Evolutionary Ecology Research, 2012, **14**: 353–360

Correlates of genetic diversity in bird nuclear genes

Tangjie Zhang and Qing Liu

College of Veterinary Medicine, Yangzhou University, Yangzhou, China

ABSTRACT

Background: The neutral theory of evolution proposes that diversity is the result of the accumulation of neutral substitutions. The term 'neutral' refers to a gene (or a genomic locus) that has no or almost no effect on fitness. The main hypotheses that would account for neutral genetic diversity are related to life-history traits.

Question: Is there a relationship between nuclear neutral diversity in birds (class Aves) and their life-history traits, including generation time, metabolic rate, longevity, and body mass?

Data: 752 groups of polymorphisms from 104 nuclear genes in 297 species of Aves belonging to 53 genera and 15 families. Data were taken from the Polymorphix database and the Popset database of GenBank.

Search method: We used logistic regression analysis and phylogenetic regression of independent variables to analyse the relationship between Watterson's estimator (θ_w) of weighted neutral sites and life-history variables, including generation time, body mass, and maximum longevity. We performed multiple regression analysis of multiple traits and natural selection efficiency. We measured natural selection efficiency as θ_n/θ_{z+i}.

Conclusions: Aves nuclear neutral diversity, represented by the mutation parameter, θ_w, was significantly negatively correlated with generation time. The other variables – metabolic rate, longevity, and body mass – did not correlate with nuclear neutral variation.

Keywords: birds, generation time, molecular evolution, mutation, nuclear genes.

INTRODUCTION

Genetic diversity in organisms is variable. Interpreting the degree of genetic diversity from diverse life and population histories can be difficult, as the mechanisms and processes that regulate that diversity are complex and poorly understood. Three main hypotheses have been advanced to explain variation in DNA substitution rate: generation time (Laird *et al.*, 1969; Wu and Li, 1985; Li, 1997), metabolic rate (Martin *et al.*, 1992; Martin and Palumbi, 1993; Nunn and Stanley, 1998), and longevity (Denham, 1957; Barja and Herrero, 2000). Researchers have tested these three hypotheses by correlating diversity with relevant life-history variables.

DNA substitution rate has been widely used for studies of genetic diversity. However, an accurate estimation of mutation rate is required for complicated molecular data, which

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

requires archaeological, fossil or complex computational support. Other difficult issues to be resolved in the study of population genetics are mutational saturation, or multiple substitutions, and the mutation parameters or degree of genetic diversity. The formula, $\theta = 4N_e\mu$, acts under the assumption that mutations are effectively neutral. Here, μ denotes the expected number of mutations for an individual DNA sequence per generation and N_e denotes the effective population size. Under the standard neutral model, higher mutation rates will lead to higher levels of polymorphism. No direct measure of N_e, μ or θ as a mutation parameter is available.

Molecular genetic data have greatly improved our ability to test hypotheses relating to the evolution of organisms. Most research has been based on the diversity of mitochondria, as a result of the higher mutation rate for mtDNA. However, nuclear gene data could provide more comprehensive information. Therefore, analysis of increasing nuclear gene polymorphisms would help us to better understand the mechanisms of molecular diversity.

Birds (class Aves) are an ideal group to study because they live longer than mammals of similar size and their phylogeny is well explored (Hackett *et al.*, 2008; Han *et al.*, 2011). In this study, we investigated the causes and correlates of Aves nuclear gene polymorphisms in 297 species of birds with diversity data, 72 species with generation time and body mass data, and 67 species with maximum longevity data. We had two objectives: (1) to determine whether the results of mitochondrial diversity studies are specific, and (2) to ascertain if there is a relationship between nuclear neutral diversity in birds (class Aves) and their life-history traits, including generation time, metabolic rate, longevity, and body mass.

