

SCIENTIFIC REPORTS



OPEN

Extreme variation in patterns of tandem repeats in mitochondrial control region of yellow-browed tits (*Sylviparus modestus*, Paridae)

Received: 18 February 2015

Accepted: 21 July 2015

Published: 19 August 2015

Xiaoyang Wang¹, Nian Liu¹, Hongli Zhang³, Xiao-Jun Yang⁴, Yuan Huang¹ & Fumin Lei^{1,2}

To investigate the evolutionary pattern and origins of tandem repeats in the mitochondrial control region of the yellow-browed tit (*Sylviparus modestus*), the control region and another four mitochondrial loci from fifteen individuals were analyzed. A 117-bp tandem repeat unit that repeated once, twice or three times in different individuals was found, and a rarely reported arrangement for this tandem repeats region that a 5' imperfect copy at its downstream and a 3' imperfect copy at its upstream was observed. The haplotype network, phylogenetic trees, and ancestral state reconstruction of the combined dataset of five loci suggested multiple origins of the same repeat number. The turnover model via slipped-strand mispairing was introduced to interpret the results, because mispairing occurred so frequently that multiple origins of certain repeat number were observed. Insertion via recombination should be a better explanation for the origin of this tandem repeat unit, considering characteristics of the combined sequence of the 3' and 5' imperfect copy, including identification of its homolog in other passerines and its predicted secondary structure.

The control region is the main non-coding region of the animal mitochondrial genome, and its length variation can affect the size of the mtDNA molecule¹. The length of the control region is highly variable due to the appearance or absence of the variable numbers of tandem repeats (VNTRs). Tandem repeats have been widely reported in the mitochondrial control region in animals¹, and the variation in the number of repeat units can be found among species^{2,3}, populations^{4,5} or even within an individual^{2,6,7}. The size and number of a tandem repeat unit differs greatly among different animals¹. Size of repeat units can range mainly from four⁸ to approximately 200 base pairs^{2,7,9} or even more than 400 base pairs in some insects^{10–12}, and the repeat number can range from two^{7,12} to more than one hundred¹³. The location of tandem repeats in control regions are mainly reported in the hypervariable domains in vertebrates^{3–5,7}.

In birds, the tandem repeats have been found in several groups, predominantly at the very end of the 3' end of the control region^{8,14–19}, and are represented as simple or short sequence repeats or microsatellite-like sequences. However, only a few tandem repeats with long motifs have been found in Domain I²⁰ and downstream of the Conserved Sequence Block I (CSB I box) in Domain III^{3,4,20}. At these two loci, length of the repeat unit has been observed longer, commonly approximately 80 bp^{3,20} or sometimes even up to 128 bp⁴.

¹Co-Innovation Center for Qinba regions' sustainable development, College of Life Sciences, Shaanxi Normal University, No. 199, South Chang'an Road, Xi'an 710062, China. ²Key Laboratory of the Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Chaoyang District, Beijing 100101, China. ³College of Life Science, Datong University, Xingyun Street, Datong 037009, China. ⁴State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, No. 32, Jiaochang East Road, Kunming 650223, China. Correspondence and requests for materials should be addressed to Y.H. (email: yuanh@snnu.edu.cn) or F.L. (email: leifm@ioz.ac.cn)

location	Voucher-specimen	Clade	repeated numbers	haplotype number for combined dataset	arrangement pattern
Tengchong	KIZgls1240	B	1	1	A
	KIZgls1241	A	1	2	A
	KIZgls1259	A	1	3	A
Luding	IOZ8837	A	2	4	2A
	IOZ8838	C	2	5	A + B
	IOZ8839	A	2	6	2A
	IOZ8840	A	2	7	2A
	IOZ8841	B	2	8	A + B
	IOZ8842	B	3	9	2A + B
Suiyang	IOZ15553	B	2	10	A + B
Pingwu	IOZ16728	B	1	11	A
Duijiangyan	IOZ18697	B	2	12	A + B
	IOZ19841	C	1	13	A
	IOZ19897	C	1	5	A
	IOZ19898	C	2	14	2A

Table 1. Detailed information about each individual in this study. Note: The haplotype number of each individual is identical with the network in Figure 1; IOZ, Institute of Zoology, Chinese Academy of Sciences; KIZ, Kunming Institute of Zoology, Chinese Academy of Sciences.

To interpret the origin and evolution of tandem repeats, mechanisms such as slipped-strand mispairing²¹, unequal crossing over¹³ or recombination²², and illegitimate elongation induced by the stem-loop structures in tRNAs or some non-replication origin sequences²³ have been proposed. Considering the likelihood and frequency of occurrence, slipped-strand mispairing is considered the most likely explanation for the variable repeat numbers in mtDNA. According to this mechanism, the nascent and parental H strand competitively binds to the parental L strand during replication. When a motif in the nascent H strand misaligned to its complementary strand due to the bending or complex folding of its upstream sequence^{3,9,24}, a repeated motif is formed in the following replication event. This mechanism can explain both the increase or decrease of repeat numbers and the variations in tandem repeat clusters where multiple types of repeat unit exist.

