Zootaxa 3620 (2): 260–272
www.mapress.com/zootaxa/

Article

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

http://dx.doi.org/10.11646/zootaxa.3620.2.4
http://zoobank.org/urn:lsid:zoobank.org:pub:9E3522DE-93F4-4EB6-98F9-704528FCEC18

The complete mitochondrial genome of a tessaratomid bug, *Eusthenes cupreus* (Hemiptera: Heteroptera: Pentatomomorpha: Tessaratomidae)

WEN SONG[1], HU LI[1], FAN SONG[1], LI LIU[1], PEI WANG[1], HUAIZHU XUN[1], XUN DAI[2], JIAN CHANG[2] & WANZHI CAI[1, 3]

[1]*Department of Entomology, China Agricultural University, Yuanmingyuan West Road, Beijing 100193, China*
[2]*Research Center of Hongta Group, Yuxi, Yunnan 653100, China*
[3]*Corresponding author. E-mail: caiwz@cau.edu.cn*

Abstract

The 16, 299 bp long mitochondrial genome (mitogenome) of a tessaratomid bug, *Eusthenes cupreus* (Westwood), is reported and analyzed. The mitogenome represents the first sequenced complete mitogenome of the heteropteran family Tessaratomidae. The mitogenome of *E. cuopreus* is a typical circular DNA molecule with a total AT content of 74.1%, and contains 13 protein-coding genes (PCGs), 22 transfer RNA (tRNA) genes, two ribosomal RNA (rRNA) genes, and a control region. The gene arrangement is identical with the most common type in insects. Most PCGs start with the typical ATN codon, except that the initiation codon for *COI* is TTG. All tRNAs possess the typical clover-leaf structure, except $tRNA^{Ser\ (AGN)}$, in which the dihydrouridine (DHU) arm forms a simple loop. Six domains with 45 helices and three domains with 27 helices are predicted in the secondary structures of *rrnL* and *rrnS*, respectively. The control region is located between *rrnS* and $tRNA^{Ile}$, including some short microsatellite repeat sequences. In addition, three different repetitive sequences are found in the control region and the $tRNA^{Ile}$-$tRNA^{Gln}$-$tRNA^{Met}$-*ND2* gene cluster. One of the unusual features of this mitogenome is the presence of one $tRNA^{Gln}$-like sequence in the control region. This extra $tRNA^{Gln}$-like sequence is 73 bp long, and the anticodon arm is identical to that of the regular $tRNA^{Gln}$.

Key words: Mitochondrial genome, *Eusthenes cupreus*, RNA secondary structure, $tRNA^{Gln}$-like sequence

Introduction

The mitogenome of insects is a typically double-stranded, circular molecule, and commonly includes 13 PCGs, 22 tRNA genes, 2 rRNA genes (*rrnL* and *rrnS*), and one non-coding region known as the control region which plays a role in initiation of transcription and replication (Wolstenholme 1992; Boore 1999). The mitogenome is becoming increasingly important for the study of population genetics and molecular evolution and has been widely regarded as the molecular maker for the phylogenetic analysis in metazoans because of its relatively simple genetic structure, high rate of evolution, low or absence of sequence recombination, and evolutionary conserved gene products (Lin *et al*. 2004; Gissi *et al*. 2008). In addition, the comparative study of mitogenome sequences can give us a better understanding of the genome structure, gene arrangement, and the evolution of arthropod lineages (Boore 1999; Shao *et al*. 2001; Hwang *et al*. 2001; Nardi *et al*. 2003).

Tessaratomidae is a family of true bugs with approximately 240 species and 55 genera (Rolston *et al*. 1994; Rider 2006). All tessaratomids are large to extremely large (often over 20 mm, some longer than 40 mm), robustly ovate or elongate, and phytophagous. They generally feed upon plants belonging to the plant orders Rosales and Sapindales, and spend most of their lives in tree leaves and stems. Many species are of economic importance as agricultural pests, such as the litchi stink bug, *Tessaratoma papillosa* (Westwood), which are destructive pests of litchi trees in China (Cassis & Gordon 2002, http://en.wikipedia.org/wiki/Tessaratomidae - cite_ref-cassisgross_7-0#cite_ref-cassisgross_7-0). A few species are also consumed as human food in some countries, such as the edible stink bug *Encosternum delegorguei* Spinola, which is a well known food in Zimbabwe and among the Venda people of South Africa (Dzerefos *et al.* 2002).

 Accepted by C. Schaefer: 4 Jan. 2013; published: 6 Mar. 2013

Up to now, mitogenome sequences of Tessaratomidae have not been reported. This lack of mitogenome data impedes our extraordinary attention to the evolutionary study in this group. In the current study, we determined the complete mitogenome of a fruit tree pest tessaratomid bug, *Eusthenes cupreus* (Westwood), and analyze the mitogenomic architecture, such as nucleotide composition, genomic arrangement, codon usage, and RNA secondary structure.

Materials and methods

Samples and DNA extraction. Adult specimens of *E. cupreus* were collected in Yaoqu, Yunnan Province, China, in May 2009. All specimens were initially preserved in 95% ethanol in the field, and transferred to -20°C for long-term storage. The genomic DNA was extracted from one male adult's muscle tissues of the thorax using a CTAB-based protocol (Aljanabi *et al*. 1997). The voucher specimen (No. VHem-00225) was deposited at the Entomological Museum of China Agricultural University (Beijing).

PCR amplification and sequencing. The complete mitogenome was amplified by overlapping PCR fragments (Table 1) using a range of universal insect mitochondrial primers (Simon *et al*., 2006). Species-specific primers were designed based on the sequenced fragments to bridge gaps. PCR and sequencing reactions were conducted following Li *et al*. (2012).

Annotation and bioinformatics analysis. PCGs and rRNAs were initially identified using BLAST searches in GenBank and subsequently by alignment with genes of other true bugs. tRNAs were identified by tRNAscan-SE Search Server v.1.21 (Lowe & Eddy 1997). Some tRNA genes that could not be determined by tRNAscan-SE were determined in the unannotated regions by sequence similarity to tRNAs of other heteropterans (Dotson & Beard 2001; Li *et al*. 2011, 2012a, 2012b). The base composition, codon usage, and nucleotide substitution were analyzed with Mega 5.0 (Tamura *et al*. 2011). AT-skew = (A-T)/(A+T) and GC-skew = (G-C)/(G+C) were used to measure the base-compositional difference between genes (Perna & Kocher 1995).

Construction of secondary structures of RNAs and non-coding regions. Secondary structures of the small and large subunits of rRNAs were inferred using models predicted for other insects (Gillespie *et al*. 2006; Zhou *et al*. 2007; Cameron & Whiting 2008; Li *et al*. 2011). Stem-loops were named according to the convention of (Gillespie *et al*. 2006; Cameron & Whiting 2008). Regions lacking significant homology and other non-coding regions were predicted using Mfold (Zuker 2003).

