ISSN 0026-8933, Molecular Biology, 2015, Vol. 49, No. 3, pp. 408–417. © Pleiades Publishing, Inc., 2015.
Published in Russian in Molekulyarnaya Biologiya, 2015, Vol. 49, No. 3, pp. 461–471.

MOLECULAR PHYLOGENETICS

UDC 577.21.06

The Complete Mitochondrial Genome of Peacock Sole *Pardachirus pavoninus* (Pleuronectiformes: Soleidae) and Comparative Analysis of the Control Region Among 13 Soles[1]

L. Gong[a, b], W. Shi[a], L.-Z. Si[a, b], Z.-M. Wang[c], and X.-Y. Kong[a]

[a] *Key Laboratory of Tropical Marine Bio-resources and Ecology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, 510000 China;*
e-mail: xykong@scsio.ac.cn

[b] *University of Chinese Academy of Sciences, Beijing, 100000 China*

[c] *Marine Fisheries Research Institute of Zhejiang, Zhoushan, 316021 China*

Received August 21, 2014; in final form, September 30, 2014

Abstract—The complete mitogenome of the peacock sole *Pardachirus pavoninus* (Lacepède, 1802) was determined. The total length is 16536 bp, containing 13 protein-coding genes, 22 tRNA genes and two rRNA genes, as well as one control region (CR). The L-strand replication origin (O_L), which is typically located in the WANCY cluster, is lost in *P. pavoninus*. The gene arrangement is identical to that in most teleosts. Comparison of the CR sequences among 13 soles reveals that a 211-bp fragment at the 5'-end of the CR is lost in the *P. pavoninus* mitogenome, responsible for its short sequence with a length of 872 bp. All typical conservative blocks (TAS, CSB-F, E, D, C, B, A, CSB-1, 2, 3) are identified. Seven out of 13 soles contain tandem repeats in the CR and the possible mechanisms of their formation are discussed. These results may provide the consensus sequences of the conserved units in the sole CR as well as molecular data for phylogenetic studies on Soleidae and Pleuronectiformes.

DOI: 10.1134/S0026893315030061

Keywords: flatfish, O_L, heteroplasmy, tandem repeat, termination associated sequence

INTRODUCTION

Soles are benthic flatfishes characterized by the localization of both eyes on the right side of the head [1]. Traditional systematic studies of the Soleidae, due to their high phenotypic similarity, have provoked great differences in the number and nomenclature of taxa [1–4]. This fact has made it necessary to develop molecular markers to figure out the controversial issues of Soleidae and flatfish systematics.

Mitochondrial DNA (mtDNA) has been extensively used for phylogenetic analysis due to its small size, high abundance in the cell, and high evolutionary rate [5–11]. The main non-coding region, the so-called control region (CR) or displacement loop (D-loop) contains the major regulatory elements for the replication and expression of the mitogenome, such as the sites of initiation of H-strand replication and both H- and L-strand transcription. Because of its fast evolutionary rate and functional importance, the CR is one of the most interesting parts of the vertebrate mitogenome [12–14]. Several detailed comparisons of the CRs in fish mtDNA have been published [15–19]. Nevertheless, it is still relatively badly known in Soleidae species.

The peacock sole *Pardachirus pavoninus* (Lacepède, 1802) belongs to the family Soleidae of Pleuronectiformes, occurring on the coast of the Indo-Pacific [20]. The most notable feature of this sole is that it is bordered by a dark rim and some have a blackish spot in the center [21]. Extracts from the sac under the skin is toxic and the mucus appears to have shark-repellent qualities [22].

In this study, the complete mitogenome of *P. pavoninus* was determined. Main characteristics and gene structure are described. The CR sequence was compared with that of 12 other soles. The features of repeat arrays and possible mechanisms of generation were analyzed. We hope that these results will provide consensus sequences of the conserved units of the CR in soles as well as useful molecular data for phylogenetic studies in Soleidae and Pleuronectiformes.

EXPERIMENTAL

Sampling, amplification and sequencing. The specimen of *P. pavoninus* was collected from Guangdong, China. A portion of the epaxial musculature was excised and the total genomic DNA was extracted using an SQ Tissue DNA Kit (Omega, Guangzhou, China) following the manufacturer's protocol. The

[1] The article is published in the original.

Table 1. Primers used in this study

Forward primer	Sequence, 5'–3'	Reverse primer	Sequence, 5'–3'
Z15	ATTAAAGCATAACHCTGAAGATGTTAAGAT	F2671	AGATAGAAACTGACCTGGAT
Z2625	GTTTACGACCTCGATGTTGGATCAGGACAT	F6746	GCGGTGGATTGTAGACCCATARACAGAGGT
YL-6746R	ACTAACCTCCTCATAAACC	YL-COIF	CGCCGATTATTAGAGGAA
R6754	CTAAGCCATCCTACCTGTG	F11089	TTTAACCAAGACCRGGTGATTGGAAGTC
YL-COIIR	GAACGACCCTCATCTCAC	YL-13347F	CTACTGGGGTGAAGATAA
Z10818	TTYGAAGCAGCCGCMTGATACTGACAYTT	F13413	TAGCTGCTACTCGGATTTGCACCAAGAGT
Z13347	AAGGATAACAGCTCATCCGTTGGTCTTAGG	H15149	AAACTGCAGCCCCTCAGAATGATATTTGTCCTCA
YL-13413R	GTTCACATTACCCGAGAC	F17147	TAGTTTARTGCGAGAATCCTAGCTTTGGG
Z17054	GYCGGTGGTTARAATCCTCCCTACTGCT	YL-17114F	TCTTGGCGGGATGTTGAT
L17114	RCGCCCAAAGCTAGDATTC	F49	GGCCCATCTTAACATCTTC

primers used to amplify the mtDNA of *P. pavoninus* are shown in (Table 1). Fragments generated from PCR amplification were sequenced by primer walking directly or if necessary, the purified PCR products were inserted into the pMD19-T vector (TaKaRa) then transformed in *E. coli* competent cells and sequenced using the ABI 3730 genetic Analyzer (Applied Biosystems). The complete mtDNA sequence was submitted to GenBank under the accession number KJ433565.

Bioinformatics. Sequenced fragments were assembled to create the complete mitochondrial genomes using CodonCode Aligner v3 (CodonCode Corporation, Dedham, MA) and BioEdit v7 [23]. Annotation and boundary determination of protein-coding and ribosomal RNA genes were performed using NCBI-BLAST (http://blast.ncbi.nlm.nih.gov). Transfer RNA genes and their potential cloverleaf structures were identified using tRNAscan-SE 1.21 [24], with cut-off values set to 1 when necessary. The gene map of the *P. pavoninus* mitogenome was generated using CGView [25]. In order to compare the characteristics of the CR sequences among the soles (Table 2), multiple alignment of 13 soles was performed using Clustal X 2.0 [30].

RESULTS

Genome Organization

The complete mitogenome of *P. pavoninus* is 16.536 bp, within the range of other mitogenomes sequenced in Pleuronectiformes to date (from 16417 bp in *Cynoglossus abbreviatus* and *Cynoglossus lineolatus* to 18139 bp in *Reinhardtius hippoglossoides*). The genome contains 13 protein-coding genes, two rRNA genes, 22 tRNA genes, and one control region. In addition, the gene arrangement is identical to that typical in fishes and most genes are encoded by the heavy strand (H-strand), except for *ND6* and eight tRNA genes (Fig. 1).

