EVOLUTIONARY BIOLOGY

Orthologous microsatellites, transposable elements, and DNA deletions correlate with generation time and body mass in neoavian birds

Yanzhu Ji[1,2], Shaohong Feng[3,4,5,6], Lei Wu[1,7], Qi Fang[3,8], Anna Brüniche-Olsen[9], J. Andrew DeWoody[10], Yalin Cheng[1], Dezhi Zhang[1], Yan Hao[1], Gang Song[1], Yanhua Qu[1], Alexander Suh[11,12], Guojie Zhang[4,5,6,8,13,14]*, Shannon J. Hackett[2]*, Fumin Lei[1,7,15]*

The rate of mutation accumulation in germline cells can be affected by cell replication and/or DNA damage, which are further related to life history traits such as generation time and body mass. Leveraging the existing datasets of 233 neoavian bird species, here, we investigated whether generation time and body mass contribute to the interspecific variation of orthologous microsatellite length, transposable element (TE) length, and deletion length and how these genomic attributes affect genome sizes. In nonpasserines, we found that generation time is correlated to both orthologous microsatellite length and TE length, and body mass is negatively correlated to DNA deletions. These patterns are less pronounced in passerines. In all species, we found that DNA deletions relate to genome size similarly as TE length, suggesting a role of body mass dynamics in genome evolution. Our results indicate that generation time and body mass shape the evolution of genomic attributes in neoavian birds.

INTRODUCTION

Mutations are the primary material for evolution. Various mutations have been identified that underlie both adaptive and nonadaptive evolution. In addition, the rates of mutation accumulation vary wildly across species [e.g., (*1*)] and covary with life history traits (see below). For example, the rate of nucleotide substitution covaries with body mass in mammals, birds, and poikilotherms (*2*–*4*). Universal mechanisms, such as effects related to metabolism and/or generation time, have been posited to explain these data (*5*, *6*).

The correlation between body mass and mutation accrual is believed to reflect two covariates of body mass, both of which are related to the mechanisms of how mutations are generated (*2*). Mutations that can be inherited arise in germline cells, either during DNA replication or when DNA is damaged but fails to be repaired. It is thus expected that, when compared across species, the accumulation of mutations would reflect the rate of DNA replication in germline cells and/or the degree of DNA damage and its repair, depending on the type of mutations (*7*). For mutations that are generated during DNA replication, the rate of germline cell replication (i.e., the number of germline cell replications over time) dictates how fast these mutations accumulate. If we assume that species all undergo a relatively fixed number of cell cycles during one generation, then the generation time would be negatively related to the rate of germline cell replication and thus negatively affect how many mutations are generated within a given time frame. In contrast, the accumulation of damage-caused mutations that are not repaired efficiently should track time elapsed or the strength of mutation genesis. Recent studies have emphasized the previously overlooked effects of DNA damage on primate mutation accumulation (*8*–*10*). As exogenous factors that cause DNA damage tend to relate to environment and are highly unpredictable and transient, one of the main endogenous factors, the generation of reactive oxygen species (ROS), is tightly linked to the physiology of animals (*11*, *12*). ROS are the by-product of mitochondrial oxidative phosphorylation (*13*, *14*) and are determined by the level of aerobic activities, which is further related to body mass [e.g., (*15*, *16*)].

We chose birds as our study system because the accumulated data on life history and phylogenomics make cross-species comparative studies feasible. Here, we focus on the effect of life history traits on three genomic attributes that reflect accumulated mutations in avian genomes. The genomic attributes include orthologous microsatellite length and transposable element (TE) length as representatives of mutations that at least partially reflect DNA replication. We also included DNA deletions as another genomic attribute to represent mutations that are caused by DNA damage and those mediated by nonallelic homologous recombination (NAHR), the latter possibly related to the density of TEs (*17*). Specifically, microsatellites mutate when DNA strands fail to match each other during DNA synthesis, forming either insertions or deletions of repeat units. Furthermore, the mutations in microsatellites are believed to be biased toward insertions (*18*–*21*). We also include TE length, because the heritable mobilization of TEs, while related to the transcriptional and/or (retro)transpositional activities of TEs themselves and host defenses

[1]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China. [2]Negaunee Integrative Research Center, Field Museum of Natural History, Chicago, IL 60605, USA. [3]BGI-Shenzhen, Beishan Industrial Zone, Shenzhen 518083, China. [4]Future Health Laboratory, Innovation Center of Yangtze River Delta, Zhejiang University, Jiaxing 314100, China. [5]Evolutionary and Organismal Biology Research Center, Zhejiang University School of Medicine, Hangzhou, China. [6]Liangzhu Laboratory, Zhejiang University Medical Center, Hangzhou 311121, China. [7]University of the Chinese Academy of Sciences, Beijing 100049, China. [8]Villum Centre for Biodiversity Genomics, Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen DK-2200, Denmark. [9]Section for Computational and RNA Biology, Department of Biology, University of Copenhagen, Copenhagen DK-2200, Denmark. [10]Departments of Forestry and Natural Resources and Biological Sciences, Purdue University, West Lafayette, IN 47906, USA. [11]School of Biological Sciences, Organism and Environment, University of East Anglia, NR4 7TU, Norwich, UK. [12]Department of Organismal Biology, Systematic Biology, Evolutionary Biology Centre (EBC), Science for Life Laboratory, Uppsala University, Uppsala SE-752 36, Sweden. [13]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China. [14]Women's Hospital, School of Medicine, Zhejiang University, Shangcheng District, Hangzhou, 310006, China. [15]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650201, China.
*Corresponding author. Email: guojiezhang@zju.edu.cn (G.Z.); shackett@fieldmuseum.org (S.J.H.); leifm@ioz.ac.cn (F.L.)

against TE invasions, is highly dependent on resetting the epigenetic marks in germline cells during replication that likely influence the rate of TE incorporation into avian genomes [e.g., (*22*)]. The length of orthologous microsatellites and TEs (under the assumption of similar activities) should reflect rounds of germline cell replication and be associated with generation time. In comparison, the formation of DNA deletions is frequently accompanied by imperfect repair of DNA double-strand breaks (*23*–*25*).

Together, the compiled data on avian generation time (*26*), body mass (*27*), and phylogenomics (*28*, *29*) enabled us to explicitly test the association between avian life history traits (i.e., generation time and body mass) and genomic evolution. Our a priori hypotheses were that (i) generation time is expected to negatively correlate with accumulated mutations with shared origins related to DNA replication (i.e., both orthologous microsatellite length and TE length) and that (ii) body mass is expected to negatively correlate with mutations associated with improper DNA repair (i.e., DNA deletions). Our data revealed that both hypotheses were mostly supported in nonpasserine Neoaves. In passerines, we found a distinct pattern with a correlation between generation time and deletion length that required alternative explanations.

RESULTS

Quantification of genomic attributes

To quantify orthologous microsatellites, we first downloaded/gathered raw sequencing reads from online sources (tables S1 and S2). Of all 278 species, after read trimming, we collected 11.3- to 25.4-Gb DNA sequences per species, with a median of 21.1 Gb.

Our effort to identify orthologous microsatellite first discovered 421 orthologous microsatellite loci between chicken and zebra finch, of which 331 loci have annotated genes in either or both assemblies, and all annotated genes were matched (data file S1). With this pipeline applied on the first 43 species (table S1), orthologous microsatellite loci were found in 23 of 45 supermotifs with the motif sizes of 2 to 4 base pairs (bp) (table S3). Last, we identified 695 loci using reads of the rest of the 235 species (table S2). After genotyping the orthologous microsatellites and filtering out 204 loci with invariant length across species, we found the species-averaged orthologous microsatellite length, which represents the number of repeat units that differs from the most common allele (Δ units), ranged from 0.006 to 0.229 for Neoaves that were analyzed subsequently (Fig. 1A and data file S2). TE length, calculated as the proportions of TEs [data from (*29*)] times the assembly sizes, ranged from 0.073 to 0.400 Gb (Fig. 1A and data file S2).

By analyzing the "hal" file of avian 363-way whole-genome alignment, mean DNA deletion length ranged from 77.3 to 182.8 bp per 1000 bp in Neoaves (Fig. 1A and data file S2). Note that sequences extracted in this way only included orthologous sequences that are present in the chicken, so that any chicken-specific deletions were not included (fig. S1). Furthermore, any lineage-specific insertions from chicken, Galliformes, Galloanserae, and Neognathae are present as well (fig. S1). As it is impossible to polarize Neognathae-specific insertions without a nonavian outgroup, we therefore excluded species of Galloanserae and Paleognathae from our analyses. After overlapping the remaining neoavian species and those with available life history traits, we ended up with 233 species for subsequent statistical analyses.

