

Purifying selection on young transposable element insertions persists in a small bird population

Li Lv

lvli@bjwlxy.edu.cn

Baoji University of Arts and Sciences

Short Report

Keywords: Transposable element, Pangenome, Effective population size, Avian genomics, Mutational hazard hypothesis

Posted Date: October 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11034987/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

The mutational hazard hypothesis predicts that non-coding DNA accumulates in small populations, yet an island scrub-jay with the smallest effective population size (N_e) has a reduced genome. Whether transposable elements (TEs) follow this trajectory is unknown. Comparing TE abundance and age structure across three *Aphelocoma* species spanning a ~ 55 -fold N_e range, we found that LINE elements were the most stable repeat class ($AI/AW = 0.960$), whereas LTR, SINE, and satellite DNA were substantially reduced. Crucially, young TE insertions were depleted in the island species across all categories (LINE – 33%, LTR – 17%, SINE – 68%; $p = 1.29 \times 10^{-8}$), declining disproportionately more than old copies—consistent with ongoing purifying selection against new insertions rather than wholesale loss. A subfamily-level constraint score correlated with stability. Thus purifying selection persists in small populations, and the fate of repetitive DNA depends on element-specific properties, not population size alone.

Introduction

Eukaryotic genome size varies by more than 60,000-fold, yet this immense variation cannot be explained by differences in gene number or organismal complexity [1]. Instead, genome size is primarily determined by the abundance of non-coding repetitive DNA, which encompasses two fundamentally distinct classes: transposable elements (TEs)—interspersed repeats that mobilize throughout the genome—and satellite DNA—tandemly repeated sequences concentrated in heterochromatic regions such as centromeres and telomeres [2]. Understanding the evolutionary forces governing the accumulation and loss of these repetitive elements is central to explaining genome size evolution.

The mutational hazard hypothesis (MHH), proposed by Lynch and Conery [3], provides the dominant population-genetic framework for understanding repetitive DNA dynamics. According to the MHH, non-coding DNA represents a class of slightly deleterious mutations whose fate is determined by the interplay between mutation, drift, and selection. In species with large effective population sizes (N_e), purifying selection efficiently removes these insertions; in species with small N_e , genetic drift overcomes weak selection, allowing non-coding DNA to accumulate. This hypothesis predicts that species with small populations should harbor larger genomes with greater repetitive DNA content—a prediction supported across diverse taxa, including salamanders [4], plants [5], and bacteria [6].

Despite its explanatory power, a critical limitation of the MHH is its implicit assumption that all non-coding DNA is functionally equivalent in its deleterious effects. In reality, different repeat classes differ fundamentally in their insertion mechanisms, genomic target preferences, and fitness consequences. LINE (Long Interspersed Nuclear Element) retrotransposons are autonomous elements that insert large copies (4–7 kb) preferentially into gene-rich euchromatic regions, where they can disrupt gene structure and cause deleterious mutations [7]. LTR (Long Terminal Repeat) retrotransposons, including endogenous retroviruses, can reach even larger sizes and carry regulatory sequences affecting host gene expression [8]. SINE (Short Interspersed Nuclear Element) elements are non-autonomous, short ($\sim$

100–400 bp), and depend on LINE-encoded machinery for mobilization [7]. At the opposite end of the spectrum, satellite DNA consists of tandem arrays largely restricted to pericentromeric and subtelomeric heterochromatin, where they contribute to centromere function but rarely interfere with gene expression [9]. These biological differences suggest that TEs and satellite DNA may experience fundamentally different selective regimes, and may therefore respond differently to changes in population size.

Birds provide an ideal system to test this hypothesis. Avian genomes are among the most compact among vertebrates (typically 1.0–1.3 Gb), yet their repetitive DNA content varies substantially across lineages [10, 11]. The Bird 10,000 Genomes (B10K) project and related efforts have produced hundreds of high-quality avian genome assemblies [10, 12], enabling comparative analyses of TE dynamics across diverse ecological and evolutionary contexts [13, 14]. Notably, the recent multispecies pangenome for three North American scrub-jays (*Aphelocoma* spp.) by Edwards *et al.* [15] provides an unprecedented opportunity to examine repeat element dynamics across a well-characterized population size gradient. Using 45 PacBio HiFi long-read genomes assembled into 90 haplotype-resolved assemblies, Edwards *et al.* analyzed structural variants across three closely related species: the continental *A. woodhouseii* (AW, largest N_e), the endangered *A. coerulescens* (AC, intermediate N_e), and the island endemic *A. insularis* (AI, smallest N_e , approximately 1/55 of AW). A central finding was that AI—contrary to MHH predictions—exhibits a genome ~ 80 Mb smaller than AW, driven primarily by the loss of major satellite DNA families. However, Edwards *et al.* focused on SVs and did not systematically investigate whether different TE classes follow satellite DNA's trajectory under population decline, nor did they explore the mechanistic basis for differential repeat class dynamics.

Here, we leverage the comprehensive repeat annotation data from the scrub-jay pangenome to test three hypotheses: (1) different repeat classes should respond divergently to population decline, rather than uniformly as the classical MHH predicts; (2) if purifying selection persists in small populations, young TE insertions should be selectively depleted to a greater extent than old, fixed insertions; and (3) the degree of stability of individual TE subfamilies should correlate with their selective constraint in the large ancestral population. By testing these hypotheses, we aim to clarify whether purifying selection operates on repetitive DNA even in small populations, and whether the evolutionary fate of repetitive DNA is shaped by element-specific properties rather than by population size alone.

