Mitonuclear divergence predicts gradual speciation in animal hybrid zones

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

*Corresponding author: siluwang.evo@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 Linear 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 116 animal hybrid zones worldwide. The Slowdown Effect Model 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. There is greater mtDNA distance and greater selection on mtDNA than the nDNA. Together, the refined Linear Model was most supported by the data, revealing a significant mitochondrial effect as well as a marginal and heterogeneous nuclear effect. 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 converge 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 by which lineage-specific substitutions predispose 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 suppressed 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 of speciation rates.

The heterogeneous patterns of speciation rate are likely attributed to variability of locus-specific effect sizes and probability of incompatibility in different parts of the genome. For example, the effect sizes of mitochondrial substitutions can be different from those in the nuclear genome. Moreover, the probability of incompatibility can be different in intergenomic versus intragenomic discordance.

More broadly, the heterogeneous patterns of selection against hybrids could also be explained by the speciation stages 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 overcome extensive gene flow (Felsenstein 1981; Wang et al. 2020; de Zwaan et al. 2022), which predicts a linear buildup of selection against hybrids. As the gene flow between diverging lineages attenuates, allowing DMIs to wire up, a Snowball Effect can be observed (Orr 1995). When species are in the late stages of divergence, selection against hybrids might appear as diminishing returns in response to further genetic divergence, reflecting a logarithmic Slowdown Effect (Gourbiere and Mallet 2010). 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.

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. 1A)(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 predisposes the speciation rate (Fig.1 B).