Results and discussion

Genome organization. The complete mitogenome of *E. cupreus* is a typically double-stranded, circular molecular, 16, 229 bp in size (GenBank accession number: JQ910983). It contains the typically gene content (13 PCGs, 22 tRNA genes, two rRNA genes, and a control region) (Fig. 1, Table 2). The orientation and gene order is the same as that in *Drosophila yakuba* (Clary *et al*. 1985), which has been hypothesized to be the ancestral arrangement for insects (Boore 1999). The majority-coding strand (J-strand) encodes 23 genes, and the other 14 genes are oriented on the minority-coding strand (N-stand). Gene overlaps are found at 12 locations and involve a total of 79 bp. The longest overlap is 44 bp and is located between *ATP6* and *COIII*. In addition, this mitogenome harbors 162 bp of intergenic spacer sequences, which are spread over 15 regions, ranging in the size from 1-55 bp. The longest intergenic spacer sequence is located between $tRNA^{Ile}$ and $tRNA^{Gln}$.

Protein-coding genes. The total length of all 13 PCGs is 11, 801 bp, accounting for 68.3% of the entire length of *E. cupreus* mitogenome. The overall AT content of PCGs is 73.6 %. Start codons of most PCGs are ATN, with the exception represented by the TTG start codon of *ND6* and *COI*, and the GTG start codon of *ND1*. The majority of PCGs has a complete termination codon of TAA (*ND2*, *ATP8*, *ND5*, *ND4*, *ND4L*, *ND6*, and *ND1*) and TAG (*ATP6*, *ND3*, and *CytB*), while the remaining three PCGs use the incomplete termination codons, TA (*COI*) and T (*COII* and *COIII*) (Table 2). The incomplete stop codons are presumably completed by post-transcriptional polyadenylation (Ojala *et al*. 1981).

TABLE 1. The primers used in this study.

No. of fragment[a]	Primer name	Nucleotide sequence (5'-3')	Reference
1	TI-J34	GCCTGATAAAAAGGRTTAYYTTGATA	Simon *et al.*, 2006
	C1-N1738	TTTATTCGTGGRAATGCYATRTC	Simon *et al.*, 2006
2	TW-J1301	GTTAAWTAAACTAATARCCTTCAAA	Simon *et al.*, 2006
	C1-N2776	GGTAATCAGAGTATCGWCGNGG	Simon *et al.*, 2006
3	F2100	AATTGGWGGWTTYGGAAAYTG	Simon *et al.*, 2006
	R3000	GGTAATCAGAGTATCGWCGNGG	Simon *et al.*, 2006
4	C1-J2756	ACATTTTTTCCTCAACATTT	Simon *et al.*, 2006
	C2-N3665	CCACAAATTTCTGAACACTG	Simon *et al.*, 2006
5	C2-J3399	TCTATTGGTCATCAATGGTACTG	Simon *et al.*, 2006
	A8-N4061	GAAAATAAATTTGTTATCATTTTCA	Simon *et al.*, 2006
6	F4050	GCTCCTTTATGATGAGAAAG	Present study
	R4740	TCCTAACTTCAATCCTGATG	Present study
7	A6-J4463	TTTGCCCATCTWGTWCCNCAAGG	Simon *et al.*, 2006
	C3-N5460	TCAACAAAATGTCARTAYCA	Simon *et al.*, 2006
8	C3-J4792	GTTGATTATAGACCWTGRCC	Simon *et al.*, 2006
	N3-N5731	TTAGGGTCAAATCCRCAYTC	Simon *et al.*, 2006
9	C3-J5470	GCAGCTGCYTGATAYTGRCA	Simon *et al.*, 2006
	TN-N6160	TCAATTTTRTCATTAACAGTGA	Simon *et al.*, 2006
10	F5670	TGAGGTAGATATCCTTTTAG	Present study
	R7470	ATCTTGTTGGGTTGAGATGG	Present study
11	N5-J7572	AAAGGGAATTTGAGCTCTTTT	Simon *et al.*, 2006
	N4-N8727	AAATCTTTRATTGCTTATTCWTC	Simon *et al.*, 2006
12	F8530	CTACGACTATGAGTACGTTC	Present study
	R10460	CCGTTTGCATGGGTATATCG	Present study
13	CB-J10621	CTCATACTGATGAAATTTTGGTTC	Simon *et al.*, 2006
	CB-N11526	TTCTACTGGTCGTGCTCCAATTCA	Simon *et al.*, 2006
14	CB-J11335	CATATTCAACCAGAATGATA	Simon *et al.*, 2006
	N1-N12067	AATCGTTCTCCATTTGATTTTGC	Simon *et al.*, 2006
15	F12030	AAGAAGTAAGATCACATCCC	Present study
	R12261	GTGGTGCTTGTATCCTTATG	Present study
16	N1-J12261	TACCTCATAAGAAATAGTTTGAGC	Simon *et al.*, 2006
	LR-N13000	TTACCTTAGGGATAACAGCGTAA	Simon *et al.*, 2006
17	LR-J12888	CCGGTTTGAACTCARATCATGTAA	Simon *et al.*, 2006
	LR-N13889	ATTTATTGTACCTTKTGTATCAG	Simon *et al.*, 2006
18	SR-J13342	CCTTCGCACRGTCAAAATACYGC	Simon *et al.*, 2006
	SR-N14220	ATATGYACAYATCGCCCGTC	Simon *et al.*, 2006
19	LR-J14197	GTAAAYCTACTTTGTTACGACTT	Simon *et al.*, 2006
	SR-N14745	GTGCCAGCAAYCGCGGTTATAC	Simon *et al.*, 2006
20	SR-J14610	ATAATAGGGTATCTAATCCTAGT	Simon *et al.*, 2006
	TM-N200	ACCTTTATAARTGGGGTATGARCC	Simon *et al.*, 2006

[a]The orientation is as shown in Fig. 1.

TABLE 2. Organization of *E. cupreus* mitogenome.

Gene	Direction	Location	Size (bp)	Anticodon	Codon Start	Stop	Intergenic nucleotides *
tRNA^{Ile}	F	1–66	66	32–34 GAT			
tRNA^{Gln}	R	122–191	69	158–160 TTG			55
tRNA^{Met}	F	206–273	68	236–238 CAT			14
ND2	F	274–1257	984		ATA	TAA	0
tRNA^{Trp}	F	1256–1324	69	1289–1291 TCA			-2
tRNA^{Cys}	R	1317–1380	64	1348–1350 GCA			-8
tRNA^{Tyr}	R	1388–1451	64	1418–1420 GTA			7
COI	F	1456–2990	1535		TTG	TA-	4
tRNA$^{Leu(UUR)}$	F	2991–3057	67	3020–3022 TAA			0
COII	F	3058–3736	679		ATA	T-	0
tRNA^{Lys}	F	3737–3810	74	3771–3773 CTT			0
tRNA^{Asp}	F	3814–3878	65	3844–3846 GTC			3
ATP8	F	3885–4043	159		ATG	TAA	6
ATP6	F	4037–4747	711		ATG	TAG	-7
COIII	F	4704–5490	787		ATG	T-	-44
tRNA^{Gly}	F	5491–5555	65	5521–5523 TCC			-3
ND3	F	5553–5909	357		ATA	TAG	14
tRNA^{Ala}	F	5924–5992	69	5952–5954 TGC			-1
tRNA^{Arg}	F	5994–6057	64	6023–6025 TCG			1
tRNA^{Asn}	F	6070–6139	70	6102–6104 GTT			12
tRNA$^{Ser(AGN)}$	F	6139–6211	73	6169–6171 GCT			-1
tRNA^{Glu}	F	6211–6275	65	6242–6244 TTC			-1
tRNA^{Phe}	R	6274–6339	66	6306–6308 GAA			-2
ND5	R	6339–8051	1713		ATA	TAA	-1
tRNA^{His}	R	8053–8114	62	8082–8084 GTG			1
ND4	R	8118–9443	1326		ATG	TAA	3
ND4L	R	9437–9718	282		ATT	TAA	-7
tRNA^{Thr}	F	9730–9795	66	9760–9762 TGT			11
tRNA^{Pro}	R	9796–9858	63	9826–9828 TGG			0
ND6	F	9872–10357	486		TTG	TAA	13
CytB	F	10359–11495	1137		ATG	TAG	1
tRNA$^{Ser(UCN)}$	F	11494–11562	69	11524–11526 TGA			-2
ND1	R	11580–12506	927		GTG	TAA	17
tRNA$^{Leu(CUN)}$	R	12507–12573	67	12540–12542 TAG			0
lrRNA	R	12574–13849	1276				0
tRNA^{Val}	R	13850–13916	67	13886–13888 TAC			0
srRNA	R	13917–14709	793				0
Control region		14710–16229	1520				0

