Reinforcement during ecological speciation

Mark Kirkpatrick

Section of Integrative Biology C0930, University of Texas, Austin, TX 78712, USA (kirkp@mail.utexas.edu)

Reinforcement of pre-zygotic isolation can result when any of several kinds of selection act against hybrids. This paper investigates the situation where hybrids are selected against for ecological reasons, for example when there is no niche for individuals that are phenotypically intermediate between the parental species. The calculations here show how much ecological selection can lead to the reinforcement of a female mating preference or an assortative mating trait that is expressed in both sexes. The model allows for the ecological trait to be affected by any number of loci, but assumes that selection is weak and the introgression rate small. The effect of selection against hybrids increases rapidly as the difference between the mean phenotypes of the two populations increases. When genetic variation in the ecological trait is caused by many loci, stabilizing selection on it further contributes to reinforcement.

Keywords: hybridization; sexual selection; pre-zygotic isolation; model

1. INTRODUCTION

Reinforcement is the strengthening of pre-zygotic isolation caused by selection against hybrids (Dobzhansky 1940). An unanswered question is how the mechanism of hybrid unfitness might affect the outcome of reinforcement. In some organisms, hybrids have low fitness because interactions between alleles from the parental species reduce fertility or cause problems in early development (Dobzhansky 1937; Muller 1942). In others, inter-specific hybrids are physiologically viable and fertile, but they have low fitness because they are poorly matched to local ecological conditions (Schluter 1998). For example, sometimes there simply is no niche for individuals that are phenotypically intermediate between the two hybridizing species.

Stickleback fishes living in small glacial lakes in western Canada illustrate this latter situation. There are two ecological forms, namely a benthic species and a limnetic species, that are adapted to foraging near the shore and in open water, respectively. Hybrids, which are morphologically intermediate, grow more slowly than the benthic morph in near-shore habitats and also than the limnetic morph in open water (Hatfield & Schluter 1999). Analogous forms of selection against hybrids that are phenotypically and ecologically intermediate between their parental species may be common in many groups of animals and plants (Schluter 1998).

This paper develops a population genetic model for the purpose of quantifying how much this kind of ecological selection against hybrids will enhance reproductive isolation through reinforcement. The main aim is to determine what features of the ecological landscape and the underlying genetics affect the outcome. To do that, I will consider reinforcement in a population that is hybridizing with migrants that arrive from elsewhere. The model also applies to two sympatric populations when there is one-way hybridization. The motivation for focusing on these situations is that they are easier to model than those involving two reciprocally hybridizing species.

The next section begins by outlining the genetic assumptions for a class of models of reinforcement. It then

briefly reviews earlier theoretical results for the amount of reinforcement expected under two types of pre-zygotic isolation, female mating preferences and assortative mating. Those results in turn depend on the intensity of hybrid incompatibility and set the stage for this paper, which finds expressions for the intensity of hybrid incompatibility when hybrid unfitness is caused by selection against ecological intermediates.

2. THE MODEL

The genetic model is of an organism with a haploid genome. There can be any number of loci contributing genetic variation to the ecological trait and to the characters involved in pre-zygotic isolation (for example, a female mating preference). Two alleles segregate at each of these loci and the phenotypic effect of replacing one allele by the other at locus j is denoted γ_j . No assumption is made about the distribution of effects that these loci have or about the recombination rates between them. The means of the ecological trait at birth in the focal population and in the migrants are given as $\bar{E}$ and $\bar{E}'$, respectively. The phenotypic variance of the ecological trait, which may include a random environmental component, is σ_E^2 .

The population receives migrants at a rate m . If the model is being applied to two species in sympatry, then m is the rate at which individuals of the other species hybridize with the focal species and all their hybrid descendants are assumed to mate with the focal species only. The model applies equally to species that are in secondary contact and to the last stages of sympatric speciation. I will assume that migration occurs before selection in each generation, but the results can be extended to the reverse sequence of events (Kirkpatrick & Servedio 1999).

