OXFORD UNIVERSITY PRESS | Genome Biology and Evolution

Reduced effective population size due to sex chromosome differentiation does not drive mitochondrial genome size expansion in paleognathous birds.

Journal:	*Genome Biology and Evolution*
Manuscript ID	GBE-230543
Manuscript Type:	Letter
Date Submitted by the Author:	31-May-2023
Complete List of Authors:	Weinstein, Brooke; UC Merced, Quantitative and Systems Biology Wang, Zongji; Zhejiang University, Life Science Institute Zhou, Qi; Zhejiang University, Life Science Institute Roy, Scott; San Francisco State University, Biology
Keywords:	drift barrier hypothesis, ZW sex determination, evolution of complexity

Main Manuscript for

Reduced effective population size due to sex chromosome differentiation does not drive mitochondrial genome size expansion in paleognathous birds.

Brooke Weinstein[1], Zongji Wang[2] and Qi Zhou[2,3]* and Scott William Roy[1,4]*

[1]Department of Molecular and Cell Biology, University of California-Merced, Merced, CA 95343

[2]MOE Laboratory of Biosystems Homeostasis and Protection and Zhejiang Provincial Key Laboratory for Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, Hangzhou, China, 310058

[3] Evolutionary & Organismal Biology Research Center, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, China, 310058

[4]Department of Biology, San Francisco State University, San Francisco, CA 94117

* To whom correspondence should be addressed

Corresponding author:

Scott Roy

1600 Holloway Ave, San Francisco, CA 94132

415-602-4933

scottwroy@gmail.com

Keywords: drift barrier hypothesis, ZW sex determination, evolution of complexity

Abstract

Eukaryotic genomes can differ by several orders of magnitude in size. The drift barrier hypothesis is the most prevalent explanation, positing that differences in genome size and complexity reflect changes in the efficacy of selection, with species with small effective population sizes experiencing inefficient selection and accumulating putatively deleterious insertions. However, the challenges of estimating effective population size have led to a need for complementary hypothesis tests. Aside from demographic and ecological factors, genetic changes can also affect effective population size: due to background selection on linked sites, more linkage leads to reduced effective population size. In species with ZW sex determination, the mitochondrial genome and the W chromosome are exclusively maternally transmitted and thus completely genetically linked. In birds, the non-recombining W-linked region of the ancestral sex chromosomes expanded independently in several lineages. We compared the evolution of protein-coding sequences in the mitochondrial genome across 15 species of flightless paleognathous birds in which the differentiated ZW regions correspond to 30% to 99% of the chromosome. We found an association between the size of the ZW differentiated region and the normalized fixation rate of nonsynonymous changes, as expected for reduced effective population size. However, we found no correlation between mitochondrial genome size and expanded ZW system, nor between mitochondrial genome size and rate of nonsynonymous change. These results suggest that differences in selection efficiency do not significantly contribute to differences in mitochondrial genome size in Palaeognathae and raise the question of the general predictive power of the drift barrier hypothesis.

Significance

The cause(s) of the striking variation in eukaryotic genome size remains controversial. The most prevalent idea is the drift barrier hypothesis, which states that mutations that lead to an increase in genome size are often at least slightly deleterious and, thus, preferentially accumulate in organisms with small effective population sizes because natural selection is too inefficient to remove them. We tested the relationship utilizing an expanded ZW chromosomal system in 15 species of paleognathous birds and, contrary to prediction, found no evidence that reduced selective efficiency leads to larger mitochondrial genomes.

Introduction

Both mitochondrial and nuclear genomes show a remarkable range of content and structure across eukaryotes, with substantial differences in the number of genes, introns, gene copy numbers, and intergenic DNA (Smith and Keeling 2015). However, the general evolutionary forces governing this genome variation remain obscure (Blommaert 2020). The leading explanation is the drift barrier hypothesis (DBH), which posits that genome size and complexity differences reflect differences in the efficacy of selection, with species with small effective population sizes (N_e) experiencing inefficient selection and thus accumulating deleterious insertions (Lynch & Conery 2003, Lynch & Walsh 2007). However, despite 20 years of work on the issue, direct tests of the DBH have proven elusive, mainly because correlated evolutionary changes may confound the results and because appropriate measures and definitions of N_e remain controversial.

