FREQUENCY CHANGES OF NEW INVERSIONS IN POPULATIONS UNDER MUTATION-SELECTION EQUILIBRIA¹

MASATOSHI NEL,² KEN-ICHI KOJIMA,³ AND HENRY E. SCHAFER

Department of Genetics, North Carolina State University, Raleigh, North Carolina 27607

Received August 21, 1967

ESTABLISHMENT of an inversion arrangement of any chromosome consists of three stages. The first stage is represented by stochastic survival of an inversion which has newly occurred in a given population. This stage is stochastic because a given inversion probably appears only in one individual in a population whether this population is large or small in size; and consequently the survival process of inversion descendants is subjected to stochastic errors. OHTA and KOJIMA (unpublished) examined such a stochastic process in large populations using a simplified genetic model.

The second evolutionary stage of inversions is the phase in which the frequency of an inversion increases in a population. For a small population, this stage is also stochastic and cannot be distinguished from the first stage. However, once the number of inversion descendants becomes large in a large population, the chance elimination of all inversion descendants becomes extremely unlikely, and the second stage can be treated deterministically using the relative frequency of inversion descendants.

The third stage concerns the genetic mechanism(s) by which inversions and their standard arrangements may be brought to a balanced polymorphism. Some conditions for this polymorphism were investigated by HALDANE (1957).

This paper deals with the second evolutionary stage of inversions under a simple condition. In particular, the condition is that the selective advantage of an inversion is caused by a favorable combination of alleles in the inverted segment of a chromosome when the population is in mutation-selection equilibrium. The specific objective of this paper is to examine whether an inversion chromosome carrying few deleterious alleles relative to the average number of such alleles per chromosome can increase in frequency in such a population.

Partially Recessive Mutations

Consider a large random mating population in which selectively advantageous alleles, A 's, and partially recessive deleterious alleles, a 's, are maintained in the

¹ Paper No. 2377 of the Journal Series of the North Carolina State University Agricultural Experiment Station, Raleigh. This work was financed by Public Health Research Grant GM-11546 and also by a grant from the Scientific Research Fund of the Ministry of Education, Japan. The computer time was financed by Public Health Service Grant No. FR-00011 to the Institute of Statistics.

² Division of Genetics, National Institute for Radiological Sciences, Chiba, Japan. The majority of this work was carried out during the senior author's two-month visit to North Carolina State University.

³ Present address: Department of Zoology, University of Texas, Austin, Texas 78712.

state of mutation-selection balance. The effects of genotypes at each locus are assumed to be equal and represented as

$$\begin{array}{ccc} AA & Aa & aa \\ 1 & 1-hs & 1-s \end{array}$$

where h and s are positive constants. Let the frequency of allele a be q , and assume the mutation of allele A to allele a with the rate of μ per allele per generation. Since μ is ordinarily a far smaller quantity than s and h , the equilibrium frequency of allele a , $\hat{q}$, is given approximately by μ/hs .

Since $\hat{q}$ is also expected to be a small quantity, the frequency distribution of chromosomes having r deleterious alleles among the total of n loci is given approximately by the Poisson distribution with mean $n\hat{q}$. Letting the frequency of chromosomes with r deleterious alleles be P_r ,

$$P_r = e^{-n\hat{q}} (n\hat{q})^r / r! \quad (1)$$

Under the assumption that the effects of individual loci determine the fitness value of an individual multiplicatively, the fitness value of an average individual is approximately given by

$$\bar{W}_{N/N} = \sum_{i=0}^n \sum_{j=0}^n P_i P_j (1-hs)^{i+j} \quad (2)$$

where the subscript N/N is used to denote noninversion zygotes. Formula (2) is exact if $(1-s) = (1-hs)^2$. The approximation is also reasonably good as long as deleterious homozygotes aa 's are rare. Formula (2) approximately reduces to

$$\bar{W}_{N/N} = e^{-2n\hat{q}hs} = e^{-2n\mu} \quad (3)$$

The above expression is also obtained by the use of the mutation load theory (MORTON, CROW and MULLER 1956).

