Clines maintained by overdominant selection in hybrid zones

MICHAEL A. D. GOODISMAN and ROSS H. CROZIER

Department of Genetics, La Trobe University, Bundoora, Australia
Department of Zoology and Tropical Ecology, James Cook University, Townsville, Australia

Hybrid zone models often consider environment-independent selection to operate against all hybrids. However, empirical studies suggest that hybrids may be as fit or fitter than the hybridizing parental taxa in some environments. In this study we develop a novel mathematical model to explore the effects of one form of hybrid superiority on the genetic structure of hybrid zones. Our primary goals were to investigate the allele frequency clines at a locus experiencing overdominant selection and at a linked neutral or underdominant locus. Our results indicate that overdominant selection results in flat equilibrium allele frequency clines throughout the hybrid zone and an excess of heterozygosity relative to Hardy-Weinberg equilibrium. However, the genetic clines at linked neutral or underdominant loci tend not to reflect this overdominance even when the loci are tightly linked. Overall, we conclude that overdominance is unlikely to be detected in genetic surveys unless many loci are assayed.

Michael Goodisman, Department of Biochemistry and Molecular Biophysics, 509 Biological Sciences West, University of Arizona, Tucson, AZ 85721-0088, USA. E-mail: goodisma@email.arizona.edu

Hybrid zones are regions where individuals originating from genetically differentiated populations successfully mate and produce offspring (HARRISON 1990; ARNOLD 1997). These areas provide prime opportunities to study fundamental evolutionary processes such as gene flow, selection, and speciation (HARRISON 1993). To properly interpret results arising from studies of hybrid zones it is important to understand the factors that maintain their structure and composition.

In general, the long-term stability of hybrid zones has been viewed as resulting from an interaction between migration of individuals through the hybrid zone and natural selection operating differentially on parental and hybrid genotypes (ENDLER 1977; BARTON and HEWITT 1985, 1989; HARRISON 1993; ARNOLD 1997). However, the exact nature of the selective processes operating within hybrid zones remains a matter of considerable controversy. For example, most genetic models of hybridization assume hybrids have lower fitness than at least one of the parental taxa irrespective of the environment in which the comparison is made (e.g., BAZYKIN 1969; BARTON 1979; BARTON and GALE 1993; BAIRD 1995; GAVRILETS and CRUZAN 1998; CAIN et al. 1999; DURRETT et al. 2000; but see GAVRILETS 1997).

However, some empirical studies suggest that hybrids attain fitness equal to, or greater than, individuals from the parental populations in at least some environments (STEBBINS 1965; MOORE 1977; ARNOLD and HODGES 1995; GRANT and GRANT 1996; ARNOLD 1997; CAMPBELL et al. 1998; ARNOLD

et al. 1999). These cases have sometimes been viewed as providing support for the hypothesis that hybridization can be a creative process leading to the generation of novel taxa (ARNOLD 1997). Unfortunately, no theoretical framework exists for interpreting results from genetic studies of these evolutionarily important hybrid zones.

The purpose of this investigation was to expand the existing mathematical theory available for analyzing regions of hybridization by developing a framework useful in interpreting results from hybrid zones where environment-dependent selection for hybrids occurs. We examined the steady-state patterns in nuclear allele frequencies and disequilibria at the focal selected locus and at a linked locus experiencing no, or alternate forms, of selection. Our overall goals were to determine the genetic signature that results when hybrids possess relatively high fitness through overdominance and to ascertain if loci conferring hybrid advantage are likely to be detected in genetic surveys of natural hybrid zones.

HYBRID ZONE MODEL

We consider the diploid genotypes of individuals at two diallelic nuclear loci, locus A and locus B. The frequencies of alleles A_1 and A_2 at locus A are p and $1 - p$, respectively, and the frequencies of alleles B_1 and B_2 at locus B are q and $1 - q$, respectively. Table 1 gives the 10 two-locus genotypes in this system and defines f_i as the frequency of genotype i, for $i = 1$ to 10. We allow for variation in the recombination rate, r ($0 \leq r \leq 0.5$), between loci A and B. If $r = 0$ then the

Table 1. *Ten two-locus genotypes when both loci are diallelic*

Genotype	$A_1B_1/$ A_1B_1	$A_1B_1/$ A_1B_2	$A_1B_2/$ A_1B_2	$A_1B_1/$ A_2B_1	$A_1B_1/$ A_2B_2	$A_1B_2/$ A_2B_1	$A_1B_2/$ A_2B_2	$A_2B_1/$ A_2B_1	$A_2B_1/$ A_2B_2	$A_2B_2/$ A_2B_2
Number	1	2	3	4	5	6	7	8	9	10

two loci are completely linked and if $r = 0.5$ then the loci are unlinked.

To document the genetic patterns throughout the hybrid zone, we monitor nonrandom associations of alleles at loci A and B. First, we consider deviations of the genotype frequencies from expectations under Hardy-Weinberg equilibrium. These deviations are quantified by $F_A = 1 - (f_4 + f_5 + f_6 + f_7)/[2p(1-p)]$ for locus A, and $F_B = 1 - (f_2 + f_5 + f_6 + f_9)/[2q(1-q)]$ for locus B. Positive values of F_A or F_B indicate a deficit of heterozygotes relative to Hardy-Weinberg equilibrium at locus A or B, respectively. Second, we measure nonrandom associations between alleles at the two loci (i.e., linkage disequilibrium), defined as $D = f_1 + \frac{1}{2}f_2 + \frac{1}{2}f_4 + \frac{1}{2}f_5 - pq$. Positive values of D indicate an excess of coupling gametes relative to expectations based on random association of alleles at loci A and B. Because values of D depend heavily on the allele frequencies, we will also monitor the normalized estimate of gametic disequilibrium, D', which always varies between -1 and $+1$ (WEIR 1996).