The overall base composition values for the L-strand are 29.1, 29.5, 16.4, and 25.0% for A, C, G, and T, respectively, with an A+T content of 54.0%. A total of 43 base pairs are found in 12 intergenic spacers ranging from 1 bp to 13 bp in length. Overlapping sites (a total of 26 bp) are observed between protein-coding genes: *ATP8* and *ATP6*, *ATP6* and *COIII*, *ND4L* and *ND4*, *ND5* and *ND6*, which are the four notable overlaps between protein-coding genes as reported in other vertebrate species, or between tRNA genes (*tRNA-Ile* and *tRNA-Gln*, *tRNA-Gln* and *tRNA-Met*, *tRNA-Thr* and *tRNA-Pro*) (Table 3).

Protein-Coding Genes, Ribosomal and tRNA Genes

The total size of the 13 protein-coding genes is 11.426 bp. Twelve of them begin with an ATG codon, except for the *COI* gene, which starts with GTG, as found in most other fishes [17, 31, 32]. Most genes use TAA, TA, or T as stop codons, while *ND6* ends with AGG, which is not commonly used as a stop codon in other vertebrate mitogenomes.

The most frequent amino acid is leucine (17.03%), while cysteine (0.79%) is the rarest one. The overall codon usage of the 13 protein-coding genes reveals a remarkable bias against the use of G (10.3%) at the third codon position, which is free from selective constraints on nucleotide substitutions. There is no bias at the first codon positions, but pyrimidines are over-represented as compared with purines (C 28.0% + T 40.6% = 68.6%) at the second position as in other vertebrate mitogenomes. These characteristics of *P. pavoninus* are congruent with those of other fish species reported previously [33–36].

Occurrences of *12S* rRNA and *16S* rRNA in *P. pavoninus* are flanked with *tRNA-Phe* and *tRNA-Leu* (UUR), and are interposed by *tRNA-Val*. The lengths of *12S* rRNA and *16S* rRNA are 965 bp and 1.701 bp, respectively. Twenty-two tRNA genes are interspersed between rRNA and protein-coding genes, which are in clusters or individually scattered in the genome. The sizes vary from 65 bp to 74 bp, similarly to other fish mitogenomes [34, 37–39]. Of the 22 tRNA genes,

Table 2. Information on the control region sequences of the 13 soles used in this study

Species	Abbr.	Accession No.	Length, bp		Tandem repeat				Reference
			total	No. TRs*	location**	copy No.	size	min. energy***	
Aesopia cornuta	*A. con*	NC_021969	1071	1071	/#	/	/	/	[26]
Brachirus orientalis	*B. ori*	KJ433558	925	925	/	/	/	/	[27]
Heteromycteris japonicus	*H. jap*	JQ639060	1356	1215	3'	2.0	139	−26.0(1)	[28]
Liachirus melanospilos	*L. mel*	KF573188	1220	1220	/	/	/	/	[37]
Pardachirus pavoninus	*P. pav*	KJ433565	872	836	5'	3.2	17	−5.0(2)	This study
Solea lascaris	*S. las*	AB271693	1200	910	3'	23.3	13	−4.5(2)	[17]
S. ovata	*S. ova*	KF142459	1153	884	3'	17.9	16	−5.3(2)	[31]
S. senegalensis	*S. sen*	AB270760	1017	953	3'	8.1	8	−3.7(3)	[17]
S. solea	*S. sol*	AB271692	1264	920	3'	44.4	8	−2.0(5)	[17]
Zebrias crossolepis	*Z. cro*	KJ433564	1079	1079	/	/	/	/	This study
Z. japonicus	*Z. jap*	JQ639060	1120	1120	/	/	/	/	This study
Z. quagga	*Z. qua*	NC_023225	1338	993	3'	3.7	117	−29.0(1)	[29]
Z. zebrinus	*Z. zeb*	NC_021377	1077	1077	/	/	/	/	[34]

* No TRs: CR length without tandem repeats.

** Location: tandem repeat at the 3' or 5'-end of the CR.

*** Min. energy: minimum free energy (kcal/mol), numbers in the parenthesis indicate the minimum number of copies necessary to form a stem-loop structure.

#/: No tandem repeat.

21 are able to fold into typical cloverleaf structures as predicted by tRNAscan-SE 1.21 [24], except for the *tRNA-Cys*, which shows a deviated secondary structure, in which the dihydrouracil loop could not be formed as in most bony fish [31, 32, 34]. All putative secondary cloverleaf structures contain 7 bp in the amino acid stem, while the majority has 4 bp in the DHU stem, and 5 bp in both the TΨC stem and the anticodon stem.

Variation of the O_L Region

Typically, the L-strand replication origin (O_L) is located in the WANCY tRNA cluster between the *tRNA-Asn* and *tRNA-Cys* genes and has the potential to fold into a stable stem-loop structure. However, in the *P. pavoninus* mitogenome, the WANCY region contains only a 13-bp intergenic spacer between the *tRNA-Asn* and *tRNA-Cys* genes, thus it is unable to form a stem-loop structure. The loss of the O_L sequence has also been observed in some vertebrate mitogenomes [40–42]. It has been suggested that some features of the tRNA gene sequence may be a functional substitute for the O_L [43, 44] and an alternative possibility would be a bidirectional origin in the control region [13].

Comparative Analysis of the Control Region among 13 Soles

The major non-coding region found in *P. pavoninus* is located between the *tRNA-Pro* and *tRNA-Phe* genes. The A+T content of the CR fragment and that of the whole strand excluding the CR are 62.0 and 53.6%, respectively. These results show that the CR sequence plays a vital role in the base content of the entire strand. Base composition of all 13 sole CRs was studied and *P. pavoninus* and *Solea lascar* showed the following abundances T > A > C > G while the rest of the species showed A > T > C > G. Nevertheless, all 13 species follow the general vertebrate patter of (A + T) > (C +G) in the CR [14]. Comparison of these 13 CR sequences revealed that the length heteroplasmy of these CR sequences was significant,