METHODS

Altogether, data for 1232 polymorphisms of nuclear genes in birds were obtained from the Polymorphix database (Bazin *et al.*, 2005) and the Popset database in GenBank. Two sequences were considered not to be clustered if there was a mismatch of >50 nt with <80% similarity using the ClustalW program (Thompson *et al.*, 1994). Such a mismatch was interpreted as evidence that the sequences represented duplicate genes in distinct genomic contexts.

Sequences were visually inspected and corrected where required. Dubious sequences were manually removed. Some gene sequences under significantly selective constraint (including MHC, genes of *Gallus gallus*, and ASLV) were not included. After repeats and noisy data were removed, 752 groups of polymorphisms from 104 nuclear genes in 297 species of Aves belonging to 53 genera and 15 families remained. Each group was aligned by eye using ClustalW. All groups are available on request. Details of the Aves species and genes sampled are shown in Appendix S1 (evolutionary-ecology.com/2718Appendix.pdf).

Polymorphism sequence data analysis

Measures of genetic diversity, represented here by a mutation parameter, Watterson's estimator ($\theta_{\mathbf{w}}$), were calculated separately from the introns and synonymous sites of the analysed fragments of the same species, expressed at the per-site level of diversity:

$$\theta_{\mathbf{w}} = \frac{P}{L\sum_{i=1}^{n-1}\frac{1}{i}}. \quad (1)$$

We assumed that there was a sample of n haploid individuals from the population of interest, there were an infinite number of possible alleles, and that $n \ll N_e$. P was the number of synonymous polymorphisms, L the number of synonymous sites, and n the number of sequences sampled.

We compared the efficiency of natural selection with generation time to ensure that there was a significant correlation to remove the effects of N_e, as species with long generation times tend to have small values of N_e and hence small θ-values. The efficiency of natural selection was calculated using the following formula:

$$\frac{\theta_n}{\theta_{s+i}} = \frac{\sum_n P_n / \sum_n L_n}{\left(\sum_s P_s + \sum_i P_i + 1\right) \Big/ \left(\sum_s L_s + \sum_i L_i\right)}, \quad (2)$$

where P_n, P_s, and P_i are the numbers of non-synonymous, synonymous, and intron polymorphisms; L_n, L_s, and L_i are the numbers of non-synonymous, synonymous, and intron sites for each gene in each species; θ_n represents the non-synonymous sites and θ_{s+i} represents the synonymous sites and introns.

Life-history data

Body mass, age of female sexual maturity, gestation duration, basal metabolic rate, and longevity data were obtained from the AnAge database (de Magalhaes *et al.*, 2005). Measuring generation time is not straightforward, as it depends on the age structure of the species (Charlesworth, 1994), for which data are lacking for most birds. We took either female sexual maturity or the sum of female sexual maturity and gestational duration as an approximation of generation time, with the two providing similar results.

Phylogenetic reconstructions

We created a phylogenetic hypothesis for the species included in this study by grafting them onto a higher-level phylogenetic supertree of Aves using PhyloWidget (Jordan and Piel, 2008). The topology was a composite of information drawn from Han *et al.* (2011), Hackett *et al.* (2008), and TreeBASE (http://www.treebase.org/treebase-web/home.html). Phylogenetically independent contrasts (PIC) were conducted using Phylogenetic Comparative Methods of COMPARE, version 4.6b (Martins, 2004).

Statistical analysis

The θ_w-values underwent arc-sine transformation (Sokal and Rohlf, 1981). Quantitative life-history variables were log-transformed. We calculated θ_s for synonymous sites and θ_i for intron sites and weighted the average of θ_s and/or θ_i from different genes for the same species. The effect of each life-history variable on the average of θ_s and/or θ_i was assessed using independent regression analysis. Multiple regression analyses were used to confirm the correlation of one or more life-history variables with the neutral mutation parameter (θ_{s+i}) because of the causal relationships between the three life-history variables.

