

ASSORTATIVE MATING AND SPATIAL STRUCTURE IN HYBRID ZONES

Leithen K. M'Gonigle^{1,2} and Richard G. FitzJohn¹

¹*Department of Zoology, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada*

²*E-mail: mgonigle@zoology.ubc.ca*

Received October 30, 2008

Accepted August 25, 2009

The spatial genetic composition of hybrid zones exhibits a range of possible patterns, with many characterized by patchy distributions. While several hypothetical explanations exist for the maintenance of these "mosaic" hybrid zones, they remain virtually unexplored theoretically. Using computer simulations we investigate the roles of dispersal and assortative mating in the formation and persistence of hybrid zone structure. To quantify mosaic structure we develop a likelihood method, which we apply to simulation and empirical data. We find that long distance dispersal can lead to a patchy distribution that assortative mating can then reinforce, ultimately producing a mosaic capable of persisting over evolutionarily significant periods of time. By reducing the mating success of rare males, assortative mating creates a positive within-patch frequency-dependent selective pressure. Selection against heterozygotes can similarly create a rare-type disadvantage and we show that it can also preserve structure. We find that mosaic structure is maintained across a range of assumptions regarding the form and strength of assortative mating. Interestingly, we find that higher levels of mosaic structure are sometimes observed for intermediate assortment strengths. The high incidence of assortment documented in hybrid zones suggests that it may play a key role in stabilizing their form and structure.

KEY WORDS: Assortative mating, dispersal, hybrid zones, long distance, mosaic.

Hybrid zones provide a natural setting in which to study the effects of selection and gene flow on alleles or combinations of alleles (Barton and Hewitt 1985, 1989). Theoretical studies have provided methods to infer important quantities such as the strength of selection against hybrids and dispersal distance from empirical measurements of the spatial distribution of genotypes within a hybrid zone. Consequently, genetic spatial structure can be an informative property of a hybrid zone (Barton 1979; Barton and Hewitt 1985, 1989). To date the majority of theory has assumed that a monotonic change in genetic composition will be observed along a transect through a hybrid zone; that is, the hybrid zone will be "clinal" (Bazykin 1969; Barton and Hewitt 1985, 1989). However, a number of hybrid zones exhibit significant departures from a cline, with patches alternating in species composition (e.g., Howard 1986; Harrison and Rand 1989; Bierne et al. 2003). These "mosaic" hybrid zones are a spatial patchwork of populations, each fixed (or nearly fixed) for only one of the parental species' types.

Of the factors capable of generating mosaic patterns, habitat heterogeneity has received the most attention (Harrison and Rand 1989; Cain et al. 1999; Bridle et al. 2001; Bridle and Butlin 2002). This hypothesis assumes that the hybrid zone consists of alternating patches of different habitats and that the individuals preferentially occupy their own parent's habitat. The patchy species distribution then reflects the patchy environmental distribution. This hypothesis has been tested in a number of empirical studies (e.g., Howard and Harrison 1984a,b; Howard et al. 1993; Harrison and Bogdanowicz 1997; Bridle et al. 2001; Bridle and Butlin 2002; Vines et al. 2003). However, where ecological patterns reflecting the patchiness of the hybrid zone have been found, they often explain only a small proportion of the deviations from a clinal distribution (e.g., Bridle et al. 2001; Bridle and Butlin 2002; Bierne et al. 2002b).

Long distance dispersal during colonization has also been shown to be capable of creating patchy population structures (see Nichols and Hewitt 1994; Ibrahim et al. 1996; Le Corre et al. 1997;

Bialozyt et al. 2006). However, without any reinforcing mechanism, the spatial structure created by this process will tend to be transient, reverting to a cline over many generations. The pattern of dispersal on its own is not, therefore, sufficient to explain the persistence of structure observed in many hybrid zones. Several authors (e.g., Cruzan and Arnold 1993; Jiggins and Mallet 2000; Bridle and Butlin 2002; Bailey et al. 2004) have suggested that assortative mating may also contribute to stabilizing the observed mosaic distribution in hybrid zones, but this idea has yet to be theoretically tested.

In this paper we develop a simulation model to investigate the role of assortative mating and colonization patterns in the formation and maintenance of mosaic spatial structure within a hybrid zone without habitat heterogeneity. We also develop a likelihood method to quantify the level of “mosaicity” within a hybrid zone. We find that when dispersal during colonization contains sufficiently many long distance dispersers, assortative mating can reinforce the initial spatial distribution of colonists to enhance the mosaic pattern. This mosaic structure is then able to persist for prolonged periods of time, even in the face of high levels of dispersal.

Model Description

We consider the population dynamics of a one-dimensional hybrid zone using a computer simulation to investigate the role of migration and assortative mating in generating mosaic structure. Our simulation proceeds with discrete generations, tracking genotypes at two loci for diploid individuals along 30 ecologically identical patches, flanked on the left and right by pure patches of two different species. Individuals at the left boundary were assumed fixed for allele *a* at the first locus and allele *b* at the second, while those at the right were assumed fixed for alleles *A* and *B*. The *A*-locus is always assumed to affect assortative mating, while the role of the *B*-locus changes depending on our analyses. Recombination occurs between the loci at rate ρ . Each generation, individuals reproduce sexually within their patch to create the next generation and then die. The offspring then may migrate to a different patch, and new individuals migrate in from the pure edge populations. Hybrids are assumed to be completely viable and fertile for the majority of our analyses, but we also consider viability selection against hybrids through under-dominance at the *B*-locus. All interior patches were initially empty, and initial colonization occurred via dispersal from the edge populations.

REPRODUCTION & POPULATION GROWTH

We assume that population growth within each patch is logistic, with growth rate *r* and carrying capacity *K* (we used $r = 1.25$ and $K = 1000$ for the majority of our simulations). Each individual has a Poisson distributed number of offspring with mean

$$1 + r(K - n)/K$$

where *n* is the current patch population size. When viability selection acts, individuals heterozygous at the *B*-locus survive to reproduce with probability $1 + s$ (*s* being the strength of selection for or against hybrids). Assortative mating occurs through female preference. Females mate assortatively using the “best of *N*” scheme presented in Seger (1985). In this model a female surveys *N* males and then chooses a mate. For brevity we will often use the term “lek” to refer to the group of *N* males sampled by a female, even though this is not the precise meaning of the term. We chose the best of *N* mating scheme for its generality. Where $N = 1$, this model reduces to random mating for all female preference strengths, whereas at the other extreme where *N* is the whole patch, the model becomes equivalent to the “fixed relative-preference” scheme of Kirkpatrick (1982). We will use $N = frp$ to denote this case. Assortment usually occurs with respect to the *A*-locus, however, when we consider two-locus assortment (see below), then both loci are assumed to affect female preferences. We let $\psi_{i,j}$ denote the probability that a female of genotype *i* mates with a male of genotype *j* from among the males in her lek, and *c* denote the strength of a female’s preference, ranging from one (random mating) to infinity (mating only with preferred males, if present). $\psi_{i,j}$ can be interpreted as a combination of all prezygotic isolating factors, including both pre-mating (e.g., female choosiness) and post-mating, prezygotic factors (e.g., conspecific sperm precedence). We consider the following three models of assortative mating.

