*Evolutionary Ecology Research*, 2010, **12**: 987–994

# Correlation of life-history traits with the molecular diversity of bird nuclear coding elements

Tangjie Zhang[1], Yuzhi Liu[2] and Zhiyue Wang[3]

[1]*College of Veterinary Medicine, Yangzhou University, Yangzhou, China,* [2]*College of Animal Science and Technology, Agricultural University of Hebei, Baoding, China and*
[3]*College of Animal Science and Technology, Yangzhou University, Yangzhou, China*

## ABSTRACT

**Question:** Is there a relationship between mutations in nuclear coding genes sites and life-history traits? Analyses of polymorphism data of Aves nuclear genes suggest that this is the case.

**Data:** Obtained from Polymorphix and Popset of GenBank: 752 groups of polymorphism data from 104 Aves nuclear genes of 15 families, 53 genera, 297 species.

**Search method:** We used Watterson's estimator ($\theta_w$), calculated separately from the non-synonymous ($\theta_a$) and synonymous ($\theta_s$) sites of the various gene fragments. We performed correlation analysis of the proportion of average non-synonymous mutation sites in coding genes sites and four life-history traits (body mass, metabolic rate, generation time, and longevity).

**Conclusions:** The proportion of non-synonymous sites is significantly negatively correlated with body mass. But the proportion of non-synonymous sites is not significantly correlated with generation time. A larger fraction of slightly deleterious mutations is fixed in species with either a bigger body mass or a low effective population size (Ne).

*Keywords*: Aves, life-history traits, mutation, nearly neutral theory.

## INTRODUCTION

Clarifying the contribution of neutral, beneficial, and deleterious changes to DNA diversity and evolution is of central importance in evolutionary genetics. Over the last several decades, the predominant model has been the neutral theory, first proposed by Kimura in 1968, which assumes that the majority of changes that we observe as polymorphism or divergence in DNA sequences are selectively neutral. Ohta (1973, 1976) later developed and consummated the neutral theory with the nearly neutral theory. The nearly neutral theory assumes that both slightly deleterious and slightly beneficial mutations (with selection coefficient $s$ such that $|s| < 1/\text{Ne}$, where Ne is effective population size) are effectively neutral and spread throughout the population because of random genetic drift. The nearly neutral theory,

*Correspondence: T. Zhang, College of Veterinary Medicine, Yangzhou University, Yangzhou 225009, China. e-mail: slx@yzu.edu.cn

which involves mainly slightly deleterious mutations, plays a major role in evolution at the molecular level. If Ne does not influence the distribution of selection coefficients, the fraction of slight mutations in small populations should be higher than in large populations (Ohta, 1976; Ohta and Gillespie, 1996). Because selection is commonly purifying, a larger fraction of slightly deleterious mutations can be fixed in species with low Ne.

The common and basic measure of the strength of purifying selection is $K_a/K_s$ [ratio of the rate of non-synonymous substitutions ($K_a$) to the rate of synonymous substitutions ($K_s$)] (Ohta, 1992). It is also known that different genes are under different selection pressures. In studies of molecular evolution of species, $K_a/K_s$ is widely used as an indicator of selective pressure acting on a certain protein-coding gene. The nearly neutral prediction was confirmed in comparative studies of primates and rodents (Li and Tanimura, 1987; Ohta, 1995; Eyre-Walker *et al.*, 2002), mammals, birds, and drosophilids (Keightley and Eyre-Walker, 2000), and large- versus small-bodied mammals (Popadin *et al.*, 2007). However, these attempts to discover the mutations of functional sequences used only a few comparisons between high taxonomic categories.

Here, we are interested in why the proportion of non-synonymous substitutions differs so much between taxa. We extend the analysis of molecular diversity of nuclear coding elements through as many nuclear genetic polymorphism comparisons as possible within Aves. The main reason we chose birds for a comparative approach is that bird species are strikingly long-lived compared with their similar-size mammalian counterparts (Holmes *et al.*, 2001; Holmes and Ottinger, 2003). This peculiarity will help us to understand the impact of life span on nuclear molecular diversity. By analysing nuclear DNA polymorphism data from 104 nuclear genes of Aves, we investigated how life-history traits affect molecular evolution on the nuclear genomic scale at synonymous and non-synonymous sites.