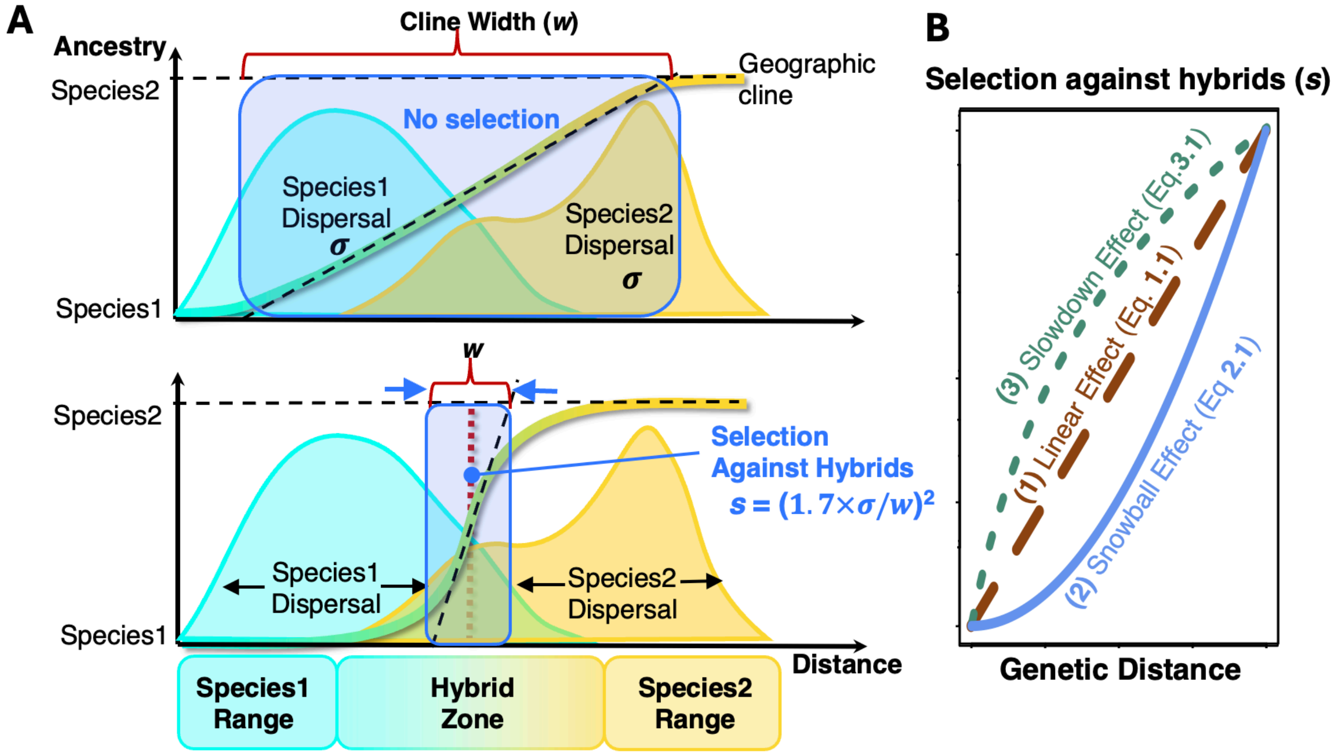

Fig.1 Hybrid zones are natural laboratories for understanding speciation. A, The cline width (*w*) of a hybrid zone is a balance of gene flow (σ) and selection (*s*) against hybrids. In the absence of s, σ will widen *w* over time as species 1 hybridizes with species 2 (top). Bottom: strong selection against hybrids counteracts with σ and shrinks w.

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. 1B). Model 2 predicts an exponential growth of *s*, as predicted by the Snowball Effect (Fig. 1B). Model 3 predicts a Slowdown Effect of genetic distance on selection against hybrids in the form of a logarithmic relationship (Fig. 1B).

We further developed three refined models to dissect mitochondrial and nuclear effects on speciation. Since the mitochondrial and nuclear genomes have distinct substitution rates (Ballard and Whitlock 2003) and coevolve to maintain critical organismal energetics, 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.