Linear-Preference (*k* loci)

While we investigate assortative mating based on at most two loci, we will describe a general version of this preference model for *k* loci. In this model preferences of parental species towards hybrids are intermediate between their preferences towards the pure types. If a female of type *i* and a male of type *j* share a total of *t* alleles across the *k* bi-allelic loci that underlie assortative mating (that is, both have *t* alleles from the same source population), then the probability that female *i* mates with male *j* in her lek is

$$\psi_{i,j} = \left(1 + \frac{t}{2k}(c - 1)\right) f_j \quad (1)$$

where f_j denotes the frequency of genotype *j* males in the female’s sample of *N*, and the $\psi_{i,j}$ ’s are standardized across all males. For example, in the one-locus case, if $c = 2$, then, if there were equal genotype frequencies, a homozygous female is twice as likely to mate with a conspecific than a heterospecific and 1.5 times as likely to mate with a heterozygous individual. A heterozygous female in this case would be 1.5 times as likely to mate with another heterozygote than with either con-specific.

Self-Preference

In this model a female prefers males of her own genotype by an equal factor (c) over all other males in her lek, discriminating equally against any other genotype. The probability that a female of genotype i mates with a male of genotype j is then given by

$$\psi_{i,j} = c^{\delta_{i,j}} f_j \tag{2}$$

where f_j again denotes the frequency of genotype j males in the female's sample of N , and $\delta_{i,j}$ is Kronecker's Delta, which is equal to one when $i = j$ and zero otherwise. Again the $\psi_{i,j}$'s are standardized across all males.

Dominant-Preference

In this model we assume that heterozygous males and females are indistinguishable from homozygous individuals characteristic of the source populations on the right side of the hybrid zone. Equation (2) then applies where $\delta_{i,j}$ is one if the phenotypes of i and j are the same.

DISPERSAL

There are two types of dispersal in our model: internal dispersal among the 30 patches and dispersal of new immigrants from the pure edge patches. In the internal patches, after reproduction each individual disperses with probability m . Individual dispersal distances were drawn randomly from a mixture of two exponential distributions, generating a leptokurtic (fat-tailed) dispersal kernel (Clark 1998):

$$pE(\mu_s) + (1 - p)E(\mu_l) \tag{3}$$

where $E(\mu)$ is an exponential distribution with mean μ . We assumed that $\mu_s < \mu_l$, so that p represents the proportion of short-distance dispersers. Dispersal distance was measured as the number of patches away from the focal patch, with an equal probability of being in either direction. Individuals that dispersed beyond the edge patches and into the pure source populations were assumed to have a negligible impact, and thus were removed from the system. In addition to local migration between patches, a fixed number I of individuals arrived each generation from each of the pure patches. These migrated according to the same dispersal kernel as above, relative to the edge of the pure patches.

ANALYZING SIMULATION RESULTS

To analyze the results of our simulation we developed a likelihood method to quantify the degree of "mosaicity" of a hybrid zone (i.e., deviation from a monotonic cline). We fit a series of horizontal steps to allele frequency data along a transect through the hybrid zone (Fig. 1). In order to compute the allele frequency within a patch we sampled every individual. A model with k steps is defined by k step locations $s = \{s_1, s_2, \dots, s_k\}$ and $k - 1$ step

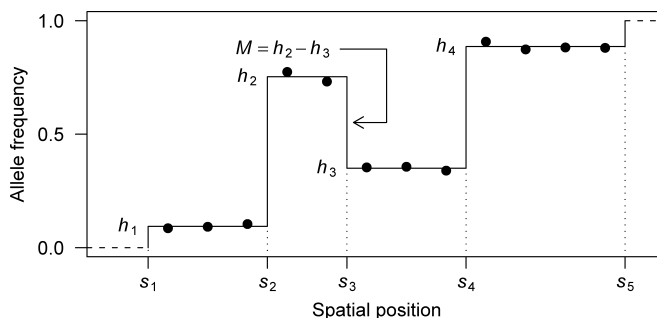


Figure 1. An example model fitted to hypothetical allele frequency data along a transect through a hybrid zone. This model has five fitted steps. The height between step s_i and step s_{i+1} is h_i . Because there is only a single down-step the mosaicity score, M , is simply the magnitude of this step (i.e., $h_2 - h_3$, see Eq. 4).

heights $h = \{h_1, h_2, \dots, h_{k-1}\}$, where h_i is the height between steps s_i and s_{i+1} . Because we assumed that the edge patches were fixed for each parental type, we added an initial step of height 0 and a final step of height 1 (Fig. 1). In appendix A we show how the likelihood of the observed allele frequency data given this model can be computed. For a set of data and given number of steps, we found the set of step locations and heights that maximize this likelihood and then used likelihood ratio tests to find the number of statistically significant steps required to best explain the data (see appendix A for details). A similar approach to ours was used by Macholán et al. (2008) in order to fit step models through data collected from the *Mus musculus musculus/M. m. domesticus* hybrid zone. However, Macholán et al. constrained their step heights to change monotonically through the hybrid zone, and thus their approach cannot be used to make inferences as to the level of mosaic structure a hybrid zone displays.

Once the step-wise model is fit to the allele frequencies, any measure of mosaicity may be investigated. Here, we focus on the sum of the magnitudes of the downward step sizes as a measure of the "mosaicity" of a data set, given by:

$$M = \sum_{i=1}^{k-2} \max(0, h_i - h_{i+1}) \tag{4}$$

(Fig. 1). This quantity has a minimum of zero for a clinal model (monotonically increasing steps), regardless of the steepness of the cline, and grows as the number and size of reversals in step height increases, attaining its maximum possible value when the hybrid zone consists of patches alternating between fixation on one or the other allele. For such cases (where each patch is fixed for either the a or the A allele), M is equal to the number of times the frequency of the A allele changes in frequency from 1 to 0 and back to 1 again. Because wider hybrid zones (e.g., those with more patches) have a greater number of possible locations where reversals in allele frequency may occur, M may be inflated

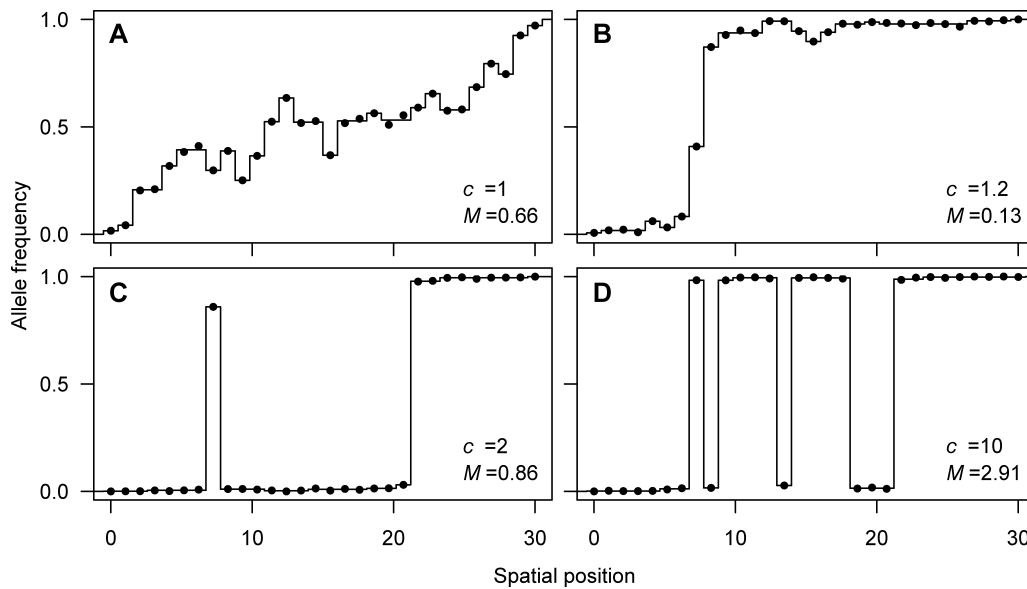


Figure 2. Example simulated hybrid zones with varying strengths of assortative mating (increasing across panels A to D), after 1000 generations. The most likely stepwise model is indicated by the line through the data. Assortative mating was based on the one-locus linear-preference model. Parameters were $\mu_s = 1$, $\mu_l = 15$, $m = 0.01$, $p = 0.75$, $r = 1.25$, $K = 1000$, $l = 50$, and $N = 5$.

in hybrid zones with finer levels of sampling. Accounting for the number of samples when computing M , however, would lead to the same hybrid zone having different mosaicism scores when sampled at different scales (an undesirable outcome). Thus we have assumed that the scale of sampling has been chosen so that it appropriately captures the spatial distribution of alleles across the hybrid zone, that is, no patches are missed during sampling. We have not, therefore, included the number of patches in our measure of mosaicism. Four sample best fit models with associated M values are shown in Figure 2.