Phylogenetic comparative analysis

Because inspection of pairwise scatterplots (e.g., Fig. 1B) further revealed that TE-related trends were not linear, we split our data into nonpasserines and passerines to better explain these data. Twelve nonpasserine species had a substantial TE burden, as the log(TE length) passed third quartile + 1.5 × interquartile range (IQR) and corresponded to the total percentage of TEs ranging from 12.85 to 31.47%. The species included all eight species sampled from Piciformes, the eurasion hoopoe (*Upupa epops*) and the common scimitarbill (*Rhinopomastus cyanomelas*) in Bucerotiformes, the hooded pitta (*Pitta sordida*) in Passeriformes, and the squirrel cuckoo (*Piaya cayana*) in Cuculiformes (Fig. 1). As the trend in nonpasserine birds was different from the trend in this clade without the species with high TEs (Fig. 1B), we further excluded the species with high TE length in nonpasserines as the third group. Last, we tested our a priori hypotheses in the three groups of species: (i) nonpasserine birds, (ii) nonpasserine birds without species with high TE length, and (iii) passerine birds (without species with high TE or deletion length; the species with high deletion length became apparent once passerines were separated). Unfortunately, we could not run models focused on species with high TE length because of the limited sample size.

We built three piecewise structural equation models (PSEMs) that represent three scenarios among the relationships of life history traits, genomic attributes, and genome (or assembly) size (Fig. 2). Model 1 describes the general basic relationships among life history traits, and between genomic attributes and genome size, without relationships between life history traits and genomic attributes (Fig. 2A). Notably, we included a correlation between TE length and deletion length, as Kapusta *et al.* (*17*) proposed an "accordion" model of genome size evolution that suggests that TE insertions drive the formation of deletions through NAHR. Model 2, as described by our a priori hypotheses, predicts that both orthologous microsatellite length and TE length are related to generation time, and DNA deletion is related to body mass, in addition to the relationships in model 1 (Fig. 2B). Last, model 3 represented a full model in which each of the three genomic attributes is related to both body mass and generation time (Fig. 2C). Note that, although the PSEMs were defined by a series of hypothetical causational (single-ended) and correlated (double-ended) relationships, here, we interpreted the relationships as correlations. Our results, as measured by correlations, do not demonstrate causations but rather lend support to these hypothetical causal models (*30*).

After running all three models using our three subsets of data, we first noticed that models 2 and 3 were supported by our data: Models 2 and 3 were both supported by all nonpasserines, whereas model 2 was supported by nonpasserines without outliers and model 3 was supported by passerines (Table 1). By examining standardized regression coefficients (coef.) as a way to measure the strength and direction of the relationships in PSEMs (for results, see table S4), we checked whether the modeling results were consistent with our hypotheses. Basically, we found that it depended on the dataset whether our three focal correlations (i.e., among generation time or body mass and genomic attributes) were significant or not. In nonpasserines, all of our three focal correlations were significant in model 2 (generation time versus orthologous microsatellite length: coef. = −0.24, $P = 0.003$; generation time versus TE length: coef. = −0.23, $P = 0.0005$; body mass versus deletion length: coef. = −0.31, $P < 0.001$), but the one between generation time and orthologous microsatellite length was insignificant in model 3 (generation time versus orthologous microsatellite length: coef. = −0.08, $P > 0.05$; generation time versus TE length: coef. = −0.19, $P = 0.04$; body mass versus deletion

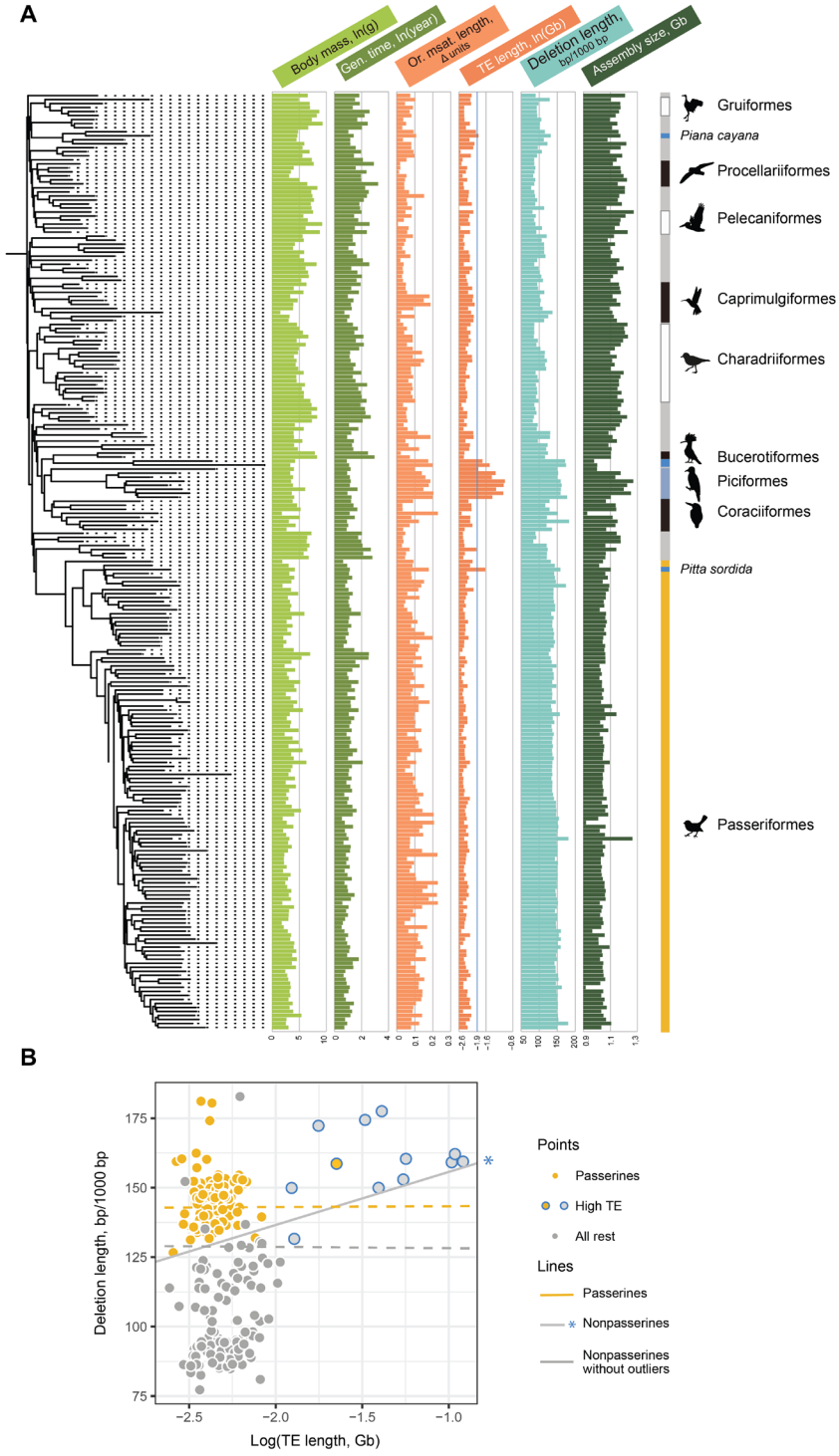

Fig. 1. Evolution of genomic attributes (orthologous microsatellite length, TE length, and deletion length), together with life history traits and genome (assembly) size. (**A**) The overview of genomic attributes evolution, together with body mass, generation time, and assembly size, across the evolutionary tree. Orders (or taxonomic groups) that include ≥5 species are highlighted with black or white bars. Species with elevated TEs are highlighted with blue bars. Silhouettes are from phylopic.org, most of them unchanged but a few are mirrored, with credit to F. Sayol, S. Wegner-Larson, B. McCormish (photo by Avenue), "annaleeblysse," M. Michaud, S. Traver, and C. Schmidt, under Public Domain Dedication 1.0 (http://creativecommons.org/publicdomain/zero/1.0/), Creative Commons Attribution-ShareAlike 3.0 Unported (https://creativecommons.org/licenses/by-sa/3.0/), or Creative Commons Attribution 3.0 Unported (https://creativecommons.org/licenses/by/3.0/). The figure is generated by the R package ggtree (*74*). Or. msat. length, orthologous microsatellite length. (**B**) Scatterplots between TE length and deletion length, highlighting the distinct trends between passerines (yellow) and nonpasserines (gray) and between those with (blue circles) or without (no circles) high TE contents. Lines represent modeling results, with passerines represented by the yellow line, nonpasserines by the gray line with a blue asterisk, and nonpasserines without outliers by the gray line. Dashed line represents statistical insignificant results, and solid lines represent significant results.