We assume that two populations overlap in an area of sympatry consisting of x linearly arranged subpopulations and produce hybrid offspring with non-zero fitness. Individuals from each of these two populations are referred to as belonging to either species 1 or species 2, although the two groups are not reproductively isolated. We further suppose that species 1 is fixed for alleles A_1 and B_1, and species 2 is fixed for alleles A_2 and B_2. Therefore, all species 1 individuals are genotype A_1B_1/A_1B_1 and species 2 individuals are genotype A_2B_2/A_2B_2. Generations are presumed to be discrete and non-overlapping, and population size is large enough so that genetic drift does not affect the genotype frequencies within the hybrid zone.

Under our model, individuals first mate and produce offspring. We assume that mating occurs at random within subpopulations and that gametes undergo recombination with probability r. After progeny production, the frequencies of the 10 two-locus genotypes are

$$\tilde{f_1} = (\tfrac{1}{2}\alpha_2 + \beta_1)^2 \tag{1A}$$

$$\tilde{f_2} = \tfrac{1}{2}f_2(f_5 + f_6) + \alpha_1(f_1 + \tfrac{1}{2}f_4) + \alpha_2(f_3 + \tfrac{1}{2}f_7) + 2\beta_1\beta_2 + \gamma \tag{1B}$$

$$\tilde{f_3} = (\tfrac{1}{2}\alpha_1 + \beta_2)^2 \tag{1C}$$

$$\tilde{f_4} = \tfrac{1}{2}f_4(f_5 + f_6) + \alpha_1(f_1 + \tfrac{1}{2}f_2) + \alpha_2(f_8 + \tfrac{1}{2}f_9) + 2\beta_1\beta_3 + \gamma \tag{1D}$$

$$\tilde{f_5} = \tfrac{1}{2}(\alpha_2)^2 + 2\beta_1\beta_4 + \alpha_2(\beta_1 + \beta_4) \tag{1E}$$

$$\tilde{f_6} = \tfrac{1}{2}(\alpha_1)^2 + 2\beta_2\beta_3 + \alpha_1(\beta_2 + \beta_3) \tag{1F}$$

$$\tilde{f_7} = \tfrac{1}{2}f_7(f_5 + f_6) + \alpha_1(\tfrac{1}{2}f_9 + f_{10}) + \alpha_2(\tfrac{1}{2}f_2 + f_3) + 2\beta_2\beta_4 + \gamma \tag{1G}$$

$$\tilde{f_8} = (\tfrac{1}{2}\alpha_1 + \beta_3)^2 \tag{1H}$$

$$\tilde{f_9} = \tfrac{1}{2}f_9(f_5 + f_6) + \alpha_1(\tfrac{1}{2}f_7 + f_{10}) + \alpha_2(\tfrac{1}{2}f_4 + f_8) + 2\beta_3\beta_4 + \gamma \tag{1I}$$

$$\tilde{f_{10}} = (\tfrac{1}{2}\alpha_2 + \beta_4)^2 \tag{1J}$$

where

$$\alpha_1 = rf_5 + (1-r)f_6 \qquad \alpha_2 = (1-r)f_5 + rf_6$$

and

$$\beta_1 = f_1 + \tfrac{1}{2}f_2 + \tfrac{1}{2}f_4 \qquad \beta_2 = \tfrac{1}{2}f_2 + f_3 + \tfrac{1}{2}f_7$$

$$\beta_3 = \tfrac{1}{2}f_4 + f_8 + \tfrac{1}{2}f_9 \qquad \beta_4 = \tfrac{1}{2}f_7 + \tfrac{1}{2}f_9 + f_{10}$$

and

$$\gamma = \tfrac{1}{2}f_5f_6[r^2 + (1-r)^2] + \tfrac{1}{2}r(1-r)[(f_5)^2 + (f_6)^2].$$

We note that these equations differ from those typically used in two-locus analyses (i.e., LEWONTIN and KOJIMA 1960) in that (1) focuses on diploid genotypes rather than on haploid gametes. Therefore, (1) does not rely on the genotype frequencies in the previous generation being in Hardy-Weinberg proportions.

After mating and reproduction, individuals within each subpopulation undergo viability selection. The fitness of an individual of genotype k, for $k = 1$ to 10, is denoted as w_k and reflects the multiplicative effects of loci A and B. We assume no multiplicative epistasis between the loci so the fitnesses of the two-locus genotypes may be parameterized as in Table 2

Table 2. *Fitness of genotypes in two-locus diallelic model with multiplicative fitness of loci and no epistasis*

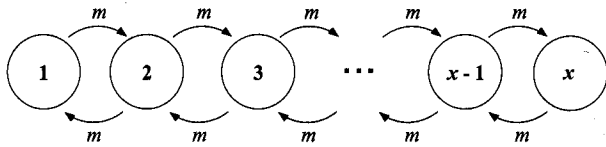|	B_1B_1	B_1B_2	B_2B_2
A_1A_1	$w_1 = ab$	$w_2 = a$	$w_3 = ac$
A_1A_2	$w_4 = b$	$w_5 = w_6 = 1$	$w_7 = c$
A_2A_2	$w_8 = bd$	$w_9 = d$	$w_{10} = cd$

Fig. 1. Stepping-stone model of migration with x subpopulations. Each generation, a constant proportion of individuals, m, from each subpopulation migrates to the adjacent subpopulations.

