POLYMORPHISM DUE TO SELECTION OF VARYING DIRECTION

By J. B. S. HALDANE and S. D. JAYAKAR
Genetics and Biometry Laboratory, Government of Orissa, Bhubaneswar

Introduction

Geneticists are showing increasing interest in genetically determined polymorphism. This is especially true of human geneticists, since polymorphism is often associated with a loss of fitness, sometimes described as genetic load. In recent discussions the following causes of more or less stable polymorphism have been distinguished.

(1) The conflict between selection and mutation. An equilibrium is reached when in each generation in a population of given number as many "mutant" genes are destroyed by selection as appear by mutation. The conflict between selection and the abnormal Mendelian segregation now called "meiotic drive" is analogous.

(2) The conflict between selection and segregation. The best known example occurs when a heterozygote is fitter than either of the two corresponding homozygotes. But where several allelomorphs are concerned matters are more complicated, and, as we hope to show later in some detail, the heterozygote need not be fitter in order to stabilize polymorphism when selective intensities differ in the two sexes.

(3) The conflict between fitness and abundance. A genotype may become less fit as it becomes more abundant. The most striking case is perhaps genetically determined self-sterility in plants, but Batesian mimicry probably acts in a similar way, mimics being better protected when greatly outnumbered by distasteful models than when mimics are common. So does resistance to infectious disease if the pathogen can adapt itself fairly rapidly so as to attack the commonest genotypes preferentially.

(4) The conflict between selection and migration. If each of two genotypes is fitter in different areas, migration between them may lead to a cline which may be polymorphic.

(5) The conflict between selections in different directions in the diploid generation and the haploid generation (usually pollen tubes) or in the two sexes.

Finally a polymorphism may be transient. A genotype which was formerly common may be disappearing as the result of changed conditions.

These are not, however, the only conditions which may lead to long-lasting genetic polymorphism, nor is the account of the five situations described above very satisfactory. We hope, in a series of papers, of which this is the first, to add some precision to the accounts of polymorphism so far given by Haldane and others. The present paper deals with polymorphism caused by selection which sometimes favours one phenotype and sometimes another, or in the general terminology of this introduction by conflict between selection in two directions.

The General Theorem

Suppose that the population consists of two phenotypes P and Q, and that the fitness of Q relative to P is F_n in the nth generation, but that F_n sometimes exceeds unity, and sometimes falls short of it. If P and Q are clones or pure lines which do not interbreed; or genotypes of a haploid differing in respect of a single gene pair, and if the ratio of P to Q in generation n is u_n, then $u_{n+1}=F_n^{-1}u_n$.

$$\text{So} \quad u_n = \left[\prod_{r=0}^{n-1} F_r\right]^{-1} u_o \tag{1.1}$$

that is to say u_n is equal to u_o divided by the product of the values of F_n. It is most improbable that this will be exactly unity. As time goes on it is almost certain that the product of the values of F_n will be much greater or much less than unity, and P or Q, as the case may be, will disappear. We can also think in terms of the geometric mean. If this differs even slightly from unity over sufficiently many generations, one or other of the phenotypes will disappear. In practice of course it is possible that the rarer phenotype will be restricted to a niche where it is at an advantage, leading to an equilibrium of type (4) considered above.

Now consider a large* diploid random mating population, so large that the probability of the extinction of a temporarily rare gene by random fluctuation is negligible. Let a pair of allelomorphs **A** and **a** at an autosomal locus occur in the gametes which form the nth generation in the ratio u_n **A**: 1 **a**. Let **a** be fully recessive as regards fitness, and let the fitness of **aa** in year n relative to **AA** and **Aa** be F_n. That is to say let P be **AA** or **Aa**, and Q be **aa**. Then (Haldane, 1924)

$$u_{n+1} = \frac{u_n\,(u_n+1)}{u_n+F_n}. \tag{1.2}$$

First consider the case when all values of u_n are large.

