Mitonuclear divergence predicts gradual speciation in animal hybrid zones

Derek Eddo+, Bailey Rouse+, Silu Wang*
Department of Biological Sciences
State University of New York, Buffalo

+ Equal contributions

*Corresponding author: siluwang.biodiv@gmail.com

Abstract

The core of speciation is the genetic incompatibilities underlying the evolution of reproductive isolation. Hybrid zones provide unique opportunities to unravel the evolutionary rate of reproductive isolation in the origin of species. The selection against hybrids accrues with increased genetic incompatibilities and drives the evolution of reproductive isolation in the face of gene flow. There have been decades of debates over the relationship between the selection against hybrids and the genetic divergence between parental lineages. The debates occur primarily among three models: (1) the Additive Effect Model predicts a linear growth of selection against hybrids with the divergence of parental lineages; (2) the Snowball Effect Model predicts exponential growth of selection; whereas (3) the Slowdown Effect Model predicts a logarithmic growth of selection. Here, we tested the three models with animal hybrid zones worldwide. The Slowdown Effect Model (3) is best supported with the full dataset. We refined the three models to consider independent and interactive effects of mitochondrial (mtDNA) and nuclear genetic (nDNA) divergence on the selection against hybrids. The refined Snowball Model (2) was best supported by the data, revealing a significant effect of nuclear genetic distance and its interaction with mtDNA distance on the selection against hybrid DNA in the hybrid zones. Collectively, this data synthesis in the early stage of animal speciation reveals a gradual development of reproductive isolation with mitonuclear genetic divergence.

Keywords: Speciation rate, Hybrid zone, Snowball Effect, Genetic Incompatibility, Speciation, Hybridization, Cline.

Introduction

The evolution of reproductive isolation is underpinned by genetic incompatibilities, which accrue as the diverging lineages accumulate lineage-specific substitutions (Dobzhansky 1934; Orr 1995). The increased genetic incompatibilities contribute to the selection against hybrids (Palopoli and Wu 1994; Orr 1995; Welch 2004), which attenuate gene flow between diverging lineages and lead to reproductive isolation and speciation (Gavrilets 2004; Nosil 2008; Westram et al. 2022). The rate at which lineage-specific substitutions drive the selection against hybrids ultimately prescribes the speciation rate (Coyne and Orr 1989). The debates over this relationship have inspired decades of investigation of genomic architecture and mechanisms of speciation (Feder et al. 2014; Seehausen et al. 2014). If divergent substitutions independently contribute to reproductive isolation, the genetic distance between the diverging lineages should predict linear selection against gene flow. If two-loci epistatic incompatibilities, or Dobzhansky & Muller Incompatibilities (DMIs)(Dobzhansky 1934; Muller 1942), are prevalent, the selection against hybrids would increase exponentially as genetic distance increases, reflecting a ‘Snowball Effect’ (Orr 1995). A greater acceleration in speciation rate would be observed if the genetic distance results in abundant higher-order incompatibilities (Orr 1995). In contrast, an early ‘slowdown’ might be observed if premating isolation is established prior to substantial genetic incompatibilities (Gourbiere and Mallet 2010).

Several taxon-specific empirical studies have supported the Snowball Effect hypothesis, where the increase of genetic distance between parental lineages exponentially suppresses hybrid fitness (Presgraves 2002; Matute et al. 2010; Moyle and Nakazato 2010). However, others reported a more gradual, linear accumulation of selection against hybrids in response to

genetic divergence (Lijtmaer et al. 2003; Gourbiere and Mallet 2010; Stelkens et al. 2010). A logarithmic relationship was also observed, reflecting a Slowdown Effect (Gourbiere and Mallet 2010). Collectively, the taxon-specific patterns concur with the heterogeneous speciation rates across the tree of life that await large-scale comparative studies to identify evolutionary principles underlying the regularities about speciation rates.

The heterogeneous patterns of speciation rate are likely attributed to the variability of locus-specific effect sizes and the probability of incompatibility in different parts of the genome. For instance, the smaller effective population size in the mitochondrial genome than the nuclear genome predicted greater drift, counteracting selection (Slatkin and Maruyama 1975). In addition, mitochondrial deleterious mutations reportedly exhibit a greater average effect size than nuclear mutations (Popadin et al. 2013). Moreover, the probability of incompatibility can differ between intergenomic and intragenomic discordance.

More broadly, the heterogeneous patterns of selection against hybrids could also be explained by the stages of speciation along the speciation continuum (Roux et al. 2016). In the early stages of divergence, large-effect single-locus genetic barriers are more effective than DMIs to overcome extensive gene flow (Felsenstein 1981; Wu 2001; Wang et al. 2020; de Zwaan et al. 2022). If the additive effect sizes of single-locus barriers were constant through time, the selection against hybrids should linearly increase as lineage-specific substitutions accumulate over time (Fig. 1, Additive Effect). If the effect sizes were larger earlier (Fisher 1930) in speciation, when the lineage-specific substitutions were scarce and exhibiting diminishing returns in response to further increments of genetic divergence, a Slowdown Effect in the form of a logarithmic function is expected (Gourbiere and Mallet 2010) (Fig. 1, Slowdown

Effect). As the gene flow between diverging lineages attenuates, allowing DMIs to build up, a Snowball Effect can be observed (Orr 1995) (Fig. 1, Snowball Effect), where the selection against hybrids grows exponentially as lineage-specific substitutions accrue. The heterogeneous tension between genetic incompatibility and gene flow across the speciation continuum is naturally encapsulated in decades of hybrid zone studies around the world (Fig. 2).

