Supplemental Information for

Dispersal predicts hybrid zone widths across animal diversity

SUPPLEMENTARY TABLES

Table S1: Key metadata collected for each cline included in this study.

Metadata	Details	Notes
type	The type of phenotypic or molecular trait for which the cline was estimated	Summarized as: morphological traits, traits involved in mating (e.g. pheromones or bird song), mtDNA markers, nuclear DNA markers, karyotypic markers, or sex-linked DNA markers (see Fig. S3C)
Number of markers	For molecular clines, the number of loci used to infer the cline	Used for hybrid indices based on multiple molecular markers
Biallelic or quantitative	Whether the cline-fitting analysis treats characters as biallelic or quantitative	Molecular cline estimates based on more than one locus were coded quantitative; all morphological data were coded quantitative
Width	Point estimate for cline width	
Center	Point estimate for cline center	
Difference between tails	The difference in trait values at the extreme tails of each hybridizing species	For biallelic traits, this reflected differences in allele frequency; for quantitative traits, this reflected standardized differences in phenotypes
Program used	The program used to estimate cline parameters	

Table S2: Key metadata collected for each hybridizing pair included in this study.

Metadata	Details	Notes
taxa	The scientific names of the hybridizing taxa	Some hybrid zones occur between taxa not currently recognized as species; in these cases, we followed the naming convention used by the study's authors
Moving or not?	Whether or not authors provided or referred to evidence that the hybrid zone is moving in position over time	Many studies did not provide this information
anthropogenic?	Whether or not the hybrid zone arose due to anthropogenic disturbance and / or is being affected by anthropogenic effects	Many studies did not provide this information
Genetic distance	Estimates of genetic distance (reported in units of substitution) between the taxa	We calculated most estimates using GenBank data. Other estimates were taken from the literature. For all estimates, we noted the marker and model of molecular evolution used.
Dispersal estimate	Estimates of dispersal, and their units, for the taxa or closely-related taxa	
Dispersal type	The methodology used to estimate dispersal	Many ecological and genetic approaches were used to estimate dispersal, including mark-recapture studies and approaches based on linkage disequilibrium and patterns of isolation-by-distance
Generation time	Estimated generation time for the taxa	
Taxonomic group	Broadly categorized into: amphibian, bird, fish, insect, mammal, non-avian reptile (NAR), and other invertebrates	
Geographic location	Broadly categorized into: Africa, Asia, Australia / New Zealand, Central America, Eurasia, Europe, Middle East, North America, South America	
citation	All papers used to summarize evidence from the hybrid zone and the hybridizing pair	

Table S3: Hybrid zones for which geographic clines were estimated in the literature but were not included in our global analysis because of missing data. Several iconic hybrid zones are neither listed here nor included in our final dataset because geographic clines were not measured. Some bird hybrid zones where dispersal data were lacking were included in bird-only analyses where a morphological proxy (hand-wing index) for dispersal was used as a predictor of cline width instead.

