Our Ideal Example / how to write about a mixed model or generalized linear model

Lotterhos lab reading resources

Methods description:

- Sample sizes per treatment (in addition to total number)
- What are the explanatory variables and what the response variable (y)
    - Of the explanatory variables, which are modeled as fixed and which are random effects, are they categorical or continuous?
    - Were variables transformed or scaled and why?
    - Interactions
    - Model terms equation (specifying intercept, slopes, etc.)
- Model family (normal, poisson, binomial, etc.)
    - If Bayesian, spell out model terms and priors
- How were the model assumptions tested?
    - List what the assumptions are
    - Describe diagnostic plots or stats that they used to test
- How was the model fit? Restricted maximum likelihood?
- Model selection
    - How was it carried out?
    - Which R package or function? Which sum of squares?
- Post-hoc testing
    - Corrections for multiple tests
- Is the decision-making process described, or is it more of a "this is what we did"
- Describe *a priori* specific hypotheses or quantities you plan to report
- Think about how to explain rationale in Methods; put specifics in Supplemental Methods

Results description:

- If outliers or anomalies
    - Report results with and without outliers, especially if they change results
- If model selection was carried out
    - Report the best model, decision-making process
- If there is an interaction and main effects
    - Describe the interaction first, what drives it?
    - Post-hoc tests for interaction
    - Then discuss the main effects in terms of the interaction
- When they describe significance (or in table of model terms)
    - The null hypothesis being rejected
    - Df
    - Test statistic or R^2
    - P-value

- TALK ABOUT NON-SIGNIFICANT RESULTS
    - Make sure to report things that you tested, even though they aren't significant. This helps to reduce publication bias towards "significant" results.

Other figures/tables:

For discussion:

- When to include interactions? Most of the time.

How to keep it short:

- Parallel sentence structure
- Focus on the story, put the details in the supplement

Example: Rockfish

Your name: Katie

Paper Citation: Comparative thermal performance among four young-of-the-year temperate reef fish species. Schaal and Lotterhos

Paper Link: unpublished

Methods description:

The following analyses were carried out using model selection starting with the most complex model including any interaction terms and dropping terms based on significance from Likelihood Ratio Tests (LRT). Residuals of the models were tested for meeting assumptions using Shapiro-Wilk tests for normality, and diagnostic plots for autocorrelation to ensure repeated measurements on each tank were appropriately modeled. If any model violated assumptions (even after data transformation), we used Monte Carlo permutations to test the null hypothesis (Whitlock and Schluter, 2019). Additionally, all models were evaluated for significant interaction terms to ensure the correct sums of squares were implemented in the Analysis of Deviance model (i.e., Type II sums of squares without a significant interaction term and Type III sums of squares with a significant interaction). These were implemented using the *Anova()* function from the *car* package in the R statistical environment version 3.4.0 (Core Development Team and Others, 2019).

Survival

We recorded the number of days survived for each fish. These data were converted to binary data for the number of days alive vs. not alive and logistic regression was carried out using Generalized Linear Model (GLM) with a logit link function and a quasibinomial error distribution to account for overdispersion in the data. Fish survival (log odds of death) was modeled as a function of the independent and interactive effects of temperature (covariate) and species complex (factor) (Model 1). In addition, we used the DNA barcoding data to conduct an analogous model with the species factor replacing the species complex factor (Model 2). A significant interaction term would indicate that the relationship between temperature and survival varied across species complexes (Model 1) or species (Model 2) (i.e., differences in the slope of the regression).

Foraging Activity

To assess whether prolonged exposure to temperature stress impacted foraging activity, we used a general linear model to relate the activity rate to the independent and all two- and three-way interactive effects of species complex (factor), the continuous predictor of temperature (covariate), and time since beginning of experiment (covariate). The 3-way interaction term allows us to test whether the effect of increasing temperature on foraging activity varies with time and whether that varies between species complexes. For example, if higher temperature treatments decrease activity immediately and stay consistent through time for one complex, but show a more extreme decline over time for another complex, this would result in a significant 3-way interaction. Because this was a repeated measures analysis, we also inspected the residuals for any evidence of autocorrelation with the acf() function in R. In addition, because we found deviations from normality that could not be resolved by transformation (see Results), we used Monte Carlo permutations to compute the *P*-value for each model parameter.