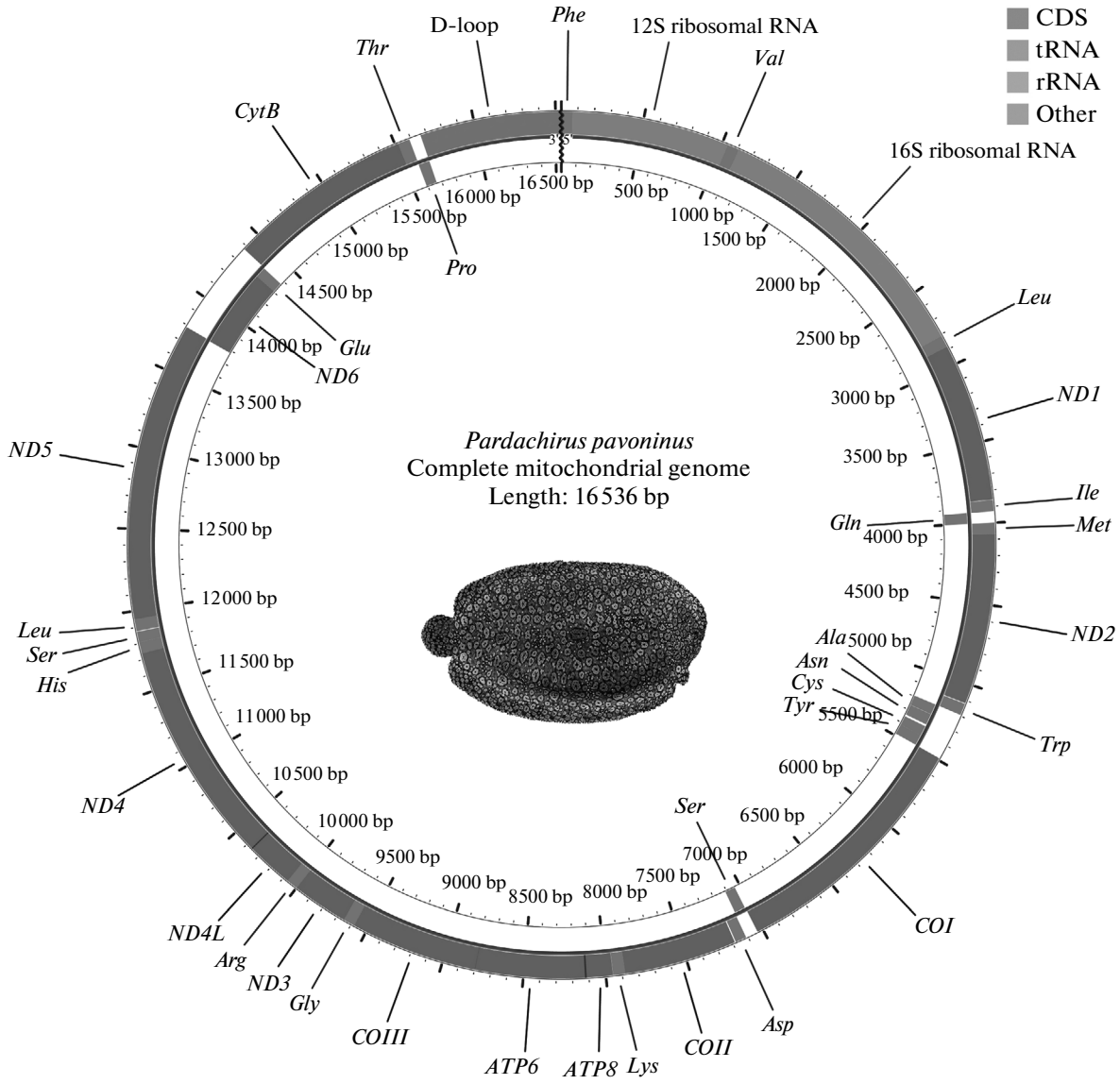


Fig. 1. Gene map of the *P. pavoninus* mitogenome.

ranging from 872 bp (*P. pavoninus*) to 1.356 bp (*Heteromycteris japonicus*). Analysis showed that the reason for *P. pavoninus*'s short CR sequence was mainly due to a 211-bp fragment loss in the 5'-end (Fig. 2). Seven out of 13 species own tandem repeats (TRs). Comparison of these arrays showed that the TRs were at the 5'-end in *P. pavoninus* CR, while in the other six soles at the 3'-end. Each species possesses various motifs with different numbers of copies. Further analysis found that the lengths of the motifs ranged from 8 bp (*Solea senegalensis* and *Solea solea*) to 139 bp (*H. japonicus*). *S. solea* showed the maximum number of copies (44.4) of an 8-bp motif, while *H. japonicas* had the minimum number of copies (two) for a 139-bp one. Secondary structure analyses showed that these TRs were able to form potential secondary structures. For example, a single motif in *Zebrias quagga* (117 bp) and *H. japonicas* CR (139 bp) was able to form a long stable stem-loop structure (Min. energy –29.0 and –26.0 kcal/mol, 37°C, respectively). In other species with short motifs (8–17 bp), such as *S. solea* and *Solea ovata*, several consecutive motifs were able to form hairpin structures (Min. energy –2.0 to –5.3 kcal/mol, 37°C) (Table 2).

In order to explore the characteristic of the CR, an alignment of the 13 sole CRs was conducted and the typical tripartite structure was exhibited, including the termination associated sequence (TAS), the central conserved blocks (CSB-F, E, D, C, B and A) and the conserved sequence blocks (CSB-1, 2, and 3). All these regulatory elements are highly conserved among these soles (Fig. 2). The TAS motif is located at the 5'-end of the CR with a typical TAS-complementary TAS block sequence (TAS-cTAS: TACAT-ATGTA).

Table 3. Features of the *P. pavoninus* mitogenome

Gene	Position		Length, bp	AA*	Anticodon	Start/Stop codon	Intergenic region**	Strand
	from	to						
tRNA-Phe	1	69	69		GAA		0	H
12S rRNA	70	1034	965				0	H
tRNA-Val	1035	1108	74		TAC		0	H
16S rRNA	1109	2809	1701				0	H
tRNA-Leu(UUA)	2810	2883	74		TAA		0	H
ND1	2884	3858	975	324		ATG/TAA	2	H
tRNA-Ile	3861	3930	70		GAT		−1	H
tRNA-Gln	3930	4000	71		TTG		−1	L
tRNA-Met	4000	4070	71		CAT		0	H
ND2	4071	5117	1047	348		ATG/TAA	4	H
tRNA-Trp	5122	5194	73		TCA		1	H
tRNA-Ala	5196	5264	69		TGC		1	L
tRNA-Asn	5266	5338	73		GTT		13	L
tRNA-Cys	5352	5416	65		GCA		0	L
tRNA-Tyr	5417	5486	70		GTA		1	L
COI	5488	7035	1548	515		GTG/TAA	3	H
tRNA-Ser(UCA)	7039	7109	71		TGA		3	L
tRNA-Asp	7113	7181	69		GTC		6	H
COII	7188	7881	694	231		ATG/T	0	H
tRNA-Lys	7882	7955	74		TTT		1	H
ATP8	7957	8124	168	55		ATG/TAA	−10	H
ATP6	8115	8798	684	227		ATG/TAA	−1	H
COIII	8798	9581	784	261		ATG/T	0	H
tRNA-Gly	9582	9652	71		TCC		0	H
ND3	9653	9998	346	115		ATG/T	0	H
tRNA-Arg	9999	10067	69		TCG		0	H
ND4L	10068	10364	297	98		ATG/TAA	−7	H
ND4	10358	11738	1381	460		ATG/T	0	H
tRNA-His	11739	11807	69		GTG		0	H
tRNA-Ser(UGC)	11808	11872	65		GCT		4	H
tRNA-Leu(CUA)	11877	11949	73		TAG		0	H
ND5	11950	13788	1839	612		ATG/TAA	−5	H
ND6	13784	14305	522	173		ATG/AGG	0	L
tRNA-Glu	14306	14374	69		TTC		4	L
Cytb	14379	15519	1141	380		ATG/T	0	H
tRNA-Thr	15520	15592	73		TGT		−1	H
tRNA-Pro	15592	15664	73		TGG		0	L
Control region	15665	16536	872				0	H

* AA—Amino acid;

** Intergenic region—non-coding bases between the feature on the same line and the line below, with a negative number indicating an overlap.