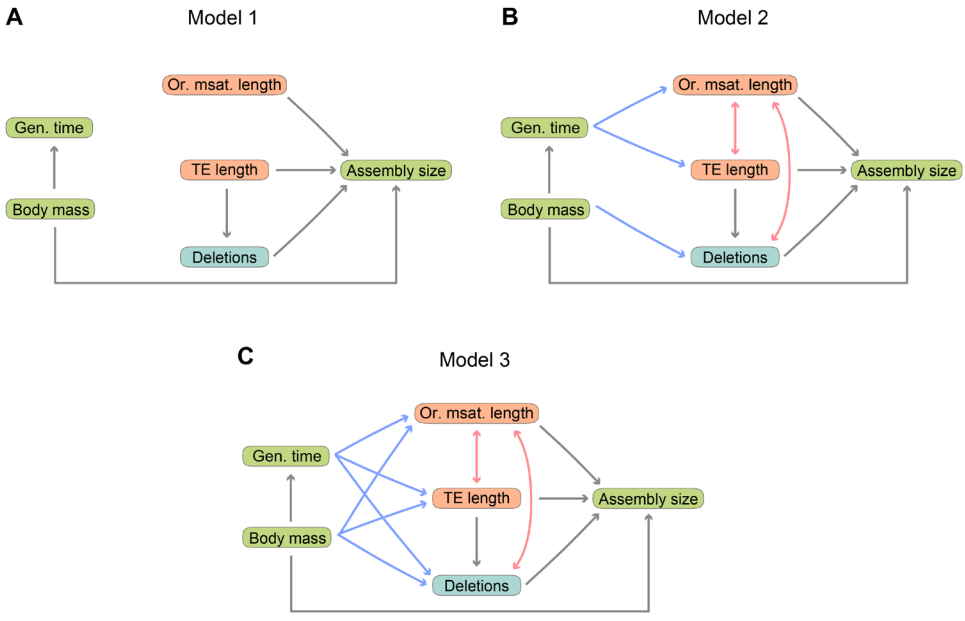

Fig. 2. Models that are tested by PSEMs. The models describe how avian genomic attributes, including orthologous microsatellite length, TE length, and deletion length, evolved under the covariation of DNA loss and gain (the arrow between TE length and deletion length), effect of life history traits on genomic attributes (arrows between body mass, generation time, and genomic attributes), and contributions of genomic attributes to genome size (or assembly size). Gray arrows indicate relationships that are common across all three models. Blue/red arrows refer to model-specific relationships. Single-directional arrows indicate relationships that are presumably causational, and bidirectional arrows indicate correlated errors among variables. (**A**) Model 1: The evolution of genomic attributes is not related to life history traits. (**B**) Model 2: Life history traits correlate with the evolution of genomic attributes as predicted by a priori hypothesis. (**C**) Model 3: A full model that makes every possible connection between life history traits and genomic attributes.

Table 1. Summary of piecewise structural equation model (PSEM) results across datasets. In the table, df denotes the degree of freedom, *C* denotes *C* statistic, *P* denotes *P* values for d-separation tests (note that *P* values of >0.05 indicate proper model fit), AICc denotes Akaike information criterion corrected for small sample sizes, and ΔAICc denotes changes in ΔAICc. Models with ΔAICc < 2 are shown in bold.

Nonpasserines

Models	df	*C*	*P*	AICc	ΔAICc
Model 2	**8**	**6.66**	**0.57**	**54**	**0**
Model 3	**2**	**0.45**	**0.79**	**55.4**	**1.4**
Model 1	12	55.7	<0.05	89.8	35.8

Nonpasserines without outliers

Models	df	*C*	*P*	AICc	ΔAICc
Model 2	**8**	**5.73**	**0.67**	**54**	**0**
Model 3	2	0.92	0.63	57.5	3.5
Model 1	12	50	<0.05	84.5	30.5

Passerines

Models	df	*C*	*P*	AICc	ΔAICc
Model 3	**2**	**1.73**	**0.42**	**57.4**	**0**
Model 1	12	34.61	<0.05	66.2	8.8
Model 2	8	25.53	<0.05	77.2	19.8

length: coef. = −0.27, P = 0.004; Figs. 3A and 4). In nonpasserines without outliers, we found that all three correlations were significant (generation time versus orthologous microsatellite length: coef. = −0.19, P = 0.03; generation time versus TE length: coef. = −0.29, P = 0.005; body mass versus deletion length: coef. = −0.40, P < 0.001; Figs. 3B and 4). Last, in passerines, the model showed that all three aforementioned correlations were insignificant (all P > 0.05; table S4), but the correlation between generation time and deletions was significant (coef. = −0.38, P = 0.001; Fig. 3C, fig. S2, and table S4). The correlation between TE length and deletion length, reflecting that increases in TE length cause larger values of deletion length, was significant only in all nonpasserines when including outliers (model 2: coef. = 0.26, P = 0.001; model 3: coef. = 0.25, P = 0.002; table S4; Fig. 3, A versus B), consistent with the accordion model that TEs may facilitate the removal of DNA by NAHR (*17*). Given that NAHR may result in larger deletions, this is further supported by the presence of longer deletions in the TE-rich clades than those in the rest clades (fig. S3).

Our modeling results also showed that the two relationships reflecting correlated errors between orthologous microsatellite length versus TE length and deletion length, respectively, were also consistent across datasets and models, although they were stronger in all nonpasserines (Fig. 3, fig. S4, and table S4). Last, comparable coefficients of deletion length and TE length to the assembly size in nonpasserines were identified in all models (Fig. 3, fig. S5, and table S4).

Potential impact of sequencing depth on PSEMs

To explore a possible effect of the sequence data on our results, we used sequencing depth as an indicator of sequencing quantity and quality. We reasoned that most genomes were generated by the

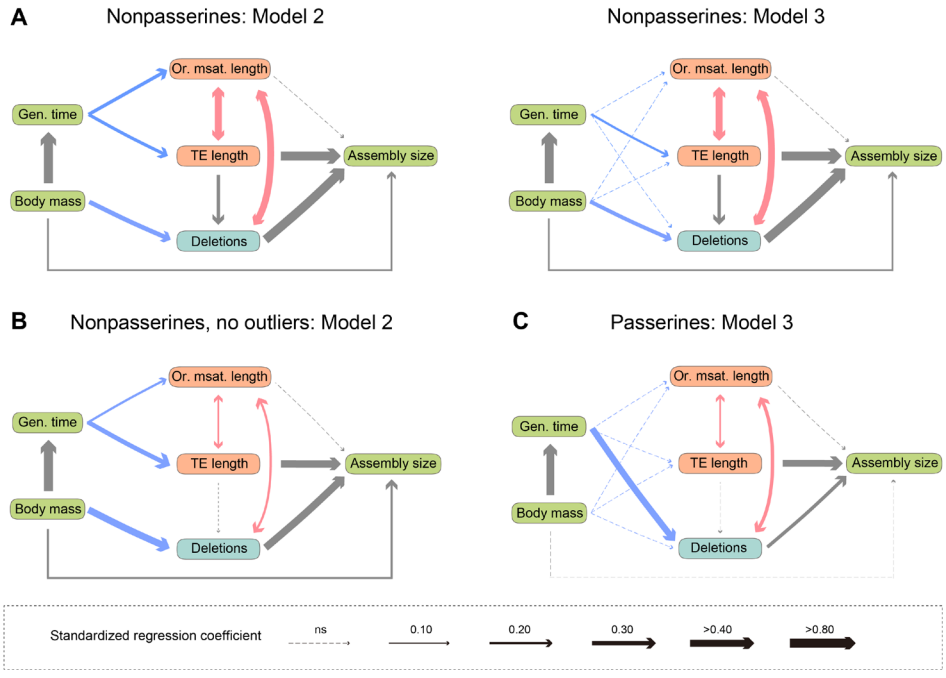

Fig. 3. PSEM results in different taxa being tested. The panels represent modeling results in (**A**) nonpasserines, (**B**) nonpasserines without outliers, and (**C**) passerines. Gray arrows are used to indicate shared correlations among models, whereas red/blue arrows indicate model-specific correlations with positive/negative directions. The thickness of arrows represents the strength of standardized regression coefficient (also see table S4). This figure shows that the correlations of our three predictions are supported in nonpasserines but not in passerines. ns, not significant.

B10K project using similar sequencing platforms and assembly methods, so that the variation in sequencing and assembly methods should be relatively minor. In comparison, the depth of focal species ranged from 24X to 122X, with a median of 49X, suggesting that the impact of depth should be considered.

Further analyses showed that passerines and nonpasserines did not differ in sequencing depth [phylogenetic generalized least squares (PGLS): $r = 0.004$, $P > 0.05$; fig. S6A]. Similarly, species with high TE length did not have higher sequencing depth (PGLS: $r = 0.03$, $P > 0.05$; fig. S6B). When correlating sequencing depth with genomic attributes and assembly size using each dataset, we did find depth to be positively correlated with orthologous microsatellite length in nonpasserines without outliers (PGLS: $r = 0.17$, $P = 0.05$; fig. S6C and table S5) outliers and depth to be positively correlated with assembly size in nonpasserines with and without outliers (PGLS: $r = 0.23$, $P = 0.01$ and $r = 0.24$, $P = 0.01$, respectively; fig. S6D and table S5).

