STADLER SYMP. Vol. 15 (1983) University of Missouri, Columbia

Measures of Fitness in *Drosophila*

(population genetics, *Drosophila*, fitness, natural selection)

Daniel L. Hartl and David S. Haymer

Department of Genetics
Washington University School of Medicine
St. Louis, Missouri 63110

SUMMARY

Ten proposed experimental measures of fitness in Drosophila have been estimated in 8 to 38 strains of D. melanogaster with 1 to 6 replications in order to assess the degree of association among the measures. One measure is a composite of the classical fitness components viability, fecundity, and mating speed. Two tests--the Knight-Robertson method and the compound-autosome method--are based on intraspecific competitive ability; three tests are based on interspecific competitive ability with D. simulans, D. mauritiana, or D. ananassae; two measures are based on the productivity and biomass of equilibrium populations; and two measures are based on the dynamics of change in frequency of a balancer chromosome in experimental populations. The tests fall into four groups with significant correlations among tests within a group but weaker or nonsignificant correlations between groups. The first group consists of the composite index and the two intraspecific tests. We infer that these methods measure attributes strongly allied with classical darwinian fitness. The second group consists of the interspecific tests, which apparently emphasize somewhat different characteristics than those associated with darwinian fitness. The third group consists of the measures based on productivity or biomass of equilibrium populations, and these measures may be allied with Wright's mean selective value, although this interpretation is speculative. The fourth and final group consists of the measures resulting from changes in chromosome frequency in experimental populations. Each group of tests measures distinct characteristics that are probably important in particular contexts, but only the intraspecific tests correspond to darwinian fitness.

INTRODUCTION

Fitness is one of the most subtle and elusive concepts in

population genetics. Not that it is conceptually difficult. FISHER (1930) defines fitness laconically as "expectation of offspring"; DOBZHANSKY (1970) elaborates the definition as "the relative contribution of a genotype to the pool of genotypes in the next generation"; and a well-known contemporary textbook of population genetics (HARTL 1980) defines the fitness of a genotype as "the average number of offspring produced by individuals of that genotype."

All of these definitions are talking about the same thing, namely the concept of fitness as used in theoretical population genetics, where fitness is merely a parameter, usually assumed to be constant, used in calculating allele frequencies recursively from generation to generation. In this context fitness is usually, though not always, synonymous with viability, which refers to the relative probability of a genotype surviving from fertilization to reproductive age. When theoretical usage of the term is broadened to include fecundity, which refers to the relative production of offspring by an individual of reproductive age, the theory inevitably becomes more complex because fitness is then no longer necessarily a property of single individuals but rather of mating pairs (see EWENS 1979 for examples). Mathematical complexity aside, the concept of fitness as used in theoretical population genetics is well defined, biologically plausible, and useful in its function of quantifying differences among genotypes in the context of natural selection.

THODAY (1953) has put forward a substantially different definition of fitness. He defines the fitness of a group of individuals in terms of the probability that the group will have living descendants after a very long period of time, his example being 100 million years. Fitness is then a measure of long-term evolutionary success, and Thoday discusses the concept in terms of its various components such as adaptation, stability, variability, and the rate of environmental change, relating these to the notion of biological progress. Unfortunately, this long-term definition of fitness is beyond the possibility of measurement, except in retrospect, and it leads to such curious paradoxes as that Cambrian trilobites were very fit, as their descendants lasted about 330 million years, but Triassic dinosaurs were not,as their descendants lasted only about 90 million years. About such organisms as hoofed mammals, primates in general, and *Homo sapiens* in particular, their fitness would of course be unknown because it is much too soon to say, and it will require another 50 million years or so to tell the tale. Although some measure of long-term evolutionary persistence may be useful for some purposes, it is our view that the term "fitness" should be restricted to its established usage in population genetics as embodied in the definitions cited earlier.

The problem with fitness is not, therefore, conceptual. The problem is in translating the conceptually simple definition into a practicable experimental scheme for purposes of estimation. WALLACE (1981) emphasizes the horrendous difficulty of estimating fitness by noting that fitness involves a