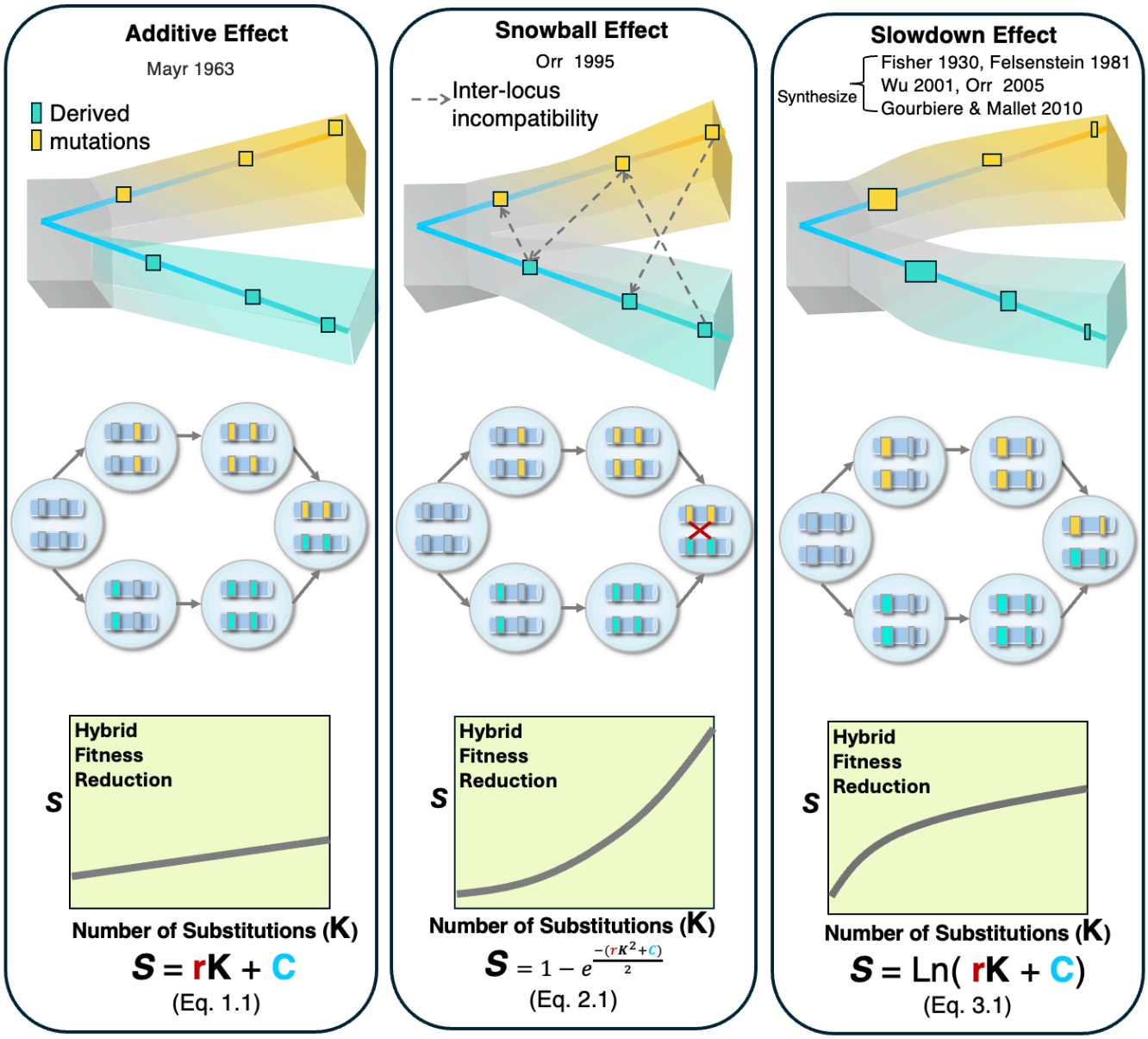

Fig. 1 Three speciation models: (1) Additive Effect, **(2)** Slowdown Effect, and **(3)** Snowball Effect.

Hybrid zones are geographic areas where diverging lineages hybridize (Barton 1979). They provide unique opportunities to observe the progression of speciation as the tension between genetic incompatibility and gene flow is resolved (Barton and Hewitt 1985). This tension is effectively theorized as the tension zone model, where the width of the cline (w) is a balance of selection (s) and dispersal (σ) (Fig. 2A) (Barton 1979). The selection (s) is a composite expression of genetic incompatibility and divergent selection, or reproductive isolation in the face of gene flow. The rate at which *s* increases with the extension of genetic distance between diverging lineages ultimately translates to the speciation rate (Fig.2B).

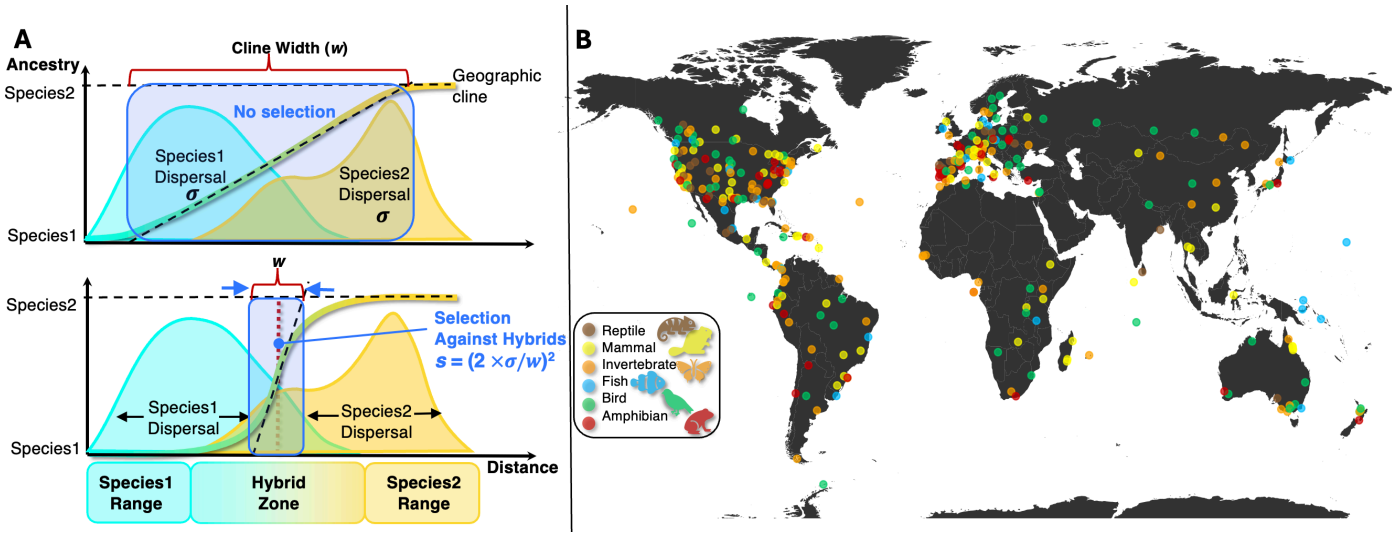

Fig.2 Hybrid zones are natural laboratories for understanding speciation. A, The cline width (*w*) of a hybrid zone is a balance of gene flow (σ) and selection (*s*) against hybrids. In the absence of s, σ will widen *w* over time as species 1 hybridizes with species 2 (top). Bottom: strong selection against hybrids counteracts with σ and shrinks *w*. **B**, World distribution of animal hybrid zones. Over 330 animal hybrid zones have been reported. Color categories correspond to six different organismal groups: mammals in yellow (65), birds in green (71), reptiles in brown (23), amphibians in red (45), invertebrates in orange (100), and fish in blue (26).

Here, we conduct a data synthesis to examine the buildup rate of selection against hybrids (*s*) in animal hybrid zones (Fig. 2). We tested three alternative models that predict the selection against hybrids with the genetic distance between parental lineages. Model 1 predicts a linear increase of *s* as genetic distance increases (Fig. 1, Additive Effect); Model 2

predicts an exponential growth of *s* (Fig. 1, Snowball Effect); while Model 3 predicts a logarithmic relationship (Fig. 1, Slowdown Effect).

We further developed three refined models to dissect mitochondrial and nuclear genetic effects on speciation. Since the mitochondrial and nuclear genomes have distinct substitution rates (Ballard and Whitlock 2003) and coevolve to maintain critical eukaryotic energetics (Eddo et al. 2025), 'Endosymbiotic Snowball Effect' was articulated (Wang et al. 2025). Therefore, we further partitioned the genetic distance into mtDNA and nDNA distance to understand their independent and interactive effects on speciation. Together, this study will inform the mode of speciation in the origin of animal biodiversity.

Method

Hybrid zones

We studied 330 animal hybrid zones reported to date (McEntee et al. 2020; Nieto Feliner et al. 2023) and categorized them into six taxonomic groups: amphibians, birds, fishes, invertebrates, reptiles, and mammals (Fig. 2B). We filtered the data with the following criteria: (1) there has been sufficient spatial sampling to fit the genetic and/or phenotypic clines across the hybrid zone; (2) there have been direct estimations of cline widths with genetic markers; (3) there is sufficient information on dispersal distance in the unit of Kilometer/generation or Kilometer/generation$^{0.5}$; (4) for the dispersal distance reported in Kilometer/generation, there is information on generation time to calculate the Kilometer/generation$^{0.5}$. These criteria resulted in 148 animal hybrid zones for model testing (see below).

