

RESEARCH

Open Access

# $K_r/K_c$ but not $d_N/d_S$ correlates positively with body mass in birds, raising implications for inferring lineage-specific selection

Claudia C Weber<sup>1</sup>, Benoit Nabholz<sup>2</sup>, Jonathan Romiguier<sup>2,3</sup> and Hans Ellegren<sup>1\*</sup>

## Abstract

**Background:** The ratio of the rates of non-synonymous and synonymous substitution ( $d_N/d_S$ ) is commonly used to estimate selection in coding sequences. It is often suggested that, all else being equal,  $d_N/d_S$  should be lower in populations with large effective size ( $N_e$ ) due to increased efficacy of purifying selection. As  $N_e$  is difficult to measure directly, life history traits such as body mass, which is typically negatively associated with population size, have commonly been used as proxies in empirical tests of this hypothesis. However, evidence of whether the expected positive correlation between body mass and  $d_N/d_S$  is consistently observed is conflicting.

**Results:** Employing whole genome sequence data from 48 avian species, we assess the relationship between rates of molecular evolution and life history in birds. We find a negative correlation between  $d_N/d_S$  and body mass, contrary to nearly neutral expectation. This raises the question whether the correlation might be a method artefact. We therefore in turn consider non-stationary base composition, divergence time and saturation as possible explanations, but find no clear patterns. However, in striking contrast to  $d_N/d_S$ , the ratio of radical to conservative amino acid substitutions ( $K_r/K_c$ ) correlates positively with body mass.

**Conclusions:** Our results in principle accord with the notion that non-synonymous substitutions causing radical amino acid changes are more efficiently removed by selection in large populations, consistent with nearly neutral theory. These findings have implications for the use of  $d_N/d_S$  and suggest that caution is warranted when drawing conclusions about lineage-specific modes of protein evolution using this metric.

## Background

It has long been established that different lineages evolve at heterogeneous rates [1,2] and that differences in organismal life history are reflected by rates of molecular evolution. This is readily observed in terms of lineage-specific nucleotide divergence, with small-bodied species with shorter generations tending to evolve more quickly than their larger relatives [3-10]. While this has been proposed to be a consequence of the higher number of germ cell divisions per unit time [11], the precise cause for the pattern remains unclear [12].

Another side effect of body size variation between lineages manifests in population size differences, as small

species tend to have larger populations [13]. This in turn might impact the prevalence of both positive and negative selection in genome evolution relative to drift. Over long timescales, the distribution of weakly selected mutations is expected to be affected by population size, with fixation probability depending on the product of  $N_e$  and the selection coefficient,  $s$  [14]. Mutations with small selective effects close to the reciprocal of  $N_e$  will behave effectively neutrally [3,15-17]. Non-synonymous substitutions are on average under greater selective constraint than synonymous substitutions. As a consequence, nearly neutral theory predicts that  $d_N/d_S$  should be lower in large populations [18], as relatively more slightly deleterious non-synonymous changes are removed due to increased selection efficiency when  $N_e$  is high. Consistent with this idea, pathogenic bacteria and endosymbionts have accelerated rates of protein evolution compared to their free-living relatives, as might be expected given their smaller

\*Correspondence: hans.ellegren@ebc.uu.se

<sup>1</sup>Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18 D, Uppsala, Sweden

Full list of author information is available at the end of the article

population sizes [19,20]. Empirical studies on mammals employing body mass as a proxy for  $N_e$  in absence of actual census counts [21] present a similar pattern, with smaller-bodied species tending to have lower  $d_N/d_S$  ratios than their larger counterparts [22-27] (though not all authors report a significant relationship; see [9]).

Nevertheless, this trend may not be universal. It was previously reported that there is no significant relationship between  $d_N/d_S$  and life history in a data set containing 7.6 kb of coding sequence from 19 avian genes [28]. It is conceivable that this result was simply owing to insufficient statistical power, as the expected relationship between body mass and substitution rates was also not retrieved. However, more recently, Nabholz *et al.* [29] found that avian mitochondrial  $d_N/d_S$  was negatively correlated with body mass. One might expect that inherent differences between inheritance, mutation rate, recombination and gene density in the mitochondria and nucleus could lead to differences in the modulation of substitution patterns by selection [30]. However, in mammals, signatures of purifying selection are reported to be congruent between nuclear and mitochondrial genes [27,31]. A complementary approach to studying the effect of population size on purifying selection is to compare island to mainland species. In principle, island endemic lineages ought to maintain life histories similar to those of their mainland relatives, while experiencing a reduction in  $N_e$  [32] (but see [33]). However, here the evidence is similarly equivocal for birds with both increases and decreases in  $d_N/d_S$  being reported [34,35], possibly owing to lack of power [36]. Moreover, an increase in  $d_N/d_S$  is predicted with increasing environmental change [37], which is expected when a species colonises an island. It therefore does not follow that an increase in  $d_N/d_S$ , if at all present, need necessarily be owing to a reduction in population size.

Thus, whether or not large populations generally exhibit lower  $d_N/d_S$  as predicted by theory is not clear at present. To determine this, we need to study additional taxa in depth, considering possible confounding variables. One notable feature of avian genomes in particular is the strong variation in GC content between lineages associated with life history [38-40]. This may be of relevance, as estimation of  $d_N/d_S$  is known to be impacted by non-stationary base composition. A degree of caution may therefore be warranted when comparing genomes that differ substantially from one another in terms of base composition.

Making use of nuclear sequences from 48 fully sequenced bird genomes, which were recently generated to resolve the phylogeny of modern birds [40], we aim to characterise the relationship between life history,  $d_N/d_S$  and the efficacy of selection in birds. In doing so, we also examine to what extent method artefacts might influence our conclusions, examining in turn non-stationary

base composition, divergence time, saturation and how examining different classes of amino acid change in relation to population size may help answer these questions.