The approach that I follow here uses the 'quasi-linkage equilibrium' approximations developed by Barton & Turelli (1991). The key assumption for the results that follow is that the linkage disequilibria are small relative to their maximum possible values. That implies that migration is weak and that the selection coefficients for

individual alleles and groups of alleles are small (see Kirkpatrick & Servedio 1999). However, qualitatively different results can emerge if individual genes have very large effects on fitness (Kirkpatrick 2000).

(a) *Reinforcement under two kinds of pre-zygotic isolation*

This section briefly reviews previous results on the amount of reinforcement that occurs in two situations, namely when pre-zygotic isolation is based on a female mating preference and when it depends on an assortative mating trait.

Kirkpatrick & Servedio (1999) studied the case of a mating preference. Genetic variation in the preference and the male trait it acts on is assumed to be caused by any number of genes with additive effects. The model applies to any kind of female preference, which is simply viewed as a trait that in some way biases a female's choice of mate. The mating preference genes are assumed to be free of direct natural selection. The general result for the amount of reinforcement at equilibrium is quite complicated, but it simplifies greatly when stabilizing selection on the male trait is weak and the preference loci are unlinked to loci causing hybrid unfitness.

Under those conditions, at equilibrium the amount by which the average preference in the population will differ from that in the migrants is

$$D_P \approx 0.5\rho h_P^2 \left(1 + I + 2m \frac{D_T}{h_T^2}\right) D_T, \quad (1)$$

when $m > 0$. The new quantities appearing here are ρ , the phenotypic correlation between the female preference and the male trait among mating pairs, h_P^2 , the heritability of the preference, h_T^2 , the heritability of the male trait, D_T , the difference between the average male trait in the two populations and I , the intensity of selection against hybrid incompatibility (which is described further below). The correlation ρ can be calculated for any kind of preference; Hall *et al.* (2000) calculate ρ for three specific preference functions. The quantities D_P and D_T are measured in units of their phenotypic standard deviations, for example $D_T = (\bar{T} - \bar{T}')/\sigma_T$, where $\bar{T}$ and $\bar{T}'$ are the means of the male trait at birth among the residents and migrants, respectively and σ_T is the phenotypic standard deviation of the male trait in the focal population. Equation (1) slightly generalizes Kirkpatrick & Servedio's (1999) equation (22) in order to include the migration term, which they assumed to be negligible.

A second kind of pre-mating isolation is caused by assortative mating. In this case, there is a single trait expressed by both males and females that influences their probability of mating. Flowering time in plants is a good example: early-flowering plants fertilize other early-flowering individuals more frequently than they fertilize late-flowering individuals. Kirkpatrick (2000) studied reinforcement under assortative mating. The model assumes that there is some pre-existing mechanism that causes a correlation between the phenotypes of mating individuals for a trait influenced by additive genes. Reinforcement occurs when the mean values of the assortment trait diverge in the two populations.

At equilibrium, the difference between the means of the assortment trait in the residents and the migrants is

$$D_A \approx \frac{2h_A^2 \tilde{\Gamma}_A \Delta_A}{2h_A^2 \tilde{\Gamma}_A - (1 - I)m}, \quad (2)$$

when $m > 0$ (see equations (3) and (4) in Kirkpatrick 2000). Here, h_A^2 is the heritability of the assortment trait and Δ_A is the number of phenotypic standard deviations between the local optimum for the assortment trait and its mean value among migrants: $\Delta_A = (\theta_A - \bar{A}')/\sigma_A$. The stabilizing selection gradient $\tilde{\Gamma}_A$, which is negative, measures the intensity of selection favouring θ_A , the assortment trait optimum (the value of the assortment trait that maximizes lifetime fitness). This gradient is defined as the regression of relative fitness onto the squared deviation of an individual's phenotype from the population mean (Lande & Arnold 1983). It has been standardized here (made non-dimensional) by multiplying it by the assortment trait's phenotypic variance σ_A^2 . As with the previous result for a mating preference, equation (2) has been simplified from a more general expression by assuming that stabilizing selection is weak and that the genes for the assortment trait are unlinked to those causing post-zygotic isolation (see Appendix A in Kirkpatrick 2000).