Taxon 1	Taxon 2	reason excluded	representative citation
Albinaria hippolyti, lineage 1	*Albinaria hippolyti*, lineage 2	no mtDNA data	Schilthuizen and Lombaerts (1995) *Biol J of Linn Soc* 54: 111 - 138.
Allonemobius fasciatus	*Allonemobius socius*	cline widths not reported as point estimates	Howard (1986) *Evolution* 40: 34 - 43.
Aponomma hydrosauri	*Aponomma libatum*	cline widths not reported as point estimates	Bull and Burzacott (2001) *Mol Ecol* 10: 639 - 49.
Barytettix humphreysii humphreysii	*Barytettix hymphreysii cochisei*	cline data not given in terms of geographic distance	Knowles *et al*. (2016) *J Orthoptera Research* 25: 75 - 82.
Caledia captiva "Moreton" lineage	*Caledia captiva* "Torresian" lineage	no mtDNA data	Shaw and Wilkinson (1980) *Chromosoma* 80: 1 - 31.
Chorthippus albomarginatus	*Chorthippus oschei*	cline data not formalized	Vedenina (2011) *Biol J of Linn Soc* 102: 275 - 291.
Cottus gobio E lineage	*Cottus gobio* W lineage	cline widths not reported as point estimates	Kontula and Vainola (2004) *Biol J of Linn Soc* 81: 535-552.
Eopsaltria australis North	*Eopsaltria australis* South	mtDNA data cannot be assigned to taxa	Morales *et al*. (2017) *J Biogeography* 44: 522 - 536.
Gasterosteus aculeatus, marine population	*Gasterosteus aculeatus*, freshwater population	no mtDNA data	Pederson *et al*. (2017) *BMC Evol Bio* 17: 130.
Gasterosteus aculeatus, stream population	*Gasterosteus aculeatus*, anadramous population	no mtDNA data	Vines *et al*. (2016) *Evolution* 70: 1023 - 1038.
Geomys breviceps	*Geomys bursarius*	cline widths not reported as point estimates	Cothran and Zimmerman (1985) *J Mammology* 66: 489 - 497.
Jacana spinosa	*Jacana jacana*	no dispersal data	Miller *et al.* (2014) *BMC Evol Bio* 14: 227.
Manacus vitellinus	*Manacus candei*	no dispersal data	Brumfield *et al.* (2001) *Evolution* 55: 2070 - 2087.
Nectarinia moreaui	*Nectarinia fuelleborni*	no dispersal data	McEntee et al. (2016) Evolution 70: 1307 - 1321.
Paratya australiensis, Kilcoy Ck	*Paratya australiensis*, Branch Ck	cline widths not reported	Wilson, Schmidt and Hughes (2016) *J Heredity* 107: 413 - 422.
Patella ulyssiponensis	*Patella rustica*	cline center reported only	Sa-Pinto *et al* (2012) *PLOS One* 7: e50330.
Plethodon jordani	*Plethodon glutinosus*	cline data not formalized	Hairston *et al*. (1992) *Evolution* 46:930 - 938.
Poephila acuticauda, yellow-beaked	*Poephila acuticauda*, red-beaked	mtDNA data from GenBank cannot be assigned to taxa	Griffith and Hooper (2016) *Emu* 2: 141 - 150.
Quiscalus quiscula versicolor	*Quiscalus quiscula quiscula*	no formal cline analysis	Yang and Selander (1968) *Systematic Biology* 17: 107 - 143.
Ranitomeya imitator banded	*Ranitomeya imitator* striped	no mtDNA data	Twomey *et al.* (2015) *Am Nat* 187: 205 - 224.
Ranitomeya imitator spotted	*Ranitomeya imitator* striped	no mtDNA data	Twomey *et al.* (2015) *Am Nat* 187: 205 - 224.
Ranitomeya imitator striped	*Ranitomeya imitator* vardero	no mtDNA data	Twomey *et al.* (2015) *Am Nat* 187: 205 - 224.
Sorex araneus - Novosibirsk	*Sorex araneus* - Tomsk	no mtDNA data	Polyakov *et al*. (2011) *J Evol Bio* 24: 1393 - 1402.
Xiphophorus birchmanni	*Xiphophorus malinche*	cline widths not reported as point estimates	Culumber *et al*. (2011) *Mol Ecol* 20: 342 - 356.
Zosterops borbonicus, grey-headed	*Zosterops borbonicus*, brown-naped	no mtDNA data	Delahie *et al.* (2017) *J Evol Bio* 30: 2132 - 2145.
Zosterops pallidus	*Zosterops virens capensis*	no dispersal data	Oatley *et al.* (2017) *Biol J Linn Soc* 121: 670 - 684.

