

Population structure, fitness surfaces, and linkage in the shifting balance process

AVIV BERGMAN^{1,2*}, DAVID B. GOLDSTEIN^{1,2,3}, KENT E. HOLSINGER⁴,
AND MARCUS W. FELDMAN^{1,2}

¹Interval Research Corporation, 1801 Page Mill Rd. Building. C, Palo Alto, CA 94304

²Department of Biological Sciences, Stanford University, Stanford, CA 94305

³Present address: Department of Biology, Pennsylvania State University, University Park, PA 16802

⁴Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs CT 06268

(Received 22 November 1994 and in revised form 7 February 1995)

Summary

Wright first introduced the idea that random genetic drift and classical mass-action selection might combine in such a way as to allow populations to find the highest peak in complicated adaptive surfaces. His theory assumes large but structured populations, in which mating is spatially local. If gene flow is sufficiently low, and the subpopulations (demes) are small enough, they will be subject to genetic drift. Distant demes drift independently, allowing many independent searches of the adaptive surface to take place. A deme that has shifted to a higher peak can, by emigration, cause the rest of the demes to shift to the higher peak. The probability of this shift depends on the migration rate. Previous studies have concluded that very little migration is necessary to effect the shift in adaptive peaks that characterizes the last phase of Wright's Shifting Balance Process (SBP).

Here we present the results of a computer study that investigates the roles of dispersal distance, the degree of epistasis in the fitness surface, and recombination on the shifting balance process. In particular, we measure their effect on the population's mean fitness. We show that over a range of dispersal distances the advantage of the SBP is a monotonically increasing function of the amount of epistasis. Our results show that the extent of dispersal that results in the greatest effect of the SBP in increasing mean fitness depends on the extent of epistasis. Finally, for low levels of epistasis, higher recombination performs better, while for intermediate levels, lower recombination results in a greater advantage of the SBP.

1. Introduction

Fisher (1930) demonstrated that under certain idealized conditions, natural selection should cause population mean fitness to increase monotonically over time. Wright (1931, 1965) recognized that under these conditions, natural selection may fail to increase the mean fitness of a population if there is epistasis among loci. In the presence of epistasis, populations may become fixed on suboptimal genotypes. To describe this problem he introduced the metaphor of an 'adaptive landscape'. In one version of the adaptive landscape (Provine, 1986), the mean fitness of a population is plotted as a function of the frequencies of alleles at the loci controlling the trait. Fig. 1 shows such an adaptive landscape for a trait controlled by two loci with two alleles each. The classical view

depicted in this figure is a highly simplified view of the actual state of affairs since it ignores statistical association between the loci (linkage disequilibrium) and assumes random mating. In other words, the gene frequency axes should more properly be replaced by either chromosome or genotype frequencies.

In terms of the 'adaptive landscape', the classical models imply that selection can only move a population uphill. That is, a population with gene frequencies starting at the origin of Fig. 1, will move to peak A and stay there, even though peak B is higher. This result holds under a set of strong assumptions, including an effectively infinite population.

When populations are smaller, random changes in gene frequency (genetic drift) become important. These random effects may shift a population from the domain of attraction of one local maximum to that of another. Metaphorically, the population crosses the

* Corresponding author.

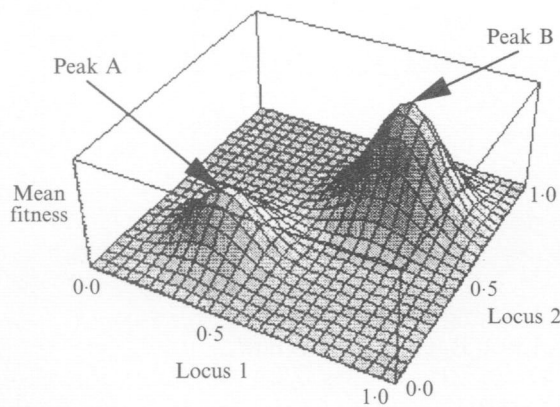


Fig. 1. An adaptive fitness landscape as defined by two interacting loci. Peak A is located where the allele frequencies at the two loci are each 0.3 and its altitude is half of that of peak B located where the allele frequencies at the two loci are 0.7.

valley of lower mean fitness separating the two 'adaptive peaks', thus bringing the population into the domain of attraction of the higher peak. In itself, reducing the population size is not a solution to the limitations recognized by Wright. For small populations selection is generally inefficient, and the population will wander all over the surface. For larger populations, however, the waiting time before a population crosses to a higher peak can be very large (Lande, 1985).