Other measures of structure

Two additional measures that are commonly applied to quantify other aspects of genetic structure within a hybrid zone are linkage disequilibrium and bimodality. Linkage disequilibrium measures the association between alleles at different loci, and bimodality measures the lack of heterozygotes at a single locus, compared with the expectation under Hardy-Weinberg equilibrium (Jiggins and Mallet 2000). We also assess these measures in our simulations.

To compute linkage disequilibrium we use Lewontin's D' , which is linkage disequilibrium standardized by its maximum possible value given the observed allele frequencies (see Lewontin 1988, for details). This statistic varies from -1 to 1 , with -1 representing a population composed entirely of A/b and a/B genotypes, 1 representing a population composed entirely of A/B and a/b genotypes, and 0 representing a population with no associations between loci. To measure bimodality we use F_{IS} , as presented in Jiggins and Mallet (2000), which is computed as

$$F_{IS} = 1 - \frac{p_{Aa}}{2p_A p_a} \quad (5)$$

where p_{Aa} is the frequency of heterozygotes and p_A and p_a denote the frequencies of the two alleles. This statistic lies between -1 and 1 , with 0 implying that a population is at Hardy-Weinberg equilibrium, and 1 (-1) representing a complete lack of heterozygotes (homozygotes).

To compute both F_{IS} and D' for a given population we pooled individuals from all patches. Alternatively, we could have computed a patch average of each statistic. Because many patches were often fixed or nearly fixed for a single genotype, however, we found that such an approach did not provide meaningful results. In particular a mosaic hybrid zone consisting of patches alternating in state, from fixation on one allele to fixation on the other, would have a maximum mosaicism score, but in each patch both F_{IS} and D' would be undefined. At the other extreme, if hybrids never survive, F_{IS} and D' would be maximal within each patch, regardless of whether the mosaicism score were low or high. Thus, we see that the mosaicism score provides a distinct measure of the "patchiness" of the hybrid zone.

Results

Because our findings are largely insensitive to which model of assortative mating was used, we primarily present results from the one-locus linear-preference model, and will only explicitly state when this is not the case. We compute mosaicism for the one-locus linear model at the A -locus, which governs assortative mating (Fig. 2).

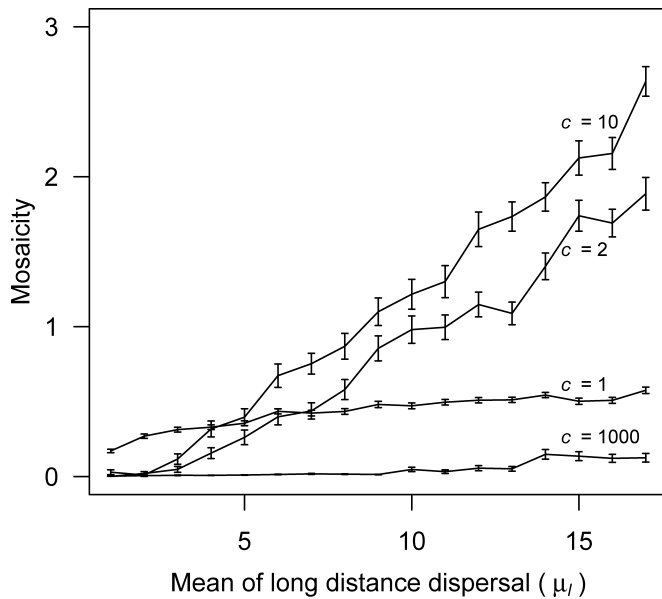


Figure 3. Mosaicity as a function of long distance dispersal distance (μ_l) in the one-locus linear-preference model. Each point is the mean of 100 replicate simulations. Different curves correspond to different female preference strengths (indicated by labels). Error bars denote \pm one standard error. Other parameters were as in Figure 2 with $N = frp$. Curves were plotted after 1000 generations.

We found that a combination of long distance dispersal and assortative mating allow mosaic population structure to form (Figs. 2, 3) and persist for long periods of time (Fig. 4A). The initial mosaic structure is generated by individuals leap-frogging over heterospecific populations during colonization to found new populations. These processes have been discussed in detail elsewhere (e.g., Nichols and Hewitt 1994; Ibrahim et al. 1996). In contrast with these previous studies, which assumed random mating, we found that once a mosaic population structure establishes, female preference acts to preserve the mosaic patterns (Fig. 4A). This is a consequence of a within-patch mating advantage to males of the more abundant female type, leading to within-patch fixation (or near fixation) on whichever mating type attains a higher frequency during the early stages of colonization. Once patches have reached carrying capacity, invasion by the other mating type becomes unlikely, and thus the final mosaic pattern persists, even with relatively high levels of between-patch dispersal ($m = 0.01$, $Nm \approx 10$).

The strength of assortative mating strongly affected the final level of mosaic structure observed in any particular hybrid zone (Fig. 2). Where mating was random, the constant arrival of foreign migrants biases patches on the left toward one allele and patches on the right toward the other, creating a gradual, roughly monotonic cline (Fig. 2A). With weak female preference mosaic structure was not stable over time. While sexual selection was a sufficiently strong force to reduce the extent of within patch

coexistence of the two mating types, it was not strong enough to prevent the occasional invasion by the rare mating type. The constant arrival of different migrant types from each end of the zone, combined with migration among patches eventually removes any traces of the founding population structure, producing a cline that was both steeper and more monotonic than when mating was random (Fig. 2B). Weak female preference thus resulted in slightly lower final mosaicity values than those obtained in the neutral case (Fig. 4A). For stronger assortative mating the hybrid zone became a mosaic of alternating patches of each genotype (Fig. 2C, 2D). Due to the near fixation of each patch on one allele type or the other, the mosaicity scores in these cases can be interpreted as approximately counting the number of complete reversals in patch genotype frequency (e.g., Fig. 2D has three nearly complete down-steps and a mosaicity score of 2.91).

The effects of assortative mating extended to other loci in the genome. Linkage disequilibrium between the assortative mating locus and an unlinked neutral locus persisted at higher levels with stronger female preference (Fig. 4B). Tighter linkage further increases the level of disequilibrium (Fig. 4B). Bimodality at both loci was also higher with stronger female preferences (Fig. 4C), achieving nearly the maximum possible value at the assortative mating locus when $c = 10$.

When migration distances were short, dispersal did not create a sufficient level of genetic patchiness for assortative mating of any level to reinforce, and thus no mosaics were observed. Increasing the mean of the long distance dispersal component (μ_l) led to a roughly linear increase in the observed level of mosaicity, when assortative mating was sufficiently strong enough to maintain a mosaic pattern (Fig. 3). To account for changes in mean dispersal distance that occur when increasing μ_l we ran a separate set of simulations where we decreased μ_s whenever we increased μ_l , in order to maintain a constant mean dispersal distance. Results remained qualitatively identical and thus have not been included. Similarly, increasing the proportion of individuals within a patch that disperse each generation (m), or the number of foreign migrants arriving each generation (I), sped up the rate at which colonization occurred, but did not qualitatively change results. Varying either population growth rate (r) or carrying capacity (K) also had little effect, and thus further exploration of these parameters has not been included.

