Data Article

Data on taxonomic status and phylogenetic relationship of tits

Xue-Juan Li a, Li-Liang Lin a, Ai-Ming Cui a, Jie Bai a, Xiao-Yang Wang a,b, Chao Xin a, Zhen Zhang a, Chao Yang c, Rui-Rui Gao a, Yuan Huang a,* Fu-Min Leib

a Co-Innovation Center for Qinba regions’ sustainable development, College of Life Sciences, Shaanxi Normal University, Xi’an 710062, China
b Key Laboratory of the Zoological Systematics and Evolution, Institute of Zoology, The Chinese Academy of Sciences, Beijing 100101, China
c Shaanxi Institute of Zoology, Xi’an 710032, China

A R T I C L E   I N F O

Article history:
Received 21 July 2016
Received in revised form
17 November 2016
Accepted 23 November 2016
Available online 28 November 2016

A B S T R A C T

The data in this paper are related to the research article entitled “Taxonomic status and phylogenetic relationship of tits based on mitogenomes and nuclear segments” (X.J. Li et al., 2016) [1]. The mitochondrial genomes and nuclear segments of tits were sequenced to analyze mitochondrial characteristics and phylogeny. In the data, the analyzed results are presented. The data holds the resulting files of mitochondrial characteristics, heterogeneity, best schemes, and trees.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Specifications Table

| Subject area | Biology, Genetics and Genomics |
|--------------|--------------------------------|
| More specific subject area | Phylogenetics and Phylogenomics |
| Type of data | Figures, Tables, Trees |

DOI of original article: http://dx.doi.org/10.1016/j.ympev.2016.07.022
* Corresponding authors.
E-mail addresses: yuanh@snnu.edu.cn (Y. Huang), leifm@ioz.ac.cn (F.-M. Lei).
How data was acquired

The analyses of A+T contents, conserved site percentages and P-distances, were obtained in MEGA 4.1[2]. Sequences were aligned in Muscle[3]. The best schemes were analyzed in Partitionfinder v1.1.1[4]. The heterogeneity was inferred with AliGROOVE[5]. The gene trees based on six datasets (one mitochondrial dataset and five nuclear segments) were constructed in RAxML 7.0.3[6]. A species tree was obtained with employing these gene trees in ASTRAL[7].

Data format

Experimental factors

The RY-coding method was employed for the third sites of protein-coding genes, while nuclear dataset was divided into different parts (exons and introns).

Experimental features

The phylogeny employed the best schemes inferred by PartitionFinder v1.1.1[4]. A species tree was obtained by employing gene trees in ASTRAL[7].

Data source location

Shaanxi Normal University

Data accessibility

Data is with this article

Value of the data

- The provided files of comparative mitochondrial characteristics of tits can be valuable to further summarize.
- The files of phylogenetic relationships would help to further study the phylogeny of tits and even Passeriformes.
- The provided `.tree` files can be directly used to compare with other results.

1. Data

In the data, Figs. 1 and 2 show base compositions and conserved site percentages of tits, respectively. Fig. 3 is the result of heterogeneity. Fig. 4 shows gene trees and a species tree. Table 1 describes the taxonomic samples. Table 2 lists the primer sequences. Table 3 is the P-distance based on mitochondrial dataset. Table 4 shows the best schemes.

2. Experimental design, materials and methods

This study sampled 13 individuals of tits by using *Sylviparus modestus* and *Remiz consobrinus* as outgroups. Each gene was aligned in Muscle[3] independently. The mitochondrial characteristics, including A+T contents, conserved site percentages and P-distances, were analyzed by using MEGA 4.1[2], and the results can be found in Figs. 1 and 2 and Table 3, respectively.