Results

Differential responses of repeat classes to population decline

We first quantified genome-wide abundance of eight major repeat classes across the three focal species using the RepeatMasker annotation data from Edwards *et al.* [15]. For each individual ($N = 15$ AW, 14 AC, 15 AI; averaging the two haploid assemblies per diploid genome), we calculated the genomic fraction occupied by each repeat class. We first assessed overall differences among the three species using Kruskal-Wallis tests, then performed pairwise Mann-Whitney U tests between AW and AI with Benjamini-

Hochberg FDR correction for multiple comparisons. Boxplots of four key repeat classes are shown in Fig. 1A, and complete statistics for all eight classes are provided in **Table S1**.

The eight repeat classes responded divergently to population decline. LINE elements were the most stable major class: the mean genomic fraction was 4.14% (SD = 0.13%) in AW and 3.98% (SD = 0.04%) in AI, yielding an AI/AW abundance ratio of 0.960—a reduction of only 4.0% ($p = 4.25 \times 10^{-5}$, Mann-Whitney U test, FDR-corrected; Fig. 1A, **Table S1**). While this small difference reaches statistical significance owing to the low inter-individual variance in AI (SD = 0.04%), the AI/AW ratio of 0.960 is the closest to unity among all repeat classes except DNA transposons (0.975). In contrast, satellite DNA decreased from 9.96% (SD = 0.34%) to 8.32% (SD = 0.45%) (AI/AW = 0.835, a 16.5% reduction; FDR-corrected $p = 3.44 \times 10^{-8}$; Fig. 1A). The full spectrum of responses is summarized by the AI/AW abundance ratios shown in Fig. 1B: LINE (0.960) and DNA transposons (0.975) cluster near the stability line, whereas LTR (0.757), SINE (0.776), Satellite (0.835), and Unknown (0.515) all fall substantially below it.

Notably, LTR retrotransposons and SINE elements were reduced even more strongly than satellite DNA in relative terms. LTR retrotransposons decreased from 7.27% (SD = 0.74%) to 5.50% (SD = 0.57%) (AI/AW = 0.757, -24.3%; FDR-corrected $p = 1.16 \times 10^{-6}$; **Fig. S1**), and SINE elements decreased from 0.049% (SD = 0.003%) to 0.038% (SD = 0.002%) (AI/AW = 0.776, -22.4%; FDR-corrected $p = 3.44 \times 10^{-8}$; **Fig. S1**). Satellite DNA nevertheless contributed a large absolute loss to genome size reduction (~ 17 Mb) owing to its high baseline abundance (~ 10% of the genome), even though its relative reduction (16.5%) was smaller than that of LTR and SINE. DNA transposons as a group remained statistically indistinguishable between AW and AI (AI/AW = 0.975; FDR-corrected $p = 0.934$; **Fig. S1**); however, at the subfamily level, DNA/hAT-Charlie showed a marked expansion in AI (AI/AW = 1.695), representing a notable outlier that may reflect either genuine amplification or stochastic effects in the small population (**Table S3**). The 'Unknown' repeat class—which likely contains unclassified TE fragments and degenerate repeats—showed the largest relative reduction (48.5%), from 3.12% (SD = 0.13%) in AW to 1.61% (SD = 0.06%) in AI (AI/AW = 0.515; FDR-corrected $p = 3.44 \times 10^{-8}$; **Fig. S1**), contributing approximately 16 Mb to the genome size reduction. The overall repeat composition shift from AW to AI is visualized in the stacked bar chart in Fig. 1C.

These results demonstrate that different repeat classes respond in fundamentally different ways to population size decline—the MHH assumption that all non-coding DNA is equivalent in its evolutionary dynamics is not supported.

Young TE insertions are depleted in the small population across all categories

The contrasting stability of LINE elements versus satellite DNA raises a critical mechanistic question: is LINE stability due to reduced transposition activity or ongoing purifying selection? To distinguish between these alternatives, we analyzed TE age structure using Kimura two-parameter divergence [16]

as a molecular clock proxy. For each TE copy in each haplotype, we calculated total Kimura divergence (substitutions + deletions + insertions relative to the family consensus) and classified insertions as young (< 5% divergence), medium (5–15%), or old (> 15%). Under reduced activity, all age classes would be proportionally reduced; under ongoing selection, young insertions would be selectively depleted while old, fixed insertions are retained.

The age composition of TE copies, aggregated across LINE, SINE, and LTR at the individual level, revealed a consistent depletion of young insertions in the small population (Fig. 2A). The mean proportion of young TE copies (< 5% divergence) declined monotonically from AW (5.69%) through AC (5.12%) to AI (4.22%), while the proportion of old copies (> 15% divergence) rose correspondingly from 76.10% (AW) to 78.30% (AI). The medium age class (5–15% divergence) showed essentially no species-level signal and is therefore omitted from the figure to keep the focus on the two informative bins. Quantitative analysis at the individual level, broken down by TE category, confirmed this young-insertion depletion across all three major TE classes (Fig. 2B, **Table S2**). The proportion of young TE insertions was significantly lower in AI than AW for LINE (1.82% vs. 1.21%, -33.3%; Mann-Whitney U = 225, $p = 1.29 \times 10^{-8}$), LTR (12.10% vs. 10.07%, -16.7%; Mann-Whitney U = 225, $p = 1.29 \times 10^{-8}$), and SINE (1.51% vs. 0.48%, -68.2%; Mann-Whitney U = 225, $p = 1.29 \times 10^{-8}$); the full young/medium/old composition for each category is shown in **Fig. S2**. Conversely, old TE proportions were elevated in AI across all categories (Fig. 2C). The magnitude of young TE depletion varied substantially among categories: SINE showed the most severe reduction (-68%), followed by LINE (-33%) and LTR (-17%). SINE's extreme depletion is consistent with its non-autonomous transposition, short length, and weak individual-level selective effects, making it particularly vulnerable to stochastic loss under reduced N_e . Sensitivity analysis using alternative Kimura thresholds confirmed robustness: the relative depletion of young insertions in AI ranged from -17.6% (young < 7%) to -29.6% (young < 4%), and AC was consistently intermediate across all thresholds (**Table S4**).

