand, respectively, and the orientation of transcription was shown by the arrows. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

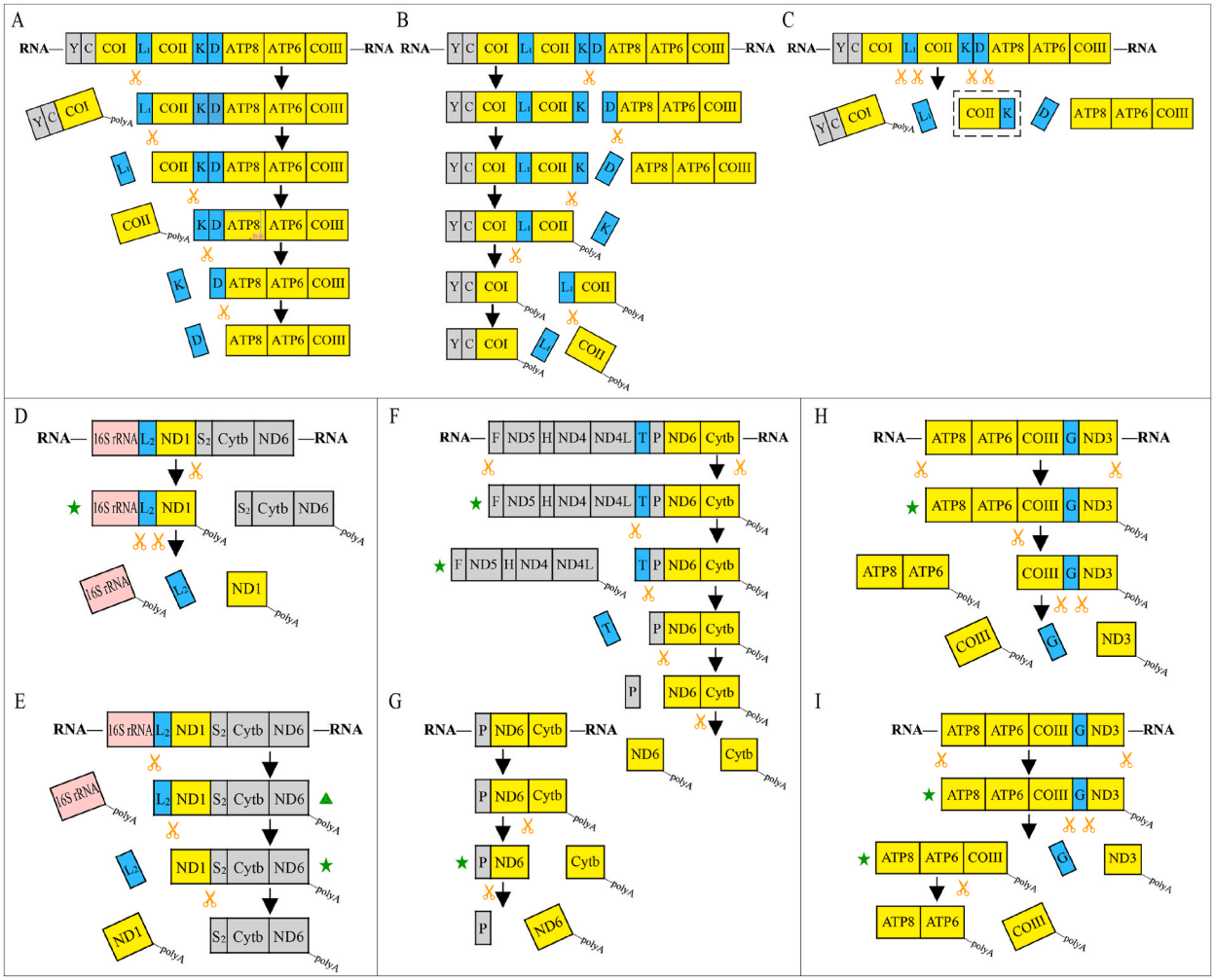

Fig. 4. The post-transcriptional cleavage process of mitogenome in *Aphidius gifuensis*. **(A)** The simulated cleavage process of primary transcript in terms of the reverse cleavage model; **(B)** The simulated cleavage process of primary transcript in terms of the forward cleavage model; **(C)** The cleavage process inferred from transcripts obtained in this study; **(D and E)** Two modes of cleavage of transcript *16S rRNA/trnL$_2$/ND1/AstrnS$_2$/AsCytb/AsND6*; **(F and G)** Two modes of cleavage of transcript *AstrnP/ND6/Cytb*; **(H and I)** Two modes of cleavage of transcript *ATP8/ATP6/COIII/trnG/ND3*. The incomplete transcripts were indicated by stars and the putative transcripts that were not observed, were represented by triangles.