Mitonuclear genetic distance

To infer the genetic distance between hybridizing species, we calculated the genetic distance of the most sequenced mitochondrial and nuclear genes in animals. We searched for

COI, ND2, ND3, ND4, CYTB, and MSH1 for mitochondrial genes and RAG1, RAG2, MC1R, BDNF, ODC1, ASIP, and ACTB for nuclear genes. We fetched DNA sequences of the target genes for each parental species from GenBank and only kept the genes if both species had at least three sequences per species. To ensure even sampling depths among the species pairs in our meta-analysis, for species with more than 10 sequence records of a gene, we randomly sampled 10 sequences among all the records. We then aligned all the sequences of each gene in each species pair with the msa function in the msa package (Bodenhofer et al. 2015) in R (R Core Team 2023). Specifically, we used MUSCLE (Edgar 2004) with gapOpening = 20 and gapExtension = 10. With the alignment, we calculated the mean pairwise distance between parental species with dist.dna function with the TN93 substitution rate model in the ape package (Paradis and Schliep 2019). Then, we averaged the between-species distance of all the mitochondrial genes and nuclear genes separately to result in the mean pairwise mtDNA distance ($\overline{K_m}$) and the mean pairwise nDNA distance ($\overline{K_n}$) between each pair of parental species. Finally, we calculated the mean pairwise genetic distance ($\overline{K}$) between each pair of parental species by averaging the nDNA and mtDNA distances. To account for the uncertainty in $\overline{K}$ estimations, we calculated the standard deviation of $\overline{K_m}$, $\overline{K_n}$, and $\overline{K}$. Among the 148 hybrid zones, 77 hybrid zones had sufficient genetic information to calculate the genetic distance between parental species. To evaluate the reliability of our genic estimation of parental species divergence, we computed the ratio of between-species distance over the sum of between- and within-species distance and conducted correlation tests of F_{ST} estimates for several pairs of species reported in the literature (see supplementary data).

Selection against hybrids

We extracted the dispersal, center, and width of hybrid zones from McEntee et al. (2020), Nieto Feliner et al. (2023), and additional literature (supplementary data). We unified the unit of dispersal rate as kilometers (KM) /$\sqrt{generation}$ with the average generation time information of each species pair. To account for sex-specific dispersal, we conducted another round of literature search for female vs male dispersal rate in each species pair (supplementary information). We calculated the average cline width (w) and center for mtDNA and nDNA data, as well as phenotypic data separately. We calculated selection (s) against heterozygotes as a function of cline width (w) and dispersal (σ): $s = (2 \times \frac{\sigma}{w})^2$ (Fig. 1) (Barton and Gale 1993). To infer the cumulative selection against hybrids, we substituted w with averaged cline width ($\overline{w}$) among mtDNA and nDNA clines (Fig. 3B). To account for uncertainty of selection estimates due to heterogeneous dispersal rates and cline width estimates, we monitored sex-specific dispersal and the variance of cline width estimates for each hybrid zone. For hybrid zones with known sex-specific dispersal rates, we calculated selection on mtDNA and nDNA, respectively, as $s_{mt} = (2 \times \frac{\sigma(female)}{w})^2$ and $s_{mt} = (2 \times \frac{\sigma(average)}{w})^2$. We derived the standard deviation of s, SD(s), as ($8\sigma^2 / w^3$) SD(w).

Speciation models

To investigate the relationship between genetic distance and strength of selection against hybrids (s) in the hybrid zones, we test three conventional hybrid incompatibility models: (1) Linear Effect (Eq. 1.1), (2) Snowball Effect (Eq. 2.1; Fig. S1), and (3) Slowdown Effect (Eq. 3.1). In each model, selection (s) is a function of genetic distance (K) with a rate

parameter *r* and a constant parameter *C* as divergent plasticity or genetic assimilation (Pigliucci and Murren 2003; Ehrenreich and Pfennig 2016; Mueller et al. 2025). The rate parameter corresponds to the average effect size of each pair of incompatibilities.

$$s = rK + C \quad \text{(Eq. 1.1)}$$

$$s = 1 - e^{-(rK^2 + C)} \quad \text{(Eq. 2.1)}$$

$$s = ln(rK + C) \quad \text{(Eq. 3.1)}$$

To account for the effects of mtDNA and nuclear genetic distance between the parental species (Fig. 3), we further refined the conventional expressions (Eq. 1.1, 2.1, and 3.1) into mitonuclear models (Eq. 1.2, 2.2, and 3.2) to partition *K* into K_m (mtDNA distance) and K_n (nuclear DNA distance) that correspond to rate parameters r_m and r_n, respectively, besides a constant of *C*.

$$s = r_m K_m + r_n K_n + C \quad \text{(Eq. 1.2)}$$

$$s = 1 - e^{-((r_m K_m + r_n K_n)^2 + c)} \quad \text{(Eq. 2.2)}$$

$$s = ln(r_m K_m + r_n K_n + C) \quad \text{(Eq. 3.2)}$$

We conducted model fitting of the above three models with the nls2 function of nls2 package (Grothendieck, 2007) in R (R Core Team 2023). With the default algorithm and starting value of $r = 0.01$, $C = 0.1$, model fitting was conducted. The AIC of each fitted model is obtained with the AIC function. In each conventional model fitting (Eq. 1.1, 2.1, 3.1), we estimated the rate parameter *r* and constant *C*. To evaluate taxon-specific patterns, we conducted conventional model fitting (1.1, 2.1, 3.1) within each taxonomic group (Amphibians, Birds, Invertebrates, Mammals, and Reptiles), except for Fishes, due to data insufficiency, Furthermore, we estimated r_m, r_n, and *C* in the refined model fitting (1.2, 2.2, 3.2). Only 18

hybrid zones had sufficient mtDNA and nDNA sequence data to participate in refined model fitting, thus further taxon-specific modeling fitting was omitted.

