Different Host–Endogenous Retrovirus Relationships between Mammals and Birds Reflected in Genome-Wide Evolutionary Interaction Patterns

Wanjing Zheng[1,2], Jun Gojobori[1,3], Alexander Suh[4,5], and Yoko Satta[1,3,*]

[1]Department of Evolutionary Studies of Biosystems, School of Advanced Sciences, SOKENDAI (The Graduate University for Advanced Studies), Kanagawa 240-0193, Japan

[2]School of Life Sciences, Fudan University, Shanghai 200438, China

[3]Research Center for Integrative Evolutionary Science, SOKENDAI (The Graduate University for Advanced Studies), Kanagawa 240-0193, Japan

[4]Department of Organismal Biology—Systematic Biology, Evolutionary Biology Centre (EBC), Uppsala University, Uppsala 75236, Sweden

[5]School of Biological Sciences—Organisms and the Environment, University of East Anglia, Norwich, UK

*Corresponding author: E-mail: satta@soken.ac.jp.

Accepted: March 21, 2024

Abstract

Mammals and birds differ largely in their average endogenous retrovirus loads, namely the proportion of endogenous retrovirus in the genome. The host–endogenous retrovirus relationships, including conflict and co-option, have been hypothesized among the causes of this difference. However, there has not been studies about the genomic evolutionary signal of constant host–endogenous retrovirus interactions in a long-term scale and how such interactions could lead to the endogenous retrovirus load difference. Through a phylogeny-controlled correlation analysis on ∼5,000 genes between the dN/dS ratio of each gene and the load of endogenous retrovirus in 12 mammals and 21 birds, separately, we detected genes that may have evolved in association with endogenous retrovirus loads. Birds have a higher proportion of genes with strong correlation between dN/dS and the endogenous retrovirus load than mammals. Strong evidence of association is found between the dN/dS of the coding gene for leucine-rich repeat-containing protein 23 and endogenous retrovirus load in birds. Gene set enrichment analysis shows that gene silencing rather than immunity and DNA recombination may have a larger contribution to the association between dN/dS and the endogenous retrovirus load for both mammals and birds. The above results together showing different evolutionary patterns between bird and mammal genes can partially explain the apparently lower endogenous retrovirus loads of birds, while gene silencing may be a universal mechanism that plays a remarkable role in the evolutionary interaction between the host and endogenous retrovirus. In summary, our study presents signals that the host genes might have driven or responded to endogenous retrovirus load changes in long-term evolution.

Key words: evolutionary interaction, host–pathogen interaction, retrovirus, transposable element, *LRRC23*, gene silencing, immunity.

Significance

Studies of endogenous retroviruses (ERVs) in mammals have suggested the important roles they could play in host evolution. However, the roles that ERVs play would not be fully understood without a focus on their nature of being able to proliferate in the host genome. The interaction between the ERV load and the host gene during the host evolution remains unrevealed. Through this comparative study of mammals and birds, we found signals for the constant host–ERV interactions that might shape the divergence of ERV loads between different taxa of vertebrates or acting as a universal restrictive mechanism, which supports the hypothesis that ERVs could have a substantial impact on the evolution of their hosts and potentials of broad functional consequences.

Introduction

An endogenous retrovirus (ERV) is a relic sequence of a retrovirus in the host genome. Retrovirus infection in vertebrates can be traced back to over 450 million years ago with phylogeny of ERVs (Aiewsakun and Katzourakis 2017). ERVs constitute the majority of endogenous viral elements (EVE) in the host genome and form part of a larger assemblage of long terminal repeat (LTR) retrotransposons (Gifford et al. 2018). After entering a host cell, the retrovirus needs to integrate into the host genome to complete its replication cycle (Hayward and Katzourakis 2015). The retroviral sequences integrated in germline cells are potential to transmit to offspring of the host. If such a sequence reaches fixation (or spreads to a certain extent) in the host population, it is considered as an ERV (Feschotte and Gilbert 2012; Hayward and Katzourakis 2015; Kapusta and Suh 2017). ERVs can increase their copy numbers in a genome through retrotransposition or reinfection in the germline (Belshaw et al. 2004; Bannert and Kurth 2006; Magiorkinis et al. 2012). As a result, ERVs constitute 8% of the human genome (Lander et al. 2001) and 11% of the mouse genome (McCarthy and McDonald 2004). The proportion of ERVs ranges from 2.4% to 11.4% of the genome in mammals but is much lower in birds, ranging from 0.16% to 3.57% (Cui et al. 2014). Two possible explanations for the difference in the ERV proportion between mammals and birds are (i) birds prevent retrovirus integration following infection more efficiently or (ii) birds possess more efficient removal mechanisms of ERVs from the genome. The removal mechanisms could include nonallelic homologous recombination (Cui et al. 2014), selection against new insertions, and silencing of insertions (Choi and Lee 2020). The second explanation has been supported so far. At population level, most of new ERV insertions are at very low allele frequencies (Suh et al. 2018; Weissensteiner et al. 2020), but oscine passerines exhibit more ERV activity and diversity than other birds, with many recent insertions of ERVs (Warren et al. 2010; Cui et al. 2014; Smeds et al. 2015; Vijay et al. 2016; Kapusta and Suh 2017; Boman et al. 2019). These observations suggest that constant removal should be responsible for the overall scarcity of ERVs in bird genomes. It was also reported that birds have higher DNA deletion rates in general (Kapusta et al. 2017) and an enrichment of transposable elements (TEs, i.e. transposons) as well as full-length ERVs on the W chromosome in particular, which is nonrecombining across most of its length (Kapusta and Suh 2017; Peona et al. 2021). High DNA deletion rates should lead to small genome sizes of birds, and accordingly, a correlation between genome size and ERV proportion was also observed (Cui et al. 2014). However, this correlation within birds or reptiles seems not strong (Cui et al. 2014). Therefore, DNA deletion alone may not fully explain the low ERV proportion and we need to consider the unique relationship between hosts and ERVs/retroviruses. High or low ERV load in the genome may be involved in adaptation and phenotype evolution of the host. For example, the evolution of *APOBEC3* genes, the antiviral genes that target retroviruses by inducing mutations in their genomes, has been driven by long-running conflicts with ancient retroviruses (Ito et al. 2020). It was also reported that RIG-I, an innate immune receptor for intracellular viral RNA, evolved in higher conservation with increasing ERV load in birds, showing the possibility of association between host immunity evolution and ERV load change (Zheng and Satta 2018).

The relationship between host and ERVs/retroviruses contains two sides: conflict and co-option. Some ERVs can still generate transcripts, proteins, and reverse transcripts, although most ERVs no longer encode intact retroviruses (Stoye 2001). The host has to retain defense against both exogenous retrovirus and provirus/ERV reinsertion, since uncontrolled insertions will threaten the integrity of the host genome. However, host defense in this conflict can be mitigated. Expression of ERVs may require a certain degree of tolerance of the host immune system to avoid chronic inflammation (Kassiotis and Stoye 2016), and ERV expression was found pathogenic in many autoimmune diseases such as systemic lupus erythematosus (SLE) (Ogasawara et al. 2000, 2010) and multiple sclerosis (MS) (Perron et al. 1997; Antony et al. 2004; Tselis 2011). On the other hand, some ERVs are co-opted by the host and involve various physiological functions. These ERVs can promote the evolution of host gene regulatory networks and phenotypes. The most prevalent case in vertebrates is that the proteins encoded by ERV genes *env* or *gag* can restrict

later viral entry or integration (Frank and Feschotte 2017). Another notable case is the important role of ERVs in the origin and evolution of the placenta in mammals. ERV-derived proteins gained a variety of novel functions in placenta such as syncytins that mediate cell fusion to form the barrier from maternal immune cells and exogenous viruses (Sha et al. 2000; Blaise et al. 2003; Dupressoir et al. 2005), and ERVs also constitute a substantial fraction of regulatory elements in placental cells (Chuong et al. 2013). ERVs also provide ligands and regulatory elements in many other biological functions, and the cases of co-option were frequently discovered in mammals (Frank and Feschotte 2017). Frequent discoveries of ERV co-option in mammals may be due to a larger chance of co-option brought by the larger amount of ERVs in mammals than in birds. That is, different ERV loads could shape the ERV–host relationship in different ways. As another possible consequence of this shaping, mammals might have evolved to be more tolerant to ERVs/retroviruses than birds in general.