Four datasets, A: the first and second sites of protein-coding genes, B: protein-coding genes with the third sites not employing RY-coding method, C: 37 mitochondrial genes with the third sites of protein-coding genes not using RY-coding method plus one control region, D: five nuclear segments, were used to analyze the heterogeneity in AliGROOVE[5], and the results can be found in Fig. 3. The best schemes were analyzed by using Partitionfinder v1.1.1[4], and the results were in Table 4. The gene trees in Fig. 4 were constructed by using RAxML 7.0.3[6], employing 1000 replications, and these results were used to construct a species tree by using ASTRAL[7].
Fig. 1. Nucleotide compositions of different mitochondrial partitions in 10 tits species. Note: AT-skew ([A−T]/[A+T]), GC-skew ([G−C]/[G+C]), PCG-1st (the first codon positions of protein-coding genes), PCG-2nd (the second codon positions of protein-coding genes), PCG-3rd (the third codon positions of protein-coding genes), tRNA-H (the tRNA genes on H-strand), tRNA-L (the tRNA genes on L-strand).

Fig. 2. Conserved site percentages of mitochondrial genes among 10 tits species.
Fig. 3. The heterogeneity analyzed by AliGROOVE. Note: The heterogeneity continuously decreased from $-1$ (red coloring) to $+1$ (blue coloring). A: the first and second codon positions of protein-coding genes, B: protein-coding genes with the third codon positions not using RY-coding method, C: mitochondrial genome with the third codon positions not using RY-coding method, D: nuclear segments dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Fig. 4. The gene trees and species tree analyzed by ASTRAL. Note: The gene trees (A–F) were constructed based on maximum likelihood method. A: MOS; B: FGB; C: ALDOB; D: PCBD1; E: CALB1; F: mitochondrial genome; G: species tree.
| Family     | Genus   | Species and subspecies            | Sample locality/source                                  | GenBank accession Nos. |
|------------|---------|-----------------------------------|-------------------------------------------------------|------------------------|
|            |         |                                   | Mitogenome    | MOS      | FGB      | ALDOB    | PCBD1    | CALB1    |
| Paridae    | Parus   | Parus major                       | Beach forestry centre, Zhouqu County, Gansu Province   | KX388473              | KX388398 | KX388413 | KX388428 | KX388443 | KX388458 |
|            |         | Parus major                       | Baihuling, Gaoligongshan, Yunnan Province              | KX388480              | KX388405 | KX388420 | KX388435 | KX388450 | KX388465 |
|            |         | Parus monticolus                  | Beach forestry centre, Zhouqu County, Gansu Province   | KX388474              | KX388399 | KX388414 | KX388429 | KX388444 | KX388459 |
|            |         | Parus monticolus                  | Dahaoping, Gaoligongshan, Yunnan Province              | KX388481              | KX388406 | KX388421 | KX388436 | KX388451 | KX388466 |
|            |         | Poecile                           |                                                     |                        |          |          |          |          |          |
|            |         | Parus montanus                    | Liaocheng, Yongdeng County, Gansu Province             | KX388478              | KX388403 | KX388418 | KX388433 | KX388448 | KX388463 |
|            |         | Parus montanus                    | Maoershan, Shangzhi City, Heilongjiang Province        | KX388479              | KX388404 | KX388419 | KX388434 | KX388449 | KX388464 |
|            |         | Parus palustris                    | Beach forestry centre, Zhouqu County, Gansu Province   | KX388475              | KX388400 | KX388415 | KX388430 | KX388445 | KX388460 |
|             | Cyanistes | Parus cyanus                     | Kizil, Baicheng County, Xinjiang                      | KX388472              | KX388397 | KX388412 | KX388427 | KX388442 | KX388457 |
|             | Machlolophus | Parus spilonotus          | Longqishan Nature Reserve, Fujian Province             | KX388476              | KX388401 | KX388416 | KX388431 | KX388446 | KX388461 |
|             | Lophophanes | Parus dichrous              | Sanguanmiao, Shaanxi Province                          | KX388477              | KX388402 | KX388417 | KX388432 | KX388447 | KX388462 |
|             | Periparus | Parus ater                        | Wen County, Gansu Province                             | NC_026223             | KX388408 | KX388423 | KX388438 | KX388453 | KX388468 |
|             | Pardaliparus | Parus venustulius            | Yangxin County, Huangshi City, Hubei Province          | NC_026701             | KX388410 | KX388425 | KX388440 | KX388455 | KX388470 |
|             | Pseudopodoces | Pseudopodoces humilis         | Bird Island, Qinghai Lake, Qinghai Province           | KP001174              | KX388407 | KX388422 | KX388437 | KX388452 | KX388467 |
|             | Sylviparus | Sylviparus modestus             | Luding County, Sichuan Province                        | NC_026793             | KX388409 | KX388424 | KX388439 | KX388454 | KX388469 |
| Remizidae  | Remiz   | Remiz consobrinus                | Xinxing Town, Panjin City, Liaoning Province          | NC_021641             | KX388411 | KX388426 | KX388441 | KX388456 | KX388471 |
Table 2
The primers used in this study.