In our previous study on the complete mitochondrial genome of the yellow-browed tits (*Sylviparus modestus*)²⁵, we found a 117 bp tandem repeat which repeated twice in the Domain III of its mitochondrial control region. To determine whether the number of this tandem repeat region varies among or within different individuals of this species, we sequenced the control region of 14 more individuals. We then discussed the potential evolutionary dynamics and the possible origin of this tandem repeat region by mapping the number of repeat unit in different individuals onto the phylogenetic relationships based on a combined datasets containing control region sequence and 4 more mitochondrial loci.

Result

Phylogenetic relationships of yellow-browed tits. Five loci were amplified from all 15 surveyed individuals, and the combined dataset was 5323 bp in length (1143 bp for cyt b, 1041 bp for ND2, 684 bp for atp6, 1213 bp for COI, and 1242 bp for control region), while 55 sites were variable and 20 sites were parsimony informative. These polymorphic sites defined 14 haplotypes in all surveyed individuals, and only one haplotype was shared by two individuals (Table 1). The haplotype network of the combined dataset showed three main haplotype clusters (Cluster A, B, and C) (Fig. 1), while some haplotypes showed loosely connection, so these clusters appear to be formed by several haplotype groups. Cluster A consisted of three groups (labeled Group I-III), Cluster B was composed of two groups (labeled Group IV and V), while Cluster C contained only one group (labeled Group VI) (Fig. 1).

Both Bayesian inference and Maximum Parsimony trees produced the similar tree topology, and revealed that Cluster B and Cluster C were clustered together, while Cluster A was a paraphyletic group at the basal clades (Fig. 2). For most nodes, the posterior probabilities were high, whereas the parsimony bootstrap values were slightly lower. Both Cluster A and B could be divided into stable haplotype groups as the haplotype network. Cluster B was a monophyletic cluster consisting of Group IV and V, and Group II and III of Cluster A formed a monophyletic cluster sister to Cluster B and C, while the two haplotypes in Group I were paraphyletic at the basal clades.

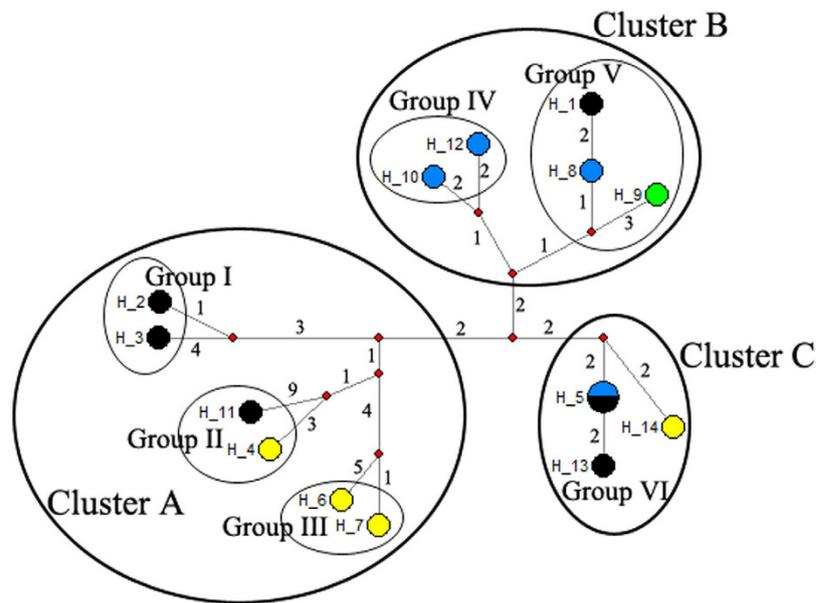


Figure 1. The haplotype network for the combined datasets. Colors stand for different repeat numbers: black for one, yellow (arrangement type 2A) and blue (arrangement type A + B) for two, and green for three. Numbers on the line indicate for the mutations between different haplotypes or nodes.

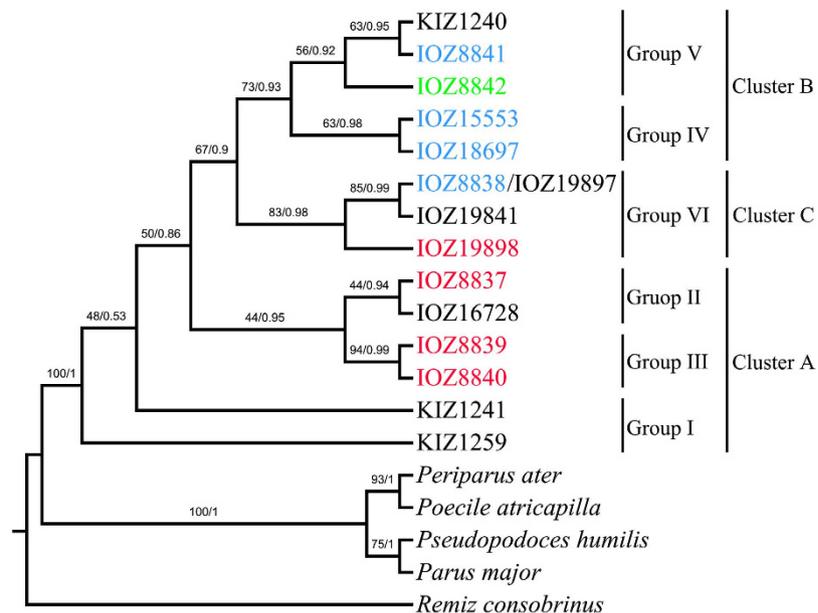


Figure 2. The phylogenetic tree for the combined dataset. The node supports are shown as bootstrap values from Maximum Parsimony analysis (former) and posterior probabilities for Bayesian inference analysis (latter). Colors for each individual stand for different repeat numbers of *S. modestus*: black for one, red for arrangement type 2A, blue for arrangement type A + B, and green for three.