Wright first introduced the idea that random genetic drift and classical mass-action selection might combine in such a way as to allow populations to find the highest peak in complicated adaptive surfaces. His theory assumes large, but structured populations, in which mating is spatially local. If gene flow is sufficiently low, and the subpopulations (demes) are small enough, they will be subject to genetic drift. Distant demes drift independently (Malecot, 1969; Goldstein & Holsinger, 1992), allowing many independent searches of the adaptive surface. Thus, drift allows small populations to cross adaptive valleys, and genetic isolation of the populations increases the area of the adaptive surface that can be visited. This is the first phase of Wright's Shifting Balance Process (SBP) (Wright, 1931; Barton & Rouhani, 1993). In the second phase, mass-action selection carries the deme up the slope to the peak. That is, the most fit genotypes fix in the deme. In the third phase, the deme exports these successful genotypes to other demes resulting in the transformation of the entire population.

A sophisticated mathematical theory has been developed to describe the behaviour of small populations under natural selection (Wright, 1931; Kimura, 1964; Ewens, 1979), and clearly shows the feasibility of the first phase while providing a quantitative picture of its operation. The second phase is regular mass-action selection, described by some of the earliest models in population genetic theory (Haldane, 1932;

Fisher, 1930). The third phase has received considerably less attention, but has been recently studied both analytically (Barton, 1992) and by computer simulation (Crow *et al.* 1990). While some dispute remains as to why the third phase works (Crow *et al.* 1990; Barton, 1992; Moore & Tonsor, 1994), the general conclusion is that it does in fact work (see also Sumida & Hamilton, 1994). That is, a deme that has shifted to a higher peak can, through migration, cause the rest of the demes to shift to the higher peak. The probability of this shift depends on the migration rate. Crow *et al.* (1990) conclude from their deterministic numerical analysis of up to nine independent loci that very little migration is necessary to effect the shift in adaptive peaks that characterizes phase III.

The discussions pertaining to the success of the third phase of the SBP have made it clear that the problem must be considered as a whole. An integrated approach is needed especially because migration has opposing effects on the different phases of the SBP. The first and second phases require low or no migration between the different demes. When migration is high, the population can be assumed to be panmictic and mass-action selection prevents the discovery and attainment by at least one deme of the higher adaptive peaks. The effectiveness of the third phase, however, requires sufficient migrants from the most fit deme. Similar tradeoffs exist for other parameters such as deme size, strength of selection, and in our analysis, dispersal distance. In order to analyze the effects of these parameters, it is necessary to consider all three phases simultaneously.

Recently, two approaches have been used to study all three phases in concert. Barton & Rouhani (1993) considered an island model with many demes undergoing soft selection (i.e. the number of migrants contributed by each deme is independent of its genetic makeup). A diffusion approximation was used to determine the distribution of allele frequencies, with a second equation relating the state of the migrant pool to this distribution. Several assumptions were required for tractability: (1) migration does not affect the genetic variance; (2) selection does not depend on gene frequencies in the population; (3) selection is sufficiently weak relative to recombination that linkage disequilibrium can be neglected. They considered two genetic models: disruptive selection on a trait controlled by the additive contributions of many loci of equal effect, and underdominance at a single locus. Based on these two models, Barton & Rouhani conclude that adaptation is fastest at intermediate levels of gene flow. More specifically, they find: (1) a population can 'shift' to a higher peak even if this second peak is only slightly higher than the one at which the population starts; (2) migration must be below a critical value for the SBP to take place at all; (3) for very strong or very weak selection the SBP operates only very slowly, while for some intermediate strengths of selection it can operate much faster.

Moore & Tonsor (1994) use computer simulations to determine the range of migration rates that allows all three phases of the process to operate and to cause a shift in fitness peaks. They measure the population mean fitness at the middle and at the end of each simulation run (6000 and 12000 generations). Their model uses a pair of loci, with free recombination and the epistatic fitness matrix introduced by Crow *et al.* (1990). It was also assumed that growth in each deme follows the logistic equation, in which the intrinsic rate of increase and the carrying capacity are both independent of genotype. As the authors point out, these assumptions are not favourable to the SBP because most demes are usually near carrying capacity. Therefore, there is little potential for variation in the number of migrants coming from each deme.

These studies by Barton & Rouhani, and Moore & Tonsor, tend to support the importance of the SBP, and conclude that there is a moderately broad range of migration rates which can result in a shift between one fitness peak and another. Both sets of authors, however, point out that the importance of the SBP will depend on the details of the population demography. They also stress the importance of assumptions about the genetic system, including both the recombination system and the amount and type of epistasis, although neither study investigated these in any detail. The treatment by Barton & Rouhani (1993) is limited to either a purely additive genetic model with linkage equilibrium, or a single, underdominant locus. The treatment of Moore & Tonsor (1994) is restricted to a single epistatic fitness matrix for two loci undergoing free recombination.

Here, our primary focus is on the interaction of the SBP with the genetic system. The earlier studies have shown that the SBP can increase the rate of adaptive evolution as measured by changes in mean fitness. We would like to quantify this increase, and determine how it depends on the complexity of the fitness landscape. For this, we need to construct a series of landscapes that differ in their degrees of ruggedness. Furthermore, the ruggedness should be expressed at the level of the genotype and not the phenotype. For this purpose, we use a modified version of the *NK* model introduced by Kauffman (1989). In the following analysis we explore the operation of the SBP under this selection scheme with a model that includes isolation by distance at the level of individuals rather than the more commonly studied discrete populations. Our purpose is to move away from the classical analysis of peak shifts in simple two-peak models to more complex selection regimes, and to investigate how this complexity interacts with the degree of isolation by distance and recombination to affect the SBP.