To exclude the effect of depth on models, we incorporated depth as a confounding variable in two of the three datasets that showed significant correlations, assuming that increasing sequencing depth would result in longer orthologous microsatellites and larger assembly sizes (fig. S7). The updated PSEMs showed that the addition of depth had limited influence on the standardized regression coefficients and significance (fig. S8 and table S4). This evidence suggests that, while sequencing depth does have impact on some genomic attribute and assembly size, the relationships among body mass, generation time, genomic attributes, and assembly size remain unchanged.

Evidence supporting the correlation between deletions and body mass

We further explored whether between-node changes in ancestral deletions correlated with between-node changes in ancestral body mass. Briefly, ancestral deletions were defined as deletions that can be confidently assigned to a branch in the phylogeny where all descendent species (but no others) harbor the deletion. As running through all 100,000 1-kb blocks to assign the phylogenetic position of each deletion was time-consuming, we randomly subsampled 20,000 1-kb blocks for subsequent analyses. The result with 10,000 blocks (fig. S9A) was consistent with that of 20,000 blocks (Fig. 4D), suggesting that we have sampled sufficient data for accurate statistical inferences. After filtering out blocks with missing species, we ended up with 34.8% ($n = 6{,}958$) blocks for further analyses. Deletions that can be confidently traced back to a branch consisted of 79.8% of all deletions ($n = 3{,}244{,}677$), with deletions of 0 to 20 bp dominating the overall length distribution (proportion of deletions of 0 to 20 bp: 90.8%; fig. S9B).

Pearson's correlations revealed that the length of deletions mapped back to each branch of the phylogeny was negatively correlated with the change in ancestral body mass as estimated by maximum likelihood (see Materials and Methods; $r = -0.27$, $P < 0.001$; Fig. 4D). The correlation in nonpasserines was stronger than that in passerines (nonpasserines: $r = -0.33$, $P < 0.001$; passerines: $r = -0.12$, $P = 0.02$; Fig. 4D), consistent with our PSEMs where, in passerines, a correlation between deletion and body mass was too weak to be detected. The correlation in nonpasserines was higher than that without the species/clades with high TE length ($r = -0.30$, $P < 0.001$).

We also examined whether the correlation between life history and genomic attributes holds in other taxa when such data were available. In mammals where both whole-genome alignment (*31*) and body mass data (*32*) were available for 186 species (data file S2), a negative trend between body mass and deletion length was also detected ($r = -0.17$, $P < 0.01$; fig. S10).

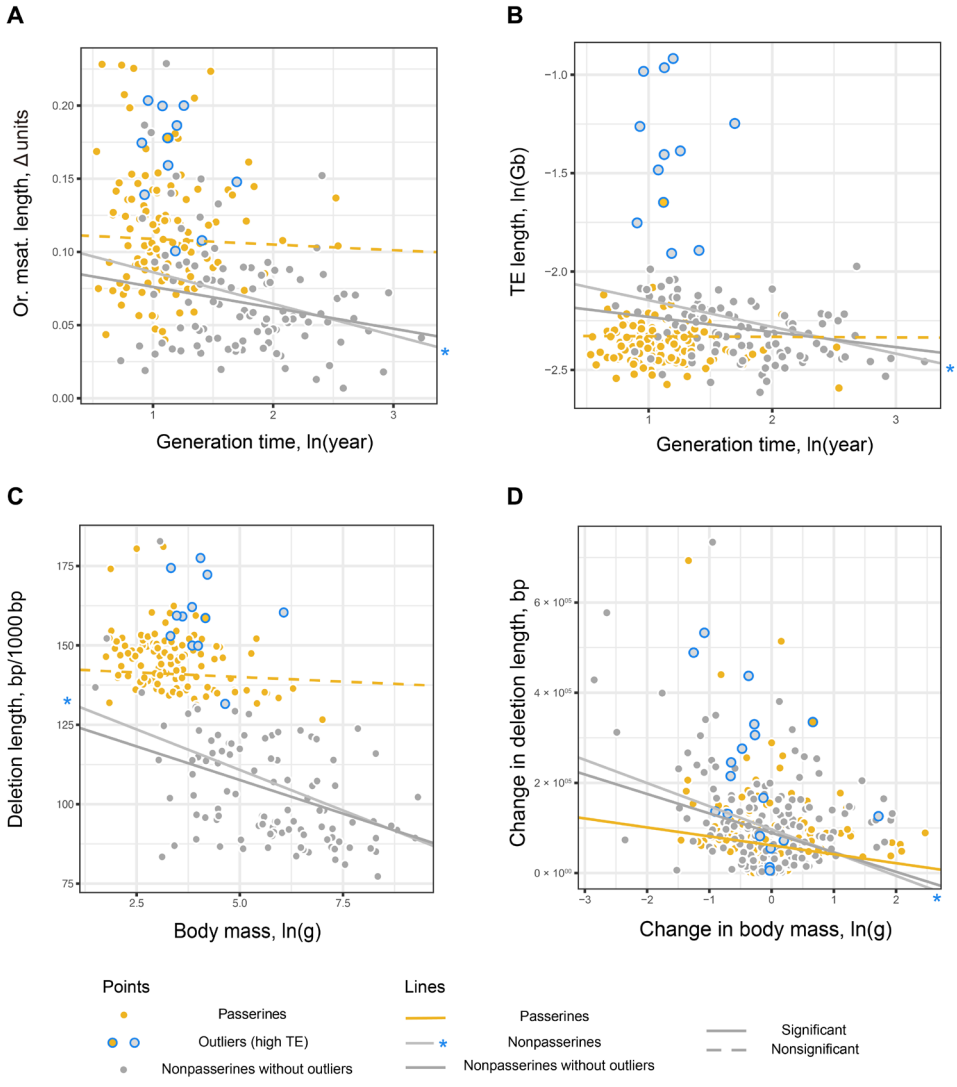

Fig. 4. Intercorrelation among life history traits and three genomic attributes. (**A**) Generation time [ln(year)] and orthologous microsatellite length (Δ units); (**B**) generation time [ln(year)] and TE length [ln(Gb)]; (**C**) body mass [ln(g)] and deletion length (bp/1000 bp); and (**D**) change in ancestral body mass [ln(g)] and change in deletion length (bp). In these plots, we used colors to differentiate species from nonpasserines (gray) and passerines (yellow), and those with high TE length (circled with blue) and without (no circles). Regression lines from PGLS models are plotted in each, with gray lines with asterisks representing the dataset with all nonpasserine species, gray lines without asterisks representing species without high TE lengths, and yellow lines representing passerine species. Significant regressions are represented by solid lines, and insignificant results are represented by dashed lines. This figure shows the three correlations between life history traits and the genomic attributes that we focused on (A to C) and the correlation between ancestral body mass and change in deletion length (D) further supported the correlation between deletion length and body mass (C).

Possible effects of longevity on DNA deletions

Mutations may be repaired more efficiently in long-lived species [e.g., (*33*, *34*)]. In our dataset, we found that long-lived species from Procellariiformes (as represented by albatrosses, petrels and shearwaters, and storm petrels) exhibit the shortest deletions compared to most other birds, which is evident after controlling for phylogeny (fig. S11A). Using a subset of 36 species with data on maximum longevity of wild animals collected from AnAge database (data file S2) (*35*), we observed a positive correlation between the residuals of deletion length versus the residuals of maximum longevity, both of which were calculated when regressed against body mass ($r = -0.34$, $P = 0.01$; fig. S11B). However, the lack of data of maximum longevity for all of our focal species excluded us from incorporating longevity into the PSEM.

DISCUSSION

The intercorrelations among genomic attributes, life history traits, and genome size among 233 neoavian bird species show that our three a priori hypotheses are mostly supported in nonpasserines, whereas in passerines, the patterns are less pronounced. We also show that DNA deletions covary with TEs in nonpasserines when species with high TE content are included, which is consistent with the study of Kapusta *et al.* (*17*). Our results highlight how life history traits influence cellular processes that ultimately shape genomic architecture in birds.

Plausible effects of effective population size

Our study has focused on the effect of the generation of mutations, which is reflected by generation time and body mass, on the

accumulation of mutations. However, processes that affect the fixation and/or repair of mutations may also play significant roles in the observed pattern. First, whether the probability of fixation has an effect on the accumulation of mutations depends on the neutrality of mutations. For neutral mutations in a diploid population of the effective population size of N_e with mutation rate of μ, the fixation rate of mutations is calculated as the rate of neutral mutations ($2N_e\mu$) multiplied by the probability of fixation ($1/2N_e$). Therefore, the fixation rate of neutral mutations completely depends on the mutation rate as the effect of N_e is canceled out.