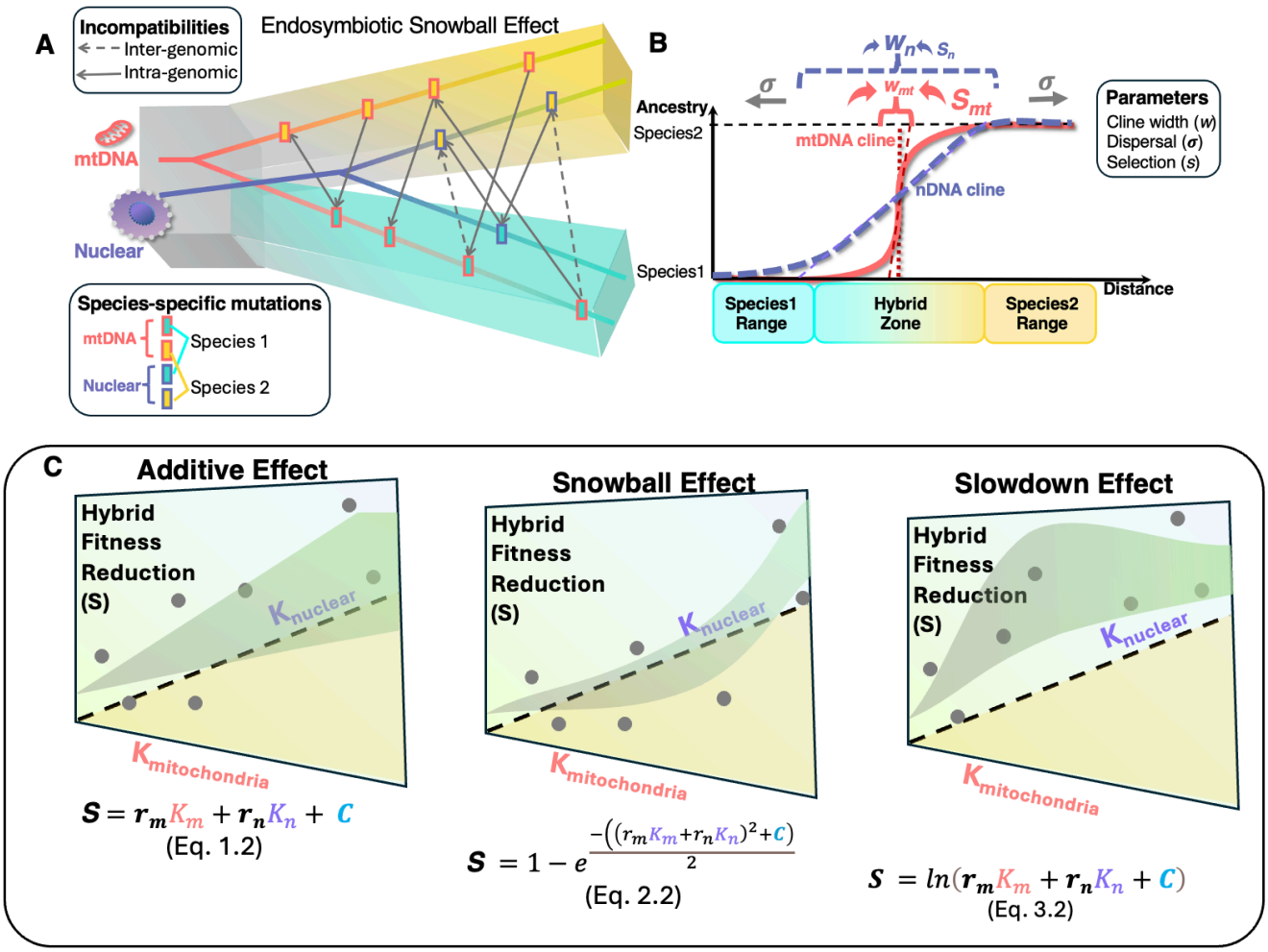

Fig. 3 Mitonuclear genetic divergence underlying the selection against hybrids. A. Endosymbiotic Snowball Effect (Wang et al. 2025): Lineage-specific substitutions in the mitochondrial genome (red) and nuclear genome (purple) can lead to intragenomic and intergenomic Dobzhansky-Muller Incompatibilities, which emergently contribute to selection against hybrids. **B.** This selection counteracts gene flow (σ) across hybrid zones and reduces the cline width (w). Greater substitution rates in the mitochondrial genome than the nuclear genome can result in a greater rate of incompatibility and thus stronger selection on mtDNA clines than the nDNA clines. **C**, Refined speciation models with independent and interactive mitonuclear effects: (1) Additive (Eq. 1.2), Snowball (Eq. 2.2), Slowdown (Eq. 3.2).

To test the significance of observed regression coefficients, we conducted 1,000 bootstrap iterations to generate a bootstrap distribution of regression coefficients $\hat{r}$ and $\hat{C}$. In each bootstrap iteration, we randomly sampled N = 148 hybrid zones with replacement from

the full dataset. For each bootstrap sample of $K^* = [K_1^*, K_2^*, \ldots K_i^*, \ldots K_{148}^*]$ and $s^* = [s_1^*, s_2^*, \ldots s_i^*, \ldots s_{148}^*]$, we simultaneously fitted the sample to (1) Additive, (2) Snowball, and (3) Slowdown Effect Models (Fig.1; Eq. 1.1, 2.1, & 3.1) to calculate $\hat{r}$ and $\hat{C}$. To account for the uncertainty in K and s, we randomly sampled K_i^* and s_i^* respectively from normal distributions: $N(K_i, sd(K_i))$ *and* $N(s_i, sd(s_i))$. To generate a null distribution of $\hat{r}$ and $\hat{C}$, we randomly shuffled and mismatched K^* and s^* in each bootstrap sample and simultaneously fitted the three models (Eq. 1.1, 2.1, & 3.1) again. We used paired-T tests to compare the observed distribution of *r* and *C*, *r_Obs* and *C_Obs, to* the corresponding null distributions, *r_H0* and *C_H0*.

A similar bootstrapping test was applied to the refined models (Fig.3C; Eq. 1.2, 2.2, & 3.2), where (1) Additive, (2) Snowball, (3) Slowdown Models were compared in terms of AICs, and the observed bootstrap distribution of $\widehat{r_m}$, $\widehat{r_n}$, and $\hat{C}$, r_m*_Obs*, r_n*_Obs* and *C_Obs*, was respectively compared to r_m*_H0*, r_n*_H0*, and *C_H0*.

Results

Discordant mitonuclear selection

There is significantly greater genetic distance in mtDNA than nDNA between parental species (mean difference = 0.036, 95% CI: 0.023-0.049, $p < 10^{-11}$). Concordantly, there is greater selection on mitochondrial clines than nuclear clines in hybrid zones (mean difference = 0.110, 95% CI: 0.024-0.196, $p < 0.05$), reflected by the narrower mtDNA cline width (Fig. 4). The discrepancy of selection on mitochondrial and nuclear genetic clines is significantly lower when there is greater genetic distance between parental species ($R^2 = 0.08$, $p = 0.04$). Our

mitonuclear divergence estimates were significantly correlated with F_{ST} estimates from heterogeneous sequencing datasets reported in the literature ($r = 0.50$, $p < 0.05$).