In cases of both a female mating preference (equation (1)) and an assortment trait (equation (2)), reinforcement is enhanced by larger values of hybrid incompatibility, which is represented by I . What exactly is that quantity? It is the effect of selection on all genes other than those affecting pre-zygotic isolation (e.g. the preference and male trait or the assortment trait) on reinforcement. Quantitatively, it is defined as

$$I \approx - \sum_{U \subseteq H} \frac{s_U}{r_{iU}} \prod_{j \in U} (p'_j - p_j) \quad (3)$$

(see equation (17) in Kirkpatrick & Servedio 1999). The summation is taken over all loci and sets of loci U in the genome H that affect hybrid fitness, exclusive of those involved in pre-zygotic isolation. The selection coefficient s_U , which represents the force of selection acting on the loci in set U , can be calculated for any form of selection as described by Barton & Turelli (1991) (see also Appendix B in Kirkpatrick & Servedio 1999). The quantities p'_j and p_j are the allele frequencies at locus j among migrants and in the local population, respectively. The recombination rate r_{iU} is the probability that recombination breaks apart the loci in the set that includes U and i , where i represents any of the loci responsible for the mating preference or the assortment trait and U represents a set of one or more of the loci affecting hybrid incompatibility. (For example, if U is the single locus j , then r_{iU} is the standard two-locus recombination rate r_{ij} between loci i and j .) Equation (3) is based on an approximation that assumes that the focal population and migrants are in quasi-linkage equilibrium (see Appendix C in Kirkpatrick & Servedio 1999).

The goal of this paper is finally in view: it is to determine the value of I generated by selection on an ecological trait. To do that, I will assume that genes affecting the ecological trait have additive effects and that they are unlinked to each other and to the loci causing pre-zygotic

isolation. Under these assumptions, the selection coefficients and recombination rates that appear in equation (3) have simple forms. However, the summations and the expressions for the allele frequencies can still be fairly complicated. I will therefore focus on two extreme cases that simplify the results even further: when genetic variation in the ecological trait is caused by a single locus and when it is caused by a large number of loci of approximately equal effect. The results are expressed first in terms of the phenotypic differences between the hybridizing populations and then in terms of the differences between their ecological niches.

(b) The intensity of hybrid incompatibility in terms of phenotypic differences

Consider the situation where a single locus underlies the ecological trait. The locus is fixed for one allele among the immigrants, but a different allele has a relative fitness advantage s in the local population. Haldane (1930) showed that the locally favoured allele will reach an equilibrium frequency of $1 - m/s$. For polymorphism to be maintained, m must be smaller than s or the locally adapted allele will be lost by the swamping effect of gene flow from the migrants. (In the present model, selection on other loci will cause deviations from that value. However, those deviations will be second order under the earlier assumptions that selection and migration are weak and so they can be neglected.) Putting together those facts and equation (3) shows that the intensity of selection against hybrid incompatibility is

$$I = 2(s - m). \quad (4)$$

This simple result carries the message that reinforcement increases as a linear function of the strength of selection favouring the allele for the ecological trait that is adapted to the local conditions. On the other hand, increasing the migration rate decreases the intensity of selection against hybrid incompatibility. This is because more migration causes the two populations to become more genetically similar, thereby decreasing the opportunity for selection to distinguish between residents and hybrids.

Equation (4) assumes that there is free recombination between the ecological trait locus and the loci affecting pre-zygotic isolation (which implies that r_{iU} in equation (3) equals 0.5). However, the effect of other recombination rates can be easily found (see Kirkpatrick & Servedio 1999, p. 873). Tighter linkage between the ecological trait locus and a locus that affects pre-zygotic isolation enhances reinforcement.

The intensity of hybrid incompatibility I can also be expressed in a form that relates more directly to phenotypic data. The selection coefficient s in equation (4) can be rewritten as $s \approx \gamma \tilde{\beta}_E / \sigma_E$, where $\tilde{\beta}_E$ is the intensity of directional selection on the ecological trait, which is defined as the product of σ_E and the (linear) regression of lifetime relative fitness onto the trait value (Lande & Arnold 1983). At a migration–selection balance, it can be shown that $\tilde{\beta}_E = m D_E / h_E^2$, where D_E is the difference between the means of the ecological trait in the two populations (measured in phenotypic standard deviations) and h_E^2 is the ecological trait's heritability in the focal population.