RESULTS

Aves nuclear genetic diversity

We correlated polymorphisms in 104 nuclear genes from 297 avian species with life-history traits to test three hypotheses. Limited life-history data are available, especially for metabolic rate, which was only documented in 12 of 297 avian species. We therefore focused on the effects of generation time and maximum longevity. Body mass, however, could be used as a proxy of metabolic rate due to their causal relationship (Martin and Palumbi, 1993; Lanfear, 2007; Nabholz *et al.*, 2008a, 2008b).

To correlate life-history variables to all sites of neutral diversity of sampled nuclear genes, we weighted the average θ-value by combining intron sites with synonymous sites, and then performed non-phylogenetic regression (logistic regression) analyses and phylogenetic regression for independent comparisons. Both the application of phylogenetic independent comparisons and logistic regression analyses showed that the θ_w of weighted neutral sites was negatively associated with generation time ($n = 70$, $P = 0.004$ and $n = 70$, $P = 0.001$, respectively) (Table 1). No strong correlation was found between the θ_w of weighted neutral sites and either body mass or maximum longevity (Table 1).

As life-history variables are correlated with each other, multiple regression analysis is required. The results of the multiple regression (Table 2) showed that three life-history variables covaried significantly; generation time was the only life-history variable to correlate significantly with θ_w. These analyses showed that generation time was a major determinant of nuclear neutral mutations in the bird species studied here.

Correlation of natural selection efficiency with generation time

We observed that θ_w was significantly and negatively correlated with generation time, but it is possible that different N_e values would result in a qualitative correlation. To test whether the correlation between θ_w and generation time was due to a correlation between N_e and generation time, we examined the efficiency of natural selection because it is also expected to correlate with N_e. We thus predicted that if the correlation between θ_w and generation time was due to a correlation between N_e and generation time, there should also be a correlation between the efficiency of natural selection and generation time. A single variable regression analysis of natural selection efficiency is shown in Table 3. As no correlation was observed between generation time and the efficiency of natural selection ($P = 0.962$), we conclude that there is a significant negative correlation between the nuclear diversity of Aves species and generation time.

Table 1. Single variable regression analyses of weighted neutral sites

			Non-phylogenetic			Phylogenetic regression of independent contrasts		
θ_w	Trait	N	Slope	R^2	P-value	Slope	R^2	P-value
	Generation time	70	–0.01	0.112	0.004**	–0.017	0.153	0.001**
	Body mass	70	–0.002	0.031	0.144	–0.003	0.02	0.175
	Longevity	68	–0.003	0.011	0.396	0.001	0.00	0.912

Table 2. Multiple regression analysis of weighted neutral sites

		θ_w	Generation time	Body mass	Longevity
θ_w	Pearson correlation		– 0.375**	– 0.175	0.104
	P-value		0.001	0.144	0.396
	N		70	70	68
Generation time	Pearson correlation	– 0.375**		0.406**	0.406**
	P-value	0.001		0.001	0.002
	N	70		62	59
Body mass	Pearson correlation	– 0.175	0.406**		0.594**
	P-value	0.144	0.001		0.000
	N	70	62		67
Longevity	Pearson correlation	– 0.104	0.406**	0.594**	
	P-value	0.396	0.002	0.000	
	N	68	59	67	

** Correlation significant at the 0.01 level (two-tailed).

Table 3. Single variable regression analysis of natural selection efficiency

θ_n/θ_{s+i}	Trait	*N*	Slope	R^2	*P*-value
	Generation time	70	0.000	0.000	0.962

DISCUSSION

There are two principal types of genetic diversity, adaptive and neutral, and in this study we examined what factors affect neutral genetic diversity. Neutral diversity has been shown to depend primarily on N_e and mutation rate. Some studies on mutation rate, however, have cast doubt on the effects of N_e on genetic diversity. However, mutation rate, which was shown by Nabholz *et al.* (2009) to positively affect the level of gene polymorphisms and substitution rates in birds and mammals, was expected to be a major determinant of within-species genetic diversity, whether or not populations are in mutation–drift equilibrium (Nei and Graur, 1984; Iizuka *et al.*, 2002). In this study, examination of neutral genetic diversity was based on gene polymorphisms.