Table S4: Model fitting for data set filtered to include clines from terrestrial systems only (*N* = 1424 clines across 122 hybrid zones). Models predicting the log of cline width were fit using a linear mixed model. Predictors were the log of mtDNA distance and log of dispersal. Random effects were cline type (see Fig. S2C) and a series of nested random effects: transect, nested within taxon pair, nested within larger taxonomic group (see Fig. S1C). Shown are AICc and ΔAICc scores. R^2GLMM(*m*) shows the estimated proportion of variance explained by the fixed effects; R^2GLMM(*c*) for both fixed and random effects. The best model is shown in bold.

model	df	LnL	AICc	ΔAICc	R^2GLMM(*m*)	R^2GLMM(*c*)
~ log(mtDNA distance) × log(dispersal)	9	-2407.1	4832.3	2.3	0.157	0.723
~ log(mtDNA distance) + log(dispersal)	8	-2408.3	4832.7	2.7	0.165	0.728
~ log(mtDNA distance)	7	-2418.6	4851.4	21.3	0.007	0.759
~ log(dispersal)	**7**	**-2408**	**4830**	**-**	**0.165**	**0.73**
~ 1	6	-2418.5	4849	19	0	0.775

Table S5: Model averaging results for terrestrial models (shown in Table S4), including coefficients and relative importance.

predictor	coefficient	standard error	relative importance
log(dispersal)	0.897	0.158	1
log(mtDNA distance)	-0.067	0.125	0.37
log(mtDNA distance) × log(dispersal)	0.057	0.126	0.204

Table S6: Model fitting for full data set (*N* = 1,488 clines across 131 hybrid zones) in which taxonomic group was included as a fixed effect rather than a random effect. Models predicting the log of cline width were fit using a linear mixed model. Predictors were the log of mtDNA distance, log of dispersal, and taxonomic group. Random effects were cline type (see Fig. S2C) and transect nested within taxon pair. Shown are AICc and ΔAICc scores. R^2GLMM(*m*) shows the estimated proportion of variance explained by the fixed effects; R^2GLMM(*c*) for both fixed and random effects. The top five models are shown as determined by AICc scores; the best model is shown in bold.

model	df	LnL	AICc	ΔAICc	R^2GLMM(*m*)	R^2GLMM(*c*)
~ log(dispersal) × taxonomic group	**18**	**-2497**	**5031.4**	**-**	**0.231**	**0.75**
~ log(dispersal) × log(mtDNA distance) + log(dispersal) × taxonomic group	20	-2499	5038.7	7.266	0.233	0.753
~ log(dispersal)	6	-2517	5046.4	14.96	0.172	0.755
~ log(dispersal) + taxonomic group	12	-2511	5046.7	15.305	0.182	0.755
~ log(dispersal) + log(mtDNA distance)	7	-2517	5049.8	18.408	0.165	0.754

Table S7: Model averaging results for models in which phylogenetic group was included as a fixed effect (shown in Table S6), including coefficients and relative importance. Coefficients not reported for predictors including taxonomic group because these are calculated for each of the seven taxonomic groups.

predictor	coefficient	standard error	relative importance
log(dispersal)	0.227	0.499	1
log(mtDNA distance)	-0.001	0.026	0.026
taxonomic group	NA	NA	1
log(mtDNA distance) × log(dispersal)	0.003	0.03	0.026
taxonomic group × log(dispersal)	NA	NA	1

Table S8: Model fitting for data set filtered to include only mitochondrial DNA clines (N = 91 clines across 74 hybrid zones). Models predicting the log of cline width were fit using a linear mixed model. Predictors were the log of mtDNA distance and log of dispersal. The sole random effect was a series of nested random effects: transect, nested within taxon pair, nested within taxonomic group (see Fig. S1C). Shown are AICc and ΔAICc scores. R^2GLMM(m) shows the estimated proportion of variance explained by the fixed effects; R^2GLMM(c) for both fixed and random effects. The best model is shown in bold.

model	df	LnL	AICc	ΔAICc	R^2GLMM(m)	R^2GLMM(c)
~ log(mtDNA distance) × log(dispersal)	6	-177.9	368.8	5.6	0.34	0.367
~ log(mtDNA distance) + log(dispersal)	5	-178	366.8	3.5	0.331	0.356
~ log(mtDNA distance)	4	-190	388.5	25.2	0.001	0.197
~ log(dispersal)	**4**	**-177.4**	**363.3**	**-**	**0.325**	**0.355**
~ 1	3	-189.4	385	21.7	0	0.204

