RESEARCH ARTICLE

Seven complete mitochondrial genome sequences of bushtits (Passeriformes, Aegithalidae, *Aegithalos*): the evolution pattern in duplicated control regions

Xiaoyang Wang[1], Yuan Huang[1], Nian Liu[1], Jing Yang[1], and Fumin Lei[1,2]

[1]Co-Innovation Center for Qinba regions' sustainable development, College of Life Sciences, Shaanxi Normal University, Xi'an, China and
[2]Key Laboratory of the Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China

Abstract

The control region (CR) of the mitochondrial DNA exhibits important functions in replication and transcription, and duplications of the CR have been reported in a wide range of animal groups. In most cases, concerted evolution is expected to explain the high similarity of duplicated CRs. In this paper, we present seven complete mitochondrial genome sequences from the bushtits (genus *Aegithalos*), in which we discovered two duplicated CRs, and try to survey the evolution pattern of these duplicated CRs. We also found that the duplicated CRs within one individual were almost identical, and variations were concentrated in two sections, one located between a poly-C site and a potential TAS (termination associated sequence) element, the other one located at the 3′ end of the duplicated CRs. The phylogenetic analyses of paralogous CRs showed that the tree topology were depending on whether the two high variable regions at the upstream of TAS element and the 3′end of duplicated CRs: when they were concluded, the orthologous copies were closely related; when they were excluded, the paralogous copies in the same lineages were closely related. This may suggest the role of recombination in the evolution of duplicated CRs. Consequently, the recombination was detected, and the breakpoints were found at ∼120 bp (the upstream of the potential TAS element) and ∼1150 bp of the alignment of duplicated CRs. According to these results, we supposed that homologous recombination occurred between paralogous CRs from different mtDNA molecule was proposed as the most suitable mechanism for concerted evolution of the duplicated CRs, and the recombination took place in every replication cycle, so that most part of the duplicated regions remain identical within an individual, while the 5′ and 3′end of the duplicated CRs were not involved in recombination, and evolved independently.

Introduction

The control region (CR), also known as the D-loop region, is the main non-coding segment of animal mitochondrial DNA (mtDNA) (Boore, 1999). The CR consists of a regulatory or initiation element for replication and transcription (Boore, 1999; Moritz et al., 1987). In most cases, a normal mtDNA molecule has only one CR, while in some rearranged mtDNA molecules, two (e.g. ticks (Shao et al., 2005), sea cucumbers (Arndt & Smith, 1998), and several groups of vertebrates (Eberhard et al., 2001; Kumazawa et al., 1996; Kurabayashi et al., 2008; Singh et al., 2008; Tatarenkov & Avise, 2007) or three (e.g. thrips (Yan et al., 2012), nematodes (Azevedo & Hyman, 1993), and some frogs (Kurabayashi et al., 2008)) copies of CRs can be found. In many cases, the duplicated CRs exhibit extremely high sequence similarity between the copies, and all the copies appear functional (Eberhard et al., 2001; Tatarenkov & Avise, 2007), which may indicate that these regions have evolved in concert (Tatarenkov & Avise, 2007).

Duplicated CRs are the main type of rearrangement found in avian mitochondrial genomes (mitogenomes). So far as we know in October 2014, four mitochondrial gene orders have been found in birds (Gibb et al., 2007). Most birds present a gene organization similar to that of chicken (*Gallus gallus*, (Desjardins & Morais, 1990)), while other birds present the rearranged types, i.e. tandem duplication of a section including part of *cyt b*, *tRNA^Thr*, *tRNA^Pro*, *nad6*, *tRNA^Glu*, and most part of CR (''duplicated tThr-CR'' named by Gibb et al., 2007) (Abbott et al., 2005; Cho et al., 2009; Morris-Pocock et al., 2010; Sammler et al., 2011; Verkuil et al., 2010); only CR was duplicated (duplicated CR) (Cerasale et al., 2012; Eberhard et al., 2001; Gibb et al., 2007; Singh et al., 2008); and duplicated type with the original CR degenerated (remnant CR2) (Bensch & Härlid, 2000; Cooke et al., 2012; Gibb et al., 2007; Mindell et al., 1998; Morgan-Richards et al., 2008; Pratt et al., 2009; Slack et al., 2007). In addition, the duplicated CRs in rearranged avian mitogenomes are identical with each other across almost the entire sequences including complete Domain II, 3′ end of Domain I and 5′ end of Domain III, whereas 5′ end of Domain I and 3′ end of Domain III show high sequence variation (Cerasale et al., 2012; Morgan-Richards et al., 2008; Sammler et al., 2011; Singh et al., 2008; Verkuil et al., 2010). Thus, the extremely high similarities in most sequences between different copies may indicate that concerted evolution occurs in the

Correspondence: Yuan Huang and Fumin Lei, Co-Innovation Center for Qinba regions' sustainable development, College of Life Sciences, Shaanxi Normal University, Xi'an 710062, China. E-mail: yuanh@snnu.edu.cn (Y.H.); leifm@ioz.ac.cn (F.L.).

duplicated mitochondrial CRs of birds (Morgan-Richards et al., 2008; Sammler et al., 2011; Verkuil et al., 2010). However, recent fixation of duplication could not be rejected, especially for the ''duplicated tThr-CR'' type (Verkuil et al., 2010). In addition, concerted evolution has been detected in the flanking duplicated genes of duplicated CRs (3 tRNAs: *tRNA^Thr*, *tRNA^Pro*, *tRNA^Glu*; 2 protein-coding genes (PCGs): *cyt b* and *nad6*) in birds (Morgan-Richards et al., 2008; Sammler et al., 2011; Verkuil et al., 2010), but its exact mechanism remains unclear.

