

## A CONTINUOUS SELECTIVE MODEL FOR AN X-LINKED LOCUS\*

THOMAS NAGYLAKI

*Department of Medical Genetics, University of Wisconsin, Madison, Wisconsin 53706*

Received 30.vii.74

### SUMMARY

Neglecting age-structure, but taking into account matings with differential fertility in Mendelian reproduction, a continuous selective model is formulated for a single X-linked locus with an arbitrary number of alleles. Without restricting the mating system, differential equations are derived for the genotypic and allelic frequencies. Assuming random mating, no selection, and constant fertilities and mortalities, these differential equations are solved explicitly. For this case, in contrast to the corresponding phenomenon in the usual model with discrete, non-overlapping generations, the difference between the frequencies of any allele in males and females approaches zero without oscillation.

### 1. INTRODUCTION

AGE-STRUCTURED models with differential fertility and mortality, required for a full and accurate analysis of most natural populations, are mathematically formidable. The mathematics of the two somewhat simplified continuous-time age-structured models introduced by Charlesworth (1970), one of which takes into account differences in fecundity, but neglects mating, while the other does the reverse, is still quite difficult. Consequently, most research in population genetics has disregarded the complication of age-structure, and treated either a population with discrete, non-overlapping generations or one with continuous random births and deaths.

Population genetic models with discrete, non-overlapping generations rest on clear biological assumptions, and apply manifestly to certain types of organisms, such as annual plants. Therefore, Wright and many others have favoured them in their investigations. Except near equilibrium, however, the difference equations resulting from this formulation do not lend themselves readily to dynamical analysis. The Malthusian-parameter approach of Fisher (1930), since it yields differential equations, admits of more complete and elegant mathematical treatment, as exemplified by his development of the Fundamental Theorem of Natural Selection. But neither Fisher's nor more recent (see, *e.g.*, Crow and Kimura, 1970, pp. 190-197) derivations of the differential equations satisfied by gametic frequencies considered explicitly the mating structure of the population or fertility differences between matings.

In a previous paper (Nagylaki and Crow, 1974, hereafter referred to as I), ignoring age-structure, but permitting matings to have differences in fecundity, we constructed continuous selective models for a single autosomal locus with multiple alleles, with or without distinguishing the sexes,

\* Paper Number 1750 from the Genetics Laboratory, University of Wisconsin. Supported by the National Institutes of Health (Grant GM-15422).

and for two alleles at each of two loci in monoecious individuals. Without restricting the mating system, in each case we derived differential equations for the genotypic frequencies, and established the validity of the customary Malthusian-parameter differential equations satisfied by the gametic frequencies. We devoted particular attention to the conditions for Hardy-Weinberg proportions under random mating. Since it was shown that Hardy-Weinberg proportions will occur only in special cases, generally, the evolution of the population in these models is specified by the usual differential equations for the gametic frequencies only when these differential equations are supplemented by the ones we deduced for the genotypic frequencies. In I, the general theory was also applied to two particular assortative mating patterns, and in a subsequent paper (Nagylaki, 1974) it was extended to incorporate mutation and migration.

Our aim now is to formulate a model without age-structure, but with arbitrary mating system and differential fertility, for a single X-linked locus with multiple alleles. The general differential equations for the genotypic and allelic frequencies will be solved for the particular case of random mating, no selection, and constant birth and death rates. We shall devote special attention to the approach to equality of the gene frequencies in males and females and to the approach to Hardy-Weinberg ratios in females. An experiment to determine whether natural populations are more likely better approximated by discrete or continuous models will be suggested.

## 2. THEORY

As far as possible, we shall follow the notation of I, employing unprimed quantities to refer to males and primed ones to females. Let  $N$ ,  $N'$  be the number of males and females.  $A_i$  will denote the alleles at the X-linked locus of interest. Let  $n_i$ ,  $2n'_i$ ,  $n'_{ij}$  ( $= n'_{ji}$ ) represent the number of  $A_i$  alleles in males, in females, and of ordered  $A_iA_j$  females, respectively. Henceforth, all genotype and mating designations will be ordered for convenience. The frequencies of  $A_i$  in males and females and of  $A_iA_j$  females are

$$p_i = n_i/N, \quad p'_i = n'_i/N', \quad P'_{ij} = n'_{ij}/N'. \quad (1)$$

The time-derivatives of  $n_i$  and  $n'_{ij}$  may be written as

$$\dot{n}_i = M \sum_{kl} X_{k,il} \tilde{a}_{k,il} - d_i N p_i, \quad (2)$$

$$\dot{n}'_{ij} = (1/2)M \sum_k (X_{i,jk} \tilde{a}'_{i,jk} + X_{j,ik} \tilde{a}'_{j,ik}) - d'_{ij} N' P'_{ij}, \quad (2')$$