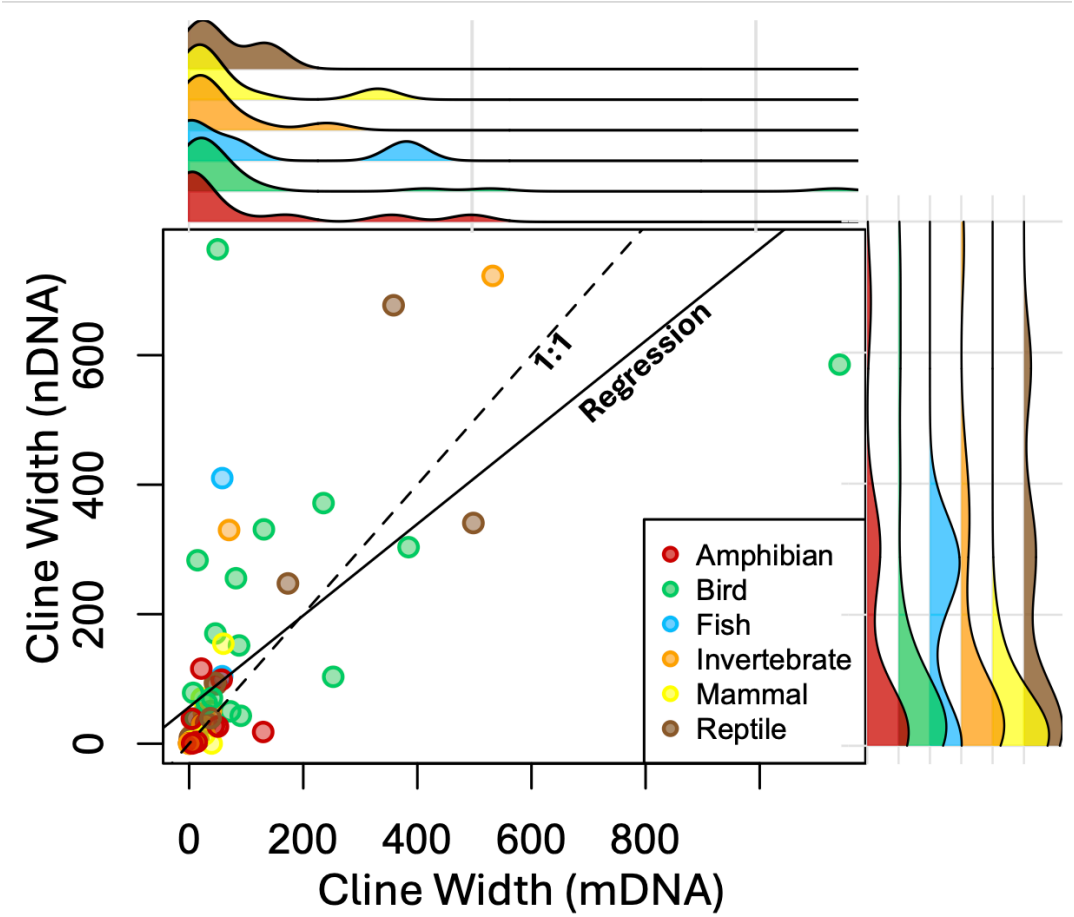

Fig. 4 Mitonuclear cline widths in animal hybrid zones. The cline widths of mtDNA were significantly narrower than nDNA (one-tailed paired-T test, $p < 0.05$). The two variables are significantly correlated ($r = 0.67$, $p < 10^{-9}$).

Speciation models

The Slowdown Effect model (3.1) (Fig. 1, 5A) has the lowest AIC (Table 1). The Slowdown model continues to fit superiorly for all the taxon-specific data except for birds and mammals, where the Snowball Effect and Linear model were respectively selected (Table 1). The regression coefficient of *r* of the Slowdown Effect model fitted to the full data was 3.078 (Table 1). Similarly, the bootstrap mean of *r_Obs* in the Slowdown Effect model (Eq. 3.1) was

3.534 (Fig. 5B), which was significantly different from *r_H0* (mean difference: 3.41, 95%CI of the difference: 3.09 to 3.74; $p < 10^{-15}$), indicating a significantly positive effect of genetic divergence on selection against hybrids. However, the estimates of *r* were heterogeneous among taxonomic groups (Table S1), with birds ($r = 17.25$) and mammals ($r = 4.82$) being the highest, invertebrates ($r = 2.59$) and reptiles ($r = 1.34$) intermediate, and lower for amphibians ($r = 0.71$) (Table 1).

Table 1 The effect of genetic divergence on the selection against hybrids among animal taxa. Standard errors (SE) of the regression coefficients for the rate parameter(*r*) and constant (*C*). The models (Fig. 1, Eq 1.1-3.1) with the lowest AICs are in bold. The Slowdown model (equation 3.1) has the lowest AIC fitted to the full dataset.

Taxa	**Models**	***r***	**SE (*r*)**	***C***	**SE (*C*)**	**AIC**
All	Additive	2.54	1.16	0.03	0.04	15.03
	Snowball	5.25	1.55	0.07	0.04	16.93
	Slowdown	**3.08**	**1.40**	**1.03**	**0.05**	**14.78**
Amphibian	Additive	0.60	2.05	0.09	0.11	-5.04
	Snowball	1.04	11.44	0.12	0.09	-4.93
	Slowdown	**0.71**	**2.30**	**1.09**	**0.12**	**-5.05**
Bird	Additive	5.87	3.62	0.04	0.11	26.04
	Snowball	**17.25**	**5.73**	**0.03**	**0.11**	**25.28**
	Slowdown	8.34	5.40	1.02	0.13	25.80
Invertebrate	Additive	2.20	2.50	0.02	0.11	4.62
	Snowball	4.81	3.40	0.06	0.10	4.94
	Slowdown	**2.59**	**2.91**	**1.02**	**0.12**	**4.58**
Mammal	**Additive**	**4.82**	**1.58**	**-0.03**	**0.04**	**-24.05**
	Snowball	8.81	1.72	0.02	0.03	-23.89
	Slowdown	5.46	1.97	0.96	0.04	-23.89
Reptile	Additive	1.17	2.11	0.02	0.11	0.55
	Snowball	2.44	4.42	0.05	0.09	0.83
	Slowdown	**1.34**	**2.31**	**1.01**	**0.11**	**0.53**

When considering the effects of mitochondrial and nuclear genetic distance separately (Fig. 3), the Snowball Effect model (Eq. 2.2) has the lowest AIC (Table 2; Fig. 5 C-D). We observed a significantly positive effect of nuclear DNA distance, a small negative effect of

mDNA distance independently, but a positive interaction of mitonuclear distance (Fig. 5C) on the selection against hybrid DNA. Specifically, the r_n was estimated to be 31.18 (Table 2). In the Snowball Effect Model, bootstrap mean of r_n, 26.82, was significantly different from r_n_H0 (mean difference: 13.27, 95%CI of the difference: 11.09 to 15.44; $p < 10^{-15}$) (Fig. 5D). In contrast, the regression coefficient of r_m was -1.85, similar to the bootstrap mean (-2.42, Fig. 5D), which was significantly more negative than r_m_H0 (mean difference: -1.43, 95%CI of the difference: -1.76 to -1.09; $p < 10^{-15}$).

Table 2 Estimates and standard errors (SE) of the r_m, r_n, and C of the refined models. The Snowball Effect model has the lowest AIC.

Models	r_m	**SE (r_m)**	r_n	**SE (r_n)**	C	**SE (C)**	**AIC**
Additive	-0.65	1.16	7.57	7.49	0.07	0.06	-19.64
Snowball	**-1.85**	**6.02**	**31.18**	**18.29**	**0.06**	**0.03**	**-20.76**
Slowdown	-0.71	1.24	7.95	8.54	1.07	0.06	-19.58

