DEGREE OF SYMPATRY AFFECTS REINFORCEMENT IN *DROSOPHILA*

Patrik Nosil[1,2]

[1]*Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, United Kingdom*

[2]*E-mail: p.nosil@sheffield.ac.uk*

In a recent paper, Yukilevich (2012) showed that asymmetries between *Drosophila* species in the strength of premating isolation tend to match asymmetries in the costs of hybridization (inferred from asymmetries in the strength of postzygotic isolation and range sizes). The results provide novel evidence that the outcome of reinforcement can depend on the strength and frequency of selection against hybridization. Here, I reanalyze the data to demonstrate that another (unconsidered) factor, namely the quantitative degree of sympatry between species, also predictably affects reinforcement. Specifically, premating isolation is strongest at intermediate degrees of sympatry. This result complements, rather than challenges, those of Yukilevich (2012). One possible explanation for this newly discovered pattern is that when the degree of sympatry is small, selection for avoidance of hybridization is rare, but when the degree of sympatry is large, homogenizing gene flow overcomes reinforcing selection. Thus, reinforcement may depend on the balance between selection and gene flow. However, the current work examined degree of sympatry, not gene flow itself. Thus, further data on gene flow levels in *Drosophila* is required to test this hypothesis, which emerged from the patterns reported here.

KEY WORDS: Comparative analysis, gene flow, reinforcement, reproductive isolation, speciation.

Speciation by reinforcement has been one of the most debated mechanisms of speciation (Dobzhansky 1951; Butlin 1995; Servedio and Noor 2003; Coyne and Orr 2004; Ortiz-Barrientos et al. 2009). A major difficulty with the study of reinforcement is that processes other than reinforcement might also generate the pattern of reproductive character displacement (i.e., stronger prezygotic isolation between sympatric relative to allopatric taxon pairs). Thus, as argued by Coyne and Orr (2004) in their book on speciation, further predictions of reinforcement need to be developed and tested.

In a recent article, Yukilevich (2012) articulated such predictions and provided novel tests of them using data from species pairs of *Drosophila*. It was argued that premating isolation should evolve to be stronger when reinforcing selection is stronger or more common. By extension, when species pairs exhibit asymmetries in the strength or frequency of reinforcing selection they are predicted to exhibit concordant asymmetries in the strength of premating isolation. Yukilevich (2012) considered two proxies for

such asymmetric reinforcing selection: (1) asymmetry between species in the degree of postzygotic isolation and (2) asymmetry in relative range sizes. It was argued that the species in a species pair suffering stronger hybrid dysfunction should evolve the stronger premating isolation. Likewise, it was argued that the species where females were "rarer" (inferred from asymmetries in range size) would experience more encounters with heterospecifics, increasing the probability of reinforcement for females of the rare species. Both these predictions were supported by the analyses. Thus, the strength and frequency of reinforcing selection appears to be a predictor of the outcome of reinforcement.

My goal here is to complement the main findings of Yukilevich (2012), which provide interesting and useful insight into reinforcement. I present new analyses of the data showing that a previously unconsidered factor also affects premating isolation in *Drosophila*. In short, I show that the strength of premating isolation between sympatric species pairs varies predictably with the quantitative degree of sympatry (percent sympatry hereafter).

Specifically, premating isolation is strongest at intermediate degrees of sympatry, potentially reflecting a balance between the opposing forces of selection and gene flow. Thus, the outcome of reinforcement appears complex and affected by factors in addition to asymmetries in selection, but is at least somewhat predictable.

Methods and Results
REGRESSION ANALYSES

Yukilevich (2012) estimated percent sympatry as follows. First, maps were used to estimate the absolute range size overlap between each species pair in square kilometers. The percent of overlap for each species is equal to absolute range size overlap between the two species/absolute range size of that species. The average percent of geographical overlap between species pairs (i.e., the percent sympatry treated here) was calculated by averaging these values across the two species in a pair.

For the data from sympatric species pairs presented by Yukilevich (2012) the association between the strength of premating isolation and percent sympatry appears curvilinear (i.e., a "hump-shaped" relationship, Fig. 1). I thus used regression analyses to test whether a quadratic model fit the data better than a linear model. Specifically, I used F-change tests to assess whether adding a quadratic term to a linear model significantly increased the r^2 value. The full results of these analyses are presented in Table 1 of this article. Unless specified otherwise, all further reference to tables refers to those published in Yukilevich (2012, see original study for details).