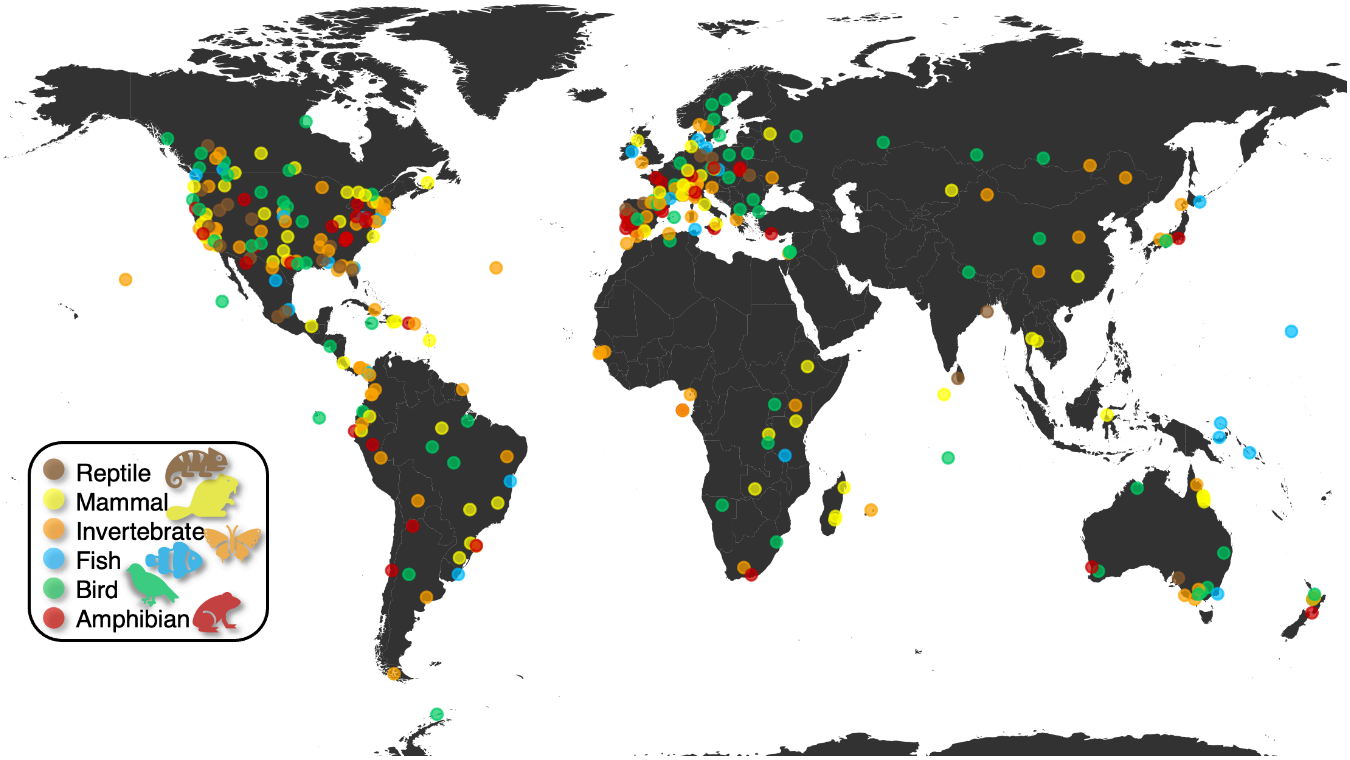

Fig. 2 World distribution of animal hybrid zones. Over 330 animal hybrid zones have been reported. Color categories correspond to six different organismal groups: mammals in yellow (65), birds in green (71), reptiles in brown (23), amphibians in red (45), invertebrates in orange (100), and fish in blue (26).

Method

Hybrid zones

To our knowledge, 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. 2). 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; (3) there is sufficient information on dispersal distance in the unit of KM/generation or KM/generation$^{0.5}$; (4) for the dispersal distance reported in KM/generation, there is information on generation time to calculate the KM/generation$^{0.5}$; (5) there are sufficient genetic sequencing data ($N > 3$) of each pure parental species. These criteria resulted in 116 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, ND4, and CYTB for mitochondrial genes and RAG1, RAG2, and MC1R 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 (Bodenhofer et al. 2015) function 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 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 mean pairwise mtDNA distance and mean pairwise nDNA distance between each pair of parental species. Finally, we calculated the mean pairwise genetic distance between each pair of parental species by averaging the nDNA and mtDNA distance.

Selection against hybrids

We extracted the dispersal, center, and width of hybrid zones from McEntee et al. (2020). We unified the dispersal (σ) unit as kilometers (KM) /$\sqrt{generation}$ with the average generation time information of each species pair. We extracted the average cline width (w) and center for mtDNA and nDNA data, as well as phenotypic data, separately. We calculated selection (s) as a function of cline width (w) and dispersal (σ): $s = (1.7 \times \sigma/w)^2$ (Fig. 1). To infer the cumulative selection against hybrids ($\bar{s}$), substituted w with averaged cline width ($\bar{w}$) among mtDNA, nDNA (Fig. 3B), and phenotypic clines.

