Large Genetic Change at Small Fitness Cost in Large Populations of *Drosophila melanogaster* Selected for Wind Tunnel Flight: Rethinking Fitness Surfaces

K. E. Weber

Department of Biological Sciences, University of Southern Maine, Portland, ME 04103

Manuscript received March 13, 1995
Accepted for publication June 8, 1996

ABSTRACT

The fitness effects of extreme genetic change by selection were studied in large populations subjected to prolonged, intense selection. Two replicate populations of *Drosophila melanogaster*, with estimated effective sizes $500 \leq N_e \leq 1000$, were selected for increased performance in a wind tunnel, selecting on average the fastest 4.5% of flies. The mean apparent flying speed of both lines increased from ~2 to 170 cm/sec and continued to respond at diminishing rates, without reaching a plateau, for 100 generations. Competitive fitness tests in generations 50 and 85 showed minimal or no fitness loss in selected lines compared to controls. Sublines relaxed in generations 65 and 85 showed minimal or no regression in apparent flying speed. Hybrid lines, from a cross of selected × control lines in generation 75, responded to reselection saltationally, showing that the chromosomes of the selected lines had been assembled from alleles at many loci, from many different chromosomes in the base population. Thus, major genetic change was achieved, but without the costs usually associated with strong directional selection. Large population size has been interpreted, in opposing models, as either a brake or an accelerator in its effects on long-term change by selection. These results favor the second model, and challenge the concept of rugged fitness surfaces underlying the first model.

IN experimentally selected traits, large genetic changes are almost invariably accompanied by strong negative fitness effects, which often limit the response to selection (LERNER 1958; AL-MURRANI 1974; NICHOLAS and ROBERTSON 1980; FALCONER 1981; ZENG and HILL 1986). This empirical fact has contributed, probably more than any other, to the establishment of the important idea that populations inhabit a rugged landscape of genetic fitness peaks (PROVINE 1985). The theory of fitness peaks, as elaborated by WRIGHT (1931, 1932, 1977a), states that a population's fitness is a function of its point coordinates in n-dimensional gene-frequency space, and that the surface of all population fitnesses in dimension n + 1 is fragmented into many separate domains of attraction of local fitness optima, determined by alternative combinations of interacting genes. According to this view, any new regime of directional selection redefines this surface and tends to trap populations on suboptimal peaks, with increased trait mean but decreases in typical fitness characters. WRIGHT (1978, 1982) interpreted the routine fitness problems of strongly selected lines as manifestations of these suboptimal peaks and thus as evidence of the reality of a landscape of many potential interaction systems, with innumerable peaks and valleys, where mass selection alone cannot locate superior systems. An alternative possibility, with a completely contrary significance for evolutionary theory, is that the negative fitness effects commonly observed in highly selected lines are only artifacts of their unrealistically small population sizes.

Corresponding author: K. E. Weber, Department of Biological Sciences, University of Southern Maine, 96 Falmouth St., Portland, ME 04103. E-mail: keweber@usm.maine.edu

In the present experiment over nine million *Drosophila melanogaster* were processed in two replicate selection lines during the 100 generations covered in this report. The scale of the experiment was made possible by a system for measuring and selecting on performance in a compartmented wind tunnel (WEBER 1988). The choice of this trait was dictated by the desire for a system that could score massive numbers, so that the breeding population could be restricted to the top few percent of measured flies but still be of unprecedented size. This made it possible to combine a mean selection intensity of ~4.5% with an effective population size estimated to vary between 500 and 1000.

Laboratory populations in the effective size range of this study may qualify as models for typical natural populations, or at least for populations that would not have to be considered particularly small if they existed as permanent genetic isolates in nature (EASTEAL 1985; FRANKHAM 1995). The capacity for sustained response to selection in populations of this effective size or larger is a fundamental parameter of evolutionary genetics that has been rarely studied. Prolonged selection of the intensity reported here must occur only infrequently in nature, at least for radically new phenotypic optima. However, intense selection is the only practicable way to probe the evolutionary limits of experimentally large populations in reasonably short times. Since higher gains are expected from large populations (ENFIELD 1980; YOO 1980a; WEBER 1990; WEBER and DIGGINS

1990), this may be the best approach available to model experimentally the fitness consequences of rapid adaptive shifts.

MATERIALS AND METHODS

The selection system: Flying speed was measured and selected in a wind tunnel (WEBER 1988). The tunnel is 1.5 m long and 12 × 12 cm square in cross-section, divided into 40 equal compartments by internal walls. Circular 4-cm holes in the internal walls form a cylindrical flight path through the axis of the apparatus. Flies are released into the downwind end from a cartridge and fly toward a bright light at the upwind end. At each compartment air escapes upward through a light-proof adjustable valve. Thus flies moving toward the light are opposed by wind speeds that increase in regular increments from zero in the first compartment to a maximum in the last. The compartment walls have a dry, slippery coat of Fluon (Northern Products, Inc., Woonsocket, RI) on the downwind sides, so that flies can only advance toward the light by flying. (Tests with dead flies in the airstream showed that flies cannot be moved upwind passively by turbulence.) To end a run, the light is blocked and the air input is switched to CO_2. Flies are anesthetized and fall to the compartment floors in a few seconds. The top of the tunnel can be removed to score the phenotypic distribution.

Performance is scaled as apparent flying speed in nominal cm/sec, based on the wind speed calculated for each compartment from the input flow rate, assuming no shear or turbulence. Various tests and observations show that the trait is actually a composite of phototaxis, activity level, flying speed, and aerial maneuvering skill. The flies can be observed during a run by uncovering windows along the sides in a darkened room. These observations show that flies that advance far upwind never fly down the center of the tunnel where the wind is strongest but skirt the edges of the airstream, advancing one compartment at a time and clinging at the lip of each opening for long periods before flying again. Leading flies that try to gain one more compartment may be blown back several instead.