| Name    | Sequence(5′–3′) | Name    | Sequence(5′–3′) |
|---------|----------------|---------|----------------|
| L1263b<sup>a</sup> | AAAGCATRRCACTGAA | H10343b | TGGGCTCATGTGACKTRACKCC |
| H1859b  | TCGATTAGAGAACGCCTCCTCA | L10236b | TCTCTGCACTTTCCACCATAG |
| L1754b  | TGGGATTAGATACCCCACTATG | H10884b<sup>a</sup> | GGGTCAAWRCRCATCTTCCATAG |
| H2294b  | TTTCAGGTGTAAGCTGAATGCTT | L10635b<sup>a</sup> | CACACTTYGAAGCTGACGAC |
| L2260ba | CAAGGTAAGTACCCCACTGATT | H11837b | ARGGTRGCCTRCAATGGRATRAAAA |
| H2891ba | TTTCAGGTGTAAGCTGAATGCTT | L10234b<sup>a</sup> | CTTCTGAGCMTCCTAGCGAC |
| L3218b  | CGAAGTCAGCACCAGCATTACC | H11047b | ATCTTACCTTTGACCATACCAA |
| H3784b  | CGGCTCTGAGATACCCCACTATG | L13040b<sup>a</sup> | ATCCATGCTTCAAGGAC |
| L3722b  | CGGCTCTGAGATACCCCACTATG | H13563b<sup>a</sup> | TCGAGGCHDCRGRRTTGC |
| H4170b  | CCAACCTAGATGACAGAGGAGGAC | L13525b | GMYGAGACATGATGATCATCTAC |
| L3803b  | CTAGATACCCCACTGAGGAACG | H14127b | CCAATTTCTGACRTCTT |
| H4644b  | CTAGATACCCCACTGAGGAACG | L14080b | TCAATCGACCCATTCCTAAAG |
| L4500b  | GTAGGCTACCAATTTACCTGCAARG | H15049b | GTCTGCTCTGCTAGTYATG |
| H5201b  | CGGCTCTGAGATACCCCACTATG | L14770b | TMGGCACCAAGAYT |
| L5143b  | CGGCTCTGAGATACCCCACTATG | H15295b | CCTCAGAATTGATATTGCTATCAG |
| H5766b  | CGGCTCTGAGATACCCCACTATG | L14996b | AACATCTAAGTACATGTAAGAACTTYGG |
| L5758b  | CGGCTCTGAGATACCCCACTATG | H15646b | GGYGGTCAATTTCTGCTTC |
| H6681b<sup>a</sup> | CGGCTCTGAGATACCCCACTATG | L15413b | GYGGATYTTGAGTACAAAC |
| L6615b<sup>a</sup> | CGGCTCTGAGATACCCCACTATG | H16064b<sup>a</sup> | CTCCAAATTTCTGACRTCTT |
| H7122b  | CGGCTCTGAGATACCCCACTATG | L15725b<sup>a</sup> | AAACCHGAGATGAGTACATG |
| H7036b  | CGGCTCTGAGATACCCCACTATG | H1530b<sup>a</sup> | GGTCTGCTGCCACAGTCAT |
| H7458b  | CGGCTCTGAGATACCCCACTATG | L14127b | CCAATTTCTGACRTCTT |
| L7525b  | CGGCTCTGAGATACCCCACTATG | L14770b | TMGGCACCAAGAYT |
| H8121b  | CGGCTCTGAGATACCCCACTATG | L14996b | AACATCTAAGTACATGTAAGAACTTYGG |
| L7987b  | CGGCTCTGAGATACCCCACTATG | H15646b | GGYGGTCAATTTCTGCTTC |
| H8628b  | CGGCTCTGAGATACCCCACTATG | L15413b | GYGGATYTTGAGTACAAAC |
| L8386b  | CGGCTCTGAGATACCCCACTATG | H16064b<sup>a</sup> | CTCCAAATTTCTGACRTCTT |
| H9235b  | CGGCTCTGAGATACCCCACTATG | H15725b<sup>a</sup> | AAACCHGAGATGAGTACATG |
| L8929b  | CGGCTCTGAGATACCCCACTATG | L14127b | CCAATTTCTGACRTCTT |
| H9726b  | CGGCTCTGAGATACCCCACTATG | L14770b | TMGGCACCAAGAYT |
| L9700b  | CGGCTCTGAGATACCCCACTATG | H15646b | GGYGGTCAATTTCTGCTTC |

