Theoretical and Applied Genetics 38, 140–143 (1968)

A Genetic Model Having Complex Linkage Behaviour

W. J. Ewens

La Trobe University, Bundoora, Vic.

Summary. A genetic model is discussed in which the position and nature of equilibrium points for gamete frequencies depends in an unusual way on the degree of linkage between the loci involved. A complete mathematical analysis is made of the model; this is followed by a verbal discussion of the effect of linkage on such models.

1. Mathematical Treatment

The quantitative investigation of the effect of linkage between two loci on the evolutionary behaviour of a population was initiated in a beautiful paper by Kimura (1956). In this paper Kimura showed that, in the genetical model introduced, the equilibrium point for the set of gamete frequencies was stable if and only if the linkage between the two loci upon which fitness was supposed to depend was sufficiently tight.

In a general discussion of two locus systems, Lewontin and Kojima (1960) also investigated the linkage behaviour of a model whose fitness matrix was of the form

$$\begin{array}{cccc} & AA & Aa & aa \\ BB & 1-a & 1-b & 1-a \\ Bb & 1-c & 1 & 1-c \\ bb & 1-a & 1-b & 1-a \end{array} \tag{1}$$

Here a, b and c are assumed to be positive quantities; note that the notation used here differs from that of Lewontin and Kojima.

Now the analysis of the behaviour of random mating populations, when fitness depends on two loci, must be carried out by considering the frequencies of the four possible gametes allowed at the two loci. Let us denote the frequencies of the gametes AB, aB, Ab and ab by x_1, x_2, x_3 and x_4 respectively. Symmetry in (1) ensures that we may take $x_1 = x_4 = x$, $x_2 = x_3 = 1/2 - x$; we do this from now on and carry out all the subsequent analysis in terms of x only.

Our first problem is to locate the equilibrium points. This has been done for the model (1) by Lewontin and Kojima and we shall not pursue the details. Symmetry in (1) ensures that

$$x = \frac{1}{4} \tag{2}$$

is always an equilibrium point; Lewontin and Kojima show that the remaining two equilibrium values are

$$x = \frac{1}{4} \pm \frac{1}{4}\sqrt{1 - \frac{4R}{b + c - a}}, \tag{3}$$

where R is the recombination fraction between "A" and "B" loci. Naturally, the equilibrium points (3) exist only when

$$R < \frac{1}{4}(b + c - a) . \tag{4}$$

By symmetry, we need examine in detail only one of the solutions (3); we shall therefore consider only that solution corresponding to the positive radical and we denote the corresponding value by x_0.

Having found any equilibrium point, we must decide whether or not it is stable, since unstable equilibrium points are of little interest. Lewontin and Kojima found that the equilibrium point (2) is stable if and only if

$$\text{(i)}\quad R > \frac{1}{4}(b + c - a) \quad \text{and} \quad \text{(ii)}\ a > |b - c| . \tag{5}$$

The condition that the equilibrium points (3) be stable is more difficult to find, and indeed this condition was not obtained by Lewontin and Kojima. The purpose of the present paper is to derive the condition that x_0 be a point of stable equilibrium, to consider this condition in some detail in the case $b = c$, and to discuss any general principles which emerge from the mathematical analysis.

We note first of all that since

$$x_0 = \frac{1}{4} + \frac{1}{4}\sqrt{1 - \frac{4R}{b + c - a}}, \tag{6}$$

we have

$$x_0\left(\frac{1}{2} - x_0\right) = R/4\,(b + c - a) , \tag{7}$$

a relation which can be used frequently in simplifying various expressions. Thus for example the mean fitness W of the population, given by

$$W = 1 - 2a\left\{x^2 + \left(\frac{1}{2} - x\right)^2\right\} - 4(b + c)\,x\left(\frac{1}{2} - x\right),$$