Methods of scoring and selection: Up to 15,000 flies (quantified volumetrically) are fractionated in a single run, for selection of the top few percent. Smaller samples (≈1000) are used to score the full distribution. Full distributions were scored every fifth generation (except for measurements in generation 96). The input air-flow rate had to be increased several times during the experiment to keep pace with the increasing performance of the flies, in generations 30 (70–110 liter/min), 35 (110–140 liter/min), and 75 (140–160 liter/min). The curve of selection response (in cm/sec) shows no effect of the upward adjustments in the flow rate at which measurements were made. All other measurements of flying speed in tests reported here were made at the input flow rate of 160 liters/min.

Flies were harvested each afternoon when most females would still be virgins, stored at 12° to prevent mating, and returned to room temperature an hour before selection. Most mating was thus postponed until after selection. To measure the phenotypic distribution, flies were stored on fresh medium 2 days at 12°, warmed for an hour at room temperature, held in empty bottles for a second hour, and allowed to fly in the tunnel for 1 hr, except in generation 75 when half-hour tests were used.

Base population and selected populations: The base population for this experiment was the line LF350, founded from 350 wild isofemale lines captured from dense fall populations at several agricultural sites in Lincoln, Massachusetts and

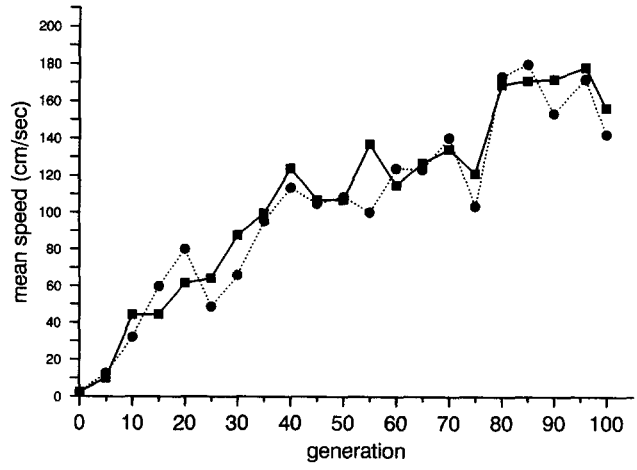

FIGURE 1.—Response to selection. Mean apparent flying speeds of lines AA1 (■) and AA2 (●), measured every five generations.

maintained at large population size. This line was divided into two large replicate populations designated CN1 and CN2, used to found selection lines AA1 and AA2, respectively. Both CN1 and CN2 were maintained as controls in large populations, in 20–30 bottles with periodic mixing. Control and selected populations were all cultured on the same medium (cornmeal-molasses-agar in half-pint bottles) and on the same schedule. Cultures were founded with 20 flies per bottle during the first half of the experiment, then with 40.

Selection was performed on all available progeny, but only the constant number required for founding was selected each generation, thus holding the percent selected to the smallest possible value. On average ~46,000 flies were harvested per generation. The number of selected parents varied between 1200 and 2400 in different periods, depending on the number of culture bottles, but was usually 2000. Effective population size can only be estimated crudely. Females eventually began to outnumber males among selected flies, and in the largest generations when the selection pressure was only 2%, the ratio of females to males could be as high as 10:1. Extra males from less intensely selected flies were usually added in these generations, to keep up the effective size. In generations where this was not done and the total number of selected flies was 2000 selected at 2%, the effective size would be 660. With variations in parental number, density, and sex ratio in mind, it seems safe to say the effective size was always at least 500, and was often as high as 1000.

Measurement of competitive fitness: Competitive fitnesses of both selected lines and both control lines were measured as the proportions of wild-type flies emerging from half-pint culture bottles founded with pregnant females of both the tested line and a white-eyed mutant stock. This method simultaneously assays relative fecundity, egg-to-adult survival, and larval competitive ability. In the first test (generation 50), 10 testee females from each line competed with 10 *w*/*w* (*white*) females per bottle, and in the second test (generation 85), 10 testee females from each line competed with 20 *cn bw*/*cn bw* (*cinnabar brown*) females, these having a similar white-eyed phenotype. A different tester stock was used the second time because the first stock, a new allele of *white*, was inadvertently lost. Higher numbers of mutant females per bottle were used in the second test because preliminary tests showed that the second mutant stock was about half as productive as the first. Because of this adjustment, the proportion of wild-type progeny from controls remained nearly the same in both tests, as desired. The total output per culture of tester and testee flies