(FELSENSTEIN 1965). Under our model, the fitnesses of the different genotypes are not constant across the hybrid zone. The 'end' subpopulations, 1 and x, preferentially support individuals that are species 1-like and species 2-like, respectively. However, the fitnesses of genotypes in the other 'central' subpopulations j, for $j = 2, 3, \ldots, x - 1$, vary (see below). Thus our framework parallels aspects of the bounded hybrid superiority model originally outlined by MOORE (1977). Selection in a given subpopulation changes the frequency of genotype k according to the equation

$$f_k = \frac{\tilde{f}_k w_k}{\bar{w}} \tag{2}$$

where

$$\bar{w} = \sum_{k=1}^{10} \tilde{f}_k w_k.$$

The final process operating within the hybrid zone is migration. Individuals move through the hybrid zone according to the one-dimensional stepping-stone model (KIMURA and WEISS 1964), so that a constant proportion of individuals, m $(0 < m < 0.5)$, within each subpopulation migrates to the adjacent subpopulation(s) each generation (Fig. 1). Note that the end subpopulations exchange migrants with only one other subpopulation and the central subpopulations exchange migrants with two other subpopulations. Therefore, each generation migrants replace a fraction m and $2m$ of end and central subpopulations, respectively.

The frequency of genotype k in the migrants entering any given subpopulation is defined as $\bar{f}_k$, which is simply the arithmetic mean of the frequency of genotype k in the neighboring subpopulations. After migration the frequency of genotype k is

$$f'_k = m\bar{f}_k + (1 - m)f_k \tag{3A}$$

for end subpopulations and

$$f'_k = 2m\bar{f}_k + (1 - 2m)f_k \tag{3B}$$

for central subpopulations. The values given by (3) represent the genotype frequencies after a single cycle of mating, selection, and migration.

HYBRID ZONE SIMULATIONS

Solving analytically for the equilibrium genotype frequencies under the stepping-stone model with arbitrary fitnesses is complex. Therefore, we resort to numerical methods to explore the distribution of genetic variation within the hybrid zone. We assume that the zone consists of $x = 40$ subpopulations and that the migration rate $m = 0.1$. We set the fitness parameters (Table 2) to $a = b = 1.2$, $c = d = 0.8$ in subpopulation 1 and $a = b = 0.8$, $c = d = 1.2$ in subpopulation x. Therefore, the alternate homozygous genotypes are strongly favored in either end of the hybrid zone in all simulations.

The area of hybridization is presumed to result from secondary contact rather than primary intergradation (ENDLER 1977). Therefore, only individuals with double homozygous genotypes initially reside within the hybrid zone. Specifically, the first $x/2 = 20$ subpopulations of the zone are occupied by individuals of genotype A_1B_1/A_1B_1 only, while the last $x/2 = 20$ subpopulations are occupied by individuals of genotype A_2B_2/A_2B_2 only.

The recursions for the two-locus genotype frequencies defined in (1)–(3) are then iterated to obtain the steady-state genotype frequencies in each subpopulation of the hybrid zone. Specifically, mating, selection, and migration, are allowed to occur until no value of f_i, for $i = 1$ to 10, changes by more than 10^{-6} in a single generation in any of the subpopulations. At this point, the system is defined as being at equilibrium.

Single-locus clines under simple selection schemes

We first explore the steady-state clines in allele frequencies ($\hat{p}$) and deviations from Hardy-Weinberg equilibrium ($\hat{F}_A$) at locus A under three baseline selection schemes. To examine the effects of selection on locus A independent of locus B, we set the recombination rate $r = 0.5$ and the fitness parameters $b = c = 1$ in the central subpopulations. The three simple cases of selection that we initially examine are selective inferiority of hybrids (underdominance at locus A), selective equality of hybrids (neutrality at locus A), and selective superiority of hybrids (overdominance at locus A).

The first of these examples essentially follows the tenets of the tension zone model of hybridization (KEY 1968; BAZYKIN 1969; BARTON and HEWITT 1985), whereby hybrids have fitness lower than the two hybridizing species in all environments. To examine genetic patterns under this scenario, we set a, $d > 1$ in the central subpopulations. The second scheme assumes that hybrid fitness is lower than the parental hybridizing species in the end subpopula-

tions only. In the central subpopulations, the fitness of hybrids equals that of the parental species ($a = d = 1$). Finally, we allow hybrid fitness to exceed that of the parental species within the central subpopulations ($a, d < 1$), thereby resulting in hybrid superiority at the selected locus.

The clines for underdominance at locus A display features expected under this selection scheme (BARTON and HEWITT 1985; BARTON and GALE 1993). The frequency, $\hat{p}$, forms a sigmoidal curve and is highest in the subpopulations that initially contained species 1 individuals (Fig. 2A). We may assess the width of this cline as the area where the allele frequencies fall between 0.2 and 0.8. In this case, we find that the cline for $\hat{p}$ is only four subpopulations wide. In general, the width of an allele frequency cline maintained by dispersal and selection is related to the characteristic scale of selection, $l = \sigma/\sqrt{s}$, where σ is the standard deviation between parent and offspring and s is the measure of selection (BARTON and HEWITT 1985). Under our stepping-stone model, $\sigma = \sqrt{m(1 - m)}$. Previous results have suggested that the width of a cline maintained by selection against hybrids is expected to equal $\sqrt{8}l$ (BARTON and HEWITT 1985), which would be approximately three subpopulations in this case. Thus our results agree qualitatively with this result. In addition, with underdominant selection we find that $\hat{F}_A$ is always positive, reflecting selection for homozygous genotypes throughout the hybrid zone (Fig. 2B). The deficit of heterozygotes becomes relatively large in magnitude in the middle subpopulations where the alleles are at intermediate frequencies.

The clines under selective neutrality of hybrids show substantially different patterns. The allele frequencies now change in a linear fashion through the hybrid zone reflecting the process of simple diffusion of alleles through the central subpopulations (Fig. 2A). The strong selection for homozygous genotypes in the end subpopulations, however, ensures that $\hat{p}$ is high near subpopulation 1 and low near subpopulation 40, as predicted under similar models where the genetic compositions of the end subpopulations are fixed (FELDMAN and CHRISTIANSEN 1975). The values of $\hat{F}_A$ are still greater than zero for all subpopulations, because of the influence of selection operating in favor of the homozygous genotypes in subpopulations 1 and x (Fig. 2B). However, in this case $\hat{F}_A$ takes on relatively small and equal values throughout most of the hybrid zone because of the absence of selection in the central subpopulations.