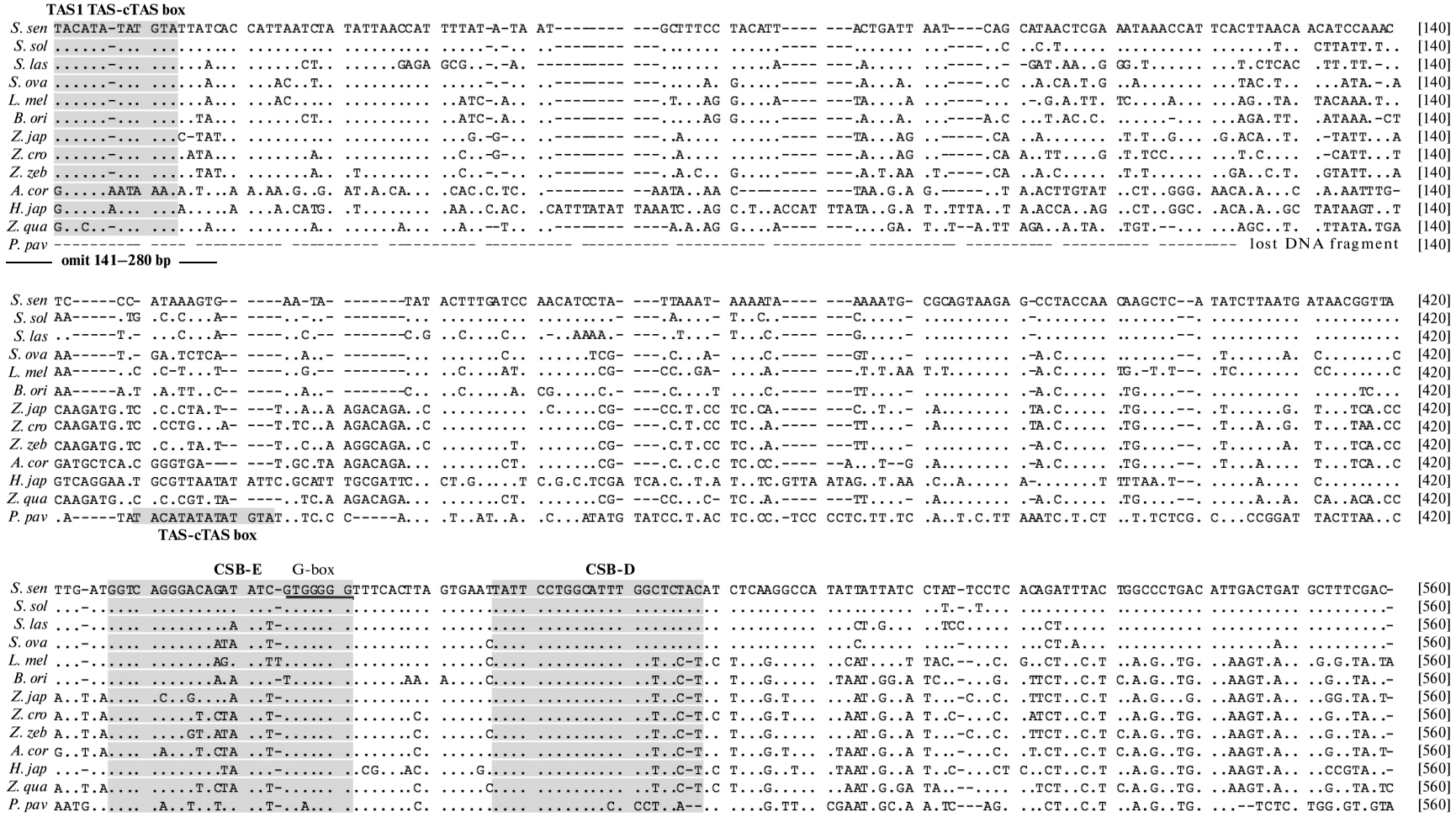
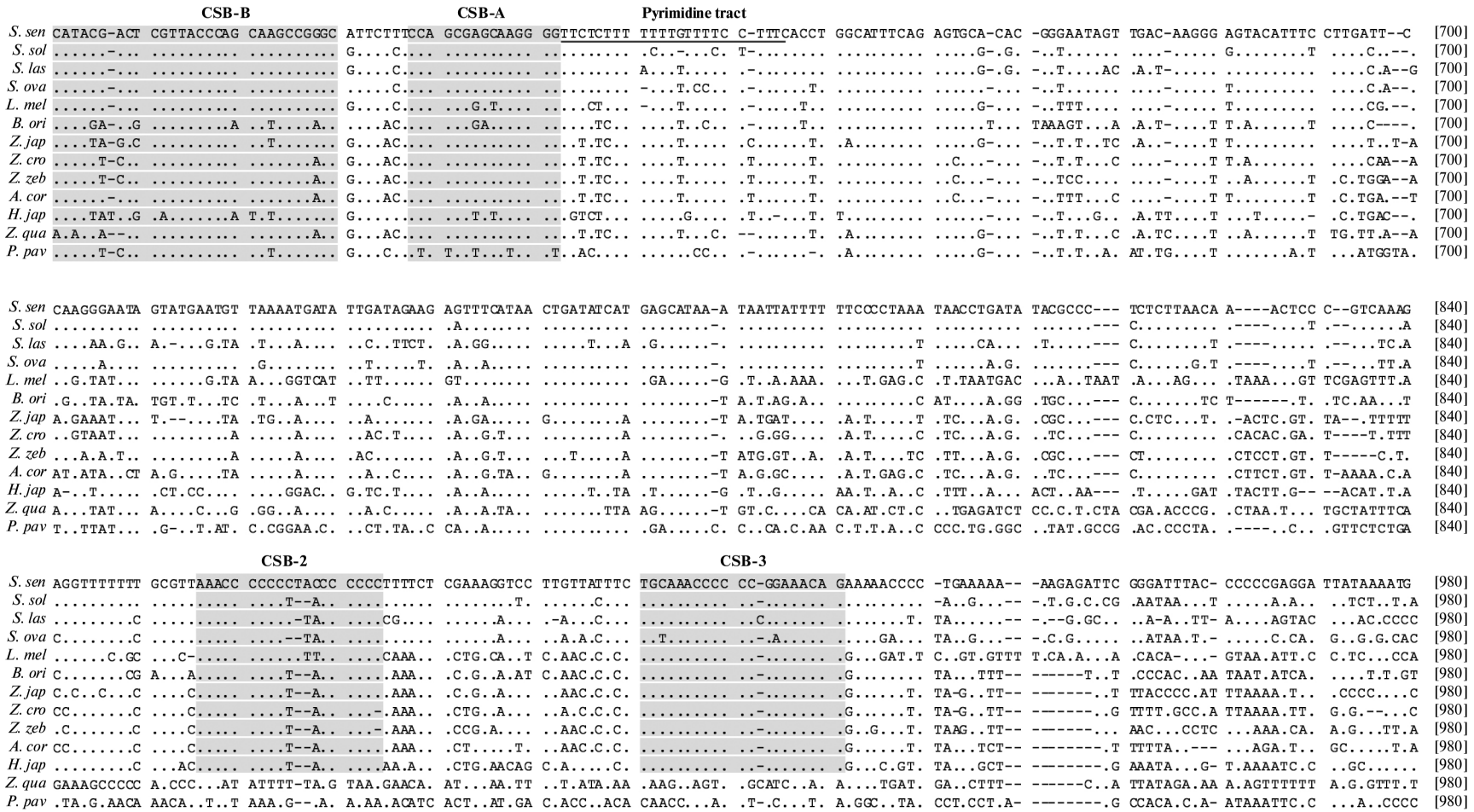


Fig. 2. Complete CR fragment alignment of the 13 soles. The boxed sequences indicate the Termination Associated Sequences (TASs), Central Conserved Blocks (CSBs) and Conserved Sequence Blocks (CSBs). The dashed line indicates the sequence lost in *P. pavoninus*.