Several hypotheses have been proposed to interpret the mechanism of concerted evolution, such as gene conversion (Tatarenkov & Avise, 2007), parallel selection (Eberhard et al., 2001), conversion between parental and nascent strands of duplicated control regions during DNA replication within the three-strandructure (Eberhard et al., 2001), exact replication mechanism based on the deletion of one copy of duplicated CRs and replacement by a duplicated copy of another CR in every replication (Ogoh & Ohmiya, 2007), and homologous or illegitimate recombination (Kurabayashi et al., 2008). All of these hypotheses can partially account for concerted evolution, but each involves some points that are difficult to explain. Morris-Pocock et al. (2010) commented that parallel selection can not clarify why putative non-functional portion in CR evolves in concert, while conversion during the initial mtDNA replication of a three-stranded D-loop structure is an unsuitable explanation for the concerted evolution of CR-flanking genes (tRNAs and PCGs) and the high variable region at the 5′ end of the duplicated CR, and homologous or illegitimate recombination requires multiple breakpoints to explain the high variations in both the 5′ and 3′ ends of the CR for the ''duplicated tThr-CR'' type.

So far, duplicated CR is main rearrangement type among avian mitogenome. In this paper, we describe seven complete mitochondrion genomes of bushtits, and try to investigate the evolutionary patterns of duplicated CRs through inferring the phylogenetic relationships among paralogous (i.e. two CRs from the same individual) and orthologous (i.e. the same CR from different individuals) CRs, and detecting evidence that recombination exists between paralogous CRs. Moreover, we discussed the most likely mechanism of concerted evolution and the potential complications that duplicated CRs may bring to phylogenetic and phylogeographic analyses.

Materials and methods

Samples collection, DNA extraction, PCR amplification and sequencing

Samples from a total of seven species and subspecies in *Aegithalos* were collected (Table 1). DNA was extracted using the phenol-chloroform method described by Zhou et al. (2007).

Five pairs of long PCR primers and two pairs of short PCR primers were selected to amplify overlapping fragments of the complete mitochondrial genome, as described by Sorenson et al.

(1999) and Yang et al. (2010) (Supplementary Appendix A). The long PCR products were purified with DNA Gel Purification Kit (U-Gene) after separation through electrophoresis in a 1.0% agarose gel. All of the purified long PCR products were used as the templates for the sub-PCR. The primer pairs for the sub-PCR assays were selected as described by Sorenson et al. (1999) and Yang et al. (2010) (Supplementary Appendix B). The short PCR and sub-PCR products were sequenced directly with the PCR primers on the ABI 3730 XL DNA Analyzer. Some internal sequencing primers were designed if the products were longer than 1400 bp.

Sequences assembly and annotation

Staden package (Staden, 1996) was used for assembling and annotating the complete sequences. The PCGs, 2 rRNAs, and 22 tRNAs were identified via comparison with the mitogenomic sequence of blackcap (*Sylvia atricapilla* (Singh et al., 2008)). Transfer RNAs were identified and secondary structures were predicted using the tRNAscan-SE Search Server version1.21 (Santa Cruz, CA) (Lowe & Eddy, 1997).

Data analysis

Two different datasets were used to infer the phylogenetic relationships among bushtits. The first dataset comprised 13 protein-coding genes and 2 ribosomal RNAs, and the second set consisted of the two duplicated CRs within an individual. These genes or regions were aligned separately with MEGA5.05 (Tamura et al., 2011) and combined using SequenceMatrix-1.7.8 (Vaidya et al., 2011). The maximum likelihood method implemented in PhyML3.0 (Guindon & Gascuel, 2003) and the Bayesian inference method in MrBayes v3.2 (Ronquist & Huelsenbeck, 2003) were used to reconstruct the phylogenetic relationships of *Aegithalos*. Modeltest3.7 (Posada & Crandall, 1998) and MrModeltest2.2 (Uppsala, Sweden) (Nylander, 2004) were employed to select the best model for maximum likelihood and Bayesian analyses, respectively.

For the analysis of duplicated CRs, the following two datasets were generated: (i) untrimmed alignment of bushtits' duplicated CRs, where both orthologous and paralogous CRs removed 5′ end of CR1 and 3′ end of CR2 which were longer than its paralogous copies, were aligned together; and (ii) trimmed alignment of bushtits' duplicated CRs, in which alignment of bushtits trimmed down both 5′ and 3′ ends that showed high variation among paralogous CRs. Unrooted trees for the three datasets were reconstructed by the Bayesian inference method implemented in MrBayes v3.2 (Ronquist & Huelsenbeck, 2003), and the best models were selected by MrModeltest2.2 (Nylander, 2004). To assess the sequence divergence in complete CRs, each domain of CRs, and all 13PCGs among bushtit were calculated by MEGA5.05 (Tamura et al., 2011) using maximum composite likelihood model.

Table 1. Data samples used in the study.

Species	Voucher-specimen	Collection locality	Sequence length	GenBank accession number
A. concinnus concinnus	IOZ1152	China, Gansu, Wen county	17,940 bp	KF951091
A. concinnus talifuensis	IOZ5344	China, Yunnan, Tengchong county	17,940 bp	KF951092
A. fuliginosus	IOZ2476	China, Shaanxi, Zhouzhi county	17,953 bp	KF951086
A. bonvaloti	IOZ3766	China, Sichuan, Yanbian county	17,953 bp	KF951087
A. caudatus vinaceus	IOZ2175	China, Shaanxi, Ansai county	17,938 bp	KF951090
A. caudatus caudatus	IOZ5879	China, Heilongjiang, Shangzhi county	17,935 bp	KF951088
A. caudatus glaucogularis	IOZ12564	China, Anhui, Jixi county	17,937 bp	KF951089

IOZ= Institute of Zoology, Chinese Academy of Sciences.