A key discriminator between the competing explanations—wholesale TE loss versus selective removal of new insertions—lies in the differential fate of young versus old insertions (Fig. 2D). If TEs were lost wholesale through drift or founder effects, young and old copies should decline proportionally; if selection removes new insertions, young copies should decline disproportionately while old copies remain relatively stable. We observed the latter pattern consistently across all three TE classes, expressed here as the percentage change in AI relative to AW (Fig. 2D). For LINE, young copies declined by 36.0% (5,579 to 3,568 per individual) while old copies declined by only 3.7% (263,318 to 253,511). For SINE, young copies declined by 68.8% (82 to 26) while old copies declined by only 0.7% (5,172 to 5,134). For LTR, young copies declined by 31.5% (22,804 to 15,632) while old copies declined by 13.4% (112,276 to 97,177). The intermediate species AC showed much smaller departures from AW (Fig. 2D, upper panel), consistent with its larger N_e relative to AI. This consistent excess of young-copy depletion over old-copy depletion argues against drift- or founder-mediated wholesale loss as the primary driver, and is most consistent with a model in which newly produced TE insertions are selectively removed by purifying selection, even in the small-population island species. We acknowledge two caveats. First,

because AI represents a single island population, we cannot fully exclude founder effects; however, founder effects alone would predict proportional reductions across all age classes, whereas we observe a disproportionate decline of young insertions with old copies remaining nearly unchanged. Second, a reduction in transposition activity could in principle also reduce the number of young insertions without necessarily invoking selection; we cannot fully distinguish this from selective removal with the present data. Nevertheless, the magnitude and consistency of young-insertion depletion across all three TE classes is most parsimoniously explained by ongoing purifying selection.

Subfamily-level selective constraint correlates with stability under population decline

The TE age analysis demonstrates that purifying selection persists in small populations, but the magnitude of young TE depletion varies among categories (SINE: -68%, LINE: -33%, LTR: -17%). This suggests that the strength of purifying selection differs among TE types. If this is the case, the degree of stability of individual TE subfamilies under population decline should be predictable from their selective constraint in the large ancestral population. To test this, we developed a data-driven constraint metric that does not rely on prior knowledge of TE biology.

For 15 TE subfamilies (from the repeat_type field, excluding low-complexity and very low-abundance types), we calculated the coefficient of variation (CV) of genomic fraction across the 30 AW haplotypes. The CV captures cross-individual abundance variability: subfamilies under strong purifying selection should exhibit narrow abundance bounds (low CV), while weakly constrained subfamilies should fluctuate more (high CV). We defined the constraint score as $1/(1 + CV)$, mapping values to a 0–1 scale where higher values indicate stronger constraint. This approach is analogous to using allele frequency variance as a measure of selective constraint, but applied to quantitative TE abundance[17].

Across the 15 subfamilies, the constraint score was positively correlated with the AI/AW abundance ratio (Spearman $\rho = 0.554$, $p = 3.23 \times 10^{-2}$; Fig. 3A). Highly constrained subfamilies tended to show AI/AW ratios close to 1.0: LINE/CR1 (constraint = 0.976, CV = 0.024) had an AI/AW ratio of 0.968; DNA transposons (constraint = 0.981) had 1.077; SINE/MIR (constraint = 0.979) had 1.066 (**Fig. S3**). In contrast, weakly constrained subfamilies were substantially depleted: DNA/Dada (constraint = 0.696, CV = 0.436) showed AI/AW = 0.195, an 80% reduction; LINE/L1 (constraint = 0.708) showed 0.549; LTR/ERV1, the most abundant LTR subfamily (3.51% of AW genome), showed moderate constraint (0.919) and substantial reduction (AI/AW = 0.680) (Fig. 3B; **Table S3**).

We note that the intermediate species AC falls between AW and AI for young TE proportion (**Fig. S4**)—consistent with a continuous relationship between N_e and selection efficiency on TE age structure—though AC does not show intermediate abundance for all repeat classes (e.g., satellite DNA is most abundant in AC, at 10.97%; **Table S1**), suggesting that repeat element abundance is shaped by additional lineage-specific factors beyond population size alone.

Discussion

Our analysis reveals two central findings: repeat classes respond divergently to population decline, and purifying selection against new TE insertions persists even in the small island population. These results add an important caveat to the mutational hazard hypothesis [3], which treated all non-coding DNA as a homogeneous class of slightly deleterious mutations.

