

## Genetic loads at a polymorphic locus which is maintained by frequency-dependent selection\*

BY MOTOO KIMURA AND TOMOKO OHTA  
*National Institute of Genetics, Mishima, Japan*

(Received 14 November 1969)

### SUMMARY

If a polymorphic locus is maintained in finite populations by frequency-dependent selection with selective neutrality at equilibrium, it is generally accompanied by two genetic loads, i.e. the dysmetric and the drift loads. The former arises because the fitness of the population may not be at a maximum at the equilibrium gene frequency and the latter because genetic drift in small populations displaces the gene frequency from its equilibrium value.

In some simple models of frequency-dependent selection considered, the drift load is independent of selection coefficients and is approximately equal to  $(n-1)/(2N_e)$ , where  $n$  is the number of alleles and  $N_e$  is the effective population size.

As a possible mechanism for maintaining genetic variability in Mendelian populations, a form of frequency-dependent selection in which each allele becomes advantageous when rare was considered many years ago by several authors such as Wright & Dobzhansky (1946), Wright (1949), and Haldane (1954). In particular, in their analysis of chromosome polymorphism in *Drosophila*, Wright & Dobzhansky (1946) considered, as an alternative to overdominance, a model of frequency-dependent selection in which fitnesses of various genotypes become equal at equilibrium.

Recently, this type of frequency-dependent selection with selective neutrality at equilibrium has attracted much attention since Kojima & Yarbrough (1967) proposed such a model to explain the prevalence of isozyme polymorphism in natural populations.

This model has an advantage of having apparently no genetic loads at equilibrium yet retaining enough selective force to keep polymorphism if gene frequencies deviate sufficiently from their equilibrium values.

In actual populations, however, it is expected that gene frequencies may deviate from their equilibrium as well as from 'optimal' values and this may create some genetic load. In fact, we intend to show in this report that such a genetic system generally is accompanied by two kinds of genetic load that may be called dysmetric and drift loads: the dysmetric load (term first suggested by J. B. S. Haldane in 1959

\* Contribution No. 750 from the National Institute of Genetics, Mishima, Shizuoka-ken, Japan. Aided in part by a Grant-in-Aid from the Ministry of Education, Japan.

in his personal communication to J. F. Crow) arises if the equilibrium gene (or genotype) frequencies differ from the optimum frequencies giving the maximum mean fitness of individuals. The drift load arises if gene frequencies deviate from their respective equilibrium values due to random genetic drift in finite populations. Recently, Robertson (1970) has shown that for an overdominant locus with a pair of alleles this load is approximately equal to  $1/(4N_e)$ , where  $N_e$  is the effective population size.

Let us first investigate the model considered by Wright & Dobzhansky (1946) in their analysis of the polymorphism involving ST and CH chromosomes. Let  $A_1$  and  $A_2$  be a pair of alleles and assume that the absolute fitnesses (measured in selective values) of the three genotypes  $A_1A_1$ ,  $A_1A_2$  and  $A_2A_2$  are respectively  $C(1+a-bp)$ ,  $C$  and  $C(1-a+bp)$ , where  $a$ ,  $b$  and  $C$  are constants and  $p$  stands for the frequency of  $A_1$ . Under random mating, the mean fitness is

$$\bar{w} = C\{1 - (a - bp)(1 - 2p)\}, \quad (1)$$

and the rate of change of  $p$  per generation by selection is

$$\Delta p = p(1-p)(a-bp)/\{1 - (a-bp)(1-2p)\}. \quad (2)$$

Thus the equilibrium gene frequency is  $\hat{p} = a/b$  and the corresponding mean fitness is  $\hat{w} = C$ . In this state the three genotypes have the same fitnesses, and also the equilibrium is stable.