Now assume that a new inversion occurs and this inversion contains r deleterious alleles. The average fitness of an inversion heterozygote (I/N) is approximately given by

$$\bar{W}_{I/N} = (1-hs)^r \sum_{i=0}^n P_i (1-hs)^i$$

which can be approximately reduced to

$$\bar{W}_{I/N} = e^{-(n\hat{q} + r)hs}. \quad (4)$$

Thus, the relative fitness of an inversion heterozygote in the population is given by

$$C_r = 1 + \bar{S}_r = \frac{\bar{W}_{I/N}}{\bar{W}_{N/N}} = e^{(n\hat{q} - r)hs} \quad (5)$$

For a small value of the exponent, $\bar{S}_r$ is approximately given as

$$\bar{S}_r \approx (n\hat{q} - r)hs. \quad (6)$$

Formula (5) or (6) can be used in determining whether a given inversion tends to increase or decrease in a population. For example, if $\mu = 10^{-5}$, $hs = 0.02$ and $n = 10^3$, an inversion with no deleterious alleles has a selective advantage of about 1% (i.e. $\bar{S}_0 = 0.01$). Therefore, such an inversion tends to increase in number. However, an inversion with more than $n\hat{q}$ deleterious alleles will be selectively eliminated from the population. In the above example, $n\hat{q}$ is 0.5 so that any

chromosome with one or more deleterious alleles is at a selective disadvantage.

At the early stage of the evolution of a new inversion, the accumulation of new deleterious mutations on the majority of inversion descendants may be ignored. Under this situation $\bar{S}_r$ remains constant. Thus, the survival probabilities of inversions in successive generations can be obtained by the use of probability generating functions (e.g. FISHER 1930; KOJIMA and KELLEHER 1962).

Over a long period of time, however, mutations will inevitably accumulate on the inversion chromosomes, and this process will lower the value of $\bar{S}_r$. This trend goes on until the frequency of deleterious alleles reaches $\hat{q}$ at all loci involved. At this point, $\bar{S}_r$ is zero; in other words the inversion chromosomes becomes selectively neutral. Thus, the inversion chromosome will not necessarily spread through the population, even if it has a selective advantage at the initial stage.

It is then important to know how frequent inversion descendants would become in the population before they become selectively neutral. Suppose that the original inversion chromosome has no deleterious alleles, and q is 0 among inversion descendants at a certain time (time = 0). The rate of change in q can be approximately given by

$$\frac{dq_t}{dt} = \mu - hsq_t \quad (7)$$

where t stands for time measured in generations and $(1-q)$ is approximately equal to one. Solving equation (7) with the initial condition of $q = 0$ at $t = 0$.

$$q_t = \hat{q} (1 - e^{-hst}) \quad (8)$$

where $\hat{q} = \mu/hs$.

The fitness values of inversion homozygote (I/I) and inversion heterozygote (I/N) can be expressed in terms of q_t , while the fitness value of noninversion homozygote (N/N) stays constant as before. These fitness values are tabulated in Table 1. In this table it should be noted that the fitness values of the three structural genotypes are approximately in the additive mode.

Denoting the frequency of inversion descendants at time t by x_t , the rate of change in x_t is now given as

$$\frac{dx_t}{dt} = n\mu e^{-hst} x_t (1-x_t). \quad (9)$$

TABLE 1

Fitness of inversion homozygote (I/I), heterozygote (I/N), and noninversion homozygote (N/N) at the 0th and tth generations, when initially there is no deleterious allele—partially recessive genes

Generation	I/I	I/N	N/N
0	1	$\frac{-n\mu}{e}$	$\frac{-2n\mu}{e}$
t	$\frac{-2n\mu(1-e^{-hst})}{e}$	$\frac{-n\mu(2-e^{-hst})}{e}$	$\frac{-2n\mu}{e}$
t (relative to I/I)	1	$\frac{-n\mu e^{-hst}}{e}$	$\frac{-2n\mu e^{-hst}}{e}$
t (approximation)	1	$1-n\mu e^{-hst}$	$1-2n\mu e^{-hst}$

The solution of equation (9) is

$$\frac{x_t}{1-x_t} = \frac{x_0}{1-x_0} e^{n\hat{q}(1-e^{-hs t})} \quad (10)$$

where x_0 is the frequency of inversion chromosomes at time 0. Equation (10) states that the ratio of $x/(1-x)$ eventually increases to $e^{n\hat{q}}$ times the original ratio. In general, $n\hat{q}$ is not a large number. For example, with $\mu = 10^{-5}$, $hs = 0.02$ and $n = 10^3$, $n\hat{q}$ is only 0.5, and the ratio $x/(1-x)$ increases only to 1.65 times the original values. Since $x/(1-x)$ is about the same as x for small values of x , the increase of inversion descendants is not considered to be important for the partially recessive mutation model.