$$u_{r+1} - u_r = \frac{(1-F_r)u_r}{u_r+F_r}$$

$$= 1-F_r + \frac{F_r\,(F_r-1)}{u_r+F_r}$$

Hence, by summation

$$u_n-u_o = n - \sum_{r=0}^{n-1} F_r + \sum_{r=0}^{n-1} \frac{F_r(F_r-1)}{u_r+F_r}. \tag{1.3}$$

If all values of u_r are sufficiently large, then only the first two terms need be considered. For any finite value of n this can be assured by choosing a large enough value of u_o. Hence if $\Sigma F_r > n$, or, what is equivalent, if the mean value of F_r is greater than unity, u_n will be less than u_o provided u_o is sufficiently large.

*Mutation gives a simple criterion for "largeness" of a population. If a population is, say, ten times the reciprocal of the frequency with which a gene appears by mutation, the probability that no mutants will appear in it in any given generation is e^{-10}, or 5×10^{-5}. In a population of this size we may base our calculations on selection and mutation rates, neglecting the possibility of random extinction. The size will be of the order of a million for many human mutation rates.

Now let u_r be small. Then from (1.2)

$$\frac{u_{r+1}}{u_r} = \frac{1+u_r}{F_r+u_r}$$

$$\log u_{r+1} - \log u_r = -\log F_r - \log\left[1 + \frac{(1-F_r)\,u_r}{F_r\,(1+u_r)}\right]$$

So by summation

$$\log u_n - \log u_o = -\sum_{r=0}^{n-1} \log F_r - \sum_{r=0}^{n-1} \log\left[1 + \frac{(1-F_r)\,u_r}{F_r\,(1+u_r)}\right] \tag{1.4}$$

As before, if all values of u_r are sufficiently small, the second term is negligible unless some value of F_r is zero. If however a value of F_r is zero the corresponding term in the sum is $\log(u_r^{-1}+1)$. This is not infinite, but can be made as large as we choose by a suitable choice of u_o. We conclude that if u_o is sufficiently small, u_n will exceed u_o provided $\Sigma \log F_r < 0$, or $\prod_{r=0}^{n-1} F_r < 1$, or the geometric mean of F_r is less than unity.

It follows that, provided the arithmetic mean of the values of F_r is greater than unity, and the geometric mean less, u_n is bounded, that is to say, it must be between certain values. Neither of the genes **A** and **a** can disappear. If the values of F_r form a cycle, so that $F_{n+c} = F_n$, then the various values of u_n in a cycle converge towards fixed values. An example will be given later.

It may be asked whether the condition that the arithmetic mean should exceed unity, and the geometric mean be less, is not highly artificial. It is easy to give an example where it would be fulfilled. Suppose that the recessive **aa** is normally 5% to 10% fitter than the dominants, but that every twenty generations, on an average, an epidemic disease kills off all the recessives; then the arithmetic mean exceeds unity, but the geometric mean is zero. After each epidemic the frequency of recessive genes is half the frequency of heterozygotes in the previous generation, that is to say somewhat less than a half. It then increases until the next epidemic. It is not of course necessary to suppose that all the recessives are killed off. And the killing agent might be a drought or frost. It is sufficient that the recessive should be fitter than the dominant on the whole, but liable to be killed or sterilized in some abnormal but rare circumstances. It is hard to suppose that this condition is never fulfilled in nature.

If both the arithmetical and geometrical means of their fitness are less than unity, recessives will disappear in the course of time (except insofar as they are replaced by mutation). And if both these means are greater than unity, dominants will disappear. Since the arithmetic mean necessarily exceeds the geometric, unless the fitnesses are constant, there is no situation where either gene may disappear according to its initial frequency, as when heterozygotes are less fit than either homozygote.

Finally we may remark that the arithmetic mean of the relative fitnesses of recessives is the reciprocal of the harmonic mean of the relative fitnesses of dominants. For stability this harmonic mean must be less than unity. As the harmonic mean is a less familiar idea than the arithmetic, we have preferred to use the latter.