The steady-state clines for $\hat{p}$ and $\hat{F}_A$ obtained with overdominant selection at locus A display important differences from the two previous examples. The cline for $\hat{p}$ now levels off in the center of the hybrid zone

(Fig. 2A). Importantly, in contrast to the case of underdominance, the width of the allele frequency cline displayed with overdominance is much larger than the scale of selection l, because the influence of
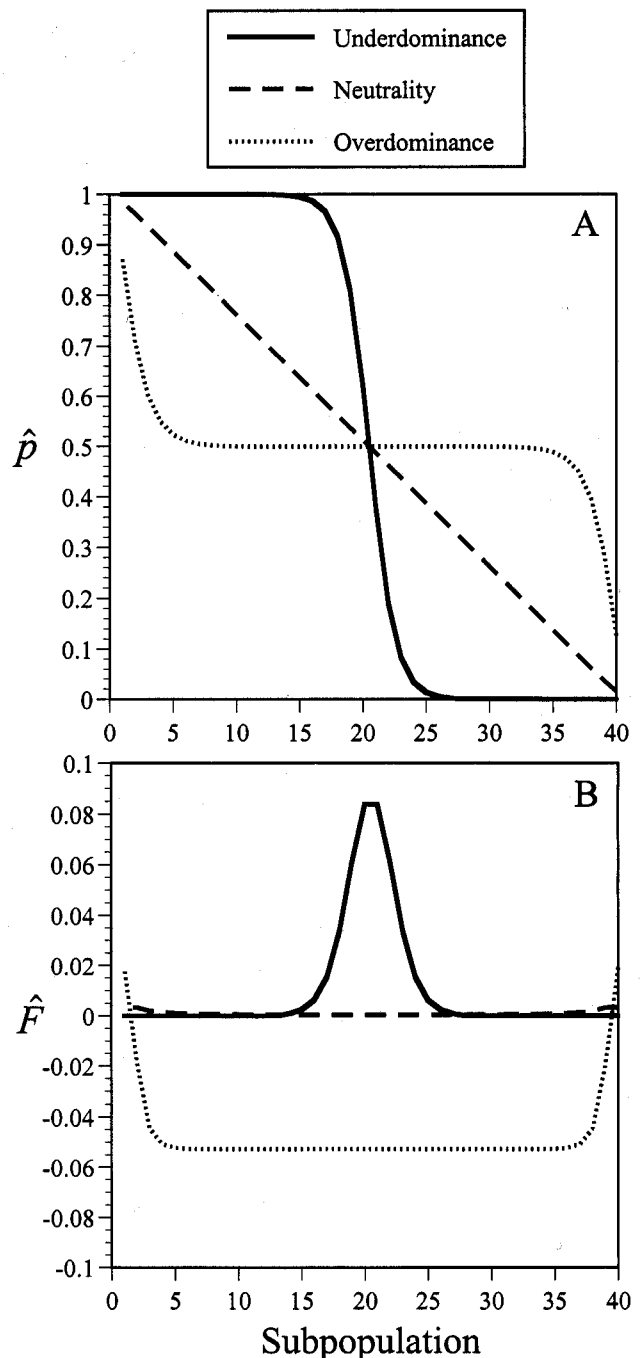

Fig. 2A and B. Effects of underdominance, neutrality, and overdominance on the equilibrium clines for **A)** the frequency of the A_1 allele, $\hat{p}$, and **B)** the deviation from Hardy-Weinberg equilibrium, $\hat{F}_A$. The selection parameters in the central subpopulations for underdominance, neutrality, and overdominance are $a = d = 1.1$, $a = d = 1$, and $a = d = 0.9$, respectively.

dispersal is low relative to selection (BARTON and HEWITT 1985). Another distinction between the two cases is that $\hat{F}_A$ takes on negative values in all but the end subpopulations (Fig. 2B). Both these patterns arise because selection maintains both alleles in heterozygous form in the central subpopulations.

Two locus clines with overdominance

We next investigate if a locus experiencing no selection or underdominant selection reflects the effects of overdominance at a second linked locus. We first consider the case where heterozygotes at locus A possess greater fitness than either homozygote while genotype at locus B does not affect fitness (a, $d < 1$, $b = c = 1$). In contrast to the previous example of heterozygous advantage, however, we now consider how variation in recombination rate between the loci affects the genetic patterns.

Our simulations indicate that clines for $\hat{p}$ flatten in the middle of the hybrid zone and are unaffected by the linkage relationship to locus B (Fig. 3A). Also, the cline for $\hat{q}$ (Fig. 3B) is not altered by selection at locus A when the two loci are unlinked ($r = 0.5$) and is virtually unaffected even when the loci are tightly linked ($r = 0.01$). Only when the recombination rate between the two loci is exceptionally low ($r = 0.0001$), does $\hat{q}$ strongly show the effects of selection operating on locus A.

The statistic $\hat{F}_A$ is negative and virtually unchanged by the linkage relationship between the two loci (Fig. 3C) and reflects the strong selection for heterozygotes in the center of the hybrid zone. In contrast, $\hat{F}_B$ is positive when $r \geq 0.01$ (Fig. 3D); the direct selection for the homozygous genotypes at locus B in the end subpopulations overwhelms the influence of indirect selection for heterozygous genotypes at the linked locus A. Again, only with very tight linkage ($r = 0.0001$) does selection at locus A cause $\hat{F}_B$ to become negative.