2. The model

(i) *The NK model and an extension to diploids*

In the *NK* model, *N* is the number of loci and *K* the number of other loci influencing the fitness contribution of each locus. To calculate the fitness value of an allele at a specific locus, denoted here as the focal locus, one first determines the relevant genetic background for that locus; that is, the state of the *K* loci that influence alleles at that locus. In our implementation of the *NK* model, for a locus at position *i* the locations of these *K* loci were chosen independently of *i*. For every possible *K*-locus genotype the focal locus is given a particular fitness value and this value is stored. This process is repeated until all loci have been assigned fitness values in all possible genetic backgrounds. Once the fitness values are assigned, the fitness of a given (*N* locus) genotype is computed as the sum of the fitness values of each of the *N* loci, where these values depend on the genetic makeup of the relevant *K* other loci.

Kauffman introduced the model for the case of haploid genotypes, but a diploid version can be constructed by assuming that each of the possible genotypes at a diploid locus corresponds to an allele in the haploid case. That is, the *NK* fitness representation for *N* diploid loci with two alleles, A_1 and A_2 , corresponds to that for *N* haploid loci with three alleles, a_1 , a_2 and a_3 , where the homozygote A_1A_1 corresponds to a_1 , the heterozygote A_1A_2 corresponds to a_2 , and the other homozygote, A_2A_2 , corresponds to a_3 . For a given locus *i*, the *K* loci epistatic to it can assume 3^K possible genetic backgrounds for each of the three possible genotypes at the focal locus *i*. The fitness landscape, therefore, is represented by $3^{K+1}N$ values. Here we assume that the linkage phase does not affect a genotype's fitness.

In our simulations, we drew the fitness values of each locus in each genetic background from a uniform distribution. We chose five different types of landscapes, with $N = 10$ and $K = (0, 2, 4, 6 \text{ and } 8)$. It is clear that the case $K = 0$ represents no additive epistasis (that is, a purely additive fitness regime) because the contribution of a locus to the overall fitness at any location is not affected by the genetic composition of any other locus. As *K* increases in magnitude, the surface becomes increasingly rugged. Ruggedness can be described in a number of ways but one obvious measure is simply the number of mutational steps, at the genotypic level, separating local optima on that surface. Local optima on the landscape surface are defined as points where all one-step mutation neighbours have a lower fitness value. The number of mutational steps separating such local optima decreases as an approximately linear function of *K* (Kauffman, 1993), suggesting that the landscapes become more rugged in a regular way as *K* is increased.

For other characteristics of *NK* landscapes, see Kauffman (1993).

(ii) *Model dynamics*

Because we are interested in how much of an improvement the SBP provides, as a function of the shape of the fitness landscape, for each fitness regime we compare the rate of evolution as a function of geographical structure represented by a dispersal distance, d . Furthermore, we consider both the effect of recombination, r , and the full evolutionary dynamics. The mean fitness of a population is denoted by \bar{w} , and we are interested in the dynamic properties of \bar{w} as a function of K , d and r . We also define the ‘advantage of the SBP’ as the difference in mean fitnesses between structured and unstructured populations at the time we terminate the simulation (10000 generations):

$$\delta = \bar{w}_{\text{structured}} - \bar{w}_{\text{unstructured}}. \quad (1)$$

The simulation method is modeled after that in Goldstein & Holsinger (1992). Consider a population of 2000 diploid individuals arranged as a 1-dimensional array of cells. Each cell houses an individual. Goldstein & Holsinger (1992) showed that the qualitative patterns of genetic differentiation are similar in one and two dimensions; only the scale over which this differentiation occurs is different (see also Kimura & Weiss, 1964). We focus on a one-dimensional model to reduce the computation time. The fitness of each individual is based upon the diploid *NK* model as described above with two alleles at each locus, arbitrarily denoted 0 and 1. In every replicate, for each parameter set, all members of the initial population had allele 0 at every position. Thus populations began without variation. It should be noted, however, that there are no special points on the fitness surface. Thus, the all-zero genotype is, on average across realizations of the fitness surfaces, the same as any other genotype.

An offspring generation is constructed from the parental generation as follows: For each cell in the array, two parents are chosen at random from the set of locations that are within $\pm d$ cells of the focal cell, where d is the maximal dispersal distance. Once the parents are chosen, mendelian inheritance with recombination is simulated. The probability that a crossover event occurs on the 10-locus chromosome is r , and the location of the crossover is chosen randomly. Multiple crossover events are not allowed. Symmetric mutation occurs at the rate of 10^{-4} per locus in each direction. The probability that an offspring survives is based on its fitness as computed according to the *NK* model. If an offspring does not survive, new parents are chosen and the process is repeated until the cell is filled. Once all cells are filled, the offspring generation becomes the adults who will give rise to the next generation, and the process is repeated for 10000 generations. At every 10th

generation, we report the mean fitness of the population.

