Recombination and Evolution of Duplicate Control Regions in the Mitochondrial Genome of the Asian Big-Headed Turtle, *Platysternon megacephalum*

Chenfei Zheng, Liuwang Nie*, Jue Wang, Huaxing Zhou, Huazhen Hou, Hao Wang, Juanjuan Liu

Provincial Key Lab of the Conservation and Exploitation Research of Biological Resources in Anhui, Anhui Normal University, Wuhu, Anhui, P.R. China

Abstract

Complete mitochondrial (mt) genome sequences with duplicate control regions (CRs) have been detected in various animal species. In Testudines, duplicate mtCRs have been reported in the mtDNA of the Asian big-headed turtle, *Platysternon megacephalum*, which has three living subspecies. However, the evolutionary pattern of these CRs remains unclear. In this study, we report the completed sequences of duplicate CRs from 20 individuals belonging to three subspecies of this turtle and discuss the micro-evolutionary analysis of the evolution of duplicate CRs. Genetic distances calculated with MEGA 4.1 using the complete duplicate CR sequences revealed that within turtle subspecies, genetic distances between orthologous copies from different individuals were 0.63% for CR1 and 1.2% for CR2app:addword:respectively, and the average distance between paralogous copies of CR1 and CR2 was 4.8%. Phylogenetic relationships were reconstructed from the CR sequences, excluding the variable number of tandem repeats (VNTRs) at the 3′ end using three methods: neighbor-joining, maximum likelihood algorithm, and Bayesian inference. These data show that any two CRs within individuals were more genetically distant from orthologous genes in different individuals within the same subspecies. This suggests independent evolution of the two mtCRs within each *P. megacephalum* subspecies. Reconstruction of separate phylogenetic trees using different CR components (TAS, CD, CSB, and VNTRs) suggested the role of recombination in the evolution of duplicate CRs. Consequently, recombination events were detected using RDP software with break points at ≈290 bp and ≈1,080 bp. Based on these results, we hypothesize that duplicate CRs in *P. megacephalum* originated from heterological ancestral recombination of mtDNA. Subsequent recombination could have resulted in homogenization during independent evolutionary events, thus maintaining the functions of duplicate CRs in the mtDNA of *P. megacephalum*.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: lwnie@mail.ahnu.edu.cn

Introduction

Gene content, genomic architecture and gene strand asymmetry have all been reported to be stable in the mitochondrial (mt) genomes of deuterostomes, particularly the vertebrates, according to the genome stabilization theory [1]. However, as more mt genomes are sequenced, genetic variations in the control regions (CRs) have been detected in many different species, including sea cucumbers, arthropods, ostracods, fish, frogs, turtles, snakes, lizards, and birds [2–11]. Because control regions are large non-coding regions (LNR) that may participate in mitochondrial genome replication and regulation of transcription, genetic variances such as duplication or degeneration occurring within CRs have significant evolutionary value in the mitochondria [3]. Although in most animals, mitochondrial genomes usually contain only one CR, duplication of mitochondrial CR does occur in more species than previously thought [12]. Even triplicate CRs have been detected in the velvet worm, *Metaperipatus inae* [13].

Duplicate CRs with highly similar sequences are proposed to have evolved in a concerted fashion [3,4,14–16]. Despite the increasing reports of concerted evolution of CRs, contradictory data have been reported in the mt genome of the *Amazona* parrot,

albatross, and three related sea birds belonging to the family Sulidae [15–17]. Eberhard documented the concerted evolution of CR within *Amazona* parrots and reported that both paralogous copies of duplicate CRs in an individual were more closely related to each other than orthologous copies of same CR in different individuals [15]. However, three exceptions to this discovery have been reported within subspecies, *Amazona ochrocephala*. The orthologous copies of CR1 and CR2 from different individuals are genetically closer to each other than to the duplicate copies (CR1 and CR2) within the same individuals. This could be attributed to independent evolution of the two copies after an ancient duplication event although the mechanism behind this is not clearly understood. Contrasting evolutionary patterns have been reported in the albatross, *Thalassarche albatrosses*, where one part of CRs evolved independently while the other showed concerted evolution [16]. Variation in duplicate CRs was detected only in two 327 bp fragments, which are referred in this study as F1 (CR1) and F2 (CR2) in five *Thalassarche* species. Conflicting phylogenetic signals were observed when different parts of these fragments were studied. The 110 bp sequence in both F1 and F2 appear to have evolved independently, but the remaining 217 bp

indicate concerted evolution. Similar patterns were detected in sea birds [17]. The 51 bp at the 5′ end of the sea birds' CRs exhibited independent evolutionary signals but the duplicate CRs evolved concertedly as a whole. These conflicting evolutionary patterns of CR duplication warrant further investigation to clarify the phylogenetic relationships by investigating multiple recombination sites [16,17].

Thus far, gene rearrangement events in mt genomes of Testudines have been discovered in only two turtles: the pancake turtle, *Malacochersus tornieri*, and the Asian big-headed turtle, *Platysternon megacephalum* [7,8,18]. The *P. megacephalum* is the sole living representative species of the Platyternon family and has an mt genome containing a segment with six-genes rearrangement ($tRNA^{His}/tRNA^{Ser}/tRNA^{Leu}/nad\ 5$ and $tRNA^{Thr}/tRNA^{Pro}$) and duplicate CRs [7,8]. The evolution of CR duplication is unclear in *P. megacephalum*. To fill this lacuna, we collected 20 individuals belonging to three subspecies of the Platyternon family and determined the evolutionary patterns of the duplicate CRs in *P. megacephalum*. Finally, we discuss the possible mechanisms by which the duplicate CRs may have originated and evolved in *P. megacephalum* mtDNA.

Materials and Methods

Ethics statement

AH1, AH2, AH3 individuals were collected from Mount Huangshan in Anhui Province. Individual Z2 was collected from Mount Huangshan that is adjacent to Zhejiang Province. Specimens from Fujian Province were provided by the Museum of Fujian Normal University. Individuals from Guangdong, Guangxi, and Yunnan Provinces were provided by Kunming Institute of Zoology, CAS, Yunnan Province.

Procedures involving animals and their care were consistent with NIH guidelines (NIH Pub. No. 85–23, revised 1996) and approved by the Animal Care and Use Committee of Anhui Normal University under approval number #20110710. Tissue samples were collected from the tails (3–5 mm from the tip) of living turtles (AH1, AH2, AH3, Z2) at the sampling location using procedures that minimized pain. These tail samples were preserved in liquid nitrogen until use. The wounds were then sterilized with 75% ethanol, and dressed with gauze and absorbent cotton wool. The turtles were then immediately released into a local habitat. Tissue samples were collected from the muscles of cryopreserved samples.