complex and unknown function of "viability, sexual activity, fertility, fecundity, longevity, developmental speed, judgement as to where and when eggs are to be deposited, and many other essentially intangible aspects of the organism's biology." Many of these fitness components, such as viability and fertility (but not, of course, the intangibles) can be measured, especially in organisms such as *Drosophila*, where there is a rich literature on fitness components (for reviews see DOBZHANSKY 1970, LEWONTIN 1974, and WALLACE 1981). Sometimes fitness components are of interest in their own right, as in studies of the effects of inbreeding and homozygosity or the effects of newly arising mutations on viability (for three among many examples see MUKAI 1964, 1969 and MUKAI, CHIGUSA & KUSAKARA 1982). However, in light of the complexity of the trait "fitness," it is evident that measures of any one fitness component or small number of components may not, and in general will not, adequately reflect the overall fitness of the individuals for purposes of predicting the results of competition or allele-frequency changes with natural selection. In particularly favorable cases the myriad details of the life cycle impinging on fitness (including the intangibles) can be lumped into a much smaller and more manageable number of *net* fitness components that are found to be adequate in prediction. This approach is exemplified in PROUT (1971a, b), who has devised a relatively simple method for estimating the components of fitness operating in adult *Drosophila* and has found in one test case that the method yields values that are reasonably good in predicting allele-frequency change with selection. Nevertheless, few protocols are as powerful as Prout's in detecting such aspects of overall fitness as mating interactions, and the fact remains that there is as yet no generally accepted operational method for estimating individual fitness, even in *Drosophila*.

Another approach to estimating fitness is to ignore the many individual components of fitness and to estimate instead a parameter that in some sense corresponds to overall fitness. After all, unless the individual components of fitness are of particular interest in their own right, it is pointless labor to measure them when all one really wants to know are their combined effects. Better to devise an experimental system in which the combined effects are directly observable and to measure these. This approach was pioneered by KNIGHT & ROBERTSON (1957), who devised a simple and elegant method for assessing the overall fitness of a *Drosophila* population that avoids the need for estimating and combining individual fitness components. Indeed, their paper was appropriately entitled "Fitness as a measurable character in *Drosophila*." The Knight-Robertson method is to allow a strain to be tested (say a wild-type strain) to compete in equal numbers against a standard tester strain carrying the dominant second-chromosomal balancers *Curly* (*Cy*) and *Plum* (*Pm*) so that the offspring of intrastrain matings can be distinguished from the offspring of interstrain matings. Knight and Robertson define the competitive index of the wild-type strain as the ratio of wild-type to *Cy/Pm* offspring (the heterozygous *Cy* or *Pm* offspring resulting from interstrain matings being ignored). We prefer to use as an

index the ratio of wild-type to total wild-type plus *Cy/Pm* (the heterozygotes again being ignored), but this difference is a matter of detail. The point is that the competitive index, however calculated, comes as close to DOBZHANSKY's (1970) definition of darwinian fitness as "the relative contribution of a genotype to the pool of genotypes in the next generation" as one can reasonably hope to achieve experimentally. The Knight-Robertson method measures a property of individuals or an average value for individuals measured in a group, and so corresponds to average individual fitness if the group in question is genetically homogeneous.

Several other direct measures of overall fitness have also been used. One, studied by JUNGEN & HARTL (1979) and HARTL & JUNGEN (1979), is a slight variant of the Knight-Robertson procedure that has a useful technical advantage. In this variant, the standard competitor is not *Cy/Pm* but rather a strain bearing compound autosomes. Compound autosomes are chromosomal rearrangements converting a normally metacentric pair of autosomes into a pair of isochromosomes having the left arms attached to a common centromere and the right arms attached to a different centromere. Compound-autosome flies mated among themselves exhibit a fertility reduction of 20 to 25 percent and sometimes much more as compared to normal flies due to the substantial proportion of grossly aneuploid and inviable zygotes that are formed. There is, however, considerable variation in fertility from strain to strain (HOLM & CHOVNICK 1975). In contrast, matings between normal and compound-autosome flies produce no surviving progeny because all of the zygotes are grossly aneuploid. The advantage of using compound-autosomes in the Knight-Robertson procedure is that, in the compound-autosome test, one need not classify and count any flies that do not enter into the competitive index because all of the progeny of interstrain matings die. The competitive index in this case is simply the proportion of wild-type flies among the total progeny. Previous studies (JUNGEN & HARTL 1979) have established that the fitness ranking of a set of normal strains is not markedly dependent on which particular compound-autosome strain is used as a standard.

CARSON (1958, 1961a, b) has proposed a measure of fitness based on the performance of a strain in a noncompetitive situation. According to CARSON (1961b): "When two genetically different populations are tested under...uniform population conditions in the laboratory, the one which is able to maintain the greater size, or biomass, is deemed to be performing better biologically under the given conditions. Size thus may be used as a measure of relative population fitness." Actually, two measurements may conveniently be made of such populations. One is biomass, which refers to the population of adults at any time; the other is productivity, which refers to the number of emerging adults per unit time. The biomass measure has several interesting characteristics. For example, the introduction of a single wild-type gamete into an equilibrium population of mutant flies results in a greater than three-fold increase in biomass in about three generations, which is accompanied, as might be expected, by a drastic reduction in the

frequency of the mutant allele (CARSON 1958). CARSON (1961a) has also shown that populations of *D. robusta* taken from the central part of their range maintain a constant (i.e., nonincreasing) biomass in laboratory populations, whether the populations are isolated or whether they periodically receive immigrants from other populations; moreover, populations derived from a single pair of wild flies and those derived from many pairs of wild flies are identical in equilibrium biomass. This measure has also been used by AYALA (1965a, b, 1966a) and by VAN DELDEN & BEARDMORE (1968), who found that hybrid populations have a greater biomass than inbred populations and that irradiation of populations with low genetic variability results in an increase in biomass, presumably due to the induced genetic variation. These observations indicate that the biomass measure has at least some of the characteristics that might be expected of any legitimate measure of fitness.