3. Results

We considered five maximal dispersal distances ($d = 1, 3, 5, 10$ and 1000). The first four distances are expected to result in spatially structured populations (Goldstein & Holsinger, 1992), while the final dispersal distance is equivalent to panmixia and involves no spatial structure. We also considered two recombination rates ($r = 0.01$ and 0.1) and five values of K (0, 2, 4, 6 and 8). With $r = 0.01$ and $r = 0.1$ the per-locus-pair recombination rates are roughly 0.001 and 0.01 respectively, much lower than is thought typical for randomly chosen pairs of eukaryotic genes. For each of the 50 parameter sets, we ran 25 independent simulations. Each simulation involved a different realization of the *NK* fitness surface. For example, with the parameter set $\{d = 1, r = 0.01$ and $K = 0\}$, 25 different realizations of an *NK* surface were constructed. Altogether we simulated 1250 independent realizations of the evolutionary process.

Fig. 2 shows the dynamic behaviour of population mean fitness for some of the parameter sets with $K = 0, 2, 6$ and 8. The curves show the average value among the 25 replications for each parameter set. Each replication is a different realization of the surface for the given K . Fig. 2a shows the results for all dispersal distances and $K = 0$. Notice first that the mean fitness in the unstructured population ($d = 1000$) increases much faster than in any of the structured populations. Furthermore, the fitness of the unstructured population is always greater than that of the population with $d = 1$. At the end of the simulation, the fitness of the unstructured population is higher than that of the most structured population ($d = 1$) by 4.3%. Thus, on the very smooth fitness surface with $K = 0$, implying no additive epistasis, unstructured populations achieve a higher mean fitness than tightly structured ones, even over very long time scales. In this case, all populations with $d > 1$ are essentially equivalent, and the tightly structured populations ($d = 1$) have lower mean fitness.

For rugged surfaces, however, we see that restricted dispersal distance can increase the rate of adaptation. Furthermore, for intermediate values of ruggedness, there appears to be an intermediate dispersal distance that maximizes the final mean fitness value. Fig. 2b shows the results for $K = 2$, and we see that an intermediate dispersal distance ($d = 5$), results in the highest mean fitness, 4.0% higher than in unstructured populations.

The importance of local mating is even greater, however, on the more rugged surface, as shown in Fig. 2c. For $K = 6$, although the unstructured population still evolves faster at first, after about 2000 generations all locally mating populations have a higher mean fitness than the unstructured population. This dif-