* Negative numbers indicate that adjacent genes overlap.

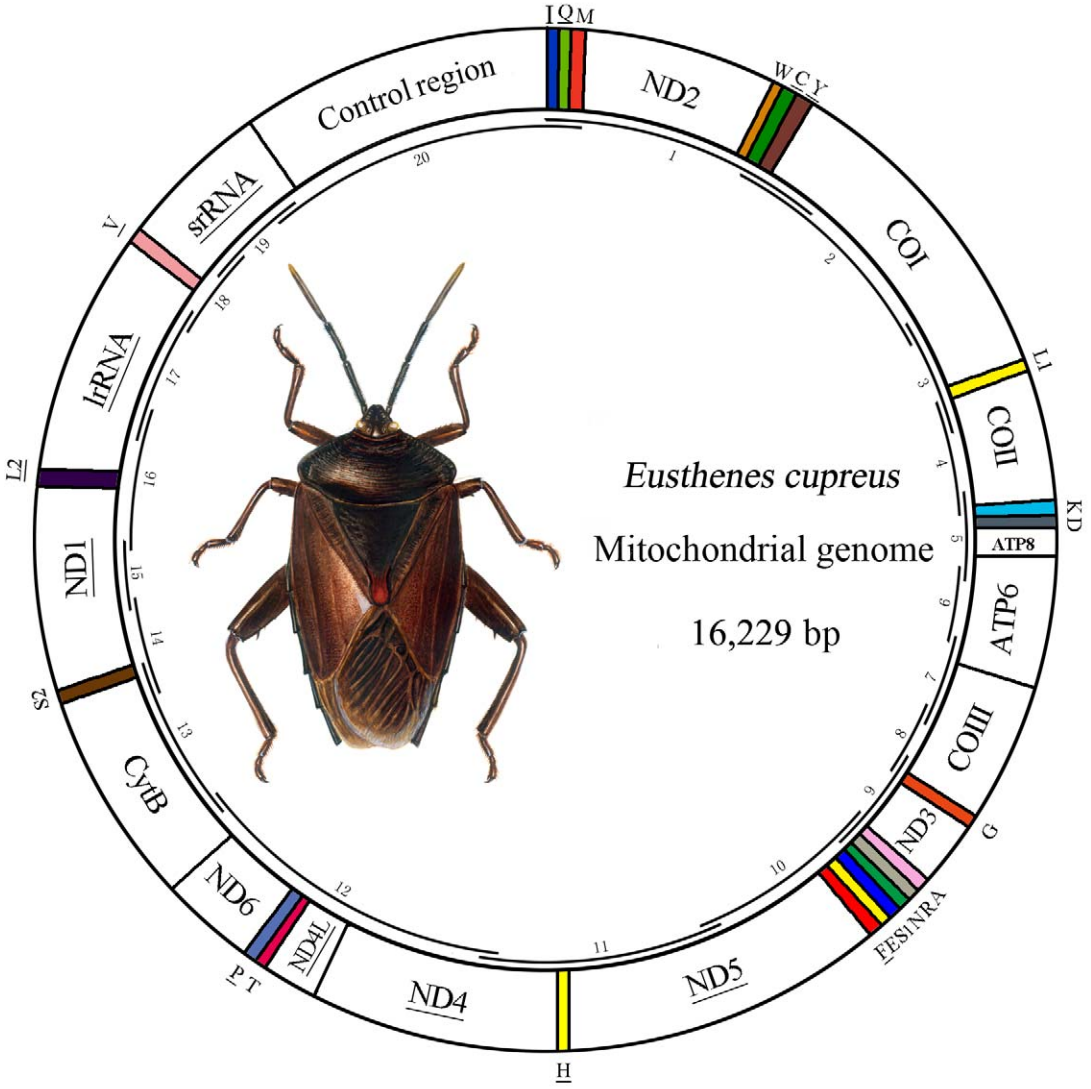

FIGURE 1. Map of the mitogenome of *E. cupreus.* The tRNAs are denoted by the color blocks and are labeled according to the IUPAC-IUB single-letter amino acid codes. Gene name without underline indicates the direction of transcription from left to right, and with underline indicates right to left. Overlapping lines within the circle denote PCR fragments amplified used for cloning and sequencing.

tRNA and rRNA genes. Twenty-two tRNAs are found in *E. cupreus* mitogenome, ranging in size from 62 to 74 bp. All tRNAs can be folded into the typical clover-leaf structure, with the exception of $tRNA^{Ser(AGN)}$, the dihydrouridine (DHU) arm of which forms a simple loop (14 bp) (Fig. 2). This phenomenon is common in the mitogenomes of true bugs (Li *et al*. 2012a, 2012b), and is also considered as a typical feature in metazoan mtDNAs (Lavrov *et al*. 2004). In most tRNAs, the amino acid acceptor (AA) arms and the anticodon (AC) arms are more conservative, except for $tRNA^{Ser\ (AGN)}$ which possessed a long optimal base pairing (9 bp in contrast to the normal 5) and a bulged nucleotide in the middle for the AC stem. The lengths of the DHU and TΨC (T) stems (2–5 bp) and loops (3–9 bp) are more variable (Fig. 2).

Based on the secondary structure, a total of 17 unmatched base pairs are found in the *E. cupreus* tRNAs. Unmatched base pairs are present in stems of 10 tRNAs, including 13 bp G-U, 3 bp U-U, and 1 bp A-C mismatches. A post-transcriptional RNA editing mechanism has been proposed to correct the errors like the mismatches, a bulged nucleotide (U) in the stem, abnormal loops and arms, and maintain the function of these tRNAs (Tomita *et al*. 2001; Lavrov *et al*. 2004; Zhang *et al*. 2008).