N_e represents the number of individuals in an ideal population that theoretically experiences genetic drift at the same rate as the actual population, accounting for real-life

ecological, demographic, and genomic complexities (Charlesworth 2009). Since it is difficult or impossible to measure N_e in natural populations directly, most studies use life history or genetic characteristics as proxies. In addition to demographic factors, the N_e of a genomic locus is influenced by the number of sites to which the locus is linked because deleterious mutations at linked sites tend to remove chromosomes from the population and thus effectively decrease population size. This is most clearly seen in differentiated sex chromosomes, wherein newly sex-linked non-recombining regions rapidly accumulate deleterious mutations and lose genes (Charlesworth and Charlesworth 2000). Therefore, comparing genome evolution between related lineages with different degrees of linkage provides a promising means for testing the DBH.

As in mammals, where nearly all mammals share the same XY chromosome pair, nearly all birds have a sex-linked region within the same pair of ZW sex chromosomes. The extent of the non-recombining (and thus completely sex-linked) region on these chromosomes differs but in most birds and all mammals, the vast majority of the sex chromosome pair is differentiated. However, paleognaths, the earliest differentiated branch of birds, represent an exception. Paleognaths include flightless ratites and semi-flighted tinamous, diverging from other birds over 110 million years ago (Jarvis et al. 2014). In many paleognathous species, the Z and W chromosomes are homomorphic, meaning that most of the ZW pair continue recombining in ZW females. In multiple paleognath lineages, however, the non-recombining sex-linked portion of the chromosome has independently expanded (Wang et al. 2022).

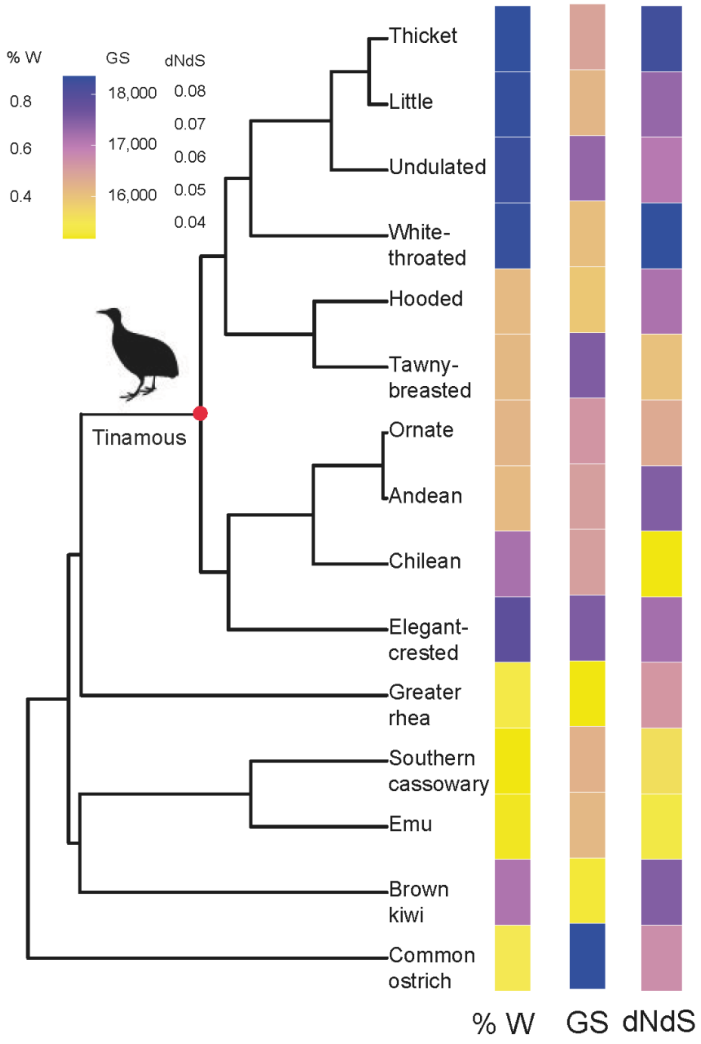
The power of this system for studying the DBH arises from a quirk of inheritance in ZW systems. Like Y chromosomes in mammals, differentiated regions of W chromosomes in birds are hemizygous and do not recombine (Charlesworth and Charlesworth 2000). Since the sex-specific (non-recombining) portion of the W chromosome is strictly maternally inherited, it is