is given by

$$W_0 = 1 - \frac{1}{2}a - R \qquad \text{at} \qquad x = x_0 . \tag{8}$$

To find whether or not the equilibrium point (6) is stable, we put

$$x_1 = x_0 + \delta_1, \quad x_2 = \frac{1}{2} - x_0 + \delta_2, \quad x_3 = \frac{1}{2} - x_0 + \delta_3,$$

where the δ_i are small deviations of the gamete frequencies from their equilibrium values. Using the recurrence relations (10a) – (10d) in Lewontin and Kojima (1960) we may find the frequencies x_i' of the gametes in the following generation, and hence the deviations δ_i' of these from the equilibrium values. Ignoring terms of order δ_i^2 we then find

$$\delta' = M\,\delta , \tag{9}$$

where M is a matrix of constants. By iterating in (9) we see that the condition that the vector δ approach zero as time goes on is that all the eigenvalues of M are less than unity in absolute value.

For the model (1) we find

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix},$$

where

$$m_{11} = W_0^{-1}\left[2\,(1-a)\,x_0 + (2-b-c)\left(\frac{1}{2}-x_0\right)\right],$$

$$m_{12} = W_0^{-1}\left[\frac{1}{2}R - b\,x_0 + x_0\,(1-4\,x_0)\,(a-b-c)\right],$$

$$m_{13} = W_0^{-1}\left[\frac{1}{2}R - c\,x_0 + x_0\,(1-4\,x_0)\,(a-b-c)\right],$$

$$m_{21} = W_0^{-1}\left[(c-b)\left(\frac{1}{2}-x_0\right)\right],$$

$$m_{31} = W_0^{-1}\left[(b-c)\left(\frac{1}{2}-x_0\right)\right],$$

$$m_{22} = W_0^{-1}\Bigg[\frac{1}{2}(1-R) + \frac{1}{2}(1+c-2\,a) + {} + x_0\,(2\,a-b-2\,c) + \left(\frac{1}{2}-x_0\right) \times {} \times (1-4\,x_0)\,(a-b-c)\Bigg],$$

$$m_{32} = W_0^{-1}\Bigg[\frac{1}{2}(1-R) - \frac{1}{2}(1-b) - b\,x_0 + {} + \left(\frac{1}{2}-x_0\right)(1-4\,x_0)\,(a-b-c)\Bigg],$$

$$m_{33} = W_0^{-1}\Bigg[\frac{1}{2}(1-R) + \frac{1}{2}(1+b-2\,a) + {} + x_0\,(2\,a-2\,b-c) + \left(\frac{1}{2}-x_0\right) \times {} \times (1-4\,x)\,(a-b-c)\Bigg].$$

Elementary operations show that with the m_{ij} defined as above, the eigenvalues of M are

$$\lambda_1 = W_0^{-1}\left[1 + R - \frac{1}{2}(b+c)\right], \tag{10}$$

$$\lambda_2 = W_0^{-1}\Bigg[1 - \frac{1}{2}a - \frac{1}{4}(b+c) + {} + \frac{1}{2}\sqrt{\left\{a-\frac{1}{2}(b+c)\right\}^2 - \frac{4R(a^2+bc-ab-ac)}{b+c-a}}\Bigg] \tag{11}$$

$$\lambda_3 = W_0^{-1}\Bigg[1 - \frac{1}{2}a - \frac{1}{4}(b+c) - {} - \frac{1}{2}\sqrt{\left\{a-\frac{1}{2}(b+c)\right\}^2 - \frac{4R(a^2+bc-ab-ac)}{b+c-a}}\Bigg] \tag{12}$$

We note first of all that λ_2 and λ_3 are both real for R in the range

$$0 < R < \frac{1}{2}(b+c-a), \tag{13}$$

which is the range within which x_0 is defined. Indeed it is not difficult to see that each λ_i is positive, so that stability behaviour will depend on whether each λ_i is less than unity.