In contrast, the fixation of slightly deleterious mutations depends on N_e (*36*, *37*). Previous studies have shown that populations with smaller N_e fix more slightly deleterious mutations compared to those with larger N_e (*38*–*41*). As species with small N_e often have large body mass and long generation time, a positive correlation between body mass and the number of fixed mutations would be expected, which is in contrast to our observation. Further, for cross-species comparisons where the direction and strength of selection are different across species, N_e is not expected to exert directional impact on the accumulation of mutations. It depends on the biology of each species whether an individual insertion or deletion is deleterious or not. Considering that TEs and deletions lead to incremental genome size changes, it is uncertain whether the species are under the same direction and strength of selection with regard to genome size (or a correlated trait). Second, N_e can also positively affect the time for mutations to fix, as it takes (on average) $4N_e$ generations for neutral mutations to be fixed in a population (*42*). It thus may take much longer time for neutral mutations to reach fixation in populations/species with larger N_e than those in species with smaller N_e. Assuming the same evolutionary time and mutation rate across species, fewer mutations are predicted to be fixed in species with large N_e than those in small N_e. This prediction translates to a positive correlation between body mass (or generation time) and accumulated mutations, which is contradictory to our observation. Together, this evidence suggests that N_e is not the major driver responsible for the pattern exhibited in our analyses.

Deletions are associated with the evolution of flight, body mass, and genome size

Previous studies have implicated the role of flight in reducing genome size, as evidenced by smaller genome sizes across flighted vertebrates including birds, bats, and pterosaurs (*43*–*45*). It has been hypothesized that bird genomes are constrained due to the metabolic demands of flight, as (i) smaller cells have higher surface-to-volume ratios and thus higher metabolic rates and (ii) small cells also tend to have small genomes (*43*, *46*, *47*). In other studies, a declining trend for both body mass (*48*–*50*) and genome size (*51*) has been documented in theropods, the lineage from which birds have evolved. Both the associations between genome size versus flight and body mass can be potentially explained by the accumulation of deletions. Both reduction in body mass and flight are tightly associated with shifts in metabolic rates, either through metabolic allometry (*52*, *53*) or through energetic demands of flight. The increase in metabolism subsequently induced double-strand breaks where deletions are formed. As it is apparent in our study that longer deletions tend to relate to smaller genome sizes (or smaller assembly sizes; fig. S5), it is thus probable that the accumulation of deletions is the link between both body mass reduction [also see (*54*)] and genome size reduction before the evolution of flight.

The correlation between body mass and DNA deletions may be applicable in taxa beyond birds. In mammals, we have also observed a negative trend between body mass and DNA deletions (fig. S10). In addition, the correlations between life history traits, genomic attributes, and genome size can also be found in amphibians, where much larger genome sizes with much larger variance are exhibited compared to birds. For example, one of the smallest genomes in frogs, found in ornate burrowing frog (*Platyplectrum ornatum*), was characterized by short introns, deletion bias in TEs, and suppressed TE activities (*55*). Moreover, *P. ornatum* and two other frog species with small genomes converge on life history traits including rapid tadpole development (*55*). On the contrary, the gigantic genome sizes of salamanders have been shown to have slow removal rates of TEs (*56*, *57*), whereas, recently, a link between large genome size and slow developmental time has been identified (*58*). These studies suggest a link between life history traits (including body mass), genomic attributes, and genome size evolution in different vertebrate groups.

The distinct pattern in Passeriformes

Compared to PSEM results in nonpasserines, the results in passerines show a distinct pattern in that none of the three expected correlations are detected. Instead, a correlation between generation time and deletions is supported (Fig. 2C and fig. S2). Recently, a study of mathematical models revealed that the accumulation of damage-caused mutations may track cell division rate as well, if some of them are repaired efficiently before DNA replication (*59*). Still, it remains unknown why this correlation is significant only in passerines, i.e., whether passerines are better in terms of DNA repair than nonpasserines (see the "Limitation of the results" section below). Meanwhile, the correlated errors between orthologous microsatellite length versus TE length and deletions are still present, suggesting that some unknown factors may be still in control of the correlations in passerines.

Notably, the observation that the assembly sizes tend to scatter once deletion lengths are longer (fig. S5B) suggests uncharacterized genome expansion events in these passerine species. For instance, in nonpasserines, most species with elevated TE levels tend to deviate from the negative correlation between assembly size and deletion. This pattern can be explained by the presence of TEs, which increased the assembly sizes. In comparison, in passerines, only one species that deviated from the negative correlation (hooded pitta, *P. sordida*) has been characterized to have more TEs than most other species. It remains to be characterized why other passerine species deviated from the trend, including white-throated oxylab (*Oxylabes madagascariensis*), chipping sparrow (*Spizella passerine*), common sunbird-asity (*Neodrepanis coruscans*), red-billed oxpecker (*Buphagus erythrorhynchus*), and Crossley's vanga (*Mystacornis crossleyi*).

One of the closely related species of the chipping sparrow, the yellow-throated bunting (*Emberiza elegans*), has exceptionally high levels of repetitive sequence in a genome assembly sequenced with PacBio and scaffolded with Hi-C (PRJNA778594). In addition, three of the five species (*O. madagascariensis*, *N. coruscans*, and *M. crossleyi*) are island birds endemic to Madagascar. Island birds are known to have small N_e and thus experience stronger levels of drift (*60*). It also happens that another endemic bird appeared to accumulate more microsatellite DNA in Piciformes (*61*), suggesting a potential link between island endemics and DNA expansion.

Limitation of the results

Overall, our results could be limited by a general lack of variation of the life history traits under investigation. Compared to mammals, birds (especially neoavian birds) are generally smaller and exhibit less variance in body masses, potentially due to the demand of flight [e.g., (*62*)]. Similarly, avian genomes sizes also show limited variation (*63*). It is thus expected that detecting correlations between constrained variables is inherently hard for our focal species. Nevertheless, the fact that we were able to find correlations between body mass and genomic attributes, and between genomic attributes and genome sizes, indicates that these signals have been strong enough to be detected. The distinct pattern between passerines and nonpasserines (Fig. 3) can also be explained by the lack of variation in body mass, generation time, genome size, and genomic attributes of passerines.

Second, the study is conducted using genomic data mostly assembled from short reads. The annotation of TEs and the estimation of genome size by assembly size, for example, can be greatly improved by third-generation sequencing (*64*). However, we think that DNA deletions (especially smaller ones; fig. S9B) are less impaired by the sequencing technology, as their identification relied on sequence similarity of nonrepetitive DNA. Regardless, we expect that the increasing available assemblies using third-generation sequencing technique will extensively advance our understanding of the repetitive part of avian genomes.

Last, because of limited data availability regarding avian longevity, we were unable to include the longevity of birds into our PSEMs to test the role of longevity systematically. Compared to the rapidly advancing field of genome sequencing, the accumulation of "traditional" data on life history of birds proceeds much slowly but continues to be critical for meaningful interpretation of genomic data.

MATERIALS AND METHODS

Identification and genotyping orthologous microsatellite loci

To account for microsatellite underrepresentation in whole-genome assemblies due to their repetitiveness, we used raw sequencing reads to identify and quantify avian microsatellites. For the 43 species published by Zhang *et al.* (*28*), we downloaded raw reads from National Center for Biotechnology Information Sequence Read Archive libraries with a targeted size of approximately 15 Gb (~12X given an average genome size of 1.3 Gb; table S1). We tried to download libraries with similar sequencing strategy as far as possible, with paired-end sequencing of read length of 100 bp and an insertion size of 800 bp. For an additional 235 bird species published by Feng *et al.* (*29*), we extracted raw sequencing reads with a target coverage of 20X (table S2) and a target insertion size of 500 bp. The reads were trimmed using Trimmomatic (*65*), from which microsatellites of motif sizes of 1 to 7 nucleotides (nt) were identified using the Tandem Repeat Finder (*66*).

We identified orthologous microsatellite loci across species using the microsatellites uncovered in the sequencing reads. Because of the limitation of a downstream software [STR-FM (short tandem repeat profiling using a flanking-based mapping approach)], we only focused on microsatellite with motif sizes of 2 to 4 nt for orthologous microsatellites. We first identified orthologous microsatellites in the 43 species published by Zhang *et al.* (*28*). We relied on the sequence similarity among the 15-bp flanking regions of orthologous loci to help identify them (for further information, see Supplementary Methods). To identify orthologous microsatellites in the 235 species (*29*), we searched for the already-identified orthologous loci from the 43 species in each of the 235 species. We imputed genotypes of the orthologous microsatellite loci from sequencing reads using STR-FM (see Supplementary Methods for details) (*67*). For each locus, we calculated the change in allele size by subtracting the allele size that was represented most frequently across species from the allele size of each allele of each locus in each species. For heterozygous loci, the changes in allele size were averaged. Last, a loci-averaged number was calculated to represent the average change in allele size of the orthologous microsatellites for each species. All related custom scripts for microsatellite quantification are available at https://github.com/yanzhu-ji/avian-msats.

