doi:10.1111/evo.12362

UNIFICATION OF METHODS FOR ESTIMATING THE STRENGTH OF REPRODUCTIVE ISOLATION

James M. Sobel[1,2,3] and Grace F. Chen[1,4]

[1]*Department of Plant Biology, Ecology, Evolutionary Biology, and Behavior Program, Michigan State University, East Lansing, Michigan 48824*

[2]*Current Address: Department of Biological Sciences, Binghamton University, State University of New York, Binghamton, New York 13902*

[3]*E-mail: jsobel@binghamton.edu*

[4]*Current Address: Department of Biology, State University of New York College at Oneonta, Oneonta, New York 13820*

Received June 13, 2013
Accepted January 14, 2014

Understanding the evolution of reproductive isolation is tantamount to describing the origin of species. Therefore, a primary goal in evolutionary biology is to identify which reproductive barriers are most important to the process. To achieve this goal, the strength of multiple forms of isolation must be compared in an equivalent manner. However, a diversity of methods has been used to estimate barrier strength, falling into several mathematically distinct categories. This study provides a unified method for calculating isolation that relates the amount of gene flow experienced by taxa to random expectations in a simple linear framework. This approach has three distinct advantages over previous methods: (1) it is directly related to gene flow, (2) it is symmetrical, such that measures in both the positive and negative range are comparable, and (3) it is equivalent between broad categories of reproductive isolation, allowing for appropriate comparisons. This linear formulation can be adjusted for use in all forms of isolation, and can accommodate cases in which null expectations for con- and heterospecific gene flow differ. Additionally, this framework can be used to calculate total reproductive isolation and the relative contributions of individual barriers.

KEY WORDS: **Gene flow, reproductive barriers, reproductive isolation indexes, speciation.**

Within the framework of the biological species concept (Mayr 1942, 1963), the process of speciation involves the accumulation of barriers to gene exchange leading to complete reproductive isolation (Coyne and Orr 2004). Reproductive barriers can take many forms, from prezygotic barriers such as ecogeographic isolation and mating isolation to postzygotic barriers such as intrinsic hybrid inviability (Dobzhansky 1937; Mayr 1942; Poulton 1908). However, general rules governing the relative strengths of different forms of isolation are largely unknown, and there has been longstanding debate about which categories of isolating barriers are most important in speciation (Rice and Hostert 1993; Schemske 2000; Coyne and Orr 2004; Sobel et al. 2010). An overlooked, but essential issue in this debate involves the methods used to calculate the strength of reproductive barriers. To make meaningful comparisons of the strength of isolation between barriers and taxa, a unified means of estimating their strength is both required and lacking.

Online enhancements: Reproductive Isolation Calculator, Appendix.

Early work on the nature of reproductive barriers sought to quantify the degree to which crossing barriers prevented hybridization or resulted in hybrid unfitness, often in an attempt to discern phylogenetic relationships (e.g. Stalker 1942; Vickery 1959; Ehrman 1965). As molecular phylogenetic methods allowed for an independent means of estimating relationships among species, barrier strength was instead examined to discern the forces responsible for species formation. For example, in their highly influential study, Coyne and Orr (1989, 1997) used data from the literature on 171 interspecific hybridization attempts in *Drosophila* to compare the relative rate of evolution between broad categories of reproductive isolation. A surge of both single species pair studies (e.g. Ramsey et al. 2003; Matsubayashi and

EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

Katakura 2009; Dopman et al. 2010) and comparative data on the strength of reproductive isolation resulted (e.g. Mendelson 2003; Moyle et al. 2004; Bolnick and Near 2005).

These studies have provided valuable insight into the forces responsible for species formation within groups. However, metrics for estimating barrier strength were not standardized, resulting in a diversity of approaches to calculate reproductive isolation. This can be problematic within studies, but especially hinders comparisons across taxa and barriers (e.g. Funk et al. 2006; Lowry et al. 2008). The lack of consensus on the methods used to calculate indexes of isolation thus impedes our ability to generalize the patterns of speciation. Therefore, our goals in this paper are to review the most common methods used to calculate reproductive isolation by previous authors, present a simple derivation of a single unified method, and provide examples to illustrate its advantages. Ultimately, we hope this unification of methods will allow researchers the opportunity to compare measures of reproductive isolation among disparate forms of isolation and/or taxa, and provide a metric of isolation that has an intuitive connection to gene flow.

CURRENTLY USED METHODS

The purpose of calculating the strength of a reproductive isolating barrier is to estimate how much gene flow is reduced by a barrier (Coyne and Orr 2004). In most cases, previous workers have used equations to describe reproductive isolation that range from 0 when there is no isolation to 1 when there is complete isolation. However, as long as a metric satisfies this one requirement, little attention has been paid to the relationship between the strength of a barrier and the extent to which gene flow is reduced by it. As a result, several algebraically nonequivalent formulas have been proposed (Table 1).

In the case of prezygotic isolation, estimates of isolation attempt to relate how a barrier will affect the probability of heterospecific zygote formation. For example, sexual barriers due to mating preferences have been widely studied in the laboratory as an agent of isolation (e.g. Rice and Hostert 1993; Gleason and Ritchie 1998). The equation used to describe the relationship between mating preferences and isolation presented in Coyne and Orr (1989) has been widely adopted, and can be expressed as:

$$\text{Repro. Iso.} = 1 - \frac{\text{freq. of heterospecific matings}}{\text{freq. of conspecific matings}};$$

which will be abbreviated as:

$$RI_1 = 1 - \frac{H}{C}. \quad (1)$$

The metric indeed results in an isolation index that equals 1 when mating preferences result in no heterospecific mating and zero when there are no preferences. Based on RI_1, the relationship between mating preference and isolation is not linear between 0 and 1, and ranges to $-\infty$ in cases of disassortative mating (Fig. 1). This lack of linearity is problematic in that measures of isolation are not easily comparable, both due to a lack of direct proportionality and asymmetry in the positive and negative range.

A similar equation that has been proposed relates heterospecific-mating frequency to the sum of both hetero- and conspecific mating. This method has been used in several examples of pollination isolation (e.g. Ramsey et al. 2003; Scopece et al. 2007), and can be abbreviated as:

$$RI_2 = 1 - \frac{H}{C + H}. \quad (2)$$

RI_2 can range from 0 (all heterospecific gene flow) to 0.5 (random mating) to 1 (no heterospecific gene flow) (Fig. 1). While this equation results in a linear relationship between isolation and frequency of hybridization, assigning a nonzero isolation value when mating is random is counter intuitive.