Detection of recombination in duplicated CRs

RDP 4.33 (Martin et al., 2010) was used to look for evidence of recombination between paralogous CRs. The seven taxa were divided into three groups (see phylogenetic results), and duplicated CRs in each group were analyzed individually. In the detection results, the major parent usually means a sequence closely related to that from which the greater part of the recombinant's sequence may have been derived, while the minor parent usually means a sequence closely related to that from which sequences in the proposed recombinant region may have been derived (Martin et al., 2010). And the minor parent is often located between the two breakpoints.

Results

The general features of *Aegithalos* mitogenome

The lengths of the 7 genomic sequences are from 17,935 bp (*A. caudatus caudatus*) to 17,953 bp (*A. bonvaloti* and *A. fuliginosus*). Each sequence comprises of 37 genes (including 13 protein-coding genes, 2 ribosomal RNA genes and 22 transfer RNA genes) and 2 non-coding regions, i.e. CRs, whose sequences are almost identical within individuals. For two duplicated CRs, one is located in the ancestral position for birds (Desjardins & Morais, 1990) (control region 2 or CR2), while the novel one is located between $tRNA^{Thr}$ and $tRNA^{Pro}$ (control region 1 or CR1). Annotations of each of these sequences are provided (Supplementary Appendix C).

In all 7 of the mitogenome sequences, the length, the start and stop codons are identical for each PCG. Moreover, the start codon of all PCGs are ATG with the exception of ND3 whose start codon is ATA (Supplementary Appendix C).

The secondary structures of the 22 tRNAs were predicted by tRNAscan-SE1.21 (Lowe & Eddy, 1997). All of the transfer RNAs formed the typical cloverleaf structure, including $tRNA^{Ser-AGY}$, which has lost the DHU arm in most metazoan mitochondrial genomes (Supplementary Appendix D).

Structure of duplicated CRs

Both of the duplicated CRs were composed of three domains based on the observed levels of sequence variation, namely, the hypervariable domain I, the conserved central domain II, and the variable domain III (Supplementary Appendix E). And among bushtits, duplicated CRs within an individual were found to exhibit almost identical sequences, except for both the 5′ end of domain I and 3′ end of domain III. Domain I of CR2 was about 410 bp in length, while that of CR1 was about 50 bp longer at the 5′ end, and both had a poly-C stretch and a potential TAS (termination associated sequence) element. Domain III of CR1

was about 260 bp in length, while that of CR2 was about 100 bp longer at the 3′ end. The variable sites among paralogous CRs were mainly concentrated between the two elements in domain I and in the last 20 bp of CR1 in domain III (Supplementary Appendix E). Domain II and the rest parts of domain I and domain III were almost identical among paralogous CRs within an individual. Several conserve blocks, such as F box, D box, C box, bird similarity box and conserved sequence block-1 (CSB-1) are found in both CRs (Supplementary Appendix E). This type of variation distribution among paralogous CRs within an individual have not been found previously in birds, while in previous researches, the high variable regions were found at the downstream of TAS element (Sammler et al., 2011), most of domain I (Abbott et al., 2005; Eberhard et al., 2001; Morris-Pocock et al., 2010), or even could not found in some birds due to no difference throughout the entire duplicated CRs (Cerasale et al., 2012; Cho et al., 2009; Singh et al., 2008). The high sequence similarity of the duplicated CRs may indicate that they evolved in concert.

Phylogeny of *Aegithalos* and duplicated CRs

The phylogenetic tree and node support deduced from the dataset comprising PCGs and rRNAs are shown in Figure 1, and the combined CR dataset produced the same tree topology, but the branch lengths are much shorter. Based on this tree topology, this genus could be divided into three lineages. The first consisted of *A. concinnus* (Group1), while the second corresponded to *A. caudatus* (Group2), and the third was formed by *A. bonvaloti* and *A. fuliginosus* (Group3). Furthermore, the second and third groups formed a sister group.

The sequence divergence among CRs showed that the substitution rate of unmatched regions among duplicated CRs (the 5′ end of CR1 and 3′ end of CR2) seemed to have a different manner compared with the rest part of CRs, and the three domains of CR had different substitution rates. In addition, the rate of duplicated CRs seemed to be slower than expected considering that of PCGs (Table 2).

For the phylogenetic analyses of duplicated CRs, the two independent analyses revealed conflicting observations (Figure 2). The results of the untrimmed alignment of the duplicated CRs showed that the orthologous CRs from the three lineages (see phylogenetic results) formed clade first, and each lineage subsequently clustered with paralogous ones. However, analysis of the trimmed bushtits' dataset suggested that paralogous ones from the same individual formed a clade first, then the seven individuals formed three lineages as the phylogenetic analysis (Figure 3).