First, repeat classes do not respond uniformly to a ~ 55 -fold reduction in N_e . LINE elements—particularly CR1, which comprises $> 95\%$ of LINE copies—are maintained within narrow abundance bounds ($CV = 0.024$) and show only a modest 4.0% reduction ($AI/AW = 0.960$), whereas LTR (0.757), SINE (0.776), and satellite DNA (0.835) are substantially reduced. The stability of LINE/CR1 is most plausibly maintained by purifying selection, evidenced by the significant depletion of young LINE insertions in AI (-33%). This aligns with comparative analyses showing that N_e does not explain long-term TE content variation across animals [18], and with lineage-specific TE dynamics [14, 19]. Our results suggest selection efficacy against repetitive DNA depends on element-specific properties—a logic akin to the nearly neutral theory [20].

Second, the disproportionate depletion of young insertions relative to old copies across LINE, LTR, and SINE provides direct evidence that purifying selection operates even in severely bottlenecked populations, and argues against founder effects or drift-mediated wholesale loss. This parallels human population genomics, where young, low-frequency TE insertions are enriched for deleterious variants [7, 21]. The particularly severe depletion of young SINE insertions (-68%) may reflect their non-autonomous transposition and weak individual selective effects, making them susceptible to stochastic loss under drift [22].

Third, our finding has practical implications for conservation genomics [23, 24]. Genome size reduction during bottlenecks may superficially appear beneficial, but it does not necessarily indicate improved genome health. Conservation genomic assessments should consider not only SNP diversity and inbreeding, but also TE composition and age structure [25, 26], as the accumulation of young, potentially active TE insertions may represent an underappreciated genetic threat [27].

Several limitations should be noted. Our analysis is restricted to three closely related *Aphelocoma* species, limiting taxonomic generalizability [10, 28, 29]; broader sampling, including other bird pangenomes [30] and additional vertebrate clades [31], will be needed to test generality. We used Kimura divergence as a TE age proxy; although lineage-specific mutation-rate variation could introduce biases [31, 32], sensitivity analysis with alternative thresholds confirmed the robustness of the young-insertion depletion (**Table S4**). We focused on repeat elements and did not examine other structural variants, such as copy-number variants, whose architecture also shapes avian genome evolution [33]. The constraint-stability relationship rests on a single correlation with a modest sample size ($N = 15$ subfamilies), and does not incorporate epigenetic silencing or chromatin context [34, 35]. Finally, AI is a single island population, so we cannot fully separate reduced long-term N_e from recent bottlenecks or founder effects; multiple independent island-mainland comparisons would help disentangle these factors.

In conclusion, purifying selection against new TE insertions persists even in a bottlenecked island population, and the evolutionary fate of repetitive DNA under population decline is shaped by element-specific properties rather than population size alone. These findings highlight the value of TE age-structure analysis for inferring selective dynamics in natural populations.

Methods

Data source

All genomic data were obtained from the Dryad repository (doi:10.5061/dryad.8pk0p2p01) associated with Edwards *et al.* [15], which contains repeat annotations for 45 PacBio HiFi genomes assembled into 90 *Aphelocoma* haplotypes (and 4 additional outgroup haplotypes from *Aphelocoma californica*, *Cyanocorax yucatanicus*, and *Cyanocitta stelleri*; 94 haplotypes in total across 6 species). We restricted all analyses to the three focal species AW, AC, and AI (44 individuals, 88 haplotypes) and used: (1) the per-haplotype RepeatMasker summary table (all_haps_repmask_nornd_sums.txt; 2,820 rows), and (2) individual BED-format RepeatMasker output files (88 files for AW, AC, AI, generated with BEDOPS rmsk2bed). Species codes: AW = *A. woodhouseii* (large Ne), AC = *A. coerulescens* (medium Ne), AI = *A. insularis* (small Ne, $\sim 1/55$ of AW Ne).

Repeat element quantification

RepeatMasker annotations were grouped into eight classes based on maj_rep_type. For each diploid individual (N = 15 AW, 14 AC, 15 AI), genomic fraction was calculated by summing annotated base pairs from both haploid assemblies and dividing by the diploid genome length ($2 \times$ haploid assembly length). Overall differences among the three species were first assessed using the Kruskal-Wallis rank-sum test. Pairwise comparisons between AW and AI were performed using two-sided Mann-Whitney U tests, with Benjamini-Hochberg FDR correction applied across all eight repeat classes to control for multiple testing. The AI/AW ratio was defined as mean genomic fraction in AI divided by mean in AW. Analyses used SciPy v1.15. All Mann-Whitney U tests were computed with the exact (permutation-based) method, appropriate for small samples without ties.

TE age structure analysis

TE age was estimated using Kimura two-parameter distance [16] as the sum of substitution, deletion, and insertion percentages relative to family consensus sequences. Copies were classified as young ($< 5\%$ divergence), medium (5–15%), or old ($> 15\%$). Sensitivity analysis re-classified young insertions using thresholds of 3–7% divergence (**Table S4**), confirming that the relative depletion in AI is robust to threshold choice. Only LINE, SINE, and LTR elements were included; satellite DNA and simple repeats were excluded. Per-individual proportions were calculated by aggregating counts from the two haploid assemblies per individual. AW vs. AI differences were tested with Mann-Whitney U.

Constraint score calculation

Fifteen subfamilies (from repeat_type) were selected based on: mean genomic fraction > 0.005% in AW, present in > = 5 AW and AI haplotypes, excluding Simple_repeat/Low_complexity/rRNA/tRNA/Unknown/RC. CV was calculated as standard deviation / mean of genomic fraction across 30 AW haplotypes. Constraint score = $1/(1 + CV)$, bounded 0–1 with higher values indicating stronger constraint. Spearman's rank correlation tested the constraint–AI/AW relationship.