4. Discussion

In this study, we constructed a quantitative transcript map of *A. gifuensis* mitogenome and ranked the expression levels of individual mitochondrial genes. The relative expression level of mitochondrial genes of *A. gifuensis* was similar to that of insect species with no mitochondrial gene rearrangement [4,8]. Seven genes of the mitogenome of *A. gifuensis* was re-annotated based on transcripts, and single T residues were identified as incomplete stop codon in five PCGs. Despite this, UAA was still present at the 3′ end of the monocistronics of these five PCGs, implying that mRNAs cleaved from the polycistronics obtained their complete stop codons by some biochemical reactions, most likely polyadenylation. In addition, genes we re-annotated were all PCGs and rRNA without tRNAs. The reason was that the insert size of cDNA library was between 0.5 and 6 kb according to the PacBio Iso-Seq protocol, which can hardly effectively contain tRNA of about 65bp in length. In further study, more hybrid sequencing strategies were needed to make use of more accurate short reads in conjunction with PacBio long reads.

We also classified the transcripts into distinct types, including the polycistronic transcripts, antisense transcripts and LncRNAs from CRs. We proposed the models of transcription and post-transcriptional cleavage process based on the specific transcripts. The model of mitochondrial gene transcription of *A. gifuensis* contained two TISs and five TTSs on the J strand, whereas one TIS and one TTS on N strand. The entire J strand was transcribed with the polycistronic primary transcript initiated from CR1 and terminated downstream of *trnM*, and four transcriptional cassettes of J strand initiated from CR1 and terminated at *12S rRNA*, *16S rRNA*, *trnL2* and *COIII* respectively. In addition, there was a primary transcript initiated from *trnV* and terminated at *trnL2*. The entire N strand was transcribed without non-coding regions downstream of antisense *ND1* and the polycistronic primary transcript initiated from CR2 and

terminated at antisense *ND1*. The proposed mode of transcription of *A. gifuensis* was similar to that of the model organism *Drosophila* [6]. The mitochondrial genes of J strand were all transcribed, whereas the mitochondrial genes of N strand were transcribed except for a few antisense genes downstream of antisense *ND1*. Most of the non-coding regions of both strands were transcribed as the primary transcript and cleaved after transcription. Modification of gene order of two gene clusters (*trnI-CR1-trnM-CR2-trnQ* and *trnW-trnY-trnC*) seemed to have little effects on the mode of transcription, possibly because the rearrangements occurred among tRNA genes rather than PCGs. As for the duplicated control regions, the current results identified TISs in CR1 and CR2, indicating that both CRs play essential role in transcriptional regulation. This result is consistent with the findings from mammals that an individual mitochondrial DNA strand usually process one TIS and may have multiple TTSs, and the presence of multiple TTS is believed to enable the production of transcripts of various length or the control of expression of different genes [27]. Further research was needed to determine the exact location of TISs and TTSs.

A detailed transcription map of HeLa cell mitogenome showed that the precise endonucieoiytic cleavages occurred almost immediately before and after a tRNA sequence, suggesting that the tRNA sequences may play an important role as recognition signals in the processing of the primary transcripts [28]. On this basis, the "tRNA punctuation" model was proposed [7]. This model applied to *Drosophila* and found that the tRNAs were removed sequentially in a 3′ to 5′ direction, which was called "reverse cleavage" model [5]. However, the tRNAs were removed by 5′ to 3′ cleavage in the polycistronic transcripts in *Eriosina fullo*, which was known as the "forward cleavage" model [8]. The 5′ end of tRNA was cleaved first by RNase P and 3′ end was cleaved later by RNase Z, and the cleave of other RNA such as mRNA depended on other protein factors [29]. The post-transcriptional cleavage process of *A. gifuensis* followed a "tRNA punctuation" model, in which tRNAs were recognized and cleaved from the primary transcripts and remaining mRNA and rRNA subunits were processed to their functional forms. In particular, we observed transcripts that followed both "reverse cleavage" and "forward cleavage" model during the post-transcriptional cleavage process as a supplement to "tRNA punctuation" model. In addition, cleavage could be processed between mRNA and AsRNA, tRNA and AsRNA, mRNA and mRNA, but it was noteworthy that cleavage could not be processed between AsRNAs. These observations indicate that RNA cleavage of mitochondrial genes does not follow a general defined processing order. And we speculated that there are different modes of cleavage in the post-transcriptional cleavage process, and the temporary order of cleavage depends on the order of binding of different restriction endonuclease at different restriction sites. During the maturation of monocistrons, we observed some UTRs at the 3′ end of mRNAs encoded by the intergenic regions. These UTRs were removed at the 5′ ends of mRNA for accurate recognition of the start codon, but were remained at 3′ end of mRNA until translation, maturation, even degradation, as it did not affect the translation with the in-frame stop codon. The process was similar to the results observed in *Drosophila* [5].