## MATERIALS AND METHODS

### Source data

We obtained data for 77 polymorphic nuclear genes of birds from Polymorphix (Bazin *et al.*, 2005) and 1155 polymorphic nuclear genes of birds from Popset of GenBank. Two sequences were not clustered if there was a mismatch of >50nt with <80% similarity flanking using the CLUSTALW program (Thompson *et al.*, 1994). Such a mismatch is interpreted as evidence that the sequences represent duplicate genes in distinct genomic contexts. Sequences without stop codons were inspected by eye and corrected when required. Dubious sequences were manually removed. There were 752 groups in 104 nuclear genes of 15 families, 53 genera, 297 species after repeat and noisy data were removed. Each group was aligned by eye using CLUSTALW. Alignments of all groups are available on request.

### Polymorphism sequence data analyses

Genetic diversity measures: Watterson's estimator ($\theta_w$) was calculated separately from the non-synonymous ($\theta_a$) and synonymous ($\theta_s$) sites of the different gene fragments analysed, expressed as per-site level of diversity:

$$\theta_w = \frac{P}{L\sum_{i=1}^{n-1}\frac{1}{i}} \quad (1)$$

where $P$ is the number of synonymous or non-synonymous polymorphisms, $L$ is the number of synonymous or non-synonymous sites, and $n$ is the number of sequences sampled.

## Life-history data

Body mass, age of female sexual maturity, gestation duration, maximum longevity, and metabolic rate were obtained from the AnAge database (de Magalhaes *et al.*, 2005), which contains values from multiple recent compilations from the literature or the compilations by Bennett (1986). Measuring generation time is not straightforward because it depends on the age structure of species (Charlesworth, 1994), about which data are lacking for most birds. We took either female sexual maturity or the sum of female sexual maturity and gestation duration as rough approximations of generation time. The two gave similar results. In some cases no estimate was available, so the mean for both sexes was used, or the geometric mean of maximum and minimum was calculated. Metabolic rates are especially difficult to obtain and we did not try to measure them directly. Instead, we used two sorts of rate estimates: (1) those from the published work of others listed in AnAge (http://genomics.senescence.info/species/); (2) the rest from the equation that relates metabolic rate to body mass.

## Statistical tests

$\theta_w$ values were arcsine transformed (Sokal and Rohlf, 1981). Quantitative life-history variables were log-transformed. First we calculated $\theta_s$ for synonymous sites and $\theta_a$ for non-synonymous sites for different genes, and then the weighted average of $\theta_s$ and $\theta_a$ from all analysed genes for a certain species. To estimate and confirm effects of life-history variables on non-synonymous mutations of nuclear coding genes, we correlated the proportion of averaged non-synonymous mutations sites in coding gene sites

$$\frac{\dfrac{P_A}{L_A}}{\dfrac{P_S + P_A}{L_S + L_A}}$$

with life-history variables using independent regression:

$$\frac{\dfrac{P_A}{L_A}}{\dfrac{P_S + P_A}{L_S + L_A}} = \frac{\theta_A}{\theta_{A+S}} \tag{2}$$

where $P_A$ and $P_S$ is the number of synonymous or non-synonymous polymorphisms, $L_A$ and $L_S$ is the number of synonymous or non-synonymous sites, and $\theta_{A+S}$ is Watterson's estimator ($\theta_w$) of Aves nuclear encoding sites.

## RESULTS

### Estimates of Aves life-history variables

Table 1 provides estimates of body mass, generation time, and metabolic rate for each species, gleaned from the AnAge database, which contains values from multiple recent compilations from the literature (Bennett, 1986). Generation times are the rough approximations based on female sexual maturity or the sum of female sexual maturity and gestation.