Support for the role of generation time in genetic diversity

As weighted sites are presumed to be neutral, we considered life-history variables relevant to nuclear neutral diversity, as represented by a mutation parameter, θ_w. Both logistic regression and phylogenetic regression analysis of independent comparisons showed that the neutral nuclear diversity appeared to be negatively correlated with generation time. This negative correlation was affirmed by further multiple regression analysis of multiple traits. No other variables, including body mass and maximum longevity, correlated with nuclear diversity before or after multiple regression.

Recently, it was reported that genetic variability in mtDNA was not correlated with population size (Bazin *et al.*, 2006, Nabholz *et al.*, 2009). The present results, based on the neutral nuclear diversity of 297 avian species, also show no significant correlation with population size. Although further research is required to support the conclusion that avian nuclear neutral diversity is not due to the effective population size, our observations suggest that avian nuclear neutral diversity is, at least in part, due to the extent of mutational input, which corroborates the conclusions of Nabholz *et al*. (2009) based on data from mtDNA.

Germline mosaicism (Woodruff *et al.*, 1996), where a fraction of nuclear mutations appear in just one or two meiotic divisions, is certain to strengthen the correlation between generation time and mutation rate per year. Where the replication of mtDNA is potentially decoupled from cell division (Ballard and Whitlock, 2004), such a dependency is weakened. We suggest that generation time might be the major cause of 'molecular clock' variation that is observed from one generation to the next.

Two other variables

The metabolic rate hypothesis, which is based on mtDNA diversity, proposes that the production of mutagenic free radicals, reactive oxygen species (ROS), increases with increasing rates of respiration; therefore, so does the rate of mutation. In contrast to the diversity of mitochondrial genes, the replication of nuclear genes might show a more pronounced generation time effect, as nuclear genes are more closely linked to cell division, whereas mitochondria can divide many times during the lifetime of a cell. The results of this study show that there was no significant correlation between the θ_w of Aves weighted neutral sites and metabolic rate, using body mass as a proxy. We inferred that the effects of metabolic rate on mtDNA diversity are specific.

The longevity hypothesis posits that long-lived organisms have evolved decreased rates of mtDNA mutation as a means to reduce the deleterious effects of somatic mutations that accumulate during their lifetime. After a comparison of rates of mtDNA evolution in birds and mammals, Nabholz *et al.* (2009) favoured maximum longevity as the main determinant of avian mtDNA mutation rates. Based on an estimate from the nine longest mitochondrial genes, Welch *et al.* (2008) found that mammalian mtDNA synonymous mutation rates were negatively correlated with maximum longevity ($n = 36$). Our results, for a larger set of species, were based on nuclear diversity in Aves. We did not find a significant correlation between neutral nuclear diversity and longevity after multiple regression analyses, indicating that generation time affected nuclear neutral diversity most strongly.

CONCLUSION

In this study, we performed a comprehensive overview of the nuclear genetic diversity and its polymorphic variations in the class Aves. The neutral nuclear diversity of 104 nuclear genes of 297 bird species was significantly correlated with generation time. The longevity hypothesis of mitochondrial diversity does not fit the patterns of neutral nuclear diversity in Aves, and the neutral nuclear diversity of Aves was unrelated to body size, longevity or metabolic rate.

ACKNOWLEDGEMENTS

We thank Prof. Adam Eyre-Walker for his kind support and advice. T.Z. thanks the School of Life Sciences, University of Sussex for hosting him as a visiting researcher. This work was supported by a grant from the Priority Academic Program Development of Jiangsu Higher Education Institutions.