Structure and content of the control region and the tandem repeat regions. The length of control regions of *S. modestus* varied from 1374bp to 1609bp, due to the number of tandem repeats and a ploy-C site at the very end of the 3' end. The control region contained several conserved blocks: a potential termination associated sequence (TAS) element in Domain I, F box, D box, C box and BSB box in Domain II, and CSB1 in Domain III (Fig. 3A).

The tandem repeat region was found downstream of the CSB1 box in Domain III, and the numbers of tandem repeats were repeated once, twice and three times, while no heteroplasmy was found within an individual (Table 1). In the only shared haplotype, the two individuals had different numbers of repeat unit, one had only one repeat unit, while the other one had two repeat units (Table 1).

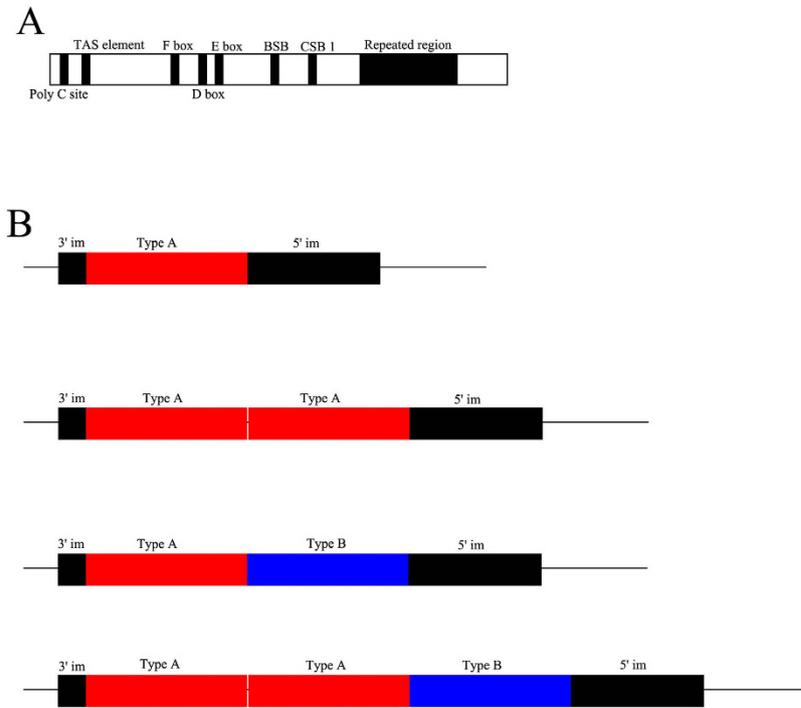


Figure 3. The structure for control region (A) and repeated region (B). Conserved blocks in control region marked with black and labeled at the top or bottom, and components of repeat region marked with different colors and labeled at the top.

Components	Sequence (5'-3')
TypeA	TTCATTTCGTTAAATATTTAACTATAAAATTCCTACCTTATTCAAACATTCATAATTCATTTCGTTTAAACAACAAACATTCCTCTATCTTCCCTATTTTTTATTATATCATTTCCTTA
TypeB C..... C.....
3' im copy CG. G..... T. . GT
5' im copy C..... C..... A..... T..... T.....
Combined sequence of CG. G..... T. . GT
3' and 5' im copy C..... C..... A..... T..... T.....

Note: Dash indicates gap; dot indicates the site is identical with *S. modestus*.

Table 2. Alignment of the two repeat units, 3' and 5' im copy of *Sylviparus modestus*. Note: Dash indicates gap; dot indicates the site is identical with *S. modestus*.

The repeat unit was 117 bp in length, and its nucleotide composition showed an unexpected high AT content (over 75%) (Table 2). At the downstream of the repeated regions, there was a 96-bp sequence that exhibited only 5 variable sites compared to the preceding 96 bp of the repeat unit (this region is hereafter termed the 5' imperfect copy, or 5' im copy for short, Table 2), while a 23-bp sequence at upstream of the repeated region had 6 variable sites compared to the last 23 bp of the repeat unit (hereafter termed the 3' imperfect copy, or 3' im copy for short, Table 2). This arrangement of the tandem repeats, with imperfect copies at both ends, has rarely been reported.

Among different individuals, two types of repeat unit with two variable sites were found (Table 2): one was found in all individuals in this study (type A), and another one was found in some individuals with two or more repeat units in which the first 22 bp at the 5' end was identical to the 5' im copy (type B). Four different patterns of arrangement in the repeated region were detected (Fig. 3B), three of which occurred in more than four individuals, with the exception of the pattern that repeated three times (Table 2). In all haplotype clusters and groups, all three clusters contained more than one repeat number or arrangement pattern, and three groups consisted of different repeat numbers (Fig. 1).

