



Short Communication

Organization and Variation of the Mitochondrial DNA Control Region in Ardeidae (Aves: Ciconiiformes) and their Phylogenetic Relationship

Zuhao Huang*, Meifang Li, Rujue Ruan and Jiawei Qin

School of Life Sciences, Jinggangshan University, Ji'an, Jiangxi Province 343009, China

ABSTRACT

Control region is a non-coding region of mitochondrial genome. To infer the organization and variation of Ardeidae mitochondrial DNA control region, the control region genes of 20 species were analyzed. All the Ardeidae species had duplicate control regions, except four species having only single region. The control region spans the region between the genes for tRNA^{Glu} - tRNA^{Phe} in the most Ardeidae species. The length of the control region sequences ranged from 1427 bp (*Ixobrychus flavicollis*) to 3871 bp (*Botaurus stellaris*). The average genetic distances among the species varied from 5.39% (between *Egretta sacra* and *Egretta garzetta*) to 41.15% (between *Egretta sacra* and *Botaurus stellaris*). Five conserved sequence boxes (F to B) in the domain II of Ardeidae sequences were localized. Maximum likelihood method was used to reconstruct the phylogenetic trees based on HKY+ G model. The phylogenetic tree grouped members of every genus were grouped into each clade, except the genus *Gorsachius*. Our results indicated that the genus *Gorsachius* might not be monophyletic.

Article Information

Received 18 January 2017

Revised 24 March 2017

Accepted 05 April 2017

Available online 25 August 2017

Authors' Contributions

ZH designed the study and wrote the article. ML, RR and JQ collected and analyzed the data.

Key words

Ardeidae, Mitochondrial DNA, Control region, Variation, Phylogeny.

The mitochondrial DNA (mtDNA) control region, the major noncoding segment of animal mitochondrial genome, is regarded as the most variable part of the vertebrate mtDNA genome, presumably because of the lack of coding constraints (Marshall and Baker, 1997). Control region has been widely used as a marker for phylogenetic relationship (Huang and Ke, 2016a, b) and phylogeographic structure studies (e.g., Donne-Goussé et al., 2002), because of fast evolutionary rate.

Control region can be divided into three domains among avian (Randi and Lucchini, 1998; Donne-Goussé et al., 2002; Ruokonen and Kvist, 2002). The structure and variation of avian control region have been long contended. But few Ardeidae species control region have been studied. The avian family Ardeidae, generally known as the herons, were a group of large wading birds. Phylogenetic relationship within the herons has been interpreted differently. In the present study, we examined the structure of the control region of Ardeidae species based on the complete mitochondrial genome collected from GenBank. The aims of this paper are: (1) to characterize

the organization and variation of the Ardeidae mtDNA control region, (2) to infer the phylogenetic relationships of Ardeidae using mtDNA control region.

Materials and methods

All the sequences were retrieved from the GenBank (species and accession numbers in Supplementary Table S1). We only analyzed the control region sequence from the whole mitochondrial genome in order to ensure the accuracy. A total of 20 species from 12 genera belonging to family of Ardeidae was analyzed (Supplementary Table S1).

Sequences were aligned by the CLUSTAL X procedure (Thompson et al., 1997). DnaSP v5.0 (Librado and Rozas, 2009) was used to define the variable sites. Only the sequences the genes for tRNA^{Glu} - tRNA^{Phe} was used to analyze, once the species had duplicate control regions. Nucleotide composition was calculated using MEGA7.0 (Kumar et al., 2016). Genetic distance between species was calculated using TN93 (Tamura and Nei, 1993) model in MEGA7.0 (Kumar et al., 2016). The conserved sequence boxes found were compared with the previously published avian boxes (e.g., Randi and Lucchini, 1998; Donne-Goussé et al., 2002; Ruokonen and Kvist, 2002).