CSB-B
CSB-A
Pyrimidine tract
CSB-2
CSB-3
S. sen CATACG-ACT CGTTACCCAG CAAGCCGGGC ATTCTTTCCA GCGAGCAAGG GGTTCTCTTT TTTTGTTTTC C-TTTCACCT GGCATTTCAG AGTGCA-CAC -GGGAATAGT TGAC-AAGGG AGTACATTTC CTTGATT--C [700]
S. sol ......-.. .... G....C.... .... .C..-...C. T-..... ....G-.G. -..T..... ....-..... G.....T ....C..--. [700]
S. las ......-.. .... G....C.... .... A...T..... .-..... ....G-.G. -..T....AC .A.T-..... ....C.A--G [700]
S. ova ......-.. ....C.... .... -...T.CC.. .-.....T. ....-... -..T..... ....-..... T..... ....C.A--. [700]
L. mel ......-.. .... G....C.... ...G.T.... ....CT... -...T..... .-.....T.. ....G-... -..TTT.... ....-..... T..... ....CG.--. [700]
B. ori ....GA-..G .....A ..T....A.. ....AC.... ...GA..... .....TC.. ....T..C.. .-.....T.. ....T... TAAAGT...A .A.T-....T T.A.....T ....C----. [700]
Z. jap ....TA-G.C ....T..... G...AC.... ....T.TC.. ....T..... .C.....T. .A..... .....G-... -..T.T..TC .A..-....T T..... ....T..T-A [700]
Z. cro .....T-C.. ....A.. G...AC.... ....T.TC.. ....T..... .T.....T. ....C...-... -..T.T...C ....-....T T.A..... ....CAA--A [700]
Z. zeb .....T-C.. ....A.. G...AC.... ....T.TC.. ....T..... .T.....T. ....C...-... -..TCC.... ....-....T ..A.....T .C.TGGA--A [700]
A. cor ......-.. ....A.. G...AC.... ....T.TC.. ....T..... .T.....T. ....C...-... -..TTT...C ....-....T T.....T .C.TGA.--T [700]
H. jap ....TAT..G .A.....A T.T..... G..... ...T.T.... ....GTCT... .....G.... .T..-...T. T..... .....-... -..T...G.. .A.TT....T ...T.....- .C.TGAC--. [700]
Z. qua A.A..A--.. ....A.. G...AC.... ....T.TC.. ....T...C. --.....T. .A..... .....G-... -..T.T...C .A.TC....T ..A.....T TG.TT.A--A [700]
P. pav .....T-C.. ....T..... G...C...T. T..T...T.. .T..AC.... ......CC.. .-.....-. .A..... .....G-... -..T.T..A. AT.TG....T .....A.T ...ATGGTA. [700]

S. sen CAAGGGAATA GTATGAATGT TAAAATGATA TTGATAGAAG AGTTTCATAA CTGATATCAT GAGCATAA-A TAATTATTTT TTCCCCTAAA TAACCTGATA TACGCCC--- TCTCTTAACA A----ACTCC C--GTCAAAG [840]
S. sol .... .A..... ....-. ....--- C..... .----....T .--.....A [840]
S. las ....AA.G.. A.-...G.TA .T...A.... ..C..TTCT. .A.GG..... .....T...A .G.....-. ....T .....CA... .T.....--- C.....T .----..... .--...TC.A [840]
S. ova .....A.... ..... .G..... ..T.....T. .A..A..... ....-. ....T ......A.G. .....--- C.....G.T .----....T .--...TT.A [840]
L. mel ..G.TAT.. ......G.TA A...GGTCAT ..TT..... GT..... ..... .GA.....-G .T..A.AAA. ...T.GAG.C .T.TAATGAC ...A..TAAT .A...AG... .TAAA...GT TCGAGTTT.A [840]
B. ori .G..TA.TA. TGT.T...TC .T...A...T ....C..... .A..A..... ....-T A.T.AG.A.. ....C AT....A.GG .TGC...--- C.....TC T------.T. .TC.AA...T [840]
Z. jap A.GAAAT... T.--....TA .TG..A.... ..A..... .A.GA..... G..... .....A ....-T A.TGAT.... .A.T.....T .TC...A.G. .CGC...--- C.CTC...T. .-ACTC.GT. TA--.TTTTT [840]
Z. cro ..GTAAT... ....A ....AC.T.... .A..G.T... ....A ....-. ..G.GG.... .A.T.....C .TC...A.G. ..TC...--- C..... .CACAC.GA. T----T.TTT [840]
Z. zeb ...A.A.T.. ....A ..... .AC..... .A..G.T... ...T.....A ....-T ATG.GT..A. .A.T....TC .TT...A.G. .CGC...--- CT..... .CTCCT.GT. T-----C.T. [840]
A. cor AT.ATA..CT A.G.....TA .....A.... ..A..C.... .A..G.TA.. G..... .....A ....-T A.G.GC.... .A.T.GAG.C .TC...A.G. ..TC...--- C..... .CTTCT.GT. T-AAAA.C.A [840]
H. jap A-..T..... .CT.CC.... ....GGAC.. G.TC.T.... .A..A..... .....T..TA .T.....-G .T..G..... AA.T..A..C .TTT..A... ACT..AA--- .T.....GAT .TACTT.G-- --ACAT.T.A [840]
Z. qua A...TAT... A....C...G .GG..A.... ..A.C..... .A..A.TA.. .....TTA AG.....-T GT.C....CA CA.AT.CT.C ..TGAGATCT CC.C.T.CTA CGA.ACCCG. .CTAA.T... TGCTATTTCA [840]
P. pav T..TTAT... .G-..T.AT. C.CGGAA.C. ..CT.TA..C CA..A..... ..... .GA.....C. C..CA.C.AA C.T.T.A..C CCC.TG.GGC ..TAT.GCCG .AC.CCCTA. .----..C.. .GTTCTCTGA [840]