expected to have complete genetic linkage with the entire mitochondrial genome (which is also maternally inherited), with rare or no recombination within either genomic element or reassortment between the two. This increased linkage predicts increased influences of background selection and hitchhiking (negative and positive selection on linked sites), leading to a decrease in N_e for the mitochondrial genome. The relationship between increased linkage and decreased N_e is supported by reduced nucleotide diversity in mitochondrial genes from ZW species (Berlin et al. 2007, Wang et al. 2022). The parallel ZW differentiation across diverse paleognathous lineages with available genomic data thus affords a rare opportunity to test the hypothesis that reduced selective efficiency leads to larger genomes.

Results

We obtained mitochondrial genomes and the percentage of the sex chromosome pair that is differentiated, as well as a maximum-likelihood chronogram estimated from the concatenated alignments of whole-genome non-coding sequences, for 15 species of paleognathous birds, all previously generated by Wang et al. 2022. Across the 15 paleognath species, the differentiated ZW regions vary widely, corresponding to between 30% and 99% of the chromosome. We first looked at mitochondrial genes by creating a concatenated alignment of the 13 protein-coding sequences across the 15 genomes. We found that mitochondrial genomic content is consistent, with 13 protein-coding genes (CDS) in each species, but the size of the protein-coding region ranges by 9.3% between species (10.0 to 11.0 kilobases). However, the size of the non-coding regions is more variable, ranging by 51% from 4.8 to 8.0 kb. In total, mitochondrial genome size ranges by 18.6%, from 15.8 to 18.9 kb. We then reconstructed branch-specific ratios of nonsynonymous changes (d_N; subject to selection) to synonymous changes (d_S; presumed

93 neutral) across the tree, which is a measure used to estimate the efficacy of natural selection and
94 is expected to scale negatively with N_e (Ohta 1995) **(Figure 1)**.
95
96
97
98
99
100