Modeltest 3.0 (Posada and Crandall, 1998) and the

* Corresponding author: hzhow@163.com

0030-9923/2017/0005-1917 \$ 9.00/0

Copyright 2017 Zoological Society of Pakistan

Akaike information criterion (AIC), (Posada and Buckley, 2004) were used to identify the appropriate nucleotide substitution models. Maximum likelihood tree (ML) (Strimmer and Haeseler, 1996) was obtained using heuristic searches, based on the substitution model proposed by Modeltest 3.0 (Posada and Crandall, 1998). We calculated ML tree using MEGA7.0 (Kumar *et al.*, 2016). Statistical support for the internodes in phylogenetic tree was tested by bootstrap percentages (BP) with 1,000 replicates (Felsenstein, 1985).

Results and discussion

Alignments

The alignment of control region of Ardeidae was straightforward. Most the Ardeidae species had duplicate control regions, except four species (*Egretta eulophotes*, *Egretta novaehollandiae*, *Ixobrychus cinnamomeus*, *Ixobrychus flavicollis*) having only single control regions. The single control region spans the region between the genes for tRNA^{Glu} - tRNA^{Phe} of the four Ardeidae species (Supplementary Table S1), which was inconsistent with most of the avian species between tRNA^{Glu} and tRNA^{Phe} (e.g., Huang and Ke, 2016a).

The length of the control region sequences are high variable, ranging from 1427 bp (*Ixobrychus flavicollis*) to 3871 bp (*Botaurus stellaris*), with an average size of 2737 bp. The control region size of Ardeidae and its variation is larger than that of avian Phasianidae family (ranged

from 1144bp to 1555 bp; Huang and Ke, 2016a). Control region are usually considered to be the most variable parts of mtDNA (Randi and Lucchini, 1998). Extensive size variation of the mtDNA control region, attributable to variation in number of tandem repeats, has been reported for many animals (e.g., Boyce *et al.*, 1989).

Base composition and genetic distance

The average nucleotide composition of Ardeidae control region sequences was: 32.70%A, 30.16%T, 14.54%G, and 22.60% C, with a bias against G. The amount of A+T was more than that of G+C among whole control region, which is same as avian control region (e.g., Marshall and Baker, 1997; Ruokonen and Kvist, 2002; Huang and Ke, 2016a, b). The lack of guanines is evident in each domain, especially in domain III. Thymines and adenines are most prevalent in the first domain, thymines and cytosines in the second, and adenines and cytosines in the third (Supplementary Table S2).

Nucleotide frequencies were not significantly different among species, and thus the TN93 model (Tamura and Nei, 1993) is an appropriate estimator of genetic distance (Randi and Lucchini, 1998). Ardeidae control region sequences were alignable with certainty within genus. Genetic distances between species ranged from 5.39% (between *Egretta sacra* and *Egretta garzetta*) to 41.15% (between *Egretta sacra* and *Botaurus stellaris*), showing a wide range of divergences. The average genetic

