Estimation of divergence times for major lineages of galliform birds: Evidence from complete mitochondrial genome sequences

Xian-Zhao Kan1*, Xi-Feng Li1, Zu-Pei Lei2, Lei Chen1, Hui Gao3, Zu-Yu Yang3, Jian-Ke Yang1, Zhi-Chun Guo1, Lei Yu1, Li-Qin Zhang1 and Chao-Ju Qian1

1The Provincial Key Laboratory of the Conservation and Exploitation Research of Biological Resources in Anhui, College of Life Sciences, Anhui Normal University, Wuhu 241000, China.
2The Administration Bureau of Wuyanling National Nature Reserve, Taishun 325500, China.
3State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China.

Accepted 26 April, 2010

Determining an absolute timescale for avian evolutionary history has been recently challenged by the relaxed molecular clock methods, that rates of molecular evolution can vary significantly among organisms. In this study, we used relaxed molecular clocks to date the divergence of major lineages of Galliformes based on complete mitochondrial genomes. A nucleotide dataset of 13 concatenated protein-coding genes from 22 species of Galliformes was used to investigate the evolutionary divergences within the group. Using Gallus bravardi, Schaubortyx and Gallinuloides fossils as calibration points, divergence times analyses were performed with four relaxed molecular clock methods as follows: (1) Bayesian method of Multidivtime; (2) Bayesian Markov chain Monte Carlo (MCMC) analysis of the Bayesian evolutionary analysis by sampling trees (BEAST); (3) local rate minimum deformation method (LRMD) of TREEFINDER; and (4) nonparametric rate smoothing (NPRS) of TREEFINDER. The various relaxed clock methods all indicated that (1) Megapodiidae originated in the Late Cretaceous; (2) Numididae, Phasianidae, Arborophila and Coturnicinae originated in the Eocene of Palaeogene; (3) Pavoninae and Gallininae originated at the Eocene-Oligocene boundary; (4) Phasianinae and Meleagridinae originated in the Oligocene; (5) divergence times estimation among most genera of Phasianidae were much older than those of the previous studies. Our results might provide a more likely time scale for evolutionary history of the galliform birds.

Key words: Relaxed molecular clock, divergence time, Galliformes, mitochondrial genome.

INTRODUCTION

The avian order Galliformes, comprising the land fowl or gallinaceous birds, is one of the most important groups of birds both for human society and for research (van Tuinen and Dyke, 2004). There are five currently recognized families within the Order Galliformes (Megapodiidae, Cracidae, Numididae, Odontophoridae and Phasianidae) and the family Phasianidae include seven subfamily which are Arborophila, Coturnicinae, Pavoninae, Gallininae, Meleagridinae, Tetraoninae and Phasianinae (Crowe et al., 2006).

The molecular clock hypothesis states that DNA and protein sequences evolve at a rate that is relatively constant over time and among different organisms. However, determining an
absolute timescale (strict molecular clock) for evolutionary history has been recently challenged by the relaxed molecular clock methods that rates of molecular evolution can vary significantly among organisms (Brown et al., 2008; Hipsley et al., 2009; Pereira and Baker, 2006b).

The relaxed molecular clock method were well used to resolve the problem of evolutionary divergence time among some modern birds (Pereira and Baker, 2006a, 2006b; Pereira et al., 2007; Pereira and Wajntal, 2008). Pereira and Baker (2006a) suggested that many phylogenetic splits among avian taxa had been underestimated and needed to be revised.

Within Galliformes, molecular time estimates have been obtained for closely related genera and species using strict molecular clock (Drovski, 2003; He et al., 2009; Huang et al., 2007; Huang et al., 2009; Lucchini et al., 2001; Pereira et al., 2002; Randi et al., 2000; van Tuinen and Dyke, 2004). However, in recent years, this approach was not proved to be ideal, as it did not account for uncertainty in estimating branch lengths and time, including the calibration time, and heterogeneity in the rate of deoxyribonucleic acid (DNA) substitution among sites and in different lineages (Pereira and Baker, 2006b). Several studies within Galliformes using relaxed molecular clock methods have shown strong departures from a strict molecular clock method (Crowe et al., 2006; Grau et al., 2005; Haddrath and Baker, 2001; Pereira and Baker, 2004, 2006b). Pereira and Barker (2006b) pointed out that galliform fossils might not be useful for point calibrations, but instead might be better employed as priors for the estimation of node ages under a Bayesian approach. However, only several mitochondrial genes and limited galliform species were included in previous studies using this approach. In order to better investigate galliformes timescales, it is necessary to adopt relaxed molecular clock and sample more taxa.