Sample collection and sequencing of duplicate CRs from the mt genome

Twenty specimens from each of the three turtle subspecies were collected from southern China and boundary areas adjacent to Vietnam and Myanmar (Table 1). These were chosen based on morphological characteristics described previously (*P.m. megacephalum* Gray 1831, *P.m. peguense* Gray 1870, *P.m. shiui* Ernst and McCord 1987) [19–22].

Genomic DNA extraction was performed using the phenol-chloroform method. Two pairs of primers were designed for each CR based on published mt genome sequences of *P. megacephalum* (DQ016387, DQ256377 [7,8]) (Table 2). Outer primers were used to determine whether the duplicate CRs occurred in all individuals at the same locations as previously described [7,8]. For CR1 whose position in the vertebrate mt genome is known, one outer primer (G17) was designed from the *cyt b* gene and the other (G20) was designed from the *12s rRNA* gene. A second pair of outer primers, G21 and G22, was designed to amplify the duplicate CR (CR2), located between the *nad1* and *nad2* genes. The amplification

Table 1. Locations of sample collection.

Subspecies	Individual codes	Locations	Abbreviation
P.m. megacephalum	A1	Anhui, China	AH
	A2	Anhui, China	AH
	A3	Anhui, China	AH
	Z2	Zhejiang, China	ZJ
	FJ	Fujian, China	FJ
	X1	Fujian, China	FJ
	X4	Fujian, China	FJ
	X5	Fujian, China	FJ
	X7	Fujian, China	FJ
	X8	Fujian, China	FJ
	X9	Fujian, China	FJ
P.m. shiui	YN	Guangxi, China (adjacent to Vietnam)	VN
	X6	Guangxi, China (adjacent to Vietnam)	VN
	YP	Guangdong, China	GD
	Y1	Guangdong, China	GD
	Y2	Guangdong, China	GD
P.m. peguense	TP	Yunnan, China (adjacent to Myanmar)	MY
	T1	Yunnan, China (adjacent to Myanmar)	MY
	ZP	Guangxi, China	GX
	ZJ	Guangxi, China	GX

doi:10.1371/journal.pone.0082854.t001

products of the outer primers were used as templates for the inner primers to ensure specificity. For CR1, inner primers, CR1S and CR1X, were located between $tRNA^{Pro}$ and $tRNA^{Phe}$ genes. The second pair of inner primers, CR2S and CR2X, was designed to amplify CR2 located between $tRNA^{Pro}$ and $tRNA^{Gln}$ genes. Long-template (LT) PCRs with outer primers were performed in a total volume of 25 μL containing 100 ng genomic DNA, 1.25 U LA Taq DNA polymerase (TaKaRa Co., Ltd, Dalian, China), 2.5 μL 10× PCR buffer, 3 mM MgCl$_2$, 0.4 mM of each nucleotide and 0.1 μM of each primer. PCR cycling was performed as follows: 94°C for 1 min for pre-denaturation plus 38 cycles of 94°C for 10 s; 50–52°C for 5 min followed by an additional extension at 72°C for 10 min. PCR products were purified using an Axygen DNA Gel Extraction Kit. Then nested PCR was performed with inner primers and products of LT PCR as templates in a 25 μL reaction volume consisting of 100 ng template DNA, 1.25 U Taq DNA polymerase (TaKaRa Co., Ltd, Dalian, China), 2.5 μL 10× PCR buffer, 3 mM MgCl$_2$, 0.4 mM of each nucleotide, and 0.1 μM of each primer. Nested PCR was performed as follows: 94°C for 1 min for pre-denaturation plus 94°C for 10 s and 50–52°C for 5 min for 30 cycles followed by an additional extension at 72°C for 10 min.

Finally, the purified nested PCR products were cloned into a pM-18T vector (TaKaRa Co., Ltd, Dalian, China) and were transformed into *Escherichia coli* strain, DH5α. Plasmids isolated from single clones were sequenced in both directions to ensure the

Table 2. Specific primers used to amplify duplicate CRs in *Platysternon megacephalum*.

Gene	Name	Sequences	Annealing temperature
CR1	G27	AGCAGCCTCCATCCTWTACTT	50–53°C
	G20	ATTGGCTACACCTTGACCTGAC	
	CR1S	AACCACATCATTCCGACCAC	50–53°C
	CR1X	CGGCTCCTTTGTCTAATAG	
CR2	Z22	TTTCTAACCGCACCTCAC	52–53°C
	G22	GTAGTTGGGTTTGGTTTAATCC	
	CR2S	CCCGAATAATAGACTCAACC	50–52°C
	CR2X	CGTGTGGGTTCATTAGTAAAGG	

fidelity of each sequence using the ABI 3700 automated sequencer (Sangon Biotechnology Co., Ltd, Shanghai, China). All sequences were aligned using BioEdit 7.0 and ClustalX 1.8 software [23,24]. Then all sequences were compared against sequences in GenBank using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Sequences with mutations were re-amplified using high-fidelity polymerase (TransGen Co., Ltd, Beijing, China), then re-cloned and re-sequenced. After ensuring their homology with published sequences by BLAST comparisons, all sequences were deposited in GenBank (Accession numbers KC476449-KC476496).

Phylogenetic analysis

Positions of the duplicate control regions were identified from locations of tRNA genes ($tRNA^{Pro}$ and $tRNA^{Phe}$ for CR1; $tRNA^{Pro}$ and $tRNA^{Gln}$ for CR2) in the genome using tRNAscan-SE 1.21. (http://lowelab.ucsc.edu/tRNA_Scan-SE). Three functional regions including the terminal-associated sequence (TAS), central conserved domain (CD), and conserved sequence blocks (CSB), were identified in duplicate CRs through recognition of sequences similar to those found in other turtles [25,26]. The mean genetic distances between CRs in all individuals were calculated from complete CR sequences using the Kimura-2 parameter model in the MEGA 4.1 program [27]. Variable number tandem repeats (VNTRs) have been reported to exist only at the 3′ end of the CR1 CSB region [7,8]. Therefore, phylogenetic analyses were performed in CR sequences excluding the VNTRs, using the neighbor-joining (NJ) and maximum likelihood (ML) algorithms with PAUP*4.0 b10 [28]. Bayesian inference (BI) analyses were performed using MrBayes ver.3.2.1 on 2000,000 generations and 100 sampled generations [29]. The robustness of the NJ and ML trees was estimated using bootstrap proportion (BP) with 1000 replicates. Statistical support for the BI tree was evaluated with the posterior probability (PP). *Chrysemys picta* was designated as the outgroup in the phylogenetic analyses. To understand the evolutionary pattern in detail, unrooted phylogenetic trees were reconstructed using separate regions (TAS, CD, CSB, and VNTRs) based on BI analyses. Nucleotide substitution model (HKY85) was made with jModeltest Version 2.1.3 [30] using Akaike Information Criterion correction (AICc) prior to phylogenetic reconstruction.