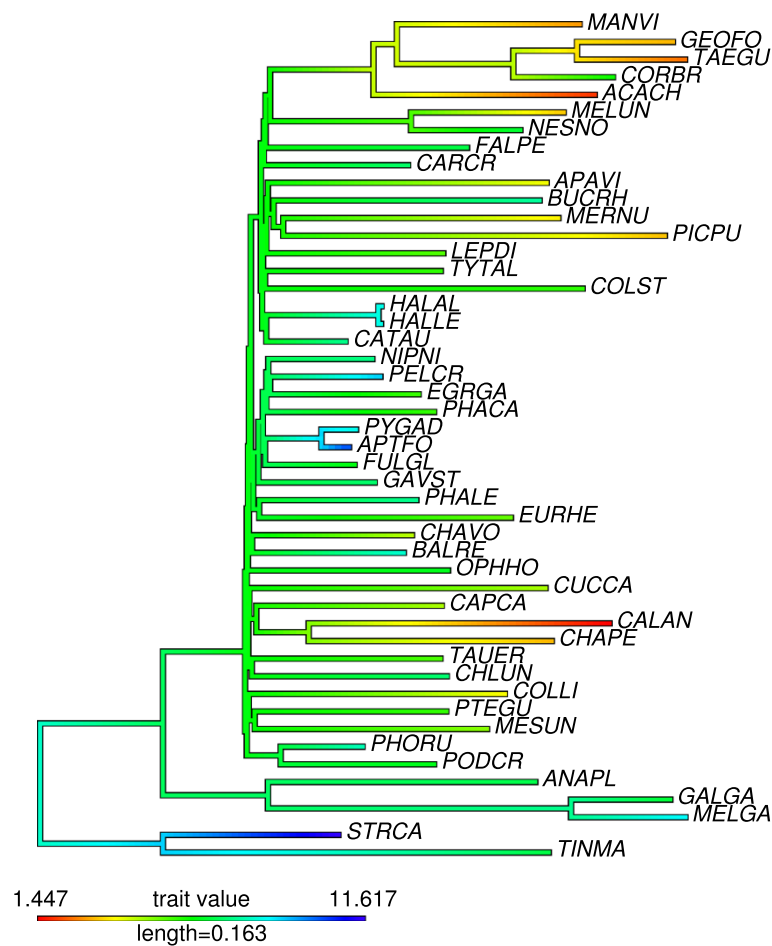
## Results

### $d_N/d_S$ is elevated, not reduced, in birds with putatively larger populations

To assess whether nuclear sequences from birds provide evidence that more efficient purifying selection in large effective populations removes a higher proportion of non-synonymous changes, we estimated lineage-specific  $d_N/d_S$  for 48 species by maximum likelihood, considering 921 out of 1,185 1:1 orthologues that did not contain internal stops. We used a consensus phylogenetic tree obtained from several types of phylogenomic analyses of these 48 genomes and focused on rates in terminal branches (Figure 1). One initial observation was that variation in  $d_N/d_S$  among lineages was relatively limited, in the range of 0.13 to 0.17. Our results appear to contradict the notion that more efficient protein-level selection in large populations is reflected by reduced  $d_N/d_S$ . Instead, body mass and  $d_N/d_S$  were significantly negatively correlated (Spearman's rank correlation:  $\rho = -0.4306$ ,  $P = 0.0027$ ; Figure 2). To ensure the robustness of this observation, we additionally considered a data set comprising 11 kb of coding sequence from 169 avian species [41]. A similar negative correlation was seen ( $\rho = -0.3807$ ,  $P = 3.3 \times 10^{-7}$ ; see Figure in Additional file 1). We hereafter refer to these data as the gene-rich and taxon-rich sets, respectively. Additionally, results from Coevol, which provides information on associations between traits and substitution patterns through evolutionary time using a Bayesian Monte Carlo framework rather than merely considering tip nodes [42], showed a similar negative correlation between  $d_N/d_S$  and body mass for the taxon-rich set ( $R = -0.302$ ,  $pp = 0.026$ ). This approach also corrects for phylogenetic non-independence between branches, suggesting that the result is not simply due to non-independence of the observations.

### $d_N$ and $d_S$ are higher in small-bodied birds

We next surveyed synonymous and non-synonymous substitution rates and their relationships with life history individually.  $d_S$  behaves as predicted if small birds with short generation times evolve more rapidly, correlating negatively with body mass in both data sets ( $\rho = -0.5208$ ,  $P = 0.0002$  for the gene-rich data set, Figure 3;  $\rho = -0.3015$ ,  $P = 6.8 \times 10^{-5}$  for the taxon-rich data set).  $d_N$  was similarly negatively correlated with body mass ( $\rho = -0.5147$ ,  $P = 0.0003$  for the gene-rich data set, Figure 3;  $\rho = -0.3814$ ,  $P = 3.1 \times 10^{-7}$  for the taxon-rich data set). This indicates that high  $d_N/d_S$  in species with putatively large populations is not due to the denominator of the ratio being smaller in absolute terms, though there



**Figure 1 Avian phylogeny.** Total evidence nucleotide tree from Jarvis *et al.* [40] displaying the 48 species considered in our study. Branches are coloured according to log (body mass) in grams. Full species names are given in Additional file 2.

must by definition be a reduction relative to  $d_N$ . Taken at face value, these findings would seem to suggest that selection is less rather than more efficient in birds with large population sizes. It is, however, possible that the negative relationship between  $d_N/d_S$  and body mass is a method artefact or is explained by another factor that covaries with life history.

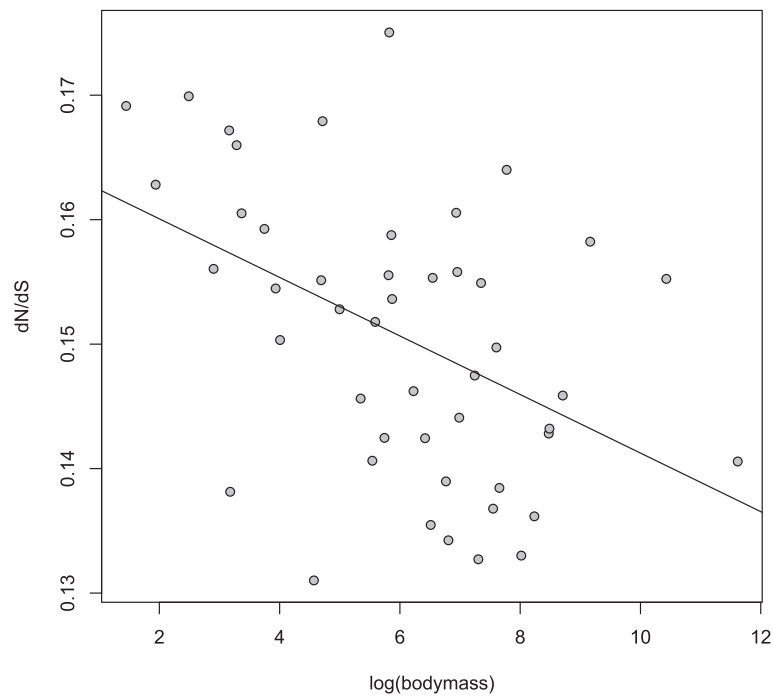
#### No evidence that non-stationary base composition accounts for elevated $d_N/d_S$

In addition to the above-mentioned correlations between substitution rates and life history traits, small birds have higher GC content than large species [38]. Non-stationary composition may lead to model misspecification if not accounted for, as the underlying models assume codon frequencies to be at equilibrium. This can impact estimates of divergence and lead to false conclusions [43,44]. Considering only orthologues with low variance in GC3 content (see Materials and methods), where we would expect less impact of compositional differences on rate

estimation, we observed a reduction in the strength of the negative correlation relative to the high-variance set ( $\rho = -0.3018, P = 0.0396$  for the low variance set;  $\rho = -0.5307, P = 0.0001$  for the high variance set; Figure 4). However, the sign of the correlation did not reverse. We additionally calculated the correlation between body mass and  $d_N/d_S$  through time controlled for equilibrium GC content using Coevol. This did not alter the correlation coefficient ( $R = -0.302, pp = 0.019$  for the taxon-rich set). Note that median  $d_N/d_S$  was lower for the high-variance subset (median 0.0939) than for the low-variance set (median 0.2301; Wilcoxon test  $P = 2.2 \times 10^{-16}$ ; Figure 4).

#### Divergence time and estimation of $d_N/d_S$

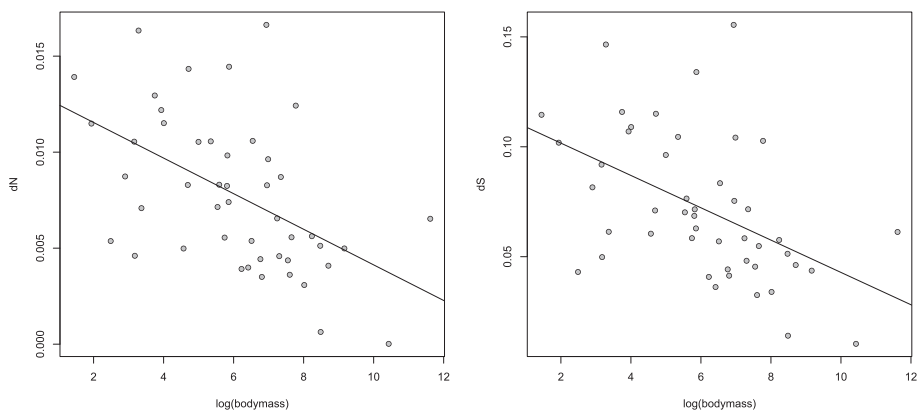
Another possibility that may explain the negative relationship between  $d_N/d_S$  and  $N_e$  is that there is a dependence of  $d_N/d_S$  on time. When divergence times are short, the ratio may be inflated owing to artefacts that can be statistical or biological in nature and do not reflect a genuine