<sup>a</sup> Means the primers used in LA-PCR.
### Table 3
The P-distance based on mitogenome dataset.

| Species            | Genus     |
|--------------------|-----------|
| Parus cyanus       | Cyanistes |
| Parus major        | Parus     |
| Parus monticolus   | Poecile   |
| Parus palustris    | Machlolaphus |
| Parus spilonotus   | Lophophanes |
| Parus dichrous     | Pseudopodoces |
| Parus montanus     | Periparus |
| Parus montanus baicalensis | Pardaliparus |
| Parus major        | 0.085     |
| Parus monticolus   | 0.084 0.051 |
| Parus palustris    | 0.092 0.087 0.088 |
| Parus spilonotus   | 0.090 0.084 0.083 0.094 |
| Parus dichrous     | 0.096 0.089 0.089 0.080 0.096 |
| Parus montanus     | 0.091 0.088 0.088 0.038 0.093 0.080 0.090 0.088 0.088 |
| Parus montanus baicalensis | 0.086 0.021 0.052 0.087 0.084 0.090 0.088 0.088 |
| Parus major        | 0.085 0.052 0.010 0.088 0.083 0.089 0.088 0.088 0.053 |
| Parus monticolus   | 0.096 0.091 0.090 0.097 0.094 0.101 0.100 0.099 0.091 0.091 |
| Pseudopodoces humilis | 0.093 0.090 0.089 0.080 0.094 0.082 0.078 0.079 0.090 0.089 0.102 |
| Parus ater         | 0.092 0.089 0.086 0.077 0.092 0.080 0.077 0.076 0.088 0.086 0.099 0.071 |
| Parus venustulus   | 0.092 0.089 0.086 0.077 0.092 0.080 0.077 0.076 0.088 0.086 0.099 0.071 |
Table 4
Best schemes analyzed by PartitionFinder.