To better understand the host–ERV relationship during long-term evolution in different vertebrate groups, we conducted a comparative study to examine genomic evolutionary signals of constant host–ERV interaction between host genes and ERV loads in mammals and birds. The genes that have very strong host–ERV interaction could be rare, but there can be many genes having weak or moderate interactions. A single gene having weak or moderate interaction with ERVs may not be distinguishable from some false positive signals, but the true signal will be amplified by multiple genes and possibly become large enough to show the difference between mammals and birds. A gene is a unit of function, and the dN/dS values per gene over the genome offer much richer information than a genome-wide average dN/dS value. In summary, our major aim is to reveal the signals of host–ERV interaction at the genome-wide scale from per gene which possibly corresponds to the difference of ERV loads between mammals and birds.

Results

Bayes Factors that Indicate Associations between Gene dN/dS and the ERV Load (G–E) Are Differently Distributed between Mammals and Birds

We used 5,178 genes of 12 placental mammals and 4,890 genes of 21 birds for studying the association between the genome-wide coding gene dN/dS ratios and the full-length ERV load (i.e. G–E association, for simplicity in this paper) in each of the 2 vertebrate groups. The species used in this study are shown in supplementary table S1, Supplementary Material online. We used a phylogenetic general least square (PGLS) (Pagel 1997, 1999) framework that is implemented in BayesTraits (Meade and Pagel 2017). The degree of association is indicated with a Bayes factor (BF). Conventionally, a logarithm of BF (log BF) greater than two indicates evidence, greater than five indicates strong evidence, and greater than ten indicates very strong evidence (Raffety 1996). In this study, we sometimes use these thresholds to define log BF value levels, but we will be more conservative for the interpretations of thresholds to strengths of evidence. After filtering out highly conserved genes (see Materials and Methods), we proceed with analyses using 5,173 genes for mammals and 4,869 genes for birds. For mammals, 1,262 genes show log BF > 2, 153 genes show log BF > 5, and 9 genes show log BF > 10. For birds, 1,735 genes show log BF > 2, 581 genes show log BF > 5, and 133 genes show log BF > 10.

For both mammals and birds, the distributions of log BF values are unimodal (Fig. 1A). The mode of the kernel density estimation is 0.935 for mammals and 0.661 for birds. To compare the shape of the distributions of the two groups, we normalized the log BF by a modified "standard deviation," for which the mean was replaced with the mode in the equation that calculates the standard deviation, so that the modes of the two samples will be aligned (Fig. 1B to C). The distribution of normalized log BF for birds has a shorter left tail and longer right tail, and the difference between the distributions is supported by the Mann–Whitney U test ($U = 5468783$, P-value $= 2.6 \times 10^{-30}$, one-sided). The result implies that genes having higher levels of G–E association are more frequently seen in birds than in mammals. We next extracted the 2,114 homologous genes between mammals and birds from our data set and use them to repeat the analysis, and the result is similar, but the difference between the two distributions is slightly smaller. This show that (i) the difference is robust and that (ii) rather than the homologous genes, the mammal/bird-specific genes (including the highly divergent genes of which the homology is difficult to determine) make a greater contribution to the difference in the log BF distribution between mammals and birds. The distributions reflect the genome-wide G–E association patterns, which have slight difference between mammals and birds. Birds may have stronger genome-wide evolutionary interaction with ERV load than mammals.

G–E Association Levels Are Mostly Independent from Signs of Correlation or dN/dS But with Exceptions

To investigate the property of the G–E associations, we split the genes into five groups according to the BF values by quintiles (Fig. 2). Group 1 has the lowest log BF values and group 5 has the highest. The root-to-tip dN/dS shows no big difference between the groups, especially when excluding group 1 (see Fig. 2A; P-values of the Kruskal–Wallis tests for the first and the last four groups are shown). The frequencies of the signs of the correlation coefficients (R) are overall stable but slightly shift toward negatives from group 3 to group 4 in birds. That is, negative correlation coefficients are a little