The level of mosaic structure not only depended on the strength of female preference, but also on her lek size (Fig. 5A). Whenever mating was nonrandom (e.g., for lek sizes greater than 1), the level of mosaicity increased with female mate preferences from $c \approx 1.5$. For small leks mosaicity scores remained high throughout the entire range of strong female preferences (e.g., $N = 2, 5$ in Fig. 5A). Surprisingly, however, large leks displayed a decline in mosaicity with very strong female preferences (e.g., $N = 20, frp$ in Fig. 5A). This implies that when a female

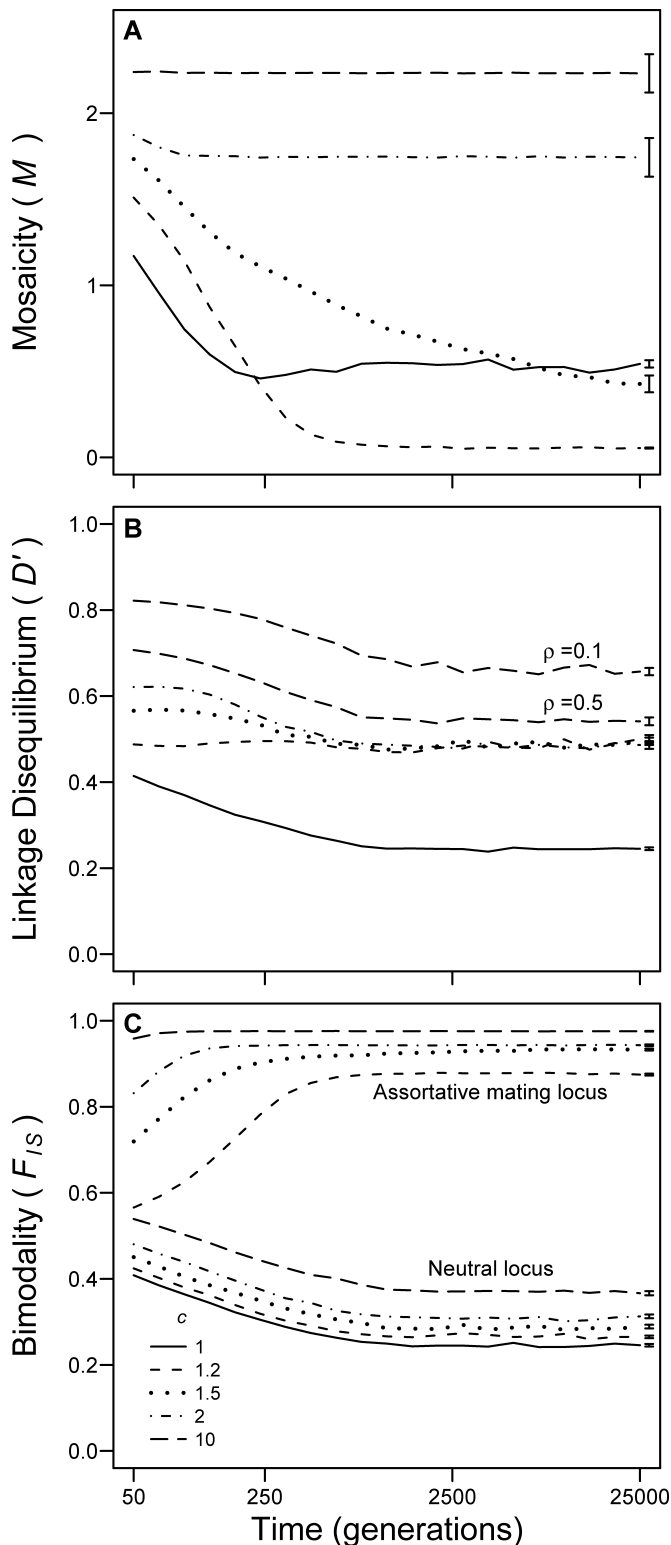


Figure 4. Mosaicity at the assortative mating locus (panel A), linkage disequilibrium between the assortative mating locus and a freely recombining neutral second locus (panel B) and bimodality at each locus (panel C) plotted over time. Each curve is the mean of 100 replicate simulations. Different curves correspond to different female preference strengths in the one-locus linear-preference model (the key in C applies to all panels). Recombination rates

samples only a small number of individuals from the population before mating, mosaic structure is more stable than when she has access to many males. This occurs because, when females only sample a few males, there is a high chance that the lek of a rare female will not contain any males of her preferred type and she will, therefore, mate with a more common male, which further reduces the expected mating success of the rare males. Conversely, when females sample many males, the lek of a rare female will often contain at least one male of her preferred type. A few males and females may, in this case, be sufficient to colonize a patch occupied by heterospecifics. At very strong mate preferences and large lek sizes the population structure reduces to a noisy cline, with different alleles often coexisting in the same patch. However, the mosaicity level is approximately 2–3 times that for random mating (Fig. 5A). The higher mosaicity score is due to a complete lack of heterozygotes, whose presence in the random mating case acts to reduce differences in allele frequencies between adjacent patches and thus the average step size. A drop in mosaicity values below those of the random mating case can be observed for intermediate mating preferences in the limiting case where a female samples the entire patch of males before mating (see Fig. 5 and the $c = 1000$ curve in Fig. 3). With intermediate c values, rare types are selected against, but not so strongly that they do not occasionally fix. The constant arrival of migrants from the boundaries thus eventually overwhelms any traces of the founding population structure, eventually creating a steep monotonic cline.

Changing the model of female preference did not qualitatively affect results (see Fig. 5B). However, with both the dominant and self-preference models the strength of female preferences required to overwhelm the mosaic structure was weaker than in the one-locus linear model (compare the $c \approx 10^2$ region in the dominant and self-preference models to the $c \approx 10^3$ region in the one-locus linear model). This occurs because the weak discrimination against hybrids combined with the reduced preference strength of hybrids in the linear model effectively reduces the efficacy of assortative mating for a given preference strength, compared to the other two models. Mean mosaicity scores were lower with the dominant-preference model, due to there being a slight bias toward patches fixing the dominant allele. In the two-locus linear model results were qualitatively unchanged when female preferences were weak. When preferences were stronger, however, curves appeared qualitatively more similar to those from the

were set to $\rho = 0.5$ (free recombination) for all curves except the topmost curve in panel B, which corresponds to $\rho = 0.1$. In panel C, under random mating ($c = 1$) both loci display nearly identical trajectories, and thus the A-locus has been omitted for clarity. Error bars denote \pm one standard error and are only shown at generation 25000 for clarity. Other parameters were as in Figure 2 with $N = frp$.

