

**Behavior of Dobzhansky-type epistatic hybridization models
under varying dominance and selection: preliminary numerical simulations**

Research report to the Worldwide Universities Network (W.U.N.)

Andrew L. Hipp
Department of Botany
University of Wisconsin – Madison
Madison, Wisconsin 53706
U.S.A.

Seth Bullock
Informatics Research Institute (IRI)
School of Computing
University of Leeds
Leeds, LS2 9JT
England

In collaboration with Roger K. Butlin (University of Leeds)
and David M. Shuker (University of Edinburgh)

Research supported by a W.U.N. Global Exchange Award
to Andrew L. Hipp for the period 1 June – 30 June 2004

Introduction

Dobzhansky (1937) formulated a two-locus model of genetic divergence under which derivative populations are reproductively isolated from one another but not from the ancestral population. Models of this type, referred to as Dobzhansky-type epistatic models, have the potential to explain hybridization and speciation scenarios in which fertile genetic and morphological intermediates exist despite the fact that F1 hybrids between the source populations are sterile. Such models can also explain asymmetrical clines in characters that vary between the source populations, as well as offsets in those clines from the hybrid zone center.

Sergey Gavrilets (1997) studied the properties of hybrid zones under Dobzhansky-type epistatic selection using mathematical perturbation methods that allowed him to analytically approximate equilibrium gamete and allele frequencies under varying selection and migration parameters. He also simulated equilibrium allele frequencies across a hybrid zone for parameter values beyond the range for which perturbation methods are expected to be robust. The conclusions of his study included the findings that Dobzhansky-type epistatic selection does indeed permit limited hybridization while posing less of a barrier to neutral gene flow.

In this study, we further investigate the properties of Dobzhansky-type hybrid zones using deterministic numerical simulations based on the two-locus model formulated in Gavrilets (1997). We investigate multivariate cline properties across a wide

range of selection parameters, including scenarios in which derived alleles are completely dominant, scenarios in which they are completely recessive, and scenarios in which they are codominant. We build on Gavrilets' model by incorporating a separate parameter for selection on derived alleles, allowing us to evaluate cases in which derived alleles became fixed by genetic drift versus those in which derived alleles became fixed by natural selection or became adaptive subsequent to fixation. We also demonstrate how varying degrees of asymmetry in selection parameters affect the frequency of recombinant gametes at the center of the hybrid zone and evaluate the effects of individual parameters on the degree of hybridization.

Methods

Gavrilets' (1997) numerical method of simulating equilibrium allele and gamete frequencies across hybrid zones was implemented in R version 2.0.0 (R Development Core Team, 2004) by iterating his equations (2) and (4a-c); neutral alleles were not analysed for this paper. To address the effects of selection on derived alleles in the case in which alleles are codominant, fitness matrix (3) an additional factor nw was added to each cell of the fitness matrix, where n = number of derived alleles in the genotype (e.g., $n = 4$ for genotype AA bb , $n = 2$ for genotypes AaBb, AABB, and aabb) and w = the increase in relative fitness for each derived allele. The matrix was normalized such that maximum fitness = 1. The effects of treating derived alleles as dominant were simulated by setting $\beta = \beta_1 = s > 0$, $\alpha = \alpha_1 = 0$; derived alleles were treated as recessive in alternative simulations by setting $\alpha = \alpha_1 = \beta = s = 0$, $\beta_1 > 0$. Simulations for direct comparison with Gavrilets were run for 1000 generations with 12 islands, with migration rates $m = m_0 = 0.005$, the value used by Gavrilets in his numerical simulations (the reported $m = m_0 = 0.05$ in Gavrilets 1997 is an error).

To simultaneously compare the effects of several parameters on hybrid zone dynamics, mean fitness across all genotypes in the central islands and frequency of recombinant (hybrid) gametes in the central islands was graphed against the parameters of interest. Initial trials demonstrated that the same trends in these two dependent variables were found across a range of eight to twenty islands. Simulations were consequently run for 1000 generations with eight islands. Initial trials indicated that 1000 generations was adequate to investigate trends, though some of the clines change subtly when run for 2000 generations. Code for performing simulations is available upon request.

Results and Discussion

Effects of disentangling selection for derived alleles from selection against hybrids

The fitness matrices used for numerical simulations in Gavrilets (1997, pp. 1032 ff.) appear to approximate the selection expected when positive selection on derived alleles = 0.01 (Table 1). This models the situation in which derived alleles have become fixed under natural selection or become adaptive subsequent to fixation. Simulations run under the assumption that $\alpha = \alpha_1 = 0$ and $w = 0.01$ recover allele frequency clines that are virtually indistinguishable from simulations run at $\alpha = 0.01$ and $\alpha_1 = 0.02$ with $w = 0$ (Figure 1), suggesting that the numerical simulations in Gavrilets 1997 may not generalize well to cases in which derived alleles arose by drift alone.

The effect of added selection for derived alleles is a decrease in the offset between allele clines, narrowing the hybrid zone and decreasing the frequency of recombinant gametes (Figures 1 – 3). The main cause of this effect appears to be that adding selection for derived alleles into the model equalizes selection against recombinant genotypes composed of ancestral-type gametes (**aB**) and derived-type gametes (**Ab**). The model thus begins to depart from the Dobzhansky model of a fitness ridge connecting two derived genotypes, with a fitness valley between. Results of our simulations demonstrate that even moderate selection for derived alleles substantially decreases frequency of the recombinant gametes in the central islands (Figures 3 and 4). Some of the most striking results are from simulations with substantial selection against F1 hybrids (e.g., $s = 0.9$, Figure 4b). With light selection for derived alleles (e.g., $\alpha = 0.01$, $\alpha_1 = 0.02$, as Gavrilets uses), even very low values of β and β_1 reduce frequency of the hybrid gamete in the central islands to well under 0.1. The lowest rates of recombination at any degree of selection against F1 hybrids (i.e., across all values of s) are found in the symmetrical selection matrices, in which $\alpha = \beta$ (Figure 4a and 4b). This effect is obscured if one looks only at cases in which selection for the derived alleles is greater than the migration rate, e.g., Gavrilets' numerical simulations (in which $\alpha = 0.01$ and $m = m_0 = 0.005$).