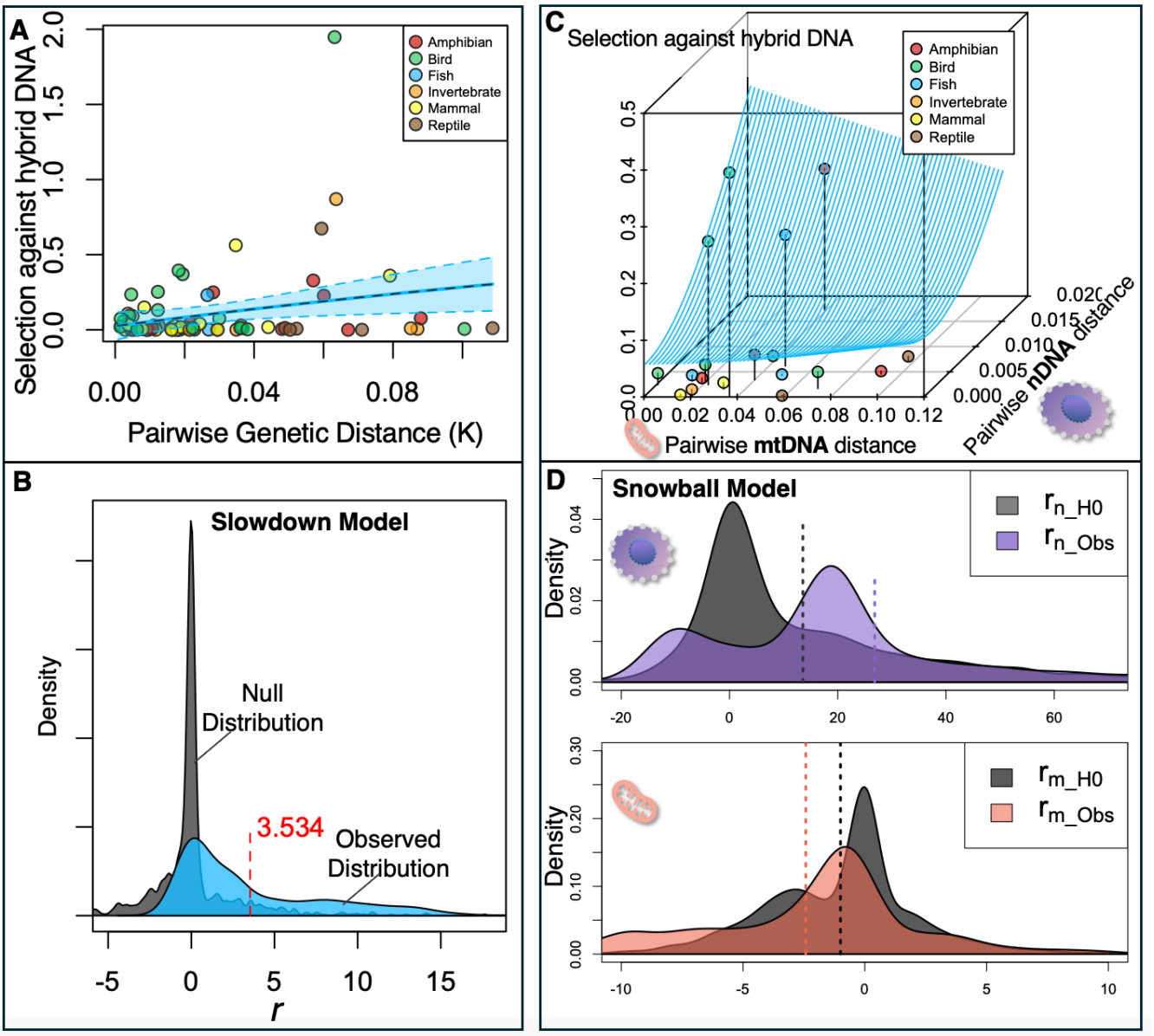

Fig. 5 Coefficients of estimates in the best-fitted models for simplified versus refined models of speciation. A, Conventional models (equations 1.1, 2.1, and 3.1) considered the total genetic distance as the predictor for selection against hybrids. The Slowdown Effect model (dotted line) has the lowest AIC. **B**, Bootstrap distributions of the rate parameter *r* (equation 1.3) estimations based observed relationship between *K* and *s* across animal hybrid zones (blue) versus *r* estimates based on randomly mismatched *K* and *s* (grey). Estimates of *r* are from the best-fitted conventional model. **C**, Refined models considered the effect of mtDNA and nDNA divergence and their interactions separately. The Snowball Effect model has the lowest AIC. **D**, The bootstrap distribution of rate parameter estimations of r_m and r_n, which corresponds to the effects of mtDNA and nDNA distance separately. Estimates of r_m and r_n are from the best-fitted refined model.

Discussion

The accumulation of selection against hybrids in the hybrid zones between diverging lineages is key to unraveling speciation rates *in situ*. Here, we leveraged decades of hybrid zone studies to understand the evolution of reproductive isolation in the face of gene flow. Overall, the selection against hybrids increased with genetic divergence between the parental species, despite the heterogeneous rates of this relationship (Table 1, 2; Fig. 5).

The Slowdown model (equation 3.1) best fitted the full data among the three conventional models (equation 1.1, 2.1, 3.1) (Fig. Table 1, 4 A-B). This result is consistent with the Slowdown Effects of parental genetic distance previously observed in sympatric starfish and fruit fly (Gourbiere and Mallet 2010). However, the delta AIC was small (Table 1), which suggests that the superiority of the Slowdown model over the other components was weak. The slowdown model predicts an opposite pattern from the Snowball Effect (2.1) expectation at the early stage of speciation, with an early decelerating buildup of selection against hybrids as the parental lineages diverge. One explanation for this early deceleration is the establishment of premating isolation prior to extensive genetic divergence (Gourbiere and Mallet 2010). Early premating isolation would contribute to the reduction in gene flow, leading to narrower cline width over dispersal across the hybrid zones. The early premating isolation without widespread genetic divergence between the parental lineages can be underpinned by a single or few large-effect substitutions (Nosil and Schluter 2011; Servedio et al. 2011), genetic assimilation (Waddington 1961), and/or cultural evolution (Lachlan and Servedio 2004). These mechanisms might experience organismal constraints and thus are heterogeneous among taxonomic groups. For example, vocal learning in songbirds might uniquely increase the

likelihood of cultural evolution as premating isolation (Slabbekoorn and Smith 2002; Lachlan and Servedio 2004), though it can also disrupt reproductive isolation. Indeed, we observed the heterogeneity of rate parameter estimations among taxonomic groups (Table 1).

In fact, the rate parameter was higher in endotherms (Table 1), birds, and mammals, with avian hybrid zones fitting the Snowball model better. The relatively lower rates in ectotherms (Table 1), amphibians, reptiles, and invertebrates could indicate a relatively smaller genetic effect on reproductive isolation at the early stage of speciation in these groups. High basal metabolic rates in endotherms could require stronger selection on genetic incompatibility, especially between the mitochondrial and nuclear genomes (Hill 2019). As endotherms exhibit faster climatic niche shifts than ectotherms (Rolland et al. 2018), the underlying genetic divergence between endothermic sister populations could disproportionately divergent adaptation. Both the proximate and ultimate differences in endotherms vs ectotherms could lead to greater effect sizes of genetic divergence on the selection against hybrids.

When we consider the mtDNA and nDNA divergence separately (Eq. 1.2, 2.2, 3.2), Snowball Effect model (Eq. 2.2) fits the data best (Fig. 1). In this model, nuclear genetic divergence and its interaction with mitochondrial divergence showed large positive effects, while the independent mitochondrial divergence exhibited a mild negative effect (Fig. 5C, Table 2). This result is concordant with the conventional model (Fig. 5A) in terms of the overall positive effect of genetic divergence on selection against hybrids, but discordant in the modes. Such discordance illuminates the heterogeneity of effects in different parts of the genomes. Combining a rapid, exponentially growing nDNA effect with a mildly negative mtDNA effect can

give rise to a Slowdown cumulative genetic effect. The three speciation models (Fig. 1) are not necessarily mutually exclusive and could be parallel, sequential, or hierarchical in the genome evolution underlying speciation. Future combinatorial examination of the three reductionist models in different parts of the nuclear genome could further unveil the evolutionary genomic mechanism of speciation.

The interaction of mtDNA and nDNA distance showed a positive effect on selection against hybrids (Fig. 5C). This pattern agrees with the growing evidence that mitonuclear incompatibility is important for early-stage speciation (Hill 2017; Wang et al. 2021; Burton 2022; Moran et al. 2024). However, the mildly negative mitochondrial independent effect is surprising, as the greater mutation rates of mtDNA than nDNA in animal cells (Ballard and Whitlock 2003; Lynch et al. 2006) were expected to contribute to greater interspecific incompatibility resulting from mitochondrial substitutions (Orr 1995). In addition, the roughly 5-fold greater effect sizes of mitochondrial deleterious mutations than nuclear mutations (Popadin et al. 2013) could further contribute to hybrid unfitness. However, greater purifying selection (Popadin et al. 2013) in the mitochondrial genome and compensatory evolution of the nuclear genome (Princepe and De Aguiar 2024) might drastically dampen the effect size of independent mitochondrial substitutions on hybrid fitness, which is consistent with our observation (Fig. 5C). This notion is also supported by the narrower mtDNA cline than the nDNA cline (Fig. S1) despite the smaller effective population size of mtDNA than nDNA expected greater drift and a wider cline under neutrality (Slatkin and Maruyama 1975).