Recombination testing

Discordant evolutionary signals were detected when phylogenetic trees were reconstructed with different regions separately. These conflicting signals may be due to recombination of duplicate CRs. To further analyze these signals, four recombination tests for duplicate CRs were conducted: 1) Recombination detection program (RDP) [31], 2) Geneconv [32], 3) Maxchi [33] and 4) Chimaera [34]. Because of the heterologous nature of the VNTR lengths, it was necessary to collect sequences with same base pair lengths for alignment. Therefore, we chose only 4 individuals (AH2, YP, ZP, and TP) to represent the three subspecies. These analyses were performed using the RDP software with the parameters described previously [35]. Similar analyses were carried out for all individuals to make sure whether recombination is widespread in mtCRs of this turtle. All analyses were performed 3 times.

Results

Duplicate control regions in mtDNA of Platysternon megacephalum

PCR of this turtle's mtDNA using outer primers amplified duplicate regions that contained both *cyt b* and *12s rRNA* for CR1 and both *nad1* and *nad2* for CR2. This showed that duplicate CRs were present in all individuals collected. The positions of duplicate CRs as described by Peng *et al.* and Parham *et al.* [7,8] are shown in Fig. 1. The length of CR1 ranged from 1,222–1,608 bp and CR2 ranged from 1,128–1,420 bp. Because of the absence of VNTRs in CR2, sequences in the 3′ end of CR1 and CR2 could not be aligned properly. However, approximately 800 bp of the 5′ end of CR1 and CR2 was aligned properly with few indels, as descried previously [7,8].
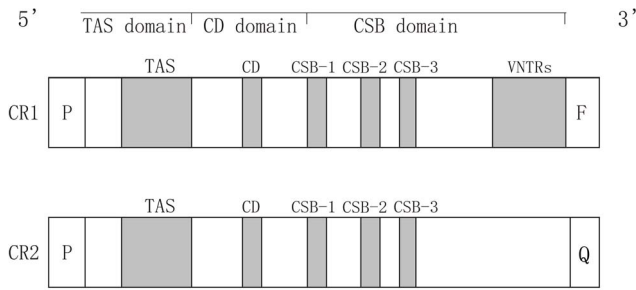
Structures of duplicate CRs

Three conserved functional sections including the TAS, CD, and CSB domains were detected in both CR1 and CR2 in all individuals (Fig. 2). The TAS domain is in general affiliated with termination of the H-strand by forming a very stable hairpin structure between its core sequence (5′-TACAT-3′) and the reversed complementary sequence (5′-ATGTA-3′). The central conserved domain (CD) is located downstream of the TAS domain and is 459 bp long. It is characterized by the conserved sequence 5′-AATCACGAGAGATAAGCAA-3′. All three conserved sequences, CSB-1, CSB-2, and CSB-3 were found in the 5′ end of the CSB domain and were recognized by the presence of 5′-TTAATGCTTGTTAGACATA-3′ for CSB-1, 5′-TTAAACCCCCCTACCCCCC-3′ for CSB-2 and 5′-TCGTCAAACCCCTAAATCC-3′ for CSB-3 [25,26,36]. All core sequences were identical in all duplicate CRs, and there was no variation among the 20 individuals.
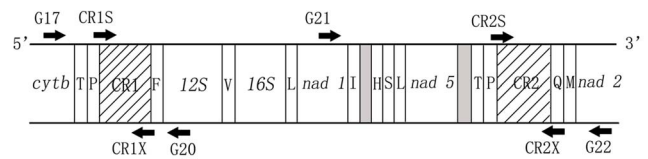

Figure 1. Locations of duplicate CRs and primers used for amplification. G17/G20 are outer primers for CR1; G21/G22 are outer primers for CR2. These primer pairs were used to determine the presence of CRs in the mtDNA of individual turtles. Inner primers, CR1S/CR1X and CR2S/CR2X, are designed to amplify specific duplicate CRs. The names of tRNA genes are shown using abbreviations of amino acids (T: $tRNA^{Thr}$, P: $tRNA^{Pro}$, F: $tRNA^{Phe}$, V: $tRNA^{Val}$, L: $tRNA^{Leu}$, I: $tRNA^{Ile}$, H: $tRNA^{His}$, S: $tRNA^{Ser}$, Q: $tRNA^{Gln}$, M: $tRNA^{Met}$). Gray boxes represent non-coding spaces between two genes. CR1 and CR2 are indicated with slashes.

Figure 2. Structures of duplicate CRs in individuals. Three functional regions, TAS, CD, and CSB, were detected in both CRs in all individuals. The core sequences of these regions were found to be identical in CR1 and CR2. Variable numbers of tandem repeats (VNTRs) were detected only in CR1.
doi:10.1371/journal.pone.0082854.g002

In the 3′ end of CSB, VNTRs with heterology in length and motif were detected in the mtDNA of this turtle as previously reported [7,8]. In all individuals, CR2 had a specific sequence that was 411 bp in length after the 3′ end of CSB. All these sequences were aligned with few indels and only 60 bp were variable in different individuals. However, significant differences in CR1 sequences were detected with respect to both length and motif. The length of VNTRs varied from 77 bp to 786 bp leading to length heterology of mtDNA. The first repeat unit, TATGTTA, was detected in all individuals, and the AT motif was the last repeat motif in all individuals. Although the beginning and ending units were identical in all subspecies, the remaining repeat units contained subspecies-specific heterology in both length and motif sequence. In CR1 within mtDNA of subspecies, *P.m. peguence*, the repeat motifs were relatively simple and contained TATTA (repeated 6–8 times) and TATGTTATATATAATATAT (repeated 2–3 times). However, the repeat units in *P.m. megacephalum* and *P.m. shiui* had similar motifs (TGTTATAGTA and TGTGTTATATATAATATAT) but differed only in the number of repeats (Fig. 3).
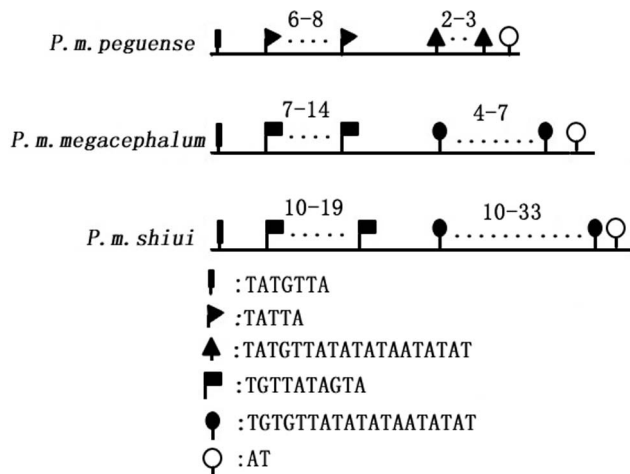
Figure 3. Heterology of VNTRs in different subspecies. The line at the bottom of the image indicates the lengths of VNTRs. The numbers above the dots represent the copy numbers of motifs.
doi:10.1371/journal.pone.0082854.g003