Results description:

Survival

Because of overdispersion in our data (dispersion parameter = 4.98), all survival analyses were carried out using a quasibinomial model. The best model according to LRT included a 2-way interaction between species complex and temperature due to a steeper slope in the CQ complex compared to the BY complex ($X^2_{df = 1}$ = 6.09, P = 0.014; Table 2). In addition, we detected a significant main effect of temperature due to the decline in survival with increased temperatures and a main effect of species complex driven by higher overall mortality in the CQ complex ($P < 0.05$; Table 2, compare Figure 2A for BY to Figure 2B for CQ).

When analyzing survival variation among species, we could not parse out black and yellowtail rockfish due to model overfitting from a low sample size of yellowtail rockfish across the temperature treatments (n = 9) and only a single mortality event within that species (standard error of model coefficients for yellowtail were inflated due to overfitting, Table S2). Therefore, we compared the BY complex with copper and quillback rockfish to determine whether either copper or quillback rockfish were driving the decreased survival in the CQ complex compared to the BY complex. The best model according to a LRT included the main effects of species and temperature, but not their interaction ($X^2_{df = 2}$ = 4.95, P = 0.084; see Supplement R Markdown for model selection results). Therefore, temperature and species both independently influenced the degree of survival, with survival declining with temperature at the same rate for all species ($P < 0.001$), but with the adjusted means showing differences among species ($P < 0.001$, Table 3). The slope of the relationship between survival and temperature did not significantly vary among BY complex, copper, and quillback rockfishes (Figure 2A, C, and D). After performing a Tukey-Kramer contrast, we found that quillback rockfish had overall lower survival (i.e., lower intercept) than both copper and BY complex rockfish, but survival for copper and BY complex rockfish did not differ from one another (i.e., curves in Figure 2A and C for BY and copper are not statistically different from each other, but the curve in Figure 2D for quillback has overall lower adjusted means, Table S3).

Foraging Activity

Foraging activity declined with temperature for BY complex rockfish throughout the experiment (Figure 4A, B), but only declined with temperature in the CQ complex after prolonged exposure (Figure 4C, D). The best model to describe foraging activity rates according to LRT included the

3-way interaction between the categorical predictor of species, the continuous predictor of temperature and continuous predictor of time ($X^2_{df=1}$ = 3.8, P = 0.048). We found no evidence of autocorrelation in the residuals of our model, but ran a Monte Carlo permutation due to model residuals not being normally distributed even after data transformation. Under this analysis, the 3-way interaction was not significant (P = 0.1; Table 4), but the 2-way interactions between both species complex and temperature (P < 0.05, Table 4) and between species complex and time (P < 0.001, Table 4) were significant. The significant species-complex-by-temperature interaction indicated that the slope of the relationship between activity and temperature varied between species complexes, and was driven by the BY complex having a steeper decline in activity with temperature compared to the CQ complex over all timepoints (Figure S3). The significant species-complex-by-time interaction indicated that the slope of the relationship between activity and time varied across species complexes, and was driven by a greater decline in mean activity over time for the CQ complex, whereas mean BY activity did not change over time (Figure S4).

To further understand these complex interactions, we analyzed foraging activity for each species complex separately. For the BY complex, the best model according to LRT included the continuous predictor of temperature (covariate) and the continuous predictor of time (covariate). After Monte Carlo permutation, foraging activity in the BY complex decreased significantly due to temperature, but there was no main effect of time ($F_{df=1}$ = 52.57, P < 0.001 and $F_{df=1}$ = 0.055, P > 0.05, respectively; Figure 4A, B). However, for foraging activity in the CQ complex, the best model according to LRT included the significant 2-way interaction between temperature and time ($F_{df=1}$ = 9.27, P < 0.01; Figure 4C, D). This interaction was caused by an initial lack of effect of temperature on foraging activity, with a shift starting on day 10 of the experiment to a decline in activity at temperatures higher than 18°C (Figure 4C, D). To identify whether the change in activity was confounded by any large mortality event on day 10 for the CQ complex, we plotted the survival in tanks above and below 18°C across time (Figure S5). This plot shows a consistent decline in survival through time (starting at day 2 for quillback rockfish and day 6 for copper rockfish), suggesting that this shift in activity on day 10 was not driven by a large mortality event.