A third general form of the equation used to calculate isolation was originally presented in Stalker's (1942) study of sexual isolation in *Drosophila* with the equation:

$$RI = \frac{\%\ \text{consp. females insem.} - \%\ \text{alien females insem.}}{\%\ \text{consp. females insem.} + \%\ \text{alien females insem.}}.$$

Using the same notation as above, this formula can be abbreviated as:

$$RI_3 = \frac{C - H}{C + H}. \quad (3)$$

This index ranges from -1 (all heterospecific mating) to 0 (random mating) to 1 (no heterospecific mating), which is intuitive; yet the biological meaning of the ratio between the numerator and the denominator is not clear or described. This equation and similar derivations have been adopted by some other workers (e.g. Merrel 1950; Mendelson 2003 Takami et al. 2007). Further modifications of RI_3 have been proposed, especially in studies linking sexual selection and sexual isolation (e.g. Gilbert and Starmer 1985; Rolan-Alvarez and Caballero 2000). However, RI_3 has been used much less frequently than RI_1 (Table 1).

Measures of postzygotic barriers have also primarily used one of these three main base equations, with the relative fitness of heterospecific and conspecific offspring in place of mating frequencies (Table 1). However, in almost all cases, the mathematical relationship between the probability of gene flow and the metric of isolation has been given little if any attention.

A Simple Linear Interpretation

As described above, most agree that metrics of reproductive isolation should be scaled to 0 when no isolation exists, and 1 when isolation is complete, but the manner in which this is achieved can greatly affect the resulting interpretation of the value

Table 1. Alternative methods for calculating reproductive isolation indexes.

Equation	Type of isolating barriers	Algebraic equivalent	Example reference
$RI = 1 - \frac{\text{frequency of heterospecific matings}}{\text{frequency of homospecific matings}}$	Prezygotic (mating preference)	$RI_1 = 1 - \frac{H}{C}$	Coyne and Orr (1989)
$RI = 1 - \frac{\text{number of cross species foraging bouts}}{\text{total number of foraging bouts}}$	Prezygotic (pollinator fidelity)	$RI_2 = 1 - \frac{H}{H+C}$	Ramsey et al. (2003)
$RI = \frac{\%\,\text{conspecific females inseminated} - \%\,\text{alien females inseminated}}{\%\,\text{conspecific females inseminated} + \%\,\text{alien females inseminated}}$	Prezygotic (mating preference)	$RI_3 = \frac{C-H}{C+H}$	Stalker (1942)
$RI = 1 - \frac{\text{observed}/\text{expected heterospecific matings}}{\text{observed}/\text{expected homospecific matings}}$	Prezygotic (flowering time)	$= RI_1$	Martin and Willis (2007)
$RI = 1 - \frac{\text{proportion of hybrid seed}}{(1 - \text{proportion of hybrid seed})}$	Prezygotic (hybridization rate)	$= RI_1$	Brock (2009)
$RI = 1 - \frac{\text{proportion of seeds sired by the minor donor}}{\text{proportion of seeds sired by the major donor}}$	Postpollination (seed production)	$\approx RI_1$	Ruane (2009)
$RI = \frac{\left(1 - \text{proportion fertile hybrid males}/\text{proportion fertile pure males}\right)}{2}$	Postzygotic (hybrid sterility)	$= \frac{1}{2} RI_1$	Bono and Markow (2009)
$RI = \frac{\%\ \text{of successful allopatric pairings} - \%\ \text{of successful sympatric pairings}}{\%\ \text{of successful sympatric pairings}}$	Prezygotic (mating preference)	$= -RI_1$	Hosken et al. (2009)
$RI = 1 - (\%\ \text{of trials where foreign habitat was chosen})$	Prezygotic (habitat isolation)	$= RI_2$	Nosil et al. (2005)
$RI = \frac{\#\ \text{of conspecific spawning events} - \#\ \text{of heterospecific spawning events}}{\#\ \text{of conspecific spawning events} + \#\ \text{of heterospecific spawning events}}$	Prezygotic (mating preference)	$= RI_3$	Mendelson (2003)
$RI = 1 - \frac{2 \times \text{average fitness of the hybrid offspring}}{\text{average fitnesses within the two allopatric populations}}$	Postzygotic (incompatibility)	$= RI_3$	Palmer and Feldman (2009)
$RI = 1 - \frac{\text{prob. mating with heterospecific male}}{\text{prob. mating under random choice}(= 0.5)}$	Prezygotic (mating preference)	$= RI_3$	Takami et al. (2007)

obtained. The most straightforward way to achieve this goal is to use a simple linear equation that describes the relationship between the probability of gene flow and reproductive isolation with the following conditions: (1) if the probability of gene flow is 0, reproductive isolation is 1, (2) if the probability of gene flow is 0.5 (e.g. random mating), reproductive isolation is 0, (3) if the probability of gene flow is 1 (e.g. complete disassortative mating), reproductive isolation is -1. These conditions are satisfied by a line with y-intercept of 1 and slope of -2 (Fig. 1), producing the linear equation:

$$RI_4 = 1 - 2x; \tag{4}$$

where RI_4 is the reproductive isolation index and x is the probability of gene flow, $P(GF)$.

For prezygotic barriers, the probability of gene flow can be estimated by relating the number or frequency of heterospecific matings to all possible matings. For simplicity, we use the terms "heterospecific (H)" and "conspecific (C)" throughout, but it is important to note that H and C could refer to comparisons between species, ecotypes, or populations depending on the application. The probability of gene flow can most easily be expressed as: $P(GF) = \frac{H}{C+H}$, which can be substituted for x in RI_4 to give a general method for calculating reproductive isolation as:

$$RI_{4A} = 1 - 2 \times \left(\frac{H}{H+C}\right). \tag{4A}$$

This equation is algebraically equivalent to forms existing in the literature (see RI_3 above), and can be incorporated into methods that rely on these previous equations (e.g. Carvajal-Rodriguez

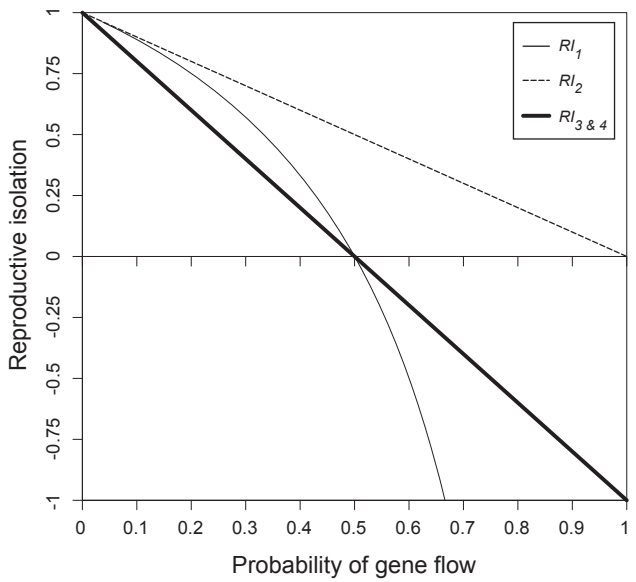

Figure 1. Relationship between the probability of gene flow and the measure of reproductive isolation obtained through the most commonly used method (RI_1), an alternative method (RI_2), and the proposed method (RI_3). RI_1 ranges from 1 at complete isolation, through 0 at random mating (probability of gene flow = 0.5), and to $-\infty$ when gene flow is facilitated. RI_2 ranges from 1 at complete isolation, 0.5 at random mating, and to 0 when gene flow is facilitated. RI_3 ranges from 1 at complete isolation, 0 at random mating, and -1 when gene flow is facilitated, and it is equivalent to the simple linear solution provided as equation RI_4.