The only statistic that displays a substantial difference between the linked and unlinked simulations is the steady-state linkage disequilibrium and normalized linkage disequilibrium (Fig. 3E and 3F). When $r = 0.5$, $\hat{D}$ is small and positive and reaches only a fraction of its maximum. The disequilibrium reaches greater magnitude with tighter linkage and is close to its maximum when $r = 0.0001$. In general, both $\hat{D}$ and $\hat{D}'$ are greatest in magnitude on the edges of the hybrid zone where selection in subpopulations 1 and x acts in favor of the homozygous genotypes. These patterns stand in sharp contrast to those expected under models of selection against hybrids, where linkage disequilibrium reaches its greatest magnitude in the center of the hybrid zone (BARTON and GALE 1993).
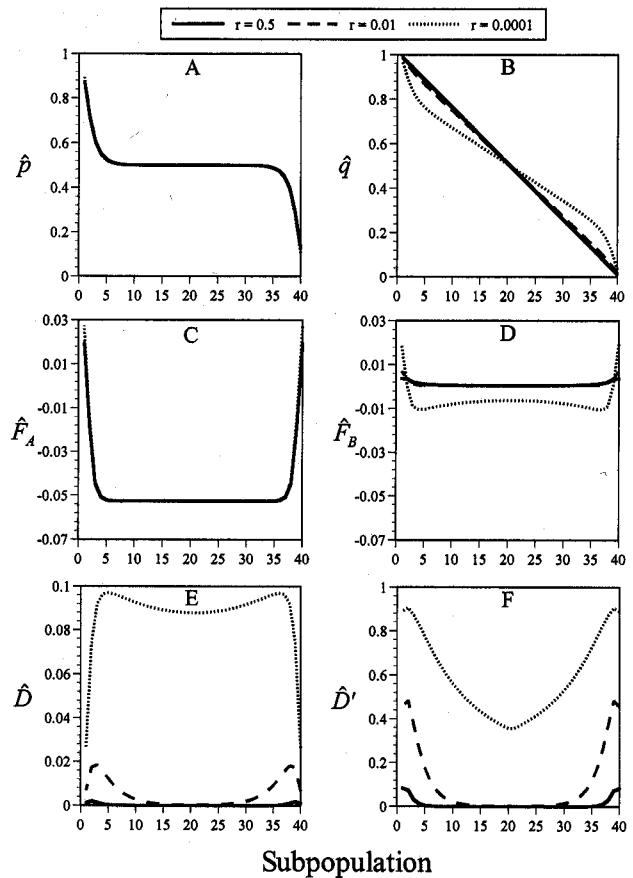

Fig. 3A–F. Effects of overdominant selection on locus A and no selection on locus B on the equilibrium clines for A) the frequency of the A_1 allele, $\hat{p}$, B) the frequency of the B_1 allele, $\hat{q}$, C) the deviation from Hardy-Weinberg equilibrium at locus A, $\hat{F}_A$, D) the deviation from Hardy-Weinberg equilibrium at locus B, $\hat{F}_B$, E) the linkage disequilibrium between loci A and B, $\hat{D}$, and F) the normalized linkage disequilibrium between loci A and B, $\hat{D}'$, for varying recombination rates, r. The selection parameters in the central subpopulations are $a = d = 0.9$, $b = c = 1.0$.

Previous theoretical studies on hybridization have determined that the maximum linkage disequilibrium in hybrid zones maintained by weak selection against hybrids is given by $\tilde{D} \approx \sigma^2(1 + r)/rW_AW_B$, where W_A and W_B are the widths of the clines for loci A and B, respectively (BARTON and GALE 1993). It is of some interest to compare this analytical result derived under a different set of assumptions to our simulated results. Under our framework, $\sigma^2 = m(1 - m)$ and the width of the clines for the loci may be approximated by the distance where the allele frequencies are between 0.2 and 0.8. Not surprisingly, our analyses indicate that maximum $\hat{D}$ observed under our simulations does not coincide with $\tilde{D}$, the value expected under weak selection against hybrids. Specifically, we find that maximum $\hat{D}$ equals 0.0020, 0.018, and 0.097 when $r = 0.5$, 0.01, and 0.0001, respectively. How-

ever, the expected estimates for these recombination rates are $\tilde{D} = 0.00029$, 0.0092, and 0.70.

We next turn our attention to an example of how overdominance at locus A affects variation at a linked locus B which is experiencing underdominant selection (a, $d < 1 < b$, c). We find that the cline for $\hat{p}$ is characteristically flat in the middle of the hybrid zone (Fig. 4A) while $\hat{q}$ follows the sigmoidal pattern expected when hybrid genotypes possess lower fitness than parental genotypes (Fig. 4B). Of substantial importance, however, is that the clines for the equilibrium allele frequencies are once again largely unaffected by variation in the recombination rate between the two loci.