REFERENCES

Ballard, J.W.O. and Whitlock, M.C. 2004. The incomplete natural history of mitochondria. *Mol. Ecol.*, **13**: 729–744.

Barja, G. and Herrero, A. 2000. Oxidative damage to mitochondrial DNA is inversely related to maximum life span in the heart and brain of mammals. *FASEB J.*, **14**: 312–318.

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

Bazin, E., Glemin, S. and Galtier, N. 2006. Population size does not influence mitochondrial genetic diversity in animals. *Science*, **312**: 570–572.

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 Res.*, **33**: D537–D543.

Denham, H. 1957. Prolongation of the normal life span by radiation protection chemicals. *J. Gerontol.*, **12**: 257–263

Hackett, S., Kimball, R., Reddy, S., Bowie, R., Braun, E., Braun, M. *et al.* 2008. A phylogenomic study of birds reveals their evolutionary history. *Science*, **20**(588): 1763–1768.

Han, K., Braun, E., Kimball, R., Reddy, S., Bowie, R., Braun, M. *et al.* 2011. Are transposable element insertions homoplasy free? An examination using the avian tree of life. *Syst. Biol.*, **60**: 375–386.

Iizuka, M., Tachida, H. and Matsuda, H. 2002. A neutral model with fluctuating population size and its effective size. *Genetics*, **161**: 381–388.

Jordan, G. and Piel, W. 2008. PhyloWidget: web based visualizations for the tree of life. *Bioinformatics*, **24**: 1641–1642.

Laird, C.D., McConaughty, B.L. and McCarthy, B.J. 1969. Rate of fixation of nucleotide substitutions in evolution. *Nature*, **326**: 93–96.

Lanfear, R., Thomas, J.A., Welch, J.J. and Bromham, L. 2007. Metabolic rate does not calibrate the molecular clock. *Proc. Natl. Acad. Sci. USA*, **104**: 15388–15393.

Li, W.-H. 1997. *Molecular Evolution*. Sunderland, MA: Sinauer Associates.

Martin, A.P. and Palumbi, S. 1993. Body size, metabolic-rate, generation time, and the molecular clock. *Proc. Natl. Acad. Sci. USA*, **90**: 4087–4091.

Martin, A.P., Naylor, G. and Palumbi, S. 1992. Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. *Nature*, **357**: 153–155.

Martins, E.P. 2004. *COMPARE*, version 4.6b. Computer programs for the statistical analysis of comparative data. Available at: http://compare.bio.indiana.edu/.

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

Nabholz, B., Mauffrey, J.F., Bazin, E., Galtier, N. and Glémin, S. 2008b. Determination of mitochondrial genetic diversity in mammals. *Genetics*, **178**: 351–361.

Nabholz, B., Glemin, S. and Galtier, N. 2009. The erratic mitochondrial clock: variation of mutation rate, not population size, affects mtDNA diversity across birds and mammals. *BMC Evol. Biol.*, **9**: 54.

Nei, M. and Graur, D. 1984. Extent of protein polymorphism and the neutral mutation theory. *Evol. Biol.*, **17**: 73–118.

Nunn, G.B. and Stanley S. 1998. Body size effects and rates of cytochrome b evolution in tube-nosed seabirds. *Mol. Biol. Evol.*, **15**: 1360–1371.

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

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 Res.*, **22**: 4673–4680.

Welch, J.J., Bininda-Emonds, O.R.P. and Bromham, L. 2008. Correlates of substitution rate variation in mammalian protein-coding sequences. *BMC Evol. Biol.*, **8**: 53.

Woodruff, R.C., Hual, H. and Thompson, J.N. 1996. Clusters of identical new mutation in the evolutionary landscape. *Genetica*, **98**: 149–160.

Wu, C.-I. and Li, W.H. 1985. Evidence for higher rates of nucleotide substitution in rodents than man. *Proc. Natl. Acad. Sci. USA*, **82**: 1741–1745.