Table S9: Model averaging results for the data set that only includes mtDNA clines (models shown in Table S8), including coefficients and relative importance.

predictor	coefficient	standard error	relative importance
log(dispersal)	1.13	0.20	1
log(mtDNA distance)	-0.02	0.09	0.19
log(mtDNA distance) × log(dispersal)	0.01	0.07	0.05

Table S10: Model fitting for data set filtered to only include taxon pairs for which mtDNA distance was estimated using the most common gene, *cytochrome b* (*N* = 259 clines across 41 hybrid zones). Models predicting the log of cline width were fit using a linear mixed model. Predictors were the log of mtDNA distance and log of dispersal. Random effects were cline type (see Fig. S2C) and a series of nested random effects: transect, nested within taxon pair, nested within larger taxonomic group (see Fig. S1C). Shown are AICc and ΔAICc scores. R^2GLMM(*m*) shows the estimated proportion of variance explained by the fixed effects; R^2GLMM(*c*) for both fixed and random effects. The best model is shown in bold.

model	df	LnL	AICc	ΔAICc	R^2GLMM(*m*)	R^2GLMM(*c*)
~ log(mtDNA distance) × log(dispersal)	9	-452	922.7	3.8	0.08	0.643
~ log(mtDNA distance) + log(dispersal)	8	-452.2	921	2.1	0.08	0.641
~ log(mtDNA distance)	7	-453.6	921.6	2.8	0.036	0.646
~ log(dispersal)	**7**	**-452.2**	**918.9**	**-**	**0.069**	**0.632**
~ 1	6	-454	920.3	1.49	0	0.638

Table S11: Model averaging results for the dataset limited to those where *cytochrome b* was used to estimate mtDNA distance (shown in Table S10), including coefficients and relative importance.

predictor	coefficient	standard error	relative importance
log(dispersal)	0.49	0.45	0.67
log(mtDNA distance)	-0.08	0.19	0.33
log(mtDNA distance) × log(dispersal)	0.02	0.11	0.07

Table S12: Model fitting for data set filtered to only include avian hybridizing pairs (*N* = 37 hybrid zones). Models predicting the log of the geometric mean of cline width were fit using a linear model. Predictors were the log of mtDNA distance and hand-wing index (HWI), a proxy for dispersal. Shown are adjusted r^2, AICc, and relative likelihood. The best model is bolded.

formula	df	AICc	adj. r^2	rel. likelihood
~ log(mtDNA distance) × log(HWI)	4,33	128	0.061	0.192
~ log(mtDNA distance) + log(HWI)	3,34	125.4	0.089	0.705
~ log(HWI)	**2,35**	**124.7**	**0.068**	**1**
~ log(mtDNA distance)	2,35	127	0.01	0.317
~ 1	1,36	126	0	0.522

Table S13: Model averaging results for the dataset limited to avian hybridizing pairs, using hand-wing index (HWI) to measure dispersal (shown in Table S12), including coefficients and relative importance.

predictor	coefficient	standard error	relative importance
log(HWI)	0.21	0.20	0.69
log(mtDNA distance)	0.10	0.16	0.45
log(mtDNA distance) × log(HWI)	0.00	0.04	0.06

Table S14: Model fitting for data set filtered to only include avian hybridizing pairs (N = 16 hybrid zones; the outlier *Sula* boobies was removed). Models predicting the log of the geometric mean of cline width were fit using a linear model. Predictors were the log of phylogenetic divergence time and log of dispersal. Shown are adjusted r^2, AICc, and relative likelihood. The best model is bolded.

formula	df	AICc	adj. r^2	rel. likelihood
~ log(divergence time) × log(dispersal)	4,13	54.4	0.277	0.449
~ log(divergence time) + log(dispersal)	3,14	56.6	0.027	0.15
~ log(dispersal)	**2,15**	**53.1**	**0.091**	**0.861**
~ log(divergence time)	2,15	55.4	-0.044	0.273
~ 1	1,16	52.8	0	1