The steady-state patterns for the deviations from Hardy-Weinberg equilibrium are also largely independent of the recombination rate between loci. The
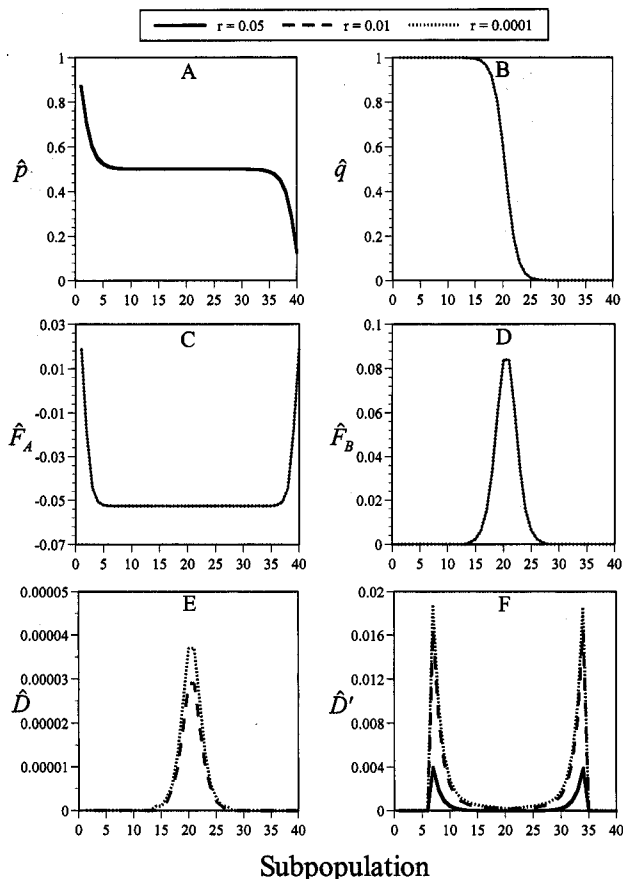
value of $\hat{F}_A$ is negative except in the end subpopulations (Fig. 4C). However, $\hat{F}_B$ is positive in all subpopulations (Fig. 4D), although it is close to zero outside of the center of the hybrid zone, in part because of the lack of variation. The steady-state gametic disequilibrium, $\hat{D}$, is positive and largest in the center of the hybrid zone regardless of the recombination rate between the loci (Fig. 4E). However, $\hat{D}$ is exceptionally small in magnitude and does not take on a particularly large fraction of its maximum value (Fig. 4F). We may once again compare the maximum disequilibrium obtained in this study to that expected, $\tilde{D}$, under models of underdominance. In this case, we expect $\tilde{D}$ to equal 0.0018, 0.060, and ~ 1.0, for $r = 0.5$, 0.01, and 0.0001, respectively. However, the observed maximum of $\hat{D}$ is always much lower than all of these values.

Effects of variation in simulation parameters

Finally, we investigate in more detail how variation in the recombination rate between loci, the migration rate among subpopulations, and the selection scheme alter the genetic structure of the hybrid zone. To document this variation we examine the match between the frequency of the genotypes in the hybrid zone and the fitnesses of those genotypes (after CAIN et al. 1999). We measure this correspondence using the equilibrium mean fitness within the hybrid zone, defined as

$$\hat{w} = \frac{1}{x} \sum_{j=1}^{x} \left(\sum_{k=1}^{10} \frac{\hat{f}_{k(j)} w_{k(j)}}{w_{\max(j)}} \right) \tag{4}$$

in a hybrid zone composed of x subpopulations, where $\hat{f}_{k(j)}$ is the equilibrium frequency of genotype k in subpopulation j, $w_{k(j)}$ is the fitness of genotype k in subpopulation j, and $w_{\max(j)}$ is the largest selection coefficient in subpopulation j. In general, relatively high values of $\hat{w}$ indicate that genotypes match their environment well and relatively low values suggest that a large fraction of unfit genotypes reside within the hybrid zone.

From Figure 5 we see that variation in the recombination rate, r, has very little effect on $\hat{w}$ regardless of the magnitude of m or the selection scheme. These results are consistent with those of previous examples (Figs. 3 and 4), which indicated that allele frequencies and deviations from Hardy-Weinberg equilibrium vary little with r. What is not evident from Figure 5, however, but is apparent from Figures 3E and 4E, is that gametic disequilibria increase in magnitude when the loci under consideration are tightly linked, because linkage of loci A and B allows selection to act more efficiently within the hybrid zone (SLATKIN 1975).
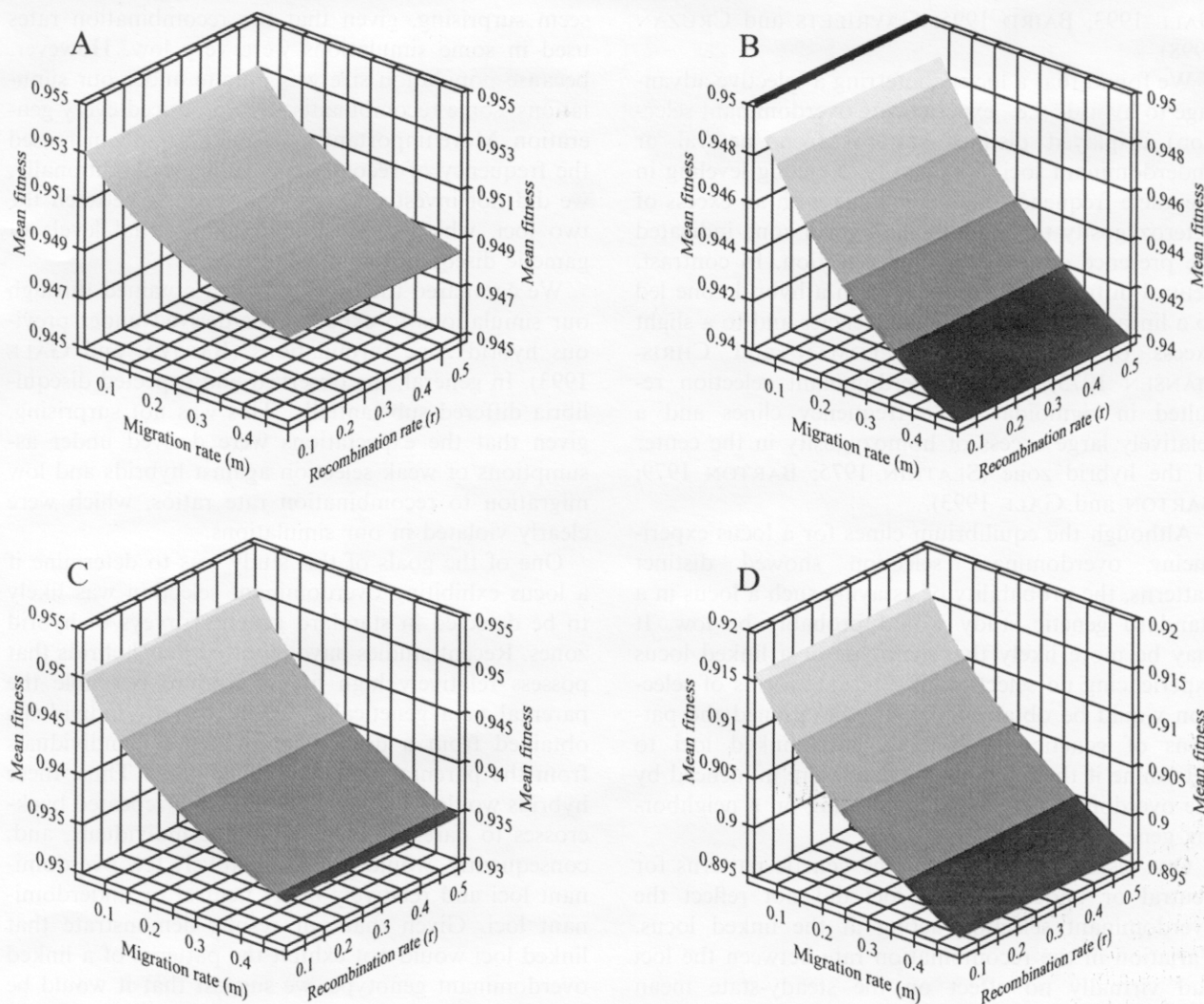
Variation in m affects the overall genetic structure