Bivoltine Organisms

Let us suppose that, like some insects, an organism has just two generations per year. In the spring generation the fitness of **aa** relative to **AA** and **Aa** is F, in the autumn generation f. Suppose the spring generation of year n to be formed from gametes with u_n **A** : 1 **a**, the autumn generation from gametes with u'_n **A** : 1 **a**. Then

$$u'_n = \frac{u_n(u_n+1)}{u_n+F},$$

$$u_{n+1} = \frac{u'_n(u'_n+1)}{u'_n+f}$$

$$= \frac{u_n(u_n+1)(u_n^2+2u_n+F)}{(u_n+F)[u_n^2+(1+f)u_n+Ff]}. \tag{2.1}$$

$$u_{n+1}-u_n = \frac{-u_n[(F+f-2)u_n^2-2(1-Ff)u_n-F(1-Ff)]}{(u_n+F)[u_n^2+(1+f)u_n+Ff]} \tag{2.2}$$

At equilibrium $u_n = U$, where $U=0$, $U=\infty$, or

$$(F+f-2)U^2-2(1-Ff)U-F(1-Ff)=0 \tag{2.3}$$

$$\text{whence } U = \frac{1-Ff \pm (1-F)(1-Ff)^{\frac{1}{2}}}{F+f-2}. \tag{2.4}$$

Since the roots are real, $Ff<1$. Since one must be positive, $F+f>2$. The figure shows the area within which the point (F, f) must lie to ensure an equilibrium with both genes present. Hence the right hand side of (2·2) is negative when u_n is large, positive when it is small. Hence the equilibrium is stable. (2·3) has only one positive root. If $F<1$ the ambiguity in (2·4) must have the negative sign. The equilibrium recessive gene frequencies in spring and autumn are:

$$Q=(U+1)^{-1}$$

$$= \frac{(1-f)-(1-Ff)^{\frac{1}{2}}}{f(F-1)} \tag{2.5}$$

$$Q' = \frac{F-1-(1-Ff)^{\frac{1}{2}}}{F(1-f)}.$$

Thus if $F=\frac{5}{9}$, $f=\frac{1}{2}$, $Q=\cdot2753$, $Q'=\cdot3101$, so the frequencies of recessives in spring and autumn are 7·58% and 9·62%.

The deviations from equilibrium fall off roughly in a geometric progression. Similar calculations could be made for a cycle of three or more generations.

Over-dominance

It is well known (Fisher, 1922) that if the relative fitnesses of **AA** : **Aa** : **aa** in each generation are F : 1 : f, being the same in each generation, there is a stable