When considering the 74 sympatric species reported in Table S1 of Yukilevich (2012), adding a quadratic term to the model resulted in a significant increase in r^2 ($P = 0.002$). This result appears robust as it persists when various covariates are included and for analyses of subsets of the data. For example, for the 74 sympatric species, adding a quadratic term significantly increased the r^2 in an analysis that also included the strength of postzygotic isolation ($P = 0.001$). The result of a better fit of the quadratic model also persisted when considering only the species pairs from Table 1 that were the focus of the article by Yukilevich (2012) (i.e., those with asymmetries in isolation), and did so in both raw and phylogenetically corrected analyses ($P = 0.019$ and 0.040, respectively). In contrast to the results for premating isolation, there was no clear association between postzygotic isolation and percent sympatry ($P = 0.458$ and 0.251 for linear and quadratic terms, respectively, in a quadratic regression model).

Yukilevich (2012) discussed how selection to avoid hybridization can be accentuated for females from the rarer species in a species pair because they more commonly encounter heterospecifics. Consistent with this hypothesis, the study reported that of 11 sympatric species pairs who had symmetrical postzy-

gotic isolation but asymmetric range sizes, nine pairs showed greater premating isolation in the reciprocal mating with females of the species with the smaller range size ($P < 0.05$ in a BiNomial test). If asymmetry in range size has a strong influence on reinforcement, one might also predict a positive correlation across species pairs between their asymmetry in premating isolation and their asymmetry in relative range sizes (all else being equal, i.e., asymmetries in hybrid unfitness could also affect the outcome of reinforcement). No such significant correlation between asymmetries in premating isolation and range size is observed in the data from Table 2 ($r = -0.27$, $P = 0.42$ in a simple linear model), and if anything, this relationship is negative (Table 1 of this study). Thus asymmetric gene flow between species with asymmetric range sizes (i.e., greater gene flow from the more common into the rarer species) might be working against the accentuated selection to avoid hybridization in the rarer species, obscuring any general relationship between asymmetry in premating isolation and asymmetry in range size. Notably, there was a marginal ($P = 0.11$) quadratic relationship between overall premating isolation and range size asymmetry in this same data.

STATISTICAL CONSIDERATIONS

Two statistical issues warranted consideration. First, in the data for sympatric taxa from Table S1 of Yukilevich (2012) there is a hump-shaped relationship between percent sympatry and genetic distance ($P = 0.011$, Table 1), raising the possibility that the quadratic relationship between premating isolation and percent sympatry reported here is confounded by genetic distance (i.e., a proxy for the age of species pairs). Second, normal regression analysis is not fully appropriate for data that is bounded between zero and one, such as the measure of premating isolation considered here. I treat issue in turn.

Several lines of evidence suggest the hump-shaped relationship between premating isolation and percent sympatry is not strongly confounded by the age of species pairs. First, a quadratic model of premating isolation on percent sympatry was almost significant in an analysis including genetic distance between species pairs (linear term only) as a covariate ($P = 0.054$, Table 1). Genetic distance itself was not significant in this model ($P = 0.46$), indicating it was not an important component to include in the model. Consistent with this suggestion, a reduced model derived using backward elimination excluded genetic distance but retained a significant quadratic effect of percent sympatry ($P < 0.05$). A model that included both linear and quadratic terms for both sympatry and genetic distance yielded a significant quadratic term for the former ($P < 0.05$), but not for the latter ($P = 0.45$). Analyzing the data via categorizing species pairs as "old" or "young" yield comparable results (here I did so using the arbitrary cutoff of young species pairs being those with $D < 0.05$, as in past work; Coyne and Orr 1989; Yukilevich 2012) (quadratic percent
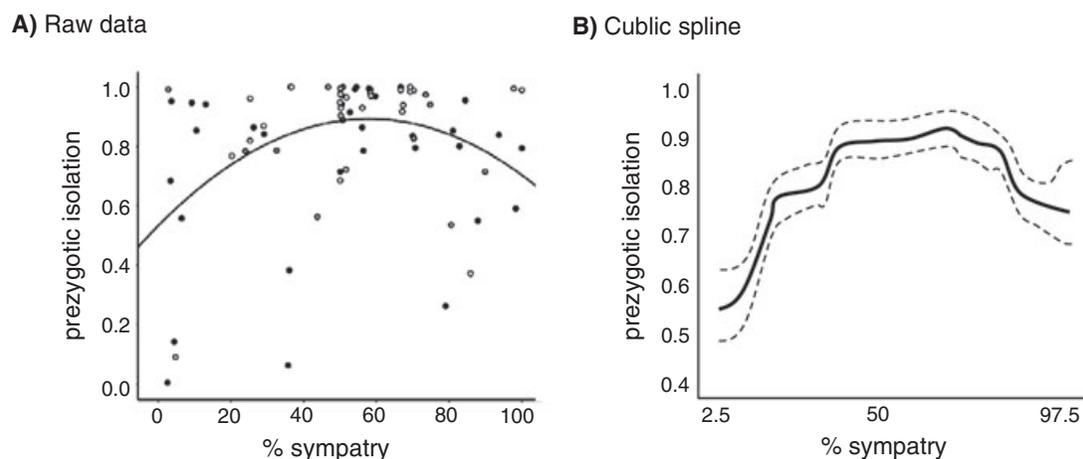
A) Raw data