S. sen AGGTTTTTTT GCGTTAAACC CCCCCTACCC CCCCTTTTCT CGAAAGGTCC TTGTTATTTC TGCAAACCCC CC-GGAAACA GAAAAACCCC -TGAAAAA-- -AAGAGATTC GGGATTTAC- CCCCCGAGGA TTATAAAATG [980]
S. sol ....T--A.. ....C... ....-..... ..... -A..G...-- -.T.G.C.CG .AATAA...T .....A.A.. ..TCT..T.A [980]
S. las .....C ....-TA.. ....CG.... ......A... .-A...C... ....C..... ....T. TA..... -- -.--G.GC.. ..A-A..TT- A....AGTAC ...AC.CCCC [980]
S. ova C.....C ....--TA.. ....A... ..A.A.C... ..T..... ..-.A..... ....GA... TA..G...-- -.C.G..... ..ATAA.T.- .....C.CA. G..G.G.CAC [980]
L. mel ......C.GC ...C-..... ....TT.. ....CAAA.. .CTG.CA..T C.AAC.C.C. ....-..... .G...GAT.T C..GT.GTTT T.CA.A...A .CACA-...- GTAA.ATT.C C.TC...CCA [980]
B. ori C.....CG A...A..... ....T--A.. .....AAA.. .C.G..A.AT C.AAC.C.C. ....-..... .G..... TA...TTT-- ------T..T .CCCAC..AA TAAT.ATCA. ....T.T.GT [980]
Z. jap C.C..C...C ....C..... ....T--A.. .....AAA.. .C.G..A.... ..AAC.C.C. ....-..... .G.....T. TA-G..TT-- -------..T TTACCCC.AT TTAAAA.T.. .CCCC....C [980]
Z. cro CC.....C ....C..... ....T--A.. ...-.AAA.. .CTG..A.... ..A.C.C.C. ....-..... .G.....T. TA-G..TT-- -------..G TTTT.GCC.A TTAAAA.TT. G.G.--...C [980]
Z. zeb .C.....C ....C..... ....T--A.. ...-.AAA.. .CCG.A.... ..AAC.C.C. ....-..... .G..G...T. TAAG..TT-- -------... AAC...CCTC ...AAA.CA. A.G...TT.A [980]
A. cor CC.....C ....C..... ....T--A.. .....AAA.. .CT.....T. ..AAC.C.C. ....-..... .G.....T. TA...TCT-- -------.T TTTTA..--- ...AGA.T.. GC.....T.A [980]
H. jap .....C ...AC..... ....T--A.. ....AA.A.. .CTG.AACAG C.A.....C. ....-..... .G...C.GT. TA...GCT-- --------.G AAATA...G- T.AAAATC.C ..GC..... [980]
Z. qua GAAAGCCCCC A.CCC...AT ATTTT-TA.G TAA.GAACA. AT...AA.TT ..T.ATA.AA .AAG..AGT. .GCATC..A. A....TGAT. GA..CTTT-- ------C..A TTATAGA.AA A.AGTTTTTT A.G.GTTT.T [980]
P. pav .TA.G.AACA AACA..T..T AAA.G--.A. A.AA.ACATC ACT..AT.GA C.ACC..ACA CAACC...A. .T-.C...T. A.GGC..TA. CCT.CCT.A- --------.G CCACA.C.A- ATAAATTC.C ...A..CCCC [980]

Fig. 2. (Contd.).

Although no TAS block exists in the uniform 5'-end of the CR in *P. pavoninus* due to the lack of a 211 bp sequence, another sequence (TACAT-ATGTA) at the location of about 290 bp in the aligned CR sequence was identified (Fig. 2). Successively, the CSB-F, E, D, C, B and A of the central conserved blocks were distinguishable and the key sequences of these blocks were as follows: CSB-F: CAGTAAGAG-ACCACCAAC; CSB-E: GGTCAGGGACA-AATT-GTGGGGG. The GT-box (GTGGGGG) was the most conservative in CSB-E; CSB-D: TATTCCTGGTATTTGG-TCCT; CSB-C was moderately conserved with a pattern as CT-CAT-C-ATGC; CSB-B: CATA-C-CGTTACCCAGCAAGC-CGGGC; CSB-A: CCAGCG-GCAAGGGG. Simultaneously, a pyrimidine tract following the CSB-A was also fixed as TTCTCTTTTTTTGTTTTCC-TTTC. The CSB-1, 2, 3 conserved sequence blocks were confirmed at the 3'-end of the CR and their consensus sequences were as follows: CSB-1: TAACTGATAT-CAAG-GCATAA; CSB-2: AAACCCCCCT-TAC-CCCCC; CSB-3: TGCAAACCCCCC-GGAAACAG, respectively. Even so, the CSB-2 and CSB-3 blocks have not been identified in *P. pavoninus* and *Z. quagga*.

DISCUSSION

The CR is the most variable region because of a faster rate of evolution as compared with the rRNA and protein-coding genes of the mitochondrial genome, but some sequences maintain a striking identity despite the influence of the rapid evolutionary changes, such as the TAS block, CSB-F, E, D, C, B and A and CSB-1, 2, and 3 blocks.

Generally, there exists only one TAS-cTAS in the CR [7, 31, 34, 38, 45]. In the present study, all soles except *P. pavoninus* had only one such block. In the *P. pavoninus* CR, although the TAS-cTAS block is missing a 211-bp counterpart, three blocks (one perfect and two imperfect TAS-cTAS blocks) still could be identified at the location of about 290–340 bp in the aligned sequences. For a long time, the TAS-cTAS block was thought to act as a termination signal for D-loop strand synthesis during the replication of mitochondrial DNA [13, 46]. Thus, the phenomenon in *P. pavoninus* CR corroborates the functional significance of this block. Furthermore, in previous studies, more than one TAS-cTAS blocks in the CR have been reported, such as two blocks in common carp, Japanese crucian carp [15] and in two species of *Rhinogobius* [47] and three blocks in *Myxine glutinosa* [48]. Generally, only one block exerts as the termination signal, however, there is no pertinent explanation for the issue of which one of these blocks functions to date. Therefore, further studies focused on these questions are needed in the future.

Generally, the differences in the TRs and the number of copies have a significant impact on the length heteroplasmy of the CR [47, 49–51]. Among the 13 soles, seven species possess TRs, and their CR lengths range from 1.017 bp (*S. senegalensis*) to 1.356 bp (*H. japonicus*) excluding the abnormal length of *P. pavoninus* CR (the same hereinafter). In the rest six species without TRs the CRs are from 925 bp (*Brachirus orientalis*) to 1220 bp (*Liachirus melanospilos*). It is not difficult to find that the tandem arrays indeed play an important role in the length heteroplasmy of CR. It is notable, however, that several sequences without TRs, on the contrary, are longer than those carrying them. Further analysis showed that seven CR sequences removing the TRs, plus the rest six sequences without TRs ranged from 884 bp (*S. ovata*) to 1220 bp (*L. melanospilos*) in length. Alignment of these 13 sole sequences suggested that the length heteroplasmy was mainly caused by the hypervariable regions downstream of the CSB-3 sequence. This data manifested that besides the variations in the TRs and copy numbers, hypervariable regions also affect the length heteroplasmy of the CR.