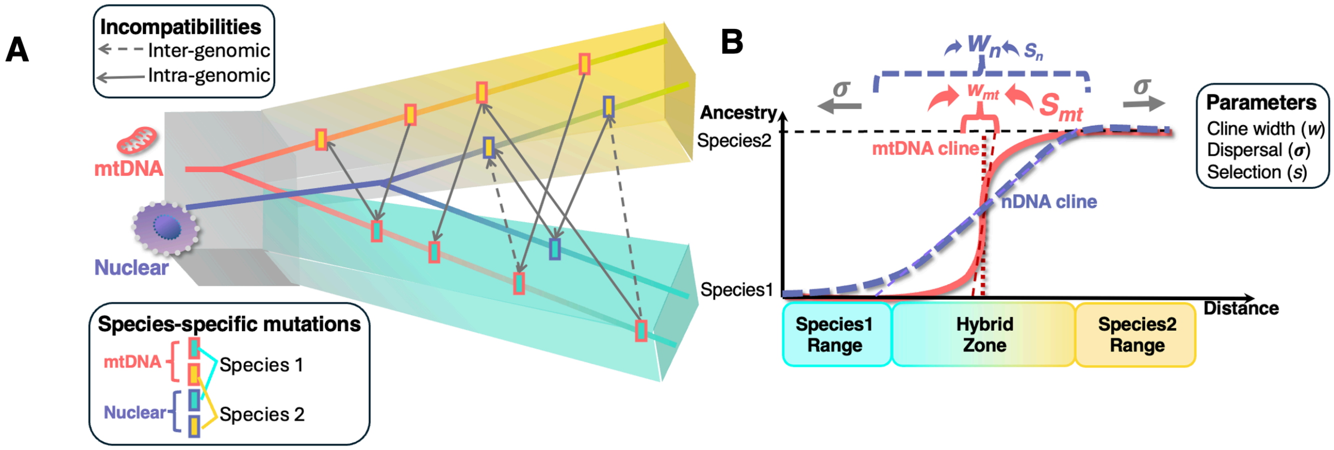

Fig. 3 Mitonuclear genetic divergence underlying the selection against hybrids. A. Lineage-specific substitutions in the mitochondrial genome (red) and nuclear genome (purple) can lead to Dobzhansky-Muller Incompatibilities, which emergently contribute to selection against hybrids. **B.** This selection counteracts with 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.

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), 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*. The rate parameter corresponds to the average effect size of each pair of incompatibility.

$$s = r \times K + C \qquad \text{(Eq. 1.1)}$$

$$s = C - e^{\frac{-r \times K^2}{2}} \qquad \text{(Eq. 2.1)}$$

$$s = Ln(r \times K + C) \qquad \text{(Eq. 3.1)}$$

To account for the independent and interactive effects of mtDNA and nuclear genetic divergence (Fig. 3), we further refined the conventional expressions (equation 1.1, 2.1, and 3.1) into mitonuclear models (equation 1.2, 2.2, and 3.2) to partition *K* into K_m (mtDNA distance) and K_n (nuclear genetic distance) that correspond to rate parameters r_m and r_n, respectively. In addition, their interactive effect has a rate parameter of r_{mn}.

$$s = r_m K_m + r_n K_n + r_{mn} K_m K_n \quad (1.2)$$

$$s = 1 - e^{\frac{-(r_m K_m + r_n K_n + r_{mn} K_m K_n)^2}{2}} \qquad (2.2)$$

$$s = Ln(r_m K_m + r_n K_n + r_{mn} K_m K_n) \qquad (3.2)$$

We conducted model fitting of the above three models with the nls2 function (Grothendieck, 2007) in R (R Core Team 2023) and computed the AIC of each fitted model with the AIC function. In each conventional model fitting (1.1, 2.1, 3.1), we estimated the rate parameter r and constant C. We estimated r_m, r_n, and r_{mn} in the refined model fitting (1.2, 2.2, 3.2). To understand the confidence level of each parameter fitting, we generated 1,000 samples (with replacement) of the species pairs and conducted model fitting of each bootstrap sample. To examine taxon-specific patterns, we separately conducted model fitting within the following organismal groups.

Results

Discordant mitonuclear selection

There is significantly greater genetic distance in mtDNA than nDNA between parental species (mean difference = 0.039, 95% CI: 0.030-0.048, $p < 10^{-11}$) (Fig. 4 A). Concordantly, there is greater selection on mitochondrial clines than nuclear clines in hybrid zones (mean difference = 0.073, 95% CI: 0.004-0.141, $p < 0.05$) (Fig. 3, 4 B), though the cline widths of mtDNA and nDNA are not significantly different ($p > 0.05$). The discrepancy of selection on mitochondrial and nuclear genetic clines is significantly lower when there is greater genetic distance between parental species ($R^2 = 0.08$, $p = 0.04$) (Fig. 4 C).