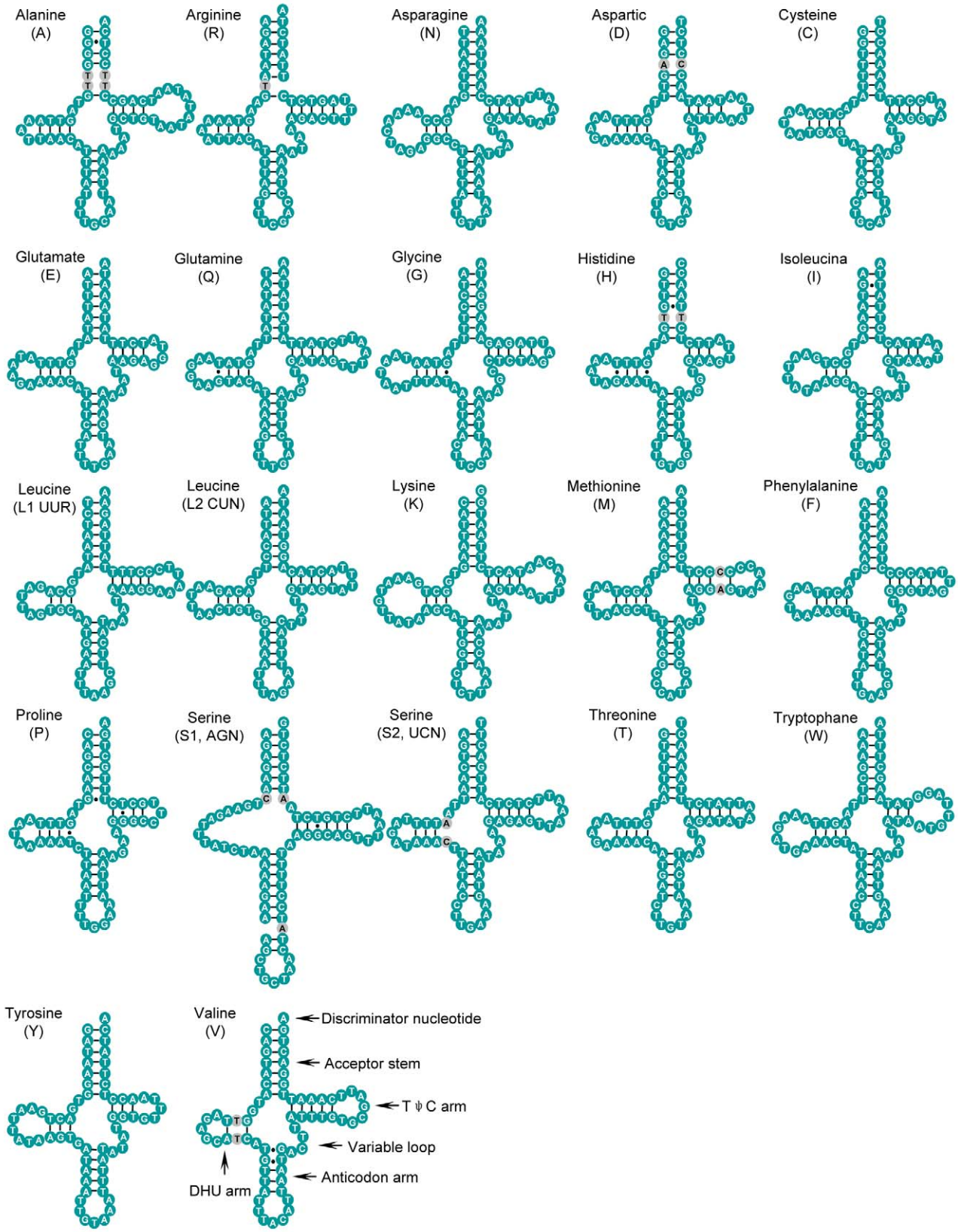

FIGURE 2. Inferred secondary structure of 22 tRNAs of the *E. cupreus* mitogenome. The tRNAs are labeled with the abbreviations of their corresponding amino acids (gray indicates the mismatches). Dashed (-) indicate Watson-Crick base pairing and (•) indicate G-U base pairing.

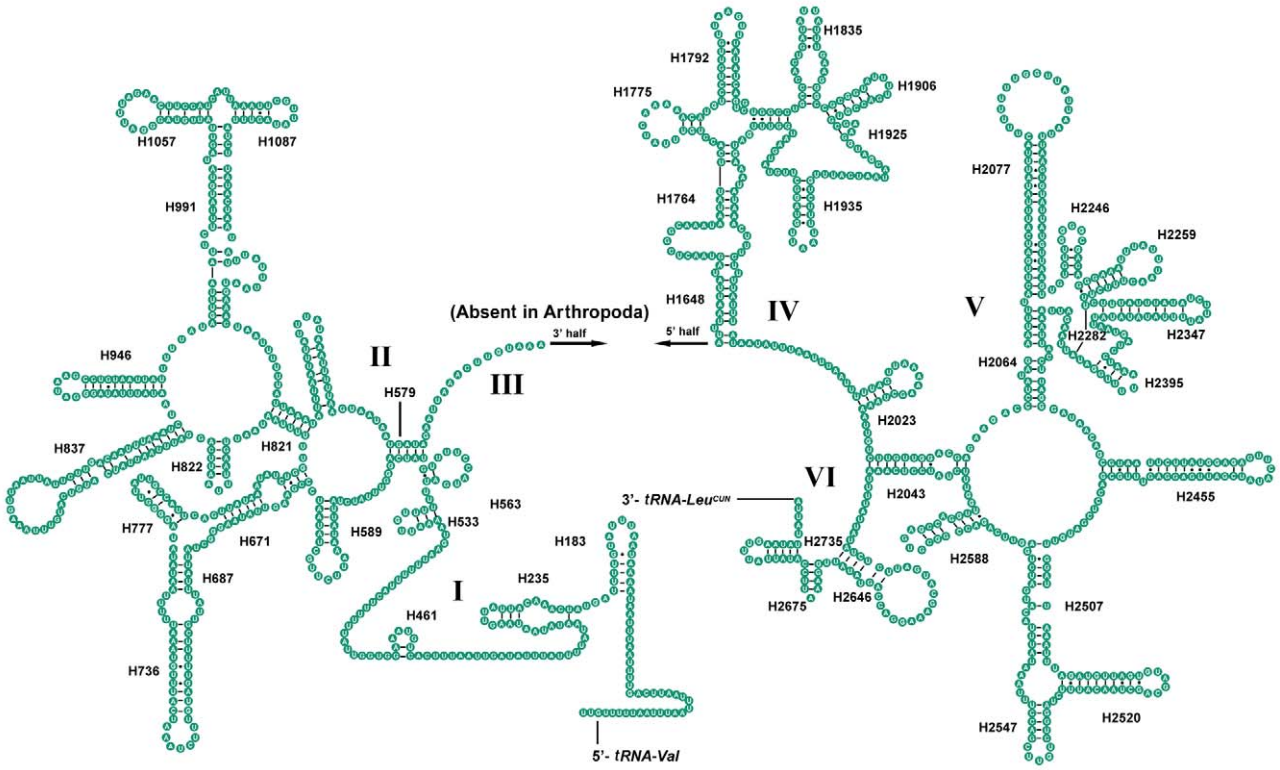

FIGURE 3. Predicted secondary structure of the *rrnL* gene in *E. cupreus*. Roman numerals denote the conserved domain structure. The numbering system follows Gillespie *et al.* (2006) (established at the Comparative RNA Website). Dashed (–) indicate Watson-Crick base pairing and dot (•) indicate G-U base pairing.