Another operational measure proposed for fitness is based on *interspecific* competitive ability. This measure was implicit in the work of MOORE (1952) and PIMENTEL (1965) and has been used systematically by AYALA (1966b, 1970). Of particular interest is a study of AYALA (1970), in which three geographic strains of *D. serrata* were examined in pairwise interspecific competition with *D. pseudoobscura*, *D. melanogaster*, and *D. nebulosa*. The fitness ranking of the geographic strains as assessed in interspecific competition was found to be the same as when the fitness ranking was assessed in terms of biomass (AYALA 1965b). Unfortunately, as we will document below, this provocative correlation does not hold in general.

Yet another measure of fitness has been used extensively by SVED & AYALA (1970), SVED (1971, 1975), TRACY & AYALA (1974), and WILTON & SVED (1979). In this procedure, a population is initiated that is segregating for a chromosome or chromosomes of interest and a dominantly marked homozygous-lethal balancer chromosome (e.g., *Cy*). The fitness effects ascribable to the segregating wild-type chromosome can then be calculated based on the equilibrium frequency that the wild-type chromosome ultimately reaches. Two measures can conveniently be calculated as an index of fitness. One is to use the raw frequency of the wild-type chromosome in the equilibrium population, which is a procedure suggested by SIMMONS & CROW (1977). The other is to use this frequency but to adjust it based on the separately measured viability effect of the wild-type chromosome; this is the calculation preferred by Sved. In any case, The Sved population procedure is eminently suited for the demonstration of heterosis, and SVED (1971, 1975) and WILTON & SVED (1979) have shown that second or third chromosomal homozygotes are reduced in this measure of fitness by 80 to 90 percent as compared to random heterozygotes, and that X-chromosomal homozygotes are reduced about 40 percent.

Based on this background, it can be seen that the complaint that fitness is not a measurable trait is simply wrong. It is a complex trait, to be sure, but it can be measured in any number of ways. These include combining individual fitness components and such methods as the Knight-Robertson method, the

compound-autosome method, and those due to Carson, Ayala, and Sved. None of these measures is by any means perfect. Individual fitness components are difficult to estimate with precision, and the intangibles must inevitably be neglected. The Knight-Robertson and the compound-autosome methods minimize certain fitness components such as developmental time and longevity. The Carson procedure is carried out in the absence of competitors, and that of Ayala is carried out in the absence of intraspecific competitors. In the Sved method, one has to be concerned with the effects of the balancer chromosome. Each of these methods has its particular strengths and weaknesses, and none is by any means perfect. Fitness is a sufficiently complex trait that no single method can realistically be expected to be absolutely reliable and uniformly practical. The best one can hope for is a method or group of methods that is accurate within realistic limits and applicable to a wide but not universal set of circumstances, bearing in mind that fitness itself is not a biological constant but will vary depending on experimental conditions. The various measures of fitness in *Drosophila* would all seem to be defensible from this point of view.

The problem with fitness in *Drosophila* is not that fitness cannot be measured but quite the opposite: there may be too many methods, and it is by no means self-evident whether the various measurements are measuring the same thing or even necessarily related things. That is to say, although all of the methods involve something that seems to relate to the rather vague notion of fitness as used in common evolutionary parlance, it is not clear whether this something corresponds to darwinian fitness in the precise sense defined by Dobzhansky and other contemporary authors in population genetics. The underlying difficulty can be summarized in an aphorism of J. W. N. Sullivan made in 1928, that "it is much easier to make measurements than to know exactly what you are measuring." (Quoted in TAYLOR 1959.) Faced with this ambiguity, we have studied a set of strains of *D. melanogaster* using all of the methods discussed earlier (HAYMER & HARTL 1982, 1983; HAYMER 1982). The purpose of the study was to determine empirically which of the methods yielded values that were strongly correlated with one another, and which gave values that could be interpreted in terms of individual fitness components. Details of the studies will be published elsewhere. Here we will focus on overall patterns and their interpretation.