Table S15: Model averaging results for the dataset limited to avian hybridizing pairs, using phylogenetic divergence time as a proxy for selection against hybrids (shown in Table S14), including coefficients and relative importance.

predictor	coefficient	standard error	relative importance
log(dispersal)	0.25	0.34	0.54
log(divergence time)	-0.12	0.31	0.32
log(divergence time) × log(dispersal)	0.10	0.25	0.17

SUPPLEMENTARY FIGURES

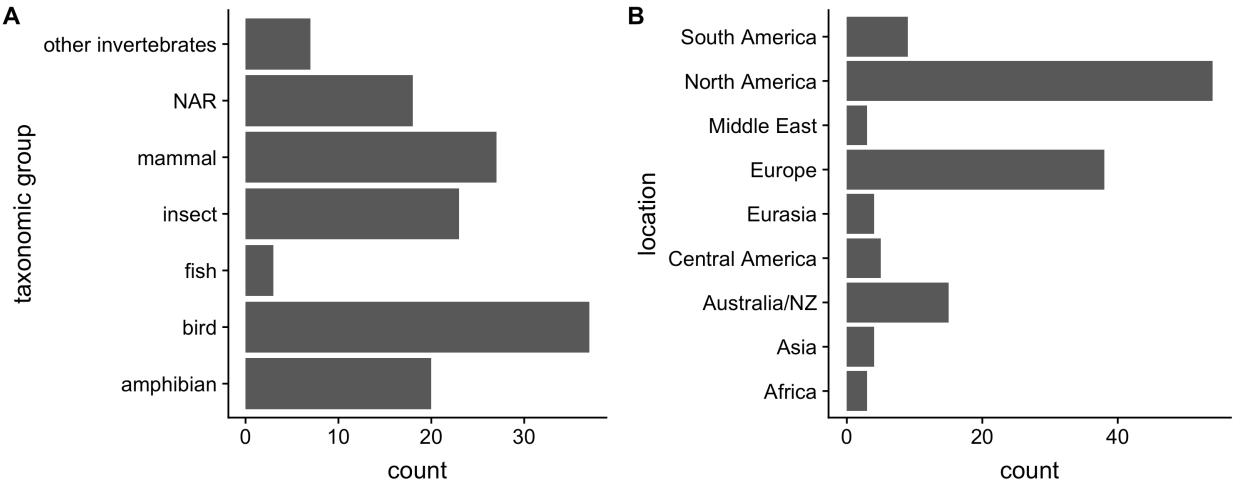

Figure S1: The (A) taxonomic and (B) geographic breadth of hybrid zones (*n*=135) included across all analyses. Hybrid zones are found in a diversity of taxa and regions, though there is a bias towards systems involving tetrapods and occurring in the temperate Northern Hemisphere.

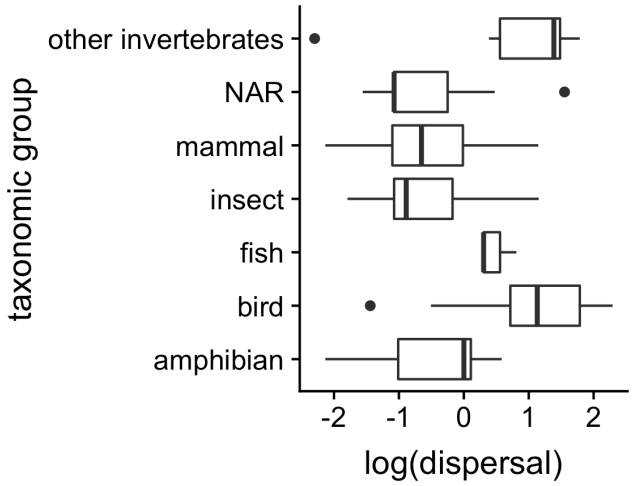

Figure S2: Dispersal length estimates across taxonomic groups (*n*=131). Dispersal varies more than four orders of magnitude. Taxonomic group is a significant predictor of variation in dispersal length ($F_{6, 124}$ = 17.1; p-val < 2.5e-14; adj. r^2 = 0.43)

