THEORY OF THE ALLELISM BETWEEN DROSOPHILA LETHALS COLLECTED AT DIFFERENT TIMES

TIMOTHY PROUT

Department of Life Sciences, University of California, Riverside, California 92502

Received February 20, 1967

IN a recent paper WALLACE (1966a) presented an analysis of the allelism of recessive lethal genes in certain population of *Drosophila melanogaster*. His specific concern was the allelism rates of lethals taken from a natural population at collection sites of varying distances apart. WALLACE argues that the allelism between two sets of lethals from a given distance apart can be related to the allelism of two sets of lethals taken from the same point in the population but separated by a given interval in time.

In connection with this latter relationship WALLACE further conjectures that the allelism rate i_{r_t} and the time interval t (in generation units) between the sampling of the first and second sets of lethals have the following functional relationship:

$$i_{r_t} = i_n + i_F(1-K)^t \quad (1)$$

where

i_{r_t} $\equiv$ the allelism between lethal genes t generations apart

i_F $\equiv$ the allelism component of i_{r_t} within a generation produced by identity by descent in a finite population

i_n $\equiv$ the allelism within a generation of the same set of lethal loci if they were in a population of infinite size

K $\equiv$ a constant described by WALLACE as "related to the rate of turnover of lethals within the population".

Note: The lower case "i" used here to denote gene allelism is comparable to the upper case "I" of WALLACE and of CROW and TEMIN (1964) which denotes chromosome allelism.

It is the purpose of this article to demonstrate that this functional relationship (equation 1) conjectured by WALLACE is approximately correct. Furthermore, as an outcome of this demonstration the parameter K will be defined in terms of the basic population parameters of selection and mutation.

One of the important approximations necessary for the derivation that follows is that for all of the lethal loci the change in lethal gene frequency, Δq between successive generations is a linear function of the gene frequency, q in the first of the two generations. This approximation will be assumed to be true at first for the purposes of the derivation and then the approximation itself will be examined.

Linear Δq approximation: The relationship between the lethal gene frequencies at a given locus between successive generations may be expressed:

$$q_1 = q + \Delta q + \delta q \quad (2)$$

where

q $\equiv$ gene frequency at generation 0

q_1 $\equiv$ gene frequency at generation 1

Δq $\equiv$ changes in gene frequency due to systematic effects

δq $\equiv$ change in gene frequency due to chance alone.

Assume that Δq is a linear function of q as follows:

$$\Delta q = k(\hat{q} - q) \quad (3)$$

where

k $\equiv$ a constant

$\hat{q}$ $\equiv$ the equilibrium gene frequency which should be achieved due to the systematic effects impinging in that locus.

Substituting (3) into (2) gives

$$q_1 = q + k(\hat{q} - q) + \delta q.$$

This is a linear recurrence equation whose general term can easily be found to be

$$q_t = \hat{q} - (\hat{q} - q)(1 - k)^t + \delta q_t \quad (4)$$

where

t $\equiv$ the t th term in the series or in this case number of generations from the start

δq_t $\equiv$ the accumulated effects of chance over t generations.

Derivation: Consider the frequency of the lethal allele at a given locus among all lethal alleles at all loci in generation 0. This is a probability, p , which can be expressed as a function of the lethal allele frequency, q , in the whole population (as opposed to among lethal alleles only) and n , the number of loci involved, thus

$$p = q/\Sigma q = q/n\bar{q}$$

where q $\equiv$ the mean lethal allele frequency over all loci. Note: $p \neq 1 - q$.

The same probability can be similarly expressed for the lethal allele at the same locus t generations hence

$$p_t = q_t/n\bar{q}$$

It is assumed the population is in equilibrium and that therefore q is constant over this interval of t generations.

The probability of choosing one and the same lethal from among lethals of generation 0 and from among lethals of generation t is pp_t ; and

$$pp_t = (qq_t)/(n\bar{q})^2. \quad (5)$$

This is the probability of allelism at a given locus between the two generations. The total probability of allelism over all loci, or just plain allelism, i_{rt} is

$$i_{rt} = \Sigma pp_t$$

substituting (5)

$$i_{rt} = (\Sigma qq_t)/(n\bar{q})^2.$$

Since, generally speaking, $\sum_{i=1}^{i=n} x_i = nE(x)$

$$i_{rt} = \frac{nE(qq_t)}{(n\bar{q})^2} = \frac{E(qq_t)}{n\bar{q}^2} \quad (6)$$

$$n\bar{q}^2 i_{rt} = E(qq_t). \quad (6A)$$

The problem, then, becomes one of finding $E(qq_t)$. This will be done in two

stages. First, it is necessary to find the conditional expectation for all those loci sharing one given equilibrium value $\hat{q}$. Then the expectation will be taken over all $\hat{q}$.