In addition, our data also revealed some interesting functions of the polyA structure of mRNA. A PolyA tail structure is normally attached to the 3′ end of a mRNA molecule and generally believed to stabilize and protect RNA from degradation [30]. However, some studies also revealed that it did not have these effects [31–33]. In the current study, we found that all transcripts such as rRNA, mRNA, even AsRNA and incomplete RNA carried a polyA tail on their 3′ ends, indicating that at least in some mitochondrial genes, the polyA tail may not serve to stabilize RNA. This supported the view of polyadenylation-dependent RNA degradation proposed by previous researchers [33], that a polyA tail added to new 3′ end of cleaved RNA to make it easier to identify and degrade faster. Polyadenylation and RNA cleavage occurred almost simultaneously.

In summary, our findings provide new insights into mitochondrial gene transcription, RNA processing and RNA degradation in insect mitogenomes with gene rearrangements and CR duplication. Our results also highlighted the need for more studies on insect lineages with derived mitochondrial structure to gain a comprehensive understanding of transcription and post-transcriptional cleavage process in insect mitochondrial genes. However, our study is rather specialized and descriptive in nature and lacks molecular insights into RNA-related mechanisms. Further molecular work focused on the mechanisms of transcriptional regulation is needed to validate our findings.

Author contribution statement

Xinjie Zhao and Shiwen Xu: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Jingrui Li: Analyzed and interpreted the data; Wrote the paper.

Hailin Yang, Li Tian, Fan Song and Wanzhi Cai: Contributed reagents, materials, analysis tools or data, Wrote the paper.

Zhonglong Lin: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Hu Li: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This work was supported by China National Tobacco Corporation of Science and Technology Major Project [110202001036[LS-05], National Natural Science Foundation of China [31922012].

Data availability statement

Data associated with this study has been deposited at the NCBI SRA database under the accession number SRR11568632.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgments

We sincerely thank Ling Ma and Bingyan Li for their constructive comments on the data analysis and Huayan Chen (Sun Yat-sen University) for providing the photo of *Aphidius gifuensis*.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e17070.