and Rolan-Alvarez 2006). However, this simple linear interpretation provides an intuitive connection to the probability of gene flow, which was lacking from previous methods. The following sections illustrate the advantages of using RI_4 to calculate all forms of reproductive isolation, and expansions on this framework will exhibit its strengths in making comparisons between barriers and taxa. Each section consists of a simplified example that illustrates the biological basis of the value obtained using this metric. RI_1 is the most common equation used in the literature for studies of speciation (Table 1), and has formed the basis of additional efforts (e.g. Martin and Willis 2007). Therefore, contrasts will often be made between the values obtained using RI_1 and RI_4. For congruency, all examples begin with the same hypothetical population of 1000 females of species A. Reproductive isolation often exhibits significant asymmetry depending on the direction of comparison (e.g. Wirtz 1999; Tiffin et al. 2001); therefore, a single isolation index is calculated that represents the degree to which this single gender and species is isolated from an alternate species B. For simplicity, each female mates only once in her lifetime, and in the case of postzygotic barriers, she produces a single offspring.

Because barriers to gene flow act sequentially throughout the life cycle of organisms (Ramsey et al. 2003; Schemske 2010), it would be preferable to discuss each barrier in the sequence in which they are experienced in nature. However, the barriers requiring the most explanation are the earliest to act; therefore, examples will build from simple to more complex rather than chronologically. Ultimately, one of the primary goals in calculating the strength of multiple forms of isolation is to reveal the relative contributions of each barrier to total isolation. Therefore, we will conclude with a discussion of combining multiple independent barriers within this linear framework.

DIRECT RELATIONSHIP WITH GENE FLOW

Perhaps the most important advantage of using RI_4 is that isolation values obtained are intuitive because they represent the proportional reduction in gene flow relative to expectations under random mating. In the hypothetical population of 1000 females, a captive mating study using equal abundances of species *A* and *B* reveals that species *A* females have strong mating preferences such that there are 800 conspecific mating events and 200 heterospecific. In such an example, previous researchers applying RI_1 would arrive at a value of reproductive isolation of 0.75. If mating had been random, approximately 500 conspecific matings and 500 heterospecific matings would have been observed. Using RI_{4A}, a value of 0.6 is obtained which represents the proportional decrease in hybridization relative to this null expectation (i.e. there are 60% fewer heterospecific matings than expected by chance) (Fig. 2A). Unlike the value obtained using RI_1, this isolation metric has biological significance, and therefore the impact of any isolation value can be intuitively understood. This example also demonstrates an alternative form of equation RI_{4A}, which can be expressed as:

$$RI_5 = 1 - \left(\frac{\text{observed hybridization}}{\text{expected hybridization}}\right), \qquad (5)$$

which also results in the value 0.6.

It is important to note that in the case of this prezygotic example, we examine how this isolation metric is related to variation in the probability of gene flow via the impact of mating preferences on hybridization. In nature, the formation of hybrids is certainly not synonymous with gene flow, as those F1's must not only be produced, but survive to adulthood and interbreed in order for gene flow to occur. However, to estimate the impact of a single barrier, we examine how it alone affects the potential for gene flow. If all other barriers are assumed to be absent, F1 hybrids will mate randomly with each other and both parents, allowing gene flow to occur at the rate of hybridization. Any nonrandom mating or fitness variation in the F1's represents postzygotic isolation, which would be calculated independently.

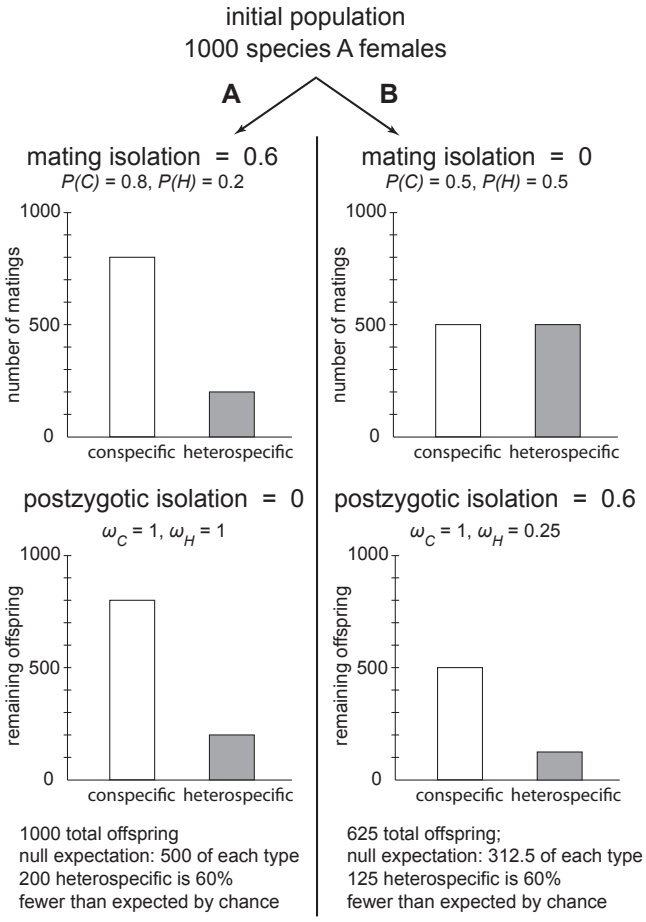

Figure 2. **Equivalency in isolation values obtained using RI_4 for different forms of isolation. A hypothetical population of 1000 females serves as the initial population, and each female mates once giving rise to a single offspring. Two scenarios are presented with an equivalent strength of isolation imposed, and the starting population experiencing mating isolation and postzygotic isolation sequentially. Mating isolation is imposed as variation in con- and heterospecific mating, and intrinsic postzygotic isolation is imposed as differential survival in the resulting offspring. In both cases a reproductive isolation value of 0.6 is used, resulting in an equivalent departure from the null expectation. (A) Mating isolation of 0.6 results in a probability of each mating type: $P(C) = 0.8$ and $P(H) = 0.2$. As a result of these matings, 800 conspecific and 200 heterospecific offspring are produced. Because there is no intrinsic postzygotic isolation, both classes of offspring have equivalent survival ($\omega_C = \omega_H = 1$). Therefore, 800 con- and 200 heterospecific offspring are the end result. Because the null expectation among 1000 total offspring would be 500 of each, the isolation value obtained using $RI_4 = 0.6$ is clearly related by the departure from this null expectation (200 is 60% fewer hybrids than expected by chance). (B) To consider an equivalent strength of postzygotic isolation, the same 1000 females are assumed to have mated randomly (mating isolation = 0), resulting in 500 conspecific and 500 heterospecific offspring. Imposing survival variation such that all conspecific offspring survive ($\omega_C = 1$) and only a quarter of heterospecific offspring survive ($\omega_H = 0.25$), RI_4 again returns a value of isolation of 0.6. As a result 500 conspecific and 125 heterospecific offspring remain after intrinsic postzygotic isolation acts. Among 625 total offspring remaining, the null expectation is 312.5 of each category. Corroborating the isolation value obtained, 125 heterospecific offspring is 60% less than this null expectation.**