So far, three main mechanisms have been proposed to account for the formation of the repeated sequences in the mitogenome, including the Illegitimate Elongation Model [52], the Improper Initiation Model [53], and the Pause-Melting Misalignment [54]. As for IEM suggested by Buroker, it is targeted at explaining the generation of the TAS motif in the 5'-end of the CR. In this study, the TRs in the *P. pavoninus* CR exactly included the TAS in the 5'-end of the CR and it was able to form a hairpin structure. Thus, when the nascent (N) heavy strand is arrested at the TAS sequence, a competitive equilibrium between the N-strand and the displaced (D) heavy strand occurs, what results in a frequent misalignment prior to elongation in the repeat region. This mechanism well explains the TRs in *P. pavoninus* CR. Nevertheless, it fails to account for the TRs in the 3'-end of the CR in the rest six species.

Another influential model, IIM, specially explains the TRs in the 3' end of the CR. Based on this model, mt-replication begins upstream of the *tRNA-Phe* gene, an improper origin of replication, and results in an identical in length repeated sequence positioned between the *tRNA-Phe* and the normal origin of replication (O_H). In this study, although the repeated arrays in six species are located at the 3'-end, they do not involve the complete sequences between *tRNA-Phe* and O_H. In fact, some of the motifs only cover a very short sequence. These features showed that none of the TRs accord with this model.

Currently, Shi et al. [54] proposed a novel model to describe the birth and indel of TRs in the mitogenome. In contrast to IEM and IIM, the PMM model can explain the formation and indel of the mt genome tandem repeats at any location. It suggests that during the pause event in mitochondrion replication, dynamic competition between the nascent (N-) strand and the displaced (D-) strand can lead to the melting of the N-strand from the template (T-) strand. When the N-strand successfully rebinds to the T-strand, mis-pairing between

complementary or near-complementary sequences can easily occur due to the ability of the TR sequence to fold into potential stem-loop structures. As a result, after the next round of mt-duplication or repair takes place, insertion or deletion of one or several motifs could occur. Comparison of these TRs indicates that the PMM model is capable to account for the tandem arrays in this study, including the *P. pavoninus* CR, which can also be explained by the IEM.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (30870283, 31071890, 3147979).

REFERENCES

1. Nelson J.S. 2006. *Fishes of the World*, 4th ed. New York: Wiley.
2. Bini G. 1968. *Atlante dei pesci delle coste italiane*, vol. VIII: *Osteitti (Pleuronettiformi, Echeneiformi, Gobiesociformi, Tetraodontiformi, Lofiiformi)*. Roma: Mondo Sommerso.
3. Hureau J.-C., Monod T. (Eds.) 1973. *Checklist of the Fishes of the North-Eastern Atlantic and of the Mediterranean*. Paris: UNESCO, vol. 1.
4. Li S., Wang H., 1995. *Fauna Sinica (Osteichthyes): Pleuronectiformes*. Beijing: Science Press.
5. Azevedo M.F.C., Oliveira C., Pardo B.G., Martínez P., Foresti F. 2008. Phylogenetic analysis of the order Pleuronectiformes (Teleostei) based on sequences of 12S and 16S mitochondrial genes. *Genet. Mol. Biol*. **31**, 284–292.
6. Pardo B.G., Machordom A., Foresti F., Porto-Foresti F., Azevedo M.F.C., Bañon R., Sánchez L., Martínez P. 2005. Phylogenetic analysis of flatfish (Order Pleuronectiformes) based on mitochondrial 16s rDNA sequences. *Sci. Mar*. (Barcelona). **69**, 531–543.
7. Cheng Y., Wang R., Xu T. 2011. The mitochondrial genome of the spinyhead croaker *Collichthys lucida*: Genome organization and phylogenetic consideration. *Mar. Genom*. **4**, 17–23.
8. Oh D.-J., Oh B.-S., Jung M.-M., Jung Y.-H. 2010. Complete mitochondrial genome of three *Branchiostegus* (Perciformes, Malacanthidae) species: Genome description and phylogenetic considerations. *Mitochondrial DNA*. **21**, 151–159.
9. Zhang X., Yue B., Jiang W., Song Z. 2009. The complete mitochondrial genome of rock carp *Procypris rabaudi* (Cypriniformes: Cyprinidae) and phylogenetic implications. *Mol. Biol. Rep*. **36**, 981–991.
10. Liu T., Jin X., Wang R., Xu T. 2013. Complete sequence of the mitochondrial genome of *Odontamblyopus rubicundus* (Perciformes: Gobiidae): Genome characterization and phylogenetic analysis. *J. Genet*. **92**, 423–432.
11. Miya M., Kawaguchi A., Nishida M. 2001. Mitogenomic exploration of higher teleostean phylogenies: A case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. *Mol. Biol. Evol*. **18**, 1993–2009.
12. Clayton D.A. 2000. Transcription and replication of mitochondrial DNA. *Hum. Reprod*. **15**, 11–17.
13. Clayton D.A. 1991. Replication and transcription of vertebrate mitochondrial DNA. *Annu. Rev. Cell Biol*. **7**, 453–478.
14. Sbisà E., Tanzariello F., Reyes A., Pesole G., Saccone C. 1997. Mammalian mitochondrial D-loop region structural analysis: Identification of new conserved sequences and their functional and evolutionary implications. *Gene*. **205**, 125–140.
15. Guo X., Liu S., Liu Y. 2003. Comparative analysis of the mitochondrial DNA control region in cyprinids with different ploidy level. *Aquaculture*. **224**, 25–38.
16. Lee W.J., Conroy J., Howell W.H., Kocher T.D. 1995. Structure and evolution of teleost mitochondrial control regions. *J. Mol. Evol*. **41**, 54–66.
17. Manchado M., Catanese G., Ponce M., Funes V., Infante C. 2007. The complete mitochondrial genome of the Senegal sole, *Solea senegalensis* Kaup. Comparative analysis of tandem repeats in the control region among soles. *Mitochondrial DNA*. **18**, 169–175.
18. Kong X., Yu J., Zhou L., Yu Z. 2007. Comparative analysis of 5'-end sequence of the mitochondrial control region of six flatfish species (Pleuronectidae) from the Yellow Sea. *Raffles Bull. Zool*. **14**, 111–120.
19. Zhao L., Zheng Z., Huang Y., Zhou Z., Wang L. 2011. Comparative analysis of the mitochondrial control region in Orthoptera. *Zool. Stud*. **50**, 385–393.
20. Randall J.E., Allen G.R., Steene R.C. 1997. *Fishes of the Great Barrier Reef and Coral Sea*. Hawaii: Univ. of Hawaii Press.
21. Carpenter K., Niem V. 1999. *The Living Marine Resources of the Western Central Pacific*, vol. 3: *Batoid Fishes, Chimaeras and Bony Fishes*, part 1: *Elopidae to Linophrynidae*. FAO Species Identification Guide for Fishery Purposes. Rome: FAO Library.
22. Masuda H., Muzik K.M. 1984. *The Fishes of the Japanese Archipelago*. Tokyo: Tokai Univ. Press, vol. 2.
23. Hall T.A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Res. Symp Ser*. **41**, 95–98.
24. Lowe T.M., Eddy S.R. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. **25**, 0955–0964.
25. Stothard P., Wishart D.S. 2005. Circular genome visualization and exploration using CGView. *Bioinformatics*. **21**, 537–539.
26. Wang S., Shi W., Wang Z., Gong L., Kong X. 2013. The complete mitochondrial genome sequence of *Aesopia cornuta* (Pleuronectiformes: Soleidae). *Mitochondrial DNA*. [Epub ahead of print]. doi 10.3109/19401736.2013.803544
27. Shi W., Jiang J., Miao X., Kong X. 2014. The complete mitochondrial genome sequence of *Heteromycteris japonicus* (Pleuronectiformes: Soleidae). *Mitochondrial DNA*. **25**, 257–258.
28. Shi W., Gong L., Wang S., Kong X. 2014. The complete mitochondrial genome of *Brachirus orientalis* (Pleu-

ronectiformes: Soleidae). Mitochondrial DNA. [Epub ahead of print]. doi 10.3109/19401736.2014.915540