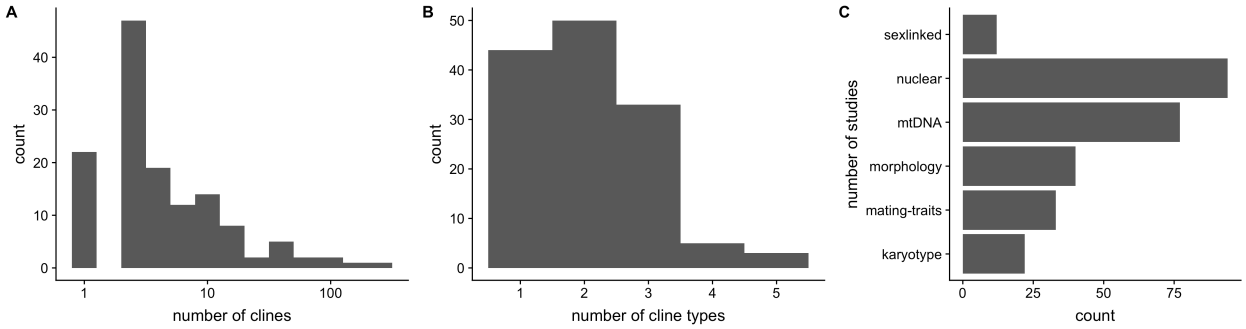

Figure S3: Clines collected across hybrid zones. (A) The number of clines estimated per hybrid zone. A median of three clines was estimated per hybrid zone. (B) The number of different cline types estimated per hybrid zone. An average of 2.1 types of clines were estimated per hybrid zone. (C) The number of hybrid zone studies that estimated each type of cline (sex-linked genetic markers, nuclear genetic markers, mtDNA genetic markers, morphological measures, mating-related traits, and karyotypes).