**Table 1.** Body mass, generation time, and metabolic rate of species studied

| Species | Generation time (days) | Body mass (g) | Maximum life span (years) | Metabolic rate (W/g) |
|---|---|---|---|---|
| *Acrocephalus arundinaceus* | 365 | 29 | 10 | 0.011735 |
| *Aegithalos caudatus* | 365 | 8.9 | 10.8 | 0.02236 |
| *Aegolius acadicus* | 392 | 124 | 17.5 | 0.005274 |
| *Aegolius funereus* | 392 | 120 | 15.9 | — |
| *Amazona aestiva* | — | 451 | 49 | — |
| *Amazona albifrons* | 730 | 215 | 25.3 | — |
| *Amazona auropalliata* | — | — | 49 | — |
| *Amazona autumnalis* | — | 416 | 27 | — |
| *Amazona barbadensis* | — | — | 49 | — |
| *Anas americana* | 388 | 793.8 | 21.3 | — |
| *Anas penelope* | 389 | 723 | 34.7 | 0.003947 |
| *Anas platyrhynchos* | 391 | 1048 | 29.1 | 0.003988 |
| *Anas strepera* | 266 | 791 | 22.3 | — |
| *Andropadus nigriceps* | 365 | 150 | 11 | — |
| *Anser caerulescens* | 754 | 2807 | 26.6 | — |
| *Anser canagica* | 1095 | — | — | — |
| *Apus melba* | 730 | 102.7 | 26 | — |
| *Athene cunicularia* | 365 | 150 | 11 | — |
| *Brachyramphus marmoratus* | 730 | 222 | 7.2 | — |
| *Branta bernicla* | 754 | 1163 | 28.7 | — |
| *Bubo scandiacus* | 730 | 2026 | 28 | 0.002085 |
| *Calidris alpina* | 751 | 52.5 | 28.8 | — |
| *Calidris pusilla* | 385 | 27 | 19.3 | — |
| *Campethera nivosa* | 365 | — | — | — |
| *Coereba flaveola* | — | 10.1 | 6.9 | — |
| *Colinus virginianus* | 388 | 194 | 6.4 | — |
| *Corvus corone* | 1095 | 518 | 19 | 0.006473 |
| *Coturnix coturnix* | 365 | 97 | 11 | 0.009186 |
| *Coturnix japonica* | 80 | 115 | 6 | — |
| *Cyanistes caeruleus* | 365 | 10.3 | 14.6 | — |
| *Dendragapus obscurus* | 391 | 1131 | 14 | 0.004383 |
| *Dendrocopos minor* | 365 | 19.8 | 9.8 | — |
| *Dendroica dominica* | 377 | 9.8 | 5.1 | 0.016367 |
| *Dixiphia pipra* | 548 | — | — | — |
| *Elanus caeruleus* | 365 | 260.5 | 6 | — |
| *Eudocimus ruber* | — | 615 | 33.2 | — |
| *Falcipennis canadensis* | 385 | 417.3 | 13 | — |