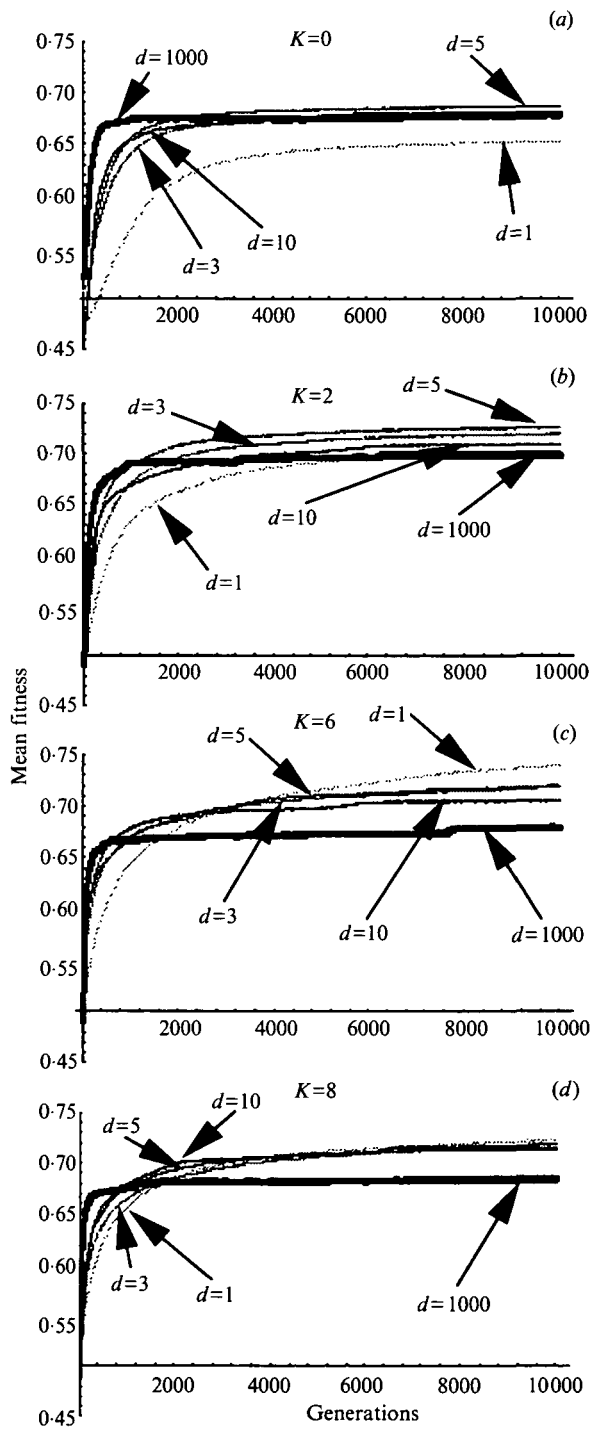


Fig. 2. Population mean fitness, averaged over 25 different randomly generated fitness landscapes (see text). (a) $K = 0$, (b) $K = 2$, (c) $K = 6$, (d) $K = 8$. Dispersal distances, from light to dark, are $d = 1, 3, 5, 10$ and 1000.

ference is maintained over the next 9000 generations and results in a mean fitness in the extremely structured population ($d = 1$) which is 8.6% higher at generation 10000 than that in the unstructured population. Also, note that for this degree of ruggedness, the most restricted dispersal distance ($d = 1$) results in the highest final mean fitness. In Fig. 2d we see that with $K = 8$, all locally mating populations ($d \leq 10$) behave

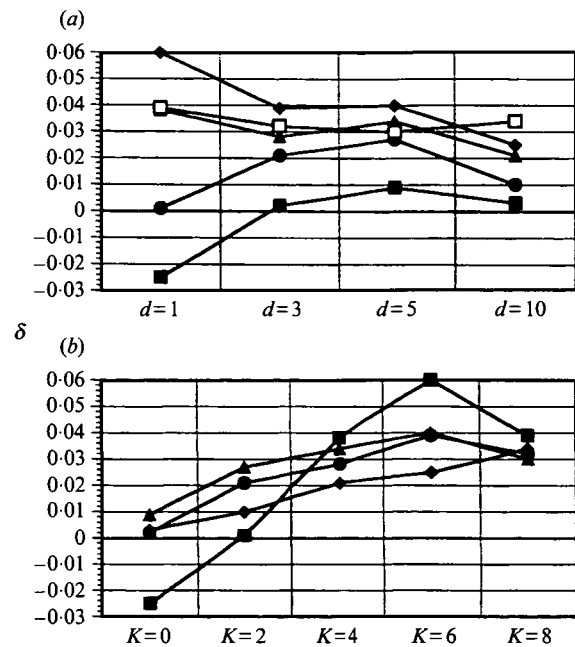


Fig. 3. Advantage of the SBP in populations with recombination rate $r = 0.01$ at generation 10000, as defined by $\delta = \bar{w}_{\text{structured}} - \bar{w}_{\text{unstructured}}$ for the set of (a): K values 0 (■), 2 (●), 4 (▲), 6 (◆) and 8 (□) and (b): dispersal distances $d = 1$ (■), 3 (●), 5 (▲) and 10 (◆).

similarly, but these are different from the panmictic one ($d = 1000$).

In order to summarize the consequences of the SBP over a biologically meaningful period of time, we compared the final mean fitness (namely, the mean fitness after 10000 generations) in the four structured populations to that observed in the unstructured population for each of the five types of fitness surfaces and each of the two recombination rates. This measure, defined earlier as δ , directly quantifies the advantage provided by the SBP. Fig. 3 shows that this advantage increases with the ruggedness of the fitness surface up to $K = 6$. This finding holds for all dispersal distances considered. Notice in particular, in panel (b) of Fig. 3, that for $d = 1$ the advantage of the SBP is an approximately linear function of K up to $K = 6$. It should be noted that for $K = 8$, δ remains positive but no longer greater than the values observed for smaller K s. That linear increase ceases at $K = 6$ probably implies that once the surface becomes sufficiently rugged, even structured populations cannot find the highest peaks. Fig. 4 represents the same surface as Fig. 3, but for $r = 0.1$. Although the patterns here are harder to interpret, we can say that recombination has a large effect on the SBP, and that it interacts with K .

Three statistical analyses of the values of \bar{w} at 10000 generations were carried out for each of $r = 0.01$ and $r = 0.1$. First, for each value of K , the values $d = 1, 3, 5, 10$ and 1000 were regarded as treatments for which each of the \bar{w} values from the 25 realizations of the fitness surface for that K were observations. The overall effect of d was then evaluated using the