Software

Analyses used Python v3.10 with NumPy and SciPy. Visualization used R v4.4 with ggplot2, ggpubr, ggrepel, and cowplot.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Funding Statement

This work was supported by the Scientific Research Program of Shaanxi Provincial Department of Education (23JK0290), the Key Scientific Research Project of Baoji University of Arts and Sciences (ZK16120), and the Research Project on Teacher Education Reform and Teacher Development (26JF27).

Author Contribution

L.L. conceptualised the study, designed the methodology, performed data collection and analysis, interpretation of findings, and wrote the manuscript.

Acknowledgements

The author thanks the School of Electrical & Electronic Engineering at Baoji University of Arts and Sciences for providing the academic environment to conduct this work.

Data Availability

All genomic data analyzed in this study were originally generated by Edwards et al. (2025) and are publicly available through the Dryad Digital Repository at <https://doi.org/10.5061/dryad.8pk0p2p01>. We used: (i) per-haplotype RepeatMasker annotation BED files (all_sj_haps_RepeatMasker_bedfiles.tar.gz), and (ii) the RepeatMasker summary table (all_haps_repmask_nornd_sums.txt).

References

1. Gregory TR. The C-value enigma in plants and animals: a review of parallels and an appeal for partnership. *Ann Bot.* 2005;95(1):133–46.
2. Lynch M. The origins of genome architecture. Sunderland, Mass: Sinauer Associates; 2007.
3. Lynch M, Conery JS. The origins of genome complexity. *Science.* 2003;302(5649):1401–4.
4. Mohlhenrich ER, Mueller RL. Genetic drift and mutational hazard in the evolution of salamander genomic gigantism. *Evolution.* 2016;70(12):2865–78.
5. Whitney KD, Garland T Jr.. Did genetic drift drive increases in genome complexity? *PLoS Genet* 2010, 6(8).
6. Lynch M. Streamlining and simplification of microbial genome architecture. *Annu Rev Microbiol.* 2006;60:327–49.
7. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvak Z, Levin HL, Macfarlan TS, et al. Ten things you should know about transposable elements. *Genome Biol.* 2018;19(1):199.
8. Mager DL, Stoye JP. Mammalian Endogenous Retroviruses. *Microbiol Spectr.* 2015;3(1):MDNA3–0009.
9. Garrido-Ramos MA. Satellite DNA: An Evolving Topic. *Genes (Basel)* 2017, 8(9).
10. 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):1311–20.
11. Kapusta A, Suh A, Feschotte C. Dynamics of genome size evolution in birds and mammals. *Proc Natl Acad Sci U S A.* 2017;114(8):E1460–9.
12. Feng S, Stiller J, Deng Y, Armstrong J, Fang Q, Reeve AH, Xie D, Chen G, Guo C, Faircloth BC, et al. Dense sampling of bird diversity increases power of comparative genomics. *Nature.* 2020;587(7833):252–7.
13. Kapusta A, Suh A. Evolution of bird genomes-a transposon's-eye view. *Ann N Y Acad Sci.* 2017;1389(1):164–85.

14. Duchene DA, Chowdhury AA, Yang J, Iglesias-Carrasco M, Stiller J, Feng S, Bhatt S, Gilbert MTP, Zhang G, Tobias JA, et al. Drivers of avian genomic change revealed by evolutionary rate decomposition. *Nature*. 2025;641(8065):1208–16.
15. Edwards SV, Fang B, Khost D, Kolyfetis GE, Cheek RG, DeRaad DA, Chen N, Fitzpatrick JW, McCormack JE, Funk WC, et al. Multispecies pangenomes reveal a pervasive influence of population size on structural variation. *Science*. 2025;390(6778):eadw1931.
16. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*. 1980;16(2):111–20.
17. Charlesworth B. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet*. 2009;10(3):195–205.
18. Marino A, Debaecker G, Fiston-Lavier A-S, Haudry A, Nabholz B. Effective population size does not explain long-term variation in genome size and transposable element content in animals. *eLife*. 2025;13:RP100574.
19. Sotero-Caio CG, Platt RN 2nd, Suh A, Ray DA. Evolution and Diversity of Transposable Elements in Vertebrate Genomes. *Genome Biol Evol*. 2017;9(1):161–77.
20. Ohta T. The Nearly Neutral Theory of Molecular Evolution. *Annu Rev Ecol Syst*. 1992;23:263–86.
21. Rishishwar L, Tellez Villa CE, Jordan IK. Transposable element polymorphisms recapitulate human evolution. *Mob DNA*. 2015;6:21.
22. Kramerov DA, Vassetzky NS. Origin and evolution of SINEs in eukaryotic genomes. *Heredity (Edinb)*. 2011;107(6):487–95.
23. Frankham R. Genetics and extinction. *Biol Conserv*. 2005;126(2):131–40.
24. Spielman D, Brook BW, Frankham R. Most species are not driven to extinction before genetic factors impact them. *Proc Natl Acad Sci U S A*. 2004;101(42):15261–4.
25. Pfenninger M, Bálint M, Pauls SU. Methodological framework for projecting the potential loss of intraspecific genetic diversity due to global climate change. *BMC Evol Biol*. 2012;12(1):224.
26. Hohenlohe PA, Funk WC, Rajora OP. Population genomics for wildlife conservation and management. *Mol Ecol*. 2021;30(1):62–82.
27. Payer LM, Burns KH. Transposable elements in human genetic disease. *Nat Rev Genet*. 2019;20(12):760–72.
28. Stiller J, Feng S, Chowdhury A-A, Rivas-González I, Duchêne DA, Fang Q, Deng Y, Kozlov A, Stamatakis A, Claramunt S, et al. Complexity of avian evolution revealed by family-level genomes. *Nature*. 2024;629(8013):851–60.
29. Jebb D, Huang Z, Pippel M, Hughes GM, Lavrichenko K, Devanna P, Winkler S, Jermiin LS, Skirmuntt EC, Katzourakis A, et al. Six reference-quality genomes reveal evolution of bat adaptations. *Nature*. 2020;583(7817):578–84.
30. Nguyen TN, Cosgrove EJ, Chen N, Lehr N, Lokey MG, Beaudry FEG, Fitzpatrick SW, Bowman R, Miller KE, Fitzpatrick JW et al. Florida scrub-jay genomes across space and time reveal impacts of