To interpret the conflicting phylogenetic results of duplicated CRs, detection of recombination was used. Recombination could
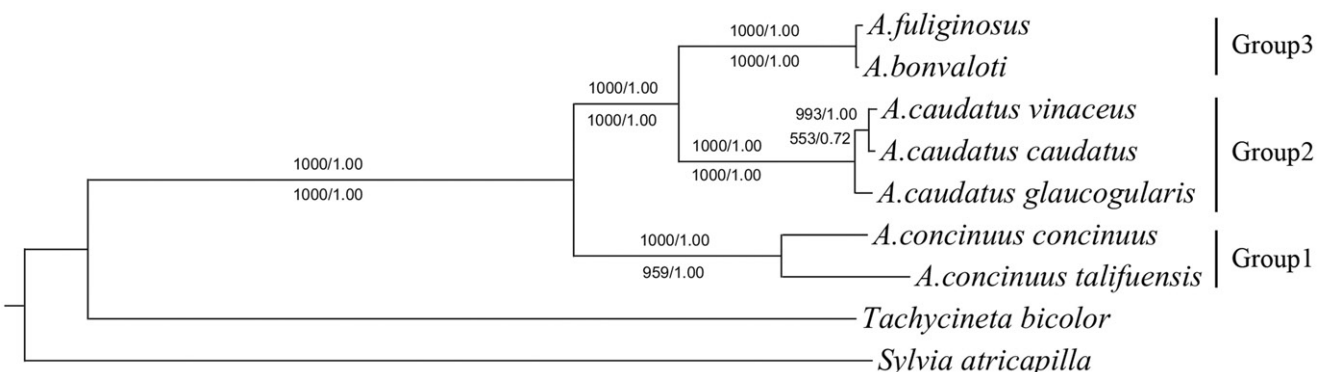

Figure 1. The Bayesian inference phylogenetic tree of genus *Aegithalos* inferred from combined PCGs and rRNA dataset. Node support is showed as bootstrap values from maximum likelihood analysis (former) and Bayesian posterior probabilities (latter). Node supports for combined PCGs and rRNA dataset are shown above the line, while that for combined control regions dataset are shown below the line.

Table 2. The sequence divergence in different component of mitogenome and duplicated CRs.

	Domain I				Domain III			
	Complete	Remove unmatched regions only	Remove unmatched region and high variable region	Domain II	Complete	Remove unmatched regions only	Remove unmatched region and high variable region	Average
cr1	0.1636	0.1414	0.1353	0.0244	0.0563	–	0.0541	0.0947
cr2	0.1488	–	0.1419	0.0233	0.1222	0.0579	0.0452	0.0938
Untrimmed alignment of all CRs (dataset i)	0.1434	–	0.1306	0.0403	0.0676	–	0.0638	0.087
PCG	–	–	–	–	–	–	–	0.0903

''–'' means the value were not calculated for certain sequences.

Figure 2. The unrooted tree of 14 control regions using Bayesian Inference method: (a) analysis with the high variation regions, (b) analysis without the high variation regions.
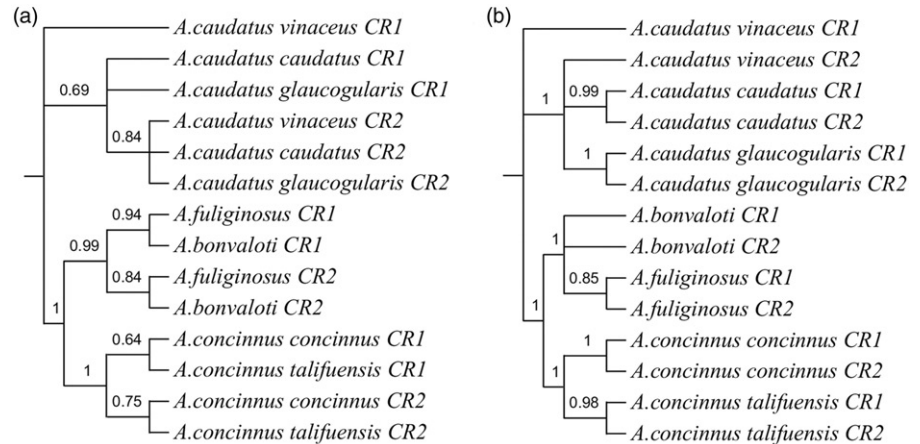
be detected in *A. concinnus* and *A. caudatus* (Table 3, Figure 3), but could not be observed between *A. fuliginosus* and *A. bonvaloti*, which may be due to the high sequence similarity in these two species. For *A. concinnus* and *A. caudatus*, recombination of paralogous CRs could be detected. However, the identity of the CR which was replaced through recombination remains unknown. The detection results showed that both CR1 and CR2 were recombinants from their paralogous and orthologous copies (Table 2), and the paralogous copies formed the proposed recombinant region (the minor parent, possibly including the 3′ end of domain I, the complete domain II, and the 5′ end of domain III) of the recombinant, whereas the 5′ and 3′ ends (the major parent, including the 5′ end of domain I, and the 3′ end of domain III) of the recombinant came from the orthologous copies. For the two breakpoint, one was often located between the poly-C stretch and the potential TAS element, while the other one was located adjacent to the 3′ end of CR1.

Discussion

The evolutionary pattern of duplicated CRs in *Aegithalos*

The present study revealed that two highly similar duplicated CRs in the mitogenomes of seven passerine species/subspecies evolved in concert. Molecular phylogenetic and dating analyses indicated that the divergence of *Aegithalos* occurred approximately 5 Mya (Päkert et al., 2010). Thus, this duplication event may have taken place before 5 Myr. However, more condensed sampling of both closely and distantly related species is needed to determine when this duplication event occurred and to examine whether all members of Sylvioidae exhibit a duplicated or remnant CR order, given that all of the available mitogenomes from this superfamily

exhibit the two aforementioned arrangement types (Bensch & Härlid, 2000; Cerasale et al., 2012; Singh et al., 2008).