LINEARITY AND SYMMETRY FACILITATE COMPARISONS

In addition to being directly applicable to gene flow, another major advantage of using RI_{4A} is that it can accommodate cases of disassortative mating or heterosis. Heterosis is relatively common among recently diverged species (e.g. Taylor et al. 2009), and could serve to facilitate gene flow between diverging populations or taxa. However, the value obtained using equation RI_1 is not proportional to the level of gene flow facilitated by the phenomenon. This is apparent in Fig. 1 as the RI_1 curve approaches negative infinity.

Because it is linear, the equation presented in RI_{4A} does not suffer the same difficulty in relating negative and positive values of reproductive isolation. If the mating preferences used above were reversed such that females of species *A* prefer males of species *B* and mate heterospecifically 80% of the time and

conspecifically only 20%, RI_1 would yield an uninterpretable value of −3. Because this value is impossible to relate to those in the positive range, these data have sometimes been omitted from comparative analysis (e.g. Coyne and Orr 1989; 1997). However, the proposed method using RI_{4A} would instead give a value of −0.6. The negative value indicates that female preference facilitates gene flow rather than confers isolation, and the magnitude denotes that there is 60% more heterospecific mating (i.e. gene flow) than expected by chance. Using an isolation metric that is linear and symmetric around zero allows these data to both be meaningfully compared to other strengths of isolation and to be incorporated into metrics of total isolation.

Because the equation is linear, proportional relationships between isolation metrics are clear (e.g. an isolation value of 0.6 is actually twice as strong as an isolation of 0.3, which is not true using RI_1). Similarly, meaningful average values and variances can be calculated from estimates obtained with RI_{4A} when replicated measurements of isolation metrics are available. Confidence intervals can be estimated for the parameters measured (e.g. heterospecific mating rates), and the lower and upper limits can be used in equation RI_{4A} to calculate the potential range of average reproductive isolation. Alternatively, equivalent values are obtained if confidence intervals are estimated directly from final isolation values. This ease of reporting uncertainty provides an opportunity to make comparisons between studies, and will facilitate future meta-analysis (Morrissey and Hadfield 2012).

$$RI_{4B} = 1 - 2 \times \left(\frac{S \times P(H_{\text{mating}}|S) + U \times P(H_{\text{mating}}|U)}{S \times P(H_{\text{mating}}|S) + U \times P(H_{\text{mating}}|U) + S \times P(C_{\text{mating}}|S) + U \times P(C_{\text{mating}}|U)} \right); \quad (4B)$$

EQUIVALENCY OF PRE- AND POSTZYGOTIC ISOLATION

In the case of postzygotic isolation, the relative fitness of hybrids can be used to calculate the probability of gene flow, again using the general formula: $P(GF) = \frac{H}{C+H}$, but substituting fitness for the offspring of heterospecific (H) and conspecific (C) crosses. This results in an isolation value with an equivalent impact on gene flow, such that values obtained from pre- and postzygotic barriers are directly comparable. Returning to the example of 1000 species A females, the connection to gene flow is apparent when considering a case of differential survival of con- and heterospecific offspring (Fig. 2B). Because we are considering this barrier independently, it will be assumed that the 1000 species A females mate randomly, producing 500 conspecific offspring and 500 heterospecific offspring. In this case, survival variation is imposed that is equivalent to the isolation example above, such that conspecific offspring have a survival rate of 0.8 and heterospecific offspring 0.2. The result is a population of 500 total offspring, 100 of which are heterospecific and 400 of which are conspecific. If these survival rates are used in equation RI_{4A}, an isolation value of 0.6 is calculated. This result can again be validated by comparing the observed number of remaining heterospecific offspring available for hybridization to the expected null value (see equation RI_5 above). Among 500 offspring, a random expectation is that 250 would be conspecific and 250 heterospecific. The observed number of 100 hybrids that remain after survival variation occurs is a 60% reduction from this null expectation, exhibiting the same intuitive connection to gene flow demonstrated above. The same value for reproductive isolation is also obtained when relative fitness is used (i.e. relative fitness of 1 for conspecifics and 0.25 for heterospecifics).

Reproductive Barriers That Affect Cooccurrence

While the simple form of equation RI_4 applies to barriers affecting cooccurrence, calculating the probability of gene flow requires differential conditioning of these probabilities in shared regions of a distribution (S) and unshared regions (U). To calculate the total probability of gene flow, the independent effects of both shared and unshared regions must be combined such that the probability of gene flow is weighted by the proportion of shared and unshared area. The basic structure relating the probability of gene flow is retained, $P(GF) = \frac{H}{C+H}$; however, it is now necessary to add separate terms for shared and unshared time or space:

where S refers to the proportion of time or space that is shared and U represents the proportion unshared. The numerator retains focus on the frequency of heterospecific gene exchange, but it is now the sum of the conditional probabilities of heterospecific mating given the appropriate spatial or temporal context. This version of the equation is unnecessarily complex under many situations, but its utility will be evident in discussion of combining multiple barriers into composites of isolation below. To simplify, when each species is present in equal abundances, random mating will result in both $P(H_{\text{mating}}|S)$ and $P(C_{\text{mating}}|S)$ having equal values of 0.5. Unless considering the impact of immigration, the probability of conspecific mating in allopatry is 1 while the probability of heterospecific mating is 0. Therefore, when considering only one barrier that affects encounter rates, RI_{4B} can ultimately be simplified as:

$$RI_{4C} = 1 - \left(\frac{S}{S+U} \right). \quad (4C)$$

While this equation appears superficially similar to RI_2, it is important to note that both heterospecific and conspecific matings

can occur in sympatry (S), so this term is not equivalent to H. It is necessary to use this form of the equation when there are shared and unshared portions of a distribution in which interactions can occur. These include geographic isolation, microhabitat isolation, temporal isolation (see Appendix A), and some forms of pollinator isolation data, such as lists of shared and nonshared pollinators (e.g. Kay 2006).