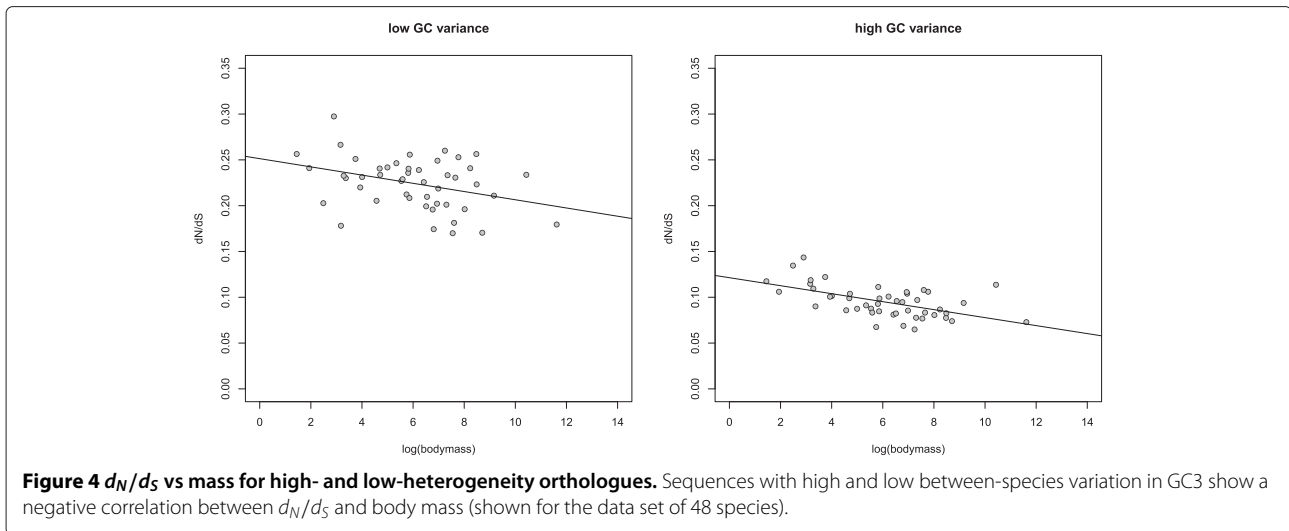
**Figure 2 Small-bodied species have higher  $d_N/d_S$ .** Body mass is significantly negatively correlated with  $d_N/d_S$  (shown for the data set of 48 species).

acceleration in the evolutionary rate. Here, both divergence times and terminal branch lengths are determined by the phylogeny considered. Explanations that have been proposed for this include segregating deleterious non-synonymous polymorphisms, the non-linear dynamics of the ratio of the two variables, and model misspecification due to failure to account for amino acid preference in different protein domains [45-50]. The time required for this effect to decay has been suggested to depend on  $N_e$  [47], which could be potentially problematic for our

data given that we find a positive correlation between body mass and time since divergence from the most recent ancestor ( $\rho = 0.406, P = 0.0127$ ; calibration points with confidence intervals in the upper quintile were excluded), indicating shorter times for birds with larger populations. Meanwhile,  $d_N/d_S$  is negatively correlated with divergence time, that is,  $d_N/d_S$  is higher for shorter branches ( $\rho = -0.3288, P = 0.047$ ; note that passerines have especially short branches, see Figure 1). However, while controlling the correlation between body mass and



**Figure 3  $d_N$  and  $d_S$  negatively correlate with mass.** Small birds exhibit more rapid rates of divergence at both synonymous ( $d_S$ ) and non-synonymous ( $d_N$ ) sites (shown for the data set of 48 species).



$d_N/d_S$  for divergence time leads to a reduction in the correlation coefficient (partial Spearman's rank correlation  $\beta = -0.3211, P = 0.0480$ , compared to  $\rho = -0.4106, P = 0.0122$  for  $d_N/d_S$  versus mass for the filtered data set), it does not altogether remove the relationship, which remains marginally significant. On the other hand, controlling the correlation between divergence time and  $d_N/d_S$  for body mass has a greater impact and renders it non-significant ( $\beta = -0.1945, P = 0.2476$ ). Finally, it should be noted that our data set mainly consists of relatively divergent lineages (> 90% have divergence times 20 to 73 million years ago), where any time dependence on  $d_N/d_S$  should be limited.

#### Saturation at third codon positions may impact estimation of $d_S$

Although  $d_S$  is often used as a proxy for the mutation rate when considering the  $d_N/d_S$  ratio, this idea ought to be treated with caution [51-53]. A reduction in  $d_S$  could be caused either by a *de facto* constraint on the rate of synonymous substitution, or methodological limitations such as saturation that lead to underestimation of the true rate. This is of particular concern for the estimation of  $d_N/d_S$  as synonymous rates might be more prone to underestimation than non-synonymous rates, because non-synonymous substitutions are generally less commonly fixed.

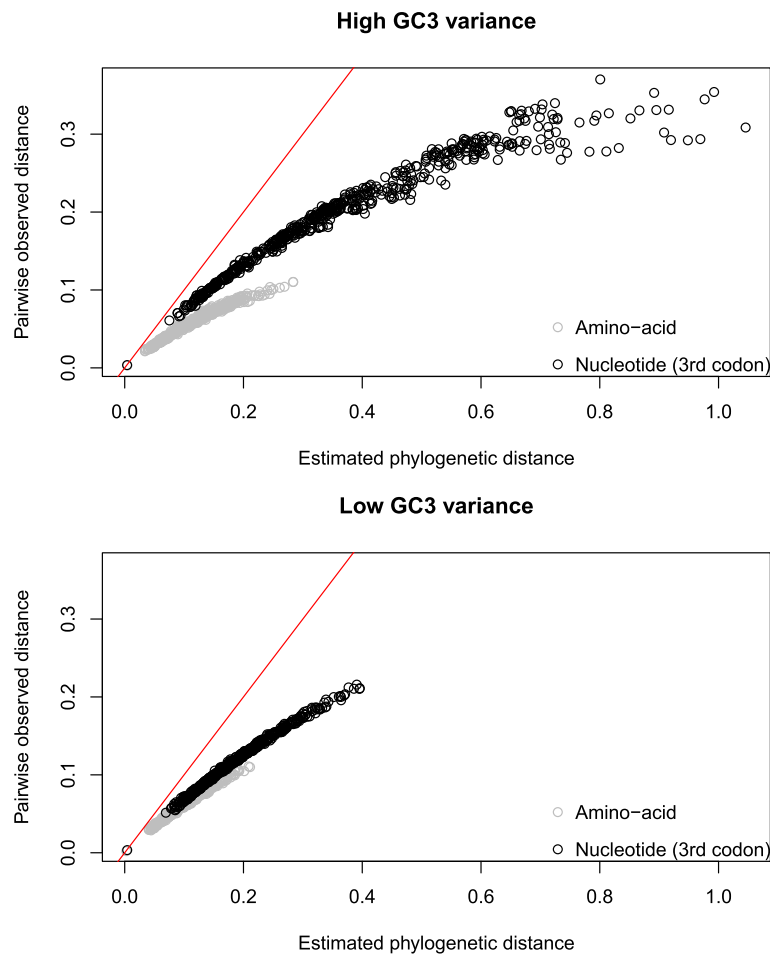
To assess whether there is evidence for saturation in our data, we compared the phylogenetic distance (the sum of branch lengths between two given species) to the number of uncorrected pairwise differences for high- and low-variance sequences, as considering the full data set would not have been computationally tractable. That the uncorrected distance does not increase linearly with the corrected distance for the high-variance subset,

instead remaining lower (Figure 5), indicates that there are multiple hits. As expected, divergence for third codon positions is greater than for amino acids. This implies that a degree of saturation and therefore underestimation of  $d_S$  relative to  $d_N$  might be of concern for our data. The weaker signal of saturation at third positions relative to amino acids in the lower-variance subset is consistent with the shorter branch lengths observed here (Figure 5). Constraint cannot explain the patterns we observe in the saturation plots, as it would affect both observed and phylogenetic distances. Nevertheless, the extent to which saturation affects our estimates of  $d_N/d_S$  is not clear.

#### Radical amino acid changes are less frequent in birds with large populations

If saturation at third sites is stronger than for amino acid substitutions and/or if many non-synonymous substitutions behave as effectively neutral, we may consider an alternative metric to assess how effectively slightly deleterious changes are purged from large bird populations. Radical amino acid changes that alter the polarity or volume of a residue are more likely to be negatively selected than conservative amino acid changes, as selective effects tend to be greater where replacements involve residues with dissimilar properties [54,55]. The ratio of radical to conservative substitutions has been suggested to be an appropriate means of testing the predictions of nearly neutral theory and overcoming saturation [29,55-57]. Here, we therefore employ  $K_r/K_c$  as our metric, where  $K_r$  and  $K_c$  respectively denote radical and conservative changes.