**Table 1.**—*continued*

| Species | Generation time (days) | Body mass (g) | Maximum life span (years) | Metabolic rate (W/g) |
|---|---|---|---|---|
| *Falco peregrinus* | 395 | 840 | 25 | — |
| *Ficedula albicollis* | 365 | 12.7 | 7.9 | — |
| *Ficedula hypoleuca* | 365 | 11.7 | 15 | 0.019829 |
| *Hemignathus virens* | — | 13.35 | 12 | — |
| *Hirundo rustica* | 380 | 18 | 16 | 0.017544 |
| *Icterus bullockii* | 376 | 36.62 | 8 | — |
| *Icterus galbula* | — | 37.5 | 14 | 0.013472 |
| *Icterus spurius* | 378 | 21 | 10.9 | — |
| *Lagopus mutus* | 365 | 535.5 | — | — |
| *Lamprotornis hildebrandti* | — | 55.8 | 16.9 | — |
| *Lepidothrix coronata* | 548 | — | — | — |
| *Leucosticte atrata* | 378 | 24 | 5.7 | — |
| *Leucosticte tephrocotis* | 379 | 35.7 | 6.6 | — |
| *Loxops coccineus coccineus* | 380 | 10.4 | 10 | — |
| *Malurus lamberti* | 280 | — | — | — |
| *Malurus leucopterus* | 280 | — | — | — |
| *Malurus melanocephalus* | 280 | — | — | — |
| *Megascops kennicottii* | 395 | 205 | 19 | — |
| *Motacilla flava* | 376 | 14.7 | 8.8 | 0.017483 |
| *Oporornis philadelphia* | — | 10 | 7.9 | — |
| *Otus flammeolus* | 387 | 55.5 | 8.1 | — |
| *Otus lettia* | 387 | — | — | — |
| *Otus senegalensis* | 387 | — | — | — |
| *Oxyura jamaicensis ferruginea* | 388 | 542 | 13.6 | — |
| *Paroreomyza montana* | 624 | 13.15 | — | — |
| *Parus major* | 365 | 17.9 | 15.4 | — |
| *Passer domesticus* | 163 | 25.3 | 23 | — |
| *Passerina amoena* | 377 | 15.15 | 9.9 | — |
| *Passerina caerulea* | 376 | 28.4 | 10.2 | — |
| *Passerina cyanea* | 377 | 15 | 11 | — |
| *Pheucticus ludovicianus* | 377 | 47.3 | 24 | — |
| *Pheucticus melanocephalus* | 377 | 42 | 25 | — |
| *Picus canus* | — | 137 | — | — |
| *Picus viridis* | 365 | 176 | 15.2 | — |
| *Ptilonorhynchus violaceus* | 1710 | 203 | — | — |
| *Riparia riparia* | 380 | 13.6 | 10 | 0.017118 |
| *Sericulus chrysocephalus* | 1440 | 203 | — | — |
| *Somateria mollissima* | 758 | 2130 | 37.8 | — |
| *Stercorarius parasiticus* | 1226 | 445.5 | 31.1 | — |
| *Sterna antillarum* | 1116 | 40 | 24.1 | — |
| *Strix aluco* | 365 | 520 | 21.5 | — |
| *Taeniopygia guttata* | 60 | 12 | 14.5 | — |
| *Thalasseus sandvicensis* | 1117 | 250 | 30.8 | — |
| *Turdus merula* | 365 | 103.2 | 21.8 | — |
| *Uria aalge* | 1856 | 956 | 38 | 0.007136 |
| *Zonotrichia albicollis* | 378 | 21.1 | 9.7 | 0.013762 |

### Genetic diversity of Aves nuclear genes

Our results showed that the mean per-site nucleotide diversity ($\theta_{A+S}$) was 0.01253 ± 0.01301 ($n = 185$). This resembles Nabholz's result, i.e. the mean per-site synonymous nucleotide diversity of cytb is 0.040 ± 0.041 ($n = 147$). This suggests that mitochondrial genes evolve approximately 5–10 times faster than nuclear genes.

Nabholz found that mutation rate is highly variable by measuring the neutrally evolving third codon positions of cytb. Compared with our downloaded data (752), our nuclear database is smaller, but we were still able to show that nuclear diversity is highly variable among bird lineages. It varies by two orders of magnitude, ranging from 0.00031 to 0.06556.

### Correlation of $\theta_A/\theta_{A+S}$ to life-history variables

The mutations of intron and synonymous sites, which are different from non-synonymous sites, are neutral, as they have no effect on the encoded protein. But are there correlations between proportion of non-synonymous mutations in all encoding genes mutations and life-history variables? In Table 2, we show the correlation between life-history variables and $\theta_A/\theta_{A+S}$ as estimated from Aves nuclear genes.

The numbers of three-variable samples are different since life-history datasets of some species are difficult to obtain, especially metabolic rate data. In terms of single variables, we see that the proportion of averaged non-synonymous mutation sites in coding genes, which is represented by $\theta_A/\theta_{A+S}$, is significantly linked to body mass ($P = 0.003$) and metabolic rate ($P = 0.003$). We found no correlations between $\theta_A/\theta_{A+S}$ and maximum life span ($P = 0.324$) or $\theta_A/\theta_{A+S}$ and generation time ($P = 0.568$).