According to the known complete control regions from Paridae, no long repeated regions were found. However, investigation of these known parid control regions revealed a 120–121 bp motif similar to the combined 3' and 5' im copy (referred to the sequence showed in Table 2 hereafter) of yellow-browed tits (Table S2). The pairwise identity between the combined motif in *S. modestus* and its homologs in other

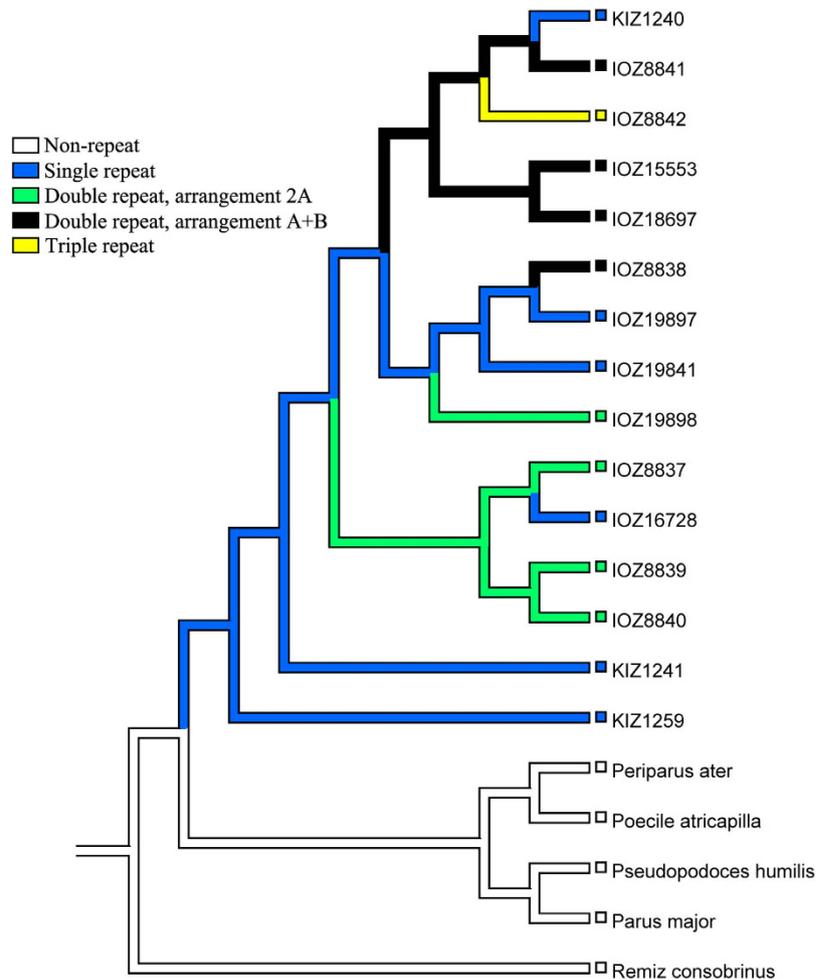


Figure 4. Ancestral state reconstruction of the numbers of repeat units in *S. modestus* on the phylogenetic tree under parsimony criterion. Squares stand for the observed repeat number for each individual, and lines depict the possible state between each nodes. Different colors are shown for different repeat numbers and arrangements: white for non-repeat, black for repeat once, green (arrangement type 2A) and blue (arrangement type A + B) for twice, yellow for repeat three times.

parids ranged from 63.9% to 72.1%. And in some other passerine lineages, the homologs of the combined 3' and 5' im copy was also observed with high similarities (Table S3).

The secondary structure of repeat regions and its adjacent sequences. Only a few short and simple stem-loop structures with low free energy (-1.1 to -3.2 kcal/M) were found in the predicted secondary structure of Type A, Type B, 3' im copy with its upstream sequence, and 5' im copy with its downstream sequence (Fig. S1 A–D), while no complex structures with multiple stem-loops were found in these sequences. For Type A, Type B and 5' im copy with its downstream sequence, two stem-loop structures were formed, while for 3' im copy with its upstream sequence, a stem-loop was formed by its 5' end and its upstream sequence. In the regions adjacent to tandem repeats, a stem-loop structure was found ~ 20 bp upstream of the 3' im copy, while another one was located ~ 20 bp downstream of the 5' im copy (Fig. S1 C,D). However, the combined 3' im and 5' im copy sometimes formed a long stem-loop structure with several internal bulges (Fig. S1 E).

Extreme variation in patterns of tandem repeats. To examine whether the four arrangements of repeat region in the control region of all surveyed individuals had a single origin, the arrangement states were mapped onto the phylogenetic tree using the parsimony method. Character-mapping analysis based on the parsimony method indicated that most of the arrangement patterns had multiple origins among all surveyed individuals (Fig. 4), revealed extreme variation in the patterns of tandem repeats in the mitochondrial control region of yellow-browed tits. The possible state for each node was shown in Table S4. This result suggests that reversions among different patterns may occur frequently and that a arrangement pattern may have multiple origins.