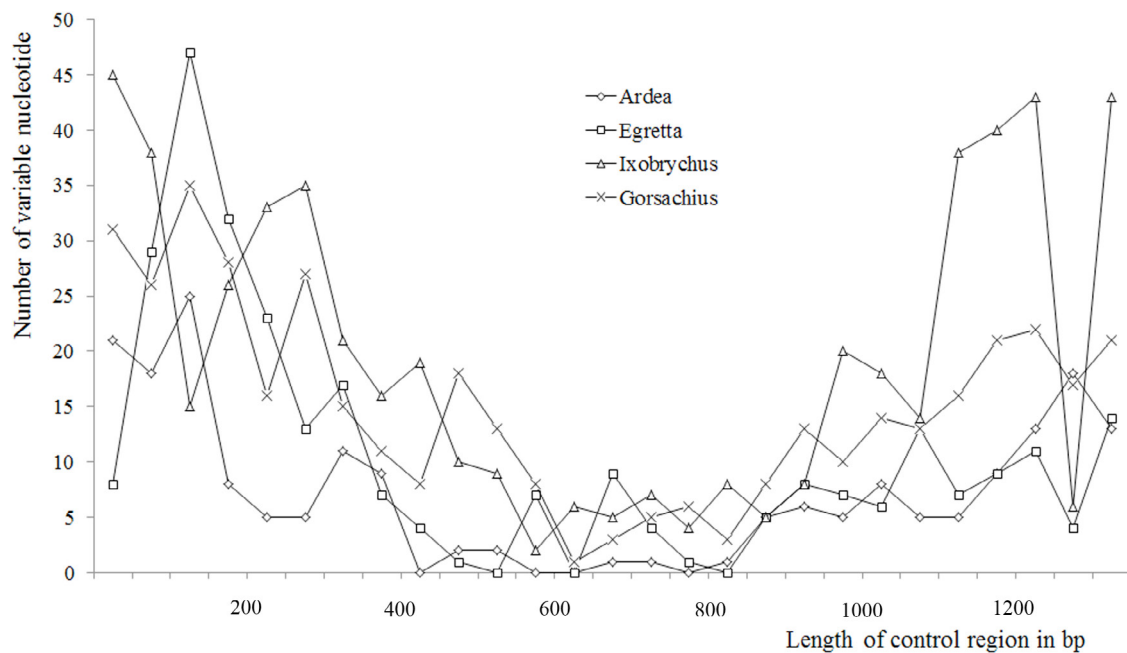


Fig. 1. Distribution of the variable sites in the control region. The number of variable sites within genera has been plotted in 50-bp intervals.

distances among the species within the genera varied from 12.64% (*Egretta*) to 25.26% (*Ixobrychus*) (Supplementary Table S3). The average genetic distances showed insignificantly negative correlation with ts/tv ($r=-0.853$, $p>0.05$).

Distribution of variable sites in the control region

Control region are usually considered to be the most variable parts of mtDNA (Randi and Lucchini, 1998). However, position mutability is not distributed randomly across the whole region, but affects particular hypervariable positions and domains (Yang, 1996). The distribution of the variation in the control region seems to be the same in all the species within genera. Nucleotide substitutions occur more frequently in peripheral domains. Average substitution rate for the three domains were 0.49: 0.12: 0.39, corresponding to relative proportions of 5:1:4, respectively. Among all the genera of Ardeidae, domain I is the most variable of the three domains (Fig. 1). Marshall and Baker (1997) believed mutational hotspots might occur in domain I, whereas more variation was expected in domain III for deeper divergences. Ruokonen and Kvist (2002) found that variation of three genera was greatest in domain III. Our results partly support the opinion of Marshall and Baker (1997).

Conserved sequence

Previous comparisons of control region sequences have identified conserved sequence elements based on greater similarity of the sequence elements compared to that of the flanking areas (e.g., Ruokonen and Kvist, 2002; Huang and Ke, 2016a). We aligned the sequences of every species. Conserved sequence blocks (CSB-1) were located in the Ardeidae (Supplementary Table S4). However, we did not find the CBS-2, 3 in Ardeidae, which were detected in fishes (e.g., Zhang *et al.*, 2011, 2013), avian (e.g., Marshall and Baker, 1997), and mammalian species (e.g., Walberg and Clayton, 1981). Five conserved sequence boxes in the domain II of Ardeidae sequences were localized (Supplementary Table S4) and identified as boxes F, E, D, C and B. Recently, six central conserved sequence boxes (F to A) were detected in fishes (Zhang *et al.*, 2011). We did not find A-box in Ardeidae control region.

In F-box, 19 of 28 nucleotide positions were fully conserved among the Anseriformes sequences. While, there were five nucleotide positions were variable in E-box, four in D-box, seven in C-box and eight in B-box (Supplementary Table S4).