As expected given that  $d_N$  is higher in small birds, both  $K_r$  and  $K_c$  correlate negatively with body mass ( $K_r$ :  $\rho = -0.5338, P = 0.0001$ ;  $K_c$ :  $\rho = -0.5872, P = 2.1 \times 10^{-5}$ )



**Figure 5 Saturation for substitution at third sites and amino acid changes.** Uncorrected observed differences are smaller than estimated phylogenetic distances for third codon sites and amino acids. The red line denotes the relationship expected in the absence of saturation ( $x = y$ ), as the phylogenetic and observed distances would be equal in this case. Divergence for third codon positions is greater than for amino acids, particularly in orthologues with high variance in GC3 between species.

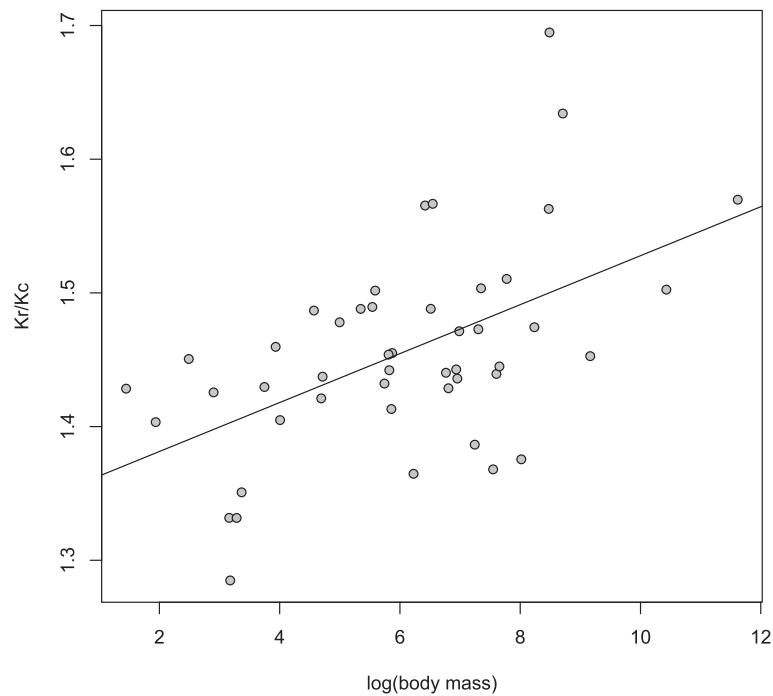
for the concatenated orthologues from the data set of 48 species. In stark contrast to  $d_N/d_S$ ,  $K_r/K_c$  is positively correlated with body mass ( $\rho = 0.4998, P = 0.0004$ ; Figure 6), suggesting that radical changes are more frequently removed from lineages with large populations. Results from Coevol confirm the positive relationship between body mass and  $K_r/K_c$  for sequences with high ( $r = 0.61, pp = 1.0$ ) and low variance in GC3 ( $r = 0.85, pp = 1.0$ ). It is also interesting to note that  $K_r/K_c$  is somewhat reduced in the high-variance subset (median 1.3599) compared to the low-variance subset (median 1.5408; Wilcoxon test  $P = 7.2 \times 10^{-12}$ ; mapNH results; Figure 7), paralleling the differences we observed for  $d_N/d_S$  (see Figure 4).

It should be noted that differences in base composition might affect the estimation of radical and conservative changes [55,56]. Given the well-characterised heterogeneity in GC content between our species, we ask whether

our results are robust to control for composition. While the partial correlation for  $K_r/K_c$  and mass controlling for GC3 is slightly reduced ( $\beta = 0.3882, P = 0.0057$ ), the correlation for GC3 and  $K_r/K_c$  controlling for mass becomes non-significant ( $\beta = -0.0431, P = 0.7770$  compared to  $\rho = -0.3215, P = 0.0298$ ). We thus find no evidence that base composition explains our observations. Note also that composition is more homogeneous between lineages in the low-variance data but this does not diminish the correlation. These results therefore support the idea that in birds radical amino acid changes are indeed more often removed from large populations than from small populations.

## Discussion

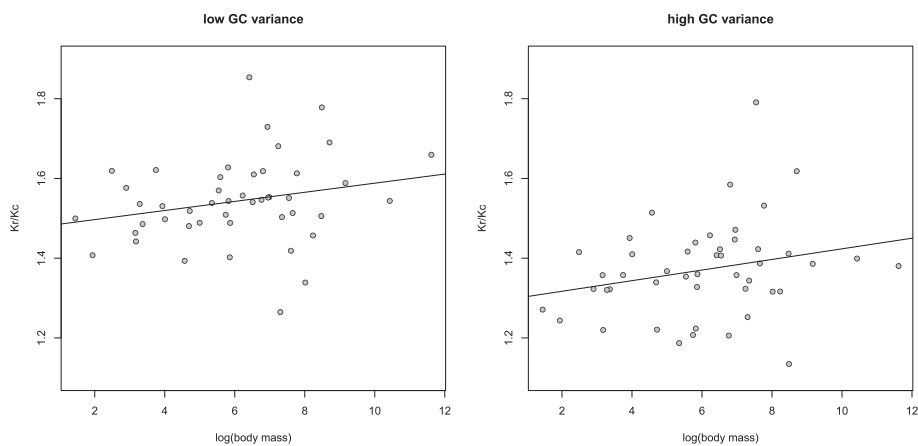
Employing a data set comprising 1,185 orthologues from 48 recently sequenced bird genomes, we examined relationships between life history and lineage-specific



**Figure 6  $K_r/K_c$  correlates positively with body mass.** Large-bodied species tend to have elevated  $K_r/K_c$  compared to small-bodied species, in principle consistent with less effective purifying selection (shown for the data set of 48 species).

patterns of substitution. We found no evidence of reduced  $d_N/d_S$  in birds with putatively higher effective population size, in apparent contradiction to nearly neutral theory. On the contrary, we consistently saw a negative correlation between body mass and lineage-specific  $d_N/d_S$ , similar to what was recently reported based on analyses of avian mitochondria [29]. This is particularly striking and not necessarily expected, given the many inherent differences between nuclear and mitochondrial sequences, as

well as the fact that we were able to consider a much larger data set here. Our observations contrast with reports of a positive correlation between body mass and  $d_N/d_S$  in mammals. However, considering the ratios of radical to conservative amino acid substitutions, we found a positive correlation between body size and  $K_r/K_c$ , meaning that lineages with putatively larger populations experience relatively fewer changes that alter the polarity and volume of a residue. That is, those differences that do occur



**Figure 7  $K_r/K_c$  correlates positively with body mass for both GC-heterogeneous and GC-homogeneous genes.** Sequences with high and low between-species variation in GC3 show a positive correlation between  $K_r/K_c$  and body mass (shown for the data set of 48 species).

in small-bodied birds may be less likely to disrupt protein function, which is in principle consistent with the notion that selection will more effectively purge deleterious changes from large populations. In contrast with  $d_N/d_S$ -based estimates, our amino acid substitution data ( $K_r/K_c$ ) therefore appear to accord with the predictions of the nearly neutral theory. If a significant proportion of non-synonymous substitutions are conservative and behave as effectively neutral, this may obscure (expected) correlations between  $d_N/d_S$  and life history. Thus, in this case,  $K_r/K_c$  could potentially be a more fine-grained measure for assessing the prevalence of protein-level selection in different lineages.