Discussion

Tandem duplication is one of the three possible mechanisms for the duplication of certain sections²⁶, and most duplication events in mtDNAs can be explained by slipped-strand mispairing^{16,17,19,21}. This scenario would be a better explanation why the three repeat arrangements containing two or three repeat units in yellow-browed tits. To explain how did the misalign of the nascent strand occur, some previous studies have suggested that slipped-strand mispairing might occur via the formation of stable and complex secondary structure within or among repeat units for both long and short repeat unit^{3,9,24}. And in several species with long repeat units, the number of repeat units can exceed twenty^{3,9}. In contrast, the tandem repeat unit in our study can only form two simple stem-loop structures and no inter-units structure (Fig. S1 A–C), with the repeat unit repeating three or fewer. This phenomenon may indicate that the number of repeat units could be affected by whether the complex secondary structures were formed by repeat units, and stable and complex structures would facilitate the occurrence of slipped-strand mispairing and more repeat numbers. For the tandem repeats in yellow-browed tits, the low occurrence of slipped-strand mispairing, providing few candidates for new repeat unit types, could explain the low numbers of repeat unit types and arrangements observed.

When discussing at the individual or population level, previous studies assumed that numbers of tandem repeat unit could be constant in certain phylogenetic lineages or geographical populations^{4–6}. However, we found two conflicting results from the phylogenetic and ancestral state reconstruction results, i) different repeat numbers were observed in close related affinities, whereas the same number was shared by individuals from different phylogenetic lineages, suggesting that the same repeat number may have multiple origins; and ii) the number of repeat units were different between the two individuals from the only shared haplotype. These contradictory results among phylogenetic lineages and the repeated numbers in each surveyed individual indicated that i) the same repeat numbers in different individuals may have multiple origins as suggested by Broughton and Dowling (1997)²⁷, with the same repeat number from different lineages experiencing a homogenization process; ii) although both nucleotide sequence and numbers of tandem repeats are accumulated and fixed by mtDNA replication, their evolution appear to be two inconsistent, independent processes; and iii) individuals in the only shared haplotype had different repeat numbers, suggesting that the accumulation of repeat units may be faster than that of nucleotide mutations. As the slipped-strand mispairing can occur in every replication, the change of repeated numbers can be fixed and inherited rapidly and readily, yielding different repeat numbers within a phylogenetic lineage or even within an individual (heteroplasmy).

Slipped-strand mispairing may explain the origin of most tandem repeats^{20,21}, but it can not readily explain the origin of the tandem repeat region in the mitochondrial control region of *S. modestus*, because it can not overcome several shortcomings described below. Though the origin of both 5' and 3' im copy can be explained by slipped-strand mispairing²⁸, this process can not readily account for the variable sites between the observed repeat unit (Type A) and the combined 3' and 5' im copy in yellow-browed tits, particularly the most variable 3' im copy. Considering the alignment and the high similarities between it and its homologs in other passerines (Tables S2,S3), the combined 3' and 5' im copy should reflect the original sequence before its duplication. If the ancestral repeat unit was that combined sequence, the observed repeat unit (Type A in Table 2), which would have arisen after the duplication of the ancestral repeat unit and been embedded within the two ancestral repeat units, should have about 10 variable sites (comprised almost 10% of the observed repeat unit) compared with that combined sequence (the ancestral repeat unit) before next duplication. If so, as described above, some intermediate repeat units between the observed repeat unit and the combined 3' and 5' im copy should exist, but none were found. More importantly, the combined 3' and 5' im copy, together with its adjacent sequences at both ends, sometimes formed a long stem-loop structure (Fig. S1 E), causing the combined sequence to be fully embedded in the structure. This structure could also possibly prevent the duplication of the observed repeat unit via slipped-strand mispairing and resulting in a longer repeat unit than the observed repeat unit if mispairing could occur via this structure.

Compared to slipped-strand mispairing, the recombination is possibly a better explanation for the tandem repeats in yellow-browed tits. Though recombination has little effect on mitochondrial genetic variability in birds²⁹, both experimental and circumstantial evidence suggest it occurs pervasively in animal mtDNA^{30–34}.

Recombination can underlie the origin of the repeated regions in yellow-browed tits, and have two possible processes. One is direct inter-molecular insertion induced by the mini-circles³³ excised from one mtDNA molecule, and the second is a series of processes involving illegitimate elongation via the stem-loop structures in some tRNAs or some non-replication origin²³, formation of mini circles and sequence insertions (or strand break and rejoin) via mini circles. The two processes described above require the components possessing both homologous and non-homologous recombination in mitochondria, and these components have been reported in mitochondria in several previous studies^{30,35,36}. Therefore, the recombination processes occurring in mtDNA have a higher probability of occurrence.

The former one is a simple process and has been reported in some long fragment insertions^{34,37,38}, whereas some studies have shown that a region adjacent to the control region is more preferentially involved in recombination induced by double-strand breaks recombination³⁹. The inter-molecular recombination process via mini-circles can explain the origin of this tandem repeat sequence. And rarity

of inter-molecular recombination³⁹ could account for why very few tandem repeats with origins interpretable by recombination were found.

The latter one is a three-step process which can also explain the origin of the tandem repeat in yellow-browed tits clearly, but direct evidence for the whole process is lacking. This process may occur as follows. First, the short fragments induced by illegitimate elongation could be dropped from the parental strand and remain free in mitochondrial matrix. Then, the free strand exposed to the reactive oxygen for a long time, causing it to exhibit a high mutation rate¹⁸ or even break down into shorter fragments, and these short fragments could be cyclized as single- or double-strand mini-circles³¹ (replication may be occurred via certain stem-loop structure in these short fragments^{40,41}). The mini circles can stably exist in the mitochondrial matrix^{31,40,41}. Once the mini-circle achieved the same length and similar sequence as the observed repeat unit was cracked and linearized, the newly formed linear fragment can be inserted into the combined 3' and 5' im copy via recombination^{31,34,37,38}. Thus the observed repeated unit was formed.