MATERIALS AND METHODS

Strains. We have studied a heterogeneous collection of wild-type strains of diverse origin. Some were well-established laboratory strains such as Hikone or Texas; some were isofemale lines derived from females collected in the southern United States and provided by Dr. Victoria Finnerty; others were lines derived from the isofemale lines by homozygosing a single second chromosome using the standard *Cy/Pm* technique; other strains derived from crosses between two or more of the lines having homozygous second chromosomes; still others were

derived by repeated sib mating in the isofemale lines for up to 20 generations. One strain was derived as a composite cross involving a number of second-chromosome homozygous lines to provide a highly heterozygous control. The purpose of all this was to establish a heterogeneous set of lines spanning a wide range of fitness for use in comparision of the methods. Since the lines were deliberately chosen to be diverse, no conclusions pertaining to the distribution of fitness effects of chromosomes in any particular natural population would be warranted.

Methods. Insofar as feasible we have used the methods originally described by the authors, and a brief summary of these methods is in order.

1. Compound-autosome test. Each experimental population was established in a half-pint milk bottle initiated with 80 unmated flies, 20 of each sex from the normal and the compound-autosome strain. After 3 days the flies were transferred to a fresh bottle for an additional 3 days, and progeny were computed periodically for 20 days. Each pair of bottles constitutes a replicate, and each strain was tested in 5 or 6 replicates. The index of fitness of a wild-type strain is simply the proportion of wild-type offspring, as hybrid progeny which do not survive. Among replicate variance is much reduced when flies to be placed in competition are reared under conditions as nearly identical as possible as regards density, etc. in the previous generation, and this sort of standardization was routinely employed. The compound-autosome strain was used C(3L)RM, *ri*; C(3R)RM, ry^2.

2. Knight-Robertson test. Prior to the experiment, a set of bottles was initiated, each containing 25 inseminated wild-type females and 25 inseminated females from a CyO/Pm^2 strain whose genetic background had been rendered heterogeneous by mating with a heterogeneous strain and re-extracting the balancers. Unmated progeny from these bottles were used to initiate the experimental populations with 25 flies of each sex from each competing strain. The index of fitness in this case is the proportion of wild-type progeny among the total wild-type plus CyO/Pm^2 progeny, the hybrid $CyO/+$ and $Pm^2/+$ progeny being ignored. These tests were carried out in 4 to 6 replicates.

3. Sved test. These tests were carried out in small plastic population cages (12 x 10 x 10cm) to which were attached 6 food vials and 2 normally empty sample vials. The food vials were replaced in sequence, one every 3½ days, so that each food vial remained with the cage for 3 weeks. Weekly samples were made by replacing the sample vials with food vials for 24 hours, and the progeny emerging from these sample vials were classified and counted. Each cage was initiated with approximately 40 $CyO/+$ males and 40 $CyO/+$ females arising from repeated backcrossing of the wild type strain to the heterogeneous CyO/Pm^2 strain. The fitness index of each strain is a function of the equilibrium frequency of $CyO/+$ and the homozygous viability of the + chromosome and was calculated according to SVED (1971).

4. Ayala test. These tests were performed exactly as the compound-autosome tests except that the compound-autosome strain

was replaced with an interspecific competitor, either *D. simulans* (marked with *yellow*), *D. mauritiana* (marked with *burgundy*), or *D. ananassae* (marked with *white*). The fitness index was calculated as the proportion of *D. melanogaster* among the total progeny, and each strain was tested in 5-6 replicates.

5. Carson test. Population cages consisted of a half-pint milk bottle capped with an inverted 500ml Nalgene polypropylene Erlenmeyer flask having a hole in its base for the insertion of a cotton plug. The food bottle was removed and replaced with a fresh bottle three times per week. The first two bottles removed in any week were treated as sampling bottles, and progeny were counted periodically for 21 days and reintroduced into the population. These progeny counts provide a measure of productivity of the population. Weekly biomass measurements were carried out by weighing the adult flies. Each population was founded with 25 males and 25 females and was carried out in 2 replicates.

6. Fitness components. We have focused on three major fitness components--time to mating (an index of mating activity), egg-to-adult viability, and female fecundity. Many potentially important fitness components, such as developmental time and longevity, have been ignored. Our hope has been that the three major fitness components would account for a substantial proportion of the variance in total fitness, so that correlations between these components and one or more of the measures of overall fitness could be detected. This has proven to be the case.

Mating parameters (of which several were studied but only one is reported here) were estimated by combining single males and single females from each strain to be tested. Time to mating is estimated as the time elapsed between initial mixing of the flies and successful mounting by the male. All animals were similarly aged and handled as in the other fitness tests, and mating tests were carried out at room temperature between 9 and 12AM.

The other fitness components were estimated using the same flies as studied in the mating experiments. Pairs that had copulated were transferred to fresh medium every day for 4 days at 25°C. Immediately after each transfer the deposited eggs were counted to provide the fecundity estimate. Egg-to-adult viability was then estimated from the number of adults emerging from these eggs.