Completely Recessive Mutations

When deleterious mutations are completely recessive to wild-type alleles, the frequency of mutant alleles, a 's, is given by $\hat{q} = (\mu/s)^{1/2}$ where s is the degree of disadvantage of recessive homozygotes. Since $\hat{q}$ is a small value compared to one, the frequency of chromosomes with r deleterious alleles is given again by P_r in formula (1). Under the same set of assumptions as before, the individuals with a chromosome carrying r deleterious alleles will have the average fitness of

$$\bar{W}_r = (1 - s\hat{q})^r \approx e^{-rs\hat{q}} \quad (11)$$

because each of the r loci has a chance of $\hat{q}$ to be in the homozygous condition. The population average of fitness values is

$$\begin{aligned} \bar{W} &= \sum_i P_i (1 - s\hat{q})^i \\ &= e^{-ns\hat{q}^2} = e^{-n\mu}. \end{aligned} \quad (12)$$

This last expression agrees with the value predicted from the mutation load theory. From (11) and (12), the relative fitness of a chromosome with r deleterious alleles is

$$C_r = 1 + \bar{S}_r = e^{(n\hat{q} - r)s\hat{q}}. \quad (13)$$

An example may serve to give a notion as to the magnitude of $\bar{S}_r$. If $s = 0.1$, $\mu = 10^{-5}$ and $n = 10^3$, then $n\hat{q} = 10$ and $\bar{S}_r$ is approximately equal to $0.001(10 - r)$. In this case, then, chromosomes with nine or less deleterious alleles will have a selective advantage over the average chromosome in the population.

The relative fitness values of I/I , I/N and N/N for inversions with 0, 1 and r deleterious alleles are found in Table 2. It should be noted in this case that the inversion chromosome with no deleterious alleles is completely dominant over the noninversions, but the inversion heterozygote shows heterosis when $1 \leq r < n\hat{q}$.

As in the previous case, new deleterious mutations occurring on the inverted segments which had an initial selective advantage will gradually reduce the degree of advantage until the average inversion becomes neutral with respect to the average noninversion chromosome. The fitness values of I/I , I/N and N/N at generations 0 and t are presented for the case of $r = 0$ in Table 3. The gene frequency, q_i , can be approximately obtained by solving the equation

$$\frac{dq_t}{dt} = \mu - s\hat{q}q_t. \quad (14)$$

The solution of (14) is

$$q_t = (\mu/s)^{1/2} (1 - e^{-(s\mu)^{1/2}t}). \quad (15)$$

The differential equation for the change in inversion frequency (corresponding to formula (9)) can be obtained by using (15). However, the analytical solution of this differential equation is not easy to obtain, and for this reason it was decided to examine the increase in inversion frequency by carrying out numerical analyses of several examples using a computer. Two approaches were taken: one was to compute cumulative changes over generations according to the fitness value expressions in Table 3 using q_t given in (15), and the other was exact deterministic simulation of populations containing inversions. The details of the latter approach are given in the next section and the results for $r = 0$ are presented in Figure 1.

The effects of the number of loci (n), the initial frequency of inversions (x_0) and the degree of selective advantage(s) on the increase of x are quite clear. Starting from a small value of x_0 (0.00001 or 0.001 in the present examples), n needs to be of the order of 1,000 rather than 100 for the frequency of inversions to become common in a population before their selective advantage disappears. Secondly, the smaller the value of s , the greater is the frequency of inversions at the end. In general, then, one may conclude that a larger mutation free inversion has better prospects of becoming abundant in a population. Under some conditions, e.g., $n = 1,000$ and $s = 0.01$, inversions can almost completely replace noninversion chromosomes.

TABLE 2

Fitness of inversion homozygote (I/I), heterozygote (I/N), and noninversion homozygote (N/N) in the 0th generation—completely recessive genes

No. of deleterious genes involved	I/I	I/N	N/N
0	1	1	$e^{-ns\hat{q}^2}$
1	e^{-s}	$e^{-s\hat{q}}$	$e^{-ns\hat{q}^2}$
r	e^{-rs}	$e^{-rs\hat{q}}$	$e^{-ns\hat{q}^2}$

TABLE 3

Fitness of inversion homozygote (I/I), heterozygote (I/N), and noninversion homozygote (N/N) in the 0th and tth generation—completely recessive genes and $r = 0$

Generation	I/I	I/N	N/N
0	1	1	$e^{-ns\hat{q}^2}$
t	$e^{-nsq_t^2}$	$e^{-ns\hat{q}q_t}$	$e^{-ns\hat{q}^2}$
t (relative to I)	1	$e^{-nsq_t(\hat{q}-q_t)}$	$e^{-ns(\hat{q}^2-q_t^2)}$
t (approximation)	1	$1 - nsq_t(\hat{q}-q_t)$	$1 - ns(\hat{q}^2 - q_t^2)$

Exact Deterministic Simulation

The deterministic changes in the population can be more exactly described in terms of the gene frequencies at each of the loci and the inversion frequency. For this description the gene frequency notation for generation t is p_{Nt} , q_{Nt} , p_{It} , q_{It} where q refers to the deleterious allele and N and I refer, respectively, to the non-inversion and inversion-bearing chromosomes. These four quantities sum to unity. The sum of p_{It} and q_{It} at each locus is equal to x_t , the frequency of the inversion chromosome. The p and q values at each locus change as a result of selection and mutation (back mutation of the deleterious allele is neglected). This selection and mutation on an individual locus basis results in relative changes of p_{Nt} versus q_{Nt} and p_{It} versus q_{It} . The relative frequencies of the deleterious allele within the two groups of chromosomes are denoted by Q_{Nt} and Q_{It} .

$$Q_{Nt} = q_{Nt}/(1-x_t) \quad (16)$$

$$Q_{It} = q_{It}/x_t.$$

The change in x_t is a function of the number of loci as well as of the gene frequencies and the selective coefficient.