It is at this point that it is necessary to introduce the assumption of linear Δq expressed by equation 4. Substituting this expression, equation 4, for q_t in equation 6A and recognizing the conditional expectation, we have

$$E(qq_t|\hat{q}) = E(q[\hat{q} - (\hat{q} - q)(1-k)^t + \delta q_t]) \quad (7)$$

$$= E(q\hat{q}) - E(q\hat{q}[1-k]^t) + E(q^2[1-k]^t) + E(q\delta q_t) \quad (7A)$$

Consider the last term first. Since q and δq_t are uncorrelated

$$E(q\delta q_t) = E(q) E(\delta q_t).$$

The chance deviations δq_t have expectation of zero so that

$$E(q\delta q_t) = E(q) E(\delta q_t) = E(q)(0) = 0.$$

Now consider the first term of (7A). Since the population is in equilibrium the expectation of the q 's is their common equilibrium value, $\hat{q}$.

$$E(q\hat{q}) = \hat{q}^2$$

Applying this relationship and the fact that $E(q^2) = \hat{q} + \sigma_q^2$, equation 6A reduces to:

$$E(qq_t|\hat{q}) = \hat{q}^2 + \sigma_q^2(1-k)^t$$

where $\sigma_q^2 \equiv$ variance of allele frequencies, q , around their mean which is $\hat{q}$.

The variance here is strictly due to the effects of chance since we are dealing with a set of loci subject to exactly the same systematic effects and thus having a common equilibrium, $\hat{q}$.

It now remains to take the expectation over all such sets.

$$\begin{aligned} E(qq_t) &= E(\hat{q}^2) + E(\sigma_q^2[1-k]^t) \\ &= \bar{q}^2 + \sigma_q^2 + \sigma^2(1-K)^t \end{aligned} \quad (8)$$

where

$$\sigma^2 = E(\sigma_q^2)$$

$\sigma^2 \equiv$ variance in equilibrium values due to real differences in systematic effects

$$K = E(k).$$

Another assumption has been made above that σ_q^2 and k are not strongly correlated.

The final expression for allelism, $i r_t$ can now be constructed by substituting (8) for $E(qq_t)$ into equation 6A.

$$\begin{aligned} i r_t &= \frac{1}{n\bar{q}^2} (\bar{q}^2 + \sigma_q^2 + \sigma^2[1-K]^t) \\ &= \frac{1}{n} + \frac{\sigma_q^2}{n\bar{q}^2} + \frac{\sigma^2}{n\bar{q}^2} (1-K)^t \end{aligned} \quad (9)$$

For the special case of allelism between lethals taken from the same generation, when $t = 0$, equation (9) reduces to the expression originally derived by WRIGHT,

DOBZHANSKY and HOVANITZ (1942). As these authors pointed out, the first two terms $\left(\frac{1}{n} + \frac{\sigma_q^2}{n\bar{q}^2}\right)$ represent the theoretical expectation for the allelism of these same loci in an infinite population. The first factor of the third term $\left(\frac{\sigma^2}{n\bar{q}^2}\right)$ represents the added increment of allelism due to chance effects or random drift produced by restriction of population (or neighborhood) size. Thus it is seen that the functional relationship, equation 1, conjectured by WALLACE is correct assuming the linear Δq relationship. The parameters of WALLACE's equation therefore have the following more fundamental and usual (WRIGHT, DOBZHANSKY and HOVANITZ) interpretation:

$$i_F = \frac{\sigma^2}{n\bar{q}^2}$$

$$i_n = \frac{1}{n} + \frac{\sigma_q^2}{n\bar{q}^2} \quad (10)$$

$K = E(k)$, i.e. the mean k over all loci expressing, as seen in equation 3, the intensity at which individual genes are driven back each generation towards their equilibrium values by systematic effects alone.

The assumption of linear Δq : It now remains to examine the assumption that Δq is linear and to find a more fundamental interpretation of K . The exact expression for Δq for a given lethal allele is:

$$\Delta q = \frac{\mu - qs(1+\mu) - q^2(1-2s)}{1 + q(1-2s)} \quad (11)$$

where $\mu \equiv$ mutation to the lethal allele from the normal allele, and $s \equiv$ selection coefficient of heterozygotes.