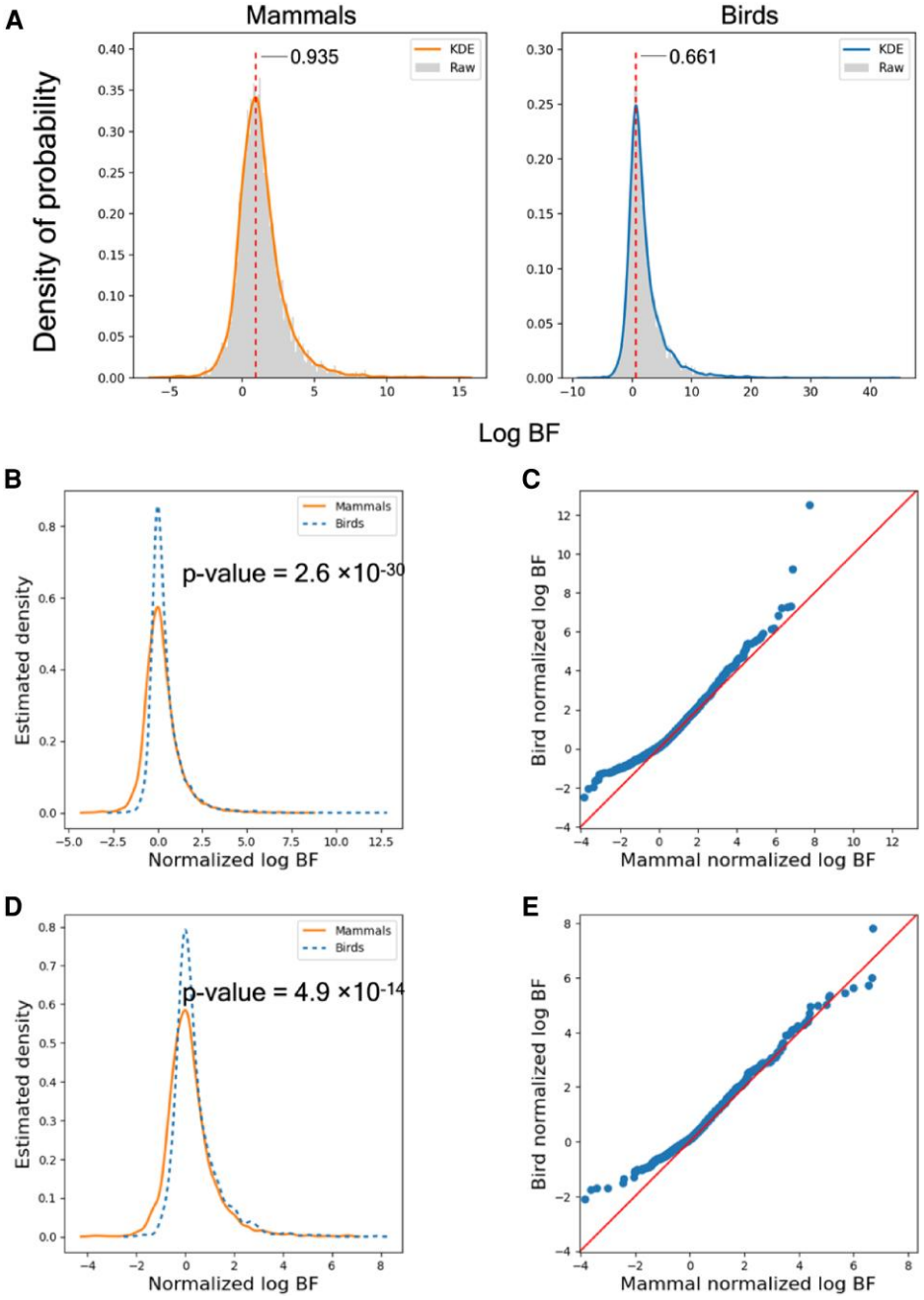

Fig. 1.—Distribution of log BF values that indicate the G–E association of genes in mammals and birds. A) Distribution of original log BF. The value shown pointed to the peak of the distribution is the mode of the kernel estimated density of log BF. B) Distribution of normalized log BF for mammals and birds; the *P*-value of Mann–Whitney *U* test is shown. (C) A Q–Q plot drawn using sampled data points from the estimated density curves in B). D) and E) are plotted in the same fashion as B) and C) but using the 2,114 homologous genes only.

more frequent among relatively high log BF–valued genes, and in birds, it is statistically significant (see Fig. 2B; *P*-values of the χ^2 tests between the two value groups forming a putative "knee point" are shown). The partial increase of negative coefficients with increasing log BF level can be indicative of a "defensive" reaction of host genes against ERV proliferation in birds that conservation of genes was promoted by the pressure of the increased ERV load. However, this trend is reversed if we only look at the genes having very large log BF values. In birds, among the 133 genes with log BF > 10, 98 (74%) show positive association and 35 show negative association. In mammals, in the nine genes with log BF > 10, seven (78%) show positive association and two show negative association. This pattern suggests the possibility

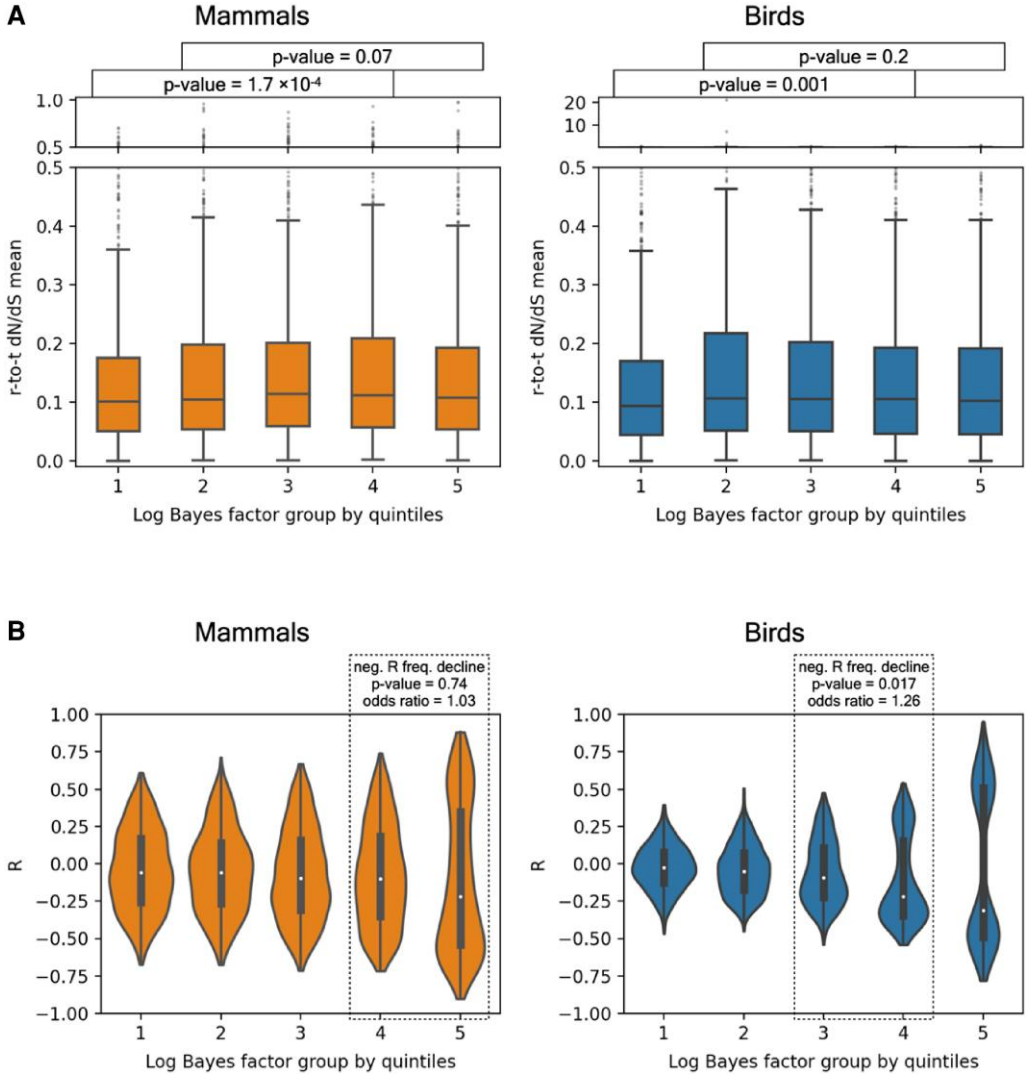

Fig. 2.—Properties of genes grouped by quintiles of log BF values from low to high. A) Root-to-tip dN/dS. *P*-values of Kruskal–Wallis tests for the first and the last four value groups are shown. (B) Correlation coefficient (R). *P*-values of χ^2 tests and odds ratios between the two value groups forming a putative "knee point" are shown.

that fast functional evolution of some genes can be associated with lower ERV loads, but a small number of slowly evolved genes can be more strongly associated with lower ERV load.