In our study, the sequence divergence showed that different components of CRs had different substitution rate, while duplicated CRs seemed to have slower rate than expected compared to that of PCGs. Together with the phylogenetic relationships among paralogous CRs that CRs from the same lineages clustered first, and the topology among each lineage was dependent on whether the highly variable regions at both the 5′ and 3′ ends were removed, these results may indicated that (i) the 5′ end of CR1 and 3′ end of CR2 evolved in an independent way, while the rest part of CRs evolved in concert; (ii) the substitution rate of duplicated CRs may be affected and slowed down by concerted evolution; and (iii) mutations among paralogous CRs were mainly accumulated between the poly-C site and a potential TAS element in domain I within an individual, while they were accumulated in almost entire domain I among paralogous copies in different individuals. This evolution pattern among duplicated CRs seemed to be a little different from previously reported duplicated CRs in birds. Previous studies showed that such high variable regions among paralogous CRs within an individual were found at the downstream of TAS element (Sammler et al., 2011), most of domain I (Abbott et al., 2005; Eberhard et al., 2001; Morris-Pocock et al., 2010), or even no difference throughout the entire duplicated CRs (Cerasale et al., 2012; Cho et al., 2009; Singh et al., 2008). So, several patterns of concerted evolution may exist in duplicated CRs. To interpret this phenomenon, two explanations were deduced: (i) exact concerted evolution regions in duplicated CRs may not the same during each replication or in different individuals, (ii) the regions evolved in concerted may be influenced by the sequence or structure adjacent to them.
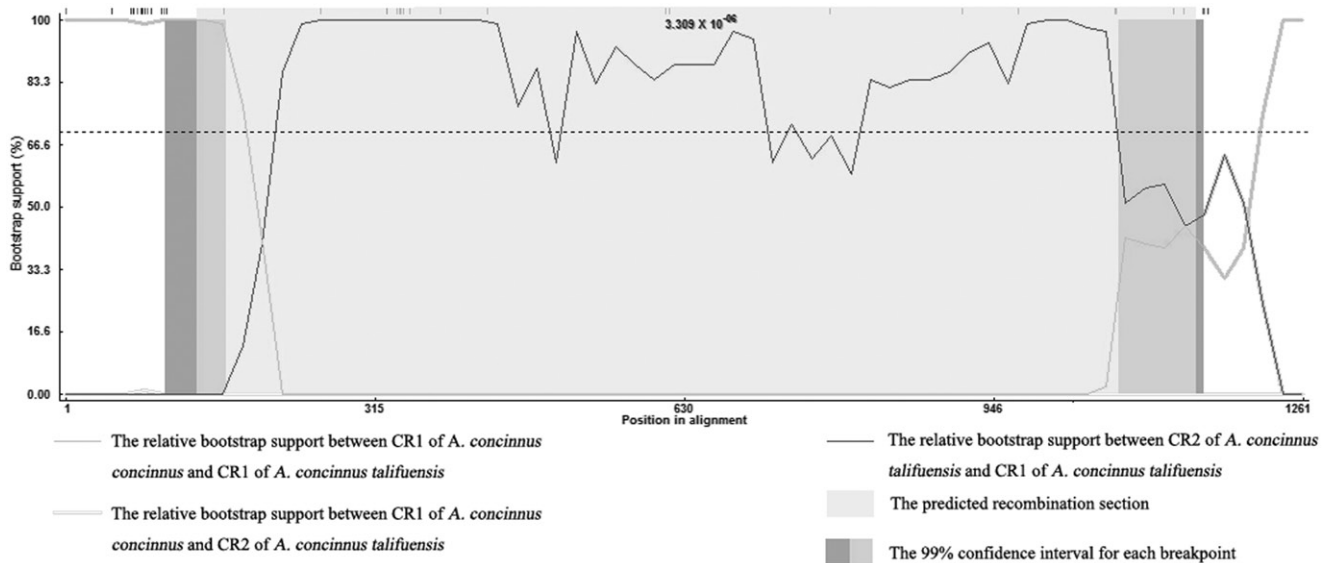
Figure 3. Recombination detection of one CR, CR1 of *A. concinnus talifuensis*, predicted by RDP4. The three lines with different gray levels indicate the pairwise relative bootstrap support between CR1 of *A. concinnus concinnus*, CR1 of *A. concinnus talifuensis* and CR2 of *A. concinnus talifuensis*. Three gray levels in the background correspond to the predicted recombination section and the 99% confidence interval for each breakpoint.

Table 3. The results of recombination detection.

	Major parent	Minor parent	Breakpoint begin	Breakpoint end
A. concinnus concinnus CR1	*A. concinnus talifuensis* CR1	*A. concinnus talifuensis* CR2	133	1150
A. concinnus concinnus CR2	*A. concinnus talifuensis* CR2	*A. concinnus talifuensis* CR1	133	1150
A. concinnus talifuensis CR1	*A. concinnus concinnus* CR1	*A. concinnus concinnus* CR2	133	1150
A. concinnus talifuensis CR2	*A. concinnus concinnus* CR2	*A. concinnus concinnus* CR1	133	1150
A. caudatus vinaceus CR1	*A. caudatus glaucogularis* CR1	*A. caudatus caudatus* CR2	103	752
A. caudatus vinaceus CR2	*A. caudatus caudatus* CR2	*A. caudatus vinaceus* CR1	354	1039
A. caudatus caudatus CR1	*A. caudatus glaucogularis* CR1	*A. caudatus caudatus* CR2	112	1059
A. caudatus caudatus CR2	–	–	–	–
A. caudatus glaucogularis CR1	*A. caudatus vinaceus* CR1	*A. caudatus caudatus* CR2	104	1059
A. caudatus glaucogularis CR2	*A. caudatus caudatus* CR2	*A. caudatus glaucogularis* CR1	103	960

Major parent usually means a sequence closely related to that from which the greater part of the recombinant's sequence may have been derived. Minor parent usually means a sequence closely related to that from which sequences in the proposed recombinant region may have been derived. ''–'' means recombination could not be detected steadily. Numbers of breakpoints indicate as sites in Supplementary Appendix E.