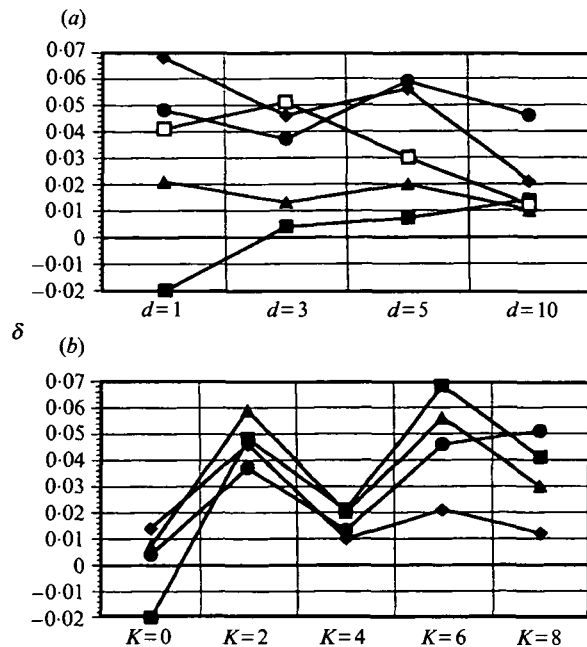


Fig. 4. Advantage of the SBP in populations with recombination rate $r = 0.1$ at generation 10000, as (a): defined by $\delta = \bar{w}_{\text{structured}} - \bar{w}_{\text{unstructured}}$ for the set of K values 0, 2, 4, 6, 8, and (b): dispersal distances $d = 1, 3, 5$ and 10 (symbols as in Fig. 3).

Table 1. Statistical results as a function of ruggedness

	$r = 0.01$		$r = 0.1$	
	K-W ^a	F ^b	K-W ^a	F ^b
$K = 0$	4.65	5.28	3.60	4.10
$K = 2$	6.66	4.03	15.47**	14.34**
$K = 4$	10.82*	14.24**	3.22	2.72
$K = 6$	24.26†	21.89†	38.43†	29.06†
$K = 8$	13.36**	10.46*	19.58†	17.70***

The null hypothesis is that \bar{w} does not depend on d .
^a Kruskal–Wallis value of H (see Sokal & Rohlf, 1995, p. 424).

^b Value of χ^2 for Friedman’s test (see Sokal & Rohlf, 1995, p. 441).

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.005$; † $P < 0.001$.

Kruskal–Wallis procedure (Sokal & Rohlf, 1995, p. 424). These results are reported in the first and third columns of Table 1 where, with $r = 0.01$, the statistical effect of d increases in magnitude as K increases up to $K = 6$. Furthermore, with $r = 0.1$ we cannot explain why there is a significant effect of d at $K = 2$ and $K = 6$, but no such effect at $K = 4$.

A bivariate analysis was also carried out with each of the 25 realizations for each K regarded as a random effect in an analysis of the role of the treatment represented by d using Friedman’s procedure (Sokal & Rohlf, 1995, p. 441). The second and fourth columns of Table 1 report the values of Friedman’s test statistic. Here $K = 4, 6$ and 8 produce significant

effects of d for $r = 0.01$ while $K = 2, 6$ and 8 do so for $r = 0.1$.

A third analysis involved non-parametric multiple comparisons among the d s using the Mann–Whitney approach (Sokal & Rohlf, 1995, p. 432). Here, as in the Kruskal–Wallis tests, for each K the 25 values of \bar{w} taken for every d were compared pairwise among the d s. It is not necessary to give all the details here; there were no surprises. With $r = 0.01$ and $K = 6$, dispersal values $d = 1, 3, 5$ and 10 were all significantly different from $d = 1000$, and $d = 1$ differed from $d = 10$. For $r = 0.1$, $d = 1, 3$ and 5 differed significantly from $d = 1000$, but $d = 10$ did not. In this case, $d = 1, 3$ and 5 were also different from $d = 10$. None of the other K s produced nearly as many significant pairwise differences among the d s as did $K = 6$.

To analyse the effect of different evolutionary trajectories for a single landscape we accumulated simulation results of 25 different runs for a single randomly chosen realization of cases $K = 2$ and $K = 6$ respectively. We examined all dispersal values, $d = 1, 3, 5, 10$ and 1000 , with recombination rate $r = 0.01$, and mutation rate 10^{-4} . The overall effect of d was then evaluated using the Kruskal–Wallis procedure (Sokal & Rohlf, 1995, p. 424) separately for the single randomly chosen realizations of $K = 2$ and $K = 6$, respectively. In each case, the d values were regarded as treatments, each producing 25 observations. Thus, for the chosen $K = 2$ case there were 125 observations and the same for the $K = 6$ case. In this analysis, both examples of $K = 2$ and $K = 6$ produced statistically significant effects of d , with the $K = 6$ example providing the stronger result. We again used a Mann–Whitney multiple comparison procedure among the d s, and this demonstrated that structured populations did better than the unstructured population at both levels of ruggedness. An anomalous result was observed in the $K = 6$ example, where the intermediate dispersal distance, $d = 10$, was significantly different from all other dispersal distances. This anomaly may, of course, be due to the particular realization of the landscape. Further examples of each K value should be studied although this is a very time-consuming procedure¹. However, the general form of the results for these two cases studied in detail supports the conclusions deduced above from Table 1.