Mitochondrial genomes have great potential for resolving ancient patterns of evolutionary history and for serving as a model of genome evolution. In this study, we used relaxed molecular clocks to date the divergence of major lineages of Galliformes based on complete mitochondrial genomes. Our goals were to (1) approximate the divergence time of major lineages of Galliformes, especially among Phasianinae species; and (2) assess the node ages using four different relaxed molecular clock methods.

MATERIALS AND METHODS

Materials

Along with the complete mtDNA sequences of *Tragopan caboti* from our research group (GU187969) (Kan et al., unpublished), all currently available 21 Galliforme complete mitochondrial sequences retrieved from GenBank were used in phylogenetic analyses as follows: *Arboophilina rufipectus* (FJ194942) (He et al., 2009), *Bambusicola thoracica* (EU165706), *Francolinus pintadeanus* (EU165707), *Lophura nycthemera* (EU417810), *Pavo muticus* (EU417811), *Polyplectron bicalcaratum* (EU417812) (Shen et al., 2009), *Coturnix chinensis* (AB073301) (Nishibori et al., 2002), *Coturnix japonica* (AP003195) (Nishibori et al., 2001), *Gallus gallus* (AP003322), *Gallus lafayetta* (AP003325) *Gallus sonneratii* (AP006741) and *Gallus varius* (AP003324) (Nishibori et al., 2005), *Lophura ignita* (AB164627) (Unpublished), *Meleagris gallopavo* (EF153719) (Guo et al., 2009), *Phasianus versicolor* (AB164626) (Unpublished), *Syrmaticus ellioti* (AB164624), *Syrmaticus humiae* (AB164625), *Syrmaticus reevesii* (AB164623) *Syrmaticus soemmerringii* (AB164622) (Unpublished), *Alectura lathami* (AY346091) (Slack et al., 2007), *Numida meleagris* (AP005595) (Nishibori et al., 2004). Two species from Anseriformes (*Anas platyrhynchos*, EU009397; *Branta canadensis*, DQ019124) were designated as out groups.

Methods

The nucleotides for each of the 13 mitochondrial protein-coding genes were aligned using the default parameters of CLUSTALX version 2.0.10 (Larkin et al., 2007). We concatenated the alignments of 13 mitochondrial protein-coding genes and an alignment of 11409 nucleotides was obtained. A substitution saturation analysis (Xia et al., 2003) was performed for subsets with the first, second and third codon positions using DAMBE 4.1.19 (Xia and Xie, 2001). According to the results, all substitutions from the three codon positions were not saturated.

Maximum likelihood (ML) analysis of the nucleotide dataset were performed with PHYML (Guindon and Gascuel, 2003) online in Montpellier bioinformatics platform (www.atgc-montpellier.fr/phyml) with 500 bootstrap replicates and using GTR model. The input topology for the time estimation was the mtDNA ML tree.

Given the importance of fossil calibration points in analyses: (1) *Gallus braviardi* (crown Gallus; 4-5 Mya; node L in Figure 1) (Lambrecht, 1933); (2) Schaubortyx (crown Gallus+Coturnix; 32-38 Mya; node D in Fig. 1) (Brodor, 1964); and (3) Gallinuloides (stem Numididae + Phasianidae; 50-54 Mya; node B in Figure 1) (Dyke et al., 2003). Divergence time analyses were performed with four relaxed molecular clock methods as follows: (1) Bayesian method of Multidivtime (Thorne and Kishino, 2002); (2) Bayesian Markov chain Monte Carlo (MCMC) analysis of Bayesian evolutionary analysis by sampling trees (BEAST) (Drummond and Rambaut, 2007); (3) local rate minimum deformation method (LRMD) of TREEFINDER (Jobb, 2008); and (4) nonparametric rate smoothing (NPRS) of TREEFINDER (Sanderson, 1997).