The large and small ribosomal RNAs are located respectively between $tRNA^{Leu(CUN)}$, and $tRNA^{Val}$, $tRNA^{Val}$, and control region (Fig. 1). The length of the *rrnL* is 1, 276 bp, and the *rrnS* is 793 bp. The secondary structure of both *rrnL* and *rrnS* can be predicted by the models of other insects (Cannone *et al.* 2002; Gillespie *et al.* 2006; Zhou *et al.* 2007; Cameron & Whiting 2008). The secondary structure of *rrnL* consists of six structural domains and 45 helices (Fig. 3), and the *rrnS* includes three structural domains and 27 helices (Fig. 4).

TABLE3. The nucleotide composition of *E. cupreus* mitogenome.

Feature	T(U)	C	A	G	AT%	AT Skew	GC Skew
Whole genome	32.1	14.3	42.0	11.6	74.1	0.133	-0.103
Protein–coding genes	40.6	13.1	33.0	13.3	73.6	-0.103	0.010
First codon position	33	12.0	34.8	20.1	67.9	0.027	0.250
Second codon position	46	19.0	20.4	14.9	66.1	-0.382	-0.119
Third codon position	43	8.2	43.7	5.0	86.7	0.008	-0.241
Protein–coding genes J-strand	35.1	14.5	37.3	13.2	72.3	0.031	-0.049
Protein–coding genes N-strand	49.5	10.8	26.1	13.6	75.6	-0.309	0.117
tRNA genes	36.7	11.3	37.7	14.3	74.4	0.013	0.120
tRNA genes J-strand	34.6	12.8	39.8	12.9	74.4	0.069	0.005
tRNA genes N-strand	49.5	10.8	26.1	13.6	75.6	-0.309	0.117
rRNA genes	31.5	14.6	44.5	9.5	75.9	0.171	-0.213
Control region	33.6	14.8	41.1	10.5	74.7	0.101	-0.169

Nucleotide composition and codon usage. The mitogenome of *E. cupreus* is typically biased toward A and T with an average AT content of 74.1% (A = 42.0%, T = 32.1%, C = 14.3 %, G = 11.6%) (Table 3). The AT content of PCGs, tRNAs, and rRNAs is 74.7%, 75.6%, and 76.2%, respectively. Asymmetry in the nucleotide composition between J-strand and N-strand was observed in this mitogenome and was a common phenomenon in the Metazoa (Perna & Kocher 1995). The PCGs and tRNAs in J-strand displayed positive AT-skew and obviously negative GC-skew, whereas the N-strand showed negative AT-skew and nearly equal G and C. This feature was probably related to the asymmetrical directional mutation pressure (Min & Hickey 2007).

The AT bias is also reflected in the codon usage. Analysis of base composition at each codon position of the concatenated 13 PCGs showed that the AT content of third codon position (81.2%) is higher than the first (68.5%) and second (66.3%) codon positions (Table 3). NNA and NNC codons are more frequent than NNU and NNG in the J-strand PCGs, whereas the N-strand genes are exactly the opposite. In addition, the most frequently used codons are the AT-rich codons, such as TTA, ATT, ATA, TTT, AAT, and TAT (Table 4).

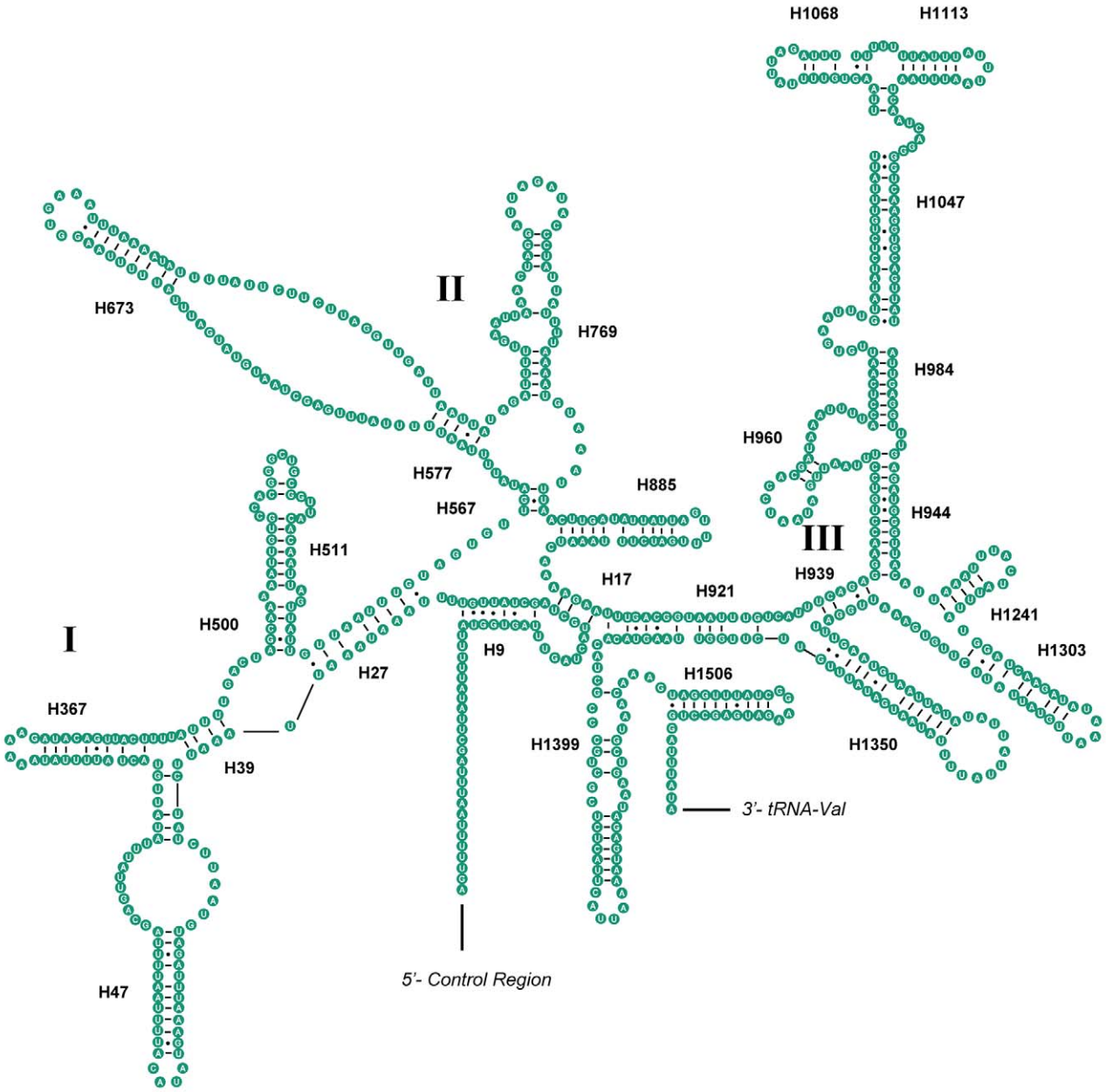

FIGURE 4. Predicted secondary structure of the *rrnS* gene in the *E. cupreus*. Roman numerals denote the conserved domain structure. Dashed (-) indicate Watson-Crick base pairing and dot (•) indicate G-U base pairing. Structural annotations follow Fig. 3.

