THE SELECTION OF DOUBLE HETEROZYGOTES

By J. B. S. HALDANE

Genetics and Biometry Unit, C.S.I.R.

The subject here treated is somewhat artificial. I shall first consider complete selection, and later partial selection in favour of double heterozygotes. By complete selection (Haldane 1924) I mean the complete elimination from the breeding population in each generation, of certain genotypes. This can occur in human breeding practice, by lethality of genes, and perhaps by exposure to insecticides. Nevertheless it is unlikely that the case here discussed ever occurs, except in Drosophila stocks, for example *Curly Dichaete*. Its interest is that, so far as I know for the first time, it reveals a situation where the course of events is qualitatively different in accordance with the intensity of linkage between the genes concerned.

Consider two pairs of allelomorphs **Aa** and **Bb** segregating in a large random mating population. Let their recombination frequencies in females and males be c and c' respectively. Let $h = \frac{1}{2}(c+c')$, the mean recombination frequency, and let $k = cc'$. If the genes are unlinked $c = c' = \frac{1}{2}$, so $h = \frac{1}{2}$ and $k = \frac{1}{4}$. In Drosophila species with negligible crossing over in males, if the genes are carried by the same chromosome $c' = 0$, so $h < \frac{1}{4}$, and $k = 0$. Further suppose that no genotypes except the double heterozygote **AaBb** survive to breed. Suppose that in the n-th generation the breeders consist of p_n**AB.ab** and q_n**Ab.aB**, where $p_n + q_n = 1$. Let $p_n - q_n = t_n$, so that $1 \geqslant t_n \geqslant -1$. An **AB.ab** female produces gametes in the frequencies:

$\frac{1}{2}[(1-c)\mathbf{AB} + c\mathbf{Ab} + c\mathbf{aB} + (1-c)\mathbf{ab}]$ and so on. It is easily calculated that

$$p_{n+1} = \frac{(1-c)(1-c')\,p_n^2 + (c+c'-2cc')\,p_n q_n + cc'q_n^2}{(1-c-c'+2cc')\,(p_n^2+q_n^2) + 2(c+c'-2cc')\,p_n q_n}$$

$$= \frac{(1-2h+k)p_n^2 + 2(h-k)\,p_n q_n + kq_n^2}{(1-2h+2k)\,(p_n^2+q_n^2) + 4(h-k)\,p_n q_n}$$

Hence $t_{n+1} = \dfrac{2(1-2h)t_n}{1+(1-4h+4k)t_n^2}$,

or $\triangle t_n = \dfrac{(1-4h)t_n - (1-4h+4k)t_n^3}{1+(1-4h+4k)t_n^2}$

There is an equilibrium when $t_n = 0$ or $\pm \left(\dfrac{1-4h}{1-4h+4k}\right)^{\frac{1}{2}}$.

The last two are however imaginary if $h > \frac{1}{4}$. Further the equilibrium at $t = 0$ (or $p = q = \frac{1}{2}$) is stable if $h \geqslant \frac{1}{4}$, and unstable if $h < \frac{1}{4}$. In this case the other two equilibria are stable. They move apart very rapidly when h, the mean frequency of recombination, falls below 25%, provided the frequencies in the two sexes are not very different. If $c' = c$, then $t = \pm(1-4c)^{\frac{1}{2}}(1-2c)^{-1}$, so when $c = \cdot24$, $t = \pm \cdot385$, so $p = \cdot31$ or $\cdot69$, and when $c = \cdot1$, $t = \cdot968$, so $p = \cdot016$ or $\cdot984$. That is to say when the mean crossover frequency falls appreciably below $\frac{1}{4}$, the stable frequencies of the two genotypes rapidly diverge from equality. In organisms where there is no recombination in one sex $c = 0$,

$k=0$, $h=\frac{1}{2}c$, so the only stable equilibria are with $t^2=1$, that is to say all individuals are **AB.ab** or all **Ab.aB**. This is in accordance with observation. A stock with, for example *Star* and *Curly* in opposite second chromosomes remains stable provided only *Star Curly* flies are selected as parents.