101

102

103

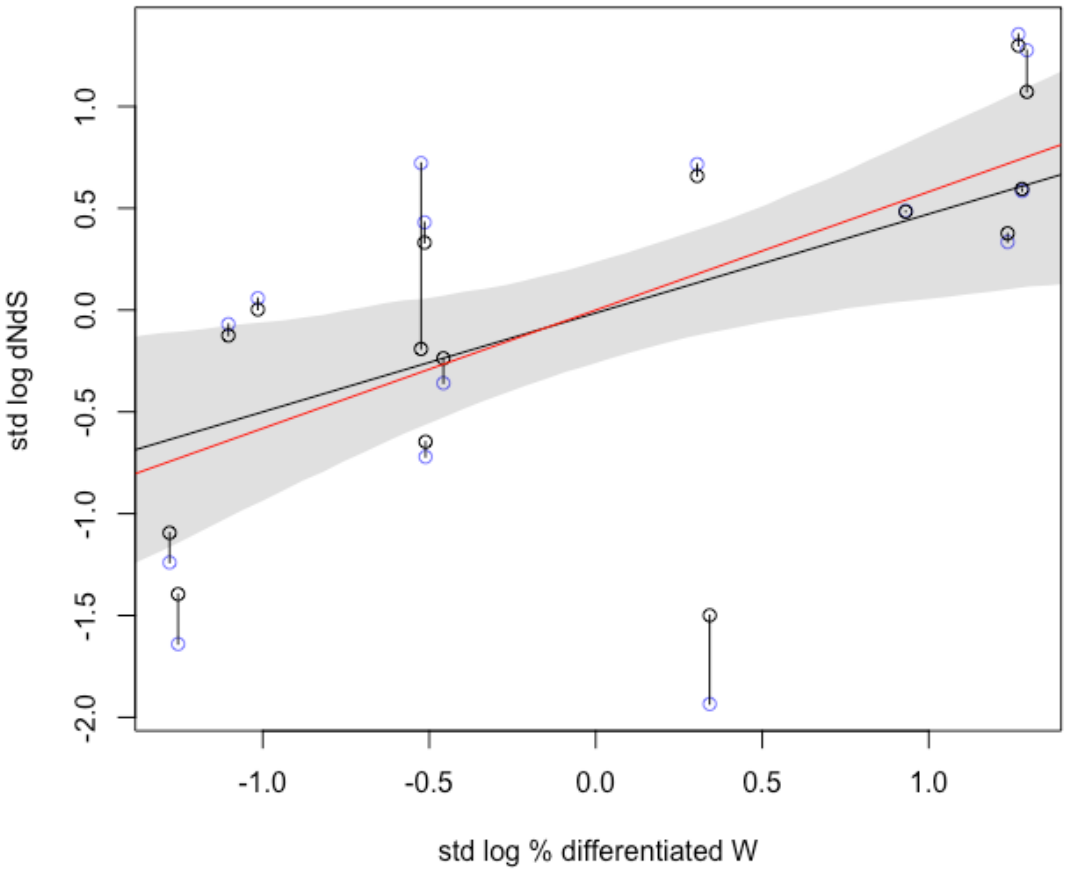
104

105

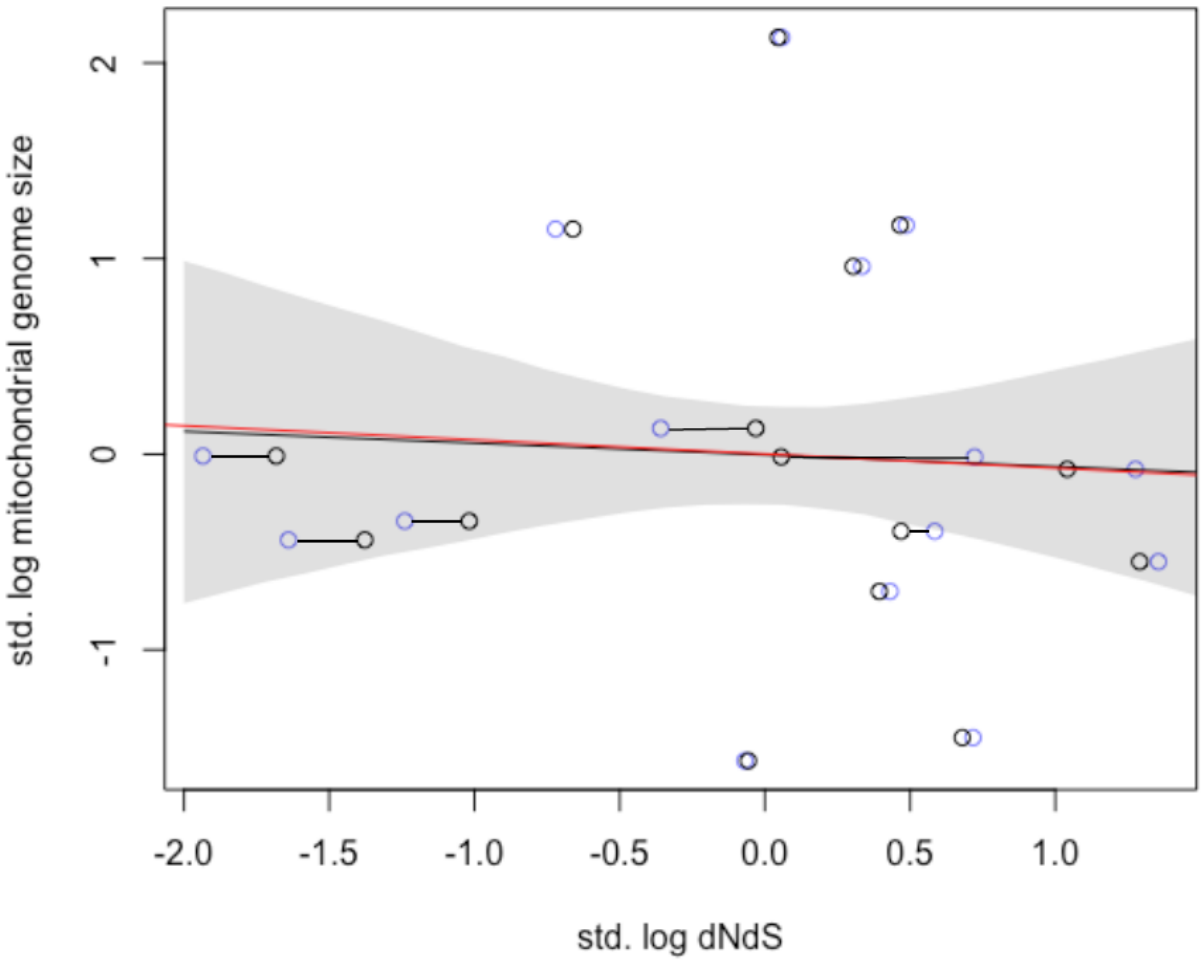
106

107

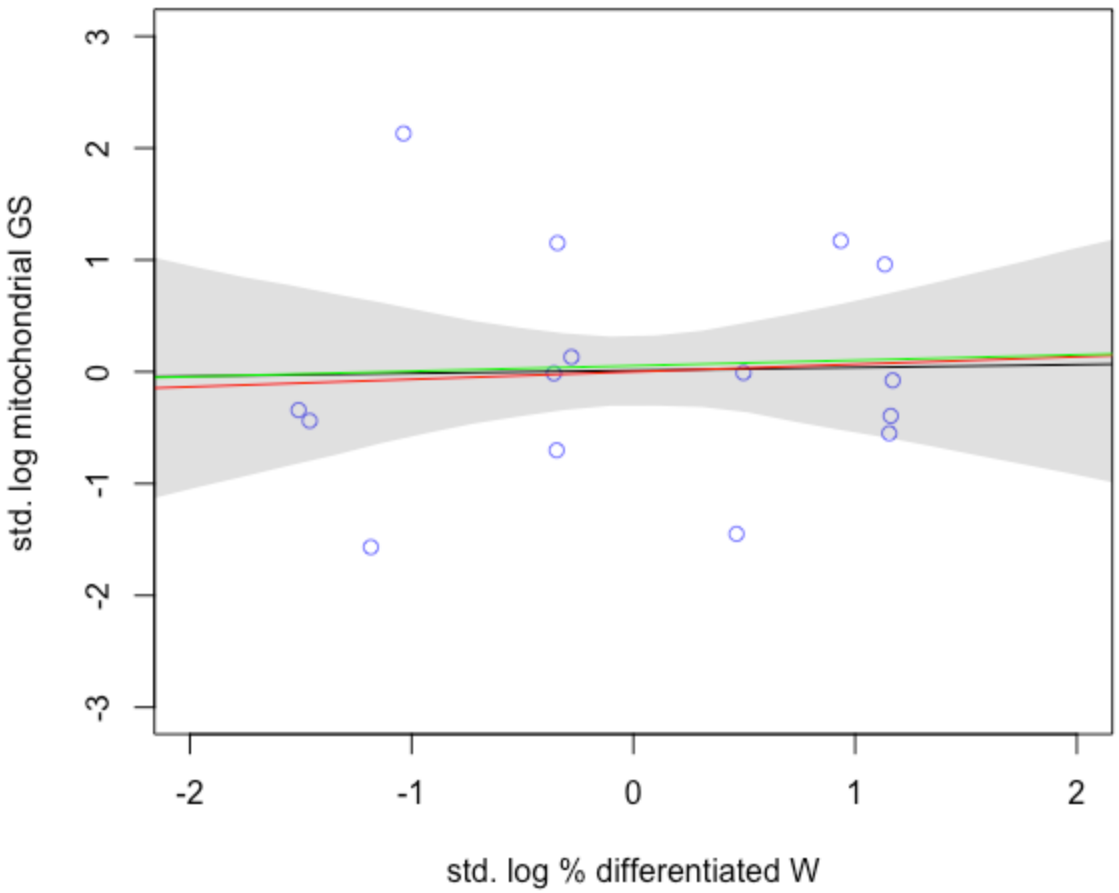
First, we tested whether increased sex chromosome differentiation leads to increased fixation of deleterious nonsynonymous variants. The genome-wide $d_N d_S$ ratio is appealing as a molecular proxy for studying N_e because it theoretically normalizes for mutation rate since synonymous mutations are often assumed to be selectively neutral. Furthermore, since $d_N d_S$ is a point estimate of a rate calculated along the phylogenetic branches, it avoids the statistical issue of phylogenetic nonindependence. Bayesian linear regression recovers a reliably positive correlation between ZW differentiation and $d_N d_S$ of mitochondrial genes (mean standardized effect = 0.48, CI = 0.15 - 0.80), although slightly less than estimated with OLS regression (0.58, p = .023) (**Figure 2**). We expect this result from Ohta's nearly neutral theory of molecular evolution (1992) because natural selection is more robust in larger populations when genetic drift is weaker, resulting in lower $d_N d_S$ ratios. Therefore, as predicted, we find a strong association between the size of the ZW differentiated region and the mutation-normalized probability of fixation of nonsynonymous changes.