As a composite index of fitness based on the individual fitness components, we have used viability times fecundity divided by time to mating. Many other alternatives are no doubt defensible. What commends this particular measure is that it is the simplest one we could think of that combines the major fitness components in what seems to be roughly the right way. Although other measures might be more appropriate or better for other purposes, our simple composite index has proven satisfactory for the matter at hand.

Statistical analysis. We have eschewed complex multivariate statistical methods in order to stay as close to the actual data as possible. Pairwise correlation coefficients provide the infor-

mation we seek. Initially the data were analyzed using Spearman's nonparametric correlation by rank (SIEGEL 1956). These results proved virtually identical to those based on the product-moment correlation coefficient, and we will deal in terms of the product-moment correlation coefficient because of its greater familiarity.

RESULTS

We begin with a note on the scale of the tests. Individual fitness components were estimated in 31 strains, the compound-autosome tests were carried out with 38 strains, the Knight-Robertson tests with 18, the interspecific competition tests with 13 (*D. simulans* and *D. mauritiana*) and 8 (*D. ananassae*), the Carson tests with 12, and the Sved tests with 9 strains. Individual fitness components were estimated on an average of 14 males and 14 females per strain, each pair being a replicate, and the other tests were replicated 5 to 6 times (compound-autosome and interspecific tests), 4 to 6 times (Knight-Robertson), 2 times (Carson), or 1 time (Sved). These data thus provide a substantial base with which to compare the various measures of fitness.

As an index of repeatability of the tests of overall fitness, we can divide the replicates randomly into two groups and examine the correlation between the two sets of replicates. The results are as follows:

Compound-autosome:	$r = 0.92$, $n = 13$, avg 5.6 reps
Knight-Robertson:	$r = 0.89$, $n = 13$, avg 5.0 reps
Carson (biomass):	$r = 0.92$, $n = 12$, avg 2.0 reps
Carson (productivity):	$r = 0.82$, $n = 12$, avg 2.0 reps
D. simulans:	$r = 0.66$, $n = 13$, avg 5.0 reps
D. mauritiana:	$r = 0.82$, $n = 13$, avg 5.0 reps

In these data the average number of replicates is the number divided into two groups for calculation of the correlation coefficients. The two tests based on intraspecific competition are highly repeatable from one set of replicates to another as judged by the high correlation coefficients. Carson's method, which does not involve competition, yields a high correlation relative to biomass and a somewhat lower correlation relative to productivity. The lowest correlation among replicates is found in the interspecific competitions involving *D. simulans*. Nevertheless, for all the tests, there are sufficiently high within-test correlations to render the among-test comparisons meaningful.

Relative to among-test comparisons revealed by pairwise correlation coefficients, the tests fall clearly into four groups having highly significant correlations within groups but generally much lower or nonsignificant correlations between groups. These groups warrant a brief individual discussion.

The first group consists of three fitness measures as assessed using the compound-autosome (CPA) method, the Knight-Robertson (KR) method, and the composite (COMP) measure of over-

all fitness calculated from individual fitness components as the product of viability and fecundity divided by time to mating. The relevant correlations are

CPA vs COMP: $r = 0.74$ $(n = 28)$
KR vs COMP: $r = 0.67$ $(n = 15)$
CPA vs KR: $r = 0.89$ $(n = 13)$

Each of these correlations is highly significant. It is perhaps surprising that the correlation between the compound-autosome method and the Knight-Robertson method is so high in light of the very different competitors that are involved in the two techniques, in one case a compound-autosome-bearing strain and in the other a *CyO/Pm*^{2} competitor, and in light of the substantial zygotic loss that occurs in the compound-autosome test. It would seem to follow that the two tests measure very nearly the same biological thing despite significant differences in detail, and, indeed, the correlation between the tests is of the same magnitude as the correlation between replicates within a test.

It is of great interest that the composite index of fitness correlates so highly with intraspecific competitive tests, as this correlation shows that the intraspecific competitive tests do measure darwinian fitness as claimed by KNIGHT & ROBERTSON (1957) and that these measures can be related to individual fitness components in a surprisingly simple way. We have also carried out a multiple regression of the form $\log(\text{CPA}) = a \log(\text{viability}) + b \log(\text{fecundity}) + c \log(\text{time to mating})$ and have estimated $a = 0.34$, $b = 0.05$, and $c = -0.17$. However, the overall correlation between CPA and COMP using these values is $r = 0.77$, which is not significantly greater than the simple-minded composite used earlier. We conclude that the composite index defined as the product of viability times fecundity divided by time to mating is preferable to the more complex index based on multiple regression because the simple composite index is biologically interpretable in spite of its somewhat lower correlation.