where  $M$ ,  $X_{i,jk}$ ,  $\tilde{a}_{i,jk}$ ,  $\tilde{a}'_{i,jk}$ ,  $d_i$ ,  $d'_{ij}$  are the number of matings per generation, the proportion of these which is between  $A_i$  males and  $A_jA_k$  females, the number of male and female progeny from one such union, and the probability per unit time (generation) of the death of an  $A_i$  male and an  $A_iA_j$  female. Except for the neglect of age structure, all of these quantities may be functions of time, genotypic frequencies, and any other relevant variables and parameters. If  $i \neq j$ ,  $n'_{ij}$  is simply one-half the number of females with unordered genotype  $A_iA_j$ ; we do *not* employ the convention (as (2') shows) that  $A_i$  comes from the male and  $A_j$  from the female parent. With our

definition, we shall have, initially,  $n'_{ij}(0) = n'_{ji}(0)$ . Then (2') implies that this symmetry is preserved for all later times:  $n'_{ij}(t) = n'_{ji}(t)$ .

Instead of the birth rates of the various types of unions per mating, defined above, it will be more convenient to use birth rates per male and female:

$$a_{i,jk} = M\tilde{a}_{i,jk}/N, \quad a'_{i,jk} = M\tilde{a}'_{i,jk}/N'. \tag{3}$$

With (3), (2) and (2') become

$$\dot{n}_i = N \sum_{kl} X_{k,il} a_{k,il} - d_i N p_i, \tag{4}$$

$$\dot{n}'_{ij} = (1/2)N' \sum_k (X_{i,jk} a'_{i,jk} + X_{j,ik} a'_{j,ik}) - d'_{ij} N' P'_{ij}. \tag{4'}$$

The mean mortalities for males and females read

$$\bar{d} = \sum_i p_i d_i, \tag{5}$$

$$\bar{d}' = \sum_{ij} P'_{ij} d'_{ij}. \tag{5'}$$

We define the average rates at which  $A_i A_j$  females give birth to males,  $A_i$  males to females,  $A_i A_j$  females to females by

$$P'_{ij} c_{ij} = \sum_k X_{k,ij} a_{k,ij}, \tag{6a}$$

$$p_i b'_i = \sum_{kl} X_{i,kl} a'_{i,kl}, \tag{6b}$$

$$P'_{ij} c'_{ij} = \sum_k X_{k,ij} a'_{k,ij}. \tag{6c}$$

The mean fertilities show at once that, as expected, males and females contribute equally to females:

$$\bar{c} = \sum_{ij} P'_{ij} c_{ij}, \tag{7a}$$

$$b'_i = \sum_i p_i b'_i, \tag{7b}$$

$$\bar{c}' = \sum_{ij} P'_{ij} c'_{ij} = b'. \tag{7c}$$

Observing that

$$N = \sum_i n_i, \quad N' = \sum_{ij} n'_{ij}, \tag{8}$$

Equations (4) to (7) yield

$$\dot{N} = \bar{m}N, \quad \dot{N}' = \bar{m}'N', \tag{9}$$

where the mean fitnesses for males and females are

$$\bar{m} = \bar{c} - \bar{d}, \quad \bar{m}' = \bar{c}' - \bar{d}'. \tag{10}$$

From (1), (4), (4'), and (10) we obtain the differential equations for the genotypic frequencies:

$$\dot{p}_i = \sum_{kl} X_{k,il} a_{k,il} - (d_i + \bar{m})p_i \tag{11}$$

$$\dot{P}'_{ij} = (1/2) \sum_k (X_{i,jk} a'_{i,jk} + X_{j,ik} a'_{j,ik}) - (d'_{ij} + \bar{m}')P'_{ij}. \tag{11'}$$

Given the fecundities, mortalities, and the mating frequencies as functions of the genotypic proportions, (11) and (11') completely specify the genotypic frequencies. Summing (11') over  $j'$ , we deduce the differential equations for the gene frequencies

$$\dot{p}_i = p'_i c_i - p_i (d_i + \bar{m}), \quad (12)$$

$$\dot{p}'_i = (1/2)(p_i b'_i + p'_i c'_i) - p'_i (d'_i + \bar{m}'), \quad (12')$$

where the allelic mortalities and fertilities are given by

$$p'_i d'_i = \sum_j P'_{ij} d'_{ij}, \quad (13a)$$

$$p'_i c_i = \sum_j P'_{ij} c_{ij}, \quad (13b)$$

$$p'_i c'_i = \sum_j P'_{ij} c'_{ij}. \quad (13c)$$

Note that the birth and death rates must be specified separately; the Malthusian parameter combination does not suffice to determine the evolution of the population.