Caveats

The available 330 animal hybrid zone studies have been concentrated in the global north (Fig. 2B), with relatively fewer studies in the more biodiverse global south. Our observation could be predominantly representing animal speciation patterns in temperate ecoregions, which may be different from those in tropical ecoregions. There are far fewer aquatic hybrid zones than terrestrial hybrid zones (Fig. 2B), which may constrain our conclusion within the terrestrial domain. Life history information, such as independently measured generation time and dispersal rate, was unavailable for a large proportion of the hybrid zones, with only 148 of the 330 cases containing sufficient information to calculate selection against hybrids. For the 148 hybrid zones for which we could calculate selection, there were insufficient sequence data in GenBank for commonly sequenced genes in the nuclear genomes of the species pairs. This sequencing gap limits the inference from the refined models to a small sample size. Despite the limited dataset, the significant mitonuclear effect on the selection against hybrid DNAs supported the Endosymbiotic Snowball Effect (Wang et al. 2025). Future studies that broaden the sequencing scope of hybrid zone studies could provide valuable modeling resolution to illuminate speciation heterogeneity across the genomes. Since the understanding of organismal gene flow is crucial for biodiversity science and conservation in the increasingly fragmented biosphere (Wang and Yoder 2025), a standardized global hybrid zone database could be productive for speciation and conservation science.

Conclusion

We investigated the evolutionary mode of reproductive isolation in the early stage of speciation, empowered by decades of animal hybrid zone studies. We found a gradual buildup

of selection against hybrids as the parental lineages diverged. Further dissection of mitochondrial and nuclear genetic divergence revealed exponential growth of selection against hybrid DNA, consistent with the Endosymbiotic Snowball Effect (Wang et al. 2025). Future dissection of the genomic heterogeneity of hybrid fitness effects would be indispensable for refining our understanding of speciation rates across the tree of life.

Reference

Ballard, W. O., and M. Whitlock. 2003. The incomplete natural history of mitochondria. Molecular Ecology 13:729–744.

Barton, N. H. 1979. The dynamics of hybrid zones. Heredity 43:341–359.

Barton, N. H., and K. S. Gale. 1993. Genetic Analysis of Hybrid Zones. Pp. 13–45 *in* Hybrid zones and the evolutionary process. Oxford University Press, Oxford, UK.

Barton, N. H., and G. M. Hewitt. 1985. Analysis of Hybrid Zones. Annu. Rev. Ecol. Syst. 16:113–148.

Bodenhofer, U., E. Bonatesta, C. Horejš-Kainrath, and S. Hochreiter. 2015. msa: an R package for multiple sequence alignment. Bioinformatics 31:3997–3999.

Burton, R. S. 2022. The role of mitonuclear incompatibilities in allopatric speciation. Cell. Mol. Life Sci. 79:103.

Coyne, J. A., and H. A. Orr. 1989. PATTERNS OF SPECIATION IN DROSOPHILA. Evolution 43:362–381.

de Zwaan, D. R., J. Mackenzie, E. Mikkelsen, C. Wood, and S. Wang. 2022. Pleiotropic opposing dominance within a color gene block contributes to a nascent species boundary via its influence on hybrid male territorial behavior. PNAS Nexus 1:pgac074.

Dobzhansky, T. 1934. Studies on hybrid sterility. Z.Zellforsch 21:169–223.

Eddo, D., Z. Hodur, and S. Wang. 2025. The Hidden Figures at Species Boundaries: The Mitochondrial Energetics Behind Mating Signal Divergence. Integrative And Comparative Biology icaf058.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792–1797.

Ehrenreich, I. M., and D. W. Pfennig. 2016. Genetic assimilation: a review of its potential proximate causes and evolutionary consequences. Ann Bot 117:769–779.

Feder, J. L., P. Nosil, A. C. Wacholder, S. P. Egan, S. H. Berlocher, and S. M. Flaxman. 2014. Genome-Wide Congealing and Rapid Transitions across the Speciation Continuum during Speciation with Gene Flow. Journal of Heredity 105:810–820.

Felsenstein, J. 1981. Skepticism Towards Santa Rosalia, or Why are There so Few Kinds of Animals? Evolution 35:124.

Fisher, R. 1930. The Genetical Theory of Natural Selection. The Clarendon Press.

G. Grothendieck, R Core Team (nls). 2007. nls2: Non-Linear Regression with Brute Force.

Gavrilets, S. 2004. Fitness landscapes and the origin of species. Princeton University Press, Princeton, N.J.

Gourbiere, S., and J. Mallet. 2010. Are species real? The shape of the species boundary with exponential failure, reinforcement, and the “missing snowball.” 64:1–24.

Hill, G. E. 2019. Mitonuclear Ecology. 1st ed. Oxford University Press.

Hill, G. E. 2017. The mitonuclear compatibility species concept. The Auk 134:393–409.

Lachlan, R. F., and M. R. Servedio. 2004. SONG LEARNING ACCELERATES ALLOPATRIC SPECIATION. Evolution 58:2049–2063.

Lijtmaer, D. A., B. Mahler, and P. L. Tubaro. 2003. HYBRIDIZATION AND POSTZYGOTIC ISOLATION PATTERNS IN PIGEONS AND DOVES. Evolution 57:1411–1418.

Lynch, M., B. Koskella, and S. Schaack. 2006. Mutation Pressure and the Evolution of Organelle Genomic Architecture. Science 311:1727–1730.

Matute, D. R., I. A. Butler, D. A. Turissini, and J. A. Coyne. 2010. A Test of the Snowball Theory for the Rate of Evolution of Hybrid Incompatibilities. Science 329:1518–1521.

McEntee, J. P., J. G. Burleigh, and S. Singhal. 2020. Dispersal Predicts Hybrid Zone Widths across Animal Diversity: Implications for Species Borders under Incomplete Reproductive Isolation. The American Naturalist 196:9–28.

Moran, B. M., C. Y. Payne, D. L. Powell, E. N. K. Iverson, A. E. Donny, S. M. Banerjee, Q. K. Langdon, T. R. Gunn, R. A. Rodriguez-Soto, A. Madero, J. J. Baczenas, K. M. Kleczko, F. Liu, R. Matney, K. Singhal, R. D. Leib, O. Hernandez-Perez, R. Corbett-Detig, J. Frydman, C. Gifford, M. Schartl, J. C. Havird, and M. Schumer. 2024. A lethal mitonuclear incompatibility in complex I of natural hybrids. Nature 626:119–127.

Moyle, L. C., and T. Nakazato. 2010. Hybrid Incompatibility “Snowballs” Between *Solanum* Species. Science 329:1521–1523.

Mueller, S. A., J. Merondun, S. Lečić, and J. B. W. Wolf. 2025. Epigenetic variation in light of population genetic practice. Nat Commun 16:1028.

Muller, H. J. 1942. Isolating mechanisms, evolution, and temperature. Biology Symposium 6:71–125.

Nieto Feliner, G., D. Criado Ruiz, I. Álvarez, and I. Villa-Machío. 2023. The puzzle of plant hybridisation: a high propensity to hybridise but few hybrid zones reported. Heredity 131:307–315.

Nosil, P. 2008. Speciation with gene flow could be common. Molecular Ecology 17:2103–2106.