It is easy enough to show that holds for λ_1, for the condition $\lambda_1 < 1$ is

$$R < \frac{1}{4}(b+c-a),$$

which is once more the conditon (13) that x_0 exists. Further, it is clear that $\lambda_3 < \lambda_2$, so that need concentrate attention on λ_2 only.

The condition $\lambda_2 < 1$ is clearly

$$\left[\left\{a-\frac{1}{2}(b+c)\right\}^2 - \frac{4\,R\,(a^2+bc-ab-ac)}{b+c-a}\right]^{1/2} < {} < \frac{1}{2}(b+c) - 2\,R. \tag{14}$$

At $R = 0$, this condition is

$$\left|a-\frac{1}{2}(b+c)\right| < \frac{1}{2}(b+c). \tag{15}$$

However, using the conditions $0 < a < b + c$, it is very easy to see that (15) always holds. Thus extremely tight linkage will always provide a point of stable equilibrium.

At the other extreme, the maximum possible value of R is

$$R = \frac{1}{4}(b+c-a).$$

At this value, the condition (14) becomes

$$|b-c| < a. \tag{16}$$

It is quite clear that if (16) does not hold, then only tight values of linkage lead to a stable equilibrium. On the other hand, if (16) holds, then both very tight and the loosest allowable linkage both lead to stable equilibrium. Surprisingly it does not follow that, when this is so, all values of R in the range (13) lead to a stable equilibrium. Thus even if the inequality (16) holds, there may well be values of R in the range (13) for which the inequality (14) does not hold. This will be the case if we can find solutions of the equation

$$\left\{a-\frac{1}{2}(b+c)\right\}^2 - \frac{4\,R\,(a^2+bc-ab-ac)}{b+c-a} = {} = \left\{\frac{1}{2}(b+c) - 2\,R\right\}^2 \tag{17}$$

in the range (13). Now the solutions of (17) are

$$R = \frac{1}{2}\Bigg[\frac{b+c}{2}\,\frac{a^2+bc-zb-ac}{b+c-a} \pm {} \pm \sqrt{\left\{\frac{a^2+bc-sb-ac}{b+c-a} - \frac{b+c}{2}\right\}^2 - a\,(b+c-a)}\Bigg]. \tag{18}$$

While it is possible to write down, from (17) and (18), the conditions that the two solutions lie in the range (13), it is probably easiest, for any particular case, merely to compute the values (18) and test whether or not this is so. That such a contingency may certainly happen is illustrated by the case

$$a = .1\,, \quad c = .78\,, \quad b = .82\,. \tag{19}$$

Here condition (16) holds so that both very tight and the loosest allowable linkage produce stable equilibria. However for the values (19), the solutions of (17) are $R = .1005$, $R = .3731$, so that for values of R in the range (.1005, .3731) no stable equilibrium is possible.

2. The case $b = c$

We can go considerably further in the analysis developed above in the particular case $b = c$. Here condition (16) is automatically satisfied, while the solutions (18) reduce to

$$R = \frac{1}{2}\left[b - \frac{(b-a)^2}{2b-a} \pm \frac{(b-a)}{2b-a}\sqrt{(b-2a)^2 - 2a^2}\right]. \qquad (20)$$

From the form of this solution, it is not difficult to see that the two values (20) lead to real values of R within the range $\left(0, \frac{1}{4}(2b-a)\right)$ only when

$$a/b < 1 - (1/\sqrt{2}). \qquad (21)$$

That is to say, when condition (21) holds, we find that the equilibrium point (8) is stable only when $0 < R < R_1$ or $R_2 < R < \frac{1}{4}(2b-a)$, where $R_1 < R_2$. Whenever equation (21) does not hold, equation (17) has no real solutions in $0 < R < \frac{1}{4}(2b-a)$, so that the inequality (14) holds for all values of R in this range. In other words, all values of R in the range lead to a stable equilibrium.

We shall consider in the next section a numerical example for the case $b = c$, which will verify the above theory; in a subsequent section we shall discuss generally the general principles indicated by the mathematical analysis.