Calculation of DNA deletion

To calculate DNA deletion length, we used multiple 1-kb syntenic alignments (n = 100,000) that were randomly sampled from the 363-way genome alignment generated by the B10K project using Progressive Cactus (*29*, *68*). Random blocks of sequences that are neither genes nor repetitive regions were extracted using hal2maf with the genome of chicken (*Gallus gallus*) as a reference (refer to Supplementary Methods for details) (*68*). Note that sequences extracted in this way only included orthologous sequences that are present in the chicken, so that any chicken-specific deletions were not included (fig. S1). Furthermore, any lineage-specific insertions from chicken, Galliformes, Galloanserae, and Neognathae are present as well (fig. S1). As it is impossible to polarize Neognathae-specific insertions without an outgroup, we therefore excluded species of Galloanserae and Paleognathae from our analyses. For each species, deletion lengths were calculated as 1000 (bp) subtracted by the syntenic alignment length, which was further averaged across blocks. Note here that the deletion length is a relative measurement instead of absolute length, as the extent of the consistent chicken-specific (and closely related species) insertions was unknown.

Collection of other biological data from literature

Overall, we compiled a dataset of generation time, body mass, assembly size, TE length, and a phylogeny from existing literature. Specifically, we collected generation time from Bird *et al.* (*26*) and body mass from the *CRC Handbook of Avian Body Masses* (*27*) and log-transformed these data before analyses. We collected the assembly size and total TE proportion (including LINE/SINE/LTR/DNA/RC/Unknown) from Feng *et al.* (*28*). The total TE length was calculated by multiplying the total TE proportion and assembly size. For the phylogeny, we relied on Burleigh *et al.* (*69*).

Phylogenetic comparative analyses

To select and describe the model among life history traits, genomic attributes, and genome size, we used PSEM as implemented in the R package piecewiseSEM (*70*), given the phylogenetic nonindependence of our data. Basically, we set up three scenarios for PSEM to test against (Fig. 1). Furthermore, we ran the models for three datasets: The first dataset included all neoavian nonpasserine birds, the second dataset excluded species with TE length as identified as outliers judged by the criteria of third quartile + 1.5 × IQR, and the third dataset consisted of all passerine birds. For each dataset, we identified the model that was best supported by the data as defined by ΔAICc < 2.

Potential impact of sequencing depth on PSEMs

We obtained sequencing depth from Feng *et al.* (*29*). To test whether sequencing depth has any effect on (or can help explain) our results, we first tested whether passerines have higher sequencing depth than nonpasserines and whether species with higher TE length have higher depth than species without using PGLS (implemented in R package phylolm) (*71*) with the status of species coded into dummy variables of 0 or 1. We also tested whether depth is correlated with any of the genomic attributes or assembly size using PGLS tests. If depth was found to be correlated with any of the variables, then PSEMs were updated with depth as a confounding variable (fig. S7).

Comparison of ancestral states of deletions and body mass

To identify whether the detected (if any) correlations between genomic attributes and life history traits can be further traced back to ancestral nodes in a phylogeny, we reconstructed ancestral states of life history traits of each node using fastAnc function in phytools (*72*), and the changes in life history traits of each branch were calculated using the adjacent node and tip values. The ancestral states of one of the genomic attributes, deletions, were reconstructed by in-house scripts. Specifically, after blocks of sequences were extracted from hal files, we first defined the deletions by the start and end positions in the block, assuming that deletions with same starts and ends originate from one deletion event. We next located the phylogenetic branch of each deletion by the branch preceding their most common ancestor. We discarded the blocks that fail to cover all species in phylogeny or the deletions that have evolved multiple times independently across the tree. The change in life history traits and the change in deletion length were subsequently correlated using Pearson correlations in R. All the above statistical analyses were performed in R v4.1.2 (*73*).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abo0099

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. W.-H. Li, L. Ellsworth, Darrell, J. Krushkal, B. H.-J. Chang, D. Hewett-Emmett, Rates of nucleotide substitution in primates and rodents and the generation–time effect hypothesis. *Mol. Phylogenet. Evol.* **5**, 182–187 (1996).
2. A. P. Martin, S. R. Palumbi, Body size, metabolic rate, generation time, and the molecular clock. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4087–4091 (1993).
3. B. Nabholz, S. Glémin, N. Galtier, Strong variations of mitochondrial mutation rate across mammals - The longevity hypothesis. *Mol. Biol. Evol.* **25**, 120–130 (2008).
4. J. S. Berv, D. J. Field, Genomic signature of an Avian lilliput effect across the K-Pg extinction. *Syst. Biol.* **67**, 1–13 (2018).
5. J. F. Gillooly, A. P. Allen, G. B. West, J. H. Brown, The rate of DNA evolution: Effects of body size and temperature on the molecular clock. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 140–145 (2005).
6. L. Bromham, The genome as a life-history character: Why rate of molecular evolution varies between mammal species. *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 2503–2513 (2011).
7. P. Moorjani, C. E. G. Amorim, P. F. Arndt, M. Przeworski, Variation in the molecular clock of primates. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10607–10612 (2016).
8. W. S. W. Wong, B. D. Solomon, D. L. Bodian, P. Kothiyal, G. Eley, K. C. Huddleston, R. Baker, D. C. Thach, R. K. Iyer, J. G. Vockley, J. E. Niederhuber, New observations on maternal age effect on germline *de novo* mutations. *Nat. Commun.* **7**, 10486 (2016).