Phylogenetic analyses of duplicate CRs

The genetic distances calculated by MEGA 4.1 software using the complete sequences of duplicate CRs (Table 3) suggested that orthologous copies were more similar than paralogous copies. Because of the absence of VNTRs in the 3′ end of the CSB domain, the genetic distance of these regions showed significantly more independent evolutionary signals with dDup = 36.4%, while dInt was only 0.7%.

Phylogenetic trees generated using the NJ, ML, and BI methods showed consensus in almost all the nodes and revealed three major clades (Fig. 4). These clades were consistent with the subspecies taxonomy with clades A and B representing *P.m. megacephalum* and *P.m. shiui*, respectively. These two subspecies were more phylogenetically familiar than the subspecies *P.m. peguense* which is represented by clade C in the phylogenetic trees. In each clade, the orthology of CR1 or CR2 from different individuals were grouped together. Each CR was distinct from the placement of paralogous copies within same individuals. Although these phylogenetic trees share similar topology in general, differences did exist (Fig. 4). For example, in NJ trees, CR1 were distinguished from CR2 in all clades without exceptions although some CR2 in clade A from the ML and BI trees showed parallel branching. CR1 and CR2 from the same individuals were always separate from each other and did not form clusters.

Conflicting evolutionary signals were observed when phylogenetic trees were reconstructed separately using the divided regions (Fig. 5). Phylogenetic trees of TAS, CD, and CSB sequences shared the same topologies, showing that paralogous copies from the same individuals always clustered together rather than with orthologous copies from different individuals. However, the opposite was observed in the phylogenetic tree built using VNTRs. Orthologous copies were more closely related to their counterparts from different individuals rather than to paralogous copies from the same individual. In addition, the divergence of duplicate CRs occurs earlier than the differentiation of subspecies, as shown in the VNTRs tree (Fig. 5).

To explain the presence of discordant signals in the phylogenetic trees constructed separately using the three different regions, multiple recombination points were investigated using the RDP software (Fig. 6). In total, 4 recombination events were detected in all tested individuals. For example, CR1, from individual AH2, displayed very high pairwise identity in common with its paralogous copy CR2, but very low pairwise identity compared to CR2 from the individual, TP. However, recombination events in AH2 were indicated by all four analyses (RDP, $P = 1.576 \times 10^{-13}$, MaxChi, $P = 5.035 \times 10^{-14}$, Chimaera, $P = 3.915 \times 10^{-13}$, Geneconv, $P = 1.622 \times 10^{-10}$). Similar recombination signal were also detected in other individuals (not shown) and the positions of recombination breakpoints that occurred in duplicate CR, without exception, were about 290 bp and 1080 bp.

Discussion

Evolutionary pattern of duplicate CRs in P. megacephalum

The mt gene order is highly conserved in Testudines [25]. Thus far, only two mt genomic rearrangements have been reported in two species, the pancake turtle and the Asian big-headed turtle [7,8,18]. However, whether duplicate CRs evolved concertedly as they did in other deuterostomes is unclear.

The mean genetic distances between duplicate CRs indicated that orthologous copies (dDup_CR1 = 0.63%; dDup_CR2 = 1.2%) were more similar to each other than to their paralogous counterparts (dDup = 4.8%) within each subspecies

Table 3. Mean genetic distance calculated using MEGA 4.1 software and different parts of CRs.

		TAS		CD		CSB		VNTRs		TOTAL[1]	
		dInt[2]	dDup[3]	dInt[2]	dDup[3]	dInt[2]	dDup[3]	dInt[2]	dDup[3]	dInt[2]	dDup[3]
P.m.	CR1	0.020	0.016	0.005	0.007	0.001	0.001	0.010	0.362	0.007	0.048
megachephalum	CR2	0.010		0.008		0.000		0.009		0.007	
P.m .shiui	CR1	0.008	0.052	0.011	0.011	0.005	0.003	0.004	0.378	0.008	0.052
	CR2	0.013		0.010		0.000		0.009		0.009	
P.m. peguense	CR1	0.008	0.064	0.002	0.005	0.003	0.002	0.005	0.352	0.004	0.043
	CR2	0.117		0.008		0.000		0.005		0.020	

[1]Total: Genetic distances calculated using complete sequences.
[2]dInt: Genetic distances of orthologous copies from different individuals.
[3]dDup: Genetic distances of paralogous copies from duplicate CRs within the same individual.
doi:10.1371/journal.pone.0082854.t003