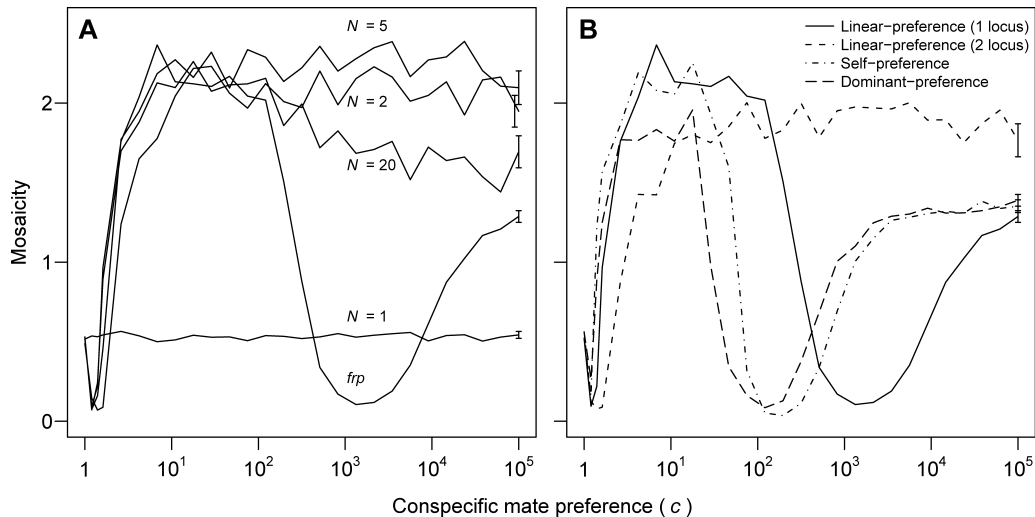


Figure 5. Mosaicity as a function of mate preference for different female lek sizes in the linear-preference model (A) and under different preference models with $N = frp$ (B). Each point is the mean of 100 replicate simulations. Error bars denote the standard error and are only plotted for the rightmost points. Other parameters were as in Figure 2 with $N = frp$ and $\rho = 0.5$. Simulations were run for 1000 generations.

one-locus model with a smaller lek size than used in the simulations. This shift is likely a consequence of recombination breaking down genotypes of rare individuals, making coexistence and thus invasion more difficult for that species.

So far we have treated hybrids as having equal fitness as parentals. In a similar manner to assortative mating, an intrinsic reduction in hybrid viability or fertility can create a frequency-dependent selective pressure against the rarer of the two parental species within a patch (Fig. 6). This is because each hybridization event represents a larger fraction of the rare species' matings, and thus each hybrid death reduces the fitness of the rare species by a proportionally larger amount. Barton and Whitlock (1997) showed that stabilizing selection on polygenic traits can similarly induce frequency-dependent selection against rare genotypes and thus, through an analogous process to that described above, can also lead to the maintenance of different allelic combinations between populations, provided migration rates are low enough. As expected, we found that as the strength of selection against heterozygotes increased a tighter correlation emerged between mosaicity scores at the assortative mating locus and the viability locus. When hybrids were lethal, recombination never occurred and the two loci were in essence completely linked. With strong assortative mating there was higher concordance between mosaicity scores at both loci, even when viability selection was relatively weak (Fig. 6C). The marginal decline in mosaicity observed for strong viability selection arises due to a slight reduction in the initial establishment of highly mosaic populations and does not represent a reduction in the stability of mosaics, once established.

In the simulation results presented, we assumed that allele frequencies were determined from the entire patch ($K = 1000$

individuals). The majority of empirical hybrid zone data most certainly contain many fewer samples. In order to investigate this sampling effect we sampled 100 individuals from each patch (without replacement) and then fit the best model to the sampled data. Results remained largely unchanged, with sampled data usually producing a model missing a few small steps but having a nearly identical mosaicity score.

Although we have only presented models consisting of 30 patches here, it is worth mentioning that the width of the cline or mosaic for any particular run often fluctuated well within the limits of these 30 patches, typically settling down to a much narrower final width than is possible with 30 patches. The resultant clines or mosaics were thus flanked on either side by large regions consisting of pure parental genotypes (e.g. Fig. 2). To assess whether the width of the hybrid zones would differ noticeably with more patches, we ran simulations with twice the number of patches (60). Indeed the hybrid zones in these cases settled into a comparable width as with 30 patches, but with larger flanking regions of pure parental species on either side. Thus our results are not likely to be affected by the number of patches in the hybrid zone.

APPLICATION TO *MYTILUS EDULIS* AND *M. GALLOPROVINCIALIS* HYBRID ZONE DATA

The smooth-shelled mussels *Mytilus edulis* and *M. galloprovincialis* form a mosaic hybrid zone that stretches around the coast of western Europe. While differences in temperature, salinity, and wave exposure affect species composition within the hybrid zone (Gardner 1994; Bierne et al. 2002b), Bierne et al. (2002b) argued that local adaptation was not sufficient on its own to explain the observed pattern in this hybrid zone. In a separate study Bierne et al.

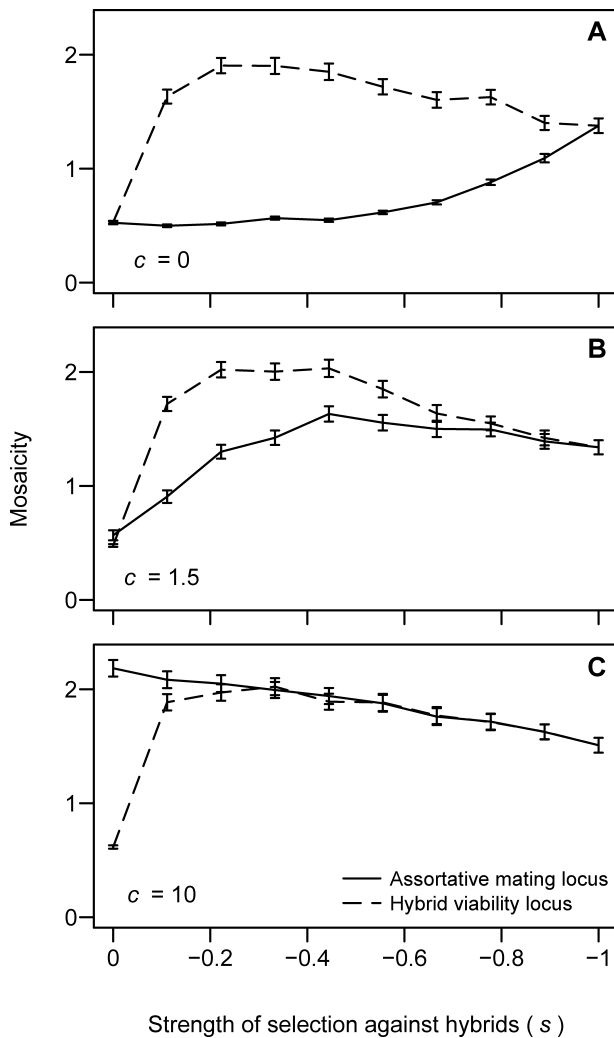


Figure 6. Mosaicity at each locus as a function of the strength of selection against hybrids in the one-locus preference model. Each curve is the mean of 250 replicate simulations. Different panels correspond to different female preference strengths (indicated by labels). Error bars denote the standard error. Other parameters were as in Figure 2 with $N = frp$ and free recombination ($\rho = 0.5$).

(2002a) suggested that the presence of assortative fertilization has likely contributed in maintaining the current population structure. To test our methods we fit models to three separate loci from this hybrid zone, using data presented in Bierne et al. (2003) (Fig. 7). Mosaicity scores and 95% bootstrap confidence intervals were 3.66 (3.08, 3.85) for *Glu-5'*, 2.43 (1.88, 2.70) for *mac-1'*, and 2.72 (2.14, 2.97) for *Efbis*. The mosaicity score for *Glu-5'* is significantly higher than that for both *mac-1'* and *Efbis*, despite similar sample sizes, suggesting that *Glu-5'* is possibly more closely linked to a locus influencing assortative mating or experiencing under-dominant selection. Furthermore, this difference between mosaicity scores demonstrates our method's ability to detect differences among loci from the same species with realistic levels of empirical sampling.