9. Z. Gao, P. Moorjani, T. A. Sasani, B. S. Pedersen, A. R. Quinlan, L. B. Jorde, G. Amster, M. Przeworski, Overlooked roles of DNA damage and maternal age in generating human germline mutations. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 9491–9500 (2019).
10. F. L. Wu, A. I. Strand, L. A. Cox, C. Ober, J. D. Wall, P. Moorjani, M. Przeworski, A comparison of humans and baboons suggests germline mutation rates do not track cell divisions. *PLOS Biol.* **18**, e3000838 (2020).
11. S. Yang, G. Lian, ROS and diseases: Role in metabolism and energy supply. *Mol. Cell. Biochem.* **467**, 1–12 (2020).
12. R. S. Balaban, S. Nemoto, T. Finkel, Mitochondria, oxidants, and aging. *Cell* **120**, 483–495 (2005).
13. J. D. Lambeth, NOX enzymes and the biology of reactive oxygen. *Nat. Rev. Immunol.* **4**, 181–189 (2004).
14. M. P. Murphy, How mitochondria produce reactive oxygen species. *Biochem. J.* **417**, 1–13 (2009).
15. C. M. Bishop, The maximum oxygen consumption and aerobic scope of birds and mammals: Getting to the heart of the matter. *Proc. R. Soc. Lond. B Biol. Sci.* **266**, 2275–2281 (1999).
16. E. M. Dlugosz, M. A. Chappel, T. H. Meek, P. Szafrańska, K. Zub, M. Konarzewski, J. H. Jones, E. Bicudo, R. F. Nespolo, V. Careau, T. Garland, Phylogenetic analysis of mammalian maximal oxygen consumption during exercise. *J. Exp. Biol.* **216** (Pt. 24), 4712–4721 (2013).
17. A. Kapusta, A. Suh, C. Feschotte, Dynamics of genome size evolution in birds and mammals. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E1460–E1469 (2017).
18. W. Amos, S. J. Sawcer, R. W. Feakes, D. C. Rubinsztein, Microsatellites show mutational bias and heterozygote instability. *Nat. Genet.* **13**, 390–391 (1996).
19. C. R. Primmer, H. Ellegren, N. Saino, A. P. Moller, Directional evolution in germline microsatellite mutations. *Nat. Genet.* **13**, 391–393 (1996).
20. C. R. Primmer, N. Saino, A. P. Møller, H. Ellegren, Unraveling the processes of microsatellite evolution through analysis of germ line mutations in barn swallows *Hirundo rustica*. *Mol. Biol. Evol.* **15**, 1047–1054 (1998).
21. M. P. Chapuis, C. Plantamp, R. Streiff, L. Blondin, C. Piou, Microsatellite evolutionary rate and pattern in *Schistocerca gregaria* inferred from direct observation of germline mutations. *Mol. Ecol.* **24**, 6107–6119 (2015).
22. S. Malki, G. W. VanderHeijden, K. A. O'Donnell, S. L. Martin, A. Bortvin, A role for retrotransposon LINE-1 in fetal oocyte attrition in mice. *Dev. Cell* **29**, 521–533 (2014).
23. G. T. H. Vu, H. X. Cao, B. Reiss, I. Schubert, Deletion-bias in DNA double-strand break repair differentially contributes to plant genome shrinkage. *New Phytol.* **214**, 1712–1721 (2017).
24. K. Rodgers, M. McVey, Error-prone repair of DNA double-strand breaks. *J. Cell. Physiol.* **231**, 15–24 (2016).
25. M. R. Lieber, The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* **79**, 181–211 (2010).
26. J. P. Bird, R. Martin, H. R. Akçakaya, J. Gilroy, I. J. Burfield, S. T. Garnett, A. Symes, J. Taylor, Ç. H. Şekercioğlu, S. H. M. Butchart, Generation lengths of the world's birds and their implications for extinction risk. *Conserv. Biol.* **34**, 1252–1261 (2020).
27. J. B. Dunning, *CRC Handbook of Avian Body Masses* (CRC Press, ed. 2, 2007).
28. G. Zhang, C. Li, Q. Li, B. Li, D. M. Larkin, C. Lee, J. F. Storz, A. Antunes, M. J. Greenwold, R. W. Meredith, A. Ödeen, J. Cui, Q. Zhou, L. Xu, H. Pan, Z. Wang, L. Jin, P. Zhang, H. Hu, W. Yang, J. Hu, J. Xiao, Z. Yang, Y. Liu, Q. Xie, H. Yu, J. Lian, P. Wen, F. Zhang, H. Li, Y. Zeng, Z. Xiong, S. Liu, L. Zhou, Z. Huang, N. An, J. Wang, Q. Zheng, Y. Xiong, G. Wang, B. Wang, J. Wang, Y. Fan, R. R. da Fonseca, A. Alfaro-Núñez, M. Schubert, L. Orlando, T. Mourier, J. T. Howard, G. Ganapathy, A. Pfenning, O. Whitney, M. V. Rivas, E. Hara, J. Smith, M. Farré, J. Narayan, G. Slavov, M. N. Romanov, R. Borges, J. P. Machado, I. Khan, M. S. Springer, J. Gatesy, F. G. Hoffmann, J. C. Opazo, O. Håstad, R. H. Sawyer, H. Kim, K.-W. Kim, H. J. Kim, S. Cho, N. Li, Y. Huang, M. W. Bruford, X. Zhan, A. Dixon, M. F. Bertelsen, E. Derryberry, W. Warren, R. K. Wilson, S. Li, D. A. Ray, R. E. Green, S. J. O'Brien, D. Griffin, W. E. Johnson, D. Haussler, O. A. Ryder, E. Willerslev, G. R. Graves, P. Alström, J. Fjeldså, D. P. Mindell, S. V. Edwards, E. L. Braun, C. Rahbek, D. W. Burt, P. Houde, Y. Zhang, H. Yang, J. Wang; A. G. Consortium, E. D. Jarvis, M. T. P. Gilbert, J. Wang, Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311–1320 (2014).
29. S. Feng, J. Stiller, Y. Deng, J. Armstrong, Q. Fang, A. H. Reeve, D. Xie, G. Chen, C. Guo, B. C. Faircloth, B. Petersen, Z. Wang, Q. Zhou, M. Diekhans, W. Chen, S. Andreu-Sánchez, A. Margaryan, J. T. Howard, C. Parent, G. Pacheco, M. H. S. Sinding, L. Puetz, E. Cavill, Â. M. Ribeiro, L. Eckhart, J. Fjeldså, P. A. Hosner, R. T. Brumfield, L. Christidis, M. F. Bertelsen, T. Sicheritz-Ponten, D. T. Tietze, B. C. Robertson, G. Song, G. Borgia, S. Claramunt, I. J. Lovette, S. J. Cowen, P. Njoroge, J. P. Dumbacher, O. A. Ryder, J. Fuchs, M. Bunce, D. W. Burt, J. Cracraft, G. Meng, S. J. Hackett, P. G. Ryan, K. A. Jønsson, I. G. Jamieson, R. R. da Fonseca, E. L. Braun, P. Houde, S. Mirarab, A. Suh, B. Hansson, S. Ponnikas, H. Sigeman, M. Stervander, P. B. Frandsen, H. van der Zwan, R. van der Sluis, C. Visser, C. N. Balakrishnan, A. G. Clark, J. W. Fitzpatrick, R. Bowman, N. Chen, A. Cloutier, T. B. Sackton, S. V. Edwards, D. J. Foote, S. B. Shakya, F. H. Sheldon, A. Vignal, A. E. R. Soares, B. Shapiro, J. González-Solís, J. Ferrer-Obiol, J. Rozas, M. Riutort, A. Tigano, V. Friesen, L. Dalén, A. O. Urrutia, T. Székely, Y. Liu, M. G. Campana, A. Corvelo, R. C. Fleischer, K. M. Rutherford, N. J. Gemmell, N. Dussex, H. Mouritsen, N. Thiele, K. Delmore, M. Liedvogel, A. Franke, M. P. Hoeppner, O. Krone, A. M. Fudickar, B. Milá, E. D. Ketterson, A. E. Fidler, G. Friis, Á. M. Parody-Merino, P. F. Battley, M. P. Cox, N. C. B. Lima, F. Prosdocimi, T. L. Parchman, B. A. Schlinger, B. A. Loiselle, J. G. Blake,