TABLE 4. Codon usage of *E. cupreus* mitochondrial PCGs.

Amino acid	Codon	N	RSCU	N+	RSCU	N-	RSCU
Phe (F)	UUU	243	1.59	103	1.36	140	1.83
	UUC	62	0.41	49	0.64	13	0.17
Leu (L)	UUA	350	4.19	205	4.44	145	3.88
	UUG	34	0.41	9	0.19	25	0.67
	CUU	59	0.71	20	0.43	39	1.04
	CUC	5	0.06	1	0.02	4	0.11
	CUA	47	0.56	40	0.87	7	0.19
	CUG	6	0.07	2	0.04	4	0.11
Ile (I)	AUU	312	1.68	191	1.61	121	1.81
	AUC	59	0.32	46	0.39	13	0.19
Met (M)	AUA	268	1.76	196	1.84	72	1.58
	AUG	36	0.24	17	0.16	19	0.42
Val (V)	GUU	85	1.65	25	0.83	60	2.82
	GUC	12	0.23	10	0.33	2	0.09
	GUA	100	1.94	81	2.68	19	0.89
	GUG	9	0.17	5	0.17	4	0.19
Ser (S)	UCU	95	2.17	26	1.04	69	3.66
	UCC	14	0.32	8	0.32	6	0.32
	UCA	99	2.26	71	2.84	28	1.48
	UCG	2	0.05	2	0.08	0	0
Pro (P)	CCU	68	1.97	42	1.66	26	2.81
	CCC	21	0.61	13	0.51	8	0.86
	CCA	47	1.36	44	1.74	3	0.32
	CCG	2	0.06	2	0.08	0	0
Thr (T)	ACU	80	1.66	46	1.29	34	2.72
	ACC	23	0.48	19	0.53	4	0.32
	ACA	87	1.8	76	2.13	11	0.88
	ACG	3	0.06	2	0.06	1	0.08
Ala (A)	GCU	69	1.74	32	1.19	37	2.9
	GCC	11	0.28	9	0.33	2	0.16
	GCA	78	1.96	67	2.48	11	0.86
	GCG	1	0.03	0	0	1	0.08
Tyr (Y)	UAU	154	1.78	65	1.76	89	1.8
	UAC	19	0.22	9	0.24	10	0.2
Stop (*)	UAA	7	1.4	3	0.5	4	2
	UAG	3	0.6	3	0.5	0	0
His (H)	CAU	50	1.37	37	1.23	13	2
	CAC	23	0.63	23	0.77	0	0
Gln (Q)	CAA	52	1.68	41	1.82	11	1.29
	CAG	10	0.32	4	0.18	6	0.71
Asn (N)	AAU	155	1.74	94	1.63	61	1.94
	AAC	23	0.26	21	0.37	2	0.06

.....continued on the next page

TABLE 4. (Continued)

Amino acid	Codon	N	RSCU	N+	RSCU	N-	RSCU
Lys (K)	AAA	76	1.52	59	1.71	17	1.1
	AAG	24	0.48	10	0.29	14	0.9
Asp (D)	GAU	62	1.72	37	1.61	25	1.92
	GAC	10	0.28	9	0.39	1	0.08
Glu (E)	GAA	69	1.59	52	1.76	17	1.21
	GAG	18	0.41	7	0.24	11	0.79
Cys (C)	UGU	31	1.59	8	1.23	23	1.77
	UGC	8	0.41	5	0.77	3	0.23
Trp (W)	UGA	96	1.94	68	2	28	1.81
	UGG	3	0.06	0	0	3	0.19
Arg (R)	CGU	17	1.24	5	0.57	12	2.4
	CGC	3	0.22	1	0.11	2	0.4
	CGA	28	2.04	24	2.74	4	0.8
	CGG	7	0.51	5	0.57	2	0.4
Ser (S)	AGU	40	0.91	15	0.6	25	1.32
	AGC	3	0.07	3	0.12	0	0
	AGA	98	2.23	75	3	23	1.22
	AGG	0	0	0	0	0	0
Gly (G)	GGU	69	1.27	15	0.44	54	2.63
	GGC	8	0.15	2	0.06	6	0.29
	GGA	112	2.06	100	2.96	12	0.59
	GGG	28	0.52	18	0.53	10	0.49

N Total number in all PCGs, N+ total number in J-strand, N- total number in N-strand, RSCU relative synonymous codon usage.

Control region. The control region of *E. cupreus* is 1, 520 bp long, locates between *rrnS* and $tRNA^{Ile}$, and contains 74.7% AT nucleotides (Fig. 5A). The stem-loop structures in the control region have been suggested as the site of the initiation of secondary strand-replication (Zhang *et al*. 1995; Clary & Wolstenholme 1987), and sequences flanking the stem-and-loop structure are highly conserved among several insect orders, possessing "TATA" consensus sequences at the 5′ end and "G(A)nT" consensus sequences at the 3′ end (Schultheis *et al*. 2002; Zhang & Hewitt 1997). Within the control region of *E. cupreus* mitogenome, three sequence stretches have the potential to form stem-loop structures (Fig. 5B).

The presence of tandem repeats in the control region has been frequently reported in many other insects and replication slippage is regarded as a dominant mechanism accounting for the existence of tandem repeats (Oliveira *et al*. 2008; Ojala *et al*. 1981). However, such tandem repeats are not found in the control region of *E. cupreus* mitogenome. In this control region, some short repeats are present, which can be considered as microsatellite such as $(TTAATAATAATA)_2$, $(AATAATTTATT)_2$, $(TTAATAAATAAAA)_2$, $(TA)_7$, and $(AT)_5$.

In addition, three different repetitive sequences are found in the control region and $tRNA^{Ile}$-$tRNA^{Gln}$-$tRNA^{Met}$-*ND2* gene cluster (Fig. 5A). First repetitive sequences have two 52 bp long repeat units; one repeat unit is located at the control region (15, 464–15, 515), and the other is found between $tRNA^{Ile}$ *and* $tRNA^{Gln}$ (66–117). Second repetitive sequences have two 26 bp long repeat units; one repeat unit is located at the downstream of the first repetitive sequences in the control region (15, 535–15, 560), and the other is the partial sequence of $tRNA^{Gln}$ (143–168). Third repetitive sequences have two 80 bp long repeat units; one repeat unit is located at the downstream of the second repetitive sequences in the control region (15, 606–15, 685), and the other is the partial sequence of *ND2* (318–397).

One of the unusual features of *E. cupreus* mitogenome is the presence of one *tRNA*^{Gln}-like sequence in the control region (Fig. 5C). The presence of tRNA-like sequences within the control region has also been reported in other insects (Cha *et al*. 2007; Hong *et al*. 2008, 2009; Kim *et al*. 2009, 2010, 2011), however, are rarely observed in Heteroptera. This extra *tRNA*^{Gln}-like sequence is 73 bp long (15, 512–15, 584), and the anticodon arm is identical to that of the regular *tRNA*^{Gln} in *E. cupreus* mitogenome. The nucleotides show 56% identities between two tRNA genes, and the variation is mainly located at the AA arm. The control region is an apparent non-coding region in vertebrate mtDNA, and has been shown in the replication origin for the heavy strand of mammalian mtDNA. The tRNA-like sequences in the control region may be remnant in the nascent DNA strand after serving as primers (Cantatore *et al*. 1987; Wan *et al*. 2012).