Note, however, that the gene frequency at equilibrium ( $\hat{p}$ ) is not generally the one which gives the maximum mean fitness of individuals. In fact, the maximum occurs when the frequency of allele  $A_1$  is midway between 0.5 and  $\hat{p}$ . Namely, we have

$$\bar{w}_{\max} = C\{1 + (2a-b)^2/(8b)\}, \quad (3)$$

with  $p_{\max} = 1/4 + a/(2b)$ . Only when  $\hat{p} = 1/2 = a/b$  we have  $\hat{w} = w_{\max} = C$ . The dysmetric load (denoted by  $L_{\text{dys}}$ ) is defined as the relative amount of loss in fitness due to  $\hat{p}$  deviating from  $p_{\max}$ . Thus, we have

$$L_{\text{dys}} = (\bar{w}_{\max} - \hat{w})/\bar{w}_{\max} = (2a-b)^2/\{8b + (2a-b)^2\}. \quad (4)$$

For example, Wright & Dobzhansky (loc. cit.) estimated  $a = 0.902$  and  $b = 1.288$  for the relative fitnesses in the polymorphism involving ST and CH chromosomes in *Drosophila pseudoobscura*. Thus, applying formula (4), we get  $L_{\text{dys}} = 0.0252$  if we treat these values as absolute fitnesses. Namely, the dysmetric load in this case amounts to about 2.5%. Also, applying the present model to Kojima and Yarbrough's study on esterase 6 polymorphism in *D. melanogaster*, we made a rough estimate of  $a$  and  $b$  using their Table 4. It gave  $a = 0.60$ ,  $b = 1.58$  and  $\hat{p} = 0.38$  where  $p$  stands for the frequency of F allele. From formula (4), we get  $L_{\text{dys}} = 0.0112$  or about 1.1%.

These examples suggest that the dysmetric load may be considerable if a large number of polymorphic loci are maintained by frequency-dependent selection, even if apparent selective neutrality is attained at equilibrium.

Next, let us investigate the drift load, that is to say, the amount of loss in fitness

in a finite population caused by random fluctuation of gene frequencies around their equilibrium values. Still using the same model, if we express the mean fitness as a function of  $p - \hat{p}$ , we have

$$\bar{w} = C\{1 + (b - 2a)(p - \hat{p}) - 2b(p - \hat{p})^2\}. \tag{5}$$

At steady state in which the tendency of  $p$  drifting away from  $\hat{p}$  by random sampling of gametes is counterbalanced by the selective force pushing  $p$  back to  $\hat{p}$ , the expected value of  $\bar{w}$  becomes

$$E(\bar{w}) = C\{1 - 2b(\mu'_2 - \hat{p}^2)\}, \tag{6}$$

where  $\mu'_2 = E(p^2)$  is the second moment of  $p$  about zero with respect to the steady state distribution. Note that  $\hat{p} = E(p)$ .

Assuming that deviation of  $p$  from  $\hat{p}$  is small,  $\mu'_2 - \hat{p}^2$  may be derived as follows. Let  $p$  be the frequency of  $A_1$  in the present generation, then the frequency of  $A_1$  in the next generation is  $p' = p + \Delta p + \xi$ , where  $\xi$  is the amount of change in gene frequency due to random sampling of gametes and it has the mean 0 and the variance  $p(1-p)/(2N_e)$ . Squaring this expression for  $p'$ , taking expectation, but neglecting the small term  $(\Delta p)^2$ , we obtain, at equilibrium in which  $E(p'^2) = E(p^2)$ ,

$$E\{2p\Delta p + p(1-p)/(2N_e)\} = 0. \tag{7}$$

At the neighbourhood of the equilibrium point, we have approximately

$$\Delta p = a(1 - \hat{p})(\hat{p} - p)$$

from (2). Therefore substituting this for  $\Delta p$  in (7), we get

$$\mu'_2 - \hat{p}^2 = \hat{p}(1 - \hat{p})/\{4N_e a(1 - \hat{p}) + 1\}, \tag{8}$$

which gives, in combination with (6),

$$\{\hat{\bar{w}} - E(\bar{w})\}/\hat{\bar{w}} = 2a(1 - \hat{p})/\{4N_e a(1 - \hat{p}) + 1\}. \tag{9}$$

We have assumed in the above treatments that deviation of the gene frequency from its equilibrium value is small. This amounts to the assumption, especially when  $\hat{p}$  is not very near to 0 or 1, that  $4N_e a(1 - \hat{p})$  is much larger than unity. Therefore we have, with good approximation

$$\{\hat{\bar{w}} - E(\bar{w})\}/\hat{\bar{w}} = 1/(2N_e), \tag{10}$$

where  $N_e$  is the 'variance' effective number of the population.