First, Bayesian method of Multidivtime represents an attempt to explicit model rate change across a tree. We used the BASEML program in the PAML ver. 3.14 package (Yang, 1997) to obtain estimates of the transition/transversion rate ratio and rates for site classes following a discrete gamma distribution with five site classes under the F84 (Felsenstein, 1984) model. Branch lengths and their variance-covariance matrix under a maximum likelihood approach were estimated with the program ESTBRANCHES. Two species used as outgroups of Galliformes were pruned from the tree and Multidivtime was used to estimate the prior and posterior ages of branching events, their standard deviations, and the 95% credibility intervals via MCMC. For MCMC runs, the parameters of the algorithm were set as: burn-in period = 1000, sample frequency = 100, and number of samples = 10,000.

Excepted time between tip and ingroup root (rttm) = 70 Mya with standard deviation (SD) = 35 Mya; rate=rate/Xitrttm, where, rate is the mean of prior distribution for the rate at the root node and x is the median amount of evolution from the ingroup root to the ingroup tips. Analyses were performed at least twice from different random starting points to check for convergence.

Second, the program BEAST 1.4.6 (Drummond and Rambaut, 2007) is currently unique in its ability to estimate the phylogenetic tree and the divergence times simultaneously. We assigned soft upper bounds to the prior distributions of all fossil calibrations using
Figure 1. Chronogram of diversification of major Galliformes lineages based on complete mitochondrial genomes. The program Multidivtime was used to estimate divergence times in conjunction with the tree topology came from the maximum likelihood (ML, 500 replicates). Letters are node names as in Table 1. The horizontal green rectangles represent the estimated 95% credibility intervals of divergence times. Scale at the bottom represents the geological time in millions of years ago according to the Geological Society of America.

log-normal distributions. We also specified a Yule prior on the rates of cladogenesis. The data set was assumed to have evolved under a GTR model with invariant sites and gamma-distributed rate heterogeneity. Five independent analyses of $25 \times 10^6$ generations each were run. Output from each run was analyzed using TRACER 1.4; 25% of the trees were discarded as burning and the remaining were combined using TreeAnnotator 1.4.6 to produce the timescale.

Finally, the program TREEFINDER infers even large trees by maximum likelihood under a variety of models of sequence evolution. It accepts both nucleotide and amino acid data and takes rate heterogeneity into account. Separate models can be assumed for use-defined data partitions, separate rates, separate edge lengths and separate character compositions. Tree search can be guided by user-supplied topological constraints and start trees. Here, we used NPRS and LRMD methods in TREEFINDER to estimate the divergence times.

RESULTS AND DISCUSSION

Divergence time estimates for major lineages of Galliformes based on complete mitochondrial genomes are given in Table 1. We found that the choice of the relaxed clock method had a substantial influence on inferred ages. Bayesian (BEAST), LRMD (TREEFINDER), and NPRS (TREEFINDER) tended to deliver similar estimates for
most clades (Table 1). In contrast, Bayesian (Multidivtime) generated considerably older dates in most clades, despite using the same bootstrapped phylogenies. The obvious differences at node A and U were also observed (Table 1). These differences of divergence time in this study might result from methodological differences. In Bayesian of BEAST, maximum likelihood (ML) tree was used as the topology for the time estimation, while the program BEAST does not require a fixed topology.

Furthermore, the method NPRS relies on minimization of ancestor-descendant local rate changes and is motivated by the likelihood that evolutionary rates are autocorrelated in time, and the method LRMD tries to keep the real rates as similar as possible to ideal local rates with well-defined dependencies. Despite these differences, all our methods strongly supported that the basal splits within Galliformes occurred deep within the Cretaceous as stated in previous studies (Crowe et al., 2006; Pereira and Baker, 2006; van Tuinen and Dyke, 2004). We estimated that crown group radiation that began in the Cretaceous between 79.9 Mya - 90.3 Mya (average value) with the 95% credibility intervals when the ancestor of Megapodiidae branched off from the remaining galliforms (Table 1, Figure 1), and were included within the 95% credibility intervals of 71 - 114 and 79 - 107 Mya for this split obtained with Bayesian approach by Pereira et al. (2006b) and Crowe et al. (2006) respectively. Moreover, the divergence times (average value) of calibration points fit the three fossil ages (Table 1).