In other words, weak selection against either the recombinant ancestral genotype ($\alpha > \beta = 0$) or the recombinant derived genotype ($\beta > \alpha = 0$) has the effect of increasing the frequency of recombinant gametes even in cases in which s is high enough to significantly depress hybridization. For example, with only moderate selection against F1 individuals ($s = 0.4$), recombinant gametes of both the ancestral and derived type form at equal and very low frequencies in the center of the hybrid zone (Figure 5). The same allele clines are seen at $s = 0.4$ with $\alpha = \alpha_1 = \beta = \beta_1 = 0.01$. (Not surprisingly, the effect of selection against the F1 is stronger than the effect of selection against F2; even strong selection against the F2 (e.g., $\alpha = \alpha_1 = \beta = \beta_1 = 0.999$) permits the formation of recombinant gametes in the hybrid zone as long as $s \ll m$.) The introduction of minor asymmetry in the degree of selection, however, greatly increases the frequency of one of the recombinant genotypes in the center of the hybrid zone, so long as either α or β is roughly equal to or less than the migration rate (Figure 6). This phenomenon is reflected in the steep gradient in recombinant gamete frequency in the plots of α vs. β (Figures 4a – b).

The relationship between migration rate and the effects of w on recombinant genotype frequency bears further investigation. At some point near $w = m$, the effect of increasing β and s from 0 to 0.2 reverses. At $w \ll m$, increases in β and s increase the frequency of recombinant genotypes in the central islands (Figure 7a). At $w \gg m$, increases in β and s decrease the frequency of recombinant genotypes (Figure 7c). At intermediate levels of w , there is an initial decrease in recombinant genotypes, followed by an increase (7b). These effects reflect the steep gradient in recombinant frequency as the selection matrix becomes asymmetrical, and one might expect the gradient to be steeper at lower migration rates.

Effects of treating derived loci as dominant vs. recessive

The cases that Gavrilets studies presume that derived alleles are completely dominant ($s = \beta = \beta_1 > 0$). Under these conditions, recombinant gamete frequencies

decrease with increases in selection across the range of values investigated in Gavrilets ($s = \beta = \beta_1 = 0.5 - 0.9$) (Figure 9). Under the assumption that derived alleles are recessive, however ($s = \beta = 0, \beta_1 > 0$), recombinant gamete frequencies increase slightly with increasing selection (Figure 10). This is presumably due at least in part to the fact that if the derived alleles are recessive, there is no selection against F1 hybrids. There is also no selection against any of the backcrosses between F1 hybrids and the source populations, which presumably make up a sizable percentage of all crosses.

At lower β_1 levels, there is a decrease in recombinant genotype frequency with increasing selection (Figure 8), paralleling the trend under the assumption that derived alleles are dominant (cf. Figure 7). Again due to the reduced selection under the case in which derived alleles are dominant – i.e., in the absence of selection against F1 hybrids and backcrosses between F1 and source populations – gamete frequency clines are not as steep under this scenario as they are when derived alleles are dominant.

Conclusions and Directions for Future Research

This study is a preliminary investigation of the properties of simple two-locus Dobzhansky-type hybrid zones across a relatively wide range of selection parameters and demonstrates that there can be reversals in gamete cline trends based on whether derived alleles are neutral or selected and whether they are dominant or recessive. It also demonstrates that the effect of selection on derived alleles is apparently stronger than the effect of dominance, though I have not endeavoured to study the possible effects of overdominance or underdominance on allele and gamete clines (this effect has been studied by Goodisman and Crozier (2001) relative to the observation that hybrids have higher fitness than parental species in novel environments).

There are several directions for future research suggested by this work. First, an investigation of the relationship between migration rate and different selection parameters is needed to evaluate how robust the cross-over points found in this study are to different migration scenarios. The reversals of trend found with increasing selection for derived alleles (parameter w) appear to correlate closely with migration rate (m), but without study across a range of migration rates this is not clear. I have also not varied the recombination rate (r) in this study. While it is reasonable to expect that two antagonistic loci might be unlinked, this expectation becomes less plausible as an increasing number of genes come into play. Moreover, the choice of dependent variables in this study (viz., average population fitness and recombinant gamete frequency of central populations) needs to be evaluated by comparison with other correlates of hybridization (e.g., asymmetry and offset of allele clines, frequency of offspring genotypes rather than gametic genotypes). When these have been evaluated, the effects of incompatibilities at sex-linked vs. autosomal loci should be compared using the methods of numerical simulation employed in this study and in Gavrilets (1997).

This study and resolution of the questions outlined above are intended to serve in part as a groundwork for future collaborative study of hybrid zone dynamics using an agent-based simulation in development by Seth Bullock. This simulation will allow us to study the effects of incompatibilities at many additional loci, both sex-linked and autosomal, and, potentially, to measure the effects of population size constraints, differential mortality of males and females, environmental selection gradients, and a wide