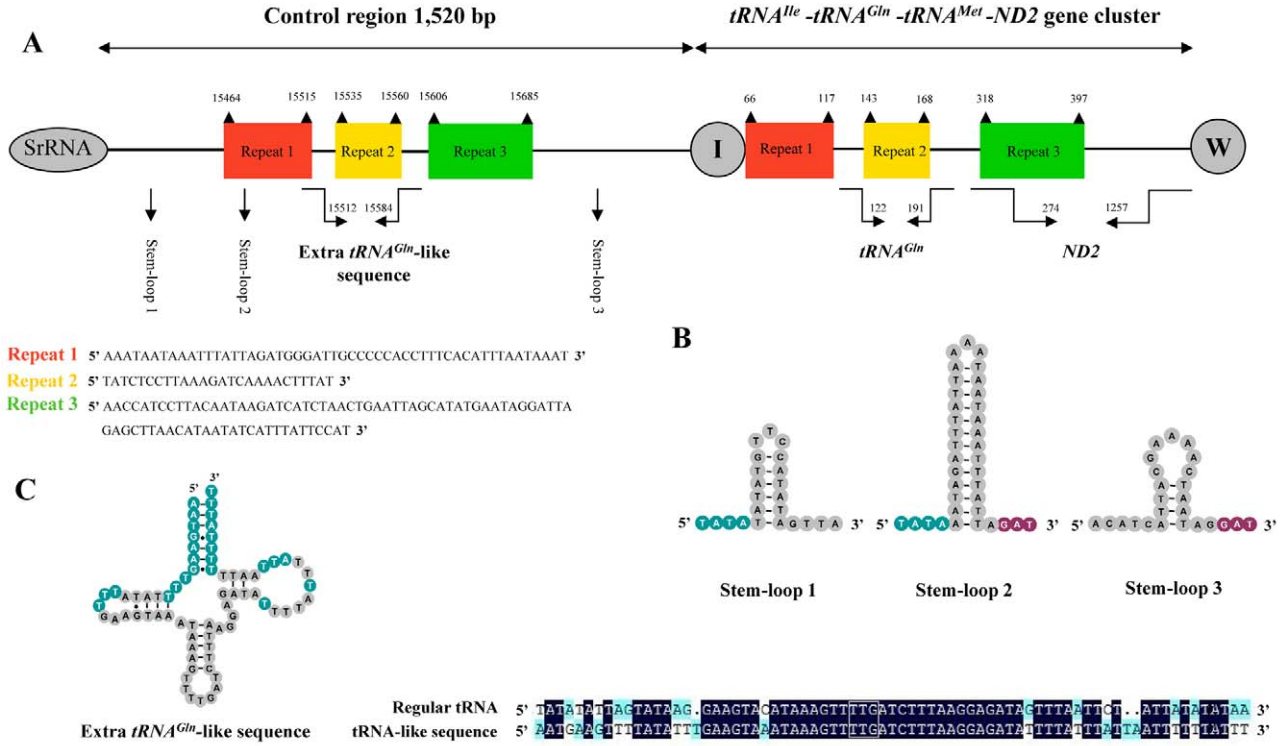

FIGURE 5. Control region of the *E. cupreus* mitogenome. (A) Structure elements found in the control region of *E. cupreus*. The red, yellow and green box represent three different repetitive sequences found in the control region and *tRNA*^{Ile} *-tRNA*^{Gln} *-tRNA*^{Met} *-ND2* gene cluste. (B) The putative stem-loops structure was found in the control region. The light green and pink indicates highly conserved flanking sequence. (C) Predicted secondary clover-leaf structure of one extra *tRNA*^{Gln}-like sequences in the control region (gray indicates the sequences are identical to the typical *tRNA*^{Gln}; the light green indicates the different sequences). Alignments with the corresponding regular *tRNA*^{Gln} sequences also provided. The boxed nucleotides indicate the anticodon, which designates the corresponding tRNA.

Acknowledgements

We sincerely thank Prof. Carl W. Schaefer of the University of Connecticut for suggestions and critical reading of the manuscript. This research is supported by grants from the National Natural Science Foundation of China (Nos. 30825006, 31061160186, 31111140015), the Special Fund for Agroscientific Research in the Public Interest (Nos. 201103012, 201103022), the Natural Science Foundation of Beijing (No. 6112013), the Key Laboratory of the Zoological Systematics and Evolution of the Chinese Academy of Sciences (No. O529YX5105).

References

Boore, J.L. & Brown, W.M. (1998) Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Current Opinion in Genetics and Development*, 8, 668–674. http://dx.doi.org/10.1016/S0959-437X(98)80035-X

Cameron, S.L. & Whiting, M.F. (2008) The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta* (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. *Gene*, 408, 112–123. http://dx.doi.org/10.1016/j.gene.2007.10.023

Cantatore, P., Gadaleta, M.N., Roberti, M., Saccone, C. & Wilson, A.C. (1987) Duplication and remoulding of tRNA genes during the evolutionary rearrangement of mitochondrial genomes. *Nature*, 329, 853–855. http://dx.doi.org/10.1038/329853a0

Cassis, G. & Gordon, F.G. (2002) *Zoological Catalogue of Australia: Hemiptera: Heteroptera (Pentatomomorpha)*. CSIRO Publishing, 741 pp. (http://en.wikipedia.org/wiki/Tessaratomidae - cite_ref-cassisgross_7-0#cite_ref-cassisgross_7-0)

Clary, D.O. & Wolstenholme, D.R. (1987) Drosophila mitochondrial DNA: conserved sequences in the A+T-rich region and supporting evidence for a secondary structure model of the small ribosomal RNA. *Journal of Molecular Evolution*, 25, 116–125. http://dx.doi.org/10.1007/BF02101753

Dotson, E.M. & Beard, C.B. (2001) Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*. *Insect Molecular Biology*, 10, 205–215. http://dx.doi.org/10.1046/j.1365-2583.2001.00258.x

Dzerefos, C.M., Witkowski, E.T.F. & Toms, R. (2009) Life-history traits of the edible stinkbug, *Encosternum delegorguei* (Hem., Tessaratomidae), a traditional food in southern Africa. *Journal of Applied Entomology*, 133, 739–759. http://dx.doi.org/10.1111/j.1439-0418.2009.01425.x

Gillespie, J.J., Johnston, J.S., Cannone, J.J. & Gutell, R.R. (2006) Characteristics of the nuclear (18S, 5.8S, 28S and 5S) and mitochondrial (12S and 16S) rRNA genes of *Apis mellifera* (Insecta: Hymenoptera): Structure, organization and retrotransposable elements. *Insect Molecular Biology*, 15, 657–686. http://dx.doi.org/10.1111/j.1365-2583.2006.00689.x

Gissi, C., Iannelli, F. & Pesole, G. (2008) Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity*, 101, 301–320. http://dx.doi.org/10.1038/hdy.2008.62

Hong, M.Y., Jeong, H.C., Kim, M.J., Jeong, H.U., Lee, S.H. & Kim, I. (2009) Complete mitogenome sequence of the jewel beetle, *Chrysochroa fulgidissima* (Coleoptera: Buprestidae). *Mitochondrial DNA*, 20, 46–60.