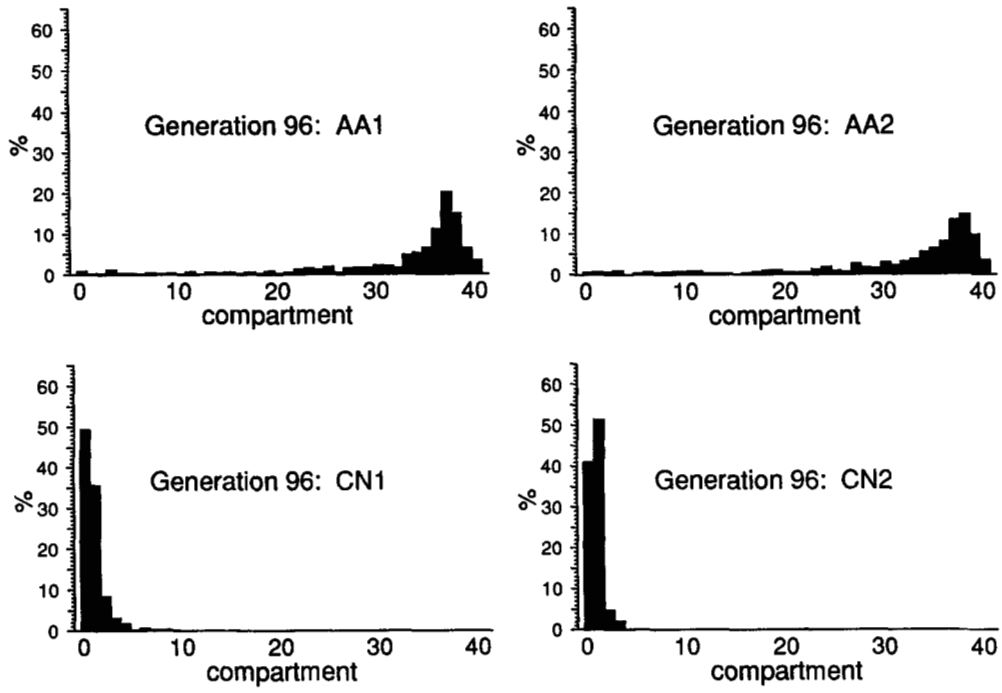

FIGURE 2.—Phenotypic distributions of selected lines (AA1 and AA2) and control lines (CN1 and CN2) in generation 96. Percents of total flies distributed in the wind tunnel compartments (numbered 1–40) and in the introduction cartridge (0), after 1 hr, with an input air flow rate of 160 liter/min. Each sample includes about 1000 flies.

was also about the same in both tests. The parents were allowed to lay eggs for 3 days, and then all progeny were harvested daily through the 20th day of culture, when very few additional flies were emerging. Six bottles were founded per line per test. All control and selected parents used to found test bottles were cultured at equal densities of 30 eggs/vial in the first test, and 60 eggs/vial in the second test. All tester stock parents were cultured at equal densities of 30 and 100 eggs/vial, respectively, in the two tests.

RESULTS

The apparent flying speeds of the selected lines were scored every five generations during the first 100 generations (Figure 1). The lines tended to remain closely matched. A period of slightly depressed performance (generations 45–75) can be accounted for by experimental changes in the water content of the medium formula. Phenotypes in generation 75 were additionally depressed because measurements were made with shorter run times. After generation 75 there was a return to previous methods of culture and measurement. Response to selection continued throughout the experiment, but at continually diminishing rates. (Recent measurements show that the lines have evolved well beyond the highest values in Figure 1, so that Figure 1 does not represent a plateau.) Control line flying speeds were also measured in 14 of the first 100 generations and remained constant, with means and standard errors of 2.22 ± 0.31 cm/sec (CN1) and 2.10 ± 0.64 cm/sec (CN2). Figure 2 shows the large differentiation between selected and control lines.

The competitive fitnesses of selected and control lines were compared in two generations (Figure 3). In each comparison, the mean proportion of surviving adult progeny was slightly higher for controls, than for the associated selected line, under identical conditions of competition. However, when the variation among treatments is tested over the variation within treatments, by ANOVA of the arcsine-transformed propor-

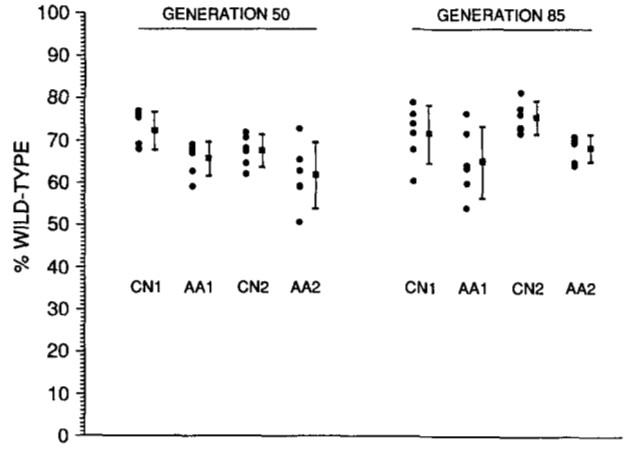

FIGURE 3.—Competitive fitnesses of selected lines (AA1, AA2) and control lines (CN1, CN2) competing against white-eyed tester stocks in generations 50 and 85. Each test of each line included six replicate culture bottles, founded with pregnant wild-type testee and white-eyed tester females, all from controlled-density cultures. Each data point represents percentage wild-type adults emerging from one culture bottle. Means and 95% confidence limits are based on arcsine-transformed proportions. Totals scored were 40,674 flies in generation 50 and 42,064 flies in generation 85.

TABLE 1

ANOVAs of competitive fitness tests

Generation	Source	d.f.	ms	F	P
50	Treatment	1	0.02575	3.999	$0.10 < P < 0.25$
	Replicates	2	0.00644	2.248	$0.10 < P < 0.25$
	Error	20	0.00287		
85	Treatment	1	0.03492	7.870	$0.10 < P < 0.25$
	Replicates	2	0.00444	1.141	$0.25 < P < 0.50$
	Error	20	0.00389		

The table shows arcsine-transformed proportions of wild-type flies emerging from culture bottles after competition with marked mutant stock. The treatment difference is between two control and two selected lines.

tions, the observed differences between treatments are not significant in either generation 50 or 85 (Table 1). Thus the differences between selected and control lines within generations might be due to the same causes as the variation among replicates within treatments. Pooling the results of both generations would probably make the difference between treatments significant, but the two tests were not carried out in exactly the same way (see MATERIALS AND METHODS). The consistent pattern of both tests taken together strongly suggests a small loss of fitness in the selected lines. The key point is that *the differences in fitness, if real, are consistently small,* averaging only six percentage points in generation 50 and seven in generation 85.

To assay the opposition of natural selection to the trait in the selected lines, large unselected subpopulations of AA1 and AA2 were initiated in generation 65 (lines RA1 and RA2), and again in generation 85 (lines RB1 and RB2). All four relaxed sublines were measured in generations 95, 96 and 97 (Table 2). The means of the relaxed lines were lower than the means of the selected lines (AA1 and AA2), measured at the same time (generation 96). However, the differences between selected and relaxed lines were small compared to intergenerational variation, and it is not clear whether the differences were due more to regression in the relaxed lines, or to continued response in the selected lines.