The yellow-browed tit was first discovered species with long tandem repeats in control regions within tits. This species was a relict, basal, and early divergent lineage in Paridae^{42–44}, and consist of three subspecies that are discontinuously distributed in mid- or high-mountain areas of the Himalayas, southwestern China and northern Indo-China⁴². We only sampled the *S. m. modestus* in southwestern China, but the preliminary results provided starting points for a future study focusing on the variation pattern of the tandem repeats of yellow-browed tits. More samples from *S. m. modestus*, the rest two subspecies and other tits could possibly test and confirm the proposed hypothesis, and more clearly make the evolution scenario and the estimated origin time of the tandem repeat of this species and other tits.

Method

Sample collection, DNA extraction, PCR amplification, sequencing, and sequence annotation. Tissue samples from 15 individuals of yellow-browed tits were collected from six locations in the southwestern mountain areas of China (Table 1), and were preserved in pure ethanol and stored at -80°C before use. The relevant specimens were kept in Institute of Zoology and Kunming Institute of Zoology and their usage were in accordance with guidelines of above institutes. All experimental protocols used in this study were approved by the Experimental Animal Ethics Committee of Institute of Zoology. Total DNA was isolated from the tissue samples using the phenol-chloroform extraction method.

To investigate the phylogenetic relationships and numbers of tandem repeat unit among *S. modestus*, we selected five mitochondrial loci, i.e., *cyt b*, *ND2*, *atp6*, *COI*, and the control region. The primers used to amplify these loci were obtained from Sorensen (2003)⁴⁵ and Li *et al.* (2008)⁴⁶ with some modifications (Table S1). PCRs were performed under the cycling condition as follow: 35 cycles of 93°C for 30s, 56°C for 30s, and 72°C for 2min (for coding genes) or 60°C for 4 min (for control region). The PCR products were purified and sequenced directly, and some internal primers were designed to overlap some possible gaps between poorly sequenced sections. The complete sequences were assembled and uploaded to GenBank.

Sequences were annotated through comparison to the complete mitochondrial genome of the yellow-browed tit²⁵, and the locations of each gene or region were determined. The conserved blocks in avian control regions were determined following previous studies^{17,47}. The tandem repeats in the control region were identified using Tandem Repeat Finder v4.07b.

Sequence analyses. The sequences of the five loci (the repeated region was removed from control region) from all the 15 individuals of yellow-browed tits were aligned individually using Clustal X1.83, and the five loci were combined together by SequenceMatrix 1.7.8. The median joining networks of the combined dataset was produced by Network 4.612.

To reconstruct the phylogenetic relationships of yellow browed tits, sequences of the five loci from the 15 sampled individuals and another five tits (*Parus major* (KP137624), *Periparus ater* (KM588075), *Pseudopodoces humilis* (KP001174), *Poecile atricapilla* (KJ909190), and *Remiz consobrinus* (KC463856)) were aligned individually and combined together by SequenceMatrix 1.7.8. The phylogenetic relationships among all the surveyed individuals were rebuilt by the Maximum Parsimony method and Bayesian Inference methods implemented in PAUP*4.0 and MrBayes3.2 using the combined dataset, and the best models for Bayesian Inference method were selected by MrModeltest2.2. The parsimony method was applied to search for the shortest tree for the combined dataset using 1000 random heuristic bootstrap replicates with tree-bisection–reconnection (TBR) branch-swapping. The Bayesian inference analysis was conducted by two runs of 1 million generations, and sampled every 100 generations with a burn-in of the first 35% generations. The secondary structure of the repeat unit and its adjacent regions were predicted by MFOLD.

Ancestral state reconstruction. Using the repeat numbers and the arrangement patterns of repeat region determined by Tandem Repeat Finder, the presence of non-repeat (0) or different repeat arrangements (1, 2, 3, 4 present for the four arrangements, repeat unit repeated once, arrangement 2A, arrangement A + B, and repeat unit repeated three times, respectively) was coded into a matrix and mapped

onto the phylogenetic trees deduced from different inference methods. Ancestral state reconstructions were implemented in Mesquite 3.01 under the parsimony criterion.