H. C. Lim, L. B. Day, M. J. Fuxjager, M. W. Baldwin, M. J. Braun, M. Wirthlin, R. B. Dikow, T. B. Ryder, G. Camenisch, L. F. Keller, J. M. DaCosta, M. E. Hauber, M. I. M. Louder, C. C. Witt, J. A. McGuire, J. Mudge, L. C. Megna, M. D. Carling, B. Wang, S. A. Taylor, G. Del-Rio, A. Aleixo, A. T. R. Vasconcelos, C. V. Mello, J. T. Weir, D. Haussler, Q. Li, H. Yang, J. Wang, F. Lei, C. Rahbek, M. T. P. Gilbert, G. R. Graves, E. D. Jarvis, B. Paten, G. Zhang, Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587**, 252–257 (2020).

30. A. von Hardenberg, A. Gonzalez-Voyer, Disentangling evolutionary cause-effect relationships with phylogenetic confirmatory path analysis. *Evolution* **67**, 378–387 (2013).
31. Zoonomia Consortium, A comparative genomics multitool for scientific discovery and conservation. *Nature* **587**, 240–245 (2020).
32. K. E. Jones, J. Bielby, M. Cardillo, S. A. Fritz, J. O'Dell, C. D. L. Orme, K. Safi, W. Sechrest, E. H. Boakes, C. Carbone, C. Connolly, M. J. Cutts, J. K. Foster, R. Grenyer, M. Habib, C. A. Plaster, S. A. Price, E. A. Rigby, J. Rist, A. Teacher, O. R. P. Bininda-Emonds, J. L. Gittleman, G. M. Mace, A. Purvis, PanTHERIA: A species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology* **90**, 2648–2648 (2009).
33. M. Keane, J. Semeiks, A. E. Webb, Y. I. Li, V. Quesada, T. Craig, L. B. Madsen, S. van Dam, D. Brawand, P. I. Marques, P. Michalak, L. Kang, J. Bhak, H. S. Yim, N. V. Grishin, N. H. Nielsen, M. P. Heide-Jørgensen, E. M. Oziolor, C. W. Matson, G. M. Church, G. W. Stuart, J. C. Patton, J. C. George, R. Suydam, K. Larsen, C. López-Otín, M. J. O'Connell, J. W. Bickham, B. Thomsen, J. P. deMagalhães, Insights into the evolution of longevity from the bowhead whale genome. *Cell Rep.* **10**, 112–122 (2015).
34. X. Tian, D. Firsanov, Z. Zhang, Y. Cheng, L. Luo, G. Tombline, R. Tan, M. Simon, S. Henderson, J. Steffan, A. Goldfarb, J. Tam, K. Zheng, A. Cornwell, A. Johnson, J. N. Yang, Z. Mao, B. Manta, W. Dang, Z. Zhang, J. Vijg, A. Wolfe, K. Moody, B. K. Kennedy, D. Bohmann, V. N. Gladyshev, A. Seluanov, V. Gorbunova, SIRT6 is responsible for more efficient DNA double-strand break repair in long-lived species. *Cell* **177**, 622–638.e22 (2019).
35. J. P. De MagalhÃes, J. Costa, A database of vertebrate longevity records and their relation to other life-history traits. *J. Evol. Biol.* **22**, 1770–1774 (2009).
36. T. Ohta, Slightly deleterious mutant substitutions in evolution. *Nature* **246**, 96–98 (1973).
37. M. Lynch, J. S. Conery, The origins of genome complexity. *Science* **302**, 1401–1404 (2003).
38. E. Figuet, M. Bonneau, E. M. Carrio, K. Nadachowska-brzyska, H. Ellegren, N. Galtier, Life history Traits, protein evolution, and the nearly neutral theory in Amniotes. *Mol. Biol. Evol.* **33**, 1517–1527 (2016).
39. T. Lefébure, C. Morvan, F. Malard, C. François, L. Konecny-dupré, L. Guéguen, M. Weiss-gayet, A. Seguin-orlando, L. Ermini, C. der Sarkissian, N. P. Charrier, D. Eme, F. Mermillod-blondin, L. Duret, C. Vieira, L. Orlando, C. J. Douady, Less effective selection leads to larger genomes. *Genome Res.* **27**, 1016–1028 (2017).
40. R. Cui, T. Medeiros, D. Willemsen, L. N. M. Iasi, G. E. Collier, M. Graef, M. Reichard, D. R. Valenzano, Relaxed selection limits lifespan by increasing mutation load. *Cell* **178**, 385–399.e20 (2019).
41. S. Mathur, J. A. DeWoody, Genetic load has potential in large populations but is realized in small inbred populations. *Evol. Appl.* **14**, 1540–1557 (2021).
42. M. Kimura, T. Ohta, The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**, 763–771 (1969).
43. A. L. Hughes, M. K. Hughes, Small genomes for better flyers. *Nature* **377**, 391 (1995).
44. C. L. Organ, A. M. Shedlock, Palaeogenomics of pterosaurs and the evolution of small genome size in flying vertebrates. *Biol. Lett.* **5**, 47–50 (2009).
45. J. D. L. Smith, T. R. Gregory, The genome sizes of megabats (Chiroptera: Pteropodidae) are remarkably constrained. *Biol. Lett.* **5**, 347–351 (2009).
46. T. R. Gregory, A bird's-eye view of the C-value enigma: Genome size, cell size, and metabolic rate in the class Aves. *Evolution* **56**, 121–130 (2002).
47. N. A. Wright, T. R. Gregory, C. C. Witt, Metabolic 'engines' of flight drive genome size reduction in birds. *Proc. R. Soc. Biol. Sci.* **281**, 20132780 (2014).
48. M. S. Y. Lee, A. Cau, D. Naish, G. J. Dyke, Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. *Science* **345**, 562–566 (2014).
49. E. L. Rezende, L. D. Bacigalupe, R. F. Nespolo, F. Bozinovic, Shrinking dinosaurs and the evolution of endothermy in birds. *Sci. Adv.* **6**, eaaw4486 (2020).
50. R. B. J. Benson, N. E. Campione, M. T. Carrano, P. D. Mannion, C. Sullivan, P. Upchurch, D. C. Evans, Rates of dinosaur body mass evolution indicate 170 million years of sustained ecological innovation on the Avian stem lineage. *PLOS Biol.* **12**, e1001896 (2014).
51. C. L. Organ, A. M. Shedlock, A. Meade, M. Pagel, S. V. Edwards, Origin of avian genome size and structure in non-avian dinosaurs. *Nature* **446**, 180–184 (2007).
52. B. K. McNab, What determines the basal rate of metabolism? *J. Exp. Biol.* **222**, jeb205591 (2019).
53. M. Kleiber, Body size and metabolic rate. *Physiol. Rev.* **27**, 511–541 (1947).
54. Y. Ji, J. A. DeWoody, Relationships among powered Flight, metabolic rate, body mass, genome size, and the retrotransposon complement of volant birds. *Evol. Biol.* **44**, 261–272 (2017).
55. S. Lamichhaney, R. Catullo, J. S. Keogh, S. Clulow, S. V. Edwards, T. Ezaz, A bird-like genome from a frog: Mechanisms of genome size reduction in the ornate burrowing frog, *Platyplectrum ornatum*. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2011649118 (2021).
56. C. Sun, D. B. Shepard, R. A. Chong, J. López Arriaza, K. Hall, T. A. Castoe, C. Feschotte, D. D. Pollock, R. L. Mueller, LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Genome Biol. Evol.* **4**, 168–183 (2012).
57. M. B. Frahry, C. Sun, R. A. Chong, R. L. Mueller, Low levels of LTR retrotransposon deletion by ectopic recombination in the gigantic genomes of salamanders. *J. Mol. Evol.* **80**, 120–129 (2015).
58. G. Lertzman-Lepofsky, A. Mooers, D. A. Greenberg, Ecological constraints associated with genome size across salamander lineages. *Proc. R. Soc. B Biol. Sci.* **286**, 20191780 (2019).
59. Z. Gao, M. J. Wyman, G. Sella, M. Przeworski, Interpreting the dependence of mutation rates on age and time. *PLOS Biol.* **14**, e1002355 (2016).
60. T. Leroy, M. Rousselle, M. K. Tilak, A. E. Caizergues, C. Scornavacca, M. Recuerda, J. Fuchs, J. C. Illera, D. H. de Swardt, G. Blanco, C. Thébaud, B. Milá, B. Nabholz, Island songbirds as windows into evolution in small populations. *Curr. Biol.* **31**, 1303–1310.e4 (2021).
61. J. D. Manthey, R. G. Moyle, S. Boissinot, Multiple and independent phases of transposable element amplification in the genomes of piciformes (woodpeckers and allies). *Genome Biol. Evol.* **10**, 1445–1456 (2018).
62. B. Jackson, K. Dial, Scaling of mechanical power output during burst escape flight in the Corvidae. *J. Exp. Biol.* **214**, 452–461 (2011).
63. T. R. Gregory, Macroevolution, hierarchy theory, and C-value enigma. *Paleobiology* **30**, 179–202 (2004).
64. V. Peona, M. P. K. Blom, L. Xu, R. Burri, S. Sullivan, I. Bunikis, I. Liachko, T. Haryoko, K. A. Jønsson, Q. Zhou, M. Irestedt, A. Suh, Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. *Mol. Ecol. Resour.* **21**, 263–286 (2021).
65. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
66. G. Benson, Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
67. A. Fungtammasan, G. Ananda, S. E. Hile, M. S.-W. Su, C. Sun, R. Harris, P. Medvedev, K. Eckert, K. D. Makova, Accurate typing of short tandem repeats from genome-wide sequencing data and its applications. *Genome Res.* **125**, 736–749 (2015).
68. J. Armstrong, G. Hickey, M. Diekhans, I. T. Fiddes, A. M. Novak, A. Deran, Q. Fang, D. Xie, S. Feng, J. Stiller, D. Genereux, J. Johnson, V. D. Marinescu, J. Alföldi, R. S. Harris, K. Lindblad-Toh, D. Haussler, E. Karlsson, E. D. Jarvis, G. Zhang, B. Paten, Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020).
69. J. G. Burleigh, R. T. Kimball, E. L. Braun, Building the avian tree of life using a large-scale, sparse supermatrix. *Mol. Phylogenet. Evol.* **84**, 53–63 (2015).
70. J. S. Lefcheck, PIECEWISESEM: Piecewise structural equation modelling in R for ecology, evolution, and systematics. *Methods Ecol. Evol.* **7**, 573–579 (2016).
71. L. s. Tung Ho, C. Ané, A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst. Biol.* **63**, 397–408 (2014).
72. L. J. Revell, phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
73. R Core Team, R: A language and environment for statistical computing (R Foundation for Statistical Computing, 2021); www.r-project.org.
74. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T.-Y. Lam, GGTREE: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
75. V. Miele, S. Penel, L. Duret, Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC Bioinformatics* **12**, 116 (2011).
76. International Chicken Genome Sequencing Consortium, Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
77. Z. N. Kronenberg, A. Rhie, S. Koren, G. T. Concepcion, P. Peluso, K. M. Munson, D. Porubsky, K. Kuhn, K. A. Mueller, W. Y. Low, S. Hiendleder, O. Fedrigo, I. Liachko, R. J. Hall, A. M. Phillippy, E. E. Eichler, J. L. Williams, T. P. L. Smith, E. D. Jarvis, S. T. Sullivan, S. B. Kingan, Extended haplotype-phasing of long-read *de novo* genome assemblies using Hi-C. *Nat. Commun.* **12**, 1935 (2021).
78. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
79. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).

Acknowledgments: We thank Q. Feng, M. Sundaram, C. Eliason, N. Crouch, Y. Zhang, W. Zhai, M. Diekhans, E. Karlsson, and two anonymous reviewers for helpful discussions/comments

throughout the development of this manuscript. We also thank G. Chen, H. Lin, and X. Chen for assistance of data extraction/transfer during the pandemic. **Funding:** This work was supported by the National Natural Science Foundation of China (32000290 to Y. J., 3213000355 to F.L., 31901214 to S.F., 32000307 to Y.C., 31900320 to D.Z., and 32100332 to Y.H.), Neugaunee Postdoctoral Fellowship (to Y.J.), Carlsberg Foundation CF19-0427 (to A.B.-O.), Carlsberg Foundation (CF16-0663 to G.Z.), National Institute for Food and Agriculture (to J.A.D.), International Partnership Program of Chinese Academy of Sciences (152453KYSB20170002 to G.Z.), Strategic Priority Research Program of the Chinese Academy of Sciences (XDB31020000 to G.Z.), and Villum Foundation (25900 to G. Z.). **Author contributions:** Conceptualization: Y.J., G.Z., S.J.H., and F.L. Methodology: Y.J., S.F., L.W., Q.F., and A.B.-O. Visualization: Y.J. and L.W. Supervision: G.Z., S.J.H., and F.L. Writing—original draft: Y.J. Writing—review and editing: J.A.D., A.S., F.L., L.W., A.B.-O., Y.C., D.Z., Y.H., G.S., and Y.Q. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Measurements of genomic attributes are available in the Supplementary Materials. Scripts and other associated data are available in the Github Repository (https://github.com/yanzhu-ji/avian-msats) and ScienceDB (www.scidb.cn/en/detail?dataSetId=672cac7e49bd4bcdac0886f0e985f639).

Submitted 9 January 2022
Accepted 18 July 2022
Published 31 August 2022
10.1126/sciadv.abo0099