Other figures/tables:

Table 4. Results of a Type II Sums of Squares ANOVA output, for the response variable of foraging activity and the explanatory variables of temperature, time in treatment, species complex, and all possible interactions. Significance codes for probability of the test statistic under the null hypothesis: *** P < 0.001, ** 0.001 < P < 0.01, * P < 0.05 where significance is determined from Monte Carlo permutation estimations of the null distribution of the F statistic under the null hypotheses. DF = degrees of freedom, SS = sums of squares, MS = mean square, F = F statistic.

Explanatory Variable	**DF**	**SS**	**F**	**Null Hypothesis**
Temperature	2	87.6	45.7 ***	Foraging activity does not vary with temperature
Time	1	26.7	27.8 ***	Foraging activity does not vary with time

Species complex	2	31.3	16.3 ***	Foraging activity does not vary between species complexes
Temperature x Time	1	4.5	4.7	The change in foraging activity across temperatures does not vary over time
Temperature x Species complex	1	5.8	6.1 *	The change in foraging activity across temperatures does not vary with species
Time x Species	1	22.2	23.2 ***	The change in foraging activity through time does not vary with species
Temperature x Time x Species	1	3.8	3.9	The interaction between any two factors differs across levels of the third factor
Residuals	522	500.9		

Example: Frogs, what else?

Your name: Molly

Paper Citation: Albecker, M.A., Pahl, M., Smith, M., Wilson, J.G. and McCoy, M.W., 2020. Influence of density and salinity on larval development of salt‐adapted and salt‐naïve frog populations. *Ecology and Evolution*, *10*(5), pp.2436-2445.

Paper Link: https://onlinelibrary.wiley.com/doi/full/10.1002/ece3.6069

Methods description:

We tested for the effects of salinity, density, and location on tadpole survival, tadpole growth, size at metamorphosis, and age at metamorphosis using generalized linear mixed effects models

using package “lme4” (Bates, Martin, Bolker, & Walker, 2015) in the R statistical programing environment version 3.5.0 (Team, 2018). For all endpoints except growth, we treated salinity, density, and location as fixed effects, and replicate as a random effect. For each analysis, we started with the full interaction model, which included the fully crossed effects of location, salinity, and density, and then reduced model complexity by sequentially dropping higher order terms based on likelihood ratio tests.

For our analysis of survival, we assumed a binomial error distribution, and we assumed a Poisson error distribution to analyze the number of days between hatching and metamorphosis (i.e., age at metamorphosis defined as Gosner stage 42, the day of forelimb emergence; Gosner, 1960). We use Poisson because these data are integers derived from a count-based sampling effort which is assumed to conform to the Poisson distribution (Bolker et al., 2009). To test for differences in size at metamorphosis, we analyzed data on total length at developmental stage 42 assuming a log-normal error distribution. We use log-normal distribution for size measurements because size is by definition bound at zero, and model diagnostics improved following log-transformation.

Results description:

Salinity and location had no effect on survival to metamorphosis (salinity: $X^2_6 = 0.29$, $p = .59$; location: $X^2_6 = 1.14$, $p = .29$), but density significantly affected survival ($X^2_6 = 15.47$, $p = .0004$) (Figure 1). Survival declined by approximately 26% in the high-density treatments relative to the low-density treatments.

Location had no effect on the time to metamorphosis ($X^2_6 = 0.60$, $p = .44$). Salinity had a marginal effect ($X^2_6 = 3.09$, $p = .07$), and density had a significant effect on the time to metamorphosis ($X^2_6 = 466.76$, $p < .0001$) (Figure 2). Tadpoles in the low-density treatment reached metamorphosis the soonest, while tadpoles took the longest time to reach metamorphosis in the high-density treatment, with individuals in the 4 ppt treatment taking approximately 4 days longer to metamorphose than freshwater individuals.