GEOGRAPHIC ISOLATION

Geographic isolation affects the cooccurrence of species, and should be estimated using equation RI_{4C}. To illustrate, spatial structure can be added to the example of 1000 species A females used previously. If the geographic extent of species A overlaps species B by 25%, 1/4 of the geographic range of species A is shared with B (S), and 3/4 is unshared (U). In the unshared area, there is no opportunity for gene flow, but in the shared area, mating can occur at random. Under a simplifying assumption that individuals are distributed evenly across their home ranges, 750 conspecific offspring will be produced in the unshared part of the range. In the shared portion of the range, random mating produces 125 species A conspecific offspring and 125 *A/B* heterospecific hybrids. In total 125 of the 1000 total offspring are *A/B* hybrids, representing a 75% reduction in heterospecific offspring relative to a null expectation of 500. Therefore, isolation is equal to 0.75, which is the value obtained when using RI_{4C}. The assumption that all individuals will be evenly distributed across the range of a species will often be violated (Brown 1984), and it will therefore sometimes be necessary to incorporate this increased complexity (see temporal isolation example in Appendix A).

In the above example, if the occupation of different geographic ranges has been demonstrated to result from intrinsic biological differences between taxa, such as through reciprocal translocation studies, this is considered *ecogeographic isolation*. However, taxa may also occupy different ranges for purely historical reasons, which also impacts gene flow. Because it does not arise from intrinsic differences between taxa, this form of geographic isolation is not considered an isolating mechanism under the BSC, and the value calculated is referred to as *effective geographic isolation* (see Sobel et al. 2010 for discussion). Both of these forms of isolation can act simultaneously, and separating their individual impacts on gene flow will require the measurement of geographically based fitness variation across the ranges of both taxa considered.

DISPERSAL AND IMMIGRANT INVIABILITY

Another instance for which it is necessary to consider the effect of encounter rate on reproductive isolation occurs when considering the strength of immigrant inviability as an isolating barrier (Nosil et al. 2005). In these cases, instead of making the simplifying assumption that gene flow does not occur in the unshared portion of the range (U), the term $U \times P(H_{\text{mating}}|U)$ can be retained. The effects of dispersal rate can be considered by using the proportion of individuals in the population that are immigrants, m, to adjust the probability of heterospecific mating in allopatry. To incorporate potential immigrant inviability, a fitness term, $\omega_{\text{immigrants}}$, must also be multiplied. The total probability of heterospecific mating in allopatric regions is therefore: $P(H_{\text{mating}}|U) \times m \times \omega_{\text{immigrants}}$, and the probability of conspecific mating is $P(C_{\text{mating}}|U) \times (1 - m \times \omega_{\text{immigrants}})$. These terms can be substituted for $P(H_{\text{mating}}|U)$ and $P(C_{\text{mating}}|U)$, respectively, in equation RI_{4B} to take this barrier into account.

Estimating Total Reproductive Isolation

An important goal in speciation studies is to assess the relative contribution of each barrier to the total isolation experienced between taxa (Ramsey et al. 2003; Schemske 2010; Sobel et al. 2010). The most widely used method for doing so was developed by Coyne and Orr (1989) and involves discounting late acting barriers by barriers that have already acted. The simple mathematical expression proposed by Coyne and Orr (1989) was:

$$RI_{\text{total}} = \text{pre} + (1 - \text{pre}) \times \text{post};$$

where the total reproductive isolation (RI_{total}) experienced is the linear sequential sum of prezygotic (pre) and postzygotic (post) isolation. This approach has been expanded to include any number of barriers measured (Ramsey et al. 2003), which has been very useful in revealing which forms of isolation have contributed most to speciation (e.g. Ramsey et al. 2003; Dopman et al. 2010; Schemske 2010). However, this straightforward formulation is accurate when combining certain combinations of barriers, but inaccurate when combining others. To illustrate, we provide two examples below where a barrier affecting cooccurrence is combined with an additional prezygotic or postzygotic form of isolation. For simplicity, the same individual strengths of isolation will be used as in the independent assessment of individual barriers presented above. In both cases, the barrier affecting cooccurrence will be ecogeographic isolation, with an individual isolation value of 0.75. We will contrast the effect of adding mating isolation versus intrinsic postzygotic isolation, both with an individual strength of 0.6. Using the framework above, either of these combined examples would result in total isolation of 0.9. The example below will show that this estimate is correct when combining ecogeographic isolation and mating isolation, but not when adding intrinsic postzygotic isolation. Finally, while incorporating these issues, a general formula for estimating both total isolation and relative contributions of barriers will be given.

COMBINING ECOGEOGRAPHIC AND MATING ISOLATION

The above derivation of equation RI_{4C} demonstrates the issues in combining different barriers into a single measure of total isolation. When both shared (S) and unshared (U) regions of space or time exist, additional barriers that contribute to total isolation can only be manifested in the shared area. In simplifying equation RI_{4B} to RI_{4C}, random mating expectations were substituted for the probability of heterospecific and conspecific mating in shared areas; $P(H_{\text{mating}}|S)$ and $P(C_{\text{mating}}|S)$, respectively. These terms can instead be retained, with observed mating frequencies used in place of random expectations. Using a strength of ecogeographic isolation of 0.75 gives U of 0.75 and S of 0.25, and a mating isolation value of 0.6 means that $P(H_{\text{mating}}|S)$ and $P(C_{\text{mating}}|S)$ are equal to 0.2 and 0.8, respectively. Substituting these values for variables in equation RI_{4B} returns a total isolation strength of 0.9, as would be found using the previous method (Fig. 3A).

COMBINING ECOGEOGRAPHIC AND INTRINSIC POSTZYGOTIC ISOLATION

When combining ecogeographic isolation with an equivalent strength of intrinsic postzygotic isolation, the result is slightly different. As above, the shared part of the geographic space is $S = 0.25$ and the unshared is $U = 0.75$, and there are again 1000 evenly dispersed single-offspring producing females. If ecogeographic and intrinsic postzygotic isolation are the only barriers acting, random mating is assumed in the region of contact. This results in 125 conspecific and 125 heterospecific offspring (Fig. 3B). Adding the 750 conspecific offspring from the unshared geographic region results in a total of 875 conspecific and 125 heterospecific offspring. If we impose an intrinsic postzygotic isolation of 0.6 in the form of survival variation, then 80% of conspecific offspring survive and only 20% of heterospecifics. The result is a total of 700 conspecific and 25 heterospecific offspring after this barrier acts. With a total of 725 individuals in the combined population, the null expectation is that 362.5 would be conspecific with an equal number of hybrids. The 25 heterospecifics observed in the example is a 93.1% reduction in gene flow; therefore, a total isolation value of 0.931 is biologically correct.