Figure 4 interprets graphically the differences between relaxed and selected lines, to estimate how much regression may have occurred. The triangles show the means of the RA lines (lower triangle) and the RB lines (upper triangle), from Table 2, superimposed on the data from Figure 1. The heavy curved line in Figure 4 is an interpretation of the response curve assuming no intergenerational environmental variation and assuming measurements under the same conditions throughout as in generation 96. Figure 4 shows that all plausible smooth curves through the data for relaxed and selected lines in generation 96 permit only slight regression to have occurred in the relaxed lines. Moreover, measurements of flying speed from the most recent generations show that response has continued since the hundredth generation, and therefore even the small difference between the RB and AA lines in generation 96 must be partly due to continued response in the AA lines. In any case, the data are sufficient to prove the essential point that *no precipitous decline in the selected trait*

TABLE 2

Phenotypes of relaxed sublines

Line	Generations relaxed	Generation 95	Generation 96	Generation 97
RA1	65	146.63	144.98	147.22
RA2	65	134.74	144.21	122.09
Mean of RAs			139.98	
RB1	85	181.05	155.05	149.17
RB2	85	154.85	163.54	164.40
Mean of RBs			161.34	

Values are expressed in cm/sec.

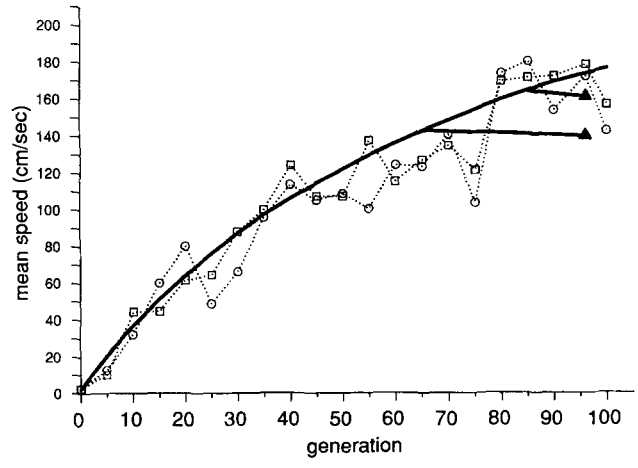

FIGURE 4.—Retention of phenotype during relaxation. Data from Table 2 and Figure 1. Lower triangle shows mean of populations RA1 and RA2, relaxed in generation 65 for 30 generations. Upper triangle shows mean of populations RB1 and RB2, relaxed in generation 85 for 10 generations. The hand-fitted curve estimates phenotypes in generations 65 and 85 assuming constant conditions of measurement and culture like those in generation 96. Phenotypes could only have fallen slightly after relaxation.

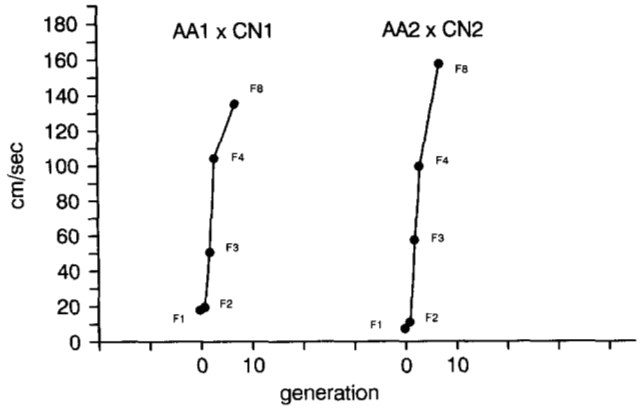

FIGURE 5.—Response to selection on hybrid lines AA1 × CN1 and AA2 × CN2. All measurements were made at standard 160 liter/min flow rate. Each point is the mean of three tests. Axes are scaled to the same ratio as in Figure 1 to permit direct visual comparison of slopes (response rates) between the two figures.

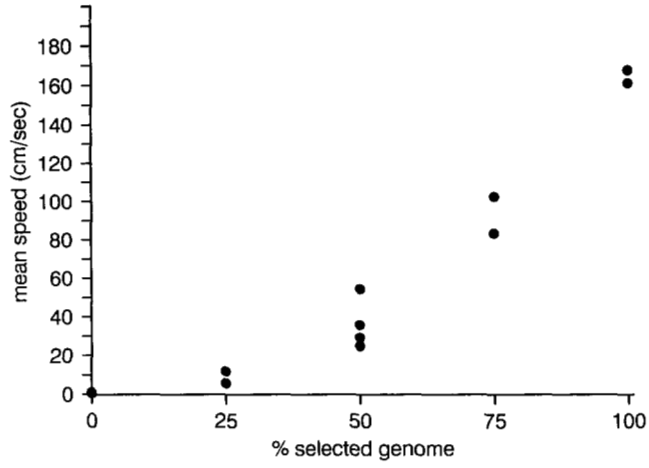

FIGURE 6.—Reciprocal crosses and backcrosses of AA2 × CN2 in generation 98. Each point is the mean of two measurements of the same cross, using flies from separate culture bottles. At 50% of the selected genome, the upper points are F_2's and the lower points are F_1's.

occurred, when lines were relaxed after a large response to selection.