Salinity did not affect length at metamorphosis ($X^2_6 = 0.11$, $p = .74$), but we observed an effect of location ($X^2_6 = 8.67$, $p = .003$) and a marginal effect of density ($X^2_6 = 3.51$, $p = .06$) (Figure 3). On average, coastal tadpoles metamorphosed approximately 1.7mm longer than inland tadpoles, and tadpoles that metamorphosed from the low-density treatments were slightly larger than the high-density individuals.

Other figures/tables:

Example: Estimating heritability in human traits

Your name: Alan / Ian

Paper Citation: Linear mixed model for heritability estimation that explicitly addresses environmental variation

Paper Link: https://www.pnas.org/content/113/27/7377

Methods description:

In this paper the authors were interested in evaluating whether including a random 'environment' variable in a linear mixed model used to estimate heritability (specifically, the additive genetic component or 'narrow-sense' heritability), would impact and potentially improve the models' estimate of heritability. To test this, they estimated heritability using a simple linear mixed model both without:

$$\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{\mu}; \sigma_g^2 \mathbf{K_{IBD}} + \sigma_r^2 \mathbf{I}\right).$$

And with:

$$\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{\mu}; \sigma_g^2 \mathbf{K_{IBD}} + \sigma_e^2 \mathbf{K_{loc}} + \sigma_{gxe}^2 \mathbf{K_{GxE}} + \sigma_r^2 \mathbf{I}\right), \qquad [8]$$

Environment (K_{Loc}) and the interaction of gene and environment (K_{GxE}) included in the model. The authors applied this model to both a simulated data set and an empirical data set that included 5,000 individuals from 9 different ethnic groups in Uganda and 34 phenotypes ranging from anthropometric indices to lipid tests. Importantly, lacking knowledge of the true causal variants for each phenotype the genetic component of the model was estimated as the fraction of the genome that is identical by descent (K_{IBD}), while the 'environment' (K_{LOC}) was estimated based on the spatial distance of each individual. Narrow sense heritability was estimated as the genetic variance divided by the sum of predictor variances in the model.

Results description:

The results of the paper found that estimates of heritability were heavily influenced by the corrections made. Both the case-study and the simulated data showed that heritability decreased after environmental effects were explicitly modeled. The corrections were most substantial in anthropometric indices although the paper states that more analysis would have to be done to determine whether specific environmental effects are really responsible for this.

Other figures/tables:

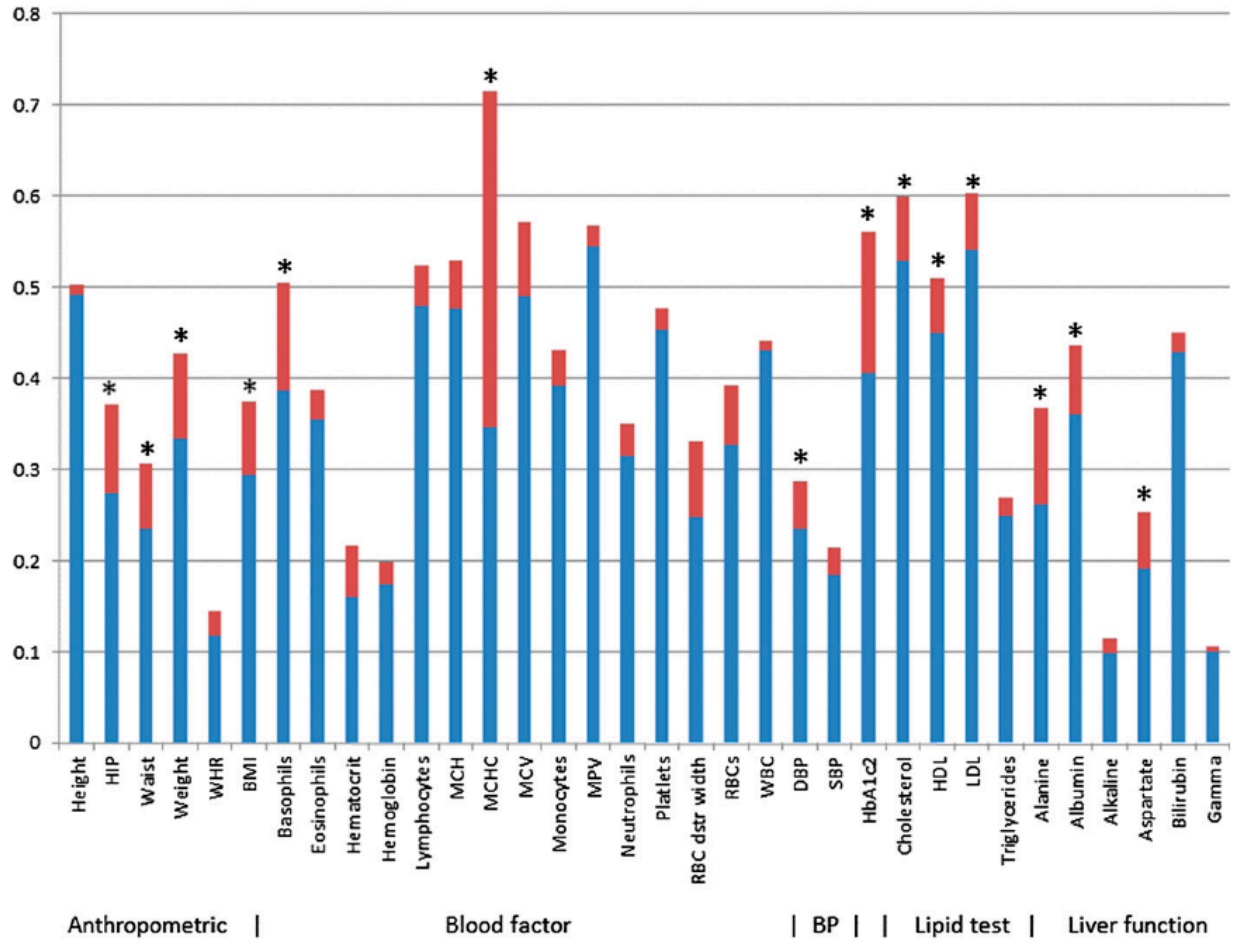

Fig. 1. Uncorrected and corrected estimates of narrow-sense heritability for phenotypes from the Ugandan cohort. The height of the blue and red bar combined corresponds to the uncorrected heritability estimate (based on Eq. **3**). The height of the blue bar corresponds to the corrected heritability estimate (based on Eq. **5**). Asterisks denote differences that are statistically significant after Bonferroni correction based on a two-sided test on the difference between uncorrected and corrected estimates from a 500-group jackknife.

Example: accumulation of reproductive isolation

Your name: Sara/Lauren

Paper Citation: Factors contributing to the accumulation of reproductive isolation: a mixed model approach

Paper Link: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5552923/

Methods description: The methods section is divided into a couple parts--the first describes the mixed model approach in general, then specifically describing the use of a mixed model for a) incorporating relatedness and categorical biogeography b) incorporating relatedness and categorical trait differences, c) incorporating multiple continuous variables with potential correlations and d) incorporating genetic distance matrix in place of phylogeny. The second part describes the interpretation of model outputs. Finally, the third piece of the methods section

utilizes four data sets to test several hypotheses including that of a significant difference in reproductive isolation between allopatric and sympatric species pairs as well as the hypothesis that the rate of accumulation of reproductive isolation differs between the two groups.

Part 1: Description of linear mixed model approach

$$Y = X\square + Z\mu + \varepsilon$$

Y is the response variable, X□ is fixed effects, Zμ is random effects, and epsilon is residuals.

Part 2: Model Description

I estimate the effects of a continuous trait (genetic distance) and a categorical trait (allopatry vs. sympatry; or trait presence vs. absence) and their interaction. In another analysis, I analyze three continuous variables simultaneously (genetic distance, geographic distance, and a phenotypic distance). The random phylogenetic effects of species used in the cross serve two purposes: to account for nonindependence due to phylogeny and account for variance that results from using individual species in multiple crosses in a dataset.