G–E Correlations Are Distinguishable from Potential Confounding Factors

We noticed that negative correlation coefficients are more frequent. It was reported that life history traits such as longevity and the body mass are negatively correlated with genome-wide dN/dS in birds and mammals (Romiguier et al. 2013; Figuet et al. 2016; Botero-Castro et al. 2017; Bolívar et al. 2019), while the body mass is positively correlated with the ERV load in mammals (Katzourakis et al. 2014). The effective population size (Ne) or other global factors acting on the genome can result in the above-mentioned two pairs of correlations. These correlations may lead to a negative dN/dS–ERV load correlation at the whole genome level and can be an explanation for our observation that negative G–E correlation coefficients appear more frequent than positive ones. In this situation, many of the negative G–E correlations can be overestimated (while the positive G–E correlations can be underestimated). However, the correlations at per gene level are not necessarily uniform and as same as what is seen at the whole genome averages and will be indicative of specific gene–ERV relationships. Since the body mass may not be independent from ERV loads, it is difficult to control one from the other; nevertheless, we can evaluate whether the body mass has dominated our G–E correlation analysis result,

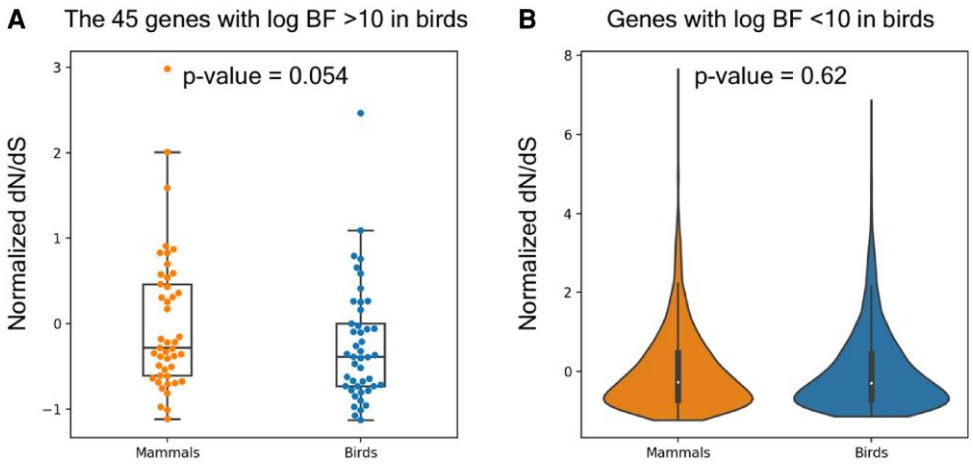

Fig. 3.—Distributions of normalized dN/dS values of 2,114 homologous genes in mammals and birds. A) Distributions for 45 genes that have a log BF value larger than 10 out of the 2,114 homologous genes. B) Distributions for the rest of the 2,114 homologous genes. *P*-values are given by Mann–Whitney *U* test.

especially in those negative correlations. This is achieved by testing whether the body mass is a significant trait in the three-trait regression model of dN/dS (the dependent variable) versus ERV load and body mass. The result shows that the effect of body mass is limited. In mammals, the body mass is not a significant (log BF > 5) trait in the three-trait regression for all the genes. In birds, the body mass is a significant trait in the three-trait regression for 581 (12% of all) genes, but only 22 of the genes (0.4% of all) are suspicious for overestimated G–E correlation, in our criteria, and have log BF values > 5 in the G–E correlation analysis, and the sign of the regression coefficient for the body mass is the same as that of the G–E correlation coefficient (see data S1 in supplementary data, Supplementary Material online), with 13 of the 22 cases being negative correlations. We conclude that most negative G–E correlations shown at the per gene level should not be largely attributed to the effect of the body mass. An explanation could be that the global and indirect negative correlation is too weak, if there is, to overtake the signals of direct dN/Ds–ERV load associations at the per gene level. In fact, the genomic dN/dS–body mass correlation was not detected when incomplete genome data of birds were used (Botero-Castro et al. 2017). Similarly, most of the negative correlation coefficients in our result have very small absolute values and are not likely to cause false positives.

The G–E Associations in Birds Do Not Reappear for the Homologous Genes in Mammals

If the mammalian homologs of the high log BF–valued genes in birds have a significantly skewed distribution of log BF values, it will imply that their evolutionary roles in mammals are correlated to those in birds, in a similar or reversed way. To examine this, we tested whether BF levels show a correlation between mammals and birds. We defined two groups of genes, which are the genes of high BF and those of low BF in birds (bird-high and bird-low) from the 2,114 bird-to-mammal homologous genes, and applied the Mann–Whitney *U* test to the BF level difference in mammals between the two groups. No difference was found in mammals between the genes with log BF > 5 in birds (bird-high genes, $n = 256$) and the genes with log BF < 2 in birds (bird-low genes, $n = 1374$) group ($U = 168{,}006$, P-value = 0.255, two-sided). The high and low log BF–valued genes in birds have the same distribution of BF in mammals, indicating that the G–E associated genes in birds are independent from the homologs in mammals. Similarly, we tested the homologs of the mammal genes of high BF (log BF > 5, $n = 67$) and those of low BF (log BF < 2, $n = 1603$) in birds and no difference was found between the two groups of genes ($U = 49{,}186$, P-value = 0.243, two-sided).

To evaluate and compare the relative functional importance of bird-high genes in birds and in mammals, we took the genes with log BF > 10 ($n = 45$) in birds and computed normalized dN/dS values (z-scores) in mammals and birds, respectively (Fig. 3A). The normalized dN/dS values of the genes with log BF > 10 are generally smaller in birds than in mammals (P-value = 0.054, two-sided). The rest of the 2,114 genes show no obvious difference between birds and mammals (P-value = 0.62, two-sided; Fig. 3). This suggests that the genes having high levels of G–E association in birds are more functionally important for birds than their homologs for mammals.

LRRC23 Shows the Strongest Evidence of the G–E Association in Birds

With caution to the conventional interpretation of log BF values, we performed a control analysis with shuffled ERV

load values randomly assigned to the species for each of the data sets of mammals and birds. The log BF values that are higher than any of the values appearing in the control analysis will be evidence indicating very strong G–E associations. As a result, 30 genes (see supplementary table S2, Supplementary Material online) in birds have log BF values higher than any of the values appearing in the control analysis, but none of the genes in mammals does so.

Among the above 30 very high log BF–valued genes in birds, *LRRC23* has the largest value 43.8, which is 1.35-fold larger than the second largest value 32.5, and the G–E correlation coefficient (R) is a positive value 0.948. In the contrast, the log BF of *LRRC23* in mammals is very small (−0.763) with $R = 0.247$. The result shows that *LRRC23* has high possibility of a positive G–E association in and only in birds.

To understand what mode of selection might underlie this strong association identified, we tested if intensified or relaxed selection of *LRRC23* has occurred in the American crow and the chicken, which have much larger ERV loads than their respective closely related species. Species in the sister clades to the American crow and the chicken are used as reference group (marked in supplementary fig. S1, Supplementary Material online). The test was performed using RELAX (Wertheim et al. 2015). Intensified selection was detected ($K = 2.35$, P-value = 0.043, likelihood ratio = 4.10). The test branches show dN/dS values larger than 1, suggesting that positive selection operated on *LRRC23* and might form the G–E association in the host–ERV evolutionary interaction. A possible scenario of this evolutionary mode is that the increase of ERV load posed a positive selection on the gene to reduce some harmful consequences of the ERV proliferation, e.g. immune disorders. Another scenario could be that positive selection had promoted functional changes of *LRRC23* historically, which led to a relaxed control on ERV proliferation as a byproduct.

However, it is difficult to tell how exactly *LRRC23* could be implicated with bird ERV load evolution due to limited knowledge about the function of *LRRC23*. LRRC23 interacts with CD28 in a pathway essential for the development and homeostasis of regulatory T cells that control autoimmune diseases (Salomon et al. 2000) and has been identified as one of the genes involved in cilia processes (Nevers et al. 2017; Han et al. 2018).

The gene having the second largest log BF is *RCOR3* and the correlation is also positive. Protein RCOR3 enables epigenetically repressing by competitively inhibiting the process enabled by its family members RCOR1 and RCOR2 that facilitates LSD1-mediated nucleosomal demethylation (Upadhyay et al. 2014). In chickens, it is found to be one of important epigenetic modulators for maintaining the integrity of primordial germ cells (Jung et al. 2018). The epigenetic modulation roll of *RCOR3* may have effects on ERV activity in germ cells and may promote an evolutionary interaction between this gene and ERVs.