In generation 75 the selected lines AA1 and AA2 were hybridized with their respective control lines, CN1 and CN2. The means of F_1 and F_2 flies from these crosses were closer to the means of unselected than of selected flies, showing strong dominance of the wild-type genome (Figure 5). Beginning with the F_2 generation, both hybrid lines were selected for six generations for increased wind tunnel performance, according to the standard selection protocol, to assay the genetic variance segregating from the crosses. During selection, the performance of the hybrid lines was measured by the standard method in the F_2, F_3, F_4, and F_8 generations (Figure 5). The response to selection was like a quantum jump, showing increased heritability of the trait in the hybrid lines. The original, gradual response seen in Figure 1 was not recapitulated. Instead, the entire previous response could now be recovered in a few generations. *Six generations of selection on the hybrid lines produced the same response as 75 generations of selection on the initial populations.*

A more conventional hybridization analysis was performed in generation 98. Lines AA2 and CN2 were crossed and backcrossed in both directions, and two replicates of all crosses were measured. Figure 6 shows a consistent pattern among crosses, confirming the earlier finding of strong dominance of wild-type alleles, shown in Figure 5. The quantity $4F_2 - 2F_1 - P_1 - P_2$, using the means of hybrids and parents, is expected to be zero if no interlocus epistasis exists (MATHER and JINKS 1977); and here equals -38.5 ± 37.5 (SE). Some epistatic component may be present but cannot be resolved with these data.

DISCUSSION

Several other very large selection experiments have been reported (WEBER 1990; WEBER and DIGGINS 1990) but not this large, nor this long, nor at this intensity. This is also the first time the fitness effects of directional selection have been tested for any populations in this size range. In typical experiments, prolonged selection produces a large fitness loss and an increasingly negative correlation between fitness and the selected trait. In this experiment, response continued for an unusually large number of generations, and apparently involved many loci, yet the genetic changes were not accompanied by a large decline in fitness, nor by a large negative fitness correlation. This combination of results has not been reported previously and is probably due to the extreme size of the populations. If it is a general rule that serious fitness problems do not occur in mass-selected populations if they are sufficiently large, then the fitness problems commonly reported in selection experiments do not support the idea of a controlling fitness surface with innumerable peaks and valleys.

Fitness loss, inbreeding, and population size: These results can be compared to those of LATTER and ROBERTSON (1962), who investigated fitness loss during selection for three different traits in *D. melanogaster.* Fitness was measured by competition against a reference strain for fecundity, larval survival, and male mating success. Lines selected for abdominal bristle number lost 78% of base-population fitness in 20 generations of selection; sternopleural bristle-number lines lost 59% of base-population fitness in 25 generations of selection, and lines selected for wing length lost 75% of fitness in only 10 generations of selection. These extreme fitness losses were sustained at typical selection pressures of 20–40%.

Inbreeding is evidently one of the main causes of fitness loss during experimental selection. Inbreeding accelerates drift, causing high frequencies of some deleterious alleles. In the study just mentioned (LATTER and

ROBERTSON 1962), the unselected control lines lost 39% of base-population fitness in 25 generations. This large fitness loss could only be due to inbreeding and indicates that much of the fitness loss in their selected lines came from the same source. Both selected and control lines were maintained with 20 breeding parents per generation. Lines of this size are typical in selection experiments.

Natural selection to maintain fitness, during artificial selection, will often be weaker in small, low-density experimental populations than in large populations where mass culture methods are used and competition is high. FRANKHAM *et al.* (1988) were able to reduce fitness loss, during selection for ethanol resistance in *D. melanogaster,* by culling the families that produced the fewest pupae. Again, fitness was estimated by reproductive competition against marker strains. In 25 generations of selection on lines with 40 parents, populations selected without culling for productivity lost 49% of base population fitness, while populations selected with culling lost only 11%. The response to selection was almost identical in both sets of lines. This result also argues that much fitness loss during selection is a result of drift and not an inevitable consequence of selection response.

Negative correlation between fitness and selected traits: When selection was relaxed, the wind tunnel lines did not steeply regress. This is previously unknown in lines that have just undergone extreme change under prolonged selection. Even conventionally large populations (ENFIELD 1977, 1980; YOO 1980a), after strong selection, have regressed precipitously as soon as selection was relaxed. In ENFIELD's study of selection for higher pupal weight in Tribolium, with estimated N_e's between 70 and 100, two sublines relaxed in generation 110 lost ~50% of their response in 10 generations (ENFIELD 1980: Figures 1 and 2). YOO's six lines of *D. melanogaster,* with estimated N_e's of 60, were relaxed after 86–89 generations of selection for abdominal bristle number and lost on average ~25% of their response in the first three generations after relaxation (YOO 1980a: Figure 1).

Regression in relaxed lines is sometimes rather gradual (DOBZHANSKY and SPASSKY 1969), and in some cases did not occur during the period of observation (*e.g.,* BROWN and BELL 1961; THODAY and BOAM 1961; ROBERTS 1966a,b; ENFIELD *et al.* 1983). However, these were cases where plateaus were reached after only a few generations of selection. Early plateaus, without apparent resistance from natural selection, should occur occasionally in small selection lines where selectable variability is soon exhausted anyway. In the present case, the lack of strong regression in the relaxed lines is unusual because the lines had continued to respond for so long.

Linkage and population size: Negative correlations between fitness and selected traits may arise from linkage disequilibria in the base population. Increasing the size of a selected population has the same effect on the fixation probabilities of linked alleles as increasing the frequency of recombination (HILL and ROBERTSON 1966). Therefore correlations due to increasingly close linkages dissolve, as population size increases.

The importance of linkage depends on the number of loci contributing to the selected trait. The number of loci involved in wind tunnel performance is large enough to cause important linkage effects, as shown by the extreme contrast between the gradual selection response in Figure 1 and the saltational response of hybrid lines in Figure 5. By generation 75, when the hybrids were created, the chromosomes of the selected lines included many selected alleles at or near fixation. These formed blocks of alleles that remained available for selection in nearly intact linkage in the hybrid lines. Most of these contributory alleles must have been present in the base population, at lower frequencies and dispersed among different chromosomes. (A few early, favorable mutations could also have risen to high frequencies by generation 75). The quantum response to selection in the hybrid lines refutes the possibility that the original gradual response of the selected lines (Figure 1) was based on a few major alleles, rising slowly in frequency because of low heritability under an inefficient selection methodology. Instead, response during the first 75 generations must have involved the gradual accumulation of many minor alleles into profoundly restructured chromosomes by a system which can concentrate available genetic variance effectively.