(Table 3). In addition, phylogenetic trees reflected a closer relationship among single CRs in different individuals than between two copies of CRs within same individual of a subspecies. In this way, orthologous CRs from different individuals were genetically more similar than paralogous CRs from the same individual within any subspecies clade. This evolutionary pattern indicated independent evolution of duplicate CRs within every subspecies of *P. megacephalum*; that is CRs evolution was not affected by each other. However, it appears that duplicate CRs evolved concertedly among different subspecies, considering that
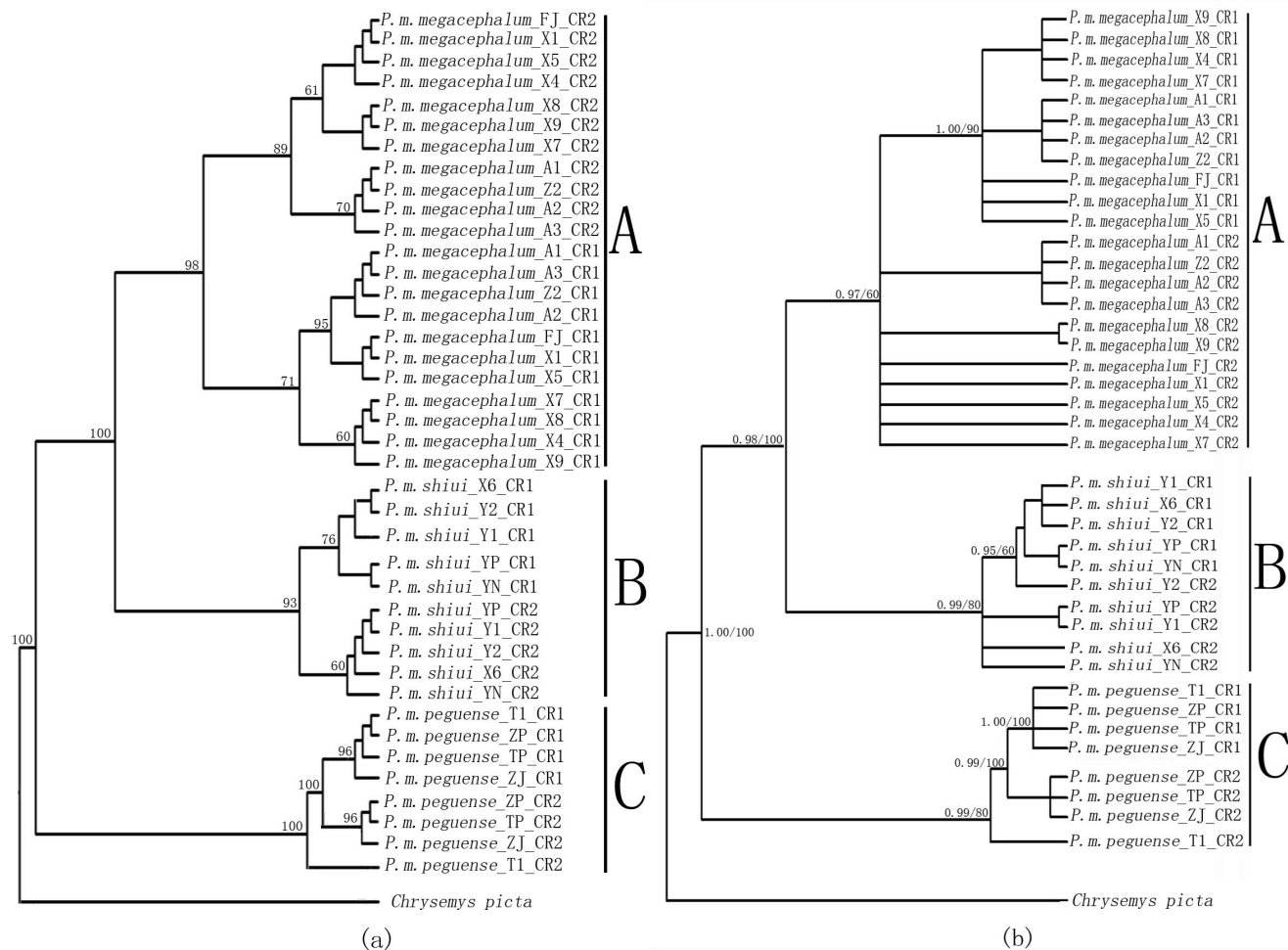

Figure 4. Phylogenetic trees reconstructed using (a) neighbor-joining, (b) maximum likelihood and Bayesian methods. Three clades were grouped. These contained the three subspecies of *Platysternon megacephalum*: (A) *P.m. megacephalum*, (B) *P.m. shiui*, and (C) *P.m. peguense*.
doi:10.1371/journal.pone.0082854.g004
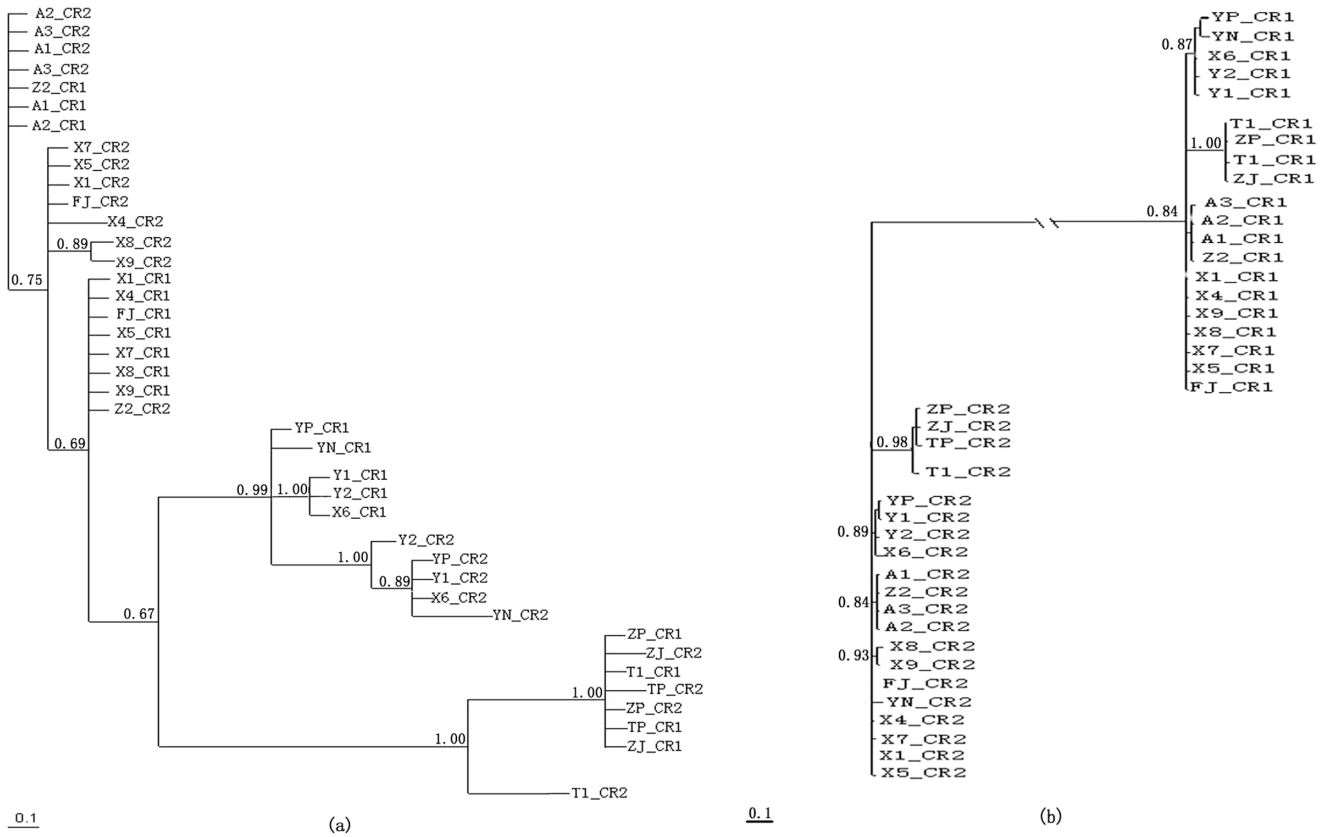
Figure 5. Phylogenetic trees reconstructed using (a) TAS, CD, and CSB domains and (b) VNTRs. All three phylogenetic trees constructed separately based on TAS, CD, and CSB shared similar topology and therefore only one tree is presented here.
doi:10.1371/journal.pone.0082854.g005