3. Numerical Example

In a more recent publication, Lewontin (1964) has carried out a numerical investigation of the behaviour of the model (1) in the case

$$a = .1\,, \quad b = .8\,. \qquad (22)$$

Lewontin's numerical results are briefly as follows. When $R < .10$ there is a stable equilibrium of the form (6); for $.10 < R < .375$ there is no stable equilibrium of any kind, while for $.375 < R < .5$ the equilibrium (2) is stable.

We now check these numerical results against the theory given above. Since

$$a/b < 1 - (1/\sqrt{2})\,,$$

our analysis asserts that there are two distinct regions where the equilibrium point (8) is stable, namely $0 < R < R_1$ and $R < R_2 < \frac{1}{4}(2b-a)$, where R_1 and R_2 are given by (18). For the numerical values (22) we find that these regions are

$$0 < R < .10061 \qquad (29)$$

and

$$.3727 < R < .3750\,. \qquad (30)$$

Note that the former interval (29) agrees with that found numerically by Lewontin; on the other hand, Lewontin's numerical method does not pick up the (admittedly very small) interval (30).

Finally, our theory asserts that when $R > .375$, then (2) provides a stable equilibrium point. This also agrees exactly with Lewontin's numerical results; indeed the theoretical prediction had already been obtained, using the model (1), by Lewontin and Kojima (1960).

It is of some interest to give a numerical example of a case where (16) does not hold. Suppose $b = .7$, $c = .4$, $a = .1$. Since with these values equation (16) does not hold, our theory indicates that the equilibrium point (6), namely

$$x = \frac{1}{4} + \frac{1}{4}\sqrt{1 - 4R}\,, \qquad (31)$$

is stable only for R sufficiently small. Now the equilibrium point (31) exists when $0 < R < .25$, but equation (14) asserts that this is stable only when

$$[.2025 - .72\,R]^{1/2} < .55 - 2\,R\,,$$

that is to say

$$0 < R < .09\,. \qquad (32)$$

The nature of the region (32) thus verifies our theoretical prediction.

4. Discussion

It will be quite clear that the effect of linkage, in the model considered, on the position and nature of the equilibrium points is rather complex; it is therefore of some interest to try to derive principles from it which appear to have general relevance.

The first result was that for a model such as (1), the equilibrium (2) is stable only if R is sufficiently large, say exceeding some value R_0. It may happen that R_0 is negative (as would happen in our model if $a > b + c$, in which case (2) is stable for all R); when R_0 is positive the condition $R_0 > R$ is a real restriction.

It is quite clear why this behaviour occurs; we illustrated by considering a population for which initially $x_1 = x_4 = \frac{1}{2}$, $x_2 = x_3 = 0$. Clearly there are initially only *AABB*, *AaBb* (in coupling phase) and *aabb* individuals. Suppose $a > b + c$; this means that *AABB* and *aabb* individuals have considerably smaller fitness than single heterozygotes. Our theory asserts that (2) is stable for all R, and clearly the explanation is that because of the small fitness of *AABB* and *aabb* with respect to that of single heterozygotes even a small amount of crossing over between "*A*" and "*B*" loci will result in a steady increase in frequency of the single heterozygotes. After sufficient time the equilibrium (2) becomes established.

However if $a < b + c$, the selective disadvantage of *AABB* and *aabb* is not so great, and starting from an initial population of $x_1 = x_4 = \frac{1}{2}$, the frequencies of single heterozygotes will increase only if linkage is sufficiently loose; we have seen the required condition is $R > \frac{1}{4}(a + b - c)$. If this is so, the stable equilibrium (2) is eventually achieved.

On the other hand, we have shown that if $R < \frac{1}{4}(a + b - c)$, it is possible to establish a stable equilibrium given by (6). This however, does not happen automatically. It will certainly occur for extremely tight linkage, when nearly all gametes are *AB* or *ab*. In this case the stability behaviour is analogous to that at a single locus "*C*", where fitness of *CC*, *Cc* and *cc* are $1 - a$, 1, $1 - a$ respectively. However we have shown that the precise nature of the region in which (6) is stable can be rather complex.