When $h>\frac{1}{4}$, t_n diminishes approximately with $(2-4h)^n$. More accurately—

$$(n+C)\ ln\ s=ln\ t_n-\frac{at_n^2}{1-s^2}\left[1-\frac{a(1+3s^2)\ t_n^2}{2(1-s^4)}+\frac{a^2(1+8s^2+5s^4+10s^6)t_n^4}{3(1-s^4)(1-s^6)}+\dots\right]$$

where $a=1-4h+4k$, and $s=2(1-2h)$. Or if $u=s^n+C$,

$$t_n=u+\frac{au^3}{1-s^2}+\frac{(2+s^2)a^2u^5}{(1-s^2)(1-s^4)}+\frac{(5+5s^2+4s^4+s^6)a^3u^7}{(1-s^2)(1-s^4)(1-s^6)}+\dots\dots$$

So except when h is only a little above $\frac{1}{4}$, this equilibrium is approached fairly rapidly. When $h=\frac{1}{2}$ it is reached in one generation. Similarly when $h<\frac{1}{4}$, t_n increases according to the same equation, nearly in a geometrical progression. If $T=\left(\dfrac{1-4h}{1-4h+4k}\right)^{\frac{1}{2}}$, and $t_n=T+x_n$, then

$$x_{n+1}=\frac{2hx_n-\frac{1}{2}[(1-4h)\ (1-4h+4k)]^{\frac{1}{2}}x_n^2}{1-2h+[(1-4h)\ (1-4h+4k)]^{\frac{1}{2}}x_n+\frac{1}{2}(1-4h+4k)x^2}$$

That is to say x_n tends to zero approximately with $\left(\dfrac{2h}{1-2h}\right)^n$.

For example if $c=c'=\cdot20$, it tends to zero with $(\frac{2}{3})^n$, if $c=c'=\cdot10$, with $(\frac{1}{4})^n$. If $c'=0$, as in Drosophila, t_n tends to ±1. $2h=c$, $k=0$, so if $t_n=x_n-1$,

$$x_{n+1}=\frac{cx_n-\frac{1}{2}(1-2c)x_n^2}{1-c+(1-2c)\ (x_n+\frac{1}{2}x_n^2)}.$$

So x_n tends to zero with $\left(\dfrac{c}{1-c}\right)^n$, for example with 3^{-n} when $c=\cdot25$.

The population considered is subject to two "forces". On the one hand segregation, if there were no selection, would force it to a composition where the two types of double heterozygote and the four types of gamete were present in equal numbers, that is to say $p=q=\frac{1}{2}$, and $t=0$. On the other hand, if the fitness of all genotypes except **AB.ab** and **Ab.aB** is zero, the mean fitness, or adaptive value, of the population, is $\frac{1}{4}[1+(1-4h+4k)t^2]$. Provided there is any linkage at all, the coefficient of t^2 is positive. Selection tends to increase the mean adaptive value, and hence to increase t^2. If it were not opposed by segregation, t^2 would become unity, that is to say all the breeding members would be **AB.ab** or all **Ab.aB**. If so the mean fitness would be raised to $\frac{1}{2}(1-2h+2k)$, or nearly $\frac{1}{2}$, if linkage was strong. So long as the mean recombination value exceeds 25%, segregation will overcome selection, and t will remain at zero. However, if the mean recombination value is under 25%, selection should have some effect, and a stable equilibrium should be reached with unequal numbers of the two genotypes, unless $c'=0$, when only one will remain.

Unfortunately, owing to the absence of crossing-over in males, Drosophila species are not suited for verifying this theory. Its interesting point is the existence of two stable equilibria which would most readily be verified if **A** and **B** had about 20% recombination. In this case there would be 13% of one type of double heterozygote,

and 87% of the other, at a stable equilibrium. It is unlikely that exactly this situation is at all frequent either in nature or in agricultural practice, though something like it must occur in the evolution of permanently heterozygous species. However, we know so little about the selective value of linkage that any indications concerning it are of value. In the particular case here considered, we note the following conclusions. In a species with no crossing over in one sex, complete selection in favour of the genotype **AaBb** will necessarily give a population consisting entirely of **AB.ab**, or entirely of **Ab.aB**. In such a population $\frac{1}{2}(1+c)$ of the zygotes are lethal. There is some selective value in a reduction of c. In organisms with crossing over in both sexes, if the mean recombination frequency, $h=\frac{1}{2}(c+c')$ exceeds 25%, the coupling and repulsion genotypes will be about equally frequent, and $\frac{3}{4}$ of all zygotes will perish without breeding. There will be no advantage in reducing the recombination frequency of **A** and **B**, say from 45% to 28%. However as soon as h falls below 25%, the fraction killed off is $\frac{1}{2}(1+2h)$, and selection will presumably tend to reduce h. Inversions may be of selective value, even if they cause some sterility.