Hong, M.Y., Lee, E.M., Jo, Y.H., Park, H.C., Kim, S.R., Hwang, J.S., Jin, B.R., Kang, P.D., Kim, K., Han, Y.S. & Kim, I. (2008) Complete nucleotide sequence and organization of the mitogenome of the silk moth *Caligula boisduvalii* (Lepidoptera: Saturniidae) and comparison with other lepidopteran insects. *Gene*, 413, 49–57. http://dx.doi.org/10.1016/j.gene.2008.01.019

Hwang, U.W., Friedrich, M., Tautz, D., Park, C.J., and Kim, W. (2001) Mitochondrial protein phylogeny joins myriapods with chelicerates. *Nature*, 413, 154–157. http://dx.doi.org/10.1038/35093090

Kim, M.I., Baek, J.Y., Kim, M.J., Jeong, H.C., Kim, K.G., Bae, C.H., Han, Y.S., Jin, B.R. & Kim, I. (2009) Complete nucleotide sequence and organization of the mitogenome of the red-spotted Apollo butterfly, *Parnassius bremeri* (Lepidoptera: Papilionidae) and comparison with other lepidopteran insects. Molecules and Cells, 28, 347–363. http://dx.doi.org/10.1007/s10059-009-0129-5

Kim, M.J., Kang, A.R., Jeong, H.C., Kim, K.G. & Kim, I. (2011) Reconstructing intraordinal relationships in Lepidoptera using mitochondrial genome data with the description of two newly sequenced lycaenids, *Spindasis takanonis* and *Protantigius superans* (Lepidoptera: Lycaenidae). *Molecular Phylogenetics and Evolution*, 61, 436–445. http://dx.doi.org/10.1016/j.ympev.2011.07.013

Kim, M.J., Wan, X., Kim, K.G., Hwang, J.S. & Kim, I. (2010) Complete nucleotide sequence and organization of the mitogenome of endangered *Eumenis autonoe* (Lepidoptera: Nymphalidae). *African Journal of Biotechnology*, 9, 735–754.

Lavrov, D.V., Brown, W.M. & Boore, J.L. (2004) Phylogenetic position of the Pentastomida and (pan) crustacean relationships. *Proceedings of the Royal Society Biological Sciences*, 271, 537–544. http://dx.doi.org/10.1098/rspb.2003.2631

Lin, C.P. & Danforth, B.N. (2004) How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. *Molecular Phylogenetics and Evolution*, 30, 686–702. http://dx.doi.org/10.1016/S1055-7903(03)00241-0

Li H, Gao, J.Y., Liu, H.Y., Liu, H., Liang, A.P., Zhou, X.G. & Cai, W. (2011) The architecture and complete sequence of mitochondrial genome of an assassin bug *Agriosphodrus dohrni* (Hemiptera: Reduviidae). *International Journal Biological Sciences*, 7, 792–804. http://dx.doi.org/10.7150/ijbs.7.792

Li, H., Liu, H., Shi, A.M., Štys, P., Zhou, X.G. & Cai, W. (2012a) The complete mitochondrial genome and novel gene arrangement of the unique-headed bug *Stenopirates* sp. (Hemiptera: Enicocephalidae). *PLoS ONE*, 7, e29419. http://dx.doi.org/10.1371/journal.pone.0029419

Li, H., Liu, H.Y., Cao, L.M., Shi, A.M, Yang, H.L. & Cai, W. (2012b) The complete mitochondrial genome of the damsel bug *Alloeorhynchus bakeri* (Hemiptera: Nabidae). *International Journal of Biological Sciences*, 8, 93–107. http://dx.doi.org/10.7150/ijbs.8.93

Li, X., Ogoh, K., Ohba, N., Liang, X. & Ohmiya, Y. (2007) Mitochondrial genomes of two luminous beetles, *Rhagophthalmus lufengensis* and *R. ohbai* (Arthropoda, Insecta, Coleoptera). *Gene*, 392, 196–205. http://dx.doi.org/10.1016/j.gene.2006.12.017

Nardi, F., Spinsanti, G., Boore, J.L., Carapelli, A., Dallai, R. & Frati, F. (2003) Hexapod origins: monophyletic or paraphyletic? *Science*, 299, 1887–1889. http://dx.doi.org/10.1126/science.1078607

Perna, N.T. & Kocher, T.D. (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *Journal of Molecular Evolution*, 41, 353–358. http://dx.doi.org/10.1007/BF01215182

Rider, D.A. (2006) Family Tessaratomidae. *In*: Aukema B. & Rieger C. (Eds.) *Catalogue of the Heteroptera of the Palaearctic Region. Vol. 5*, Netherlands Entomological Society, Amsterdam, pp. 182–189.

Rolston, L.H., Aalbu, R. L., Murphy, M. J. & Rider, D. A. (1993) A catalog of the Tessaratomidae of the world. *Papua New Guinea Journal of Agriculture, Forestry and Fisheries*, 36, 36–108.

Schultheis, A.S., Weigt, L.A., & Hendricks, A.C. (2002) Arrangement and structural conservation of the mitochondrial control region of two species of *Plecoptera*: utility of tandem repeat-containing regions in studies of population genetics and evolutionary history. *Insect Molecular Biology*, 11, 605–610. http://dx.doi.org/10.1046/j.1365-2583.2002.00371.x

Shao, R., Campbell, N.J.H., & Barker, S.C. (2001) Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera). *Molecular Biology and Evolution*, 18, 858–865. http://dx.doi.org/10.1093/oxfordjournals.molbev.a003867

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, 28, 2731–2739. http://dx.doi.org/10.1093/molbev/msr121

Wolstenholme, D.R. (1992) Genetic novelties in mitochondrial genomes of multicellular animals. *Current Opinion Genetics and Development*, 2, 918–925. http://dx.doi.org/10.1016/S0959-437X(05)80116-9

Zhang, C.Y. & Huang, Y. (2008) Complete mitochondrial genome of *Oxya chinensis* (Orthoptera, Acridoidea). *Acta Biochimica et Biophys Sinica*, 40, 7–18. http://dx.doi.org/10.1111/j.1745-7270.2008.00375.x

Zhang, D.X. & Hewitt, F.M. (1997) Insect mitochondrial control region: A review of its structure, evolution and usefulness in evolutionary studies. *Biochemical Systematic and Ecology*, 25, 99–120. http://dx.doi.org/10.1016/S0305-1978(96)00042-7

Zhang, D.X., Szymura, J.M. & Hewitt, G.M. (1995) Evolution and structural conservation of the control region of insect mitochondrial DNA. *Journal of Molecular Evolution*, 40, 382–391. http://dx.doi.org/10.1007/BF00164024

Zhou, Z., Huang, Y. & Shi, F. (2007) The mitochondrial genome of *Ruspolia dubia* (Orthoptera: Conocephalidae) contains a short A+T–rich region of 70 bp in length. *Genome*, 50, 855–866. http://dx.doi.org/10.1139/G07-057

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31, 3406–3415.http://dx.doi.org/10.1093/nar/gkg595