While the previous method fails to produce this value, it is possible to arrive at the appropriate estimate by returning to the RI_{4B} equation presented above. In calculating the total probability of gene flow, it is again necessary to include separate terms for shared and unshared regions, with the addition of terms for the fitness of con- $P(C_{\text{post}})$ and heterospecific offspring, $P(H_{\text{post}})$:

$$RI_{4D} = 1 - 2 \times \left(\frac{S \times P(H_{\text{mating}}|S) \times P(H_{\text{post}}|S)}{S \times P(H_{\text{mating}}|S) \times P(H_{\text{post}}|S) + S \times P(C_{\text{mating}}|S) \times P(C_{\text{post}}|S) + U \times P(C_{\text{mating}}|U) \times P(C_{\text{post}}|U)} \right). \quad (4D)$$

The primary difference between these two is related to the $U \times P(C \mid U)$ term in the denominator. In the case of combining ecogeographic with mating isolation, the probability of mating conspecifically in allopatry is 1, $P(C_{\text{mating}}|U) = 1$; therefore $U \times P(C_{\text{mating}}|U)$ was simplified to U. However, in the case of postzygotic isolation, conspecific offspring in allopatry and sympatry are assumed not to vary in relative fitness. Therefore, the term $P(C_{\text{post}}|U)$ is retained, and its value is equal to $P(C_{\text{post}}|S)$ (Fig. 3C). Using this equation, the correct total isolation value of 0.931 is obtained. Once again, the same value would be found using relative rather than absolute fitness measurements (e.g. $P(C_{\text{post}}) = 1$ and $P(H_{\text{post}}) = 0.25$).

A GENERAL FORMULA FOR TOTAL ISOLATION

It is relatively intuitive to combine measures of all three types of barriers presented above using RI_{4D}; however, in any given pair of taxa, multiple barriers that affect cooccurrence, prezygotic isolation, and/or postzygotic isolation can be acting. Again, the relative effects of shared (S) and unshared (U) space or time must be considered first. For any number of barriers affecting cooccurrence, the proportion of shared area for each individual barrier is multiplied to give the total space and time over which two taxa cooccur:

$$S_{\text{total}} = \prod_{i}^{n} S_i;$$

where the total shared area S_{total} is the product of each individual shared area S_i for n barriers affecting cooccurrence. The total area of unshared space U_{total} and/or time as the remaining area as:

$$U_{\text{total}} = 1 - \prod_{i}^{n} S_i.$$

This can be can be done for any number of barriers affecting cooccurrence, for example allowing the simultaneous consideration of ecogeographic isolation, temporal isolation, and microhabitat isolation. These terms can be incorporated into the basic framework presented above as:

$$RI_{4E} = 1 - 2 \times \left(\frac{S_{\text{total}} \times P(H|S) + U_{\text{total}} \times P(H|U)}{S_{\text{total}} \times P(H|S) + U_{\text{total}} \times P(H|U) + S_{\text{total}} \times P(C|S) + U_{\text{total}} \times P(C|U)} \right). \quad (4E)$$

Additional forms of reproductive isolation can be included, as the terms $P(H\,|S)$, $P(C\,|S)$, and $P(C\,|U)$ can also be considered multiplicative products of as many forms of isolation as necessary.

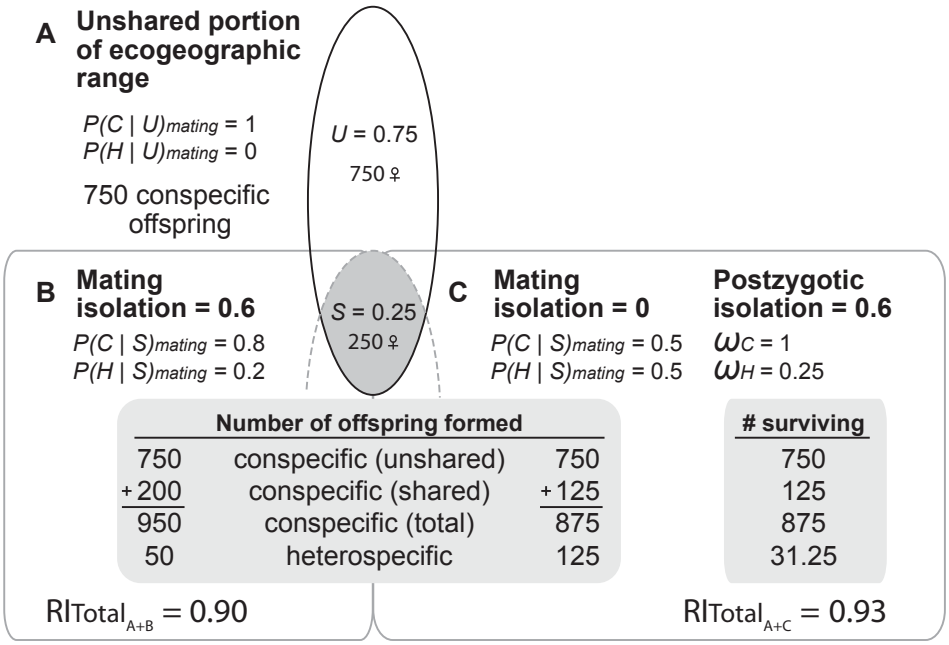

Figure 3. **Illustration of the nonequivalent effects of pre- and postmating barriers when combined with ecogeographic isolation. The ecogeographic ranges of a hypothetical species is shown with an oval, with the upper portion unshared with an alternate species and the lower portion shared. The focal species exhibits 75% of its range as unshared ($U = 0.75$) with the alternate species, and 25% shared ($S = 0.25$). The following examples will illustrate the different values obtained when adding ecogeographic isolation with either a prezygotic (mating isolation) or postzygotic barrier (survival variation). For simplicity, the starting population of the focal species is assumed to consist of 1000 evenly distributed females that all mate once and produce one offspring. (A) In the unshared portion of the ecogeographic range, there is no opportunity for heterospecific mating, $P(H \mid U) = 0$; therefore all matings are conspecific, $P(C \mid U) = 1$. This results in 750 conspecific offspring that will be added to those formed in the shared region. (B) *Ecogeographic isolation + mating isolation*. In the shared region (shaded), applying a mating isolation value of 0.6 as in the previous example results in a frequency of heterospecific matings, $P(H_{mating} \mid S) = 0.2$ and frequency of conspecific matings, $P(C_{mating} \mid S) = 0.8$. Therefore, with 250 females in the shared region, 200 conspecific and 50 heterospecific offspring are produced. Adding the 750 conspecific offspring from the unshared region gives a total of 950 con- and 50 heterospecific offspring, a 90% reduction in gene flow. Using equation RI_{4B}, an isolation value of 0.9 is indeed obtained. (C) *Ecogeographic isolation + postzygotic isolation*. Similarly to above examples, to impose a postzygotic barrier, it is necessary to pass through random mating in the shared region to form hybrids first. Therefore, the 250 females in sympatry produce 125 con- and 125 heterospecific offspring. Once again, a postzygotic barrier in the form of survival variation is imposed, such that all the conspecific offspring and 0.25 of the heterospecific offspring survive. Adding the 750 offspring from the unshared region, 875 total conspecific and 31.25 heterospecific offspring remain after this survival rate variation is imposed. This represents a 93.1% departure from the null expectation, and RI_{4C} provides this estimate.**

It is also possible to adjust this equation to accommodate studies of isolation that are conducted only shared time and space (see Appendix B).