Genetic distance as a predictor variable represents the divergence between the two taxa used in a given cross and should not be confused with the pairwise genetic distance matrix (above) that can account for nonindependence between crosses due to shared evolutionary histories

For all models, mu is the intercept--the baseline level of reproductive isolation assuming there is not a significant relationship between reproductive isolation and genetic distance. A significant mu value (from zero) means there is reproductive isolation between a pair of species.

a) Model incorporating relatedness and categorical biogeography

$$y_{\text{RI}} = \mu + x_1\beta_{\text{gen.dist.}} + x_2\gamma_{\text{sym}} + (x_1 * x_2)\beta_{\text{int}} + Z_f f + Z_M m + e$$

X1 - genetic distance between species pair and βgen.dist is the slope of the relationship between genetic distance and reproductive isolation. γsym potential difference in reproductive isolation that can be attributed to the species occurring in sympatry. βint potential change in the slope of the relationship between reproductive isolation and genetic distance for sympatric species crosses. X2 sym is a dummy variable X2 zero when pair is allopatric, one when pair is sympatric this is so that when its an allopatric species pair the terms that are only for sympatric pairs go to zero. Lastly there are two Z matrices; they represent the identity of the female parent species and male parent species used in the cross (or can be designated species 1 and species 2 if sex of the parents is not important). Level of reproductive isolation between allopatric and sympatric species pairs (μ = baseline for allopatry and γsym = change in RI for sympatry)

b) Model incorporating relatedness and categorical trait differences

$$y_{RI} = \mu + x_1\beta_{gen.dist.} + x_2\gamma_{color} + (x_1 * x_2)\beta_{int} + Z_f f + Z_m m + e$$

The change from the first model is that X2 is now a categorical variable with two or more levels that in this case represents flower color based on what the crosses were (e.g., red x red, red x white). This model would test whether specific floral colors or morphologies have increased speciation rates compared to other floral morphologies

c) Model incorporating multiple continuous variables with potential correlations

$$y_{RI} = \mu + x_1\beta_{gen.dist.} + x_2\beta_{geo.dist} + x_3\beta_{corolla} + Z_f f + Z_m m + e$$

This model includes genetic distance, geographic distance, and the difference in corolla tube length, which is one measure that captures the difference in floral size between species. The model explicitly allows for covariance between these variables in the model by changing the prior structure of the predictor variables (fixed effects) from being independent (Equation 7), to being correlated. This disentangles the effect of multiple correlated variables and using a covariance matrix they directly model the correlation of the three continuous variables. X2/X3 are both continuous variables.

Part 3: Interpreting model outputs

Used MCMC glmm.

Part 4: Datasets

a) Drosophila

Originally used in Coyne and Orr, which estimated prezygotic isolation on choice/no choice mating assays (i.e., whether two individuals chose to mate). Postzygotic isolation was defined as either hybrid sterility or F1 inviability. Author constructed the phylogeny used so they could use genetic distance as an explanatory variable.

b) Bufonidae

Primary dataset includes an estimate of postzygotic isolation from crosses based on fertilization rate, hatching rate, number of tadpoles, percentage of tadpoles that metamorphosed, fertility in backcross analysis and the stage at which eggs ceased to develop. Author enriched original dataset by determining allopatry and sympatry relationships by looking at species ranges from the IUCN Red List Database.

c) Silene

Prezygotic isolation was the total number of failed pollinations in crosses. Postzygotic isolation was pollen sterility of F1 hybrids. Added flower color as a categorical explanatory variable and whether species were sympatric or allopatric.

d) Nolana

Total postzygotic isolation was a combination of fruit set, mericarp size, and seed set. Also looked at corolla diameter differences to quantify floral distance.

Results description:

1) Drosophila

The *Drosophila* dataset was used to test two hypotheses. a) there are differences in the level of reproductive isolation between allopatric and sympatric species and b) that reproductive isolation accumulates at different rates between allopatric/sympatric species. For prezygotic isolation, the baseline level of reproductive isolation was significantly different than zero and reproductive isolation was significantly elevated in sympatric pairs. For postzygotic isolation, there was not much or no postzygotic isolation in recently diverged species regardless of geographic context. Reproductive isolation may accumulate faster between sympatric pairs of species than allopatric pairs of species. The difference increases as genetic distance (divergence time) increases.