The equilibrium, $\hat{q}$, given by (11) is

$$\hat{q} = \frac{-s(1+\mu) \pm \sqrt{s^2(1+\mu)^2 + 4\mu(1-2s)}}{2(1-2s)} \quad (12)$$

(The sign of the root is given by the sign of s .)

The graphical relationship between the exact Δq and the linear approximation is sketched in Figure 1. As indicated in the sketch the linear approximation is arranged so as to be tangent to the exact Δq curve at the point $\bar{q} = \hat{q}$.

This procedure implies that the expected equilibrium in a finite population is that given by (12), while WRIGHT (1937) has shown that for a complete recessive ($s=0$), the expected equilibrium is somewhat depressed by the inbreeding in a small population. It is believed that this discrepancy will not account for large errors in the argument which follows.

As indicated by equation 3, the slope of the linear approximation is $-K$, which can now be evaluated in terms of the parameters s and μ of the expression for the exact Δq (equation 11) in the following way:

$$K = \left. \frac{d\Delta q}{dq} \right|_{\hat{q}}$$

$$= \frac{s(1+\mu) + 2\hat{q}(1-2s)}{1 + \hat{q}(1-2s)} \quad (13)$$

FIGURE 1.—Sketch of two Δq functions—one exact, and one a linear approximation.

Substituting (12) for $\hat{q}$, approximating the denominator of (13) by unity, and dropping second order terms gives

$$K = \sqrt{s^2 + 4\mu} \quad \text{approximately.} \quad (14)$$

(The sign of the root is given by the sign of s .) Thus WALLACE's K can now be interpreted in the more fundamental terms of the selection and mutation affecting the lethals under study. K can be further simplified if it be assumed that most of the lethals are semidominant, or most are completely recessive, or most are heterotic. Under these three circumstances K simplifies in the following way:

- (A) Most loci semidominant, s positive and $\gg \mu$

$$K = s \text{ approximately}$$

- (B) Most loci completely recessive, $s = 0$

$$K = 2\sqrt{\mu} \text{ approximately}$$

- (C) Most loci heterotic, s negative and $-s \gg \mu$

$$K = -s \text{ approximately}$$

Unfortunately, estimates of K derived from data appropriate to this model will not decide between the important question of average semidominance or average heterosis, i.e. K is always positive so that the sign of s is ambiguous. Only if K turned out to be of the order of twice the square root of ordinary mutation rates, then one could conclude that neither semidominance nor heterosis prevailed.

A correction for the linear approximation: Referring to Figure 1 it can be seen that the value of Δq given by the linear approximation will be nearly correct only for those loci whose alleles are close to their equilibrium values, $\hat{q}$. For those genes which are some distance on either side of the equilibrium, the linear approximation gives a Δq which is always too large compared to the exact Δq .

Recall that the allelism, i_{T_t} , is directly proportional to $E(qq_t)$ (equation 6).

Restricting the case to allelism between successive generations, $E(qq_l)$ may be rewritten:

$$E(qq_l) = Eq(q + \Delta q) = E(q^2) + E(q \Delta).$$

If under the linear approximation, Δq is always too large then $E(q\Delta q)$ is too large and so i_{T_t} is too large. This means that the actual i_{T_t} should fall off faster with increasing t than the theoretical equation 9 predicts. A correction in equation 9 which would cause it to decrease faster with increasing t would be one which reduces the size of the factor $(1 - K)$. An approximate correction of this sort may be obtained by replacing $(1 - K)^t$ in equation (9) with $(1 - K - \bar{q})^t$. This was found by the writer to be entirely satisfactory for a few numerical cases.

It should be noted that the magnitude of the correction increases with increasing $\bar{q}$. This is so because the linear approximation to Δq progressively deteriorates as one passes from a situation in which most loci are semidominant ($s = 1/2$) through recessivity ($s = 0$) to heterosis (s is negative). Also the average equilibrium, $\bar{q}$ increases through this sequence.

Thus the correction suggested above becomes more important as the linear approximation to Δq deteriorates.

DISCUSSION

It should be emphasized that the above considerations apply only to WALLACE's conjecture concerning the functional relationship between allelism and time (equation 1). On the other hand, WALLACE's data deal strictly with allelism of lethals sampled at one point in time but at different distances apart. WALLACE further proposes, on the basis of his studies of dispersal (WALLACE 1966b), that this same functional relationship (equation 1) can be applied to his data with the substitution of the square root of distance for time. For the rationale for this interesting additional proposal the reader is referred to WALLACE's paper. It should suffice to point out here, that although the distance-allelism data fit a function of the form of equation 1 and serve well for an estimate of WALLACE's "zero point allelism", it is not possible to estimate K since there is an unknown constant of proportionality connecting time with the square root of distance. It is clear that this distance-allelism relationship, already put to good use by WALLACE, still requires additional theoretical and experimental study.