448 Nosil, P., and D. Schluter. 2011. The genes underlying the process of speciation. Trends in Ecology &
449 Evolution 26:160–167.
450 Orr, H. A. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities.
451 Genetics 139:1805–1813.
452 Palopoli, M., and C. I. Wu. 1994. Genetics of hybrid male sterility between drosophila sibling species: a
453 complex web of epistasis is revealed in interspecific studies. Genetics 138:329–41.
454 Paradis, E., and K. Schliep. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary
455 analyses in R. Bioinformatics 35:526–528.
456 Pigliucci, M., and C. J. Murren. 2003. PERSPECTIVE: GENETIC ASSIMILATION AND A POSSIBLE
457 EVOLUTIONARY PARADOX: CAN MACROEVOLUTION SOMETIMES BE SO FAST AS TO PASS US BY?
458 Evolution 57:1455–1464.
459 Popadin, K. Y., S. I. Nikolaev, T. Junier, M. Baranova, and S. E. Antonarakis. 2013. Purifying Selection in
460 Mammalian Mitochondrial Protein-Coding Genes Is Highly Effective and Congruent with
461 Evolution of Nuclear Genes. Molecular Biology and Evolution 30:347–355.
462 Presgraves, D. C. 2002. PATTERNS OF POSTZYGOTIC ISOLATION IN LEPIDOPTERA. Evolution 56:1168–
463 1183.
464 Princepe, D., and M. A. M. De Aguiar. 2024. Nuclear compensatory evolution driven by mito-nuclear
465 incompatibilities. Proc. Natl. Acad. Sci. U.S.A. 121:e2411672121.
466 R Core Team. 2023. R: A language and environment for statistical omputing. R Foundation for Statistical
467 Computing, Vienna, Austria.

Rolland, J., D. Silvestro, D. Schluter, A. Guisan, O. Broennimann, and N. Salamin. 2018. The impact of endothermy on the climatic niche evolution and the distribution of vertebrate diversity. Nat Ecol Evol 2:459–464.

Roux, C., C. Fraïsse, J. Romiguier, Y. Anciaux, N. Galtier, and N. Bierne. 2016. Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence. PLoS Biol 14:e2000234.

Seehausen, O., R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L. Peichel, G.-P. Saetre, C. Bank, Å. Brännström, A. Brelsford, C. S. Clarkson, F. Eroukhmanoff, J. L. Feder, M. C. Fischer, A. D. Foote, P. Franchini, C. D. Jiggins, F. C. Jones, A. K. Lindholm, K. Lucek, M. E. Maan, D. A. Marques, S. H. Martin, B. Matthews, J. I. Meier, M. Möst, M. W. Nachman, E. Nonaka, D. J. Rennison, J. Schwarzer, E. T. Watson, A. M. Westram, and A. Widmer. 2014. Genomics and the origin of species. Nat Rev Genet 15:176–192.

Servedio, M. R., G. S. V. Doorn, M. Kopp, A. M. Frame, and P. Nosil. 2011. Magic traits in speciation: 'magic' but not rare? Trends in Ecology & Evolution 26:389–397.

Slabbekoorn, H., and T. B. Smith. 2002. Bird song, ecology and speciation. Phil. Trans. R. Soc. Lond. B 357:493–503.

Slatkin, M., and T. Maruyama. 1975. GENETIC DRIFT IN A CLINE. Genetics 81:209–222.

Stelkens, R. B., K. A. Young, and O. Seehausen. 2010. THE ACCUMULATION OF REPRODUCTIVE INCOMPATIBILITIES IN AFRICAN CICHLID FISH. Evolution 64:617–633.

Waddington, C. H. 1961. Genetic Assimilation. Pp. 257–293 *in* Advances in Genetics. Elsevier.

Wang, S., J. E. Mank, D. Ortiz-Barrientos, and L. H. Rieseberg. 2025. Genome Architecture and Speciation in Plants and Animals. Molecular Ecology e70004.

Wang, S., M. J. Ore, E. K. Mikkelsen, J. Lee-Yaw, D. P. L. Toews, S. Rohwer, and D. Irwin. 2021. Signatures of mitonuclear coevolution in a warbler species complex. Nat Commun 12:4279.

Wang, S., S. Rohwer, D. R. De Zwaan, D. P. L. Toews, I. J. Lovette, J. Mackenzie, and D. Irwin. 2020. Selection on a small genomic region underpins differentiation in multiple color traits between two warbler species. Evolution Letters 4:502–515.

Wang, S., and A. D. Yoder. 2025. Synergies between speciation and conservation science yield novel insights for mitigating the biodiversity crisis of the Anthropocene. Proc. Natl. Acad. Sci. U.S.A. 122:e2500713122.

Welch, J. J. 2004. ACCUMULATING DOBZHANSKY-MULLER INCOMPATIBILITIES: RECONCILING THEORY AND DATA. Evolution 58:1145–1156.

Westram, A. M., S. Stankowski, P. Surendranadh, and N. Barton. 2022. What is reproductive isolation? Journal of Evolutionary Biology 35:1143–1164.

Wu, C.-I. 2001. The genic view of the process of speciation. Journal of Evolutionary Biology 14:851–865.

Acknowledgements

We thank Loren Rieseberg, Dahong Chen and members of the Silu Wang's Lab for helpful discussions. This project is supported by the startup funding from SUNY Buffalo to Silu Wang.

Data Accessibility

All the data involved in the analysis will be submitted to Dryad upon manuscript acceptance. https://github.com/setophaga/hybridzone.snowball/tree/7dcd583defec2227a624b7c9a30b2d1806e16331/Final.data.code

Author Contributions

SW conceived the idea, and BR and DE conducted data collection under the supervision of SW. SW analyzed the data with assistance from DE and BR. All authors contributed for literature data extraction and cleaning. Specifically, BR curated the GPS coordinates of the hybrid zones and DE conducted the GenBank sequence extraction and alignments. SW conducted speciation modeling while teaching DE and BR.

Supplement

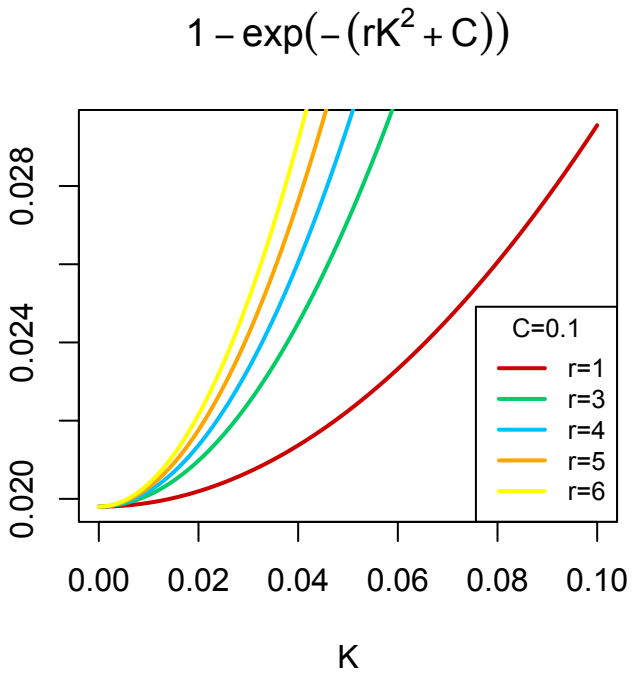

Fig. S1. **Modified Snowball Effect**. Orr (1995) derived the cumulative effect of *K* substitutions with average effect size *r* on reduction of hybrid fitness, or the "strength of reproductive isolation" as $1 - e^{-rK^2/2}$ (see Eq. 8 in Orr 1995). To incorporate epigenetic effects and be directly comparable to Additive and Slowdown effects (Eq. 1.1 and 1.3), we modified the original form and examined its form with different *r* values at moderate *C*.