Genetic distance and selection against hybrids

The Slowdown Effect model (3.1) (Fig. 5 A) has the highest bootstrap support (59.9%), followed by the Snowball model (30.2%), and then the linear model (9.9%) (Table S1). The bootstrap mean of rate parameter (r) of the Slowdown Effect model (equation 3.1) is 0.54 (Fig. 5 B), which reflects mild positive effects of genetic divergence on selection against hybrids.

The bootstrap estimates of r are heterogeneous among taxonomic groups (Fig. 5 B), with reptiles ($r = 1.84$) and invertebrates ($r = 2.25$) being the highest, birds ($r = 0.75$) being intermediate, and lower for mammals ($r = 0.16$) and amphibians ($r = -0.90$). The Snowball Effect model is selected for taxon-specific modeling. In particular, the Snowball Effect has the highest chance of being the lowest AIC in birds (43.5%), invertebrates (46.3%), reptiles (91.8%), mammals (54.8%), and amphibians (65.6%).

When considering the effects of mitochondrial and nuclear genetic distance independently and interactively (Fig. 3), the Linear model (equation 1.2) has the lowest AIC in 86.7% of the bootstrap samples. Only 13.3% of the bootstrap samples supported the Snowball Effect model (equation 2.2), and none of the bootstrap samples selected the Slowdown Effect model (equation 3.2). Both the Linear model and the Slowdown model concluded significantly positive r_m, suggesting a positive effect of mtDNA genetic divergence on selection against hybrids (Table 1, Fig. 5C and D). Only the Slowdown effect model revealed significantly positive r_n, reflecting the positive independent effect of nDNA divergence. Both the Linear and Slowdown effect models disclosed negative estimates of r_{mn}, which was significantly negative in the Slowdown model (Table 1), indicating a dampening effect of mitonuclear interaction on the selection against hybrids.