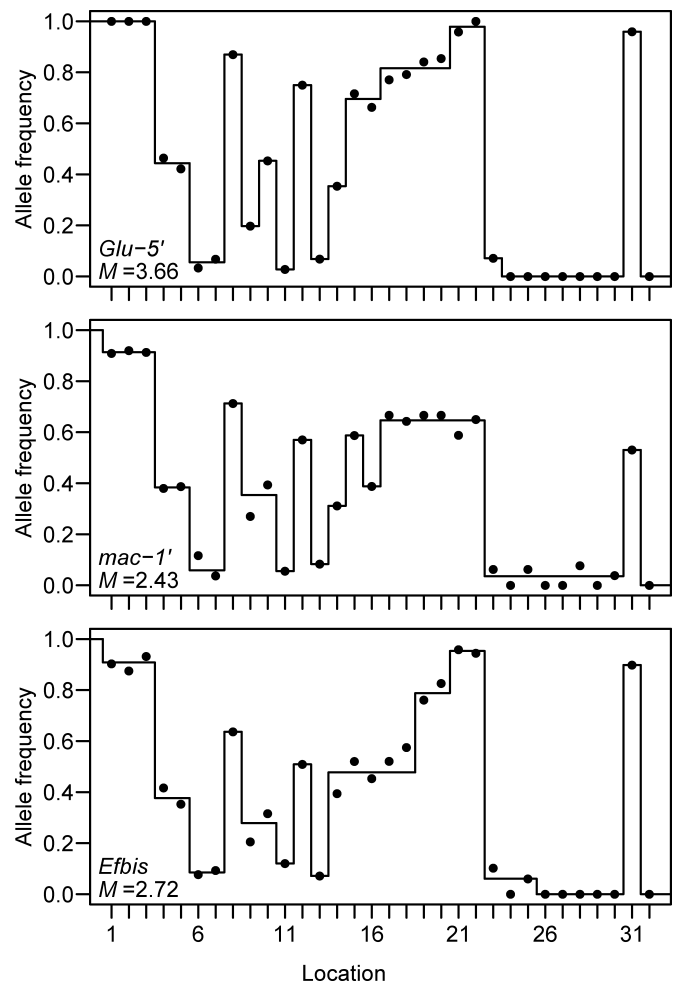


Figure 7. Best fit models for three diagnostic loci in the *Mytilus edulis*/*M. galloprovincialis* hybrid zone (data from Bierne et al. 2003). Mosaicity scores and 95% bootstrap confidence intervals were 3.66 (3.08, 3.85) for *Glu-5'*, 2.43 (1.88, 2.70) for *mac-1'*, and 2.72 (2.14, 2.97) for *Efbis*.

Discussion

Our results show that assortative mating, when coupled with long-distance dispersal during colonization, can lead to the stable persistence of mosaic patterns, even in the absence of ecological differences between incipient species. Furthermore, these results are robust to changes in a variety of assumptions about migration, female mating behavior, and the fitness of hybrids. It has been previously demonstrated that founder effects caused by long distance dispersal into vacant habitats during colonization can create spatially mosaic populations (Nichols and Hewitt 1994; Ibrahim et al. 1996), and it is this process that drives the initial mosaic patterns observed in our simulations. However, this process alone can not explain the long-term persistence of spatial structure in these populations (Fig. 4). Rather, assortative mating and/or hybrid inviability are essential to stabilize the mosaic structure in the absence of environmental heterogeneity.

We have further shown that the effects of assortative mating will likely carry over to other regions of the genome, even with high levels of recombination. The combination of continued immigration of pure *AB* and *ab* genotypes into the hybrid zone, and the slow decay of linkage disequilibrium which has been shown to characterize stepping-stone models (De and Duret 2007), led to a nonzero final value of D' , even when mating was random (Fig. 4). These equilibrium values of linkage disequilibrium were noticeably higher when female preference was stronger and/or when linkage to the assortative mating locus was tighter (Fig. 4B). Furthermore, higher levels of bimodality can be maintained at neutral loci with stronger female preferences (Fig. 4C). Sampling at a neutral locus will, therefore, not tend to reveal just how mosaic the hybrid zone may be at loci directly involved in assortative mating. However, such data can still potentially provide insight into whether a hybrid zone does have an underlying mosaic structure.

While our results indicate that assortative mating can stabilize mosaic hybrid zones for extensive periods of time, it is likely that other factors work in concert with this process. For example, habitat heterogeneity could strengthen the mosaic effect observed here. The combination of ecological differences between incipient species and a patchy environment could create a small degree of initial spatial segregation within a population. Assortative mating could then help push sub-populations towards fixation on one or the other type, whichever is locally more abundant. Similarly, in some cases the combined effects of viability selection against hybrids (which we have shown can also preserve mosaic structure) and assortative mating may allow for the preservation of a highly mosaic structure, where each force in isolation would not be sufficient to do so.

We have also presented a method that can be used to fit step-wise models through one-dimensional empirical hybrid zone data in order to objectively estimate their level of mosaicity. Our mosaicity statistic measures the number of reversals in allele frequency. It is not proposed as an alternative to the types of cline fitting that have been traditionally used in mosaic hybrid zones (e.g., see Bridle et al. 2001), but instead provides a complementary measure. To test our method we applied it to data from the *Mytilus edulis*/*M. galloprovincialis* hybrid zone (Bierne et al. 2003). Our best fit models exhibited high mosaicity scores at all three loci, demonstrating that it is informative when applied to empirical data sets. Interestingly, we found a statistically higher mosaicity score at *Glu-5'*, suggesting linkage to a locus involved in assortment, hybrid inviability, or ecological adaptation.

How common is assortment likely to be in hybrid zones? Assortative mating is thought to evolve as a consequence of divergence, with recently diverged species being more likely to hybridize (Felsenstein 1981; Coyne and Orr 1997). Bailey et al. (2004) documented strong assortative mating between the field

grasshoppers *Chorthippus brunneus* and *C. jacobsi*, which form a mosaic hybrid zone in northern Spain. The reported "isolation index" of $I = 0.59$ between these species corresponds to a c of 3.9 in our model. Bridle et al. (2006) have also recently argued that this assortative mating, through preferences for male song, likely plays an important role in maintaining the observed structure in this hybrid zone. Howard and Gregory (1993) conducted sperm competition experiments between the ground crickets *Allonemobius fasciatus* and *A. socius*, which form a mosaic hybrid zone in northeastern United States (Britch et al. 2001). When mated to both types of males *A. socius* females exhibited a conspecific sperm precedence of at least 95%, and *A. fasciatus* females of at least 98%. These translate into c values of approximately 42.5 and 49 respectively. The strength of assortative mating documented in both of these hybrid zones is sufficiently strong to lead to a high level of mosaicity in our model, provided that patchiness was initially present.

In some systems, however, assortative mating will likely play only a small role, if one at all. In the *Bombina bombina* – *B. variegata* (fire-bellied toad) hybrid zone, MacCallum et al. (1998) found that *B. bombina*-like hybrids were most often associated with pond habitats, whereas *B. variegata*-like hybrids were most often found in puddles. This strong habitat specialization explains most of the observed spatial variation, leaving little need for additional processes.

Mosaic hybrid zones may be more common than has been reported. The pattern observed in any particular hybrid zone may reflect the scale at which individuals are sampled (Schilthuizen 2000; Ross and Harrison 2002). When sampling is too coarse a mosaic pattern can appear clinal. Harrison and Bogdanowicz (1997) provided a simple characterization of hybrid zones, based on the shape of the genotypic distribution at the cline center. Mosaic hybrid zones tend to exhibit an overabundance of parental types, relative to hybrids and thus have a bimodal genotypic distribution. Resampling our mosaic hybrid zones on a coarser spatial scale would, on average, lead to a clinal pattern with a bimodal distribution of genotypes. This suggests that many bimodal hybrid zones may be mosaic at a finer spatial scale than measured. Jiggins and Mallet (2000) found a strong positive correlation between the measured strength of assortative mating and the level of bimodality in a survey of several empirical hybrid zones. While bimodality does not necessarily imply that a hybrid zone is mosaic (Cruzan and Arnold 1993; Emms and Arnold 1997), the study by Jiggins and Mallet (2000) suggests that an underlying correlation between mosaicity and assortativity may exist.