To assess the relative contribution of individual barriers to total isolation, each barrier under consideration must first be assigned a position within the linear sequence of isolation. Barriers that act simultaneously should be given equivalent ranks within the sequential series, or combined multiplicatively into a single estimate. Total isolation is first calculated using equation RI_{4E}. To calculate the absolute contribution of a barrier (as in Ramsey et al. 2003), the combined isolation including the focal barrier and all preceding barriers is calculated using RI_{4E}. The calculation is then repeated, including all preceding barriers, but excluding the focal. The latter is subtracted from the former to reveal the absolute contribution of any individual barrier:

$$AC_i = RI_{[1,i]} - RI_{[1,i-1]}$$

where $RI_{[1,i]}$ denotes the combined isolation calculated by RI_{4E} including all barriers from the first to act (1) through the focal barrier (i), and $RI_{[1,i\text{-}1]}$ denotes the same calculation omitting the focal barrier. The relative contribution of any barrier (RC_i) is simply this absolute contribution (AC_i) divided by the total isolation calculated when including all barriers (following Ramsey et al. 2003):

$$RC_i = \frac{AC_i}{RI_{\text{total}}}$$

Table 2. Reassessment of data from Ramsey et al. (2003) and Matsubayashi and Katakura (2009).

A. *Mimulus cardinalis* and *M. lewisii* (Ramsey et al. 2003).

	Original component of reproductive isolation		Reassessed component of reproductive isolation	
Isolating barriers	*M. lewisii*	*M. cardinalis*	*M. lewisii*	*M. cardinalis*
Ecogeographic isolation	0.587	0.587	0.794	0.794
Pollinator isolation	0.976	0.976	0.952	0.952
Pollen precedence	0.708	0.958	0.548	0.919
F1 seed germination	0.203	0.047	0.113	0.024
F1 survivorship	0	0	0	0
F1 percent flowering	0	0	0	0
F1 biomass	−1.393	0.056	−0.410	0.029
F1 pollen viability	0.662	0.628	0.495	0.458
F1 seed mass	0.409	0.737	0.257	0.584
Total isolation	0.99744	0.99988	0.99882	0.99996

B. *Henosepilachna vigintioctomaculata and H. pustulosa* (Matsubayashi and Katakura 2009).

	Original component of reproductive isolation		Reassessed component of reproductive isolation	
Isolating barriers	*H. vigintioctomaculata*	*H. pustulosa*	*H. vigintioctomaculata*	*H. pustulosa*
Seasonal isolation	0.09333	0.02361	0.04895	0.01195
Habitat isolation	0.84358	0.87754	0.72948	0.78181
Sexual isolation	0.86667	0.77273	0.54545	0.73333
Egg hatchability	0.68875	0.98380	0.52525	0.96812
Conspecific sperm precedence[1]	0.00000 – 0.99172	0.00000 – 0.99666	0.00000 – 0.98358	0.00000 – 0.99334
F1 reduced fitness	0.04554	−0.20980	0.02330	−0.09494
Total isolation[2]	0.98701−0.99989	0.99927−1.00000	0.96377−0.99969	0.99884 – 1.00000

[1]Sperm precedence was reported as a range of potential values due to the unknown probability of conspecific rescue.

[2]Total isolation was calculated using low and high values for conspecific sperm precedence, resulting in a range of total isolation.

Additional Considerations

For the examples presented, isolation has been experienced by a single sex within a single species. To accurately combine estimates of isolation across multiple barriers, sex-specific estimates must be maintained throughout the barriers under investigation (see Appendix C for discussion). In addition, the presentation of RI_4 and its derivatives has been simplified by assuming equal relative abundance of individuals and gametes for the species under consideration. As pointed out by Martin and Willis (2007), this assumption may often be violated, requiring an adjustment to null expectations. This can be accomplished within the framework of the equations presented by generating unequal expected values for each species considered (see Appendix D).

Reassessment of Previous Data

The magnitude of difference between the proposed method of RI_4 compared to previously employed methods varies over the range of possible isolation values. For example, Fig. 1 shows that RI_4 will be the most similar to RI_1 when isolation is either very strong or very weak. However, these two metrics will be most different at intermediate values and especially for negative values. Table 2 provides a comparison between RI_4 and previous methods using published data on the strength of isolation between species of *Mimulus* wildflowers (Ramsey et al. 2003) and ladybird beetles in the genus *Henosepilachna* (Matsubayashi and Katakura 2009). Over the positive range of isolation, RI_4 results in values that are lower in magnitude than previously estimated, with the exception of estimates that are exactly 0 or 1 (see Fig. 1). In the negative range, the nonlinear equations based on RI_1 result in a disproportionate effect of the calculated values on combined estimates of total isolation, especially for values below −1. Therefore, in the examples presented, individual estimates of isolation using the original and reassessed calculations vary considerably, while total isolation appears superficially similar (Table 2). However, regardless of the magnitude of the difference between estimates, the most significant advantage of RI_4 is that for all possible strengths

of isolation, there is an intuitive connection between the calculated metrics and gene flow.

Summary

One of the ultimate goals of speciation research is to uncover which forms of reproductive isolation impact the process most. To accomplish this objective, it is important to examine the relative contributions of each barrier to total reproductive isolation at the time of speciation. Estimating the strength of reproductive isolation is therefore a critical step toward understanding the origin of species; however there is currently no consensus on a mathematical framework for doing so. We here provide a linear interpretation of previous estimation methods with many advantages for measuring the strength of individual barriers. By deriving a metric that is directly related to gene flow, this unifying method provides a means for making comparisons of isolation within and between taxa that will aid greatly in answering fundamental questions about the nature of speciation.

ACKNOWLEDGMENTS

This work benefited greatly from discussion and comments from D. Bolnick, J. Conner, K. Gross, K. Kay, R. Lenski, A. Prather, D. Lowry, A. Orr, J. Willis, and J. Yost, and two anonymous reviewers. Additional thanks are extended to all members of the Kay lab for comments on a previous version of this manuscript. D. Schemske provided guidance and feedback throughout all parts of the project and manuscript preparation. JMS gratefully acknowledges support from a Doctoral Dissertation Improvement Grant from the National Science Foundation, DEB-0808447. JMS and GFC also thank the Plant Biology Department, the Graduate School, and the College of Natural Sciences at Michigan State University for additional support. The authors declare that there are no conflicts of interest.

LITERATURE CITED

Bolnick, D. I., and T. J. Near. 2005. Tempo of hybrid inviability in centrarchid fishes (Teleostei: Centrarchidae). Evolution 59:1754–1767.