population decline and reduced gene flow. *Curr Biol* 2026, 36(15):3794–3807 e3794.

31. Bergeron LA, Besenbacher S, Zheng J, Li P, Bertelsen MF, Quintard B, Hoffman JI, Li Z, St Leger J, Shao C, et al. Evolution of the germline mutation rate across vertebrates. *Nature*. 2023;615(7951):285–91.

32. Zhang H, Lundberg M, Tarka M, Hasselquist D, Hansson B. Evidence of Site-Specific and Male-Biased Germline Mutation Rate in a Wild Songbird. *Genome Biol Evol* 2023, 15(11).

33. da Silva VH, Laine VN, Bosse M, Oers KV, Dibbits B, Visser ME, RP MAC, Groenen MAM. CNVs are associated with genomic architecture in a songbird. *BMC Genomics*. 2018;19(1):195.

34. Peona V, Weissensteiner MH, Suh A. How complete are complete genome assemblies?-An avian perspective. *Mol Ecol Resour*. 2018;18(6):1188–95.

35. Fang B, Edwards SV. Fitness consequences of structural variation inferred from a House Finch pangenome. *Proc Natl Acad Sci U S A*. 2024;121(47):e2409943121.

Figures

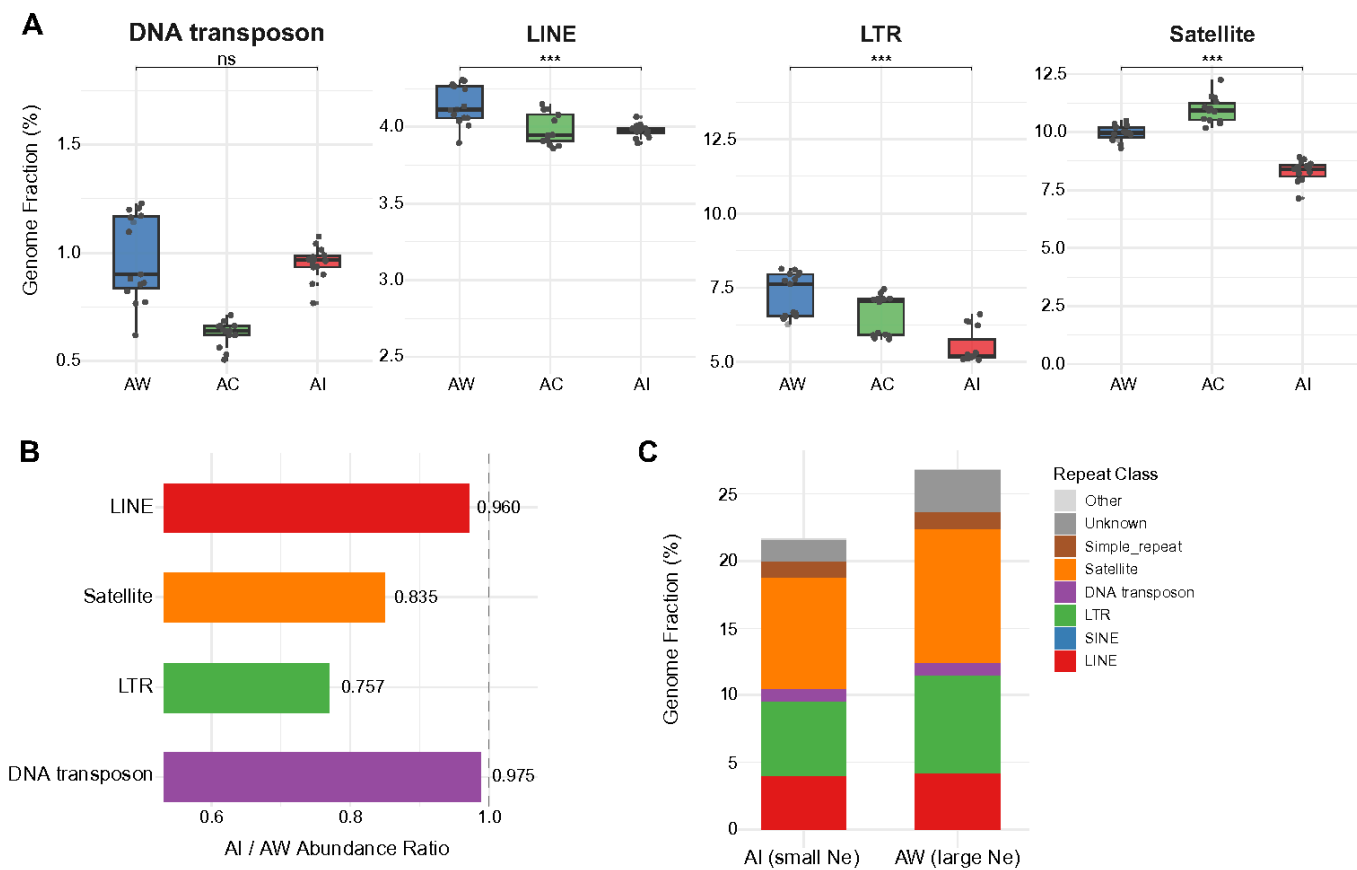


Figure 1

Repeat element composition across the population size gradient. (A) Boxplots showing genomic fraction of four key repeat classes at the individual level (N = 15 AW, 14 AC, 15 AI). Each point represents one diploid individual (average of two haploid assemblies). FDR-corrected Mann-Whitney U tests compare AW vs. AI. Exact p-values and test statistics are reported in **Table S1**. **(B)** AI/AW abundance ratios for each repeat class. Dashed line at 1.0 indicates stability; ratios < 1 indicate depletion in the island species. **(C)** Mean repeat composition stacked bar chart comparing AW and AI.