To elucidate the behaviour of the model, let us continue by assuming random mating,

$$X_{i,jk} = p_i P'_{jk}, \quad (14)$$

and no selection,

$$a_{i,jk} = c, \quad a'_{i,jk} = c', \quad (15a)$$

$$d_i = d, \quad d'_{ij} = d'. \quad (15b)$$

Substituting (14) and (15) into (6) and (13) yields

$$d'_i = d', \quad b'_i = c'_i = c', \quad c_i = c. \quad (16)$$

Inserting (16) into (12) and (12') and recalling (10), we find

$$\dot{p}_i = c(p'_i - p_i), \quad (17)$$

$$\dot{p}'_i = (1/2)c'(p_i - p'_i). \quad (17')$$

From (17) and (17') it follows at once for constant fertilities that the quantity

$$c' p_i(t) + 2c p'_i(t) = c' p_i(0) + 2c p'_i(0) \quad (18)$$

is conserved. Denoting the difference between the corresponding allelic frequencies in the two sexes by

$$q_i(t) = p_i(t) - p'_i(t), \quad (19)$$

we obtain from (17) and (17')

$$\dot{q}_i = -[c + (1/2)c']q_i. \quad (20)$$

At equilibrium, (18) and (20) yield

$$p_i(\infty) = p'_i(\infty) = [c' p_i(0) + 2c p'_i(0)] / (c' + 2c). \quad (21)$$

Substituting (14) and (15) into (11') leads to

$$\dot{P}'_{ij} = (1/2)c'(p_i p'_j + p'_i p_j) - c' P'_{ij}. \quad (22)$$

It is easy to see from (17), (17'), and (22) that the deviation

$$Q'_{ij} = P'_{ij} - p'_i p'_j \tag{23}$$

from Hardy-Weinberg proportions in females satisfies

$$\dot{Q}'_{ij} = -c' Q'_{ij}. \tag{24}$$

The death rates do not appear in equations (17) to (24) because nonselective deaths do not alter the genotypic frequencies.

If the fertilities are constant, (20) and (24) give

$$q_i(t) = q_i(0)e^{-[c+(1/2)c']t}, \tag{25}$$

$$Q'_{ij}(t) = Q'_{ij}(0)e^{-c't}. \tag{26}$$

From (9) we find readily that the sex ratio  $s = N/N'$  always satisfies

$$\dot{s} = (\bar{m} - \bar{m}')s, \tag{27}$$

which in this case reduces to

$$\dot{s} = (c - c' - d + d')s. \tag{28}$$

Hence, if births do not change the sex ratio, *i.e.* if  $c' = c$ , then (18) and (25) simplify to

$$p_i(t) + 2p'_i(t) = p_i(0) + 2p'_i(0), \tag{29}$$

$$q_i(t) = q_i(0)e^{-(3c/2)t}. \tag{30}$$

### 3. DISCUSSION

The result (29) agrees with the well-known constancy of the weighted mean of gene frequencies in the discrete, non-overlapping generations model (see, *e.g.*, Crow and Kimura, 1970, pp. 44-47). In our continuous model, however, the difference between the male and female gene frequencies approaches zero monotonically according to (30), while in the discrete case the decay is oscillatory (see, *e.g.*, Crow and Kimura, 1970, pp. 44-47):

$$q_i(t) = q_i(0)(-1/2)^t. \tag{31}$$

The difference between (30) and (31) arises because discrete changes in gene frequency allow an overshoot. The approach to Hardy-Weinberg in the discrete model is easily derived. From (23),

$$\begin{aligned} Q'_{ij}(t+1) &= (1/2)[p_i(t)p'_j(t) + p'_i(t)p_j(t)] - (1/4)[p_i(t) + p'_i(t)][p_j(t) + p'_j(t)] \\ &= -(1/4)q_i(t)q_j(t) = (1/4)Q'_{ij}(t), \end{aligned} \tag{32}$$

where we used (31) in the last step.

Therefore,

$$Q'_{ij}(t) = (1/4)^t Q'_{ij}(0). \tag{33}$$

Thus, the deviation from Hardy-Weinberg tends to zero without oscillation. It is interesting to note that in the continuous model the gene frequency difference tends to zero faster than the deviation from Hardy-Weinberg, whereas in the discrete case the opposite result holds.

A question of considerable biological importance is which of the two simple models discussed at the beginning of the paper approximates better a natural population with overlapping generations. The most sensitive test provided by the theory developed above is an experiment to look for the presence or absence of oscillation in the approach to equality of the male and female gene frequencies at an X-linked locus. It would be desirable to follow the male and female gene frequencies in a continuously-reproducing population of *Drosophila* or some other suitable organism. Nature would be simulated most closely by starting the population in demographic equilibrium, and the results would be simplest and least ambiguous if the mutant used were nearly neutral. One can monitor the amount of selection by recalling that the weighted gene frequency combination (18) remains constant as long as the gene is neutral and births do not alter the sex ratio.

*Acknowledgment.*—I thank Professor James F. Crow for helpful discussions and suggestions.

#### 4. REFERENCES

- CHARLESWORTH, B. 1970. Selection in populations with overlapping generations. I. The use of Malthusian parameters in population genetics. *Theor. Pop. Biol.*, 1, 325-370.
- CROW, J. F., AND KIMURA, M. 1970. *An Introduction to Population Genetics Theory*. Harper and Row, New York.
- FISHER, R. A. 1930. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.
- NAGYLAKI, T. 1974. Continuous selective models with migration and mutation. *Theor. Pop. Biol.*, 5, 284-295.
- NAGYLAKI, T., AND CROW, J. F. 1974. Continuous selective models. *Theor. Pop. Biol.*, 5, 257-283.