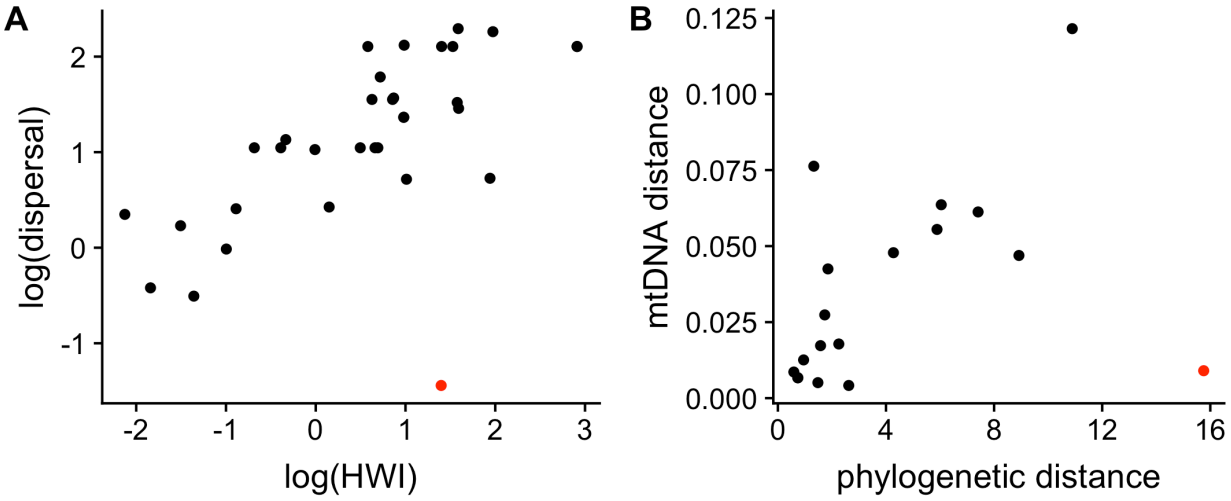

Figure S4: Correlations between dispersal and divergence estimates taken from the literature and alternative metrics, for bird hybrid zones only. (A) Correlation between our literature-based estimates of dispersal rate and a common morphological proxy for dispersal capacity, hand-wing index (HWI), with both on the natural log scale. Pearson's $r = 0.61$ ($p = 7.2 \times 10^{-8}$; $n = 32$). The taxon pair *Sula nebouxii & S. variegata* is an outlier (shown in red). Excluding this pair returns a Pearson's $r = 0.80$ ($p = 7.2 \times 10^{-8}$; $n = 31$). Log HWI also predicts log dispersal when phylogeny is taken into account using phylogenetic generalized linear modeling, with Ornstein-Uhlenbeck (OU) errors (slope = 0.44±0.14SE, t = 3.14, p = .0047; p-value is conditional on OU $\alpha = 0.40$; calculated using function phylolm with option OUrandomRoot). (B) Correlation between mtDNA distance and phylogenetic distance between hybridizing pairs based on a global phylogeny of birds (Burleigh *et al*. 2015). Pearson's $r = 0.42$ ($p = 0.08$; $n = 18$). The taxon pair *Baeolophus atricristatus & B. bicolor* is an outlier (shown in red). Excluding this pair returns a Pearson's $r = 0.78$ ($p = .0002$; $n = 17$). These results suggest our estimates of dispersal rate and divergence are robust to the use of alternate predictors as proxies for hybrid zone dispersal and selection against hybrids. However, the mtDNA data used to estimate divergence in our study and the data used in the global phylogeny overlap, so a correlation would be expected.

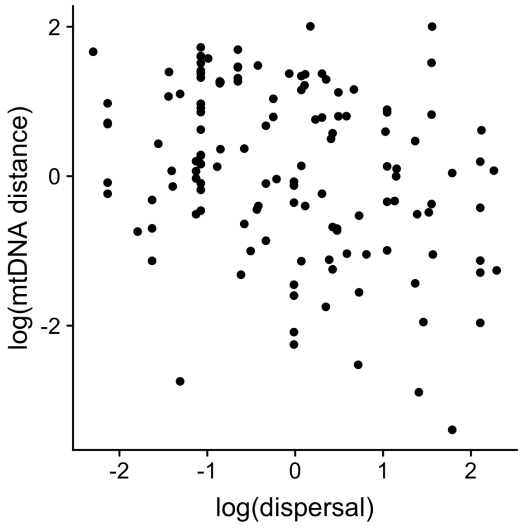

Figure S5: Dispersal length and mtDNA distance are weakly but significantly negatively correlated; Pearson's r = -0.31 (p-val = 3.1e-4, n = 131). This suggests that taxa with greater dispersal lengths either accumulate mtDNA divergence more slowly (indicative of divergence with gene flow) and / or come into secondary contact more quickly following initial divergence.

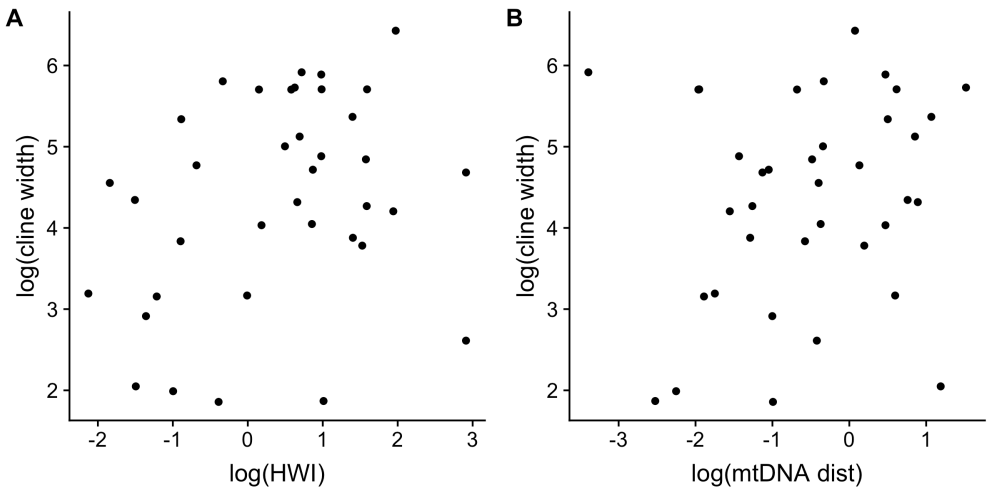

Figure S6: Scatterplots of the log of hybrid zone width (calculated as the geometric mean of all cline widths for a hybridizing pair) against (A) the log of the hand-wing index (HWI), and (B) the log of mtDNA distance, for the dataset including all bird hybrid zones for which hand-wing indices were estimated (n = 37 hybridizing pairs).