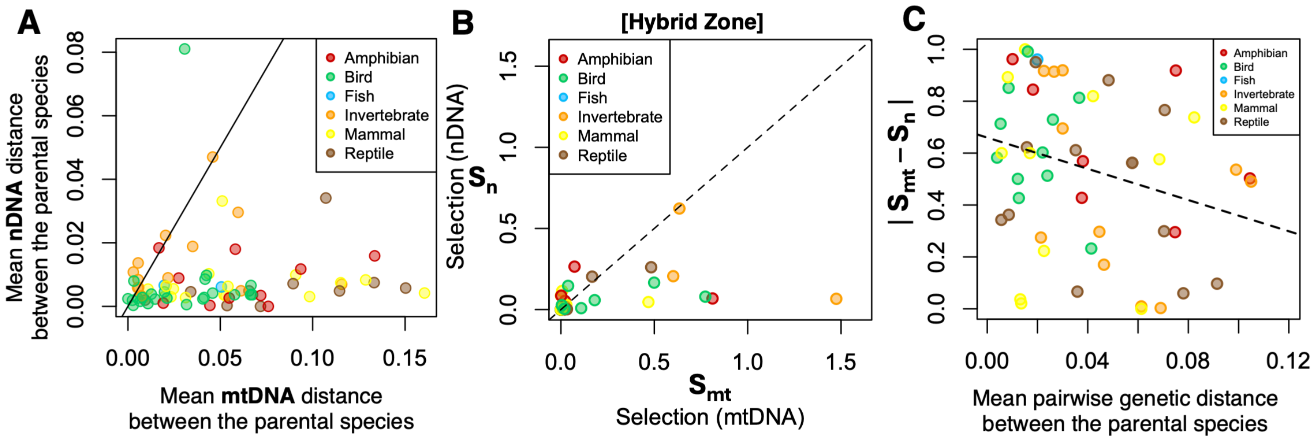

Fig. 4 Mitonuclear genetic divergence and estimated selection against hybrids. A. Most hybridizing animal species pairs demonstrate greater pairwise genetic distance in mitochondrial genes than nuclear genes ($p < 0.05$). The black line is the 1:1 line. **B.** There is greater selection on mtDNA clines than nDNA clines across animal hybrid zones ($p < 0.05$). The black dotted line is the 1:1 line. **C.** The discrepancy in selections on mtDNA versus nDNA clines decreases as the genetic distance between parental species increases. The black dotted line represents the regression line ($p < 0.05$).

Different patterns were observed among taxonomic groups. In particular, the Linear model (equation 1.2) was best fitted to birds (bootstrap probability = 68.4%) and amphibians (bootstrap probability = 75.2%) data, while the Snowball Effect (equation 2.2) was the best fitted model for mammals (bootstrap probability = 80.9%), invertebrates (bootstrap probability = 64.7%), and reptiles (bootstrap probability = 64.6%).