Concerted evolution in duplicated CRs: mechanism and complications

Gene duplication is common in rearranged mt genome of birds, and most of the duplication regions show high sequence similarity among the same molecule in most cases. The duplication of certain sections may take place through three possible mechanisms: dimerization, tandem duplication, and illegitimate recombination (Boore, 2000). For the circular mtDNA, several studies have presented evidence that tandem duplication (Abbott et al., 2005; Cho et al., 2009; Levinson & Gutman, 1987; Morris-Pocock et al., 2010; Sammler et al., 2011; Verkuil et al., 2010) and illegitimate recombination (Lunt & Hyman, 1997; Macey et al., 1997) may have played a role. According to Gibb et al. (2007), tandem duplication may be the mechanism responsible for this phenomenon in birds, and the ''duplicated tThr-CR'' type may have been derived from the avian ancestral type via tandem duplication, while the duplicated CR type originated from the ''duplicated tThr-CR'' type through the loss of duplicated genes in the flanking regions of duplicated CRs.

Most researchers try to connect the identical duplication regions with concerted evolution, but its exact mechanism is unclear. According to the existing hypotheses, we assume that the homologous recombination is a better explanation for duplicated CRs. And several research hold the same view (Hoarau et al., 2002; Kurabayashi et al., 2008; Morris-Pocock et al., 2010; Sammler et al., 2011). Together with the detection of recombination, phylogenetic analyses among paralogous CRs and the alignment of duplicated CRs, homologous recombination was occurred among paralogous CRs in different mtDNA molecules, and the conserved section (including the 3′ end of domain I, the entire domain II, and the 5′ end of domain III) of duplicated CRs was replaced by its paralogous copies, while the 5′ and 3′ end of the duplicated CRs, which encompass most of the variations between paralogous copies, were excluded from the recombinant section. Hence, we infer that inter-molecule homologous recombination between paralogous CRs is the main driving force for concerted evolution in duplicated CRs.

But for the ''duplicated tThr-CR'' type, some authors have noted that homologous recombination can not readily account for the high variations in both the 5′ and 3′ ends of CRs, because multiple breakpoints are needed to explain this phenomenon (Morris-Pocock et al., 2010). To our point of view, the mechanism

of concerted evolution related to the ''duplicated tThr-CR'' type may involve several different mechanisms. For duplicated CRs, homologous recombination is a feasible explanation, while for coding genes, parallel selection may make sense.

CRs are widely used as genetic markers in population genetics analyses, phylogeographic analyses, and phylogenetic analyses. However, for taxa with duplicated CRs, the use of CRs should be carefully examined because of concerted evolution. Due to concerted evolution, the conversion of duplicated CRs evolves at a faster rate than nucleotide substitution (Tatarenkov & Avise, 2007), or may even occurs in every replication cycle (Ogoh & Ohmiya, 2007; Sammler et al., 2011). Therefore, the substitution rate of duplicated CRs is slower than the expected rate deduced from non-duplicated CRs. In this case, the phylogenetic estimations would be severely affected, partially similar to the complications conferred on phylogenetic analyses by nuclear mitochondrial DNA (Numt) (Parr et al., 2006; Walther et al., 2011). The degree to which concerted evolution can affect phylogenetic and phylogeographic studies relies on the selected sections (Morris-Pocock et al., 2010) and the frequency of accumulating mutations in duplicated CRs (Ogoh & Ohmiya, 2007). Two sections of duplicated CRs have evolved in a disparate manner, with one evolving independently, while the other evolved in concert (Abbott et al., 2005; Cho et al., 2009; Morris-Pocock et al., 2010; Sammler et al., 2011; Verkuil et al., 2010), and the signatures they provide may be conflicting. Therefore, the independently evolved regions containing the 5′ end of domain I and the 3′ end of domain III in duplicated CRs appear to be more suitable markers for phylogenetic analyses compared with the concertedly evolved duplicated sections including domain II and the remaining parts of domain I and III. However, this section, which ranging from ∼50 bp to ∼350 bp (Abbott et al., 2005; Cho et al., 2009; Morris-Pocock et al., 2010; Sammler et al., 2011; Verkuil et al., 2010), appears to be too short to provide sufficient phylogenetic information. In addition, before using CR as a phylogenetic marker, the mtDNA arrangement should be examined for the presence of duplicates CRs to avoid the negative influence of the slow mutation rate observed in duplicated CRs.