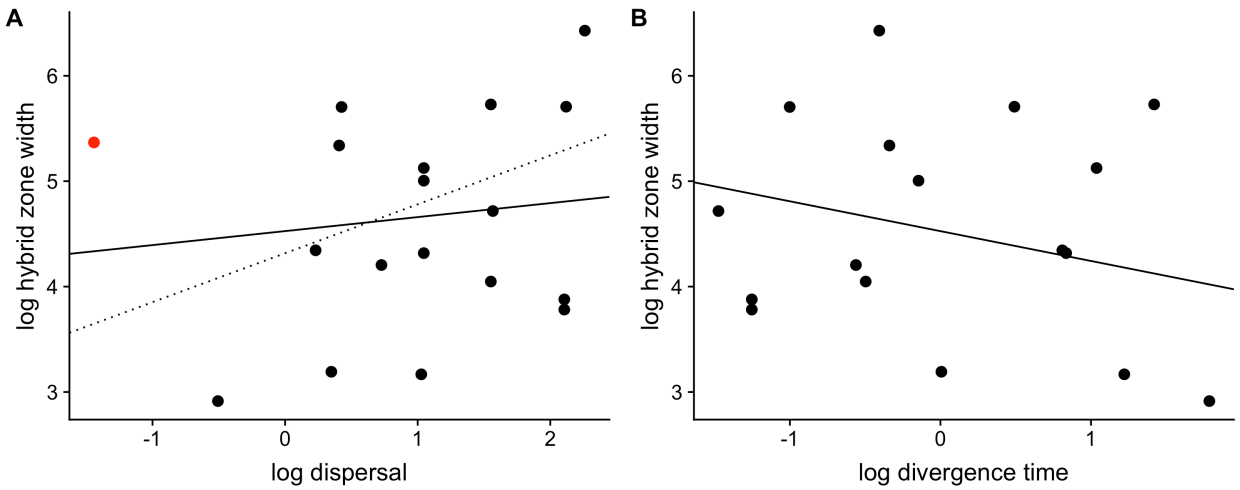

Figure S7: Scatterplots of the log of hybrid zone width (calculated as the geometric mean of all cline widths for a hybridizing pair) against (A) the log of dispersal, and (B) the log of divergence time for hybridizing species pairs from a phylogenetic analysis across bird diversity. Plotted lines in (A) are predictions from linear models, obtained via model averaging. The solid line results from modeling all data; the dotted line results after omitting the outlier shown in red (*Sula* boobies).

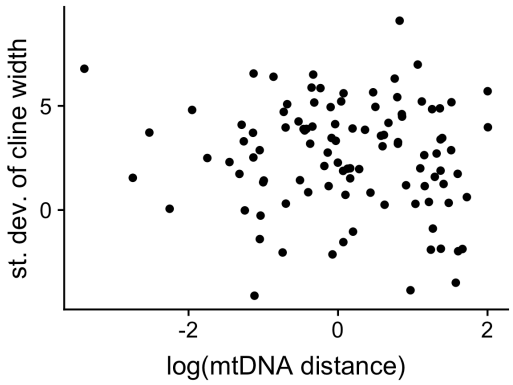

Figure 8: A test of the theoretical prediction that standard deviation of cline widths within a hybrid zone should decrease as selection strength between hybridizing pairs increases. As above, we used mtDNA divergence between taxon pairs as a proxy for selection strength. We did not recover the expected correlation r = -0.099 (p-val = 0.32, n = 104).