The difference in allele frequencies in equation (3) can be written as $(p' - p) = -D_E \sigma_E / \gamma$. These results give

$$I \approx \frac{2m}{h_E^2} D_E^2. \quad (5)$$

This shows that reinforcement increases rapidly (as the square) with D_E , the difference between the means of the ecological trait in the two populations. This is sensible because a large difference implies that hybrid offspring will be poorly adapted to the local ecological conditions. Reinforcement is also enhanced with larger values of the introgression rate m and small values of the heritability h_E^2 . However, the heritability will generally depend on m and D_E . I decreases as m increases in equation (4), while in equation (5) it seems to do the reverse. However, D_E , the difference between the trait means that also appears in equation (5), decreases as m grows, so in fact both results lead to the conclusion that selection against hybrid incompatibility declines as migration increases. The next section finds an expression for I that does not involve D_E .

Equations (4) and (5) apply when genetic variation in the ecological trait is caused by a single locus. Another genetic possibility is the situation where the trait is polygenic. I will now assume that genetic variation is caused by a large number of genes of approximately equal effect, that the phenotypic distribution is symmetric and that the fitness function is relatively smooth (specifically, that its third and higher derivatives are small over the range of phenotypic variation). These assumptions simplify the selection coefficients that appear in equation (3) (Kirkpatrick & Servedio 1999).

Appendix A uses those results in order to find that the intensity of hybrid incompatibility caused by a polygenic ecological trait is

$$I \approx \left(2 \frac{m}{h_E^2} - (4/3) \tilde{\Gamma}_E \right) D_E^2. \quad (6)$$

Comparing this result with equation (5), we see that an extra term involving stabilizing selection appears when there are many loci underlying the ecological trait. The selection gradient $\tilde{\Gamma}_E$ is negative when stabilizing selection is acting on the ecological trait and is positive when there is disruptive selection. Thus, equation (6) shows that stronger stabilizing selection leads to stronger reinforcement.

An appealing feature of equations (5) and (6) is that, in principle, all of the quantities that appear in them can be estimated from phenotypic data. If those parameters can be estimated (and if the assumptions of the model are met) then the value for I that they produce can be substituted into equations (1) or (2) in order to determine how much selection against ecologically incompatible hybrids is contributing to reinforcement.

(c) The intensity of hybrid incompatibility in terms of niche differences

The previous result focused on D_E , the difference between the means of the ecological trait in the two populations. That difference depends on the balance between introgression and selection acting on the focal population. An alternative way of expressing the results is in terms of the difference between the mean phenotype of

the immigrants and the local ecological optimum. If the population from which the immigrants arrive is at its ecological optimum, then this new measure reflects the difference between the niches of the immigrants and the local population.

In order to take this approach we need to make the additional assumption that the strength of stabilizing selection is independent of the mean phenotype. (In other words, the curvature of the log of fitness is constant, which implies that the fitness function is proportional to a Gaussian distribution.) Then the intensity of directional selection towards the local ecological optimum is $\tilde{\beta}_E = 2\tilde{\Gamma}_E(\bar{E} - \theta_E)/\sigma_E$, where θ_E is the value of the ecological trait that maximizes lifetime fitness. With that result and equation (6), we find that, when variation in the ecological trait is caused by many loci, then the intensity of hybrid incompatibility is

$$I \approx \frac{4h_E^2\tilde{\Gamma}_E^2(2m - (4/3)h_E^2\tilde{\Gamma}_E)\Delta_E^2}{(m - 2h_E^2\tilde{\Gamma}_E)^2}, \quad (7)$$

where $\Delta_E = (\theta_E - \bar{E})/\sigma_E$.

In effect, Δ_E reflects the difference between the niches of the two populations. If the mean trait value among migrants is at the ecological optimum for their population, then Δ_E is the difference between the optimal phenotypes for the two populations, which is measured in phenotypic standard deviations.