Fig. 4A–F. Effects of overdominant selection on locus A and underdominant selection on locus B on the equilibrium clines for **A**) the frequency of the A_1 allele, $\hat{p}$, **B**) the frequency of the B_1 allele, $\hat{q}$, **C**) the deviation from Hardy-Weinberg equilibrium at locus A, $\hat{F}_A$, **D**) the deviation from Hardy-Weinberg equilibrium at locus B, $\hat{F}_B$, **E**) the linkage disequilibrium between loci A and B, $\hat{D}$, and **F**) the normalized linkage disequilibrium between loci A and B, $\hat{D}'$, for varying recombination rates, r. The selection parameters in the central populations are $a = d = 0.9$, $b = c = 1.1$.

Fig. 5A–D. Equilibrium mean fitness, $\hat{\bar{w}}$, under four selection schemes for different migration rates, m, and recombination rates, r. The selection parameters in the central subpopulations are **A)** $a = d = 0.9$, $b = c = 1.0$, **B)** $a = d = 0.9$, $b = c = 1.1$, **C)** $a = d = 0.9$, $b = c = 1.8$, and **D)** $a = d = 0.1$, $b = c = 1$.

of the hybrid zone substantially. Relatively high migration rates lead to lower values of $\hat{\bar{w}}$ under all selection schemes (Fig. 5), because migration introduces unfit genotypes through the hybrid zone faster than selection can eliminate them. Moreover, the allele frequency clines with underdominant and overdominant selection become smoother and come to resemble those found under neutral diffusion (Fig. 2A) when m is large. In addition, the deviation from Hardy-Weinberg equilibrium becomes greater in more subpopulations of the hybrid zone. In contrast, the overall genetic patterns at neutral loci are virtually unaffected by variation in m.

Finally, we note that the different selection schemes also affect $\hat{\bar{w}}$ to some extent. Although it is difficult to make generalizations because of the complexity of the selection schemes considered, we note

that strong selection (e.g., Fig. 5C and 5D) tends to reduce overall mean fitness. However, even under these conditions, $\hat{\bar{w}}$ remains relatively high and never drops below 0.895.

DISCUSSION

We developed a novel two-locus model to examine the effects of hybrid superiority on the genetic structure of hybrid zones. Under our framework, selection operated in a manner similar to that proposed by MOORE (1977). Specifically, hybrid superiority was environment-dependent and confined to the middle of the hybrid zone. Our model thus differed substantially from previous models of hybridization that considered selection against hybrids only (e.g., BAZYKIN 1969; BARTON 1979; BARTON and

GALE 1993; BAIRD 1995; GAVRILETS and CRUZAN 1998).

We found that a locus conferring a selective advantage to hybrids (i.e., experiencing overdominant selection) displayed distinct patterns from neutral or underdominant loci. Specifically, a strong leveling in the allele frequency cline combined with an excess of heterozygosity throughout the hybrid zone indicated the presence of overdominant selection. In contrast, neutral diffusion of alleles through a hybrid zone led to a linear change of allele frequencies and to a slight excess of homozygotes (FELDMAN and CHRISTIANSEN 1975), while underdominant selection resulted in sigmoidal allele frequency clines and a relatively large excess of homozygosity in the center of the hybrid zone (SLATKIN 1975; BARTON 1979; BARTON and GALE 1993).

Although the equilibrium clines for a locus experiencing overdominant selection showed distinct patterns, the probability of assaying such a locus in a standard genetic study would probably be low. It may be more likely that genotype at a linked locus experiencing no selection or alternate forms of selection would be obtained. We thus examined the patterns of genetic variation at such linked loci to determine if their genetic patterns were influenced by the overdominant selection operating on a neighboring gene.

Our results indicated that the genetic patterns for neutral or underdominant loci did not reflect the overdominant selection acting at the linked locus. Variation in the recombination rate between the loci had virtually no effect on the steady-state mean fitness, allele frequency clines, or deviations from Hardy-Weinberg equilibrium. Consequently, linked neutral markers closer than 1 centimorgan to a selected locus may be required to detect a single locus experiencing overdominant selection.