However, this should not detract from the fact that the time-allelism relationship proposed by WALLACE and which is the subject this article is a new and potentially powerful tool for further analysis of *Drosophila* population structure. The experiments which this theory calls for are straight forward, although laborious. One needs only to sample lethals from a population at different times and perform very extensive allelism tests within and between such samples.

Figure 2 depicts the hypothetical data produced by such an experiment. The solid line represents the line which best fits the data assuming the functional relationship represented by equation 1 to be correct. A test of this model is obtained by a test of goodness of fit. For instance, a poor fit might indicate an interesting gradual or sudden systematic change in composition of lethals. On

FIGURE 2.—Sketch of hypothetical data, X 's, from a time-allelism study; solid line, the best fit line assuming equation 1.

the other hand, a good fit would open up a number of other possibilities. Estimates of the constants, i_n and i_F could be set equal to their expectations shown by equations (10). Also the same could be done with the constant, K (equation 14). With these, plus the information on lethal frequencies and mutation rates it would be possible to obtain information about all five of the fundamental parameters n , σ_q^2 , σ^2 , s and μ on a single population as it exists through time. Heretofore, for a given small population it has not been possible to estimate i_n , the allelism of that population's lethal loci if it were of infinite size. Without this time-allelism information in the past, it has been necessary to compare small experimental populations with large ones (PROUT 1954) assuming the populations were the same except for size, or in studies of natural populations (WRIGHT, DOBZHANSKY and HOVANITZ 1942; PAVAN and KNAPP 1954) i_n is obtained from allelism between remote localities in which case the assumption must be made that all the localities are the same.

Finally, even more information should emerge from the simultaneous use of both the time-allelism and distance-allelism tools on a given wide-spread natural population. More theoretical work is needed here not only on the distance-allelism relationship, but also on the time-allelism relationship since the considerations of this article apply only to strictly isolated populations. Nevertheless, it is conceivable that with the use of both tools one could obtain information on natural generation time and neighborhood size.

SUMMARY

WALLACE (1966a) has conjectured that recessive lethals extracted from *Drosophila* populations at different times will exhibit decreasing rates of allelism, i_{T_t} , the greater the time interval t between samples, according to the functional relationship $i_{T_t} = i_n + i_F (1 - K)^t$, where $i_F \equiv$ the component of i_{T_t} due to identity by descent in the finite population, $i_n \equiv$ the allelism shown by the same loci if the population were of infinite size, and $K \equiv$ a constant described by WALLACE as "related to the rate of turnover of lethals within the population".—

A proof is provided that WALLACE's conjecture is approximately correct. It is shown that the approximation entails the assumption that the change in gene frequency Δq is a linear function of the gene frequency q . It is further shown that under this assumption WALLACE's K is related to the selection impinging on heterozygotes, s and the mutation rate to lethality, μ in the following way: $K = \sqrt{s^2 + 4\mu}$ (approximately). (The sign of the root is given by the sign of s .)—It is pointed out that this time allelism relationship could be a powerful tool in further analyses of *Drosophila* population structure.

LITERATURE CITED

- CROW, J. F., and R. G. TEMIN, 1964 Evidence for the partial dominance of recessive lethal genes in natural populations of *Drosophila*. *Am. Naturalist* **98**: 21–33.
- PAVAN, C., and E. N. KNAPP, 1954 The genetic population structure of Brazilian *Drosophila willistoni*. *Evolution* **8**: 303–314.
- PROUT, T., 1954 Genetic drift in irradiated experimental populations of *Drosophila melanogaster*. *Genetics* **39**: 529–546.
- WALLACE, B., 1966a Distance and the allelism of lethals in a tropical population of *Drosophila melanogaster*. *Am. Naturalist* **100**: 565–578. — 1966b On the dispersal of *Drosophila*. *Am. Naturalist* **100**: 551–563.
- WRIGHT, S., 1937 The distribution of gene frequencies in natural populations. *Proc. Natl. Acad. Sci. U.S.* **23**: 307–320.
- WRIGHT, S., TH. DOBZHANSKY, and W. HOVANITZ, 1942 Genetics of natural populations. VII. The allelism of lethals in the third chromosome of *Drosophila pseudoobscura*. *Genetics* **27**: 363–395.