B) Cublic spline

Figure 1. The relationship between premating isolation and percent sympatry between species pairs of *Drosophila.* All data are from Table S1 of Yukilevich (2012). Premating isolation is strongest at intermediate degree of sympatry. (A) Raw data and the fit of a quadratic model. Filled black circles = "young species pairs," filled gray circles = "old species pairs," unfilled circles = no genetic data available for characterization of age. See text for details. (B) Fit of the relationship using the cubic spline approach. Dashed lines are ±1 SE from 1000 bootstrap replicates. See Table 1 of this study for statistics.

Table 1. Statistical analyses testing the fit of linear versus quadratic regression models (negative quadratic terms indicate a "hump-shaped" relationship, see also Fig. 1). In general, for analyses of premating isolation on percent sympatry, adding a quadratic term significantly improved the fit of the model.

					Linear			Quadratic
Comparison	Table	r^2 linear	r^2 quadratic	F- change (df)	P	B (SE)	P	B (SE)
Premating on % sympatry	**S1**	**0.043**	**0.168**	**10.65 (1, 71)**	**0.002**	**1.24(0.34)**	**< 0.001**	**−1.07(0.33)**
Postzygotic on% sympatry	S1	0.020	0.038	1.34 (1, 71)	0.251	−0.41(0.54)	0.458	0.61(0.53)
Premating on % sympatry+genetic distance	S1	0.044	0.124	3.93 (1, 43)	0.054	0.98(0.44)	0.031	−0.89(0.45)
Premating on % sympatry+postzygotic	**S1**	**0.059**	**0.199**	**12.21 (1, 70)**	**0.001**	**1.29(0.33)**	**< 0.001**	**−1.14(0.33)**
% sympatry on genetic distance	**S1**	**0.004**	**0.141**	**7.06 (1, 44)**	**0.011**	**0.91(0.38)**	**0.022**	**−0.75(0.28)**
Premating on% sympatry	**1**	**0.278**	**0.536**	**7.21 (1, 13)**	**0.019**	**1.88(0.57)**	**< 0.001**	**−1.31(0.49)**
Premating on % sympatry[1]	**S3**	**0.342**	**0.578**	**5.57 (1, 10)**	**0.040**	**1.96(0.67)**	**0.014**	**−1.43(0.61)**
Premating asymmetry on range asymmetry	2	0.071	0.197	1.22 (1, 8)	0.301	−1.08(0.87)	0.249	0.92(0.83)
Premating on range asymmetry	2	0.068	0.336	3.23 (1, 8)	0.110	1.19(0.61)	0.089	−1.06(0.59)

Abbreviations are as follows (premating = premating isolation, postzygotic = postzygotic isolation, range asymmetry = asymmetry in relative range sizes between taxon pairs with the smaller range size subtracted from the larger one). The "Table" column reports the table in Yukilevich (2012) that the data stem from. Significant quadratic models are in bold font (note that the significance of the quadratic term is identical to that of the *F*-change).
[1]Phylogenetically corrected analysis.

sympatry term in models with old/young as a covariate, full model, $P = 0.061$, reduced model, $P < 0.05$, model including only young species pairs, $P = 0.08$). Visual examination of the data illustrates this point that the trend reported here is not driven solely by young or solely by old species pairs (Fig. 1).