References

- Lunt, D. H., Whipple, L. E. & Hyman, B. C. Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. *Mol. Ecol.* **7**, 1441–1455 (1998).
- Casane, D., Dennebouy, N., de Rochambeau, H., Mounolou, J. C. & Monnerot, M. Nonneutral evolution of tandem repeats in the mitochondrial DNA control region of Lagomorphs. *Mol. Biol. Evol.* **14**, 779–789 (1997).
- Omote, K., Nishida, C., Dick, M. H. & Masuda, R. Limited phylogenetic distribution of a long tandem-repeat cluster in the mitochondrial control region in *Bubo* (Aves, Strigidae) and cluster variation in Blakiston's Fish Owl (*Bubo blakistoni*). *Mol. Phylogenet. Evol.* **66**, 889–897 (2013).
- Mundy, N. L., Winchell, C. S. & Woodruff, D. S. Tandem repeats and heteroplasmy in the mitochondrial DNA control region of the Loggerhead Shrike (*Lanius ludovicianus*). *J. Heredity*. **87**, 21–26 (1996).
- Padhi, A. Geographic variation within a tandemly repeated mitochondrial DNA D-Loop region of a North American freshwater fish, *Pylodictis olivaris*. *Gene* **538**, 63–68 (2013).
- Hernández, M. A., Campos, F., Gutiérrez-Corchero, F. & Amezcua, A. Identification of *Lanius* species and subspecies using tandem repeats in the mitochondrial DNA control region. *Ibis* **146**, 227–230 (2004).
- Wang, T. Y., Tzeng, C. S., Teng, H. Y. & Chang, T. Phylogeography and identification of a 187-bp-long duplication within the mitochondrial control region of *Formosania lacustre* (Teleostei: Balitoridae). *Zool. Stud.* **46**, 569–582 (2007).
- Wenink, P. W., Baker, A. J. & Tilanus, M. G. J. Mitochondrial control-region sequences in two shorebird species, the turnstone and the dunlin, and their utility in population genetic studies. *Mol. Biol. Evol.* **11**, 22–31 (1994).
- Bentzen, P., Wright, J. M., Bryden, L. T., Sargent, M. & Zwanenburg, K. C. T. Tandem repeat polymorphism and heteroplasmy in the mitochondrial control region of redfishes (Sebastes: Scorpaenidae). *J. Heredity*. **89**, 1–7 (1998).
- Monteforte, A., Barrio, E. & Latorre, A. Characterization of the length polymorphism in the A + T-rich region of the *Drosophila obscura* group species. *J. Mol. Evol.* **36**, 214–223 (1993).
- Lewis, D. L., Farr, C. L., Farquhar, A. L. & Kaguni, L. S. Sequence, organization, and evolution of the A + T-rich region of *Drosophila melanogaster* mitochondrial DNA. *Mol. Biol. Evol.* **11**, 523–538 (1994).
- Zhang, D. X., Szymura, J. M. & Hewitt, G. M. Evolution of structural conservation of the control region of insect mitochondrial DNA. *J. Mol. Evol.* **40**, 382–391 (1995).
- Hoelzel, A. R. Evolution by DNA turnover in the control region of vertebrate mitochondrial DNA. *Curr. Opin. Genet. Dev.* **3**, 891–895 (1993).
- Berg, T., Mourn, T. & Johansen, S. Variable numbers of simple tandem repeats make birds of the Order Ciconiiformes heteroplasmic in their mitochondrial genomes. *Curr. Genet.* **27**, 257–262 (1995).
- Abbott, C. L., Double, M. C., Trueman, J. W. H., Robinson, A. & Cockburn, A. An unusual source of apparent mitochondrial heteroplasmy: duplicate mitochondrial control regions in *Thalassarche* Albatrosses. *Mol. Ecol.* **14**, 3605–3613 (2005).
- Slack, K. E., Delsuc, F., Mclenachan, P. A., Arnason, U. & Penny, D. Resolving the root of the avian mitogenomic tree by breaking up long branches. *Mol. Phylogenet. Evol.* **42**, 1–13(2007).
- Cho, H. J. *et al.* Tandem duplication of mitochondrial DNA in the black-faced spoonbill, *Platalea minor*. *Genes. Genet. Syst.* **84**, 297–305 (2009).
- Sammler, S., Bleidorn, C. & Tiedemann, R. Full mitochondrial genome sequences of two endemic Philippine hornbill species (Aves: Bucerotidae) provide evidence for pervasive mitochondrial DNA recombination. *BMC Genomics* **12**, 35 (2011).
- Cerasale, D. J., Dor, R., Winkler, D. W. & Lovette, I. J. Phylogeny of the *Tachycineta* genus of new world swallows: insights from complete mitochondrial genomes. *Mol. Phylogenet. Evol.* **63**, 64–71 (2012).
- Haring, E., Kruckenhauser, L., Gamauf, A., Riesing, M. J., & Pinsker, W. The complete sequence of the mitochondrial genome of *Buteo buteo* (Aves, Accipitridae) indicates an early split in the phylogeny of raptors. *Mol. Biol. Evol.* **18**, 1892–1904 (2001).
- Levinson, G. & Gutman, G. A. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**, 203–221(1987).
- Pâques, F., Leung, W.-Y. & Haber, J. E. Expansions and contractions in a tandem repeat induced by double-strand break repair. *Mol. Cell. Biol.* **18**, 2045–2054 (1998).
- Broughton, R. E. & Dowling, T. E. Length Variation in Mitochondrial DNA of the Minnow *Cyprinezza Spiloptera*. *Genetics* **138**, 179–190 (1994).
- Buroker, N. E. *et al.* Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. *Genetics* **144**, 157–163 (1990).
- Wang, X. & Huang Y. Complete mitochondrial genome of yellow-browed tit (*Sylviparus modestus*). *Mitochondrial DNA* doi: 10.3109/19401736.2015.1018220 (2015).
- Boore, J. L. in *Computational Biology Series*, Vol. **1** (eds Sankoff, D. & Nadeau, J.) Ch. 13, 133–147 (Kluwer Academic Publishers, 2000).
- Broughton, R. E. & Dowling, T. E. Evolutionary dynamics of tandem repeats in the mitochondrial DNA control region of the minnow *Cyprinella spiloptera*. *Mol. Biol. Evol.* **14**, 1187–1196 (1997).
- Taylor, J. S. & Breden, F. Slipped-strand mispairing at noncontiguous repeats in *Poecilia reticulata*: a model for minisatellite birth. *Genetics* **155**, 1313–1320 (2000).
- Berlin, S., Smith, N. G. C. & Ellegren, H. Do avian mitochondria recombine? *J. Mol. Evol.* **28**, 163–167 (2004).
- Thyagarajan, B., Padua, R. A. & Campbell, C. Mammalian mitochondria possess homologous DNA recombination activity. *J. Biol. Chem.* **271**, 27536–27543 (1996).
- Lunt, D. H. & Hyman, B. C. Animal Mitochondrial DNA Recombination. *Nature* **387**, 247 (1997).
- Hoarau, G., Holla, S., Lescasse, R., Stam, W. T. & Olsen, J. L. Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish *Platichthys flesus*. *Mol. Biol. Evol.* **19**, 2261–2264 (2002).
- Sato, A. *et al.* Rare creation of recombinant mtDNA haplotypes in mammalian tissues. *Proc Natl Acad Sci USA* **102**, 6057–6062 (2005).
- Kurabayashi, A. *et al.* Phylogeny, recombination, and mechanisms of stepwise mitochondrial genome reorganization in Mantellid frogs from Madagascar. *Mol. Biol. Evol.* **25**, 874–891 (2008).
- Lakshmiathy, U. & Campbell, C. Double strand break rejoining by mammalian mitochondrial extracts. *Nucleic Acids Res.* **27**, 1198–1204 (1999).
- Lakshmiathy, U. & Campbell, C. The Human DNA Ligase III gene encodes nuclear and mitochondrial proteins. *Mol. Cell. Biol.* **19**, 3869–3876 (1999).
- Kajander, O. A. *et al.* Human mtDNA sublimons resemble rearranged mitochondrial genomes found in pathological states. *Hum. Mol. Genet.* **9**, 2821–2835 (2000).