"For prezygotic isolation, the intercept (baseline level of reproduce isolation) was significantly different than zero, and reproductive isolation was significantly elevated in sympatric pairs (γsym = (0.2533, 0.4764), (lower 95% HPD interval, upper 95% HPD interval); Table 1). There was also a significant relationship between genetic distance and reproductive isolation, as detected in the original study (Coyne & Orr, 1989), but only for allopatric species pairs. In comparison, the overall relationship between genetic distance and reproductive isolation is nonsignificant for sympatric pairs (Table 1). In this model, the coefficients are additive (Equation 11), and the relationship between genetic distance and reproductive isolation is not significantly differ- ent from zero (HPD for $\beta gen.dist + \beta int$ = (−0.1250, 0.3571)). The effect of sympatry on reproductive isolation (γsym) is so strong that most of the reproductive isolation values are near 1 across all distances (com- pletely isolated; Figure 3).

For postzygotic isolation, the intercept was not significantly different than zero; neither was the increase in reproductive isolation in sympatry were not significantly different than zero (μ and γsym had HPD overlapping zero; Table 1). This indicates that there is little to no postzygotic isolation in recently diverged species regardless of geo- graphical context. The relationship between genetic distance and reproductive isolation ($\beta gen.dist$) was significant, and the rate of increase of reproductive isolation with genetic distance was greater in sympat- ric pairs (βint = 0.1326, 0.5247). This suggests that reproductive isola- tion may accumulate more quickly between sympatric pairs of species than allopatric pairs, and the difference increases as divergence time (genetic distance) increases."

2) Bufonidae

Secondly, the analysis of the *Bufonidae* dataset suggests that though sympatric pairs are separated by small genetic distances (little reproductive isolation), reproductive isolation accumulates more quickly for sympatric pairs than allopatric pairs.

Note: Supports hypothesis gleaned from *Drosophila* dataset woo!

3) Silene

The *Silene* dataset lacked allopatric pairs for postzygotic and total reproductive isolation measurements; analysis of effects of geographic context was not possible. However, analysis of available data reveals a significant relationship between genetic distance and reproductive isolation regardless of sympatry or allopatry. Floral color differences (categorical variable) did NOT increase reproductive isolation.

4) Nolana

The analysis of this dataset utilizes a genetic matrix rather than a phylogeny for relatedness (not quite sure what this means). Genetic distance, geographic distance, and measures of flower size differences were used as explanatory variables in the *Nolana* dataset. Geographic distance was the only significant predictor of reproductive isolation. There is more reproductive isolation between pairs of species that are more geographically distant.

Other figures/tables: Tables 1-4 summarize coefficients and intercepts from each respective analysis. Ex: Table 1 for *Drosophila*

Table 1

Summary of coefficients estimated for the analysis of prezygotic (left) and postzygotic (right) reproductive isolation from the *Drosophila* data. The confidence intervals are for 95% highest posterior density (HPD) and are significant if they do not include zero (in bold)

Coefficient	Biological meaning	Prezygotic		Postzygotic	
		Lower	Upper	Lower	Upper
μ (intercept)	Average RI	**0.1975**	**0.6033**	−0.0005	0.4001
$\beta_{gen.dist.}$	Slope relating genetic distance and RI	**0.1959**	**0.4164**	**0.1906**	**0.4644**
γ_{sym}	Additional RI in sympatry	**0.2533**	**0.4764**	−0.1162	0.1355
β_{int}	Increase in slope for sympatry	**−0.3209**	**−0.0593**	**0.1326**	**0.5247**

Reproductive isolation may accumulate more quickly between sympatric pairs of species than allopatric pairs due to the rate of increase in reproductive isolation with genetic distance being greater in sympatric pairs.