References

[1] S.L. Cameron, Insect mitochondrial genomics: implications for evolution and phylogeny, Annu. Rev. Entomol. 59 (2014) 95–117, https://doi.org/10.1146/annurev-ento-011613-162007.
[2] J.L. Boore, Animal mitochondrial genomes, Nucleic Acids Res. 27 (8) (1999) 1767–1780, https://doi.org/10.1093/nar/27.8.1767.
[3] J.L. Boore, W.M. Brown, Big trees from little genomes: mitochondrial gene order as a phylogenetic tool, Curr. Opin. Genet. 8 (6) (1998) 668–674, https://doi.org/10.1016/S0959-437X(98)80035-X.
[4] F. Berthier, M. Renaud, S. Alziari, R. Durand, RNA mapping on Drosophila mitochondrial DNA: precursors and template strands, Nucleic Acids Res. 14 (11) (1986) 4519–4533, https://doi.org/10.1093/nar/14.11.4519.
[5] J.B. Stewart, A.T. Beckenbach, Characterization of mature mitochondrial transcripts in Drosophila, and the implications for the tRNA punctuation model in arthropods, Gene 445 (1–2) (2009) 49–57, https://doi.org/10.1016/j.gene.2009.06.006.
[6] M. Roberti, F. Bruni, P.L. Polosa, M.N. Gadaleta, P. Cantatore, The Drosophila termination factor DmTTF regulates in vivo mitochondrial transcription, Nucleic Acids Res. 34 (7) (2006) 2109–2116, https://doi.org/10.1093/nar/gkl181.
[7] D. Ojala, J. Montoya, G. Attardi, tRNA punctuation model of RNA processing in human mitochondria, Nature 290 (5806) (1981) 470–474, https://doi.org/10.1038/290470a0.
[8] S. Gao, Y. Ren, Y. Sun, Z. Wu, J. Ruan, B. He, T. Zhang, X. Yu, X. Tian, W. Bu, PacBio full-length transcriptome profiling of insect mitochondrial gene expression, RNA Biol. 13 (9) (2016) 820–825, https://doi.org/10.1080/15476286.2016.1197481.
[9] M. Dowton, N.J.H. Campbell, Intramitochondrial recombination-is it why some mitochondrial genes sleep around? Trends Ecol. Evol. 16 (6) (2001) 269–271, https://doi.org/10.1016/s0169-5347(01)02182-6.
[10] R.F. Shao, E.F. Kirkness, S.C. Barker, The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse, Pediculus humanus, Genome Res. 19 (5) (2009) 904, https://doi.org/10.1101/gr.083188.108.
[11] S.L. Cameron, K.P. Johnson, M.F. Whiting, The mitochondrial genome of the screamer louse Bothriometopus (Phthiraptera: Ischnocera): effects of extensive gene rearrangements on the evolution of the genome, J. Mol. Evol. 65 (5) (2007) 589–604, https://doi.org/10.1007/s00239-007-9042-8.
[12] F. Song, H. Li, G. Liu, W. Wang, P. James, D.D. Colwell, A. Tran, S. Gong, W. Cai, R. Shao, Mitochondrial genome fragmentation unites the parasitic lice of eutherian mammals, Syst. Biol. 68 (3) (2019) 430–440, https://doi.org/10.1093/sysbio/syy062.
[13] M. Mao, A.D. Austin, N.F. Johnson, M. Dowton, Coexistence of minicircular and a highly rearranged mtDNA molecule suggests that recombination shapes mitochondrial genome organization, Mol. Biol. Evol. 31 (3) (2014) 636–644, https://doi.org/10.1093/molbev/mst255.
[14] Z. Feng, Y. Wu, C. Yang, X. Gu, J. Wilson, H. Li, W. Cai, H. Yang, F. Song, Evolution of tRNA gene rearrangement in the mitochondrial genome of ichneumonoid wasps (Hymenoptera: Ichneumonoidea), Int. J. Biol. Macromol. 164 (2020) 540–547, https://doi.org/10.1016/j.ijbiomac.2020.07.149.
[15] A. Rhoads, K.F. Au, PacBio sequencing and its applications, Dev. Reprod. Biol. 13 (5) (2015) 278–289, https://doi.org/10.1016/j.gpb.2015.08.002.
[16] S. Gordon, E. Tseng, A. Salamov, J. Zhang, X. Meng, Z. Zhao, D. Kang, J. Underwood, I.V. Grigoriev, M. Figueroa, et al., Widespread polycistronic transcripts in mushroom-forming fungi revealed by single-molecule long-read mRNA sequencing, PLoS One 10 (7) (2015), e0132628, https://doi.org/10.1371/journal.pone.0132628.
[17] J. Liu, H. Jiang, J. Zan, Y. Bao, J. Dong, L. Xiong, L. Nie, Single-molecule long-read transcriptome profiling of Platysternon megacephalum mitochondrial genome with gene rearrangement and control region duplication, RNA Biol. 15 (9) (2018) 1244–1249, https://doi.org/10.1080/15476286.2018.1521212.
[18] C. Zuo, M. Blow, A. Sreedasyam, R.C. Kuo, G.K. Ramamoorthy, I. Torres-Jerez, G. Li, M. Wang, D. Dilworth, K. Barry, et al., Revealing the transcriptomic complexity of switchgrass by PacBio long-read sequencing, Biotechnol. Biofuels 11 (2018) 170, https://doi.org/10.1186/s13068-018-1167-z.
[19] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, et al., Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data, Bioinformatics 28 (2012) 1647–1649, https://doi.org/10.1093/bioinformatics/bts199.
[20] V.M. Margam, B.S. Coates, R.L. Hellmich, T. Agunbiade, M.J. Seufferheld, W. Sun, M.N. Ba, A. Sanon, C.L. Binso-Dabire, I. Baoua, et al., Mitochondrial genome sequence and expression profiling for the legume pod borer Maruca vitrata (Lepidoptera: Crambidae), PLoS One 6 (2) (2011), e16444, https://doi.org/10.1371/journal.pone.0016444.
[21] H. Wang, J. Yang, L. Boykin, Q. Zhao, Q. Li, X. Wang, S. Liu, The characteristics and expression profiles of the mitochondrial genome for the Mediterranean species of the Bemisia tabaci complex, BMC Genom. 14 (2013) 401, https://doi.org/10.1186/1471-2164-14-401.
[22] I. Kuznetsova, S.J. Siira, A.J. Shearwood, J.A. Ermer, A. Filipovska, O. Rackham, Simultaneous processing and degradation of mitochondrial RNAs revealed by circularized RNA sequencing, Nucleic Acids Res. 45 (9) (2017) 5487–5500, https://doi.org/10.1093/nar/gkx104.
[23] D.A. Clayton, Transcription and replication of mitochondrial DNA, Hum. Reprod. 15 (2000) 11–17, https://doi.org/10.1093/humrep/15.suppl_2.11.
[24] S. Gao, X. Tian, H. Chang, Y. Sun, Z. Wu, Z. Cheng, P. Dong, Q. Zhao, J. Ruan, W. Bu, Two novel lncRNAs discovered in human mitochondrial DNA using PacBio full-length transcriptome data, Mitochondrion 38 (2018) 41–47, https://doi.org/10.1016/j.mito.2017.08.002.
[25] E. Yus, M. Güell, A.P. Vivancos, W. Chen, M. Lluch-Senar, J. Delgado, A. Gavin, P. Bork, L. Serrano, Transcription start site associated RNAs in bacteria, Mol. Syst. Biol. 8 (2012) 585, https://doi.org/10.1038/msb.2012.16.
[26] I. Kuznetsova, S.J. Siira, A.J. Shearwood, J.A. Ermer, A. Filipovska, O. Rackham, Simultaneous processing and degradation of mitochondrial RNAs revealed by circularized RNA sequencing, Nucleic Acids Res. 45 (9) (2017) 5487–5500, https://doi.org/10.1093/nar/gkx104.
[27] X. Jin, Z. Cheng, B. Wang, T.O. Yau, Z. Chen, S.C. Barker, D. Chen, W. Bu, D. Sun, S. Gao, Precise annotation of human, chimpanzee, rhesus macaque and mouse mitochondrial genomes leads to insight into mitochondrial transcription in mammals, RNA Biol. 17 (3) (2020) 395–402, https://doi.org/10.1080/15476286.2019.1709746.
[28] D. Ojala, C. Merkel, R. Gelfand, G. Attardi, The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA, Cell 22 (2) (1980) 393–403, https://doi.org/10.1016/0092-8674(80)90350-5.