Bono, J. M., and T. A. Markow. 2009. Post-zygotic isolation in cactophilic *Drosophila*: larval viability and adult life-history traits of *D. mojavensis/D. arizonae* hybrids. J. Evol. Biol. 22:1387–1395.

Brock, M. T. 2009. Prezygotic barriers to gene flow between *Taraxacum ceratophorum* and the invasive *Taraxacum officinale* (Asteraceae). Oecologia 161:241–251.

Brown, J. H. 1984. On the relationship between abundance and distribution of species. Am. Nat. 124:255–279.

Carvajal-Rodriguez, A., and E. Rolan-Alvarez. 2006. JMATING: a software for the analysis of sexual selection and sexual isolation effects from mating frequency data. BMC Evol. Biol. 6:40.

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

———. 1997. "Patterns of speciation in *Drosophila*" revisited. Evolution 51:295–303.

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

Dobzhansky, T. 1937. Genetics and the origin of species. Columbia Univ. Press, New York.

Dopman, E. B., P. S. Robbins, and A. Seaman. 2010. Components of reproductive isolation between North American pheromone strains of the European corn borer. Evolution 64:881–902.

Ehrman, L. 1965. Direct observation of sexual isolation between allopatric and between sympatric strains of the different *Drosophila paulistorum* Races. Evolution 19:459–464.

Funk, D. J., P. Nosil, and W. J. Etges. 2006. Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. Proc. Natl. Acad. Sci. USA 103:3209–3213.

Gilbert, D. G., and W. T. Starmer. 1985. Statistics of sexual isolation. Evolution 39:1380–1383.

Gleason, J. M., and M. G. Ritchie. 1998. Evolution of courtship song and reproductive isolation in the *Drosophila willistoni* species complex: do sexual signals diverge the most quickly? Evolution 52:1493–1500.

Hosken, D. J., O. Y. Martin, S. Wigby, T. Chapman, and D. J. Hodgson. 2009. Sexual conflict and reproductive isolation in flies. Biol. Lett. 5:697–699.

Jennings, J. H., D. Mazzi, M. G. Ritchie, and A. Hoikkala. 2011. Sexual and postmating reproductive isolation between allopatric *Drosophila montana* populations suggest speciation potential. BMC Evol. Biol. 11:68.

Kay, K. M. 2006. Reproductive isolation between two closely related hummingbird-pollinated Neotropical gingers. Evolution 60:538–552.

Lowry, D. B., J. L. Modliszewski, K. M. Wright, C. A. Wu, and J. H. Willis. 2008. The strength and genetic basis of reproductive isolating barriers in flowering plants. Philos. Trans. R Soc. B Biol. Sci. 363:3009–3021.

Martin, N. H., and J. H. Willis. 2007. Ecological divergence associated with mating system causes nearly complete reproductive isolation between sympatric *Mimulus* species. Evolution 61:68–82.

Matsubayashi, K. W., and H. Katakura. 2009. Contribution of multiple isolating barriers to reproductive isolation between a pair of phytophagous ladybird beetles. Evolution 63:2563–2580.

Mayr, E. 1942. Systematics and the origin of species. Columbia Univ. Press, New York.

———. 1963. Animal species and evolution. Harvard Univ. Press, Cambridge, MA.

Mendelson, T. C. 2003. Sexual isolation evolves faster than hybrid inviability in a diverse and sexually dimorphic genus of fish (Percidae: *Etheostoma*). Evolution 57:317–327.

Merrel, D. J. 1950. Measurement of sexual isolation and selective mating. Evolution 4:326–331.

Morrissey, M. B., and J. D. Hadfield. 2012. Directional selection in temporally replicated studies is remarkably consistent. Evolution 66:435–442.

Moyle, L. C., M. S. Olson, and P. Tiffin. 2004. Patterns of reproductive isolation in three angiosperm genera. Evolution 58:1195–1208.

Nosil, P., T. H. Vines, and D. J. Funk. 2005. Perspective: reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution 59:705–719.

Palmer, M. E., and M. W. Feldman. 2009. Dynamics of hybrid incompatibility in gene networks in a constant environment. Evolution 63:418–431.

Poulton, E. G. 1908. What is a species? Pp. 46–94 *in* E. G. Poulton, ed. Essays on Evolution 1889–1907. Clarendon Press, Oxford, U.K.

Ramsey, J., H. D. Bradshaw, and D. W. Schemske. 2003. Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). Evolution 57:1520–1534.

Rice, W. R., and E. E. Hostert. 1993. Laboratory experiments on speciation: what have we learned in 40 years? Evolution 47:1637–1653.

Ruane, L. G. 2009. Post-pollination processes and non-random mating among compatible mates. Evol. Ecol. Res. 11:1031–1051.

Schemske, D. W. 2000. Understanding the origin of species. Evolution 54:1069–1073.

———. 2010. Adaptation and the origin of species. Am. Nat. 176:S4–S25.
Scopece, G., A. Musacchio, A. Widmer, and S. Cozzolino. 2007. Patterns of reproductive isolation evolution in Mediterranean deceptive orchids. Evolution 61:2623–2642.
Sobel, J. M., G. F. Chen, L. R. Watt, and D. W. Schemske. 2010. The biology of speciation. Evolution 64:295–315.
Stalker, H. D. 1942. Sexual isolation studies in the species complex *Drosophila virilis*. Genetics 27:238–257.
Takami, Y., N. Nagata, M. Sasabe, and T. Sota. 2007. Asymmetry in reproductive isolation and its effect on directional mitochondrial introgression in the parapatric ground beetles *Carabus yamato* and *C. albrechti.* Popul. Ecol. 49:337–346.
Taylor, S. J., M. Arnold, and N. H. Martin. 2009. The genetic architecture of reproductive isolation in *Louisiana irises*: hybrid fitness in nature. Evolution 63:2581–2594.
Tiffin, P., M. S. Olson, and L. C. Moyle. 2001. Asymmetrical crossing barriers in angiosperms. Proc. Biol. Sci. 268:861–867.
Vickery, R. K. 1959. Barriers to gene exchange within *Mimulus guttatus* (Scrophulariaceae). Evolution 13:300–310.
Wirtz, P. 1999. Mother species-father species: unidirectional hybridization in animals with female choice. Anim. Behav. 58:1–12.

Associate Editor: T. Lenormand

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Figure S1. A simple example dataset illustrates the use of equations RI_{4S1} and RI_{4S2} for calculating temporal isolation when different types of data can be collected.
Figure S2. The effect of changes in expected null frequency of hybrid formation on the value of reproductive isolation using equation RI_{4G}.
Online Appendix A: *Temporal Isolation*
Online Appendix B: Combining multiple sympatric barriers
Online Appendix C: Combining sexes with multiple barriers
Online Appendix D: Accommodation of Unequal Null Expectations