It is possible that the other genes with quite high log BF values like >5 evolve together with ERV loads, but the strength of the level of G–E association for these genes seems not comparable with *LRRC23*. We list all the analyzed genes with their log BF values ranked from high to low as data S2, supplementary data, Supplementary Material online (mammals), and data S3, supplementary data, Supplementary Material online (birds).

Gene Silencing Shows Importance to the G–E Association for Both Mammals and Birds

Since mammals and birds show a difference in ERV load distribution, relevant biological processes (BPs) may show differences in the pattern of G–E association between the two groups. To understand what BPs are involved in the G–E association with ERV load in mammals and birds, we performed gene set enrichment analyses (GSEA) based on the results of association analyses and ranked the gene sets of gene ontology (GO) terms by the normalized enrichment score (NES). We used the classic GSEA method, which yields conservative results (see Materials and Methods). We also tried a weighted GSEA method, and the result is available in the Supplementary Material online. None of the GO terms show a statistically significant signal in the classic GSEA. Nevertheless, we could use existing knowledge about physiological processes involved in the host restrictions to ERV load to assist in obtaining information from the enrichment scores. Immune response, gene silencing, and DNA recombination are the major known physiological processes involved in the host restrictions to ERV load, including resistance to retroviruses and removal after integration, and thus are potential causal factors of the ERV load difference between mammals and birds. We compared between mammals and birds the rankings of the GO terms representative for the above three BPs, containing keywords "immune response," "silencing," or "recombination" in the 3,261 shared GO terms between the bird and mammal data sets. When sorting all the GO terms by NE scores from low to high, the selected GO terms show that gene silencing ranks very high in the 95th percentile in both mammals and birds (Fig. 4; supplementary fig. S5, Supplementary Material online). Adaptive immune response also ranks relatively high in mammals and birds. The result suggests a role of gene silencing and immune systems in ERV load evolution and calls for further study.

Discussion

Phylogenetic association analyses can reveal evolutionary association between genes and ERV loads, but they cannot differentiate whether accelerated or conserved gene evolution is the cause or consequence of ERV load changes or both ways. Therefore, the association alone can only be indicative of a certain form of interaction, in which the ERV

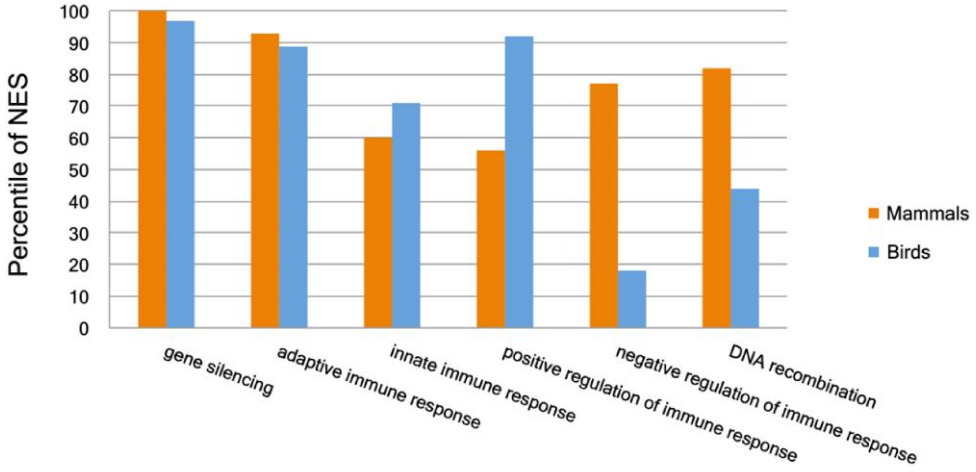

Fig. 4.—Percentiles of NES from low to high of potential ERV load–related GO terms. The charts are based on 3,261 BP GO terms shared between the mammal and bird data sets from the GSEA.

load can be the cause (driving force), the consequence (response), or both (coevolution). ERV load change can be driven by precedent gene evolution. For example, an ecological microbial (including viral) environment or an infectious disease breakout may drive the functional evolution of the immune system of the host and this evolution coincidently changed the reaction toward ERV expression or retroviral infections and thus changed the integration rate of ERVs/retroviruses. ERV load changes can also be the driving force of gene evolution. For example, accumulation of ERVs may increase the risk of disorders (Kassiotis and Stoye 2016) and become selection pressure on some genes, or some ERVs are domesticated by the host (Frank and Feschotte 2017) and change the roles of some host genes. Despite the explanatory limitation of the correlation analysis, the distribution of G–E association statistics and ranking of BPs enriched in larger G–E association statistics, as our result presents, revealed the commonalities and differences between mammals and birds in the host–ERV interactions during the evolution and provided information for hypothesizing about host–ERV interactions.

The GSEA result suggests a role of mechanisms that suppress ERV replication in the host–ERV evolutionary interaction. The levels of enrichment in G–E association of the GO term gene silencing and some GO terms about immune response are higher than that of DNA recombination in both mammals and birds and show differences between the two groups. Genes involved in the cellular response to lipopolysaccharide in birds, including *CD36*, *FN1*, and *MAP2K3*, show considerable degrees of G–E association. Since lipopolysaccharide is a major component of the cell wall of gram-negative bacteria, those genes might evolve in response to bacteria environment and might coincidently result in an influence on the immunity against retroviruses. On the other hand, the GO term DNA recombination in mammals shows a medium ranking of enrichment in G–E association (see Fig. 4). This supports the possibility that recombination-dependent deletion can influence the ERV load through an effect on deletion rate, as proposed in previous studies (Cui et al. 2014; Kapusta et al. 2017; Ji et al. 2022). In contrast to mammals, the GO term DNA recombination shows a quite low ranking of enrichment in G–E association in birds (see Fig. 4). We speculated that the recombination machinery might have a larger effect on the load of TEs other than full-length ERVs resulting from their different relationships with hosts during vertebrate evolution. We took a quick look into the total LTR retrotransposon (including ERVs, some solo LTRs, and truncated LTR retrotransposons) load (differently defined as the length percentage of the host genome as the literature offered) in birds. We found that compared with ERV load, LTR retrotransposon load reported (Cui et al. 2014) ranks slightly higher for DNA recombination as well as for negative regulation of immune response but slightly lower in positive regulation of immune response (supplementary fig. S2, Supplementary Material online). However, the load of LTR retrotransposons reported (Zhang et al. 2014) shows a drastic decline of ranking. Their LTR retrotransposon estimation includes many solo LTRs and truncated LTR retrotransposons and is likely responsible for this drastic decline, since solo LTRs and truncated LTR retrotransposons are immobile ERV remnants of which the copy numbers were not affected by gene silencing (supplementary fig. S3, Supplementary Material online). We also found that the distributions of load–rate association for LTR retrotransposon load in the host genome are slightly left-shifted compared with that for ERV load (supplementary fig. S4, Supplementary Material online), suggesting the content of solo LTRs and truncated LTR retrotransposons may have weaker evolutionary interaction with the host genome of birds.