Phylogeny of *Aegithalos*

After knowing characteristics of the whole mitogenome, especially the two duplicated CRs, our phylogenetic analyses of combined coding regions (PCGs and rRNAs) and combined non-coding regions (the two duplicated CRs) revealed that the genus *Aegithalos* could be divided into three distinct groups, as suggested by Päckert et al. (2010) and Dai et al. (2010). And some described unusual phenomenon (Dai et al., 2010, 2011; Päckert et al., 2010; Wang et al., 2014) were also found from the complete mitogenomes. But unexpected conflicting conclusions of the phylogenetic and morphological analyses for long-tailed tits (*A. caudatus*) were found. According to the phylogenetic analysis of mitochondrial genes, *A. c. vinaceus* and *A. c. caudatus* were sister taxa, while the morphological data suggested that the black-headed subspecies *A. c. vinaceus* and *A. c. glaucogularis* were sister taxa, whereas the white-headed subspecies *A. c. caudatus* was more distantly related. This may indicate that (1) the contact zones between the subspecies *A. c. vinaceus* and *A. c. caudatus* may be larger than expected (i.e. they may be not restricted to Beijing and northern Hebei during winter, as recorded in Li et al. (1982)), and gene flows may occur frequently; or that (2) *A. c. vinaceus* may represent the hybrid offspring of *A. c. glaucogularis* and *A. c. caudatus*. The two hypotheses need to be examined further with more detailed phylogeographic analyses.

Duplicated mitochondrial CRs in Passeriformes

As reported in several other passerine groups, the gene order of bushtits had two duplicated CRs. In birds, the rearrangement site was only observed at sections adjacent to the 5′-end of CR. In most cases, the rearranged gene order in birds presented as duplicated or remnant CRs. These types of arrangements have arisen independently several times in different orders (Gibb et al., 2007; Mindell et al., 1998) and even in a single order (Schirtzinger et al., 2012). In Passeriformes, this type of gene order was found in several groups (suboscine (Mindell et al., 1998; Slack et al., 2007); lyrebird (Slack et al., 2007); Australasian robins (Cooke et al., 2012); Sylvioidea (Bensch & Härlid, 2000; Cerasale et al., 2012; Singh et al., 2008)) and their relationships seemed unrelated (according to Jetz et al., 2012) which may indicate that this rearrangement type may originated more than once in this order. However, this assumption should be carefully further examined with more arrangement data and dense taxa sampling, like studies in Psittaciformes (Schirtzinger et al., 2012).

Acknowledgements

Declaration of interest

References

Abbott CL, Double MC, Trueman JW, Robinson A, Cockburn A. (2005). An unusual source of apparent mitochondrial heteroplasmy: Duplicate mitochondrial control regions in *Thalassarche* albatrosses. Mol Ecol 14:3605–13.

Arndt AL, Smith MJ. (1998). Mitochondrial gene rearrangement in the sea cucumber genus *Cucumaria*. Mol Biol Evol 15:1009–16.

Azevedo JLB, Hyman BC. (1993). Molecular characterization of lengthy mitochondrial DNA duplications from the parasitic nematode *Romanomermis culicivorax*. Genetics 133:933–42.

Bensch S, Härlid A. (2000). Mitochondrial genomic rearrangements in songbirds. Mol Biol Evol 17:107–13.

Boore JL. (1999). Animal mitochondrial genomes. Nucleic Acids Res 27: 1767–80.

Boore JL. (2000). The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. In: Sankoff D, Nadeau JH, editors. Comparative genomes: Empirical and analytical approaches to gene order dynamics, map alignment and the evolution of gene families. Dordrecht: Kluwer Academic Publishers. p 133–48.

Cerasale DJ, Dor R, Winkler DW, Lovette IJ. (2012). Phylogeny of the *Tachycineta* genus of new world swallows: Insights from complete mitochondrial genomes. Mol Phylogenet Evol 63:64–71.

Cho HJ, Eda M, Nishida S, Yasukochi Y, Chong JR, Koike H. (2009). Tandem duplication of mitochondrial DNA in the black-faced spoonbill, *Platalea minor*. Genes Genet Syst 84:297–305.

Cooke GM, King AG, Johnson RN, Boles WE, Major RE. (2012). Rapid characterization of mitochondrial genome rearrangements in

Australian songbirds using next-generation sequencing technology. J Hered 103:882–6.

Dai C, Chen K, Zhang R, Yang X, Yin Z, Tian H, Zhang Z, et al. (2010). Molecular phylogenetic analysis among species of Paridae, Remizidae and *Aegithalos* based on mtDNA sequences of COI and cyt b. Chinese Birds 1:112–23.

Dai C, Zhao N, Wang W, Lin C, Gao B, Yang X, Zhang Z, Lei F. (2011). Profound climatic effects on two east Asian black-throated tits (Ave: Aegithalidae), revealed by ecological niche models and phylogeographic analysis. PLoS One 6:e29329.

Desjardins P, Morais R. (1990). Sequence and gene organization of the chicken mitochondrial genome. J Mol Biol 212:599–634.

Eberhard JR, Wright TF, Bermingham E. (2001). Duplication and concerted evolution of the mitochondrial control region in the parrot genus *Amazona*. Mol Biol Evol 18:1330–42.

Gibb GC, Kardailsky O, Kimball RT, Braun EL, Penny D. (2007). Mitochondrial genomes and avian phylogeny: Complex characters and resolvability without explosive radiations. Mol Biol Evol 24: 269–80.

Guindon S, Gascuel O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696–704.

Hoarau G, Holla S, Lescasse R, Stam WT, Olsen J. (2002). Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish *Platichthys flesus*. Mol Biol Evol 19:2261–4.

Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. (2012). The global diversity of birds in space and time. Nature 491:444–8.

Kumazawa Y, Ota H, Nishida M, 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 13:1242–54.