both CR1 and CR2 from each subspecies were more similar or more closely related to each other than either of their own counterparts from other subspecies. In addition, duplicate CRs

divided with each other much earlier than the differentiation of subspecies as shown in phylogenetic trees constructed using VNTRs (Fig. 5b). Based on these findings, the following
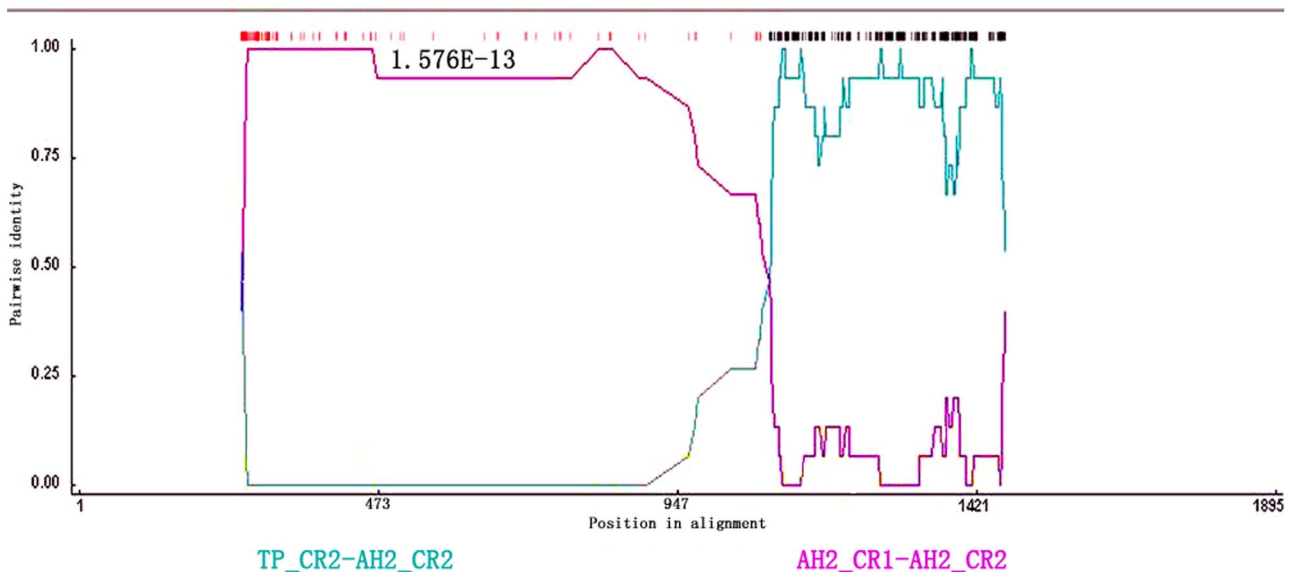
Figure 6. Assessment of recombination in duplicate CRs detected by RDP software in an individual, AH2. Green shade indicates pairwise identity between orthologous CR2 genes in individuals, TP and AH2. Purple shade depicts pairwise identity between paralogous CRs from AH2. Similar analysis was performed for all individuals.
doi:10.1371/journal.pone.0082854.g006

evolutionary scenarios were deduced: 1) originally, the ancestors of Asian big-headed turtles had only one CR in their mtDNA. After a mutation event, the CR could have duplicated, which may have offered some advantages to this turtle. Because of this, the mtDNA with duplicate CRs may have been preserved during reproduction. 2) mtDNA with duplicate CRs may have been acquired by subspecies during the separation process. This could be another way the duplicate CRs could have evolved among different subspecies in a concerted manner. The duplicate CRs could play different roles in the replication of mtDNA so that the duplicate CRs may have evolved independently within any subspecies. Whether different functions are performed by the duplicate CRs requires further investigation at the transcriptome and proteome levels.

Possible role of recombination in the evolutionary pattern of duplicate CRs

The data collected here generated specific questions such as how did the ancestral mtDNA of Asian big-headed turtles obtain duplicate CRs. It has been suggested that tandem duplication, dimerization, or illegitimate recombination could have played a role in that process [1]. Tandem duplication is caused by errors in replication, such as slipped strand mispairing, over-running the terminal signal of the mt genome, and initiation of replication from a secondary structure [1,37–40]. Two linearized monomeric mt genomes could have dimerized through head to tail joining to form a large circular mt genome containing two copies of each gene, including the CR. Finally, illegitimate recombination may have given rise to tandem duplication when a section mt genome was spliced out and introduced into another region within the same mitochondrion [41]. Based on our data from this study, duplicate CRs in mtDNA of *P. megacephalum* could have originated by heterologous recombination between two divergent CRs.

The discordant signals revealed by phylogenetic trees using different CR components suggested that recombination could have influenced the evolution of duplicate CRs in mtDNA of *P. megacephalum*. Previously, Abbott proposed a contrasting pattern of CR evolution in *Thalassarche albatrosses*: a section containing two copies could evolve independently and another section could evolve in a concerted fashion [16], which was indicated by a specific point of recombination found in F1. A recent study on seabirds (Sulidae) suggested that the 5′ end of CRs evolved independently while the duplicate CRs, as a whole sequence, evolved concertedly [15]. Further, Morris-Pocock *et al.* proposed that differences in CR evolution could be difficult to resolve without testing multiple recombination points, as described by Abbott [16,17].

Moreover, four recombination points were tested using RDP software with specific attention to distinct signals for the evolution of duplicate CRs. These data showed the break points of recombination at 290 bp and 1,080 bp upstream of the TAS domain or the VNTRs, respectively. Sequencing of the complete mtDNA of two Philippine hornbill species (Aves: Bucerotidae) showed a replication fork barrier (RFB) downstream of ETAS [42]. During each replication cycle, the 3′ end of the nascent L-strand was suspected to remain free at the RFB region until replication restarted. During the relatively long period of exposure, the free strands could be easily exchanged, leading to a high rate of recombination [42].