In conclusion, from this comparative study between mammals and birds, we discovered (i) a higher proportion of genes showing relatively strong correlation between their dN/dS and the ERV load in birds, (ii) strong evidence of positive association between the dN/dS of the gene *LRRC23* and ERV loads in birds, and (iii) a remarkable role of gene silencing played in the evolutionary interaction between the host and ERVs. The above findings form a glimpse of the mysterious evolutionary force—the constant host–ERV interaction that control the ERV loads, shaping the divergence of ERV loads between different taxa of vertebrates or acting as a universal restrictive mechanism. The signals for constant host–ERV interaction supported the hypothesis that ERVs could have a substantial impact on the evolution of their hosts and potentials of broad functional consequences.

Materials and Methods

Selected Mammal and Bird Species

Association analyses for dN/dS and traits require reliable ERV copy number and genome size data, as well as species representative of large ranges of birds or mammals. Since this study uses published data, the quality of ERV load estimation depends on the reliability of published data of ERV copy number and genome size. To ensure that the estimated ERV load is reliable for association analyses, we selected the species of which the ERV copy numbers were obtained from whole genome data of good quality. Our threshold for good quality is set as scaffold N50 $\geq$ 0.5 Mb (5×10^5 bp) for a mammal and scaffold N50 $\geq$ 1 Mb (10^6 bp) for a bird. Under this condition, 12 placental mammals and 21 birds (for the species, see supplementary table S1, Supplementary Material online) were selected from whole genome–sequenced species for association analyses. All data treatments and analyses were performed separately for birds and mammals.

Coding Sequence Alignments

We retrieved aligned coding sequences of orthologous genes of 48 birds and outgroup species from Jarvis et al. (Jarvis et al. 2015) and those of 39 placental mammals and outgroup species from Douzery et al (Douzery et al. 2014). First, we removed the sequences containing premature termination codons (PTCs). Next, a gene was used in this study if its alignment data meets the following criteria: it contains sequences of all the selected species, and it contains at least one species of the outgroup of the selected species. The outgroups for birds include the American alligator (*Alligator mississippiensis*), green sea turtle (*Chelonia mydas*), green anole lizard (*Anolis carolinensis*), and human (*Homo sapiens*). The outgroups for placental mammals include tammar wallaby (*Macropus eugenii*), Tasmanian devil (*Sarcophilus harrisii*), gray short-tailed opossum (*Monodelphis domestica*), and platypus (*Ornithorhynchus anatinus*). In this way, we got alignments of 5178 genes for mammals and 4890 genes for birds.

Identification of Homologous Genes

Homologous genes were identified by BLASTn searches implemented in Blast2GO v5 (Götz et al. 2008). The consensus sequences of bird genes were used as queries for the BLASTn searching against the reference database made of the consensus sequences of mammal genes. The top-hit mammal sequence for each bird sequence was taken as the homologous gene for that bird sequence. If multiple bird sequences get the same top-hit mammal sequence, then only the bird sequence having the largest *e*-value is kept.

ERV Loads

ERV load is defined as the copy number divided by genome size in Gb for each species. Copy numbers of ERVs in bird species were retrieved from Cui et al. (Cui et al. 2014), where ERVs were defined as ERV-like sequences that unambiguously matched known retroviral proteins. Copy numbers of ERVs in mammal species were retrieved from Katzourakis et al (Katzourakis et al. 2014), where ERV were defined as ERV-like sequences containing the *pol* gene. In both of those, two literatures the ERVs were screened using tBLASTn and a library of representative viral protein sequences built by each research group. Note that this means that non-full-length ERVs such as solo LTRs (remnants of ERVs after intraelement nonallelic homologous recombination) and partially deleted ERVs are likely not included in these ERV copy number estimates. Genome sizes corresponding to the genome versions used in the above literatures were retrieved from National Center for Biotechnology Information (NCBI) Genome (NCBI 2018) and Archive Ensembl (Ensembl 2018). The retrieved copy numbers of ERVs and genome sizes are included in supplementary data, Supplementary Material online.

G–E Correlation Analysis

We applied a method of phylogenetic traits correlation analysis to test for an association between the dN/dS of coding genes and ERV load (G–E association, for simplicity in this paper), which is based on the PGLS model (Pagel 1997). This method estimates the correlation of the variations of two continuous traits, and the effect of phylogenetic relationships on the variations is controlled. Applied to this study, the root-to-tip dN/dS is the first trait and the ERV load is the second trait. The phylogenetic information used, including tree topologies and divergence times, is from Jarvis et al. 2014 (Jarvis et al. 2014) for birds and TimeTree (Kumar et al. 2017) for mammals. Evidence of association is evaluated with the BF or its logarithmic form (log BF), which is the ratio of marginal likelihood of correlated model over

that of noncorrelated model estimated from Markov chain Monte Carlo (MCMC). Positive or negative association with a detected gene can be reflected in the positive or negative sign of the average correlation coefficient (R) of the correlated model. This method is implemented in BayesTraits v3.0.1 (Meade and Pagel 2017). In the case that a BayesTraits analysis would not complete in a reasonable time period, the gene as the input was manually checked and we confirmed that all such genes are highly conserved; therefore, the correlation analysis requiring variation should not succeed. These conserved genes were then excluded for the analyses in this study. Root-to-tip dN/dS (ratio of nonsynonymous substitution rate over synonymous substitution rate) (Pagel 1997) were estimated under a branch model with PAML v4.9 (Yang 2007) using all the sequences (species) available in the coding sequence alignment data for each gene. Kernel density estimations of log BF distributions are performed with SciPy v1.7.1.

Brown rat (*Rattus norvegicus*) was absent in the original data set for *LRRC23* of mammals. To compare with mammals, we additionally collected the brown rat *LRRC23* sequence from NCBI Nucleotide database (Nucleotide ID: NM_001394347.1) and performed BayesTrait analysis for *LRRC23* for mammals.

Control Analyses

The test for evaluating whether the body mass dominates the G–E correlation was performed using BayesTrait v3.0.1 (Meade and Pagel 2017). The test model is a three-trait regression model that uses the ERV load and the body mass as two independent variables and the dN/dS of the gene as the independent variable. The null model is a two-trait regression model that is the same as the test model but without the body mass. A BF is computed to measure the marginal likelihood of the test model over that of the null model.

The control analysis of G–E correlation was performed using BayesTrait v3.0.1 (Meade and Pagel 2017). The true ERV load values are shuffled and randomly assigned to the species used in the true G–E correlation analysis, for mammals and birds, separately. The other part of the analysis is exactly the same as the true G–E correlation analysis.

Test for Intensified or Relaxed Evolution

With a limitation of data availability, only five branches were used for the test. The American crow and the chicken are the only cases in this study which have apparently larger ERV loads compared with their respective sister groups (marked in supplementary fig. S1, Supplementary Material online); therefore, the leaf branches in these two groups were chosen to be the test group, and the branches in their sister groups became the reference group. The test was performed using RELAX, which evaluates the bidirectional change of dN/dS of the test group from the reference group (Wertheim et al. 2015).

GO Annotation and GSEA

GO annotation and GSEA was performed in Blast2GO v5 (Götz et al. 2008). GSEA (Subramanian et al. 2005) were performed using preranked lists of rescaled log BF. The classic method of GSEA takes the rank alone into account, while the weighted method additionally takes the expression differentiation values into account. Classic GSEA was recommended for preranked data (Subramanian et al. 2005) since the magnitude of differences in log BF values may bias the weighted GSEA caused by the impact of occasional extreme values. In addition to the classic GSEA, we also used a weighted method in which the expression differentiation is represented with a rescaled value of log BF to reduce impact of extreme values (see supplementary materials and methods, Supplementary Material online, and supplementary results, Supplementary Material online).