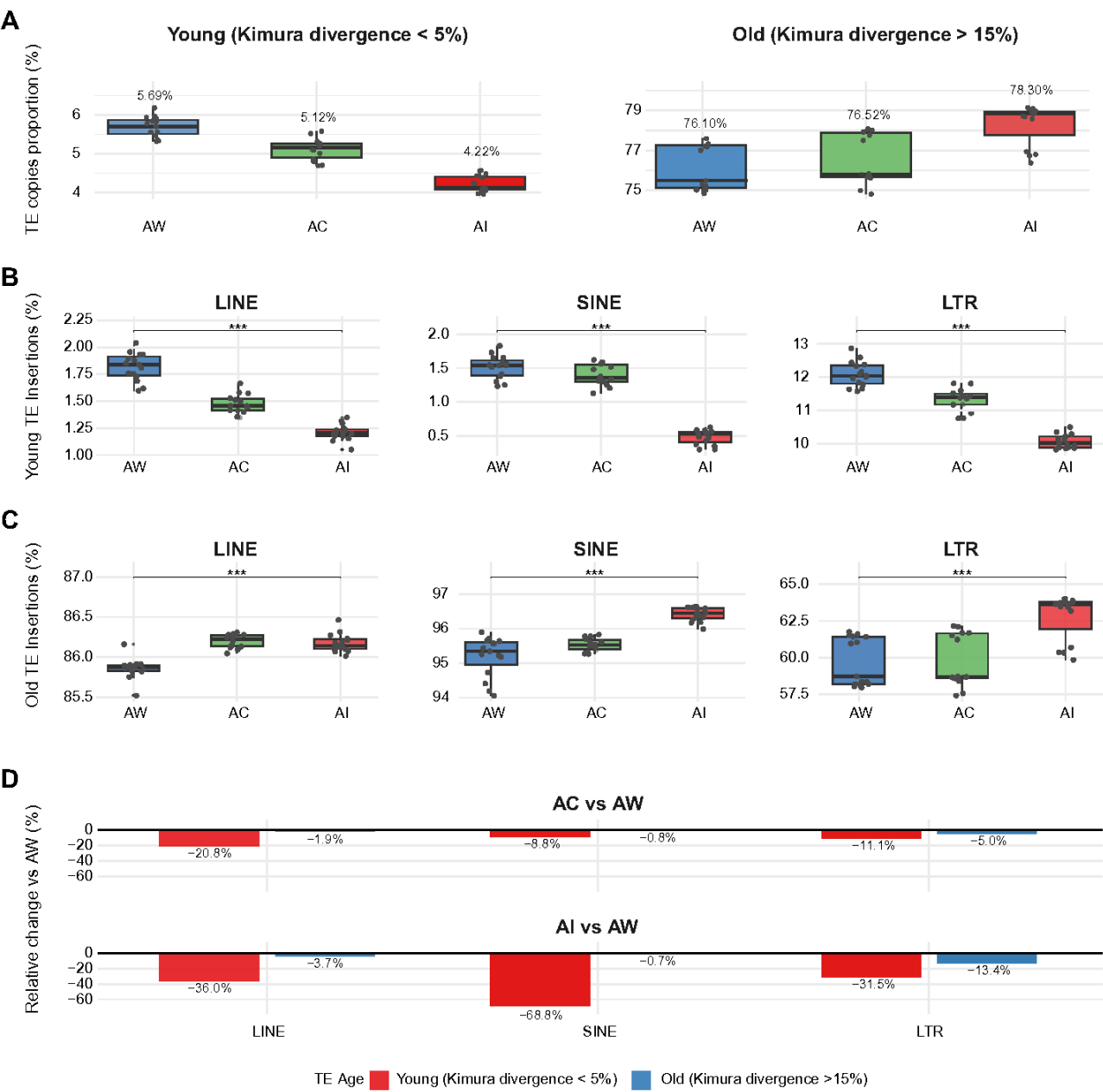


Figure 2

TE age structure reveals ongoing purifying selection in the small population. **(A)** Age composition of TE copies (combined LINE, SINE, LTR) at the individual level, showing the proportion of young (<5% divergence, red) and old (>15% divergence, blue) insertions. Panels are faceted with independent y-axis scales so that both the small young fraction and the large old fraction are visible; numbers indicate the mean percentage per individual. The medium class (5-15%) showed no species-level signal and is omitted. Young copies decline monotonically from AW to AI, while old copies increase. **(B)** Proportion of young TE insertions by TE category at the individual level (N = 15, 14, 15). All three categories show significant depletion in AI (Mann-Whitney U = 225, $p = 1.29 \times 10^{-8}$ for each). Note different y-axis scales reflecting different baseline young proportions. **(C)** Proportion of old TE insertions, showing