The quadratic relationship between premating isolation and percent sympatry is not an artifact of analyzing data bounded between zero and one using normal regression. For example, the relationship was also "hump-shaped" when visualized using the cubic spline approach of Schluter (1988), which makes no a priori assumptions about the form of the relationship (Fig. 1). Addi-

tionally, and perhaps most convincingly, a quadratic association was supported using regression analyses that assume a beta error distribution, rather than a normal distribution. Beta errors are continuous and bounded by zero and one and thus "beta regression" is appropriate for analyzing data on reproductive isolation (see Cribari-Neto and Zeileis 2010). Beta regression of premating isolation on percent sympatry was implemented in R and yielded three main results. First, a linear model was nonsignificant ($z = 1.84$, $P > 0.05$, pseudo-$r^2 = 0.04$, log-likelihood $= 49.32$ on 3 df). Second, the quadratic model was significant ($z = -2.80$, $P < 0.01$, pseudo-$r^2 = 0.16$, log-likelihood $= 53.15$ on 4 df). Third, a

log-likelihood test indicated the quadratic model was significantly better than the linear model ($\chi^2 = 7.66$, $P = 0.006$).

Discussion

"*reinforcement requires some gene flow, but not too much*". Coyne and Orr (2004, p. 371).

In summary, the most robust new result to emerge is that premating isolation between sympatric species of *Drosophila* is strongest at intermediate degrees of sympatry. This result complements those of Yukilevich (2012) to increase our understanding of premating isolation in *Drosophila*. Below I discuss whether this could be due to effects of gene flow. My goal here is not to be exhaustive, as reviews of reinforcement can be found elsewhere (Butlin 1995; Servedio and Noor 2003; Coyne and Orr 2004; Ortiz-Barrientos et al. 2009), but rather to focus on the most salient points.

Gene flow between populations is often a homogenizing force that prevents or constrains population divergence. Indeed, theoretical models have demonstrated that high levels of gene flow between diverging populations can erode the effects of reinforcing selection, preventing reproductive character displacement (Sanderson 1989; Servedio and Kirkpatrick 1997; Cain et al. 1999; Servedio and Noor 2003). However, gene flow also generates the opportunity for selection against hybridization to occur in the first place. Thus, gene flow can exert a dual effect during reinforcement, as modeled by Kirkpatrick (2000). The quotation above exemplifies how a potential prediction is that the effects of reinforcement are maximized when gene flow is intermediate; that is, high enough to allow the evolution of reinforcement, but low enough to prevent homogenization of adaptive divergence in mate choice.

Few empirical studies have examined the effects of gene flow on the outcome of reinforcement (Servedio and Noor 2003 for review). Perhaps the clearest example from nature is a study by Nosil et al. (2003) examining the effects of gene flow on the outcome of reinforcement during ecological speciation in walking-stick insects. The results demonstrate the dual effects of gene flow: the magnitude of female mating discrimination against males from other populations was greatest when gene flow between populations adapted to alternate host plants was intermediate. The generality of this result is unknown, although at least three other studies have documented stronger effects of reinforcement on the species within a hybridizing pair that is less abundant, and thus undergoes more frequent encounters with heterospecifics (Waage 1979; Noor 1995; Peterson et al. 2005). Additionally, a laboratory experimental evolution study has demonstrated that reinforcement can evolve in the face of gene flow, so long as gene flow levels are not too high (Matute 2010). Finally, even after reinforcement occurs, gene flow may be required to spread alleles for mating

discrimination outside of the immediate zone of hybridization and across the species range (Coyne and Orr 2004; Ortiz-Barrientos et al. 2009).

I stress that I propose a relationship between gene flow and reinforcement in *Drosophila* not as conclusion, but rather as a testable hypothesis that emerges from the pattern reported here and that is consistent with theory and work in other systems. The only factor examined here was sympatry, not gene flow, and thus further data on gene flow levels between species and populations of *Drosophila* is required to now test the hypothesis. It is very possible that factors other than gene flow contribute to the results reported here, such as the evolution of habitat isolation, which could affect gene flow levels independent of degree of sympatry and also either promote or interfere with the reinforcement of mating preference (Yukilevich and True 2006; Nosil and Yukilevich 2008).