While radical amino acid mutations should be subject to stronger negative selection, it has been suggested that adaptive evolution may lead to similar proportions of radical and conservative fixation [55]. Could increased rates of adaptive evolution in small birds be responsible for our observation that  $d_N/d_S$ , but not  $K_r/K_c$ , increases with decreasing body mass? Given a high proportion of effectively positively selected mutations, we might predict that the rate of fixation will increase with population size [16]. However, to affect the genome-wide average substantially, positive selection would need to be common, which is unlikely to be the case in vertebrate species with modest population sizes. The absence of high-resolution diversity data limits our ability to quantify directly the prevalence of adaptive non-synonymous substitutions in our study species. Although a past survey of chicken and zebra finch divergence and diversity data estimated the frequency of amino acid changes driven to fixation by positive selection ( $\alpha$ ) to be around 20% [58], this value did not differ significantly from zero. Further, simulations indicate that the influence of  $N_e$  on the proportion of adaptive amino acid changes is limited, impacting mainly populations under 10,000 [37]. Since birds typically have larger  $N_e$  than this, we might not necessarily expect differences in  $N_e$  to lead to adaptive changes being more common in smaller-bodied species. This prediction is for instance reflected in the similar percentage of fixations driven by positive selection in *Drosophila miranda* and *D. melanogaster* despite a fivefold difference in population size [59].

There are several conceivable explanations for the discrepancy between our results for the relationship between  $d_N/d_S$  and  $N_e$  and theoretical expectations. One possibility is that body mass is a poor proxy for population size in birds [60], but it is not clear how this alone could lead to a reversal in the sign of the correlation, though it could in principle introduce noise. Moreover, the fact that we correlated body size of a single extant species with substitution rates reflecting evolutionary processes in multiple ancestors over significant periods of time naturally means that strong relationships cannot be expected. Another is

that there was limited variation in  $d_N/d_S$  (0.13 to 0.17), again weakening the signal in the data. Further, there was some evidence that third sites could be moderately saturated, indicating that we tend to underestimate synonymous changes for greater divergences, such as those observed in small-bodied bird lineages. How much of the variation this might explain is not clear, and divergence appears somewhat low for saturation alone to have a large impact. Given significant constraint on fourfold degenerate sites in birds [61], a reduction in  $d_S$  could also be caused by selection on silent sites. However, there is currently no evidence for a correspondence between constraint and population size [51,61]. Interestingly, we find that species  $d_N/d_S$  and  $d_S$  are positively correlated ( $\rho = 0.535, P = 0.0001$ ), counter to what one might expect given that  $d_S$  is the denominator of  $d_N/d_S$ . This could either indicate a bias in rate estimation or merely be an artefact of the correlations between rates and life history. It is possible that multiple factors work together to produce the pattern observed. Indeed, restricting analyses to orthologues conserved across multiple species can in itself reverse already weak correlations between genomic parameters [62].

Further, non-stationary GC content can affect estimation of substitution rates, but we detect no clear evidence for this. Given the well-established role of GC-biased gene conversion (gBGC), in driving heterogeneity in avian base composition [38,63,64], it could also impact substitution rates. gBGC is associated with the rate of meiotic recombination and leads to the preferential fixation of GC over AT alleles [65-67].  $d_N$  in particular has been suggested to increase near mammalian recombination hotspots in the absence of positive selection as a result [68-72]. Since small-bodied bird species tend to have increased GC content [38], it is tempting to speculate that  $d_N/d_S$  could be inflated in these lineages. In mammals, correlations between body mass and  $d_N/d_S$  are partly masked by the effects of gBGC overcoming weak selection [9]. However, the impact of gBGC on global  $d_N/d_S$  is difficult to assess conclusively given that we do not have relevant information on rates of recombination for the majority of our study species. This should be further investigated once detailed estimates of recombination rates become available. Interestingly, no AT  $\rightarrow$  GC bias is seen in rapidly diverging sequences between chicken and zebra finch [73].

An additional issue that could affect the estimation of  $d_N/d_S$  is the quality of the sequence alignments from which rates are estimated. In principle, if aligned sequences from small-bodied birds were more prone to false positive homology calls, spurious non-synonymous substitutions may be inferred, resulting in a potentially upward-biased  $d_N/d_S$ . While theoretically possible [74], removing the impact of alignment uncertainty on inferred



substitution rates is currently prohibitively computationally costly. Several authors have previously discussed the impact of aligner choice on the rate of false positive inference of positive selection [75-77], and report that certain algorithms perform better than others. We emphasise that the first pass of alignments for the data set of 48 species was performed using SATé+PRANK (see Materials and methods), and that the class of aligners that PRANK belongs to appears less prone to false positives than others [75-77]. As such, our approach ought to be as robust as is currently feasible for a data set of this size. To address these limitations conclusively, comprehensive studies on the impact of sequence divergence on alignment uncertainty as well as further advances in alignment and rate estimation methods will be needed.

We finally note that an alternative explanation might be that the discrepancy between  $K_r/K_c$  and  $d_N/d_S$  is not merely owing to methodological artefacts relating to measuring  $d_N/d_S$  accurately but that our naive model of how substitution rates ought to relate to population size is incomplete. The range of  $N_e$  across which nearly neutral dynamics are expected to hold depends on the distribution of selective effects that is assumed [16]. Some models propose that the distribution of selection coefficients for mutants depends on current fitness, impacting the rate of acceptance of slightly deleterious mutations [78-80]. Accordingly, it has been suggested that dependence of  $d_N/d_S$  on  $N_e$  may be weak [78], with changes in population size rather than population size *per se* modulating  $d_N/d_S$  and both expansions and contractions leading to increases in the ratio [33,79]. The rate of diversification appears to correlate positively with the rate of molecular evolution in bird but not mammalian lineages [28,81], tempting speculation that rapidly evolving birds are especially prone to frequent population size fluctuations. However, to explain our observations, under the size fluctuation model  $K_r/K_c$  would have to be relatively less sensitive than  $d_N/d_S$  to changes in  $N_e$  and more sensitive to  $N_e$  itself.

## Conclusions

Although branch-specific estimates of  $d_N/d_S$  show no evidence for more efficient selection in large bird populations,  $K_r/K_c$  estimates appear to conform to the predictions of nearly neutral theory in birds, with small-bodied birds tending to have fewer radical amino acid changes. If, as one interpretation of our work suggests,  $K_r/K_c$  is more robust in certain scenarios, gathering deeper insight into the dynamics of this measure will be of broad relevance for inference of protein-level selection. Further, we suggest that the role of gBGC and how the distribution of selective effects differs between different populations will need to be elucidated to determine

conclusively to what extent  $d_N/d_S$  is determined by population size under the nearly neutral theory of molecular evolution.

The practical implications of our observations depend partly on the precise mechanisms responsible. How, for instance, might tests for positive selection be influenced? One might imagine that an upward bias in  $d_N/d_S$  within a given lineage could lead to the naive assumption that a higher proportion of coding sequences with an average  $d_N/d_S > 1$  indicates more frequent adaptation. How branch-site tests might be affected is difficult to predict without knowing the distribution of sites that violate our assumptions of how  $d_N$  and  $d_S$  ought to behave. It has been suggested that branch-site models may lack power when saturation is present, but are less likely to yield false positives [82]. This contrasts with the higher expected rate of false positives caused by alignment problems [75-77]. We also note that comparisons between species and comparisons of different classes of sequence within genomes are expected to be affected differently by certain artefacts. For instance, ecological shifts might affect lineage-specific rate estimates to a greater extent than gene-specific rates [16], while a constraint on  $d_S$  [53] could impact  $d_N/d_S$  in both cases.

Overall, our observations suggest that a careful examination of potential sources of error is called for when interpreting evolutionary rate estimates, and that this must be done with the specific questions and data set in mind. Further, while we cannot presently conclude that radical and conservative rates are inherently more reliable for detecting negative selection, the fact that  $d_N/d_S$  does not consider the effects of different classes of non-synonymous change suggests that it likely presents an incomplete picture of selective processes.

## Materials and methods

### Sequence alignments

#### Data for 48 genomes

Coding sequence alignments for 48 bird species (see Additional file 2) were obtained from a recent initiative to resolve the phylogeny of modern birds; see Jarvis *et al.* [40] and Zhang *et al.* [83] for a detailed description of how these data were generated. Briefly, this data set comprises 8,295 orthologous protein-coding sequences identified by propagating chicken and zebra finch annotations to the remaining species and classifying orthology by combining information from alignment statistics, reciprocal best hits and synteny. Multiple sequence alignments were generated by running SATé+PRANK followed by SATé+MAFFT on concatenated exon sequences [40]. Of 1,185 1:1 orthologues present in all species, 921 contained no internal stop codons. Concatenated alignments comprising the highest and lowest variance in GC3 from the same study were also considered [40].