the complementary increase in AI across all categories. **(D)** Percentage change in AI (upper) and AC (lower) relative to AW for young (<5%) and old (>15%) copies, per category. Negative bars indicate depletion, positive bars enrichment. In all three categories, young insertions decline disproportionately more than old insertions in AI relative to AW, consistent with selective removal of new insertions rather than wholesale TE loss.

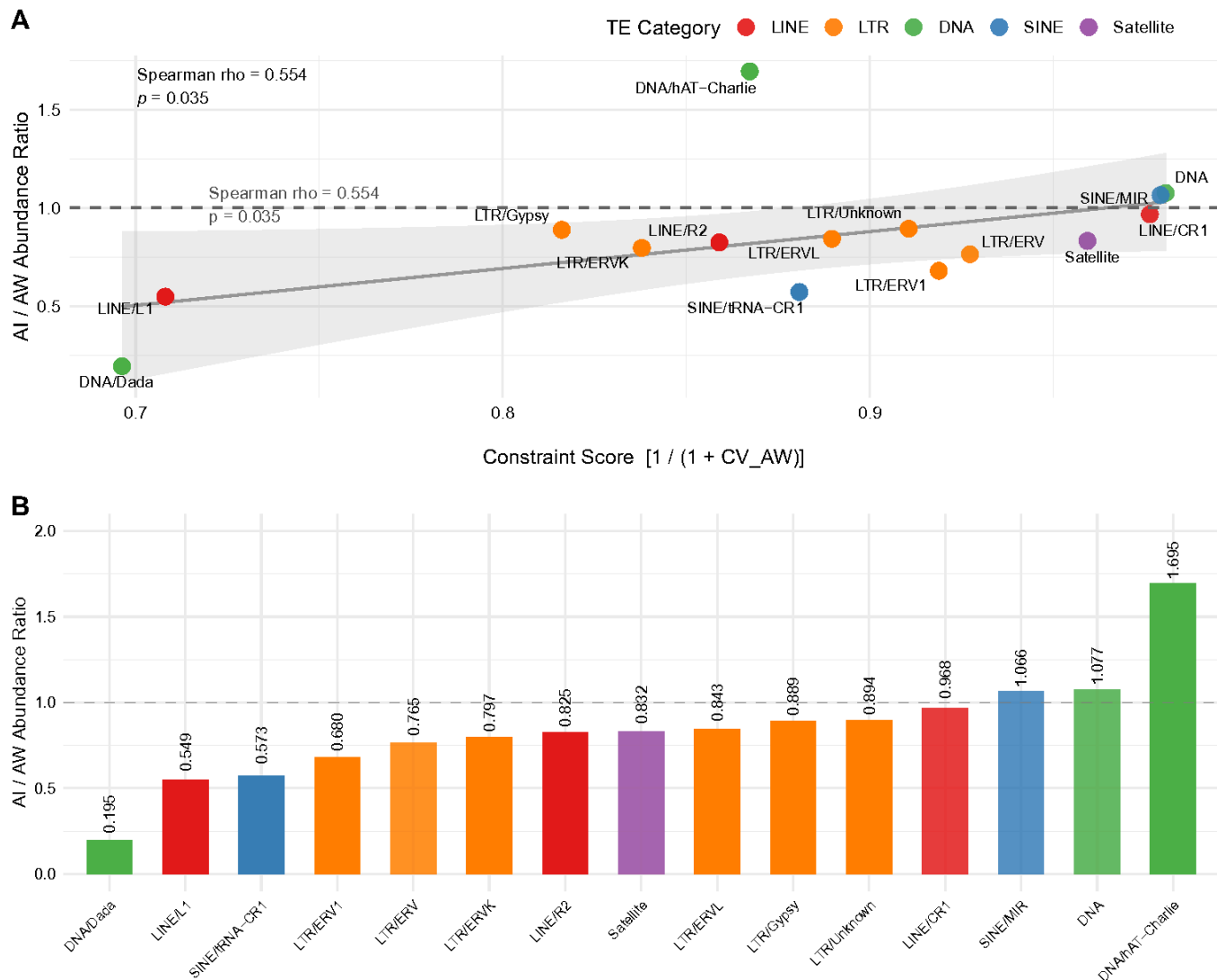


Figure 3

Subfamily-level constraint correlates with TE stability under population decline. (A) Scatter plot showing the relationship between constraint score $[1/(1+CV_AW)]$ and AI/AW abundance ratio for 15 TE subfamilies. Spearman rho = 0.554, $p = 3.23 \times 10^{-2}$. (B) AI/AW abundance ratios for all 15 subfamilies, sorted in ascending order. Dashed line at 1.0 indicates stability.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTables.xlsx](#)

- [Supplementarymaterials.docx](#)