[29] W. Rossmanith, P. Of, Z: mitochondrial tRNA processing enzymes, Biochim. Biophys. Acta 1819 (9–10) (2012) 1017–1026, https://doi.org/10.1016/j.bbagrm.2011.11.003.
[30] R.J. Temperley, S.H. Seneca, K. Tonska, E. Bartnik, L.A. Bindoff, R.N. Lightowlers, Z.M. Chrzanowska-Lightowlers, Investigation of a pathogenic mtDNA microdeletion reveals a translation-dependent deadenylation decay pathway in human mitochondria, Hum. Mol. Genet. 12 (18) (2003) 2341–2348, https://doi.org/10.1093/hmg/ddg238.
[31] A. Bratic, P. Clemente, J. Calvo-Garrido, C. Maffezzini, A. Felser, R. Wibom, A. Wedell, C. Freyer, A. Wredenberg, Mitochondrial polyadenylation is a one-step process required for mRNA integrity and tRNA maturation, PLoS Genet. 12 (5) (2016), e1006028, https://doi.org/10.1371/journal.pgen.1006028.
[32] B.K. Mohanty, V.F. Maples, S.R. Kushner, The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in Escherichia coli, Mol. Microbiol. 54 (4) (2004) 905–920, https://doi.org/10.1111/j.1365-2958.2004.04337.x.
[33] S. Slomovic, D. Laufer, D. Geiger, G. Schuster, Polyadenylation and degradation of human mitochondrial RNA: the prokaryotic past leaves its mark, Mol. Cell Biol. 25 (15) (2005) 6427–6435, https://doi.org/10.1128/MCB.25.15.6427-6435.2005.