### Data for 169 species

To extend our taxon sampling, we also analysed 11,160 bp of sequence from 169 avian species, consisting of the coding sequences of the Hackett *et al.* [41] data set and two additional widely used phylogenetic markers, RAG1 and RAG2, which were downloaded from GenBank (see Additional file 3 for accession numbers). The marker sequences were translated into amino acids, aligned using MUSCLE [84] and subsequently converted back to nucleotides. These data are what we refer to as the taxon-rich set.

### Life history traits

Body mass data were extracted from the *CRC Handbook of Avian Body Masses* [85] for all available tip nodes. Where multiple entries for a given species were present, the mean value was used.

### Phylogenetic trees

For the taxon-rich data set, we used the tree of Hackett *et al.* [41]. For the 48 genomes, the total evidence nucleotide tree estimated by Jarvis *et al.* [40] was used, along with corresponding time calibration points, which we considered for our divergence time analyses.

### Sequence divergence

#### Maximum likelihood estimation

Given the difference in the sizes of the two alignment data sets, as well as in the evolutionary distances between the sampled taxa, we employed two different methods of maximum likelihood estimation. To make the analyses on the larger gene-rich data set with less dense taxon sampling tractable, we approximated branch-specific  $d_N/d_S$  ratios by substitution mapping using mapNH [24,86]. We did this by fitting a homogeneous YN98 [87] model to coding sequence alignments and subsequently mapping synonymous and non-synonymous substitutions onto individual branches. This was done separately for each orthologue from the 1:1 set that did not include an internal stop, and  $d_N/d_S$  was obtained by summing substitution counts prior to dividing to avoid low count numbers introducing noise. To make these numbers comparable to those from Codeml, the ratio of non-synonymous to synonymous counts was divided by 3. As the branches leading to the two eagles were too short to estimate  $d_N/d_S$  reliably, we considered only *Haliaeetus albicilla*.

$d_N$  and  $d_S$  were obtained by fixing  $\omega = 1$  in mapNH (following the rationale presented in Yang and Nielsen [87], p. 411) and multiplying the resulting normalised substitution counts by the corresponding branch lengths. This feature is implemented in the development version of Bio++ [88], available online [89].

On the other hand, for the 11-kb taxon-rich data set, rates were estimated using Codeml [90] with lineages grouped by taxonomic order to reduce variance in  $d_N/d_S$

owing to short branches. We assigned one local  $d_N/d_S$  for every avian order, resulting in 53 local values (see Additional file 4 for groups). Concatenating the alignments further served to reduce noise.

The ratio of radical to conservative amino acid changes ( $K_r/K_c$ ) for the taxon-rich data set was calculated by concatenating 1,185 1:1 orthologues, fitting a Jukes–Cantor model and mapping radical and conservative substitution counts onto the tree using mapNH. Radical changes are those that alter the polarity or volume of the residue. Here, L, I, F, M, Y, W, H, K, R, E and Q were classified as having large volumes, while Y, W, H, K, R, E, Q, T, D, N, S and C were classified as polar. Results using a WAG01 model were qualitatively similar to those calculated using the Jukes–Cantor model. Considering each orthologue individually before summing counts yielded noisy results, presumably owing to low numbers of radical amino acid substitutions in individual alignments. Overall, performance was better where a greater number of substitution counts was available, as using the full set of 8,295 orthologues yielded a slightly stronger correlation between body mass and  $K_r/K_c$  than when smaller subsets were considered ( $\rho = 0.513$ ,  $P = 0.0003$ ). Due to the short eagle branches, *Haliaeetus leucocephalus* was excluded.

#### Bayesian estimation of coevolution between substitution and life history

Coevol [42] was used on subsets of the gene-rich data set to calculate  $K_r/K_c$  and  $d_N/d_S$ . As above, the polarity and volume definition (-polvol) was used to classify amino acid changes as radical or conservative. To control the relationship between body mass and  $d_N/d_S$  for equilibrium base composition, we also ran Coevol with equilibrium GC as a parameter. A more detailed description of the methods used, as well as priors and calibration points, is given in Nabholz *et al.* [29].

#### Saturation analysis

From the 830 orthologues with the highest and lowest variance in GC3, 200 genes were randomly selected [40]. The pairwise divergence was computed from the number of observed differences between two sequences without correction for multiple substitutions. The phylogenetic distance (that is, the patristic distance) was obtained from the sum of branch lengths between two species, computed using a phylogenetic tree estimated by maximum-likelihood using PAML. We used a GTR+GAMMA model in baseml [90] for the third codon position data set and WAG, an empirical substitution matrix, in Codeml [90] for the protein data set.

#### Statistics and data availability

Statistical analyses were performed in R. The genome data from the 48 bird species are available online [91].

## Additional files

**Additional file 1:  $d_N/d_S$  versus mass for the species-rich set.** Supplementary information.

**Additional file 2: Species table.** Table of species names and abbreviations.

**Additional file 3: GenBank accession numbers.** Accession numbers for RAG1 and RAG2 sequences used in taxon-rich set.

**Additional file 4: Taxon grouping.** Groups of taxa used for Codeml  $d_N/d_S$  analysis.

## Abbreviations

bp: base pair; gBGC: GC-biased gene conversion; kb: kilobase.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

CCW and HE initiated and conceived the project. CCW, BN and JR designed and performed the analyses. CCW wrote the manuscript together with HE, and coordinated the project. All authors read, edited and approved the paper.

## Acknowledgments

Computational analyses were performed using resources provided by the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) of Uppsala University, supported by the Swedish National Infrastructure for Computing (SNIC). The study was funded by the European Research Council (AdG 249976), Knut and Alice Wallenberg Foundation (Wallenberg Scholar Grant and the Swedish Research Council (2010-5650; all to HE). We thank Erich Jarvis for coordinating the avian phylogenomics project and, along with Guojie Zhang and Tom Gilbert, providing access to genomes. We also thank Laurent Guéguen for help with estimating  $d_N$  and  $d_S$  in mapNH, Carina Mugal, Jochen Wolf, Nicolas Galtier, Julien Duthel, Simon Whelan and Laurence Hurst for helpful discussions, and two reviewers for their comments on the manuscript. This is publication ISE-M 2014-189.

## Author details

<sup>1</sup>Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18 D, Uppsala, Sweden. <sup>2</sup>Institut des Sciences de l'Evolution-Montpellier, UMR CNRS-UM2 5554, University Montpellier II, 34095 Montpellier, France. <sup>3</sup>Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland.