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

We thank Drs. So Nakagawa, Hirohisa Kishino, and Tatsuya Ota for kindly offering advice to this study. Computations were partially performed on the National Institute of Genetics (NIG) supercomputer at Research Organization of Information and Systems (ROIS) National Institute of Genetics.

Funding

The article processing charge (APC) was funded by SOKENDAI (The Graduate University for Advanced Studies). The Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) and the China Scholarship Council (CSC) jointly offered the Japanese Government Scholarship to one of the authors, W.Z., during this study.

Data Availability

Alignments of mammal coding sequences were downloaded from https://www.orthomam.univ-montp2.fr/orthomam_v8/html/index.php?article=Request (Douzery et al. 2014). Alignments of bird coding sequences were downloaded from http://gigadb.org/dataset/101041 (Jarvis et al. 2015).

Literature Cited

Aiewsakun P, Katzourakis A. Marine origin of retroviruses in the early Palaeozoic era. Nat Commun. 2017:8(1):13954. https://doi.org/10.1038/ncomms13954.

Antony JM, Van Marle G, Opii W, Butterfield DA, Mallet F, Yong VW, Wallace JL, Deacon RM, Warren K, Power C. Human endogenous retrovirus glycoprotein-mediated induction of redox reactants causes oligodendrocyte death and demyelination. Nat Neurosci. 2004:7(10):1088–1095. https://doi.org/10.1038/nn1319.

Bannert N, Kurth R. The evolutionary dynamics of human endogenous retroviral families. Annu Rev Genomics Hum Genet. 2006:7(1): 149–173. https://doi.org/10.1146/annurev.genom.7.080505.115700.

Belshaw R, Pereira V, Katzourakis A, Talbot G, Paces J, Burt A, Tristem M. Long-term reinfection of the human genome by endogenous retroviruses. Proc Natl Acad Sci USA. 2004:101(14):4894–4899. https://doi.org/10.1073/pnas.0307800101.

Blaise S, de Parseval N, Benit L, Heidmann T. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. Proc Natl Acad Sci USA. 2003:100(22):13013–13018. https://doi.org/10.1073/pnas.2132646100.

Bolívar P, Guéguen L, Duret L, Ellegren H, Mugal CF. GC-biased gene conversion conceals the prediction of the nearly neutral theory in avian genomes. Genome Biol. 2019:20(1):5. https://doi.org/10.1186/s13059-018-1613-z.

Boman J, Frankl-Vilches C, da Silva dos Santos M, de Oliveira EHC, Gahr M, Suh A. The genome of blue-capped cordon-bleu uncovers hidden diversity of LTR retrotransposons in zebra finch. Genes (Basel). 2019:10(4):301. https://doi.org/10.3390/genes10040301.

Botero-Castro F, Figuet E, Tilak MK, Nabholz B, Galtier N. Avian genomes revisited: hidden genes uncovered and the rates versus traits paradox in birds. Mol Biol Evol. 2017:34(12):3123–3131. https://doi.org/10.1093/molbev/msx236.

Choi JY, Lee YCG. Double-edged sword: the evolutionary consequences of the epigenetic silencing of transposable elements. PLoS Genet. 2020:16(7):e1008872. https://doi.org/10.1371/journal.pgen.1008872.

Chuong EB, Rumi MAK, Soares MJ, Baker JC. Endogenous retroviruses function as species-specific enhancer elements in the placenta. Nat Genet. 2013:45(3):325–329. https://doi.org/10.1038/ng.2553.

Cui J, Zhao W, Huang Z, Jarvis ED, Gilbert MTP, Walker PJ, Holmes EC, Zhang G. Low frequency of paleoviral infiltration across the avian phylogeny. Genome Biol. 2014:15(12):539. https://doi.org/10.1186/s13059-014-0539-3.

Douzery EJP, Scornavacca C, Romiguier J, Belkhir K, Galtier N, Delsuc F, Ranwez V. OrthoMaM v8: a database of orthologous exons and coding sequences for comparative genomics in mammals. Mol Biol Evol. 2014:31(7):1923–1928. https://doi.org/10.1093/molbev/msu132.

Dupressoir A, Marceau G, Vernochet C, Benit L, Kanellopoulos C, Sapin V, Heidmann T. Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. Proc Natl Acad Sci USA. 2005:102(3): 725–730. https://doi.org/10.1073/pnas.0406509102.

Ensembl. 2018. Archive Ensembl. Available from: http://www.ensembl.org/info/website/archives/index.html

Feschotte C, Gilbert C. Endogenous viruses: insights into viral evolution and impact on host biology. Nat Rev Genet. 2012:13(4):283–296. https://doi.org/10.1038/nrg3199.

Figuet E, Nabholz B, Bonneau M, Mas Carrio E, Nadachowska-Brzyska K, Ellegren H, Galtier N. Life history traits, protein evolution, and the nearly neutral theory in amniotes. Mol Biol Evol. 2016:33(6): 1517–1527. https://doi.org/10.1093/molbev/msw033.

Frank JA, Feschotte C. Co-option of endogenous viral sequences for host cell function. Curr Opin Virol. 2017:25:81–89. https://doi.org/10.1016/j.coviro.2017.07.021.

Gifford RJ, Blomberg J, Coffin JM, Fan H, Heidmann T, Mayer J, Stoye J, Tristem M, Johnson WE. Nomenclature for endogenous retrovirus (ERV) loci. Retrovirology 2018:15(1):59. https://doi.org/10.1186/s12977-018-0442-1.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 2008:36(10):3420–3435. https://doi.org/10.1093/nar/gkn176.

Han X, Xie H, Wang Y, Zhao C. Radial spoke proteins regulate otolith formation during early zebrafish development. FASEB J. 2018:32(7):3984–3992. https://doi.org/10.1096/fj.201701359R.

Hayward A, Katzourakis A. Endogenous retroviruses. Curr Biol. 2015:25(15):R644–R646. https://doi.org/10.1016/j.cub.2015.05.041.

Ito J, Gifford RJ, Sato K. Retroviruses drive the rapid evolution of mammalian APOBEC3 genes. Proc Natl Acad Sci U S A. 2020:117(1): 610–618. https://doi.org/10.1073/pnas.1914183116.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT, et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 2014:346(6215): 1320–1331. https://doi.org/10.1126/science.1253451.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT, et al. Phylogenomic analyses data of the avian phylogenomics project. GigaScience 2015:4(1):4. https://doi.org/10.1186/s13742-014-0038-1.

Ji Y, Feng S, Wu L, Fang Q, Brüniche-Olsen A, DeWoody JA, Cheng Y, Zhang D, Hao Y, Song G, et al. Orthologous microsatellites, transposable elements, and DNA deletions correlate with generation time and body mass in neoavian birds. Sci Adv. 2022:8(35): eabo0099. https://doi.org/10.1126/sciadv.abo0099.

Jung HG, Hwang YS, Park YH, Cho HY, Rengaraj D, Han JY. Role of epigenetic regulation by the REST/CoREST/HDAC corepressor complex of moderate NANOG expression in chicken primordial germ cells. Stem Cells Dev. 2018:27(17):1215–1225. https://doi.org/10.1089/scd.2018.0059.

Kapusta A, Suh A. Evolution of bird genomes—a transposon's-eye view. Ann N Y Acad Sci. 2017:1389(1):164–185. https://doi.org/10.1111/nyas.13295.