## DISCUSSION

In the present study, the nuclear diversity among bird lineages varied from 0.00031 to 0.06556. Variation has two sources: (a) sampling variance in a finite sample of sites, and (b) differences in the true level of variation across loci. The estimates of variability suggest that the mean per-site encoding nucleotide diversity ($\theta_{A+S}$) was 0.01253 ± 0.01301 ($n = 185$), although this conclusion depends on the assumption that these 4133 protein-coding genes are representative of the entire Aves genome. While a survey of 4133 protein-coding genes is certainly not large enough to provide a precise measure, there is no reason to suspect that these loci are not representative. Furthermore, the mean per-site nucleotide diversity we sampled from Aves is approximately one-fifth to one-tenth of mitochondrial genetic diversity, which is consistent with evolutionary rates of nuclear and mitochondrial genes. Therefore, we are confident that substantial DNA-sequence variation exists among birds

**Table 2.** Single-variable regressions

| | Trait | *n* | Slope | $R^2$ | *P*-value |
|---|---|---|---|---|---|
| $\theta_A/\theta_{A+S}$ | Body mass | 32 | 0.085 | 0.246 | 0.003 |
| | Maximum life span | 31 | 0.095 | 0.034 | 0.312 |
| | Generation time | 36 | 0.133 | 0.061 | 0.139 |
| | Metabolic rate | 6 | – 1.140 | 0.369 | 0.148 |

and the influence of life history on polymorphism appears to be pervasive. In fact, the theoretical predictions rely on assumptions that are unlikely to hold at all loci (Gillespie, 1991).

Nucleotide substitutions in genes that code for proteins can be either synonymous or non-synonymous. Non-synonymous changes are usually eliminated by purifying selection but, under certain conditions, Darwinian selection may lead to their retention. Investigating the number and rate of non-synonymous substitutions in coding genes may therefore provide information about the degree of selection. Our studies show that nuclear protein-coding genes of large birds have a higher proportion of average non-synonymous mutation sites relative to synonymous sites. Because non-synonymous substitutions are more harmful than synonymous substitutions, large birds with high metabolic rates appear to experience less efficient purifying selection than small birds with low metabolic rates. As a result, large birds seem to accept more harmful mutations, causing a higher level of amino acid dissimilarity. This suggests that purifying selection performs better in species with larger effective sizes (Ne), and it represents stronger and more pervasive evidence for the nearly neutral theory. This is consistent with the observations of Popadin, that organisms with large body mass have small effective population sizes and thus inefficient selection against slightly deleterious mutations by mitochondrial protein-coding genes (Popadin *et al.*, 2007). Slightly deleterious mutations that are fixed in populations have been reported previously (Bustamante *et al.*, 2005; Williamson *et al.*, 2005; Nabholz *et al.*, 2008). However, previous attempts to determine evolutionary patterns used only a few comparisons on the basis of higher category or mitochondrial protein-coding genes. Here, we deal with the nuclear gene mutations of birds as much as possible. Birds not only provide quite a large and pervasive comparison, but also better represent a genome than do a few mitochondrial genes.

Many studies have shown that larger animals have lower molecular rates of evolution in absolute time, possibly because of the interaction of both the biotic and physical environment with phenotypic variation (McKinney, 1997; Cardillo *et al.*, 2005; Liow *et al.*, 2008). By using a large dataset of fossil mammals, Liow *et al.* (2008) found that large-sized mammal genera and species have higher origination and extinction rates, and therefore shorter durations, which is surprisingly consistent with the results of molecular studies. Generally speaking, large mammals suffer from compound disadvantages. They have more deleterious mutations (Popadin *et al.*, 2007), smaller population sizes (Blackburn and Gaston, 1999), longer generation times (Brook and Bowman, 2005), and lower metabolic rates (Symonds, 1999). It makes sense that bigger animals have more slightly deleterious mutations than do smaller animals. Otherwise, bigger animals would have greater extinction risk.

## ACKNOWLEDGEMENTS

We thank Adam Eyre-Walker for helpful advice, and the Centre for the Study of Evolution and School of Biological Sciences, University of Sussex for sharing and supporting works. This research was also supported by the Science & Technology Foundation of Jiangsu and a Study-Abroad Scholarship from Jiangsu Province, China.

## REFERENCES

Bazin, E., Duret, L., Penel, S. and Galtier, N. 2005. Polymorphix, a sequence polymorphism database. *Nucleic Acids Research*, **33**: D481–D484.

Bennett, P.M. 1986. *Comparative studies of morphology, life history and ecology among birds*. DPhil thesis, University of Sussex, UK.