Pleiotropy and population size: Negative correlation between fitness and selected traits can also arise from pleiotropy, but the combination of desirable and undesirable effects in the same allele need not create a permanent stand-off. Instead, the negative side effects of selected alleles will cause secondary selection for modifiers to alleviate those effects. It is commonly observed that natural selection gradually suppresses many of the effects of major mutations, when they are maintained in small stock cultures. Therefore in large populations, where variability is plentiful, the alleviation of negative pleiotropic effects of selected alleles by available modifiers must be relatively rapid.

In the simplest model of this process, an allele with positive effects on the selected trait has negative effects on fitness, which are compensated by an epistatic allele at a second locus. As the first allele rises in frequency, it raises the frequency of the second by increasing its mean fitness. By this mechanism, many independently acting alleles under direct selection should acquire trailing clusters of modifiers. In genetically diverse populations, any apparently additive response to mass selection must drag along with it an increasing epistatic component. In larger populations, the weak second-order effects of selected-gene frequencies on modifier-gene frequencies can permeate the genetic background with greater efficiency.

A summary of the arguments: Large population size prevents inbreeding and drift, overcomes the random unfavorable linkages that necessarily occur in highly polygenic traits, and increases the efficiency of selection on modifiers that lessen negative side effects of selected genes. All identifiable causes of fitness loss during selection would be alleviated by these effects of large population size, predicting the results that were actually obtained. According to these arguments, the unfavorable developments that accompany long-term selection in most experiments should be increasingly postponed in larger lines, or even eliminated. *Negative fitness effects are, in this view, not an inherent feature of radical change by mass selection.*

Exactly the same factors also explain the empirical rule that selection response increases with population size (JONES *et al.* 1968; ENFIELD 1980; YOO 1980a; WEBER 1990; WEBER and DIGGINS 1990). Thus, the high retention of allelic diversity in large populations not only prevents inbreeding depression, but also supports trait gains (ROBERTSON 1960). The increase in recombinant diversity improves the focus of selection on positive alleles, just as it permits disentanglement from negative ones (HILL and ROBERTSON 1966). The more efficient recruitment of modifiers in large populations applies not only to modifiers that reduce negative effects, but equally to modifiers that enhance the effect of selected alleles on the trait. The greater effect of mutation in larger lines should also be mentioned (FRANKHAM 1980, 1983; FRANKLIN 1982; HILL 1982a,b; WEBER and DIGGINS 1990). Mutation would become a major factor at longer time scales.

A direct comparison of fitness effects would require the same operations on populations of two sizes, with more replication, yet a persuasive comparison can be made between the present results and the results of previous long-term selection experiments on other traits, at smaller population sizes. There is no reason to suspect that this trait is unique. It might be argued that because wind tunnel performance demands not only phototaxis but also vigor and flying ability, it should be closely associated with general fitness, more than, say, abdominal bristle number or other peripheral traits might be. But wind tunnel performance has no functional connection to the fitness components that were tested, which included only fecundity, egg-to-adult survival, and larval competitive ability.

Alternative models of selection limits: Not all models predict that larger population size is the most effective way to maximize selection response and minimize fitness costs. According to WRIGHT (1977a), artificial selection limits correspond to the many phenotypic peaks in gene-frequency space caused by epistatic interactions. Regardless of population size, the expected outcome of selection is the same local optimum, ascended by the steepest possible path from the starting-point of base-population gene frequencies. Because of drift, smaller lines have increased probabilities of other outcomes, both better and worse than the expected outcome, while larger lines have higher probabilities of the expected outcome. Therefore the larger the line, the lower its chances are of attaining a superior outcome, *i.e.*, a limit with maximal trait increase and minimal fitness trade-off. This model of selection outcomes dictated by innumerable fitness peaks is opposed to the arguments presented here, which emphasize the dramatic advantage of larger population size.

The idea of multiple selective peaks was inspired partly by the results of early small-scale selection experiments (WRIGHT 1982), where response plateaued quickly and was usually accompanied by severe fitness problems. In these experiments it was obvious that internal factors resisted change by mass selection. Rather than seeing the drift-based inefficiency of selection in small populations as the *source* of the problems, WRIGHT saw drift, operating in small demes, as the *solution.* He concluded that long-term progress under natural selection (WRIGHT 1988) and in domestic breeding (WRIGHT 1982) arises primarily not by simple mass selection, but by exploitation of the variety exposed among small somewhat inbred demes, or locally selected stocks (WRIGHT 1977b). The present experiment is an indirect test of this idea (see COHAN *et al.* 1989). Results so far are more easily interpreted to support the alternative view, that fitness problems are best overcome and response maximized by the more deterministic action of selection in large populations with the avoidance of subdivision or inbreeding.

Imagining the fitness surface: The genetic fitness surface of a selected trait could be revealed, partially, by a simple experiment. One could get breeding pairs of a widespread species from distantly separated places, grow them into large populations, and mix these in different proportions to create a large, replicated array of composite populations, with different initial gene frequencies. Uniform mass selection on these populations for some trait might show replicably different outcomes, and this could be taken as preliminary evidence of a fixed surface of peaks controlling the evolution of the trait. The essence of the experiment would be to confirm that major contributory alleles in each population were also present initially in other populations where they did not contribute to the selection response, being in a different peak domain. But even without a genetic analysis, the response to selection in these population arrays would already provide an immediate test of WRIGHT's boldest assumption. This is the assumption that by randomly varying initial gene frequencies, the magnitude of change available to mass selection becomes *vastly amplified,* compared to the magnitude that is always available regardless of initial frequencies. If this crucial assumption could not be validated in experiments of this simple design, at least occasionally, then WRIGHT's shifting balance in its three phases

(WRIGHT 1977a), and the controlling fitness surface itself, would no longer merit the attention they have received in the past in evolutionary theory.