Kapusta A, Suh A, Feschotte C. Dynamics of genome size evolution in birds and mammals. Proc Natl Acad Sci USA. 2017:114(8): E1460–E1469. https://doi.org/10.1073/pnas.1616702114.

Kassiotis G, Stoye JP. Immune responses to endogenous retroelements: taking the bad with the good. Nat Rev Immunol. 2016:16(4):207–219. https://doi.org/10.1038/nri.2016.27.

Katzourakis A, Magiorkinis G, Lim AG, Gupta S, Belshaw R, Gifford R. Larger mammalian body size leads to lower retroviral activity. PLoS Pathog. 2014:10(7):e1004214. https://doi.org/10.1371/journal.ppat.1004214.

Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol. 2017:34(7):1812–1819. https://doi.org/10.1093/molbev/msx116.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. Initial sequencing and analysis of the human genome. Nature 2001:409(6822): 860–921. https://doi.org/10.1038/35057062.

Magiorkinis G, Gifford RJ, Katzourakis A, De Ranter J, Belshaw R. Env-less endogenous retroviruses are genomic superspreaders. Proc Natl Acad Sci USA. 2012:109(19):7385–7390. https://doi.org/10.1073/pnas.1200913109.

McCarthy EM, McDonald JF. Long terminal repeat retrotransposons of *Mus musculus*. Genome Biol. 2004:5(3):R14. https://doi.org/10.1186/gb-2004-5-3-r14.

Meade A, Pagel M. 2017. BayesTraits V3.0.1. Available from: https://www.evolution.reading.ac.uk/BayesTraitsV3.0.1/BayesTraitsV3.0.1.html.

NCBI. 2018. NCBI Genome. Available from: https://www.ncbi.nlm.nih.gov/genome

Nevers Y, Prasad MK, Poidevin L, Chennen K, Allot A, Kress A, Ripp R, Thompson JD, Dollfus H, Poch O, et al. Insights into ciliary genes and evolution from multi-level phylogenetic profiling. Mol Biol Evol. 2017:34(8):2016–2034. https://doi.org/10.1093/molbev/msx146.

Ogasawara H, Hishikawa T, Sekigawa I, Hashimoto H, Yamamoto N, Maruyama N. Sequence analysis of human endogenous retrovirus clone 4-1 in systemic lupus erythematosus. Autoimmunity. 2000:33(1):15–21. https://doi.org/10.3109/08916930108994105.

Ogasawara H, Kageyama M, Yamaji K, Takasaki Y. The possibility that autoimmune disease can be induced by a molecular mimicry mechanism between autoantigen and human endogenous retrovirus. Lupus 2010:19(1):111–113. https://doi.org/10.1177/0961203309106767.

Pagel M. Inferring evolutionary processes from phylogenies. Zool Scr. 1997:26(4):331–348. https://doi.org/10.1111/j.1463-6409.1997.tb00423.x.

Pagel M. Inferring the historical patterns of biological evolution. Nature 1999:401(6756):877–884. https://doi.org/10.1038/44766.

Peona V, Palacios-Gimenez OM, Blommaert J, Liu J, Haryoko T, Jønsson KA, Irestedt M, Zhou Q, Jern P, Suh A. The avian W chromosome is a refugium for endogenous retroviruses with likely effects on female-biased mutational load and genetic incompatibilities. Philos Trans R Soc B Biol Sci. 2021:376(1833):20200186. https://doi.org/10.1098/rstb.2020.0186.

Perron H, Garson JA, Bedin F, Beseme F, Paranhos-Baccala G, Komurian-Pradel F, Mallet F, Tuke PW, Voisset C, Blond JL, et al. Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis. Proc Natl Acad Sci U S A. 1997:94(14):7583–7588. https://doi.org/10.1073/pnas.94.14.7583.

Raffety AE. Hypothesis testing and model selection. In: Gilks WR, Richardson S, Spiegelhalter D, editors. Markov chain monte carlo in practice. New York (NY): Springer; 1996. p. 163–188.

Romiguier J, Ranwez V, Douzery EJP, Galtier N. Genomic evidence for large, long-lived ancestors to placental mammals. Mol Biol Evol. 2013:30(1):5–13. https://doi.org/10.1093/molbev/mss211.

Salomon B, Lenschow DJ, Rhee L, Ashourian N, Singh B, Sharpe A, Bluestone JA. B7/CD28 costimulation is essential for the homeostasis of the CD4+CD25+ immunoregulatory T cells that control autoimmune diabetes. Immunity 2000:12(4):431–440. https://doi.org/10.1016/S1074-7613(00)80195-8.

Sha M, Lee X, Li XP, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY, Edouard P, Howes S, et al. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature 2000:403(6771):785–789. https://doi.org/10.1038/35001608.

Smeds L, Warmuth V, Bolivar P, Uebbing S, Burri R, Suh A, Nater A, Bureš S, Garamszegi LZ, Hogner S, et al. Evolutionary analysis of the female-specific avian W chromosome. Nat Commun. 2015:6(1):7330. https://doi.org/10.1038/ncomms8330.

Stoye JP. Endogenous retroviruses: still active after all these years? Curr Biol. 2001:11(22):R914–R916. https://doi.org/10.1016/S0960-9822(01)00553-X.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005:102(43):15545–15550. https://doi.org/10.1073/pnas.0506580102.

Suh A, Smeds L, Ellegren H. Abundant recent activity of retrovirus-like retrotransposons within and among flycatcher species implies a rich source of structural variation in songbird genomes. Mol Ecol. 2018:27(1):99–111. https://doi.org/10.1111/mec.14439.

Tselis A. Evidence for viral etiology of multiple sclerosis. Semin Neurol. 2011:31((03|3)):307–316. https://doi.org/10.1055/s-0031-1287656.

Upadhyay G, Chowdhury AH, Vaidyanathan B, Kim D, Saleque S. Antagonistic actions of Rcor proteins regulate LSD1 activity and cellular differentiation. Proc Natl Acad Sci U S A. 2014:111(22):8071–8076. https://doi.org/10.1073/pnas.1404292111.

Vijay N, Bossu CM, Poelstra JW, Weissensteiner MH, Suh A, Kryukov AP, Wolf JBW. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. Nat Commun. 2016:7(1):13195. https://doi.org/10.1038/ncomms13195.

Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A, Searle S, White S, Vilella AJ, Fairley S, et al. The genome of a songbird. Nature 2010:464(7289):757–762. https://doi.org/10.1038/nature08819.

Weissensteiner MH, Bunikis I, Catalán A, Francoijs KJ, Knief U, Heim W, Peona V, Pophaly SD, Sedlazeck FJ, Suh A, et al. Discovery and population genomics of structural variation in a songbird genus. Nat Commun. 2020:11(1):3403. https://doi.org/10.1038/s41467-020-17195-4.

Wertheim JO, Murrell B, Smith MD, Pond SLK, Scheffler K. RELAX: detecting relaxed selection in a phylogenetic framework. Mol Biol Evol. 2015:32(3):820–832. https://doi.org/10.1093/molbev/msu400.

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007:24(8):1586–1591. https://doi.org/10.1093/molbev/msm088.

Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. Science 2014:346(6215):1311–1320. https://doi.org/10.1126/science.1251385.

Zheng W, Satta Y. Functional evolution of avian RIG-I-like receptors. Genes (Basel). 2018:9(9):456. https://doi.org/10.3390/genes9090456.

Associate editor: Cristina Vieira