29. Li D., Shi W., Miao X., Kong X. 2014. The complete mitochondrial genome of *Zebrias quagga* (Pleuronectiformes: Soleidae). *Mitochondrial DNA*. [Epub ahead of print]. doi 10.3109/19401736.2014.933331
30. Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J., Higgins D.G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*. **23**, 2947–2948.
31. Shi W., Gong L., Wang S., Kong X. 2013. The complete mitochondrial genome of *Solea ovata* (Pleuronectiformes: Soleidae). *Mitochondrial DNA*. **25**, 454–455.
32. Gong L., Shi W., Wang Z., Miao X., Kong X. 2013. Control region translocation and a tRNA gene inversion in the mitogenome of *Paraplagusia japonica* (Pleuronectiformes: Cynoglossidae). *Mitochondrial DNA*. **24**, 671–673.
33. Manchado M., Catanese G., Infante C. 2004. Complete mitochondrial DNA sequence of the Atlantic bluefin tuna *Thunnus thynnus*. *Fisheries Sci*. **70**, 68–73.
34. Wang Z., Shi W., Jiang J., Wang S., Miao X., Huang L., Kong X. 2013. The complete mitochondrial genome of a striped sole *Zebrias zebrinus* (Pleuronectiformes: Soleidae). *Mitochondrial DNA*. **24**, 633–635.
35. Peng Z., Wang J., He S. 2006. The complete mitochondrial genome of the helmet catfish *Cranoglanis bouderius* (Siluriformes: Cranoglanididae) and the phylogeny of otophysan fishes. *Gene*. **376**, 290–297.
36. Kong X., Dong X., Zhang Y., Shi W., Wang Z., Yu Z. 2009. A novel rearrangement in the mitochondrial genome of tongue sole, *Cynoglossus semilaevis*: Control region translocation and a tRNA gene inversion. *Genome*. **52**, 975–984.
37. Gong L., Shi W., Wang S., Kong X. 2013. The complete mitochondrial genome of *Liachirus melanospilos* (Pleuronectiformes: Soleidae). *Mitochondrial DNA*. **1–2**. [Epub ahead of print]. doi 10.3109/19401736.2013.845763
38. Cui Z., Liu Y., Li C.P., You F., Chu K.H. 2009. The complete mitochondrial genome of the large yellow croaker, *Larimichthys crocea* (Perciformes, Sciaenidae): Unusual features of its control region and the phylogenetic position of the Sciaenidae. *Gene*. **432**, 33–43.
39. Xu T., Cheng Y., Sun Y., Shi G., Wang R. 2011. The complete mitochondrial genome of bighead croaker, *Collichthys niveatus* (Perciformes, Sciaenidae): Structure of control region and phylogenetic considerations. *Mol. Biol. Rep*. **38**, 4673–4685.
40. Seligmann H., Krishnan N.M., Rao B.J. 2006. Possible multiple origins of replication in primate mitochondria: alternative role of tRNA sequences. *J. Theor. Biol*. **241**, 321–332.
41. Seligmann H., Krishnan N.M. 2006. Mitochondrial replication origin stability and propensity of adjacent tRNA genes to form putative replication origins increase developmental stability in lizards. *J. Exp. Zool. B*. **306**, 433–449.
42. Seligmann H. 2010. Mitochondrial tRNAs as light strand replication origins: similarity between anticodon loops and the loop of the light strand replication origin predicts initiation of DNA replication. *Biosystems*. **99**, 85–93.
43. Desjardins P., Morais R. 1990. Sequence and gene organization of the chicken mitochondrial genome: A novel gene order in higher vertebrates. *J. Mol. Biol*. **212**, 599–634.
44. Seligmann H., Labra A. 2014. The relation between hairpin formation by mitochondrial WANCY tRNAs and the occurrence of the light strand replication origin in Lepidosauria. *Gene*. **542**, 248–257.
45. He A., Luo Y., Yang H., Liu L., Li S., Wang C. 2011. Complete mitochondrial DNA sequences of the Nile tilapia (*Oreochromis niloticus*) and blue tilapia (*Oreochromis aureus*): Genome characterization and phylogeny applications. *Mol. Biol. Rep*. **38**, 2015–2021.
46. Southern Š.O., Southern P.J., Dizon A.E. 1988. Molecular characterization of a cloned dolphin mitochondrial genome. *J. Mol. Evol*. **28**, 32–42.
47. Chen I.S., Hsu C.H., Hui C.F., Shao K.T., Miller P., Fang L.S. 1998. Sequence length and variation in the mitochondrial control region of two freshwater gobiid fishes belonging to *Rhinogobius* (Teleostei: Gobioidei). *J. Fish Biol*. **53**, 179–191.
48. Delarbre C., Rasmussen A., Arnason U., Gachelin G. 2001. The complete mitochondrial genome of the hagfish *Myxine glutinosa*: Unique features of the control region. *J. Mol. Evol*. **53**, 634–641.
49. Ravago R.G., Monje V.D., Juinio-Meñez M.A. 2002. Length and sequence variability in mitochondrial control region of the milkfish, *Chanos chanos*. *Mar. Biotechnol*. **4**, 40–50.
50. Chen C.A., Ablan M.C.A., McManus J.W., Bell J.D., Tuan V.S., Cabanban A.S., Shao K. 2004. Variable numbers of tandem repeats (VNTRs), heteroplasmy, and sequence variation of the mitochondrial control region in the three-spot Dascyllus, *Dascyllus trimaculatus* (Perciformes: Pomacentridae). *Zool. Stud*. **43**, 803–812.
51. Lunt D.H., Whipple L.E., Hyman B.C. 1998. Mitochondrial DNA variable number tandem repeats (VNTRs): Utility and problems in molecular ecology. *Mol. Ecol*. **7**, 1441–1455.
52. Buroker N., Brown J., Gilbert T., O'Hara P., Beckenbach A., Thomas W., Smith M. 1990. Length heteroplasmy of sturgeon mitochondrial DNA: An illegitimate elongation model. *Genetics*. **124**, 157–163.
53. Broughton R.E., Dowling T.E. 1994. Length variation in mitochondrial DNA of the minnow *Cyprinella spiloptera*. *Genetics*. **138**, 179–190.
54. Shi W., Kong X.-Y., Wang Z.-M., Yu S.-S., Chen H.-X., De Stasio E.A. 2013. Pause-melting misalignment: A novel model for the birth and motif indel of tandem repeats in the mitochondrial genome. *BMC Genomics*. **14**, 103. doi 10.1186/1471-2164-14-103