Innumerable possibilities exist of incomplete selection for double heterozygotes. I shall only consider one in which the fitness of all genotypes except the double heterozygotes is equal, that of the double heterozygotes being $(1+K)$ times greater. Clearly at equilibrium in such a population the frequencies of each of the four genes will be $\frac{1}{2}$.

Let the n-th generation be formed from female gametes:

$$\tfrac{1}{2}(x_n\mathbf{AB}+y_n\mathbf{Ab}+y_n\mathbf{aB}+x_n\mathbf{ab}).$$

where $x_n+y_n=1$,

and male gametes: $\frac{1}{2}(x_n'\mathbf{AB}+y_n'\mathbf{Ab}+y_n'\mathbf{aB}+x_n'\mathbf{ab})$.

Then in the next generation the frequencies, at fertilization, of **AB.ab** and **Ab.aB** are $\frac{1}{2}x_nx_n'$ and $\frac{1}{2}y_ny_n'$. These have an extra fertility measured by K, so we find

$$x_{n+1} = \frac{1+(1+K)\ (1-c)x_nx_n'-(1-c-cK)y_ny_n'}{2+K(x_nx_n'+y_ny_n')}$$

$$y_{n+1} = \frac{1-(1-c-cK)x_nx_n'+(1+K)\ (1-c)y_ny_n'}{2+K(x_nx_n'+y_ny_n')}$$

with a similar pair of equations for x_{n+1}' and y_{n+1}'. Now put $x_n-y_n=u_n$, $x_n'-y_n'=u_n'$. Then

$$u_{n+1} = \frac{(2+K-2c-2cK)\ (u_n+u_n')}{4+K+Ku_nu_n'}$$

$$u_n'_{+1} = \frac{(2+K-2c'-2c'K)\ (u_n+u_n')}{4+K+Ku_nu_n'}.$$

Let $u_n = (2+K-2c-2cK)s_n$, $u_n'=(2+K-2c'-2c'K)s_n$, and let $h=\frac{1}{2}(c+c')$, $k=cc'$, as before. Then

$$s_{n+1} = \frac{2[2+K-2h(1+K)]\ s_n}{4+K+K[(2+K)^2-4h(1+K)(2+K)+4k(1+K)^2]s_n{}^2}$$

The equilibria are given by $s=0$, or

$$s^2 = \frac{K-4h-4hK}{K[(2+K)^2-4h(1+K)\ (2+K)+4k(1+K)^2]}$$

So $s=0$, or $x_n=y_n=x_n'=y_n'=1$, is a stable equilibrium if

$$h > \frac{K}{4(1+K)}, \quad \text{or} \quad K \leqslant \frac{4h}{1-4h}.$$

If h is less than this value, the other two equilibria are stable. The frequency of **AB.ab** at equilibrium is then:

$$\tfrac{1}{4}[(1-2h-2hK^{-1}\pm(2-2h+K-2hK)s],$$

where s is the equilibrium value of s_n. The advantage of double heterozygosity must therefore be fairly large, or the linkage fairly tight, before selection becomes operative. It would be possible to verify these conclusions in Drosophila. Thus if double heterozygotes were given twice the chance of other genotypes to leave offspring both types of double recessive should persist if the recombination frequency in females exceeds 25%. Similar calculations could be made in more complicated cases, for example where single heterozygotes have fitnesses intermediate between those of homozygotes and of double heterozygotes. It is clear that the greater the advantage enjoyed by double heterozygotes, the higher will be the mean recombination frequency at which the equilibrium with coupling and repulsion equally frequent becomes unstable. But this frequency can never exceed 25%.

SUMMARY

Selection in favour of double heterozygotes leads to different types of equilibrium according to the linkage between the loci concerned. For slight selection or loose linkage coupling and repulsion should be equally frequent. When the selection or the linkage becomes sufficiently intense there is an equilibrium when almost all or almost none of the double heterozygotes are in coupling.

REFERENCE

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