Received: 30 May 2014 Accepted: 13 November 2014

Published online: 11 December 2014

## References

1. Britten R: **Rates of DNA sequence evolution differ between taxonomic groups.** *Science* 1986, **39**:1393–1398.
2. Wu CI, Li WH: **Evidence for higher rates of nucleotide substitution in rodents than in man.** *Proc Natl Acad Sci USA* 1985, **82**:1741–1745.
3. Ohta T: **Evolutionary rate of cistrons and DNA divergence.** *J Mol Evol* 1972, **1**:150–157.
4. Bromham L: **Why do species vary in their rate of molecular evolution?** *Biol Lett* 2009, **5**:401–404.
5. Nabholz B, Glémin S, Galtier N: **Strong variations of mitochondrial mutation rate across mammals – the longevity hypothesis.** *Mol Biol Evol* 2008, **25**:120–130.
6. Martin AP, Palumbi SR: **Body size, metabolic rate, generation time, and the molecular clock.** *Proc Natl Acad Sci USA* 1993, **90**:4087–4091.
7. Welch JJ, Bininda-Emonds ORP, Bromham L: **Correlates of substitution rate variation in mammalian protein-coding sequences.** *BMC Evol Biol* 2008, **8**:53.
8. Wilson Sayres MA, Venditti C, Pagel M, Makova KD: **Do variations in substitution rates and male mutation bias correlate with life-history traits? A study of 32 mammalian genomes.** *Evolution* 2011, **65**:2800–2815.
9. Lartillot N, Delsuc F: **Joint reconstruction of divergence times and life-history evolution in placental mammals using a phylogenetic covariance model.** *Evolution* 2012, **66**:1773–1787.
10. Bromham L: **The genome as a life-history character: why rate of molecular evolution varies between mammal species.** *Philos Trans R Soc Lond B: Biol Sci* 2011, **366**:2503–2513.
11. Li WH, Ellsworth DL, Krushkal J, Chang BH, Hewett-Emmett D: **Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis.** *Mol Phylogenet Evol* 1996, **5**:182–187.
12. Thomas GWC, Hahn MW: **The human mutation rate is increasing, even as it slows.** *Mol Biol Evol* 2014, **31**:253–257.
13. Damuth J: **Population density and body size in mammals.** *Nature* 1981, **290**:699–700.
14. Kimura M: **On the probability of fixation of mutant genes in a population.** *Genetics* 1962, **47**:713–719.
15. Ohta T: **Slightly deleterious mutant substitutions in evolution.** *Nature* 1973, **246**:96–98.
16. Akashi H, Osada N, Ohta T: **Weak selection and protein evolution.** *Genetics* 2012, **192**:15–31.
17. Ohta T, Gillespie J: **Development of neutral and nearly neutral theories.** *Theor Popul Biol* 1996, **49**:128–142.
18. Ohta T: **Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory.** *J Mol Evol* 1995, **40**:56–63.
19. Woolfit M, Bromham L: **Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes.** *Mol Biol Evol* 2003, **20**:1545–1555.
20. Warnecke T, Rocha EPC: **Function-specific accelerations in rates of sequence evolution suggest predictable epistatic responses to reduced effective population size.** *Mol Biol Evol* 2011, **28**:2339–2349.
21. Lanfear R, Kokko H, Eyre-Walker A: **Population size and the rate of evolution.** *Trends Ecol Evol* 2014, **29**:33–41.
22. Nikolaev SI, Montoya-Burgos JI, Popadin KY, Parand L, Margulies EH, Antonarakis SE: **Life-history traits drive the evolutionary rates of mammalian coding and noncoding genomic elements.** *Proc Natl Acad Sci USA* 2007, **104**:20443–20448.
23. Kosiol C, Vinar T, da Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, Siepel A: **Patterns of positive selection in six mammalian genomes.** *PLoS Genet* 2008, **4**:1000144.
24. Romiguier J, Figueat E, Galtier N, Douzery EJP, Boussau B, Duthel JY, Ranwez V: **Fast and robust characterization of time-heterogeneous sequence evolutionary processes using substitution mapping.** *PLoS One* 2012, **7**:33852.
25. Lartillot N: **Interaction between selection and biased gene conversion in mammalian protein-coding sequence evolution revealed by a phylogenetic covariance analysis.** *Mol Biol Evol* 2013, **30**:356–368.
26. Romiguier J, Ranwez V, Douzery EJP, Galtier N: **Genomic evidence for large, long-lived ancestors to placental mammals.** *Mol Biol Evol* 2013, **30**:5–13.
27. Popadin KY, Polishchuk LV, Mamirova L, Knorre D, Gunbin K: **Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals.** *Proc Natl Acad Sci USA* 2007, **104**:13390–13395.
28. Lanfear R, Ho SYW, Love D, Bromham L: **Mutation rate is linked to diversification in birds.** *Proc Natl Acad Sci USA* 2010, **107**:20423–20428.
29. Nabholz B, Uwimana N, Lartillot N: **Reconstructing the phylogenetic history of long-term effective population size and life-history traits using patterns of amino acid replacement in mitochondrial genomes of mammals and birds.** *Genome Biol Evol* 2013, **5**:1273–1290.
30. Ballard JWO, Whitlock MC: **The incomplete natural history of mitochondria.** *Mol Ecol* 2004, **13**:729–744.
31. Popadin KY, Nikolaev SI, Junier T, Baranova M, Antonarakis SE: **Purifying selection in mammalian mitochondrial protein-coding genes is highly effective and congruent with evolution of nuclear genes.** *Mol Biol Evol* 2013, **30**:347–355.
32. Woolfit M, Bromham L: **Population size and molecular evolution on islands.** *Proc R Soc B: Biol Sci* 2005, **272**:2277–2282.
33. Charlesworth J, Eyre-Walker A: **The other side of the nearly neutral theory, evidence of slightly advantageous back-mutations.** *Proc Natl Acad Sci USA* 2007, **104**:16992–16997.