std log dNdS
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-1.0
-0.5
0.0
0.5
1.0
std log % differentiated W

We next sought to determine whether reduced N_e (proxied by the $d_N d_S$ ratio and % differentiated W chromosomal sequence) is associated with changes in mitochondrial genome size. We found no correlation between mitochondrial genome size and our genetic proxy for N_e, $d_N d_S$, using Bayesian linear regression (mean standardized effect = -0.06, CI = -0.49 - 0.36) or OLS (-0.07, p = 0.80; **Figure 3).**


std. log mitochondrial genome size
2
1
0
-1
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
std. log dNdS

Looking at the relationship between % W chromosome differentiation and mitochondrial genome size, Bayesian linear regression does not recover any reliable correlation (mean standardized effect = 0.05, CI = -0.33 - 0.44). However, in this case, phylogenetic non-independence due to shared ancestry creates a statistical issue that should be accounted for because any relationship between the two traits could simply result from shared phylogenetic history (Felsenstein 1985). We inferred that % W chromosome differentiation shows a significant, strong phylogenetic signal corresponding to a Brownian motion evolution model (Pagel's lambda = 1.00, p-value = .0001). Mitochondrial genome size also has the same substantial phylogenetic signal, but the hypothesis test is insignificant (Pagel's lambda = 1.00, p-value = 0.17). Therefore, we used a Bayesian mixed model to include information about phylogenetic relationships. For comparison, but without the uncertainty in $d_N d_S$, we also fit a maximum likelihood linear phylogenetic model with generalized least squares (PGLS). We found there is still no effect with Bayesian phylogenetic correction (mean standardized effect = 0.02, CI = -0.44 - 0.49), using PGLS (0.04, p = 0.93), nor with OLS (0.06, p = 0.84) **(Figure 4)**. Our analysis suggests that N_e, as proxied by increased genetic linkage and the rate of molecular evolution, has no impact on mitochondrial genome size in paleognathous birds, contrary to the predictions of the DBH.