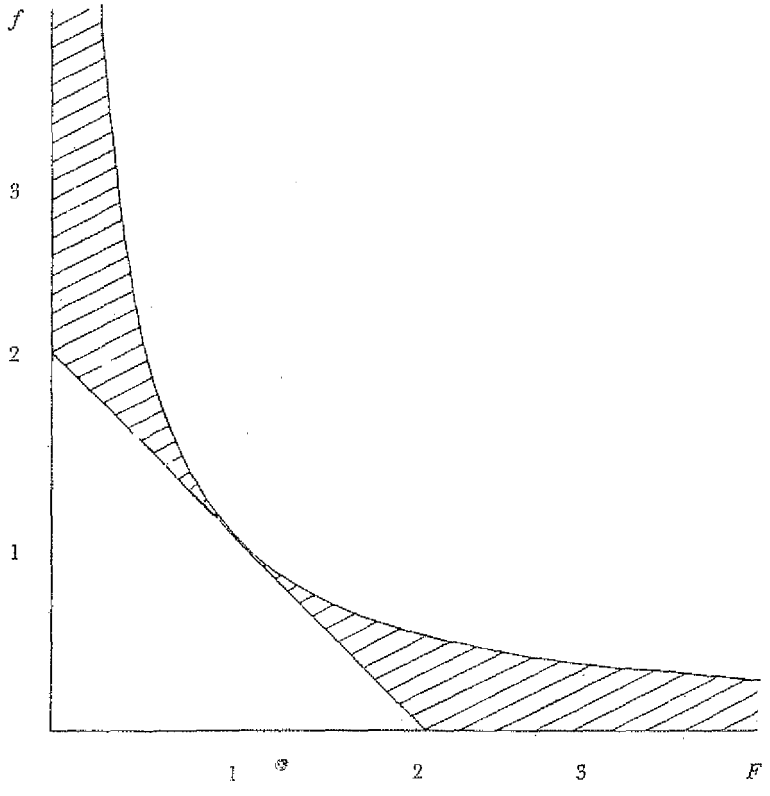

Fig. 1. A bivoltine population reaches a stable annual oscillation if the point (F, f) is in the shaded area.

equilibrium if and only if F and f are both less than unity. Let us now consider the case when F and f are variable.

$$u_{n+1} = \frac{F_n u_n^2 + u_n}{u_n + f_n} \tag{3.1}$$

$$\text{So } \frac{u_{n+1}}{u_n} = F_n + \frac{1 - F_n f_n}{u_n + f_n} \tag{3.2}$$

$$= f_n^{-1} + \frac{(F_n f_n - 1) u_n}{f_n (f_n + u_n)} \tag{3.3}$$

We see that we can choose u_0 so large that $u_{n+1} u_n^{-1}$ is as close as we wish to F_n. Hence $u_o > u_n$ if the geometric mean of F_r is less than unity. Similarly we can choose u_o so small that $\frac{u_{n+1}}{u_n}$ is as close as we wish to f_n^{-1}. Hence $u_n > u_o$ if the geometric mean of f_r is less than unity. Thus polymorphism will persist if the geometric mean fitnesses of each homozygote relative to the heterozygotes are each less than unity. Once again this may occur, even if the mean fitness of homozygotes relative to heterozygotes exceeds unity, provided the relative fitness of homozygotes occasionally falls very low.

It is not sufficient or necessary that the arithmetic mean fitness of the heterozygotes relative to homozygotes (that is to say the harmonic mean of F_r or f_r) should exceed unity.

For example an occasional severe epidemic of *falciparum* malaria might suffice to keep the gene for haemoglobin S from disappearance, even if for generations on end heterozygotes for sickling were at a disadvantage compared with persons homozygous for haemoglobin A.

Discussion

There is no great difficulty in extending the argument to a pair of sex-linked genes when dominance is complete in the homogametic sex. Here however a condition for equilibrium is that the relative fitness of heterogametic "mutants" should exceed unity, that of homogametic mutants being less than unity. It will be best discussed along with other cases where selection is in opposite directions in the two sexes.

We think that we have proved two novel points, first that a mere series of changes in the direction of selection may be enough to secure polymorphism, secondly that one result of Mendelian inheritance is that polymorphism may be permanent, if not exactly stable, when it is not so for a difference inherited clonally, cytoplasmically, or in a haploid. Finally we have shown that a statistical treatment of the intensity of selection can lead to definite results.

The results of this paper can, and doubtless will, be extended. Thus it is clear that inbreeding makes the conditions for polymorphism more stringent, and that the sums in (1·3) and (1·4) can be replaced by integrals when generations overlap.

Summary

In a large random mating population segregating for a pair of allelomorphs with full dominance, the condition that neither allelomorph should disappear is that the arithmetic mean of the fitnesses of recessives in different generations should exceed unity, their geometric mean fall below it.

REFERENCES

Haldane, J. B. S. (1924). A mathematical theory of natural and artificial selection, Part I. *Trans. Camb. Phil. Soc.*, **23**, 19-41.

Fisher, R. A. (1922). On the dominance ratio. *Proc. Roy. Soc. Edin.*, **42**, 321-341.