Our results also strongly supported that the families Numididae, Phasianidae and all six subfamilies of Phasianidae in this study originated in the Paleogene of tertiary (Crowe et al., 2006; He et al., 2009; Pereira and Baker, 2006; van Tuinen and Dyke, 2004) (Table 1, Figure 1). The split between Numididae and Phasianidae occurred in the Eocene of Paleogene between 39.5 - 44.1 Mya with the 95% credibility intervals, and this result was younger than those estimated by Crowe et al. (2006), Pereira and Baker (2006b); van Tuinen and Dyke (2004) (Table 1, Figure 1). The split between Numididae and Phasianidae occurred in the Eocene of Paleogene between 39.5 - 44.1 Mya with the 95% credibility intervals, and this result was younger than those estimated by Crowe et al. (2006), Pereira and Baker (2006b); van Tuinen and Dyke (2004). Furthermore, our results indicated that the subfamilies Arborophilinae and Coturnicinae originated in the Eocene of Paleogene and Pavoninae and Gallininae of Phasianidae originated at the Eocene-Oligocene boundary. However, divergence times estimation among most genera of Phasianidae were much older than those of the previous studies (Bloomer and Crowe, 1998; Huang et al., 2007; Huang et al., 2009; Kimball et al., 1997; Randi et al.,

Table 1. Estimated divergence times for major lineages of Galliformes compared with different methods.