Kurabayashi A, Sumida M, Yonekawa H, Glaw F, Vences M, Hasegawa M. (2008). Phylogeny, recombination, and mechanisms of stepwise mitochondrial genome reorganization in mantellid frogs from Madagascar. Mol Biol Evol 25:874–91.

Levinson G, Gutman GA. (1987). Slipped-strand mispairing: A major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–21.

Li G, Zheng B, Liu G. (1982). Aves Vol. 13 Passeriformes: Paridae - Zosteropidae. In: Chen Y, editor. Fauna Sinica. Beijing, China: Science Press. p 51–4.

Lowe TM, Eddy SR. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–64.

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

Macey JR, Larson A, Ananjeva NB, Fang Z, 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.

Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. (2010). RDP3: A flexible and fast computer program for analyzing recombination. Bioinformatics 26:2462–3.

Mindell DP, Sorenson MD, Dimcheff DE. (1998). Multiple independent origins of mitochondrial gene order in birds. Proc Natl Acad Sci USA 95:10693–7.

Morgan-Richards M, Trewick SA, Bartosch-Harlid A, Kardailsky O, Phillips MJ, McLenachan PA, Penny D. (2008). Bird evolution: Testing the Metaves clade with six new mitochondrial genomes. BMC Evol Biol 8:20.

Moritz C, Dowling TE, Brown WM. (1987). Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. Proc Natl Acad Sci USA 84:7183–7.

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

Nylander JAA. (2004). MrModeltest v2 [Computer software and manual]. Evolutionary Biology Centre. Uppsala: Uppsala University.

Ogoh K, Ohmiya Y. (2007). Concerted evolution of duplicated control regions within an ostracod mitochondrial genome. Mol Biol Evol 24: 74–8.

Päkert M, Martens J, Sun Y-H. (2010). Phylogeny of long-tailed tits and allies inferred from mitochondrial and nuclear markers (Aves: Passeriformes, Aegithalidae). Mol Phylogenet Evol 55:952–67.

Parr RL, Maki J, Reguly B, Dakubo GD, Aguirre A, Wittock R, Robinson K, et al. (2006). The pseudo-mitochondrial genome influences mistakes in heteroplasmy interpretation. BMC Genomics 7:185.

Posada D, Crandall KA. (1998). Modeltest: Testing the model of DNA substitution. Bioinformatics 14:817–18.

Pratt RC, Gibb GC, Morgan-Richards M, Phillips MJ, Hendy MD, Penny D. (2009). Toward resolving deep neoaves phylogeny: Data, signal enhancement, and priors. Mol Biol Evol 26:313–26.

Ronquist F, Huelsenbeck JP. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–4.

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

Schirtzinger EE, Tavares ES, Gonzales LA, Eberhard JR, Miyaki CY, Sanchez JJ, Hernandez A, et al. (2012). Multiple independent origins of mitochondrial control region duplications in the order Psittaciformes. Mol Phylogenet Evol 64:342–56.

Shao R, Barker SC, Mitani H, Aoki Y, Fukunaga M. (2005). Evolution of duplicate control regions in the mitochondrial genomes of metazoa: A case study with Australasian *Ixodes* ticks. Mol Biol Evol 22:620–9.

Singh TR, Shneor O, Huchon D. (2008). Bird mitochondrial gene order: Insight from 3 warbler mitochondrial genomes. Mol Biol Evol 25: 475–7.

Slack KE, Delsuc F, McLenachan PA, Arnason U, Penny D. (2007). Resolving the root of the avian mitogenomic tree by breaking up long branches. Mol Phylogenet Evol 42:1–13.

Sorenson MD, Ast JC, Dimcheff DE, Yuri T, Mindell DP. (1999). Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Mol Phylogenet Evol 12:105–14.

Staden R. (1996). The Staden Sequence Analysis Package. Mol Biotechnol 5:233–41.

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. Mol Biol Evol 28:2731–9.

Tatarenkov A, Avise JC. (2007). Rapid concerted evolution in animal mitochondrial DNA. Proc R Soc Lond B Biol Sci 274:1795–8.

Vaidya G, Lohman DJ, Meier R. (2011). SequenceMatrix: Concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 27:171–180.

Verkuil YI, Piersma T, Baker AJ. (2010). A novel mitochondrial gene order in shorebirds (Scolopacidae, Charadriiformes). Mol Phylogenet Evol 57:411–16.

Walther E, Schöfl G, Mrotzek G, Haryanti, Sugama K, Saluz HP. (2011). Paralogous mitochondrial control region in the giant tiger shrimp, *Penaeus monodon* (F.) affects population genetics inference: A cautionary tale. Mol Phylogenet Evol 58:404–8.

Wang W, Dai C, Alström P, Zhang C, Qu Y, Li S-H, Yang X, et al. (2014). Asymmetric introgression between two East Asian long-tailed tits (*Aegithalos*), revealed by multilocus analysis and coalescent simulations. Front Zool 11:40.

Yan D, Tang Y, Xue X, Wang M, Liu F, Fan J. (2012). The complete mitochondrial genome sequence of the western flower thrips *Frankliniella occidentalis* (Thysanoptera: Thripidae) contains triplicate putative control regions. Gene 506:117–24.

Yang C, Lei F, Huang Y. (2010). Sequencing and analysis of the complete mitochondrial genome of *Pseudopodoces humilis* (Aves, Paridae). Zool Res 31:333–44.

Zhou Z, Huang Y, Shi F. (2007). The mitochondrial genome of *Ruspolia dubia* (Orthoptera: Conocephalidae) contains a short A+T region of 70bp in length. Genome 50:855–66.

Supplementary material available online
Supplementary Appendix A–E