34. Wright SD, Gillman LN, Ross HA, Keeling DJ: **Slower tempo of microevolution in island birds: implications for conservation biology.** *Evolution* 2009, **63**:2275–2287.
35. Johnson KP, Seger J: **Elevated rates of nonsynonymous substitution in island birds.** *Mol Biol Evol* 2001, **18**:874–881.
36. Loire E, Chiari Y, Bernard A, Cahais V, Romiguier J, Nabholz B, Lourenço JM, Galtier N: **Population genomics of the endangered giant Galápagos tortoise.** *Genome Biol* 2013, **14**:136.
37. Lourenço JM, Glémin S, Galtier N: **The rate of molecular adaptation in a changing environment.** *Mol Biol Evol* 2013, **30**:1292–1301.
38. Weber CC, Boussau B, Romiguier J, Jarvis ED, Ellegren H: **Evidence for GC-biased gene conversion as a driver of between-lineage differences in avian base composition.** *Genome Biol* 2014. doi:10.1186/s13059-014-0549-1.
39. Nabholz B, Künstner A, Wang R, Jarvis ED, Ellegren H: **Dynamic evolution of base composition: causes and consequences in avian phylogenomics.** *Mol Biol Evol* 2011, **28**:2197–2210.
40. Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT, Suh A, Weber CC, da Fonseca RR, Li J, Zhang F, Li H, Zhou L, Narula N, Liu L, Ganapathy G, Boussau B, Bayzid MS, Zavidovych V, Subramanian S, Gabaldón T, Capella-Gutiérrez S, Huerta-Cepas J, Rekepalli B, Munch K, Schierup M, et al.: **Whole-genome analyses resolve early branches in the tree of life of modern birds.** *Science* 2014. doi:10.1126/science.1253451.
41. Hackett SJ, Kimball RT, Reddy S, Bowie RCK, Braun EL, Braun MJ, Chojnowski JL, Cox WA, Han K-L, Harshman J, Huddleston CJ, Marks BD, Miglia KJ, Moore WS, Sheldon FH, Steadman DW, Witt CC, Yuri T: **A phylogenomic study of birds reveals their evolutionary history.** *Science* 2008, **320**:1763–1768.
42. Lartillot N, Poujol R: **A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters.** *Mol Biol Evol* 2011, **28**:729–744.
43. Bay RA, Bielawski JP: **Inference of functional divergence among proteins when the evolutionary process is non-stationary.** *J Mol Evol* 2013, **76**:205–215.
44. Weber CC, Hurst LD: **Protein rates of evolution are predicted by double-strand break events, independent of crossing-over rates.** *Genome Biol Evol* 2009, **1**:340–349.
45. Rocha EPC, Smith JM, Hurst LD, Holden MTG, Cooper JE, Smith NH, Feil EJ: **Comparisons of  $dN/dS$  are time dependent for closely related bacterial genomes.** *J Theoretical Biol* 2006, **239**:226–235.
46. Ho SYW, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigo AG, Cooper A: **Time-dependent rates of molecular evolution.** *Mol Ecol* 2011, **20**:3087–3101.
47. Peterson GI, Masel J: **Quantitative prediction of molecular clock and  $K_A/K_S$  at short timescales.** *Mol Biol Evol* 2009, **26**:2595–2603.
48. Dos Reis M, Yang Z: **Why do more divergent sequences produce smaller nonsynonymous/synonymous rate ratios in pairwise sequence comparisons?** *Genetics* 2013, **195**:195–204.
49. Mugal CF, Wolf JBW, Kaji I: **Why time matters: codon evolution and the temporal dynamics of  $dN/dS$ .** *Mol Biol Evol* 2014, **31**:212–231.
50. Wolf JBW, Künstner A, Nam K, Jakobsson M, Ellegren H: **Nonlinear dynamics of nonsynonymous ( $d_N$ ) and synonymous ( $d_S$ ) substitution rates affects inference of selection.** *Genome Biol Evol* 2009, **1**:308–319.
51. Eory L, Halligan DL, Keightley PD: **Distributions of selectively constrained sites and deleterious mutation rates in the hominid and murid genomes.** *Mol Biol Evol* 2010, **27**:177–192.
52. Chamary J-V, Parmley JL, Hurst LD: **Hearing silence: non-neutral evolution at synonymous sites in mammals.** *Nat Rev Genet* 2006, **7**:98–108.
53. Drummond DA, Wilke CO: **Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution.** *Cell* 2008, **134**:341–352.
54. Yampolsky LY, Kondrashov FA, Kondrashov AS: **Distribution of the strength of selection against amino acid replacements in human proteins.** *Hum Mol Genet* 2005, **14**:3191–3201.
55. Smith NGC: **Are radical and conservative substitution rates useful statistics in molecular evolution?** *J Mol Evol* 2003, **57**:467–478.
56. Wernegreen JJ: **Reduced selective constraint in endosymbionts: elevation in radical amino acid replacements occurs genome-wide.** *PLoS One* 2011, **6**:28905.
57. Eyre-Walker A, Keightley PD, Smith NGC, Gaffney D: **Quantifying the slightly deleterious mutation model of molecular evolution.** *Mol Biol Evol* 2002, **19**:2142–2149.
58. Axelsson E, Ellegren H: **Quantification of adaptive evolution of genes expressed in avian brain and the population size effect on the efficacy of selection.** *Mol Biol Evol* 2009, **26**:1073–1079.
59. Bachtrög D: **Similar rates of protein adaptation in *Drosophila miranda* and *D. melanogaster*, two species with different current effective population sizes.** *BMC Evol Biol* 2008, **8**:334.
60. Nee S, Read A, Greenwood J, Harvey P: **The relationship between abundance and body size in British birds.** *Nature* 1991, **351**:312–313.
61. Künstner A, Nabholz B, Ellegren H: **Significant selective constraint at 4-fold degenerate sites in the avian genome and its consequence for detection of positive selection.** *Genome Biol Evol* 2011, **3**:1381–1389.
62. Weber CC, Hurst LD: **Intronic AT skew is a defensible proxy for germline transcription but does not predict crossing-over or protein evolution rates in *Drosophila melanogaster*.** *J Mol Evol* 2010, **71**:415–426.
63. Webster MT, Axelsson E, Ellegren H: **Strong regional biases in nucleotide substitution in the chicken genome.** *Mol Biol Evol* 2006, **23**:1203–1216.
64. Mugal CF, Arndt PF, Ellegren H: **Twisted signatures of GC-biased gene conversion embedded in an evolutionary stable karyotype.** *Mol Biol Evol* 2013, **30**:1700–1712.
65. Webster MT, Hurst LD: **Direct and indirect consequences of meiotic recombination: implications for genome evolution.** *Trends Genet* 2012, **28**:101–109.
66. Duret L, Galtier N: **Biased gene conversion and the evolution of mammalian genomic landscapes.** *Annu Rev Genomics Hum Genet* 2009, **10**:285–311.
67. Duret L, Eyre-Walker A, Galtier N: **A new perspective on isochore evolution.** *Gene* 2006, **385**:71–74.
68. Kostka D, Hubisz MJ, Siepel A, Pollard KS: **The role of GC-biased gene conversion in shaping the fastest evolving regions of the human genome.** *Mol Biol Evol* 2012, **29**:1047–1057.
69. Ratnakumar A, Mousset S, Glémin S, Berglund J, Galtier N, Duret L, Webster MT: **Detecting positive selection within genomes: the problem of biased gene conversion.** *Philos Trans R Soc Lond B: Biol Sci* 2010, **365**:2571–2580.
70. Galtier N, Duret L, Glémin S, Ranwez V: **GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates.** *Trends Genet* 2009, **25**:1–5.
71. Berglund J, Pollard KS, Webster MT: **Hotspots of biased nucleotide substitutions in human genes.** *PLoS Biology* 2009, **7**:26.
72. Galtier N, Duret L: **Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution.** *Trends Genet* 2007, **23**:273–277.
73. Capra JA, Pollard KS: **Substitution patterns are GC-biased in divergent sequences across the metazoans.** *Genome Biol Evol* 2011, **3**:516–527.
74. Redelings B: **Erasing errors due to alignment ambiguity when estimating positive selection.** *Mol Biol Evol* 2014:1979–1993.
75. Markova-Raina P, Petrov DA: **High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 *Drosophila* genomes.** *Genome Res* 2011, **21**:863–874.
76. Jordan G, Goldman N: **The effects of alignment error and alignment filtering on the sitewise detection of positive selection.** *Mol Biol Evol* 2012, **29**:1125–1139.
77. Blackburne BP, Whelan S: **Class of multiple sequence alignment algorithm affects genomic analysis.** *Mol Biol Evol* 2013, **30**:642–653.
78. Cherry JL: **Should we expect substitution rate to depend on population size?** *Genetics* 1998, **150**:911–919.
79. Goldstein RA: **Population size dependence of fitness effect distribution and substitution rate probed by biophysical model of protein thermostability.** *Genome Biol Evol* 2013, **5**:1584–1593.
80. Wylie CS, Shakhnovich EI: **A biophysical protein folding model accounts for most mutational fitness effects in viruses.** *Proc Natl Acad Sci USA* 2011, **108**:9916–9921.

81. Goldie X, Lanfear R, Bromham L: **Diversification and the rate of molecular evolution: no evidence of a link in mammals.** *BMC Evol Biol* 2011, **11**:286.
82. Gharib WH, Robinson-Rechavi M: **The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC.** *Mol Biol Evol* 2013, **30**:1675–1686.
83. Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, Odeen A, Cui J, Zhou Q, Xu L, Pan H, Wang Z, Jin L, Zhang P, Hu H, Yang W, Hu J, Xiao J, Yang Z, Liu Y, Xie Q, Yu H, Lian J, Wen P, Zhang F, Li H, *et al.*: **Comparative genomics across modern bird species reveal insights into avian genome evolution and adaptation.** *Science* 2014. doi:10.1126/science.1251385.
84. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**:1792–1797.
85. Dunning JBJ: *CRC Handbook of Avian Body Masses*. 2nd edition. Boca Raton, Florida: CRC Press; 2007.
86. Dutheil JY, Galtier N, Romiguier J, Douzery EJP, Ranwez V, Boussau B: **Efficient selection of branch-specific models of sequence evolution.** *Mol Biol Evol* 2012, **24**:1–15.
87. Yang Z, Nielsen R: **Synonymous and nonsynonymous rate variation in nuclear genes of mammals.** *J Mol Evol* 1998, **46**:409–418.
88. Gueguen L, Gaillard S, Boussau B, Gouy M, Groussin M, Rochette NC, Bigot T, Fournier D, Pouyet F, Cahais V, Bernard A, Scornavacca C, Nabholz B, Haudry A, Dachary L, Galtier N, Belkhir K, Dutheil JY: **Bio++: efficient extensible libraries and tools for computational molecular evolution.** *Mol Biol Evol* 2013, **30**:1745–1750.
89. **Bio++ Wiki.** [<http://biopp.univ-montp2.fr/>]
90. Yang Z: **PAML 4: phylogenetic analysis by maximum likelihood.** *Mol Biol Evol* 2007, **24**:1586–1591.
91. Zhang G, Li B, Li C, Gilbert MTP, Jarvis E, The Avian Genome Consortium, Wang J: *The avian phylogenomic project data*; 2014. <http://dx.doi.org/10.5524/101000>.

doi:10.1186/s13059-014-0542-8

**Cite this article as:** Weber *et al.*:  $K_r/K_e$  but not  $d_N/d_S$  correlates positively with body mass in birds, raising implications for inferring lineage-specific selection. *Genome Biology* 2014 **15**:542.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