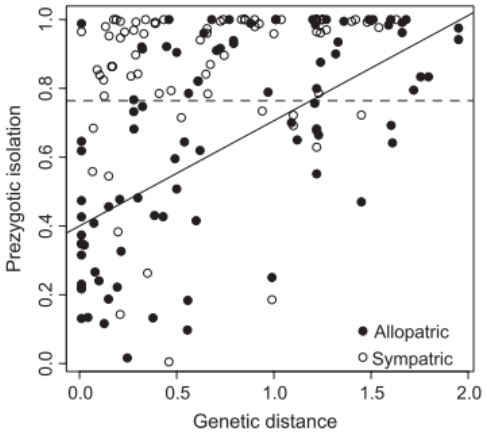

FIGURE 3 The relationship between genetic distance and prezygotic isolation for *Drosophila* species pairs that are either allopatric or sympatric, demonstrating that most sympatric species pairs show almost complete isolation, resulting in no relationship between genetic distance and reproductive isolation for the sympatric context. The best fit lines were constructed using the mode of the parameters from the MCMCglmm analysis. The solid line represents allopatric species pairs, and the dashed line represents sympatric species pairs

TABLE 2 Summary of coefficients estimated for the analysis of postzygotic reproductive isolation from the Bufonidae data. The original index of reproductive isolation (left) was calculated by Malone and Fontenot (2008) following the procedure of Coyne and Orr (1989) and Presgraves (2002). The new index (right) takes into account that reproductive barriers are sequential following Ramsey et al. (2003). The confidence intervals are for 95% highest posterior density (HPD) and are significant if they do not include zero (in bold)

		Postzygotic (Original)		Postzygotic (New)	
Coefficient	Biological meaning	Lower	Upper	Lower	Upper
μ (intercept)	Average RI	**0.2980**	**0.6945**	**0.8398**	**0.9686**
$\beta_{gen.dist.}$	Slope relating genetic distance and RI	**3.6212**	**5.8819**	**0.4472**	**1.2592**
γ_{sym}	Additional RI in sympatry	**−0.2816**	**−0.0427**	**−0.0988**	**−0.0105**
β_{int}	Increase in slope for sympatry	**0.4679**	**3.5163**	**0.2665**	**1.3992**

The effect of sympatry actually decreases the level of reproductive isolation. Intercept greater than zero. The effect of genetic distance on RI is quite steep (3.62, 5.88) and increased rate of accumulation in sympatric pairs was significant.

TABLE 3 Summary of coefficients estimated for the analysis of prezygotic reproductive isolation when considering geographical context (left) or floral divergence (right) from the *Silene* data. The confidence intervals are for 95% highest posterior density (HPD) and are significant if they do not include zero (in bold)

		Prezygotic (Geographic model)	
Coefficient	**Biological meaning**	**Lower**	**Upper**
μ (intercept)	Average RI	−1.2138	1.9094
$\beta_{gen.dist.}$	Slope relating genetic distance and RI	**0.9126**	**7.1054**
γ_{sym}	Additional RI in sympatry	−0.3643	0.1510
β_{int}	Increase in slope for sympatry	−1.5671	3.5080
		Prezygotic (Floral differences model)	
Coefficient	**Biological meaning**	**Lower**	**Upper**
μ (intercept)	Average RI	−1.0166	1.4672
$\beta_{gen.dist.}$	Slope relating genetic distance and RI	**3.4368**	**9.5154**
γ_{color}	Additional RI due to color differences	−0.1556	0.3158
β_{int}	Increase in slope for color differences	−2.4138	2.2990

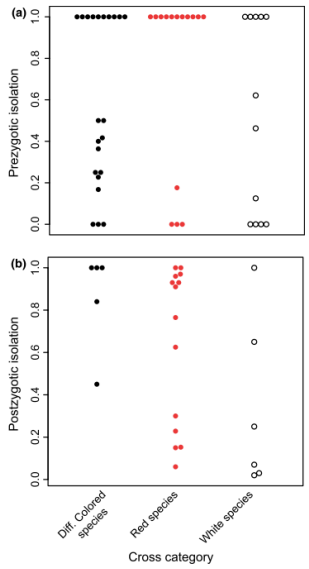

FIGURE 4 Variation in the level of reproductive isolation between *Silene* species pairs that either had different floral colors or shared floral colors (separated into pairs where both species were either white or red) for both (a) prezygotic and (b) postzygotic isolation demonstrating no effect of floral differences on reproductive isolation

Sympatry had no effect on baseline pre or postzygotic isolation or the rate of accumulation of reproductive isolation, but there was a significant effect of genetic distance and reproductive isolation that didn't vary between sympatric and allopatric pairs.