Phylogenetic relationships

On the basis of hierarchical likelihood-ratio tests

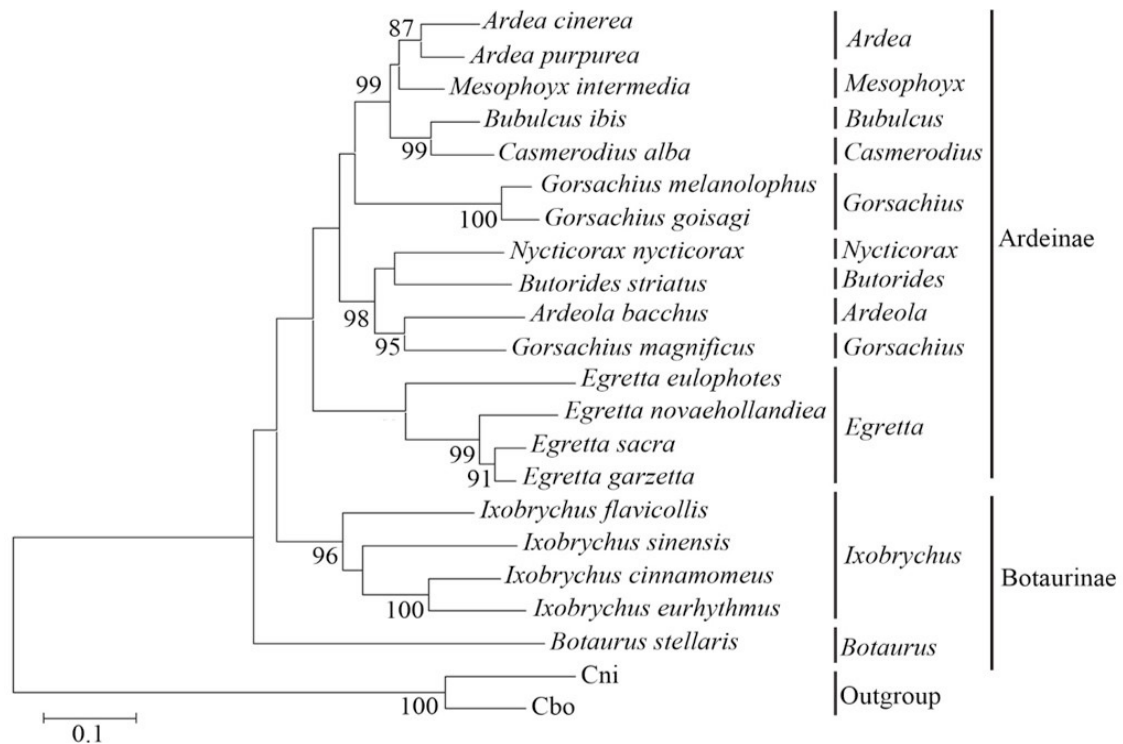


Fig. 2. Phylogenetic tree of Ardeidae constructed from control region sequences. Numbers (in internodes) represent bootstrap values (>90%) from 1000 replications. The codes are shown in Supplementary Table S1.

(hLRTs) as implemented in Modeltest 3.0, the model Hasegawa-Kishino-Yano (HKY) model + Gamma distribution was used (HKY+ G, $-\ln L = 11272.92$, $p < 0.001$, AIC = 22638.04, BIC = 23006.38). We set the shape of the Gamma distribution as 0.55 (estimated by Modeltest). Maximum likelihood method was used to reconstruct the phylogenetic trees based on HKY+ G model. Many clades were supported by bootstrap values of more than 90%. Member of Botaurinae was formed an alone clade, as a sister to Ardeinae (Fig. 1). Species of *Botaurus* was the first to split from the Ardeidae lineage.

All the genera could be discriminated by their distinct clades in the phylogenetic tree except *Gorsachius* (Fig. 1). Bock (1956) considered the four species *gosiagi*, *melanolophus*, *magnificus*, and *leuconotus* to comprise a genus "*Gorsachius*". Comparing to *Egretta* and *Ardea*, the systematics of *Gorsachius* has a little been contended. Control region gene supported that *gosiagi* and *melanolophus* formed one clade. However, *magnificus* was sister to *Ardeola bacchus* (Fig. 2). *Leuconotus* was generally solitary like *gosiagi* and *melanolophus*; observations of *magnificus* were lacking (Payne and Risley, 1976). McCracken and Sheldon (1998) found that *melanolophus* and *leuconotus* did not form a clade based on osteological characteristics. The present study did not include *leuconotus*. Our results indicated that the genus *Gorsachius* might not be monophyletic. To better resolve phylogenetic relationships of the genus *Gorsachius*, more taxon sampling as well as multiple genetic markers are needed for future studies.

Acknowledgements

The study was supported by National Natural Science Foundation of China (31260088, 31560590), Jiangxi Province Talent Project 555, Jiangxi Province Talent Project 555, Jiangxi Province Major Disciplines Academic Leaders (20133BCB22010), the Natural Science Foundation of Jiangxi Province (20171BAB204011), and the Science and Technology Foundation of Jiangxi Provincial Department of Education (GJJ150768).