The second group of tests consists of the methods based on interspecific competitive ability with either *D. simulans* (SIM), *D. mauritiana* (MAU), or *D. ananassae* (ANA). The relevant correlations in this case are

SIM vs MAU: $r = 0.62$ $(n = 13)$
SIM vs ANA: $r = 0.90$ $(n = 6)$
MAU vs ANA: $r = 0.90$ $(n = 6)$

Each of these correlations is significant, suggesting that the methods all measure basic biological characteristics related to interspecific competition. It is of interest that the correlation involving the sibling species *D. simulans* and *D. mauritiana* is smaller than the others, but this could be a statistical artifact of the relatively small number of strains tested. With one exception, none of the interspecific tests is significantly correlated with any of the other fitness tests. The one exception involves the compound-autosome test, which correlates with both SIM and MAU ($r = 0.58$, $n = 13$ and $r = 0.72$, $n = 13$, respectively). Whether these correlations are real or spurious we do not

know, but it is interesting that they involve the sibling species of *D. melanogaster*. In any case, the interspecific tests are not correlated with the Knight-Robertson test or with the composite index of fitness.

The third group of tests consists of the two fitness measures emerging from Carson's noncompetitive population cage method, namely biomass (BIO) and productivity (PROD). The correlation between these measures is

$$\text{BIO vs PROD: } r = 0.62 \quad (n = 12)$$

which is significantly different from 0. However, neither of these measures is significantly associated with any of the others, including the intraspecific tests, the composite fitness index, the interspecific tests, and the Sved test. Evidently the Carson method emphasizes those aspects of the life cycle that lead to large, productive equilibium populations in the absence of competitors, and these aspects are not completely congruent with those involved in the other tests. We note again that all tests except the Carson test and the Sved test minimize developmental time and longevity as components of fitness, and this includes the composite index of fitness. Differences among strains in these components might account for the lack of a correlation between the Carson method and the others. Alternatively, as will be discussed below, the Carson method may well be measuring something other than darwinian fitness.

The fourth and final group of measures corresponds to the two indices growing out of the Sved population test. One index is just the equilibrium frequency of *CyO/+* adults (EQUIL); the other is this same frequency corrected for the homozygous viability of the chromosome in question (WSVED). These would be expected to be negatively correlated, as the larger the value of EQUIL the smaller must be the overall fitness (WSVED) of the chromosome. Indeed,

$$\text{EQUIL vs WSVED: } r = -0.74 \quad (n = 9)$$

which is significant. Neither of these indices is significantly correlated with any of the others, including the Carson indices. It is hard to account for the results of the Carson test and the Sved test based on developmental time and longevity alone because the two tests are uncorrelated with each other. Whatever the Sved test may be measuring, it is clear that it is something not easily related to the results of any of the other measures.

DISCUSSION

We began by estimating 10 proposed indices of fitness among a heterogeneous collection of *Drosophila* strains and have found that the indices are associated in four groups. One group consists of the compound-autosome test, the Knight-Robertson test, and a composite index of fitness based on the components viability, fecundity, and time to mating. All of these tests are highly correlated, and the fact that the composite index is in-

volved strongly implies that these methods measure the same thing and that this thing is darwinian fitness in the sense that this term is used in contemporary population genetics. We do not wish to argue that our composite index of fitness is necessarily the best possible one or all inclusive. Indeed, it ignores such potentially important fitness components as developmental time, longevity, oviposition-site preference, and a multitude of intangibles. Neither do we wish to argue that the compound-autosome method or its forerunner, the Knight-Robertson method, provides a panacea for all problems in the study of fitness. These methods, too, minimize certain fitness components that in other contexts might be critically important. It is a great disappointment that neither test correlates with the results of the Sved test, which tends to undermine their potential usefulness in population prediction. On the other hand, the Sved test fails to correlate with individual fitness components, which suggests that conclusions based on this method should be accepted with reserve until corroborated by independent methods. In spite of their imperfections, the compound-autosome method and the Knight-Robertson method are adequate for measuring individual fitness under the specified experimental conditions on the grounds that they are highly correlated with the composite index based on direct estimates of individual fitness components.

The tests based on interspecific competitive ability constitute a group by themselves generally uncorrelated with other measures. This is not unexpected. Tests based on interspecific competition give little or no weight to components of fitness having to do with overall mating activity or mate competition. Valid measures of darwinian fitness must take these fitness components into account. We conclude that interspecific competitive ability is a characteristic in and of itself and distinct from darwinian fitness.

It is not entirely clear from our data what is involved in the Carson measures of biomass and productivity. They do not correlate well with other measures of fitness, particularly the composite index. One possible explanation is that the Carson method so heavily weighs developmental time, longevity, and perhaps other components not included in the composite that the expected correlation simply vanishes. A more interesting explanation is that the Carson method does not measure darwinian fitness at all. WRIGHT (1969 and earlier) has defined the "mean selective value" of a population as that attribute of a population which determines its success in competition among demes. He argues that the mean selective value of a population is a property of the population as a whole and distinct from the average darwinian fitness of its members, as the latter is determined by the outcome of selection within demes and the former by the outcome of selection among demes. One component of a deme's mean selective value is its ability to sustain a large and productive population to serve as a source of migration, colonization, or competition relative to other demes. Since the Carson method measures equilibrium biomass and productivity, it is reasonable to speculate that the measures are more closely related to mean selective value than to darwinian fitness, and this point has been emphasized previously by CARSON (1961b). This interpretation requires inde-

pendent verification, however.