Heterologous sequences are considered essential to mtDNA recombination [43]. In the mtDNA of flounder, *Platichthys flesus*, two heterogeneous patterns (C and T arrays) were found in the VNTRs motif all 168 individuals analyzed [44]. Most individuals had only pure C or T arrays, but one individual had a compound

CT array. This is a direct evidence of recombination between mtDNA in the flounder. Other evidence of heterologous recombination of mtDNA has been reported in various animals [45–48]. Our data showed that the sequences at the 3′ end of the CSB domain of CR1 and CR2 differed significantly. Specifically, the VNTRs in the CR1 of different subspecies were heterologous with respect to the length and motif sequences. These findings may explain how recombination of duplicate CRs was induced in the **mtDNA of P. megacephalum**

The data from this study facilitates the hypothesis that CR1 and CR2 could have originated from different ancient individuals with only one CR, and then crossing over may have occurred between heterologous mt genomes, allowing the insertion of several genes containing CRs into other genomes through an occasional paternal linkage [49]. Such genetic variation could be maintained in the mt genome of *P. megacephalum* over time due to potential advantages by maintaining duplicate CRs [50,51]. In addition, duplicate CRs may contribute to different functions, so that evolutionary pressure that could have existed earlier during evolution to change the CRs may have diminished. This could be a plausible way the CRs would have evolved independently within each subspecies after the divergence of subspecies and accumulated mutations independently over time.

Another question raised by our findings is what maintains over time the main functions of CRs, such as initiation of replication and transcription of mtDNA during the independent evolution of duplicate CRs. It has been proposed that intra-molecular recombination may occur through gene conversion when mtDNA has three strands [14]. The nascent H-strand of one CR then recombines with homologous parental strands of another CR, leading to the homogenization of both CRs. Our data show that homogenization of the TAS, CD, and CSB domains could have occurred via recombination upstream of TAS and VNTRs. During every replication cycle, a nascent H-strand with one CR (CR1 or CR2) may have crossed over to the parental strand of another CR at the break points. Then homogenization may have taken place in the TAS, CD, or CSB domains, leading to concerted evolution of these two areas while allowing independent evolution of TAS and VNTRs. Some essential elements, such as the termination and initiation sites of the H-strand, were included in TAS, CD, and CSB [3]. In this way, sequence homogenization of these three regions would have been critical to ensure the functions of these CRs during periods of independent evolution.

Conclusion

This is the first report describing the evolution of duplicate CRs within the Testudines mt genome. Our findings show that in *P. megacephalum* duplicate CRs evolved independently within subspecies but concertedly across subspecies. Sequence homogenization in duplicated CRs may have resulted from recombination to ensure their function during evolution.

Author Contributions

Conceived and designed the experiments: LN CZ. Performed the experiments: CZ JW. Analyzed the data: CZ JW HZ. Contributed reagents/materials/analysis tools: HH. Wrote the paper: CZ LN. Photographs and Line Drawings: HZ HW JL.

References

1. Boore JL (2000) The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. In: SankoffD, NadeauJH. (Eds.), Comparative Genomes: Empirical and Analytical Approaches to Gene Order Dynamics, Map Alignment and the Evolution of Gene Families. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 133–148.

2. Fan SG, Hu CQ, Wen J and Zhang LP (2011) Characterization of mitochondrial genome of sea cucumber *Stichopus horrens*: A novel gene arrangement in Holothuroidea. Science China Life Sciences.54: 434–441.

3. Shao R, Barker SC, Mitani H, Aoki Y and Fukunaga M (2005) Evolution of duplicate control regions in the mitochondrial genomes of Metazoa: a case study with Australasian Ixodes ticks. Mol.Biol.Evol. 22: 620–629.

4. Ogoh K and Ohmiya Y (2004) Complete mitochondrial DNA sequence of the sea-firefly, *Vargula hilgendorfii* (Crustacea, Ostracoda) with duplicate control regions. Gene.327: 131–139.

5. Zhuang X and Cheng CHC (2010) ND6 Gene "Lost" and Found: Evolution of Mitochondrial Gene Rearrangement in Antarctic Notothenioids. Mol.Biol. Evol.27:1391–1403.

6. Zhang JF, Nie LW, Wang Y and Hu LL (2009) The complete mitochondrial genome of the large-headed frog, *Limnonectes bannaensis* (Amphibia: Anura), and a novel gene organization in the vertebrate mtDNA. Gene. 442:119–127

7. Peng QL, Nie LW and Pu YG (2006) Complete mitochondrial genome of Chinese big-headed turtle, *Platysternon megacephalum*, with a novel gene organization in vertebrate mtDNA. Gene. 380: 14–20.

8. Parham JF, Feldman CR. and Boore JL (2006a) The complete mitochondrial genome of the enigmatic big-headed turtle (*Platysternon*): Description of unusual genomic features and the reconciliation of phylogenetic hypotheses based on mitochondrial and nuclear DNA. BMC Evol. Biol. 6: 11.

9. Jang HK and Hwang UW (2011) Complete mitochondrial genome of the black-headed snake *Sibynophis collaris* (Squamata, Serpentes, Colubridae). Mitochondrial DNA. 22: 77–79.

10. Okajima Y and Kumazawa Y (2010) Mitochondrial genomes of acrodont lizards: timing of gene rearrangements and phylogenetic and biogeographic implications. BMC Evol. Biol. 10: 141.

11. Verkuil IY, Piersma T and Baker JA (2010) A novel mitochondrial gene order in shorebirds (Scolopacidae, Charadriiformes). Mol. Phylogenet. Evol. 57: 411–416.

12. Schirtzinger EE, Tavares ES, Gonzales LA, Eberhard JR, Miyaki CY, et al. (2012) Multiple independent origins of mitochondrial control region duplications in the order *Psittaciformes*. Mol. Phylogenet. Evol.64: 342–356.

13. Braband A, Podsiadlowski L, Cameron SL, Daniels S and Mayer G (2010) Extensive duplication events account for multiple control regions and pseudogenes in the mitochondrial genome of the velvet worm *Metaperipatus inae* (Onychophora,Peripatopsidae). Mol.Phylogenet.Evol.57: 293–300.

14. Kumazawa Y, Ota H, Nishida M and Ozawa T (1998) The complete nucleotide sequence of a snake (Dinodon semicarinatus) mitochondrial genome with two identical control regions. Genetics 150: 313–329.