<table>
<thead>
<tr>
<th>Node</th>
<th>tMRCA</th>
<th>Bayesian (Multidivtime)</th>
<th>Bayesian (BEAST)</th>
<th>LRMD (TREEFINDER)</th>
<th>NPRS (TREEFINDER)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Galliformes</td>
<td>77.3 (74.4, 78.6)</td>
<td>78.8 (70.2, 90.2)</td>
<td>90.8 (87.5, 94.1)</td>
<td>94.2 (87.6, 98.1)</td>
</tr>
<tr>
<td>B</td>
<td>Numidiae-Phasianidae</td>
<td>52.9 (52.0, 55.0)</td>
<td>51.5 (48.6, 53.4)</td>
<td>52.0 (50.1, 53.9)</td>
<td>52.6 (50.4, 53.8)</td>
</tr>
<tr>
<td>C</td>
<td>Phasianidae</td>
<td>44.3 (41.3, 47.3)</td>
<td>41.2 (38.2, 44.4)</td>
<td>42.8 (41.2, 44.4)</td>
<td>38.9 (37.1, 40.1)</td>
</tr>
<tr>
<td>D</td>
<td>Coturnicinae, Pavoninae, Gallininae, Phasianidae, and Meleagridae</td>
<td>40.7 (37.5, 44.0)</td>
<td>36.8 (34.3, 38.8)</td>
<td>35.0 (32.2, 37.9)</td>
<td>35.0 (33.3, 36.1)</td>
</tr>
<tr>
<td>E</td>
<td>Pavoninae, Gallininae, Phasianidae, and Meleagridae</td>
<td>38.3 (35.1, 41.7)</td>
<td>34.3 (31.8, 36.5)</td>
<td>32.6 (29.8, 35.5)</td>
<td>33.0 (31.3, 34.0)</td>
</tr>
<tr>
<td>F</td>
<td>Pavoninae and Gallininae</td>
<td>37.5 (34.2, 41.0)</td>
<td>33.3 (30.8, 35.8)</td>
<td>32.2 (29.6, 35.1)</td>
<td>32.3 (30.7, 33.3)</td>
</tr>
<tr>
<td>G</td>
<td>Pavo and Gallininae</td>
<td>36.2 (32.9, 39.7)</td>
<td>31.7 (29.3, 34.5)</td>
<td>31.3 (28.7, 34.0)</td>
<td>31.3 (29.7, 32.3)</td>
</tr>
<tr>
<td>H</td>
<td>Gallininae</td>
<td>29.0 (25.8, 32.5)</td>
<td>24.3 (21.3, 27.2)</td>
<td>25.6 (23.6, 27.8)</td>
<td>25.4 (24.2, 26.1)</td>
</tr>
<tr>
<td>I</td>
<td>Bambusicola and Gallus</td>
<td>25.2 (22.2, 28.7)</td>
<td>20.5 (17.7, 23.5)</td>
<td>22.5 (20.8, 24.4)</td>
<td>22.1 (21.1, 22.7)</td>
</tr>
<tr>
<td>J</td>
<td>Gallus</td>
<td>9.40 (7.80, 11.6)</td>
<td>8.10 (6.90, 9.30)</td>
<td>8.90 (8.10, 9.70)</td>
<td>8.2 (8.0, 8.3)</td>
</tr>
<tr>
<td>K</td>
<td>G. varius, G. gallus and G. lafayettii</td>
<td>7.80 (6.70, 9.70)</td>
<td>6.70 (5.80, 7.70)</td>
<td>7.00 (6.30, 7.60)</td>
<td>6.5 (6.4, 6.6)</td>
</tr>
<tr>
<td>L</td>
<td>G. varius and G. gallus</td>
<td>5.10 (4.50, 6.30)</td>
<td>4.40 (3.80, 4.80)</td>
<td>4.50 (4.00, 5.00)</td>
<td>4.2 (4.1, 4.2)</td>
</tr>
<tr>
<td>M</td>
<td>Phasianinae, including Meleagridinae</td>
<td>31.3 (27.9, 34.8)</td>
<td>28.5 (25.7, 31.3)</td>
<td>29.3 (27.6, 30.9)</td>
<td>27.3 (25.9, 28.1)</td>
</tr>
<tr>
<td>N</td>
<td>Phasianus, Lophura, Synmaticus, and Meleagrides</td>
<td>29.7 (26.2, 33.3)</td>
<td>27.0 (23.8, 29.7)</td>
<td>28.0 (26.4, 29.6)</td>
<td>26.2 (24.9, 27.0)</td>
</tr>
<tr>
<td>O</td>
<td>Phasianus, Lophura, and Synmaticus</td>
<td>19.1 (16.1, 22.4)</td>
<td>17.7 (15.6, 20.0)</td>
<td>18.8 (17.5, 19.9)</td>
<td>18.1 (17.2, 18.7)</td>
</tr>
<tr>
<td>P</td>
<td>S. reevesii, S. elliottii, and S. humiae</td>
<td>14.8 (12.2, 17.9)</td>
<td>13.6 (11.5, 15.9)</td>
<td>14.3 (13.3, 15.4)</td>
<td>14.4 (13.7, 14.8)</td>
</tr>
<tr>
<td>Q</td>
<td>S. elliottii and S. humiae</td>
<td>13.4 (10.9, 16.4)</td>
<td>12.2 (10.0, 14.2)</td>
<td>13.2 (12.2, 14.1)</td>
<td>13.2 (12.5, 13.6)</td>
</tr>
<tr>
<td>R</td>
<td>S. elliottii and S. humiae</td>
<td>0.12 (0.03, 0.24)</td>
<td>0.14 (0.06, 0.23)</td>
<td>0.14 (0.13, 0.15)</td>
<td>0.20 (0.19, 0.20)</td>
</tr>
<tr>
<td>S</td>
<td>Phasianus and Lophura</td>
<td>15.6 (12.9, 18.9)</td>
<td>14.2 (12.0, 16.2)</td>
<td>15.4 (14.3, 16.4)</td>
<td>15.0 (14.3, 15.5)</td>
</tr>
<tr>
<td>T</td>
<td>Lophura</td>
<td>8.10 (6.30, 10.5)</td>
<td>7.10 (5.00, 8.90)</td>
<td>8.2 (7.70, 8.80)</td>
<td>7.9 (7.5, 8.1)</td>
</tr>
<tr>
<td>U</td>
<td>Coturnicinae</td>
<td>20.6 (17.3, 24.6)</td>
<td>19.2 (15.3, 22.8)</td>
<td>7.90 (3.70, 15.0)</td>
<td>18.6 (17.6, 19.1)</td>
</tr>
</tbody>
</table>

Note: Estimated time to most recent common ancestor (tMRCA) of clades of non-controversial monophyletic status. 95% credibility intervals are given in parentheses.
ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (No. 30870172), the State Key Laboratory of Systematic and Evolutionary Botany, China (2009), the Key Program of Natural Science Foundation of the Anhui Higher Education Institutions (No. KJ2008A132C), the Natural Science Foundation of Anhui Normal University (No. 2008xzx14), the Key Laboratory of Biotic Environment and Ecology Safety in Anhui Province and the Program for Innovative Research Team in Anhui Normal University. The authors also wish to thank the two anonymous reviewers for their insightful comments on the manuscript.

REFERENCES


Nishibori M, Tsuzuki M, Hayashi T, Yamamoto Y, Yazue H (2002). Complete nucleotide sequence of the Coturnix chinsensis (blue-breasted quail) mitochondrial genome and a phylogenetic