All of our fitness tests are relatively short-term tests ranging from one to 10 or 12 generations. The tests requiring multiple generations are the Carson tests and the Sved tests, and there is the possibility that significant changes in fitness could occur during this period. As WRIGHT (1931) emphasized long ago in connection with single-gene models in population genetics: "Selection, whether in mortality, mating or fecundity, applies to organisms as a whole and thus to the effects of the entire gene system rather than to single genes. A gene which is more favorable than its allelomorph in one combination may be less favorable in another. Even in the case of cumulative effects, there is generally an optimum grade of development of the character and a given plus gene will be favorably selected in combinations below the optimum but selected against in combinations above the optimum. Again the greater the number of unfixed genes in a population, the smaller must be the average effectiveness of selection for each one of them. The more intense the selection in one respect, the less effective it can be in others. The selection coefficient for a gene is thus in general a function of the entire system of gene frequencies. *As a first approximation relating to a given population at a given moment* [emphasis ours], one may, however, assume a constant net selection coefficient for each gene." Since the fitness effects of single genes may change, it follows that the average fitness of an entire population may change unless the population in question is very highly inbred. This being the case, we have rather more confidence in the single-generation measures of fitness than in the multigeneration measures.

The overall story, then, is that the 10 measures of fitness do not assess the same biological phenomena. Only the intraspecific tests are allied with classical darwinian fitness as evidenced by their high correlation with the composite index based on individual fitness components. The interspecific tests obviously measure interspecific competitive ability, and this is weakly correlated or uncorrelated with darwinian fitness. The Carson population tests are noncompetitive and seem to measure something quite distinct from the others, possibly though not necessarily related to mean selective value. The interpretation of the Sved test is problematical. It measures something distinct from all the other tests that is likely to be important in some particular contexts, but, beyond the context of experimental measurement, it is not clear to what extent the test results are associated with any presently understood aspect of fitness or evolutionary success. This in no way diminishes the value of the Sved test in demonstrating chromosomal heterosis, which, after all, was its original purpose.

ACKNOWLEDGEMENT

This research was supported by grant number GM30351 from the National Institutes of Health.

LITERATURE CITED

AYALA, F. J. 1965a Evolution of fitness in experimental populations of *Drosophila serrata*. Science *150*: 903-905.

AYALA, F. J. 1965b Relative fitness of populations of *Drosophila serrata* and *Drosophila birchii*. Genetics *51*: 527-544.

AYALA, F. J. 1966a Evolution of fitness. I. Improvement in the productivity and size of irradiated populations of *Drosophila serrata* and *Drosophila birchii*. Genetics *53*: 883-895.

AYALA, F. J. 1966b Reversal of dominance in competing species of *Drosophila*. Am. Nat. *100*: 81-83.

AYALA, F. J. 1970 Population fitness of geographic strains of *Drosophila serrata* as measured by interspecific competition. Evolution *24*: 483-494.

CARSON, H. L. 1958 Increase in fitness in experimental populations resulting from heterosis. Proc. Nat. Acad. Sci. U.S.A. *44*: 1136-1141.

CARSON, H. L. 1961a Relative fitness of genetically open and closed populations of *Drosophila robusta*. Genetics *46*: 553-567.

CARSON, H. L. 1961b Heterosis and fitness in experimental populations of *Drosophila melanogaster*. Evolution *15*: 496-509.

DOBZHANSKY, Th. 1955 A review of some fundamental concepts and problems of population genetics. Cold Spring Harbor Symp. Quant. Biol. *20*: 1-15.

DOBZHANSKY, Th. 1970 Genetics of the Evolutionary Process. Columbia University Press, New York.

DYKHUIZEN, D. and D. L. HARTL 1980 Selective neutrality of 6PGD allozymes in *E. coli* and the effects of genetic background. Genetics *96*: 801-817.

EWENS, W. J. 1979 Mathematical Population Genetics. Springer-Verlag, New York.

FISHER, R. A. 1930 The Genetical Theory of Natural Selection. Clarendon Press, Oxford.

HARTL, D. L. 1980 Principles of Population Genetics. Sinauer Associates, Sunderland, Massachusetts.

HARTL, D. L. and D. E. DYKHUIZEN 1981 Potential for selection among nearly neutral allozymes of 6-phosphogluconate dehydrogenase in *Escherichia coli*. Proc. Nat. Acad. Sci. U.S.A. *78*: 6344-6348.