15. Eberhard JR, Wright TF and Bermingham E (2001) Duplication and concerted evolution of the mitochondrial control region in the parrot genus Amazona. Mol.Biol.Evol. 18: 1330–1342.

16. Abbott CL, Double MC, Trueman JWH, Robinson A and Cockburn A (2005) An unusual source of apparent mitochondrial heteroplasmy: duplicate mitochondrial control regions in *Thalassarche albatrosses*. Mol. Ecol.14: 3605–3613.

17. Morris-Pocock JA, Taylor SA, Birt TP and Friesen VL (2010) Concerted evolution of duplicated mitochondrial control regions in three related seabird species. BMC Evol. Biol. 10: 14.

18. Parham JF, Macey JR, Papenfuss TJ, Feldman CR, Türkozan O, et al. (2006b) The phylogeny of Mediterranean tortoises and their close relatives based on complete mitochondrial genome sequences from museum specimens. Mol.Phylogenet.Evol.38: 50–64.

19. Gray JE (1831) Characters of a new genus of freshwater tortoise from China. Proceedings of the Zoological Society of London 831:106–107.

20. Gray JE (1870) Supplement to the catalogue of shield reptiles in the collection of the British Museum. part 1.Testudinata(tortoises).Taylor and Francies,London,120 pp.

21. Ernst CH and McCord WP (1987) Two new turtles from south Asia. Proceedings of the biological society of Washington. 100:624–628.

22. Ernst CH and Laemmerzahl AF (2002) Geographic variation in the Asian big-headed turtle, Platysternon megacephalum (Reptilia: Testudines: Platysternidae). Proceedings of the biological society of Washigton. 115: 18–24

23. Hell TA (1999) Bioedit A user friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucl. Acids Sym. Posit. Series 41: 95–98.

24. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F and Higgins DG (1997) The ClustalX windows in terface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research. 24: 4876–4882.

25. Xiong L, Nie LW, Li XS and Liu X (2010) Comparison research and phylogenetic implications of mitochondrial control regions in four soft-shelled turtles of Trionychia (Reptilia, Testudinata). Genes & Genomics. 32: 291–298.

26. Zhang YY, Nie LW, Huang YQ, Pu YG and Zhang L (2009) The Mitochondrial DNA Control Region Comparation Studies of Four Hinged Turtles and Its Phylogentic Significance of the Genus Cuora sensulato (Testudinata: Geoemydidae). Genes & Genomics. 31: 349–359.

27. Kumar S, Tamura K and Nei M (2004) An integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. 5: 150–163.

28. Swofford DL (2002) PAUP*:Phylogenenc Analysis using Parsimony (*and Other Methods), Version 4.0b l0. Sunderland: Sinauer Associates.

29. Huelsenbeck J and Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. Bioinformatics. 17: 754–755.

30. Darriba D, Taboada GL, Doallo R and Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9(8): 772.

31. Martin D and Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. Bioinformatics.16: 562–563.

32. MaynardSmith J (1992) Analysing the mosaic structure of genes. J.Mol.Evol. 35: 126–129.

33. Sawyer S (1989) Statistical tests for detecting gene conversion. Mol.Biol.Evol. 6: 526–538.

34. Posada D and Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc.Natl.Acad.Sci.USA.98: 13575–13762.

35. Ujvari B, Dowton M and Madsen T (2007) Mitochondrial DNA recombination in a free-ranging Australian lizard. Biology Letters. 3: 189–192.

36. Li XS, Nie LW, Wang L, Xiong L and Zhou K (2010) The mitochondrial genome complete sequence and organization of the Pig-nosed Turtle *Carettochelys insculpta*(Testudines, Carettochelyidae) and its phylogeny position in Testudines. Amphibia-Reptilia. 31: 541–551.

37. Levinson G and Gutman GA (1987) Slipped-strand mispairing: A major mechanism for DNA sequence evolution. Mol.Biol.Evol. 4(3): 203–221.

38. Madsen CS, Ghivizzani SC and Hauswirth WW (1993) Protein-binding to a single termination-associated sequence in the mitochondrial DNA D-loop region. Mol. Cell Biol. 13: 2162–2171.

39. Lunt DH and Hyman BC (1997) Animal mitochondrial DNA recombination. Nature.387: 247–247.

40. Zhang J (2003) Evolution by gene duplication: AN update. Trends in Ecology & Evolution.18: 292–298.

41. Macey JR, Larson A, Ananjeva NB, Fang Z and Papenfuss TJ (1997) Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. Mol.Biol.Evol. 14: 91–104.

42. Sammler S, Bleidorn C and Tiedemann R (2011) Full mitochondrial genome sequences of two endemic Philippine hornbill species (Aves: Bucerotidae) provide evidence for pervasive mitochondrial DNA recombination[J]. BMC genomics, 12(1): 35.

43. White DJ, Wolef JN, Pierson M and Gemmell NJ (2008) Revealing the hidden complexities of mtDNA inheritance. Mol.Ecol. 17: 4925–4942.

44. Hoarau G, Holla S, Lescasse R, Stam WT and Olsen J (2002) Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish *Platichthys flesus*. Mol. Biol. Evol. 19(12): 2261–2264.

45. Kajander OA, Rovio AT, Majamaa K, Poluton J, Spelbrink JN, et al. (2000) Human mtDNA sublimons resemble rearranged mitochondrial genomes found in pathological states. Hum. Mol. Genet. 9(19): 2821–2835.

46. Phadnis N, Sia RA and Sia EA (2005) Analysis of repeat-mediated deletions in the mitochondrial genome of Saccharomyces cerevisiae. Genetics. 171(4): 1549–1559.

47. Gibson T, Blok VC and Dowton M (2007) Sequence and characterization of six mitochondrial subgenomes from Globodera rostochiensis: multipartite structure is conserved among close nematode relatives. J. Mol. Evol. 65(3): 308–315.

48. Guo X, Popadin KY, Markuzon N, Orlove YL, Kraytsberg Y, et al. (2010) Repeats, longevity and the sources of mtDNA deletions: evidence from 'deletional spectra'. Trends in Genetics.26(8): 340–343.

49. Gyllensten U, Wharton D, Josefsson A and Wilson AC (1991) Paternal inheritance of mitochondrial DNA in mice. Nature 352, 255–257.

50. Kumazawa Y, Nishida M, Ota H and Ozawa T (1996) Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA gene cluster. Mol.Biol.Evol. 13: 1242–1254.

51. Arndt A and Smith MJ (1998) Mitochondrial gene rearrangement in the sea cucumber genus Cucumaria. Mol.Biol.Evol.15: 1009–1016.