Equation (7) shows that the intensity of selection on hybrid incompatibility grows rapidly as the difference Δ_E between the niches grows: I is proportional to the square of Δ_E . Analysis shows that I also grows with the intensity of stabilizing selection $\tilde{\Gamma}_E$. When the hybridization rate is sufficiently small ($m < -2h_E^2\tilde{\Gamma}_E/3$), the intensity of hybrid incompatibility increases with m and decreases with h_E^2 . These conclusions assume that all else is equal, but some parameters (such as heritability) will of course be influenced by others (such as migration and the intensity of stabilizing selection).

If only a single locus contributes variation to the ecological trait, then the second term inside the parentheses in the numerator of equation (7) disappears. This again shows that stabilizing selection acting on polygenic traits contributes to reinforcement.

3. DISCUSSION

These results show how much selection against ecological traits in hybrids will enhance the reinforcement of pre-zygotic isolation between two populations. The answer is phrased in terms of a quantity I , which I refer to as the intensity of selection against hybrid incompatibility. Qualitatively, the basic trends are what one expects intuitively: reinforcement is strengthened by larger differences between the mean phenotypes of the two species and by stronger stabilizing natural selection around the local ecological optimum. Other effects are perhaps less intuitive: for example, reinforcement is stronger when there are many loci contributing to the ecological trait than when there is just a single locus.

A quantity that is strangely absent from the expressions for I is the fitness of the F_1 hybrids. That is, the depth of the adaptive valley between the species does not appear

explicitly in the results for the intensity of selection against hybrids. In fact it does appear, but in a cryptic form. If the fitness function has the simple shape of two peaks connected by a smooth valley, then the fitness of F_1 s is approximately proportional to the product $\Delta_E^2\tilde{\Gamma}_E$. That quantity appears in equation (7), so we can interpret I as indeed depending on the fitness of the F_1 hybrids.

The assumptions used in the model developed here are restrictive in several ways, and it would be helpful to understand what happens when they are relaxed. The model assumes haploidy. One expects that the conclusions also apply to diploids when there is no dominance. However, we currently have no analytic results to suggest what happens when any of the alleles involved in reinforcement show strong dominance. A second situation for which it would be useful to have analytic results is when there is two-way hybridization. The simulations of Servedio (2000) suggested that reinforcement might be easier when there is reciprocal introgression. The basic reason for this is that alleles that increase pre-zygotic isolation are not simply diluted out of each population by immigration: they themselves can emigrate to the other population and then return, making it less likely they will be lost.

The intensity of hybrid incompatibility has also been calculated for the situation where epistatic incompatibilities between the two species reduce fertility or cause problems in early development (Kirkpatrick & Servedio 1999). Because the quantities involved in that case are different from the ones studied here, it is difficult to draw general conclusions about the relative contributions that these two mechanisms might make towards reinforcement.

Empirical data are therefore needed in order to evaluate their relative importance. The models point to the quantities about which we need more information. When reinforcement is caused by selection against ecological intermediates, we need to know the difference between the mean trait values for the hybridizing species, the heritabilities, the strength of stabilizing selection and the frequency of hybridization. When reinforcement is caused by selection against epistatic incompatibilities that disrupt development, we need to know the number of epistatic sets of genes, the number of loci within each set and the strength of selection against discordant combinations of alleles in each set (Kirkpatrick & Servedio 1999). Until those kinds of data become available, it will be difficult to have any idea about which type of selection against hybrids is the more powerful driving force behind reinforcement in natural populations.

I am very grateful to S. Otto, D. Schluter, M. Servedio, M. Whitlock and two anonymous reviewers for discussion and comments and to the National Science Foundation for support from grant DEB-9973221.