171

172

173

174

175

176

177

178

179

Discussion

These data represent a rare direct test of the predictions of the drift barrier hypothesis, under which unnecessary genomic elements accumulate when selection is inefficient. Here, we find no evidence for mitochondrial genome expansion under a confirmed case of reduced N_e due to increased transmission linkage to extensive non-recombining regions, incorporating confidence intervals and using multiple methods to help address our small sample size. Notably, it is impossible to confidently infer causality between ZW chromosomal differentiation and reduced mitochondrial N_e. Possible explanations are reduced mitochondrial N_e due to increased linkage or ZW differentiation due to reduced N_e. However, the correlation strongly suggests that for whatever reason, mitochondrial genomes in birds with differentiated sex chromosomes experienced reduced N_e.

A lack of association between N_e and mitochondrial genome size is less surprising given birds' general lack of mitochondrial genome size variation. However, the objection that a lineage simply does not have much variation in mitochondrial genome size despite variation in N_e itself rejects the drift barrier hypothesis in favor of some alternative hypothesis, for instance, a cryptic lineage-specific tendency towards small genome size, such as in most birds from the energetic challenges of flight (Wright et al. 2014). In this study, all the birds included are semi-flighted or flightless, which we also expect imposes constraints due to the high metabolic cost of running (Bundle et al. 1999). Put another way, to contend that a hypothesis has explanatory power but only under a limited subset of circumstances chosen post hoc is to concede that the hypothesis has limited explanatory power at best.

This study represents a rare opportunity to study the relationship between N_e and organelle genome size through a controlled case in which variation in N_e is due to genetic

changes in a single genomic region rather than global changes in demography and in which standard measures of selective efficiency confirm the expected changes in N_e. Our failure to find the expected association suggests limits on the explanatory power of the drift barrier hypothesis and points to the urgency of further controlled tests.

Materials and Methods

We estimated the genome-wide $d_N d_S$ ratio using HyPhy v. 2.5.36 (Pond and Muse 2005) with an MG94 model of codon evolution plus a GTR model of nucleotide substitution, including confidence intervals. A key component of HyPhy methods is that d_S can vary across branches, and we chose the MG94 model because it explicitly models both synonymous and nonsynonymous site variability. Additionally, the FitMG94 workflow uses a corrected empirical estimator (CF3x4) that provides improved estimates of several parameters in the evolutionary model by accounting for biases in nucleotide composition induced by stop codons (Goldman and Yang 1994). Before proceeding, to ease prior specification plus linear model fit and computation, all three variables were logged and then rescaled to have a mean of zero and a standard deviation of one.

All Bayesian regressions were performed using rstan (Stan Development Team 2023). For the phylogenetic mixed model, we used ape (Paradis et al. 2004) to calculate the distance matrix and a covariance matrix linearly related to the phylogenetic distance between the species (Brownian motion). Priors were checked via prior predictive simulation and weakly regularized to penalize extreme parameter values: intercept ~ Normal (0, 0.2), slope ~ Normal(0,0.5), and standard deviation ~ Exponential(1). We accommodated the inferred measurement error in each observed HyPhy value (d_N/d_{SOBS}, i) by adding the additional parameter (d_N/d_{STRUE}, i): d_N/d_{SOBS}, i

~ Normal((d_N/d_{STRUE}, i), (d_N/d_{SSD}, i)). Maximum-likelihood PGLS regressions were implemented using nlme (Pinheiro et al. 2017).

Acknowledgements

BNW and SWR were supported by the National Science Foundation [grant number 1616878].

Data Availability

The data underlying this article are available in its online supplementary material.

Literature Cited

Berlin, S., Tomaras, D., & Charlesworth, B. 2007. Low mitochondrial variability in birds may indicate Hill–Robertson effects on the W chromosome. Heredity, 99(4):389-396.

Blommaert, J. 2020. Genome size evolution: towards new model systems for old questions. Proc. Royal Soc. B, 287(1933): 20201441.