HARTL, D. L. and H. JUNGEN 1979 Estimation of average fitness of populations of *Drosophila melanogaster* and the evolution of fitness in experimental populations. Evolution *33*: 371-380.

HAYMER, D. S. 1982 Experimental measures of fitness in *Drosophila*. Ph. D. Thesis, Purdue University, West Lafayette, Indiana.

HAYMER, D. S. and D. L. HARTL 1982 The experimental assessment of fitness in *Drosophila*. I. Comparative measures of competitive reproductive success. Genetics *102*: 455-466.

HAYMER, D. S. and D. L. HARTL 1983 The experimental assessment of fitness in *Drosophila*. II. A comparison of competitive and noncompetitive measures. Submitted to Genetics.

HOLM, D. G. and A. CHOVNICK. 1975 Compound autosomes in *Drosophila melanogaster*: The meiotic behavior of compound thirds. Genetics *81*: 293-311.

JUNGEN, H. and D. L. HARTL 1979 Average fitness of populations of *Drosophila melanogaster* as estimated using compound-auto-

some strains. Evolution *33*: 359-370.
KNIGHT, G. R. and A. ROBERTSON 1957 Fitness as a measurable character in *Drosophila*. Genetics *42*: 524-530.
LEWONTIN, R. C. 1974 The Genetic Basis of Evolutionary Change. Columbia University Press, New York.
MOORE, J. A. 1952 Competition between *Drosophila melanogaster* and *Drosophila simulans*. II. Improvement of competitive ability through selection. Proc. Nat. Acad. Sci. U.S.A. *38*: 813-817.
MUKAI, T. 1964 The genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rate of polygenes controlling viability. Genetics *50*: 1-19.
MUKAI, T. 1969 The genetic structure of natural populations of *Drosophila melanogaster*. VII. Synergistic interaction of spontaneous mutant polygenes controlling viability. Genetics *61*: 749-761.
MUKAI, T., S. I. CHIGUSA and S. I. KUSAKABA 1982 The genetic structure of natural populations of *Drosophila melanogaster*: XV. Nature of developmental homeostasis for viability. Genetics *101*: 279-300.
PIMENTEL, D., E. H. FEINBERG, P. W. WOOD and J. T. HAYES 1965 Selection, spacial distribution, and the coexistence of competing fly species. Am. Nat. *99*: 97-109.
PROUT, T. 1971a The relation between fitness components and population prediction in *Drosophila*. I: The estimation of fitness components. Genetics *68*: 127-149.
PROUT, T. 1971b The relation between fitness components and population prediction in *Drosophila*. II: Population prediction. Genetics *68*: 151-167.
SIEGEL, S. 1956 Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill, New York.
SIMMONS, M. J. and J. F. CROW 1977 Mutations affecting fitness in *Drosophila* populations. Ann. Rev. Genet. *11*: 49-78.
SVED, J. A. 1971 An estimate of heterosis in *Drosophila melanogaster*. Genet. Res., Camb. *18*: 97-105.
SVED, J. A. 1975 Fitness of third chromosome homozygotes in *Drosophila melanogaster*. Genet. Res., Camb. *25*: 197-200.
SVED, J. A. and F. J. AYALA 1970 A population cage test for heterosis in *Drosophila pseudoobscura*. Genetics *66*: 96-113.
TAYLOR, L. W. 1959 Physics: The Pioneer Science. Dover, New York.
THODAY, J. M. 1953 Components of fitness. Symp. Soc. Exptl. Biol. *7*: 96-113.
TRACEY, M. L. and F. J. AYALA 1974 Genetic load in natural populations: Is it compatible with the hypothesis that many polymorphisms are maintained by natural selection? Genetics *77*: 569-589.
VAN DELDEN, W. and J. S. BEARDMORE 1968 Effects of small increments of genetic variability in inbred populations of *Drosophila melanogaster*. Mutat. Res. *6*: 117-127.
WADE, M. J. 1976 Group selection among laboratory populations of *Tribolium*. Proc. Nat. Acad. Sci. U.S.A. *73*: 4604-4607.
WADE, M. J. 1977 An experimental study of group selection. Evolution *31*: 134-153.
WADE, M. J. 1978 A critical review of the models of group selection. Quart. Rev. Biol. *53*: 101-104.
WALLACE, B. 1981 Basic Population Genetics. Columbia University Press, New York.

WILTON, A. N. and J. A. SVED 1979 X-chromosomal heterosis in *Drosophila melanogaster*. Genet. Res., Camb. *34*: 303-315.
WRIGHT, S. 1931 Evolution in Mendelian populations. Genetics *16*: 97-159.
WRIGHT, S. 1969 Evolution and the Genetics of Populations. Vol. 2. The Theory of Gene Frequencies. University of Chicago Press, Chicago.