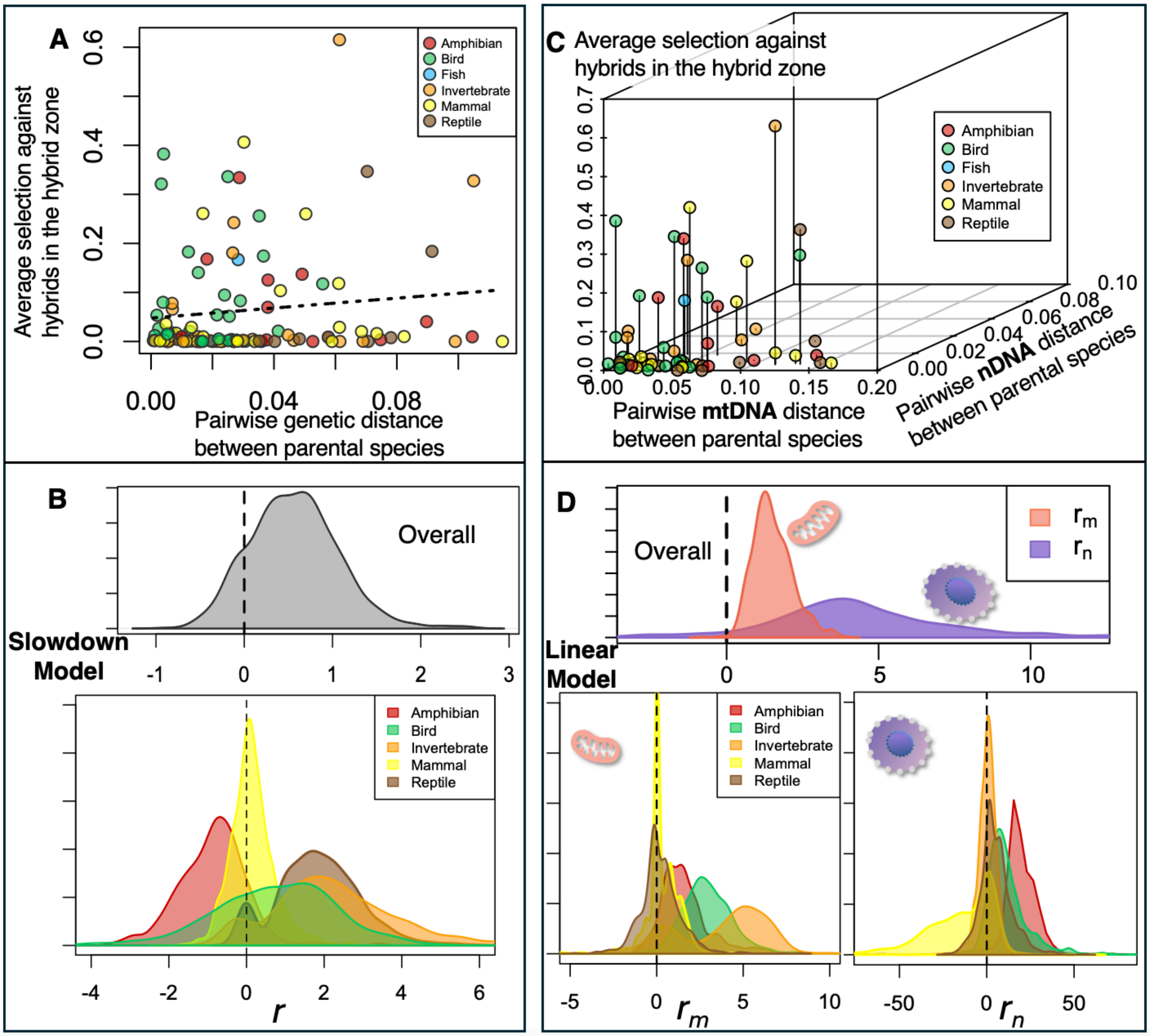

Fig. 5 Coefficients of estimates in the best-fitted models for simplified versus refined models of speciation. A, Conventional models (equation 1.1, 2.1, and 3.1) considered the total genetic distance as the predictor for selection against hybrids. The slowdown model (dotted line) has the lowest AIC. **B**, Bootstrap distributions of the rate parameter ***r*** (equation 1.3) estimations based on all the hybrid zones together, or individual taxonomic groups separately. **C**, Refined models considered the effect of mtDNA and nDNA divergence and their interactions separately. The linear 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. Variations in bootstrap estimates of r_m and r_n are observed among taxonomic groups. In particular, greater r_m estimates occurred in invertebrates and birds, while amphibians showed greater r_n.

Table 1 The effect of mitochondrial and nuclear genetic divergence on the selection against hybrids in animal hybrid zones. 95% confidence intervals (CIs) of the coefficients of estimations for the rate parameters of mtDNA (r_m), nDNA (r_n), and mitonuclear interaction (r_{mn}) effects. The bold 95% CIs correspond to significant effects.

Model	r_m	r_m 95%CI (2.5%, 97.5%)	rn	r_n 95%CI (2.5%, 97.5%)	r_{mn}	r_{mn} 95%CI (2.5%, 97.5%)	AIC
Linear (1.2)	**1.50**	0.50 2.96	3.86	-2.26 12.15	-84.32	-220.24 73.81	**-75.43**
Snowball (2.2)	16.86	-18.12 78.81	23.45	-1129.71 1044.32	18.00	-460.10 833.32	-65.61
Slowdown (3.2)	**17.71**	13.53 32.82	**135.33**	71.88 229.28	-**1642.80**	-4170.37 -1172.54	120.62

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. Different from the expectation of the Snowball Effect, we observed a gradual increase of selection against hybrids over genetic divergence between the parental species. Mitochondrial genetic divergence tends to be more pronounced than nuclear genetic divergence and significantly contributes to speciation. The effect of nuclear genetic divergence is also positive, though more heterogeneous (Fig. 5 C and D).

The Slowdown model (equation 3.1) best fitted the full data among the three conventional models (equation 1.1, 2.1, 3.1) (Fig. 5 A and B). This pattern is opposite 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 behavioral/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 predispose the likelihood of cultural evolution as premating isolation, though it can also disrupt reproductive isolation. Indeed, we observed the heterogeneity of rate parameter estimations among taxonomic groups (Fig. 5 B). The lower rate parameter estimations in mammals and amphibians could indicate a relatively mutated genetic effect on reproductive isolation at the early stage of speciation in these groups.