Supplementary material

There is supplementary material associated with this article. Access the material online at: <http://dx.doi.org/10.17582/journal.pjz/2017.49.5.sc3>

Statement of conflict of interest

Authors have declared no conflict of interest.

References

- Bock, W.J., 1956. *Amer. Mus. Novitat.*, 1779. Available at: <http://hdl.handle.net/2246/5404> (Accessed on May 29, 2017).
Boyce, T.M., Zwick, M.E. and Aquadro, C.E., 1989.

- Genetics*, **123**: 825-836.
Donne-Goussé, C., Laudet, V. and Hänni, C., 2002. *Mol. Phylogenet. Evol.*, **23**: 339-356. [https://doi.org/10.1016/S1055-7903\(02\)00019-2](https://doi.org/10.1016/S1055-7903(02)00019-2)
Felsenstein, J., 1985. *Evolution*, **39**: 783-791. <https://doi.org/10.2307/2408678>
Huang, Z.H. and Ke, D.H., 2016a. *Mitoch. DNA*, **27**: 350-354.
Huang, Z.H. and Ke, D.H., 2016b. *Mitoch. DNA*, **27**: 2036-2039.
Librado, P. and Rozas, J., 2009. *Bioinformatics*, **25**: 1451-1452. <https://doi.org/10.1093/bioinformatics/btp187>
Marshall, H.D. and Baker, A.J., 1997. *Mol. Biol. Evol.*, **14**: 173-184. <https://doi.org/10.1093/oxfordjournals.molbev.a025750>
McCracken, K.G. and Sheldon, F.H., 1998. *Auk*, **115**: 127-141. <https://doi.org/10.2307/4089118>
Kumar, S., Stecher, G. and Tamura, K., 2016. *Mol. Biol. Evol.*, **33**: 1870-1874. <https://doi.org/10.1093/molbev/msw026>
Payne, R.B. and Risley, C.J., 1976. *Misc. Publ. Univ. Mich. Mus. Zool.*, **150**: 1-115.
Posada, D. and Buckley, T.R., 2004. *Syst. Biol.*, **53**: 793-808. <https://doi.org/10.1080/10635150490522304>
Posada, D. and Crandall, K.A., 1998. *Bioinformatics*, **14**: 817-818. <https://doi.org/10.1093/bioinformatics/14.9.817>
Randi, E. and Lucchini, V., 1998. *J. mol. Evol.*, **47**: 449-62. <https://doi.org/10.1007/PL00006402>
Ruokonen, M. and Kvist, L., 2002. *Mol. Phylogenet. Evol.*, **23**: 422-432. [https://doi.org/10.1016/S1055-7903\(02\)00021-0](https://doi.org/10.1016/S1055-7903(02)00021-0)
Strimmer, K. and Haeseler, A.V., 1996. *Mol. Biol. Evol.*, **13**: 964-969. <https://doi.org/10.1093/oxfordjournals.molbev.a025664>
Tamura, K. and Nei, M., 1993. *Mol. Biol. Evol.*, **10**: 512-526.
Thompson, J.D., Gibson, T.J. and Plewniak, F., 1997. The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tool. *Nucl. Acids Res.*, **24**: 4876-4882.
Walberg, M.W. and Clayton, D.A., 1981. *Nucl. Acids Res.*, **9**: 5411-5421. <https://doi.org/10.1093/nar/9.20.5411>
Yang, Z., 1996. *Trends Ecol. Evol.*, **11**: 367-372. [https://doi.org/10.1016/0169-5347\(96\)10041-0](https://doi.org/10.1016/0169-5347(96)10041-0)
Zhang, H., Zhang, Y., Zhang, X.M., Song, N. and Gao, T.X., 2013. *Mitoch. DNA*, **24**: 151-157.
Zhang, Y., Zhang, H., Gao, T.X. and Miao, Z.Q., 2011. *Biochem. Syst. Ecol.*, **39**: 627-634. <https://doi.org/10.1016/j.bse.2011.01.004>