Formula (10) means that the drift load relative to the fitness at the equilibrium in an infinite population is equal to  $1/(2N_e)$ . Note that this load depends only on the effective population size but is independent of selection coefficients  $a$  and  $b$ . Note also that this value is twice as large as the corresponding value at an over-dominant locus (cf. Robertson, 1970). If the drift load is defined relative to the maximum fitness, we have

$$L_{\text{drift}} = \left(\frac{1}{2N_e}\right) \frac{\hat{\bar{w}}}{\bar{w}_{\text{max}}}$$

Unless  $\bar{w}_{\max}$  differs much from  $\hat{w}$ , this load is approximately equal to  $1/(2N_e)$ , i.e.

$$L_{\text{drift}} = 1/(2N_e). \tag{11}$$

In the above model, we have assigned absolute fitnesses  $C(1+a-bp)$ ,  $C$ , and  $C(1-a+bp)$  to three genotypes  $A_1A_1$ ,  $A_1A_2$ ,  $A_2A_2$ , but we can show, using the same procedure as above, that the drift load is unchanged if we assign  $C$ ,  $C(1-a+bp)$  and  $C\{1-2(a-bp)\}$  to the three genotypes, though the dysmetric load changes with this new assignment. These models of frequency-dependent selection may be called ‘gene-frequency dependent’ selection. On the other hand, selection may be genotype-frequency-dependent as considered by Clarke & O’Donald (1964). Two examples are as follows: (1) The heterozygotes are distinct and the absolute fitnesses of  $A_1A_1$ ,  $A_1A_2$  and  $A_2A_2$  are respectively  $C(1-sp^2)$ ,  $C(1-2sp(1-p))$  and  $C\{1-s(1-p)^2\}$  in which  $s$  is a positive constant. (2) There is complete dominance between alleles so that the absolute fitnesses of the recessive homozygote ( $A_1A_1$ ) and of the two genotypes involving the dominant ( $A_1A_2$ ,  $A_2A_2$ ) are respectively  $C(1-sp^2)$  and  $C\{1-s(1-p^2)\}$ . In both cases, we can show that under random mating the dysmetric load is zero, while the drift load is again approximately  $1/(2N_e)$  if  $s$  is small.

We will now show that the total load for frequency-dependent selection with selective neutrality at equilibrium may be expressed in general as a sum of dysmetric and drift loads, still assuming a pair of alleles.

Let  $p_{\max}$  be the gene frequency giving the maximum mean fitness. Then, at the neighbourhood of  $p_{\max}$ , the mean fitness of individuals as a function of  $p$  may be expressed in the form

$$\bar{w} = \bar{w}_{\max} - \alpha(p - p_{\max})^2, \tag{12}$$

where  $\alpha > 0$ . For the model of Wright & Dobzhansky (1946) considered above, this relation holds exactly for any gene frequency with  $\alpha = 2bC$ , but in general this is an approximation which is valid at the neighbourhood of  $p_{\max}$ . Writing

$$p - p_{\max} = (p - \hat{p}) + (\hat{p} - p_{\max})$$

and noting  $E(p - \hat{p}) = 0$ , the expected value of  $\bar{w}$  becomes

$$E(\bar{w}) = \bar{w}_{\max} - \alpha(\hat{p} - p_{\max})^2 - \alpha E\{(p - \hat{p})^2\}. \tag{13}$$

The total load is then

$$L = \frac{\bar{w}_{\max} - E(\bar{w})}{\bar{w}_{\max}} = \frac{\alpha(\hat{p} - p_{\max})^2}{\bar{w}_{\max}} + \frac{\alpha E\{(p - \hat{p})^2\}}{\bar{w}_{\max}}. \tag{14}$$

This is a sum of the dysmetric and the drift loads, that is

$$L_{\text{dys}} = (\hat{p} - p_{\max})^2 / \bar{w}_{\max}, \tag{15}$$

and

$$L_{\text{drift}} = \alpha E\{(p - \hat{p})^2\} / \bar{w}_{\max}. \tag{16}$$

Since, at the neighbourhood of the equilibrium point, we can write approximately  $\Delta p = -k(p - \hat{p})$  in which  $k$  is a constant, we get from (7),

$$E\{(p - \hat{p})^2\} = \mu'_2 - \hat{p}^2 = \hat{p}(1 - \hat{p}) / (1 + 4N_e k),$$

so that the formula for the drift load becomes

$$L_{\text{drift}} = \frac{\alpha \hat{p}(1 - \hat{p})}{\bar{w}_{\text{max}}(1 + 4N_e k)} \tag{17}$$

So far, we have assumed a pair of alleles, but the above treatments may readily be extended to multi-allelic cases.