Understanding the mechanisms by which assortative mating occurs in hybrid zones would provide insight into the applicability of our model, as well as possibly reveal interesting theoretical extensions to other models of sexual selection. Our findings depend on sexual selection inducing positive frequency-dependent

selection; rare males are always at a disadvantage in the best of N mating scheme. This is not necessarily the case with other models of assortative mating. In a “grouping-based” model, females mate within some group of individuals with a fixed probability and otherwise mate with a male drawn randomly from the population (Felsenstein 1981; Otto et al. 2008). Because group membership may be frequency independent (e.g., groups could be chosen based on spatial or temporal proximity), this model does not necessarily induce a rare-type disadvantage, and thus we would expect it to yield a qualitatively different outcome.

Our model assumes an underlying demic system. Consideration of a similar model in continuous space may provide an interesting avenue for future research. In a homogeneous environment, where there are no patch boundaries stabilizing the sizes of pure (or nearly pure) populations, we would expect fluctuations in population size to lead to the eventual loss of mosaic structure. This may not be the case, however, in heterogeneous environments. Ecologically heterogeneous environments that favor some degree of local adaptation, or environments where the carrying capacity and thus the density of individuals varies in space, may stabilize population sizes of the different mating types, and thus allow for the preservation of mosaic structure. Temporal fluctuations in population size, or regularly occurring local extinctions followed by recolonization via long distance dispersal may also allow for the long term persistence of mosaic structure, although, in this case it would vary spatially in time.

Despite having received significant empirical attention, mosaic hybrid zones have remained largely unexplored in the theoretical literature, and the majority of models so far have assumed an underlying clinal structure. Both theoretical and empirical work have demonstrated, however, that clinal models do not always make accurate predictions about evolutionary processes occurring in a mosaic hybrid zone. For example, Cain et al. (1999) found that reinforcement evolved under a much wider set of circumstances for a mosaic hybrid zone than for a clinal one, and the empirical hybrid zones described in Bridle and Butlin (2002) and Cruzan and Arnold (1993) did not conform well to the clinal model expectations. Given that our current estimate of the prevalence of mosaic hybrid zones in nature is probably underestimated (because their detection is sensitive to the scale of sampling) and that predictions based on clinal models may not apply to mosaic hybrid zones, it is important that theory be developed to help understand the forces creating and maintaining mosaic hybrid zones and the effects that these mosaic hybrid zones have on evolutionary processes.

ACKNOWLEDGMENTS

We thank Evan Braswell, Tim Vines, Christina Burch, the OLM/SOWD groups and the anonymous reviewers for many helpful comments throughout the development of this manuscript. We thank Nicolas Bierne for

kindly sharing his data. We especially thank Sally Otto for her encouragement, insightful comments, and discussions. This work was supported by funding from the Natural Science and Engineering Research Council of Canada (CGS-D to LKM and Discovery Grant to Sally Otto) and by a University Graduate Fellowship from the University of British Columbia and the Capability Fund from Manaaki Whenua Landcare Research to RGF.

LITERATURE CITED

- Bailey, R. I., C. D. Thomas, and R. K. Butlin. 2004. Premating barriers to gene exchange and their implications for the structure of a mosaic hybrid zone between *Chorthippus brunneus* and *C. jacobsi* (Orthoptera: Acrididae). *J. Evol. Biol.* 17:108–119.
- Barton, N. H. 1979. The dynamics of hybrid zones. *Heredity* 43:341–359.
- Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.* 16:113–148.
- . 1989. Adaptation, speciation and hybrid zones. *Nature* 341:497–503.
- Barton, N. H., and M. C. Whitlock. 1997. The evolution of metapopulations. Pp. 183–210 in I. A. Hanski and M. E. Gilpin, eds. *Metapopulation biology: ecology, genetics, and evolution*. Academic Press, San Diego.
- Bazykin, A. D. 1969. Hypothetical mechanism of speciation. *Evolution* 23:685–687.
- Bialozyt, R., B. Ziegenhagen, and R. J. Petit. 2006. Contrasting effects of long distance seed dispersal on genetic diversity during range expansion. *J. Evol. Biol.* 19:12–20.
- Bierne, N., P. David, P. Boudry, and F. Bonhomme. 2002a. Assortative fertilization and selection at larval stage in the mussels *Mytilus edulis* and *M. galloprovincialis*. *Evolution* 56:292–298.
- Bierne, N., P. David, A. Langlade, and F. Bonhomme. 2002b. Can habitat specialisation maintain a mosaic hybrid zone in marine bivalves? *Mar. Ecol.-Prog. Ser.* 245:157–170.
- Bierne, N., P. Borsa, C. Daguin, D. Jollivet, F. Viard, F. Bonhomme, and P. David. 2003. Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. *Mol. Ecol.* 12:447–461.
- Bridle, J. R., and R. K. Butlin. 2002. Mating signal variation and bimodality in a mosaic hybrid zone between *Chorthippus* grasshopper species. *Evolution* 56:1184–1198.
- Bridle, J. R., S. J. E. Baird, and R. K. Butlin. 2001. Spatial structure and habitat variation in a grasshopper hybrid zone. *Evolution* 55:1832–1843.
- Bridle, J. R., C. I. Saldamando, W. Koning, and R. K. Butlin. 2006. Assortative preferences and discrimination by females against hybrid male song in the grasshoppers *Chorthippus brunneus* and *Chorthippus jacobsi* (Orthoptera: Acrididae). *J. Evol. Biol.* 19:1248–1256.
- Britch, M. L., M. L. Cain, and D. J. Howard. 2001. Spatio-temporal dynamics of the *Allonemobius fasciatus*—*A. socius* mosaic hybrid zone: a 14-year perspective. *Mol. Ecol.* 10:627–638.
- Cain, M. L., V. Andreasen, and D. J. Howard. 1999. Reinforcing selection is effective under a relatively broad set of conditions in a mosaic hybrid zone. *Evolution* 53:1343–1353.
- Clark, J. S. 1998. Why trees migrate so fast: Confronting theory with dispersal biology and the paleorecord. *Am. Nat.* 152:204–224.
- Coyne, J. A., and H. A. Orr. 1997. “Patterns of speciation in *Drosophila*” revisited. *Evolution* 51:295–303.
- Cruzan, M. B., and M. L. Arnold. 1993. Ecological and genetic associations in an iris hybrid zone. *Evolution* 47:1432–1445.
- De, A., and R. Durrett. 2007. Stepping-stone spatial structure causes slow decay of linkage disequilibrium and shifts the site frequency spectrum. *Genetics* 176:969–981.
- Emms, S. K., and M. L. Arnold. 1997. The effect of habitat on parental and hybrid fitness: Transplant experiments with Louisiana irises. *Evolution* 51:1112–1119.