Fig. 5 shows the difference between the curves in Figs 3 and 4, reflecting the difference in the advantage of the SBP for the two values of r . It appears that the effect of recombination on the magnitude of the advantage of the SBP depends on the degree of ruggedness. For $K = 2$, the higher recombination rate ($r = 0.1$) results in a greater advantage, i.e. $\delta|_{r=0.1} > \delta|_{r=0.01}$ (see definition (1)), while for $K = 4$ the situation is reversed. For even more rugged surfaces, however, recombination appears to have little effect on average.

¹ We restricted our analysis to a limited set of parameters due to computing limitations. Each parameter setting took the equivalent of about 24 h of an SGI R4400 CPU time.

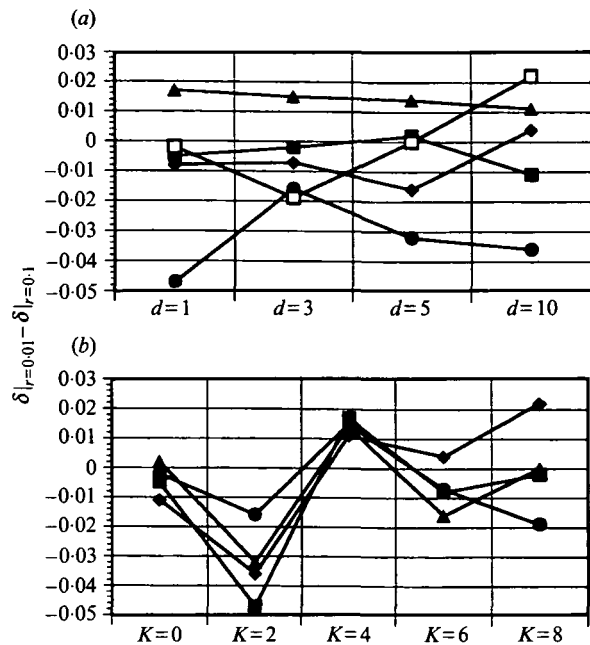


Fig. 5. Difference between the advantage of the SBP in populations with recombination rates $r = 0.01$ and $r = 0.1$. Consistent behaviour can be observed only at intermediate levels of ruggedness, $K = 2$ and 4 (symbols as in Fig. 3).

There is, however, a strong interaction with the dispersal distance. For example, at $K = 8$ and $d = 10$ higher recombination hinders the SBP, while the situation is reversed for $d = 3$. This behaviour is difficult to interpret, but an explanation of the general non linearity in the dependence on r may be that recombination affects two aspects of evolution on rugged surfaces in opposing ways. First, higher recombination decreases the waiting time until the appearance of new multi-locus genotypes (Christiansen *et al.*, 1995). Secondly, recombination breaks up adapted gene complexes, and, if the epistasis among the loci is strong, the recombinant genotypes will be selected against in the genetic backgrounds of those demes that have not yet shifted. In this case, fit demes cannot export their superior genotypes and the SBP will not operate as well. Some of the results observed here would follow if, for sufficiently small K , the effect of recombination on the generation of novel multilocus genotypes dominates, and recombination results in an ultimately higher mean fitness. For larger K , the effect of recombination in retarding the emigration of superior chromosomes becomes more important, and recombination has the opposite effect. For sufficiently rugged surfaces there is too great an interaction with dispersal distance to allow a general description of the effect of recombination.

4. Discussion

Wright's shifting balance process facilitates the discovery of new adaptive peaks over a moderate range of conditions, both in terms of the migration

parameters and the characteristics of the fitness landscape. In this work we focused on the dynamics of the SBP and the advantage that it may provide in facilitating a population's attainment of a globally optimum mean fitness. We have shown that the degree of population structure and the ruggedness of the fitness surface together influence the population's mean fitness and the rate at which higher mean fitness is attained. The use of the NK model (Kauffman, 1989) as a means of generating the adaptive landscape allowed us to create fitness surfaces that are tunably rugged at the genetic level. Thus, ruggedness can be characterized by the number of mutational steps separating local optima. This way of representing fitness constitutes a departure from the usual analysis of shifts between two peaks. Here, there may be many peaks and the population might transit through several peaks during the dynamics that we observe. We have shown that over a range of dispersal distances, the advantage of the SBP is a monotonically increasing function of the ruggedness, and increases approximately linearly with K up to a critical level of complexity ($K = 6$). This result has two important implications: First, it suggests that K is a reasonable measure of ruggedness and has a predictable influence on the dynamics of mean fitness in structured and unstructured populations. Second, for intermediate levels of ruggedness, the SBP substantially increases the level of improvement in the population's mean fitness.

How population structure influences the effect of the SBP has also been examined. In our framework, migration and deme size have been combined in a single parameter, the dispersal distance. Our results show that the extent of dispersal that results in the highest mean fitness after 10000 generations depends on the degree of ruggedness, K .