Let us consider an  $n$ -allelic system with alleles  $A_1, A_2, \dots, A_n$ , and assume that the absolute fitness ( $w_{ij}$ ) of genotype  $A_i A_j$  is given by  $C(1 - s_i p_i - s_j p_j)$ , where  $s_i$  and  $s_j$  are positive constants and  $p_i$  and  $p_j$  are the frequencies of alleles  $A_i$  and  $A_j$  ( $i, j = 1, 2, \dots, n$ ). An equivalent model of frequency-dependent selection was considered by Wright (1949).

Under random mating, the mean fitness of individuals and the rate of change in gene frequency are respectively

$$\bar{w} = C(1 - 2v),$$

and

$$\Delta p_i = -p_i(s_i p_i - v)/(1 - 2v),$$

where

$$v = \sum_{i=1}^n s_i p_i^2.$$

At equilibrium we have  $\hat{w} = C(1 - 2\hat{v})$  and  $\hat{p}_i = \hat{v}/s_i$ , where

$$\hat{v} = 1 / \sum_{i=1}^n (1/s_i). \tag{18}$$

In this case the set of equilibrium frequencies ( $\hat{p}_i$ 's) is the one which gives the maximum mean fitness of individuals, as may easily be seen by writing  $\bar{w}$  in the form

$$\bar{w} = C \left\{ 1 - 2\hat{v} - 2 \sum_{i=1}^n s_i (p_i - \hat{p}_i)^2 \right\}. \tag{19}$$

Therefore, there is no dysmetric load, i.e.  $\bar{w}_{\text{max}} = \hat{w}$ . The drift load is

$$L_{\text{drift}} = \{\hat{w} - E(\bar{w})\}/\hat{w} = 2 \left\{ \sum_{i=1}^n s_i E(p_i^2) - \hat{v} \right\} / (1 - 2\hat{v}). \tag{20}$$

If we note that  $\Delta p_i \approx -k(p_i - \hat{p}_i)$ , in which  $k = \hat{v}/(1 - 2\hat{v})$ , and use equation (7) in which  $p_i$  is substituted for  $p$ , we get

$$\sum_i s_i E(p_i^2) = \hat{v}(n + 4N_e k)/(1 + 4N_e k).$$

Thus, we obtain from (20),

$$L_{\text{drift}} = \frac{2(n - 1)\hat{v}}{(4N_e - 2)\hat{v} + 1},$$

or more simply, as a good approximation,

$$L_{\text{drift}} = (n - 1)/(2N_e), \tag{21}$$

assuming that  $N_e$  is large enough so that  $4N_e \hat{v} \gg 1$ . This reduces to (11) if the number of alleles is two ( $n = 2$ ). Note that the drift load given by (21) is twice as

large as the corresponding value obtained by Robertson (1970) for multi-allelic overdominant system.

We would like to thank Dr Alan Robertson for helpful suggestions with regard to the manuscript.

#### REFERENCES

- CLARKE, B. & O'DONALD, P. (1964). Frequency-dependent selection. *Heredity* **19**, 201–206.
- HALDANE, J. B. S. (1954). The statics of evolution. In *Evolution as a Process* (Ed. J. Huxley *et al.*), pp. 109–121. London: George Allen and Unwin.
- KOJIMA, K. & YARBROUGH, K. M. (1967). Frequency-dependent selection at the esterase 6 locus in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, **57**, 645–649.
- ROBERTSON, A. (1970). The reduction of fitness from genetic drift at heterotic loci in small populations. *Genetical Research* **15**, 257–260.
- WRIGHT, S. (1949). Adaptation and selection. In *Genetics, Paleontology and Evolution* (Ed. G. Jepsen *et al.*), pp. 365–389. Princeton: Princeton University Press.
- WRIGHT, S. & DOBZHANSKY, TH. (1946). Genetics of natural populations. XII. Experimental reproduction of some of the changes caused by natural selection in certain populations of *Drosophila pseudoobscura*. *Genetics* **31**, 125–156.