- Felsenstein, J. 1981. Skepticism towards Santa Rosalía, or why are there so few kinds of animals? *Evolution* 35:124–138.
- Gardner, J. P. A. 1994. The structure and dynamics of naturally occurring hybrid *Mytilus edulis* and *M. galloprovincialis* (Bivalvia, Mollusca) populations: review and interpretation. *Arch. Hydrobiol.* 99 (supp):37–71.
- Harrison, R. G., and S. M. Bogdanowicz. 1997. Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. *Evolution* 51:493–505.
- Harrison, R. G., and D. M. Rand. 1989. Mosaic hybrid zones and the nature of species boundaries. Pp. 111–133 in J. A. Otte and D. Endler, ed. *Speciation and its Consequences*. Sinauer.
- Howard, D. J. 1986. A zone of overlap and hybridization between two ground cricket species. *Evolution* 40:34–43.
- Howard, D. J., and P. G. Gregory. 1993. Post-insemination signalling systems and reinforcement. *Philos. Trans. Roy. Soc. B Biol. Sci.* 340:231–236.
- Howard, D. J., and R. G. Harrison. 1984a. Habitat segregation in ground crickets – experimental studies of adult survival, reproductive success, and oviposition preference. *Ecology* 65:61–68.
- . 1984b. Habitat segregation in ground crickets – the role of interspecific competition and habitat selection. *Ecology* 65:69–76.
- Howard, D. J., G. L. Waring, C. A. Tibbets, and P. G. Gregory. 1993. Survival of hybrids in a mosaic hybrid zone. *Evolution* 47:789–800.
- Ibrahim, K. M., R. A. Nichols, and G. M. Hewitt. 1996. Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity* 77:282–291.
- Jiggins, C. D., and J. Mallet. 2000. Bimodal hybrid zones and speciation. *Trends Ecol. Evol.* 15:250–255.
- Kirkpatrick, M. 1982. Sexual selection and the evolution of female choice. *Evolution* 36:1–12.
- Le Corre, V., N. Machon, R. J. Petit, and A. Kremer. 1997. Colonization with long-distance seed dispersal and genetic structure of maternally inherited genes in forest trees: a simulation study. *Genet. Res.* 69:117–125.
- Lewontin, R. C. 1988. On measures of gametic disequilibrium. *Genetics* 120:849–852.
- MacCallum, C. J., B. Nurnberger, N. H. Barton, and J. M. Szymura. 1998. Habitat preference in the *Bombina* hybrid zone in Croatia. *Evolution* 52:227–239.
- Macholán, M., S. J. E. Baird, P. Munclinger, P. Dufková, B. Bímová, and J. Piálek. 2008. Genetic conflict outweighs heterogametic incompatibility in the mouse hybrid zone? *BMC Evolutionary Biology* 8:271.
- Nichols, R. A., and G. M. Hewitt. 1994. The genetic consequences of long-distance dispersal during colonization. *Heredity* 72:312–317.
- Otto, S. P., M. R. Servedio, and S. L. Nuismer. 2008. Frequency-dependent selection and the evolution of assortative mating. *Genetics* 179:2091–2112.
- R Development Core Team. 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- Ross, C. L., and R. G. Harrison. 2002. A fine-scale spatial analysis of the mosaic hybrid zone between *Gryllus firmus* and *Gryllus pennsylvanicus*. *Evolution* 56:2296–2312.
- Schilthuizen, M. 2000. Bimodal hybrid zones and the scale of a snail. *Trends Ecol. Evol.* 15:469.
- Seeger, J. 1985. Unifying genetic models for the evolution of female choice. *Evolution* 39:1185–1193.
- Vines, T. H., S. C. Kohler, A. Thiel, I. Ghira, T. R. Sands, C. J. MacCallum, N. H. Barton, and B. Nurnberger. 2003. The maintenance of reproductive isolation in a mosaic hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*. *Evolution* 57:1876–1888.

Associate Editor: M. Servedio

Appendix A

HYBRID ZONE STRUCTURE LIKELIHOOD METHOD

Here we describe a likelihood method for fitting a series of steps to allele frequency data, as in Figure 1. We consider a one-dimensional transect through a hybrid zone, with m “patches”. We assume an initial step height at zero, before the first patch, and a final step height at one, after the last patch. A stepwise model through the hybrid zone will consist of k step locations $s = \{s_1, s_2, \dots, s_k\}$ and $k - 1$ step heights (in addition to the first and last step heights fixed at zero and one), $h = \{h_1, h_2, \dots, h_{k-1}\}$, where h_i is the height between steps s_i and s_{i+1} . The stepwise model “partitions” the patches, and we will refer to a particular partition as the set of patches between two adjacent model steps. The step heights correspond to the estimated allele-frequency for each partition. Our method aims to quantify both the number and placement of these steps using a maximum likelihood approach.

Suppose that the expected genotype frequencies within the i th patch are $E_i[AA]$, $E_i[Aa]$, and $E_i[aa]$, and the i th patch has observed genotype counts of $x_{i,AA}$, $x_{i,Aa}$, and $x_{i,aa}$ for the genotypes AA , Aa and aa . The likelihood of sampling (with replacement) from the underlying frequencies is given by the multinomial probability

$$\binom{n_i}{x_{i,AA} \ x_{i,Aa} \ x_{i,aa}} E_i[AA]^{x_{i,AA}} E_i[Aa]^{x_{i,Aa}} E_i[aa]^{x_{i,aa}}$$

where $n_i = x_{i,AA} + x_{i,Aa} + x_{i,aa}$ is the number of individuals in the i th patch and $\binom{n}{x_1 \ x_2 \ x_3}$ is the multinomial coefficient defined as $n!/(x_1!x_2!x_3!)$.

The expected genotypic frequencies are a function of the allele frequencies given by the partition allele frequency and the inbreeding coefficient for the patch, f_i . Since f_i is not directly of interest we set it to its most likely value, given the local genotypic values; that is, we define f_i as

$$f_i = 1 - \frac{x_{i,Aa}}{2p_i(1 - p_i)n_i}$$

where p_i is the observed allele frequency in the i th patch. If the i th patch is within the j th partition, it shares that partition’s step height, h_j , and its expected frequencies are

$$\begin{aligned} E_i[AA] &= h_j^2 + f_i h_j (1 - h_j) \\ E_i[Aa] &= 2h_j(1 - h_j)(1 - f_i) \\ E_i[aa] &= (1 - h_j)^2 + f_i h_j (1 - h_j) \end{aligned}$$

In cases where an expected frequency was negative, that frequency was set to zero and the other expectations were standardized appropriately.

The likelihood of observing the data across all m patches can then be calculated as

$$\Pr(x | s, h) = \prod_{i=1}^m \binom{n_i}{x_{i,AA}, x_{i,Aa}, x_{i,aa}} \times E_i[AA]^{x_{i,AA}} E_i[Aa]^{x_{i,Aa}} E_i[aa]^{x_{i,aa}} \quad (\text{A1})$$

where the product is taken over all m patches. For a given set of step locations s we find the heights h that maximize equation (A1) using univariate optimization in R (R Development Core Team 2008).

For a given number of steps, k , equation (A1) must be maximized with respect to step locations. For large data sets it is not feasible to exhaustively search for the best model, as the number of possible models is on the order of $m!$ for m patches. Instead, we used a genetic algorithm to identify the best model for a given number of steps. The algorithm begins with a randomly generated initial pool of k -step models. It then runs through multiple generations of mutation, recombination and selection. Mutation

randomly replaces one or more steps within a model with other possible steps, recombination switches steps between different models (while maintaining step number), and selection samples the best models, weighted by their log-likelihood, to initiate the next generation.

To find the number of statistically significant steps required to best explain the data, we started with the best single step model ($k = 1$) and added steps until the difference between the best k step model and $k + 1$ step model was not statistically significant following a likelihood ratio test. Each step requires two additional parameters (a step location and a step height), so we compared the likelihood ratio with a Chi-squared distribution with two degrees of freedom. Simulations revealed that our use of two degrees of freedom was in fact a conservative assumption, while one degree of freedom was not. This method to fit mosaic hybrid zone data is available as an R package at <http://www.zoology.ubc.ca/prog/mosaic/>