Another rarely examined aspect of fitness surfaces is their rate of change by mutation. Even on the short time scale of selection experiments, replicate selection lines become differentiated by different mutations (FRANKHAM 1980, 1983); sometimes these affect fitness or involve interactions affecting the selected trait (YOO 1980a,b), thus potentially creating new peaks in the course of the experiment. The following short argument suggests that, in fact, there is probably no way to preserve the notion of evolution by peak shifts without incorporating frequent mutation. One should first recall that a step upward in Wrightian fitness space requires (among other providential circumstances) a higher peak that is conveniently nearby, and an intervening saddle that is conveniently shallow. It is easy enough to take both these preconditions for granted in models of a single peak-shift. But rapid cumulative change by repeated peak-shifts (WRIGHT 1963, 1977a, 1988) envisions not just two adjacent peaks, but a stairway of peaks, all serially adjacent, each shallowly separated from the next, and each ascending to a higher phenotypic level. This idea very soon becomes miraculous, if it is based on a fixed fitness surface without mutation. Certainly the complete interaction systems of pigs and camels could not have been present already as rare alleles in the first artiodactyl gene pool. Thus, the question is not whether mutation ever restructures a fitness surface, but only how mutation should be incorporated. For example, how long can a fitness peak persist as a constraint? Mutation completely renews the additive genetic variance of typical quantitative traits in 1000 or fewer generations (LYNCH 1988). Nothing restricts mutation to the additive component of the genome. Just as it creates additive genetic variance, it must eventually rearrange the fitness surface, so that new possibilities emerge. Even with occasional strong blockage, the mean half-life of Wrightian genetic constraints may be fleeting on the time-scale of evolutionary change.

If one begins to list the factors that could affect the characteristics of fitness surfaces, it becomes apparent that natural selection has great room to play upon the mechanics of the genome itself, to mold its basic features, not only sex and recombination but genetic interactions as well, to create a system that is responsive to selection. One must first set aside the literature that treats peaks as ecologically possible phenotypes, instead of local optima in gene-frequency space caused by genetic interactions. (The first impediment to change is extrinsic, and the second intrinsic, like the flatness of a road *vs.* the circularity of a wheel.) Focusing then only on the mechanics of the genome, one sees that many factors could raise or lower the relief on the Wrightian terrain. A major factor is the frequency of those interactive gene groups whose effects are not simply nonlinear across genotypes but have emergent properties, creating isolated fitness domains. (An analogy is the complete metamorphosis of meaning in a word by the change of one letter. Thus, starting from the word BOAT, neither COAT nor BOLT is an intermediate step toward COLT, because neither word is closer to COLT in the "phenotypic" space of meanings.) Another factor is the average number of loci within such integral units. Another factor is the tightness of linkage between interactive loci. Another factor is the tendency for regulatory loci to be *cis*-acting. An important factor is the frequency of underdominance, between interactive units as haplotypes, or indeed at individual loci. Other general factors are the degree of redundancy among genes and systems, the degree of modular control of localized parts of the phenotype, the number of ways regulatory cascades can diversify, and the average number of modifiers required to neutralize negative pleiotropic effects of favorable alleles. These aspects, and certainly others, could all be arbitrarily tuned so as to guarantee the extreme fissuring of the fitness surface into terrain too rugged to be negotiated by mass selection, but an extremely rugged fitness surface is not a logically inescapable property of complex genetic systems. WRIGHT assumed that critical allelic combinations resist permutation, like the faces of a Rubik cube, where good combinations may have to be replaced by bad ones temporarily, to get to better ones. But the Rubik cube was designed and "selected" precisely for its resistance to linear change; one can imagine other systems selected to be easy. Thus, simplified fitness surfaces may be selectively favored during long histories of adaptive remodeling.

This research was supported by U.S. Public Health Service grant 5-T32-GM07620-07-0131 and National Institutes of Health grants 5-RO1.GM 21,179; KO4.HD 00638; and RO1.GM 40907.

LITERATURE CITED

AL-MURRANI, W. K., 1974 The limits to artificial selection. Anim. Breed. Abstr. **42:** 587–592.

BROWN, W. P., and A. E. BELL, 1961 Genetic analysis of a "plateaued" population of *Drosophila melanogaster*. Genetics **46:** 407–425.

COHAN, F. M., A. A. HOFFMAN and T. W. GAYLEY, 1989 A test of the role of epistasis in divergence under uniform selection. Evolution **43:** 766–774.

DOBZHANZKY, T., and B. SPASSKY, 1969 Artificial and natural selection for two behavioral traits in Drosophila pseudoobscura. Proc. Natl. Acad. Sci. USA **62:** 75–80.

EASTEAL, S. 1985 The ecological genetics of introduced populations of the giant toad *Bufo marinus*. II. Effective population size. Genetics **110:** 107–122.

ENFIELD, F. D. 1977 Selection experiments in Tribolium designed to look at gene-action issues, pp. 177–190 in *Proceedings of the International Conference on Quantitative Genetics*, edited by E. Pollak, O. Kempthorne and T. B. Bailey. Iowa State University Press, Ames, IA.

ENFIELD, F. D., 1980 Long-term effects of selection: the limits to response, pp. 69–86 in *Selection Experiments in Laboratory and Domestic Animals*, edited by A. Robertson. Commonwealth Agricultural Bureaux, Slough, UK.

ENFIELD, F. D., D. T. NORTH, R. ERICKSON and L. ROTERING, 1983 A selection response plateau for radiation resistance in the cotton boll weevil. Theoret. Appl. Genet. **65:** 277–281.

FALCONER, D. S., 1981 *Introduction to Quantitative Genetics,* Ed. 2. Longman, New York.