When we consider the independent and interactive effects of mtDNA and nDNA divergence (equation 1.2, 2.2, 3.2), the full dataset fitted the linear model (equation 1.2) best with around 87% bootstrap support. In this model, mtDNA distance had a significant effect, while nDNA distance showed heterogeneous effect sizes (Fig. 5C and D). The significant mtDNA effect is consistent with the relatively greater mtDNA distance between parental lineages (Fig. 4 A) and the stronger selection on mtDNA clines in the hybrid zones (Fig. 3, 4B). The greater mutation rates of mtDNA than nDNA in animal cells (Ballard and Whitlock 2003; Lynch et al. 2006) may predispose the chance of incompatibility resulting from mitochondrial substitutions. 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 might dampen this contribution.

The interaction of mtDNA and nDNA distance did not show a significant effect on selection against hybrids (Table 1), as the rate parameter (r_{mn}) estimates are highly heterogeneous. This pattern may appear contradictory to the growing evidence that mitonuclear incompatibility is important for early-stage speciation (Hill 2017; Wang et al. 2021; Moran et al. 2024). However, the interaction term (r_{mn}) between mtDNA and nDNA reflects mitonuclear discordance, which may not necessarily cause mitonuclear incompatibility. In particular, among the mitonuclear discordant substitutions, the synonymous mitochondrial or nuclear substitution that does not disrupt mitonuclear co-function would not lead to mitonuclear incompatibility. The large variation in the bootstrap estimates of r_{mn} (Fig. 5D) is consistent with the heterogeneous effect of mitonuclear discordance, which calls for future investigations. As the genetic divergence increases, discordance in selections on mtDNA clines and nDNA clines decreases (Fig. 4C), which indicates genome-wide mitonuclear coevolutionary resolution as the parental lineages diverge.

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 parent lineages diverged. Further dissection of mitochondrial and nuclear genetic divergence revealed relatively greater divergence and selection in the mitochondrial genome than the nuclear genome, which significantly contributed to the cumulative selection against hybrids across species boundaries. This observation would shed light on the evolutionary mechanisms of speciation rate heterogeneity 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 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.

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. Zellforsch 21:169–223.

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

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.

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. 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.

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.

Nosil, P., and D. Schluter. 2011. The genes underlying the process of speciation. Trends in Ecology & Evolution 26:160–167.

Orr, H. A. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics 139:1805–1813.

Palopoli, M., and C. I. Wu. 1994. Genetics of hybrid male sterility between drosophila sibling species: a complex web of epistasis is revealed in interspecific studies. Genetics 138:329–41.

Paradis, E., and K. Schliep. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526–528.

Popadin, K. Y., S. I. Nikolaev, T. Junier, M. Baranova, and S. E. Antonarakis. 2013. Purifying Selection in Mammalian Mitochondrial Protein-Coding Genes Is Highly Effective and

Congruent with Evolution of Nuclear Genes. Molecular Biology and Evolution 30:347–355.

Presgraves, D. C. 2002. PATTERNS OF POSTZYGOTIC ISOLATION IN LEPIDOPTERA. Evolution 56:1168–1183.

R Core Team. 2023. R: A language and environment for statistical omputing. R Foundation for Statistical Computing, Vienna, Austria.

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.

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., 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.

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.

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 curation under the supervision of SW. SW analyzed the data with assistance from DE and BR. All authors contributed to the manuscript.

Supplement

Table S1 The effect of genetic divergence on the selection against hybrids. 95% CIs of the coefficients of estimates for the rate parameters (*r*) and constant. The Slowdown model (equation 3.1) has the lowest AIC.

Model	***r***	***r* 95%CI (2.5%, 97.5%)**	***C***	***C* 95%CI (2.5%, 97.5%)**	**AIC**
Linear (1.1)	0.50	-0.31 1.44	**0.05**	0.02 0.08	-164.06
Snowball (2.1)	7.98	-8.05 32.02	**1.06**	1.03 1.08	-163.41
Slowdown (3.1)	0.54	-0.33 1.61	**1.05**	1.02 1.08	**-164.07**