Bundle, M. W., Hoppeler, H., Vock, R., Tester, J. M., & Weyand, P. G. 1999. High metabolic rates in running birds. Nature, 397(6714):31-32.

Charlesworth, B. 2009. Effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet., 10(3), 195-205.

Charlesworth, B., & Charlesworth, D. 2000. The degeneration of Y chromosomes. Philos. Trans. R. Soc. of Lond. B, Biol. Sci., 355(1403):1563-1572.

Felsenstein, J. 1985. Phylogenies and the comparative method. Am. Nat., 125(1), 1-15.

Goldman, N., & Yang, Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol., 11(5):725-736.

Jarvis, E. D. et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science, 346(6215):1320-1331.

Lynch, M., & Conery, J. S. 2003. The origins of genome complexity. Science, 302(5649):1401-1404.

Lynch, M., & Walsh, B. 2007. *The origins of genome architecture* (Vol. 98). Sunderland, MA: Sinauer Associates.

Ohta, T. 1992. The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst., p. 263-286.

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

Paradis, E., Claude, J., & Strimmer, K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics, 20(2):289-290.

Pinheiro, J. et al. 2017. Package 'nlme'. Linear and nonlinear mixed effects models, version, 3(1).

Pond, S. L. K., & Muse, S. V. 2005. HyPhy: hypothesis testing using phylogenies. In: *Statistical methods in molecular evolution*. Springer, New York, NY. p. 125-181

Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol., 2:217-223.

Smith, D. R., & Keeling, P. J. 2015. Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. Proc. Natl. Acad. Sci., 112(33):10177-10184.

Stan Development Team 2023. "RStan: the R interface to Stan." R package version 2.21.8, https://mc-stan.org/.

Wang, Z. et al. 2022. Phylogeny and sex chromosome evolution of Palaeognathae. J. Genet. Genomics, 49(2):109-119.

Whitney, K. D., & Garland Jr, T. 2010. Did genetic drift drive increases in genome complexity?. PLoS Genetics, 6(8):e1001080.

Wright, N. A., Gregory, T. R., & Witt, C. C. 2014. Metabolic 'engines' of flight drive genome size reduction in birds. Proc. Royal Soc. B: Biol. Sci., 281(1779):20132780.

Fig. 1. Paleognathae phylogeny and raw mitochondrial trait data. Chronogram of the 15 paleognathous birds included in the analysis with heat maps showing the raw trait data. GS is the mitochondrial genome size, and the red dot denotes the tinamou order/family, of which 2/3^{rds} of the species are members.

Fig. 2. Strong positive correlation between the amount of genetic linkage and the rate of molecular evolution. 89% confidence interval and mean (black line) of the estimated effect of % differentiated W chromosome on $d_N d_S$ using Bayesian regression in 15 species of paleognathous birds. Blue points show standardized d_N/d_S estimates from HyPhy, whereas black points incorporate confidence intervals from HyPhy. The connecting vertical lines represent shrinkage in the posterior distribution resulting from measurement uncertainty in d_N/d_S. The red line is the mean estimated effect from OLS regression.

Fig. 3. No reliable correlation between the rate of molecular evolution and mitochondrial genome size. 89% confidence interval and mean estimated effect of mitochondrial genome size (GS) as a function of $d_N d_S$ in 15 species of paleognathous birds using a Bayesian phylogenetic mixed model. Blue points show standardized $d_N d_S$ estimates from HyPhy, whereas black points incorporate confidence intervals. The red line is the mean estimated effect from OLS regression, and the green line is from PGLS regression.

Fig. 4. No reliable correlation between the amount of genetic linkage and mitochondrial genome size. 89% confidence interval and mean estimated effect of mitochondrial genome size (GS) as a function of Z/W differentiation in 15 species of paleognathous birds using a Bayesian

317 phylogenetic mixed model. The red line is the mean estimated effect from OLS regression, and
318 the green line is from PGLS regression.
319
320
321
322
323