| Dataset          | Subset | Subset partitions                                                                 | Optimal model |
|------------------|--------|-----------------------------------------------------------------------------------|---------------|
| Protein-coding   | P1     | atp6_pos1, nad1_pos1, nad2_pos1, nad3_pos1, nad4L_pos1, nad4_pos1, nad5_pos1     | GTR + I + G   |
|                  | P2     | atp6_pos2, atp8_pos2, cox3_pos2, cox2_pos2, cox1_pos2, cytb_pos2, nad1_pos2,     | GTR + I + G   |
|                  |        | nad2_pos2, nad3_pos2, nad4L_pos2, nad4_pos2, nad5_pos2                          |               |
|                  | P3     | atp6_pos3, atp8_pos1, atp8_pos3, cox3_pos3, cox2_pos3, cox1_pos3, cytb_pos3,     | GTR + G       |
|                  |        | nad1_pos3, nad2_pos3, nad3_pos3, nad4L_pos3, nad4_pos3                          |               |
|                  | P4     | cox3_pos1, cox2_pos1, cox1_pos1, cytb_pos1                                     | GTR + I + G   |
|                  | P5     | nad5_pos3, nad6_pos3                                                            | GTR + G       |
|                  | P6     | nad6_pos1, nad6_pos2                                                            | GTR + G       |
| Mitogenomes      | P1     | rrnS, rrnL, atp6_pos1, nad1_pos1, nad2_pos1, nad3_pos1, nad4L_pos1, nad4_pos1, | GTR + I + G   |
|                  |        | nad5_pos1, trnR, trnD, trnG, trnH, trnL, trnM, trnF, trnS(agy), trnW, trnV        |               |
|                  | P2     | atp6_pos2, atp8_pos2, cox3_pos2, cox2_pos2, cox1_pos2, cytb_pos2, nad1_pos2,     | GTR + I + G   |
|                  |        | nad2_pos2, nad3_pos2, nad4L_pos2, nad4_pos2, nad5_pos2                          |               |
|                  | P3     | atp6_pos3, atp8_pos3, cox3_pos3, cox2_pos3, cox1_pos3, cytb_pos3, D_loop, nad1_- | GTR + I + G   |
|                  |        | pos3, nad2_pos3, nad3_pos3, nad4L_pos3, nad4_pos3                              |               |
|                  | P4     | cox3_pos1, cox2_pos1, cox1_pos1, cytb_pos1, trnN, trnL(uur), trnL(cun), trnS(ucn), | GTR + I + G   |
|                  |        | trnT, trnY                                                                      |               |
|                  | P5     | atp8_pos1, nad5_pos3, nad6_pos3                                                 | GTR + I + G   |
|                  | P6     | nad6_pos1, nad6_pos2, trnA, trnC, trnQ, trnE, trnP                               | GTR + I + G   |

Nuclear segments

| P1 | ALDOB_exon, CALB1_exon, MOS_exon, PCBD1_exon, PCBD1_intron, FGB_exon | GTR + I + G |
|    |                                                                      | GTR + I + G |
| P2 | ALDOB_intron, CALB1_intron, FGB_intron                               | GTR + G     |

Note: Pos1, pos2, and pos3 indicate the first, second and third codon positions of protein-coding genes in mitogenomes, respectively.

Transparency document. Supporting information

Transparency data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2016.11.079.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2016.11.079.

References

[1] X.J. Li et al., Taxonomic status and phylogenetic relationship of tits based on mitogenomes and nuclear segments, Mol. Phyloinet. Evol. (2016) 14–20.
[2] K. Tamura et al., MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0, Mol. Biol. Evol. 24 (8) (2007) 1596–1599.
[3] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797.
[4] R. Lanfear et al., Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses, Mol. Biol. Evol. 29 (2012) 1695–1701.
[5] P. Kück et al., AliGROOVE – visualization of heterogeneous sequence divergence within multiple sequence alignments and detection of inflated branch support, BMC Bioinform. 15 (2014) 294.
[6] A. Stamatakis, RAxML-VI-HPC: maximumlikelihood-based phylogenetic analyses with thousands of taxa and mixed models, Bioinformatics 22 (2006) 2688–2690.
[7] S. Mirarab et al., ASTRAL: genome-scale coalescent-based species tree estimation, Bioinformatics 30 (17) (2014) i541–i548.