Blackburn, T.M. and Gaston, K.J. 1999. The relationship between animal abundance and body size: a review of the mechanisms. *Adv. Ecol. Res.*, **28**: 181–210.

Brook, B.W. and Bowman, D. 2005. One equation fits overkill: why allometry underpins both prehistoric and modern body size-biased extinctions. *Popul. Ecol.*, **47**: 137.

Bustamante, C.D., Fledel-Alon, A.,Williamson, S., Nielsen, R., Hubisz, M.T., Glanowski. S. *et al.* 2005. Natural selection on protein-coding genes in the human genome. *Nature*, **437**: 1153–1157.

Cardillo, M., Mace, G.M., Jones, K.E., Bielby, J., Bininda-Emonds, O.R.P., Sechrest, W. *et al.* 2005. Multiple causes of high extinction risk in large mammal species. *Science*, **309**: 1239–1241.

Charlesworth, B. 1994. *Evolution in Age-structured Populations*. Cambridge: Cambridge University Press.

de Magalhaes, J.P., Costa, J. and Toussaint, O. 2005. Hagr: the human ageing genomic resources. *Nucleic Acids Research*, **33**: D537–D543.

Eyre-Walker, A., Keightley, P.D., Smith, N.G.C. and Gaffney, D. 2002. Quantifying the slightly deleterious model of molecular evolution. *Mol. Biol. Evol.*, **19**: 2142–2149.

Gillespie, J.H. 1991. *The Causes of Molecular Evolution*. Oxford: Oxford University Press.

Holmes, D.J. and Ottinger M.A. 2003. Birds as long-lived animal models for the study of aging. *Experimental Gerontology*, **38**: 1365–1375.

Holmes, D.J., Flückiger, R. and Austad, S.N. 2001. Comparative biology of aging in birds: an update. *Experimental Gerontology*, **36**: 869–883.

Keightley, P.D. and Eyre-Walker, A. 2000. Deleterious mutations and the evolution of sex. *Science*, **290**: 331–333.

Kimura, M. 1968. Evolutionary rate at molecular level. *Nature*, **217**: 624–626.

Li, W.H. and Tanimura, M. 1987. The molecular clock runs more slowly in man than in apes and monkeys. *Nature*, **326**: 93–96.

Liow, L.H., Fortelius, M., Bingham, E., Lintulaakso, K., Mannila, H., Flynn, L. *et al.* 2008. Higher origination and extinction rates in larger mammals. *Proc. Natl. Acad. Sci. USA*, **105**: 6097–6102.

McKinney, M.L. 1997. Extinction vulnerability and selectivity: combining ecological and paleontological views. *Annu. Rev. Ecol. Syst.*, **28**: 495–516.

Nabholz, B., Glémin, S. and Galtier, N. 2008. Strong variations of mitochondrial mutation rate across mammals – the longevity hypothesis. *Mol. Biol. Evol.*, **25**: 120–130.

Ohta, T. 1973. Slightly deleterious mutant substitutions in evolution, *Nature*, **246**: 96–98.

Ohta, T. 1976. Role of very slightly deleterious mutations in molecular evolution and polymorphism. *Theor. Popul. Biol.*, **10**: 254–275.

Ohta, T. 1992. The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.*, **23**: 263–286.

Ohta, T. 1995. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *J. Mol. Evol.*, **40**: 56–63.

Ohta, T. and Gillespie, J.H. 1996. Development of neutral and nearly neutral theories. *Theor. Popul. Biol.*, **49**: 128–142.

Popadin, K., Polishchuk, L.V., Mamirova, L., Knorre, D. and Gunbin, K. 2007. Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals. *Proc. Natl. Acad. Sci. USA*, **104**: 13390–13395.

Sokal, R.R. and Rohlf, F.J. 1981. *Biometry*. San Francisco, CA: W.H. Freeman.

Symonds, M.R.E. 1999. Life histories of the Insectivora: the role of phylogeny, metabolism and sex differences. *J. Zool.*, **249**: 315–337.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**: 4673–4680.

Williamson, S.H., Hernandez, R., Fledel-Alon, A., Zhu, L., Nielsen, R. and Bustamante, C.D. 2005. Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proc. Natl. Acad. Sci. USA*, **102**: 7882–7887.