In the discussion of the role of migration, Moore & Tonsor attempt to separate the operation of the different phases of the SBP. Their computations produced an unexpected non-monotonicity in the effect of migration on the probability that a deme finds and climbs the higher peak (that is, phases one and two). In their figure 2, the broad effect of migration is to retard the operation of phases one and two, but for the low migration rates, an increased rate briefly leads to an apparent increase in the operation of phases one and two. We suspect that this non-monotonicity is due to a confounding of phases one and two with phase three. Moore & Tonsor report the percentage of those simulations in which one or more demes have made the switch to the higher peak after 6000 generations. They do not report the number of simulations in which a deme finds the higher peak at some point during the evolution. We suspect that in many cases, demes find the higher peak but are unable to stay there through generation 6000 because of migration in exactly those cases where phase three does not operate. Barton & Rouhani's arguments

about the power of migration are also consistent with this interpretation.

Finally, our results show that the effect of the recombination rate on the SBP is complicated, even for the low recombination rates studied here. For low levels of ruggedness, high recombination performs better ($K = 2$), while for intermediate levels, low recombination results in a greater advantage of the SBP ($K = 4$). For more rugged surfaces, there is little consistent effect of the recombination rate. To draw more definite conclusions concerning the effect of recombination on the SBP, a larger range of recombination values must be considered as well as other ways of realizing rugged landscapes (e.g. Bergman & Feldman, 1992). These studies are underway. The interaction between K and r in influencing the advantage of the SBP suggests that it would be interesting to investigate the fate of genes that modify the recombination rate in populations subject to this form of epistasis (see also Bergman & Feldman, 1992). In particular, it is important to determine whether the direction of selection on the rate of recombination depends on an interaction between the degree of ruggedness and the dispersal distance.

The authors are grateful to Dr Mony Slatkin, and an anonymous reviewer for helpful comments on an earlier version of the manuscript. This research was supported in part by NIH grant GM 28016 and NASA grant NC 2-778.

References

- Barton, N. H. (1992). On the spread of new gene combinations in the third phase of Wright's shifting balance. *Evolution* **46**, 551–557.
- Barton, N. H. & Rouhani, S. (1993). Adaptation and the 'shifting balance'. *Genetical Research* **61** (1), 57–74.
- Bergman, A. & Feldman, M. W. (1992). Recombination dynamics and the fitness landscape. *Physica D* **56**, 57–67.
- Christiansen, F. B., Otto, S. P., Feldman, M. W. & Bergman, A. (1994). Waiting with and without recombination: The time to production of a double mutant. (in preparation).
- Crow, J. F., Engels, W. R. & Denniston, C. (1990). Phase three of Wright's shifting-balance theory. *Evolution* **44** (2), 233–247.
- Ewens, W. J. (1979). *Mathematical Population Genetics*. Berlin: Springer-Verlag.
- Fisher, R. A. (1930). *The Genetic Theory of Natural Selection*. Oxford: Clarendon Press.
- Goldstein, D. B. & Holsinger, K. H. (1992). Maintenance of polygenic variation in spatially structured populations: Roles for local mating and genetic redundancy. *Evolution* **46** (2), 412–429.
- Haldane, J. B. S. (1932). *The Causes of Evolution*. New York: Harper [Reprinted, 1966, Cornell University Press, Ithaca].
- Kauffman, S. A. (1989). Adaptation on rugged fitness landscapes. pp. 527–618. In *Lectures in the Sciences of Complexity*. (ed. Stein, D. L.), Redwood City, California: Addison-Wesley.
- Kauffman, S. A. (1993). *The Origin of Order*. Oxford: Oxford University Press.
- Kimura, M. (1964). Diffusion models in population genetics. *Journal of Applied Probability* **1**, 177–232.
- Kimura, M. & Weiss, G. H. (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**, 561–576.
- Lande, R. (1985). The fixation of chromosomal rearrangements in subdivided population with local extinction and recolonization. *Heredity* **54**, 323–332.
- Malecot, G. (1969). *The Mathematics of Heredity*. San Francisco: W. H. Freeman and Company.
- Moore, F. B. G. & Tonsor, S. J. (1994). A simulation of Wright's shifting-balance process: migration and the three phases. *Evolution* **48** (1), 69–80.
- Provine, W. B. (1986). *Sewall Wright and Evolutionary Biology*. Chicago: The University of Chicago Press.
- Sokal, R. R. & Rohlf, F. J. (1995). *Biometry*, 3rd edition. New York: Freeman and Co.
- Sumida, B. H. & Hamilton W. D. (1994). Both Wrightian and "parasite" peak shifts enhance genetic algorithm performance in the travelling salesman problem. pp. 254–279. In *Computing with Biological Metaphors*. (ed. Paton, R.), London: Chapman and Hall.
- Wright, S. (1931). Evolution in mendelian populations. *Genetics* **16**, 97–156.
- Wright, S. (1965). Factor interaction and linkage in evolution. *Proceedings of the Royal Society of London Series B* **162**, 80–104.
- Wright, S. (1982). Character change, speciation and the higher taxa. *Evolution* **36**, 427–443.