Firstly, take the case where (16) does not hold. Here the region for stability of (6) is of the form $0 < R < R_1$. It is not difficult to see why the region is of this form and we exemplify this by supposing $x_1 \approx 1$. Now *AaBb* is superior in fitness to *AABB*, so that with tight linkage its frequency can increase and eventually the equilibrium (6) is achieved. However if linkage is not tight, then the imbalance of fitness between *AaBB* and *AABb* means that *aB* and *Ab* gametes are produced in sufficiently unequal numbers to upset the rate of increase of *AaBb* relative to the rate of increase of other zygote. What then happens is that the frequency of one of *AABb* or *AaBB* increase to the detriment of the other and also of *AaBb*. But the superior fitness of *AABB* means that eventually this frequency decreases and we find $x_1 \to 1$.

Thus we see that (16) introduces an important concept of "balance" in fitness of single heterozygotes. It is as though the increase of *AaBb* occurs in a "walking on a tightrope" manner; if linkage is sufficiently loose and the "pull" of *AaBB* to one side does not sufficiently balance that of *AABb* to the other, stability will be lost. Equation (16) shows numerically how strong the "imbalance" can be before stability is lost for loose linkage.

Our further discussion will refer to the case $b = c$. Here we have found that the region for stability of (6) is of the form $\left[0, \frac{1}{4}(b + c - a)\right]$ if $a/b > 1 - (1/\sqrt{2})$, and of the form $\left[(0, R_1), \left(R_2, \frac{1}{4}(b + c - a)\right)\right]$ if $a/b < 1 - (1/\sqrt{2})$. To discuss the reason for this, we concentrate attention on the latter contingency, where a gap $R_1 < R < R_2$ exists in values of R leading to stability of (6).

We illustrate in the case $a = .1$, $b = .8$, $x_1 \approx 1$. Initially nearly all individuals are *AABB*, *AaBB*, *AABb* or *AaBb* (coupling phase). If linkage is extremely tight ($0 < R < .10061$ in our case) the frequency of double heterozygotes (which have the highest fitness of all zygotes) will increase due to an increase in frequency of the gamete *ab*. However if linkage is moderate ($.10061 < R < .3727$), the double heterozygotes will produce a non-negligible number of single heterozygote offspring; the selective disadvantage of the latter ensures the gradual loss of the genes *a* and *b* and we find $x_1 \to 1$. However for $.3727 < .375$, linkage is sufficiently loose to ensure a sufficiently large number of balanced homozygotes *AAbb* and *aaBB* to be formed; the comparatively high fitness of the latter, relative to that of single heterozygotes, ensures that the equilibrium (6) is eventually reached. In this case this equilibrium is quite similar to the equilibrium (2) and may thought of as being a "continuation" of the latter. The above arguments explain the seemingly strange behaviour of both tight and loose linkage leading to stability, with moderate linkage leading to instability.

Zusammenfassung

Es wird ein genetisches Modell behandelt, in dem die Lage und Art der Gleichgewichtspunkte für die Gametenfrequenz in ungewöhnlicher Weise von dem Grad der Kopplung zwischen den beiden in Frage kommenden Loci abhängen. Für das Modell wird eine vollständige mathematische Analyse vorgelegt und anschließend die Wirkung besprochen, welche die Kopplung in derartigen Modellen hat.

References

1. Kimura, M.: A model of a genetic system which leads to closer linkage by natural selection. Evolution **10**, 278–287 (1956). — 2. Lewontin, R. C.: The interaction of selection and linkage. 1. General considerations: heterotic models. Genetics **49**, 49–67 (1964). — 3. Lewontin, R. C., and K. Kojima: The evolutionary dynamics of complete polymorphisms. Evolution **14**, 458–472 (1960).