In contrast, we found that variation in the migration rate of individuals through the hybrid zone did affect the genetic structure of the hybrid zone. Our results agree with previous theoretical studies that noted that dispersal rate plays a critical role in determining population genetic structure in hybrid zones. High migration tends to homogenize allele frequencies and swamp the effects of selection (SLATKIN 1973; NAGYLAKI 1975, 1976; ENDLER 1977; BARTON and HEWITT 1985). In our study, this was evidenced by a decrease in the equilibrium standardized mean fitness through the hybrid zone, because high migration tended to reduce the ability of selection to remove unfit genotypes.

Linkage disequilibrium within the hybrid zone was generally negligible, particularly when selection operated differently on the two selected loci. This may

seem surprising, given that the recombination rates used in some simulations were very low. However, because population size was infinite under our simulations, some recombinants were produced every generation. More importantly, selection rapidly increased the frequency of recombinant gametes. Additionally, we did not investigate cases of epistasis between the two loci, which could lead to substantial levels of gametic disequilibria.

We compared the disequilibrium obtained through our simulations to expectations derived under previous hybrid zone formulations (BARTON and GALE 1993). In general, the observed and expected disequilibria differed substantially. This was not surprising, given that the expectations were derived under assumptions of weak selection against hybrids and low migration to recombination rate ratios, which were clearly violated in our simulations.

One of the goals of this study was to determine if a locus exhibiting overdominant selection was likely to be detected in standard genetic surveys of hybrid zones. Recent studies have reported that hybrids that possess relatively high fitness tend to resemble the parental taxa genetically rather than F_1 individuals obtained from a direct mating between individuals from the parental taxa (ARNOLD 1997). Thus, these hybrids would presumably result from repeated backcrosses to parental or parental-like individuals, and, consequently, would possess relatively few overdominant loci and relatively many neutral or underdominant loci. Given that our results demonstrate that linked loci would not exhibit the patterns of a linked overdominant genotype, we suggest that it would be very difficult to detect evidence of this type of selection unless very many loci were assayed.

ACKNOWLEDGEMENTS

REFERENCES

Arnold ML, (1997). Natural hybridization and evolution. Oxford University Press, Oxford.

Arnold ML and Hodges SA, (1995). Are natural hybrids fit or unfit relative to their parents? Trends Ecol. Evol. 10: 67–71.

Arnold ML, Bulger MR, Burke JM, Hempel AL and Williams JH, (1999). Natural hybridization: How low can you go and still be important? Ecology 80: 371–381.

Baird SJE, (1995). A simulation study of multilocus clines. Evolution 49: 1038–1045.

Barton NH, (1979). The dynamics of hybrid zones. Heredity 43: 341–359.

Barton NH and Gale KS, (1993). Genetic analysis of hybrid zones. In: Hybrid zones and the evolutionary process, (ed. RG Harrison), Oxford University Press, New York, pp. 13–45.

Barton NH and Hewitt GM, (1985). Analysis of hybrid zones. Annu. Rev. Ecol. Syst. 16: 113–148.

Barton NH and Hewitt GM, (1989). Adaptation, speciation and hybrid zones. Nature 341: 497–503.

Bazykin AD, (1969). Hypothetical mechanism of speciation. Evolution 23: 685–687.

Cain ML, Andreasen V and Howard DJ, (1999). Reinforcing selection is effective under a relatively broad set of conditions in a mosaic hybrid zone. Evolution 53: 1343–1353.

Campbell DR, Waser NM and Wolf PG, (1998). Pollen transfer by natural hybrids and parental species in an Ipomopsis hybrid zone. Evolution 52: 1602–1611.

Durrett R, Buttel L and Harrison R, (2000). Spatial models for hybrid zones. Heredity 2000: 9–19.

Endler JA, (1977). Geographic variation, speciation, and clines. Princeton University Press, Princeton.

Feldman MW and Christiansen FB, (1975). The effect of population subdivision on two loci without selection. Genet. Res. 24: 151–162.

Felsenstein J, (1965). The effect of linkage on directional selection. Genetics 52: 349–363.

Gavrilets S, (1997). Hybrid zones with Dobzhansky-type epistatic selection. Evolution 51: 1027–1035.

Gavrilets S and Cruzan MB, (1998). Neutral gene flow across single locus clines. Evolution 52: 1277–1284.

Grant BR and Grant PR, (1996). High survival of Darwin's finch hybrids: Effects of beak morphology and diets. Ecology 77: 500–509.

Harrison RG, (1990). Hybrid zones: Windows on evolutionary process. In: Oxford surveys in evolutionary biology, (eds D Futuyma and J Antonovics), Oxford University Press, Oxford, pp. 69–128.

Harrison RG (ed), (1993). Hybrid zones and the evolutionary process. Cornell University Press, Ithaca.

Key KHL, (1968). The concept of stasipatric speciation. Syst. Zool. 17: 14–22.

Kimura M and Weiss GH, (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 49: 561–576.

Lewontin RC and Kojima K-I, (1960). The evolutionary dynamics of complex polymorphisms. Evolution 14: 458–472.

Moore WS, (1977). An evaluation of narrow hybrid zones in vertebrates. Quart. Rev. Biol. 52: 263–277.

Nagylaki T, (1975). Conditions for the existence of clines. Genetics 80: 595–615.

Nagylaki T, (1976). Clines with variable migration. Genetics 83: 867–886.

Slatkin M, (1973). Gene flow and selection in a cline. Genetics 75: 733–756.

Slatkin M, (1975). Gene flow and selection in a two-locus system. Genetics 81: 782–802.

Stebbins GL, (1965). Colonizing species of the native California flora. In: The genetics of colonizing species, (eds HG Baker and GL Stebbins), Academic Press, New York, pp. 173–195.

Weir BS, (1996). Genetic data analysis II. Sinauer associates, Sunderland.