APPENDIX A

This appendix derives the result for the intensity of hybrid incompatibility I when genetic variation in the ecological trait is caused by many loci of approximately equal effect. When the phenotypic distribution is

symmetric and when third and higher derivatives of the fitness function are small, the selection coefficients appearing in equation (3) are

$$s_j = \frac{\gamma_j \tilde{\beta}_E}{\sigma_E} + \frac{\gamma_j^2 (1 - 2p_j) \tilde{\Gamma}_E}{\sigma_E^2}, \quad (\text{A1})$$

and

$$s_{jk} = \frac{2\gamma_j \gamma_k \tilde{\Gamma}_E}{\sigma_E^2}. \quad (\text{A2})$$

Selection coefficients involving three or more loci vanish. (These results are from equation (B8) in Kirkpatrick & Servedio (1999), but here the terms involving the stabilizing selection gradient $\tilde{\Gamma}_E$ are multiplied by a factor of 2 in order to make them consistent with the definition given by Lande & Arnold (1983).) Substituting equations (A1) and (A2) into equation (3) gives

$$I \approx -2 \sum_{j \in H} \left[\frac{\gamma_j}{\sigma_E} \tilde{\beta}_E + \frac{\gamma_j^2}{\sigma_E^2} (1 - 2p_j) \tilde{\Gamma}_E \right] (p'_j - p_j) - (4/3) \sum_{j,k \in H; j \neq k} \frac{\gamma_j \gamma_k}{\sigma_E^2} \tilde{\Gamma}_E (p'_j - p_j) (p'_k - p_k). \quad (\text{A3})$$

(The factor of 2 in the expression for s_{jk} has been cancelled by a factor of 0.5, which appears because the double summation in equation (A3) counts s_{jk} separately from s_{kj} , whereas in equation (3) the term involving that selection coefficient appears only once.)

For a trait with a given (and finite) amount of genetic variance, γ becomes smaller as the number of loci grows. Consequently, the term involving γ_j^2 in equation (A3) becomes negligible. Further, the double summation converges on one in which terms with $j=k$ are included. Then

$$I \approx -2 \sum_{j \in H} \frac{\gamma_j}{\sigma_E} \tilde{\beta}_E (p'_j - p_j) - (4/3) \sum_{j,k \in H} \frac{\gamma_j \gamma_k}{\sigma_E^2} \tilde{\Gamma}_E (p'_j - p_j) \times (p'_k - p_k) \approx 2\tilde{\beta}_E D_E - (4/3) \tilde{\Gamma}_E D_E^2. \quad (\text{A4})$$

Last, using the expression for $\tilde{\beta}_E$ given in the text just before equation (5) we arrive at equation (6).

REFERENCES

- Barton, N. H. & Turelli, M. 1991 Natural and sexual selection on many loci. *Genetics* **127**, 229–255.
- Dobzhansky, T. 1937 *Genetics of the evolutionary process*. New York: Columbia University Press.
- Dobzhansky, T. 1940 Speciation as a stage in evolutionary divergence. *Am. Nat.* **74**, 312–321.
- Haldane, J. B. S. 1930 A mathematical theory of natural and artificial selection. VI. Isolation. *Proc. Camb. Phil. Soc.* **26**, 220–230.
- Hall, D. W., Kirkpatrick, M. & West, B. 2000 Runaway selection when female preference is directly selected. *Evolution* **54**, 1862–1869.
- Hatfield, T. & Schluter, D. 1999 Ecological speciation in sticklebacks: environmental-dependent hybrid fitness. *Evolution* **53**, 866–873.
- Kirkpatrick, M. 2000 Reinforcement and divergence under assortative mating. *Proc. R. Soc. Lond. B* **267**, 1649–1655.
- Kirkpatrick, M. & Servedio, M. R. 1999 The reinforcement of mating preferences on an island. *Genetics* **151**, 865–884.
- Lande, R. & Arnold, S. J. 1983 The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226.
- Muller, H. J. 1942 Isolating mechanisms, evolution and temperature. *Biol. Symp.* **6**, 71–125.
- Schluter, D. 1998 Ecological causes of speciation. In *Endless forms: species and speciation* (ed. D. J. Howard & S. H. Berlocher), pp. 114–129. Oxford University Press.
- Servedio, M. 2000 Reinforcement and the genetics of non-random mating. *Evolution* **54**, 21–29.