As noted above, an interesting pattern is that there is in fact a hump-shaped relationship between percent sympatry and genetic distance. This could arise if the species pairs with intermediate degrees of sympatry, which I have argued also have the strongest premating isolation, undergo less gene flow than species with greater degrees of sympatry, resulting in higher genetic distances. However, a mechanism other than gene flow, such as secondary contact between recently diverged lineages, is required is required to explain the low genetic distance of species pairs with the smallest degrees of sympatry. Some studies testing for gene flow in *Drosophila* do exist, with gene flow detected between some species pairs (Hey and Nielsen 2004) but not others (Counterman and Noor 2006). Thus, the variability in gene flow required to test its effects on reinforcement may exist and recent advances in DNA sequencing make it feasible to quantify gene flow in numerous species pairs (Ellegren 2008; Hohenlohe et al. 2010). Thus, future work explicitly testing the effects of the balance between selection and gene flow on reinforcement is warranted.

LITERATURE CITED
Butlin, R. K. 1995. Reinforcement—an idea evolving. Trends Ecol. Evol. 10:432–434.

Cain, M. L., V. Andreasen, and D. J. Howard. 1999. Reinforcing selection is effective under a relatively broad set of conditions in a mosaic hybrid zone. Evolution 53:1343–1353.

Counterman, B. A., and M. A. F. Noor. 2006. Multilocus test for introgression between the cactophilic species *Drosophila mojavensis* and *Drosophila arizonae*. Am. Nat. 168:682–696.

Coyne, J. A., and H. A. Orr. 1989. Patterns of speciation in *Drosophila*. Evolution 43:362–381.

———. 2004. Speciation. Sinauer Associates, Inc., Sunderland, MA.

Cribari-Neto, F., and A. Zeileis. 2010. Beta regression in R. J. Stat. Softw. 34:1–24.

Dobzhansky, T. 1951. Genetics and the origin of species. Columbia Univ. Press, New York, NY.

Ellegren, H. 2008. Sequencing goes 454 and takes large-scale genomics into the wild. Mol. Ecol. 17:1629–1631.

Hey, J., and R. Nielsen. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167:747–760.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko. 2010. Population genomics of parallel adaptation in three-spine stickleback using sequenced RAD tags. PLoS Genet. 6:e1000862. doi:10.1371/journal.pgen.1000862.

Kirkpatrick, M. 2000. Reinforcement and divergence under assortative mating. Proc. R. Soc. Lond. Ser. B 267:1649–1655.

Matute, D. R. 2010. Reinforcement can overcome gene flow during speciation in *Drosophila* Curr. Biol. 20:2229–2233.

Noor, M. A. 1995. Speciation driven by natural selection in *Drosophila*. Nature 375:674–675.

Nosil, P., B. J. Crespi, and C. P. Sandoval. 2003. Reproductive isolation driven by the combined effects of ecological adaptation and reinforcement. Proc. R. Soc. Lond. Ser. B 270:1911–1918.

Nosil, P., and R. Yukilevich. 2008. Mechanisms of reinforcement in natural and simulated polymorphic populations. Biol. J. Linn. Soc. 95:305–319.

Ortiz-Barrientos, D., A. Grealy, and P. Nosil. 2009. The genetics and ecology of reinforcement implications for the evolution of prezygotic isolation in sympatry and beyond. Ann. N Y Acad. Sci. 1168:156–182.

Peterson, M. A., B. M. Honchak, S. E. Locke, T. E. Beeman, J. Mendoza, J. Green, K. J. Buckingham, M. A. White, and K. J. Monsen. 2005. Relative abundance and the species-specific reinforcement of male mating preference in the *Chrysochus* (Coleoptera: Chrysomelidae) hybrid zone. Evolution 59:2639–2655.

Sanderson, N. 1989. Can gene flow prevent reinforcement. Evolution 43:1223–1235.

Schluter, D. 1988. Estimating the form of natural selection on a quantitative trait. Evolution 42:849–861.

Servedio, M. R., and M. Kirkpatrick. 1997. The effects of gene flow on reinforcement. Evolution 51:1764–1772.

Servedio, M. R., and M. A. F. Noor. 2003. The role of reinforcement in speciation: theory and data. Annu. Rev. Ecol. Evol. Syst. 34: 339–364.

Waage, J. K. 1979. Dual function of the damselfly penis—sperm removal and transfer. Science 203:916–918.

Yukilevich, R. 2012. Asymmetrical patterns of speciation uniquely support reinforcement in *Drosophila*. Evolution 66:1430–1446.

Yukilevich, R., and J. R. True. 2006. Divergent outcomes of reinforcement speciation: the relative importance of assortative mating and migration modification. Am. Nat. 167:638–654.

Associate Editor: K. Dyer