FRANKHAM, R., 1980 Origin of genetic variation in selection lines, pp. 56–68 in *Selection Experiments in Laboratory and Domestic Animals,* edited by A. Robertson. Commonwealth Agricultural Bureaux, Slough, UK.

FRANKHAM, R., 1983 Origin of genetic variation in selection lines. Proceedings of the 32nd Annual National Breeders' Roundtable 1–18.

FRANKHAM, R., 1995 Conservation genetics. Ann. Rev. Genet. **29:** 305–327.

FRANKHAM, R., L. P. JONES and J. S. F. BARKER 1968 The effects of population size and selection intensity in selection for a quantitative character in Drosophila. I. Short-term response to selection. Genet. Res. **12:** 237–248.

FRANKHAM, R., B. H. YOO and B. L. SHELDON, 1988 Reproductive fitness and artificial selection in animal breeding: culling on fitness prevents a decline in reproductive fitness in lines of *Drosophila melanogaster* selected for increased inebriation time. Theoret. Appl. Genet. **76:** 909–914.

FRANKLIN, I. R., 1982 Population size and the genetic improvement of animals, pp. 181–196 in *Future Developments in the Genetic Improvement of Animals,* edited by J. S. F. Barker, K. Hammond and A. E. McClintock. Academic Press, New York.

HILL, W. G., 1982a Rates of change in quantitative traits from fixation of new mutations. Proc. Natl. Acad. Sci. USA **79:** 142–145.

HILL, W. G., 1982b Predictions of response to artificial selection from new mutations. Genet. Res. **40:** 255–278.

HILL, W. G., and A. ROBERTSON, 1966 The effect of linkage on limits to artificial selection. Genet. Res. **8:** 269–294.

JONES, L. S., R. FRANKHAM and J. S. F. BARKER, 1968 The effects of population size and selection intensity in selection for a quantitative character in Drosophila. II. Long-term response to selection. Genet. Res. **12:** 249–266.

LATTER, B. D. H., and A. ROBERTSON, 1962 The effects of inbreeding and artificial selection on reproductive fitness. Genet. Res. **3:** 110–138.

LERNER, I. M., 1958 *The Genetic Basis of Selection.* Wiley, New York.

LYNCH, M., 1988 The rate of polygenic mutation. Genet. Res. **51:** 137–148.

MATHER, K., and J. L. JINKS, 1977 *Introduction to Biometrical Genetics.* Cornell University Press, Ithaca, NY.

NICHOLAS, F. W., and A. ROBERTSON, 1980 The conflict between natural and artificial selection in finite populations. Theoret. Appl. Genet. **56:** 57–64.

PROVINE, W. B., 1985 The R.A. Fisher-Sewall Wright controversy, pp. 197–219 in *Oxford Surveys in Evolutionary Biology,* Vol. 2, edited by R. DAWKINS and M. RIDLEY. Oxford University Press, London.

ROBERTS, R. C., 1966a The limits to artificial selection for body weight in the mouse. I. The limits attained in earlier experiments. Genet. Res. **8:** 347–360.

ROBERTS, R. C., 1966b The limits to artificial selection for body weight in the mouse. II. The genetic nature of the limits. Genet. Res. **8:** 361–375.

ROBERTSON, A., 1960 A theory of limits in artificial selection. Proc. R. Soc. London B **153:** 234–249.

THODAY, J. M., and T. B. BOAM, 1961 Regular responses to selection. I. Description of responses. Genet. Res. **2:** 161–176.

WEBER, K. E., 1988 An apparatus for selection on flying speed. Drosophila Inf. Serv. **67:** 92–93.

WEBER, K. E., 1990 Increased selection response in larger populations. I. Selection for wing-tip height in Drosophila melanogaster at three population sizes. Genetics **125:** 579–584.

WEBER, K. E., and L. T. DIGGINS, 1990 Increased selection response in larger populations. II. Selection for ethanol vapor resistance in Drosophila melanogaster at two population sizes. Genetics **125:** 585–597.

WRIGHT, S., 1931 Evolution in Mendelian populations. Genetics **16:** 97–159.

WRIGHT, S., 1932 The roles of mutation, inbreeding, crossbreeding, and selection in evolution. Proceedings, VIth International Congress of Genetics **1:** 356–366.

WRIGHT, S., 1963 Plant and animal improvement in the presence of multiple selective peaks, pp. 116–122 in *Statistical Genetics and Plant Breeding,* edited W. D. HANSON and H. F. ROBINSON, Natl. Acad. Sci., Nat. Res. Council.

WRIGHT, S., 1977a *Evolution and the Genetics of Populations,* Volume 3. University of Chicago Press, Chicago.

WRIGHT, S., 1977b Modes of evolutionary change of characters, pp. 459–473 in *Proceedings of the International Conference on Quantitative Characters,* edited by E. POLLAK, O. KEMPTHORNE and T. B. BAILEY. Iowa State University Press, Ames, IA.

WRIGHT, S., 1978 The relation of livestock breeding to theories of evolution. J. Anim. Sci. **46:** 1192–1200.

WRIGHT, S., 1982 Character change, speciation, and the higher taxa. Evolution **36:** 427–443.

WRIGHT, S., 1988 Surfaces of selective value revisited. Am. Nat. **131:** 115–123.

YOO, B. H., 1980a Long-term selection for a quantitative character in large replicate populations of *Drosophila melanogaster.* I. Response to selection. Genet. Res. **35:** 1–17.

YOO, B. H., 1980b Long-term selection for a quantitative character in large replicate populations of Drosophila melanogaster. II. Lethals and visible mutants with large effects. Genet. Res. **35:** 19–31.

ZENG, Z. B., and W. G. HILL, 1986 The selection limit due to the conflict between truncation and stabilizing selection with mutation. Genetics **114:** 1313–1328.

Communicating editor: A. G. CLARK