38. Shao, R. *et al.* Novel mitochondrial gene content and gene arrangement indicate illegitimate inter-mtdna recombination in the chigger mite, *Leptotrombidium pallidum*. *J. Mol. Evol.* **60**, 764–773 (2005).
39. Bacman, S. R., Williams, S. L. & Moraes, C. T. Intra- and inter-molecular recombination of mitochondrial DNA after *in vivo* induction of multiple double-strand breaks. *Nucl. Acids Res.* **13**, 4218–4226 (2009).
40. Shao, R. *et al.* Evolution of extensively fragmented mitochondrial genomes in the Lice of Humans. *Genome Biol. Evol.* **4**, 1088–1101 (2012).
41. Jiang, H., Barker, S. C. & Shao, R. Substantial variation in the extent of mitochondrial genome fragmentation among Blood-Sucking Lice of Mammals. *Genome Biol. Evol.* **5**, 1298–1308 (2013).
42. Gosler, A. & Clement, P. in *Handbook of the Birds of the World*, Vol. **12**. (Eds del Hoyo, J., Elliot, A. & Christie, D.) Ch. 14, 662–744 (Lynx Edicions, 2007).
43. Tietze, D. T. & Borthakur, U. Historical biogeography of tits (Aves: Paridae, Remizidae). *Org. Divers. Evol.* **12**, 433–444 (2012).
44. Johansson, U. S. *et al.* A complete multilocus species phylogeny of the tits and chickadees (Aves: Paridae). *Mol. Phylogenet. Evol.* **69**, 852–860 (2013).
45. Sorenson, M. D. *et al.* Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Mol. Phylogenet. Evol.* **12**, 105–114 (1999).
46. Li, S. H. *et al.* Sailing through the Late Pleistocene: unusual historical demography of an East Asian endemic, the Chinese Hwamei (*Leucodioptron canorum canorum*), during the last glacial period. *Mol. Ecol.* **18**, 622–633 (2009).
47. Randi, E. & Lucchini, V. Organization and evolution of the mitochondrial DNA control region in the avian genus *Alectoris*. *J. Mol. Evol.* **47**, 449–462 (1998).

Acknowledgements

We are grateful to Zuohua Yin, Dezhi Zhang who have assisted with field specimen collections. We own many thanks to Fei Ye for his valuable comments and suggestions. This research was supported by National Natural Science Foundation of China (No. 31201711), Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDB13020300), Shaanxi Normal University (No. 2013CXB008), A grant from China MOST (No. 2014FY210200), State Key Program of NSFC (No. 31330073) and the Key Laboratory of the Zoological Systematics and Evolution of the Chinese Academy of Sciences (No. O529YX5105).

Author Contributions

X.W. designed and carried out the study, accomplished the laboratory work and statistical analyses, and wrote the manuscript. N.L. and H.Z. drafted the manuscript. X.Y. provided the samples from Yunnan for this study. F.L. and H.Y. conceived and elaborated the design of the whole study, including sample collection, paper wrote and revision of the manuscript.

Additional Information

Accession codes: Sequences have been deposited in NCBI GenBank under Accession Numbers KT327932-KT328001.

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Wang, X. *et al.* Extreme variation in patterns of tandem repeats in mitochondrial control region of yellow-browed tits (*Sylviparus modestus*, Paridae). *Sci. Rep.* **5**, 13227; doi: 10.1038/srep13227 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>