The changes in the population from generation t to generation $t+1$ are given by (17-21).

$$x_{t+1} = x_t [x_t(1-sQ_{It}^2)^n + (1-x_t)(1-sQ_{It}Q_{Nt})^n] / \bar{w} \quad (17)$$

$$q_{I(t+1)} = \frac{q_{It} [1 - s(q_{Nt} + q_{It})] + \mu p_{It} \bar{w}_i}{x_t - s q_{It}(q_{Nt} + q_{It})} x_{t+1} \quad (18)$$

$$p_{I(t+1)} = x_{t+1} - q_{I(t+1)} \quad (19)$$

$$q_{N(t+1)} = \frac{q_{Nt} [1 - s(q_{Nt} + q_{It})] + \mu p_{Nt} \bar{w}_i}{1 - x_t - s q_{Nt}(q_{Nt} + q_{It})} (1 - x_{t+1}) \quad (20)$$

$$p_{N(t+1)} = 1 - x_{t+1} - q_{N(t+1)} \quad (21)$$

where

$$\bar{w} = x_t^2(1-sQ_{It}^2)^n + 2x_t(1-x_t)(1-sQ_{It}Q_{Nt})^n + (1-x_t)^2(1-sQ_{Nt}^2)^n \quad (22)$$

and

$$\bar{w}_i = 1 - s(q_{It} + q_{Nt})^2. \quad (23)$$

$\bar{w}$ is the average fitness of the population with respect to chromosomes and $\bar{w}_i$ is with respect to a single locus.

The frequency of the deleterious allele at a locus in the noninversion chromosomes is initially $Q_{N0} = \hat{q}$. The initial frequency in the inversion chromosomes is $Q_{I0} = 0$. The value of x_0 used in the computations which produced Figure 1 is varied, as are the values of n and s , to produce the separate curves.

In this formulation Q_{Nt} may vary instead of being assumed to remain equal to $\hat{q}$. Since $Q_{I0} = 0$, x_t increases and the effect is to decrease the frequency of the deleterious alleles in the population. The mutation load at each locus, $s(q_{Nt} + q_{It})^2$, then becomes less than its equilibrium value, $s\hat{q}^2$, and the imbalance between mutation and selection causes the frequency of deleterious alleles to increase within noninversion chromosomes. Among the non-inversion chromosomes

FIGURE 1.—The pattern of increase in inversion frequency (x) over generations (t). The frequency is expressed on the logarithmic scale (base 10). The 12,000 generation values of x are also shown for cases 1 and 5 (indicated by asterisks). The specifications for each case are as follows:

Case	x_0	s	n
1	10^{-5}	0.01	100
2	10^{-5}	0.01	1000
3	10^{-5}	0.10	100
4	10^{-5}	0.10	1000
5	10^{-3}	0.01	100
6	10^{-3}	0.01	1000
7	10^{-3}	0.10	100
8	10^{-3}	0.10	1000

Q_{Nt} will be affected as shown by (16) and (20) and will become greater than $\hat{q}$. With Q_{Nt} greater than $\hat{q}$ the inversion has a larger selective advantage than when Q_{Nt} equals $\hat{q}$ and so the differential equation always underestimates the frequency of the inversion. In later generations the inversion approaches selective neutrality as Q_{Nt} and Q_{It} both approach $\hat{q}$.

In most cases the underestimation of x_t is negligible. This can be seen from the

TABLE 4

Inversion frequencies after 6,000 generations (x_{6000}) as given by the differential equation and exact deterministic simulation approaches. (For identification of cases see Figure 1)

Case	Differential equation	Deterministic simulation
1	.000147	.000150
2	.999942	.999996
3	.000027	.000027
4	.169745	.184201
5	.014411	.014821
6	.999989	.999997
7	.002703	.002703
8	.922554	.957763

x_{6000} values given in Table 4. When x_t increases slowly (cases 1, 3, 5, 7 in Figure 1) Q_{Nt} remains almost constant and the underestimation is very small. It is when x_t increases rapidly that x_t is more severely underestimated because of the increase of Q_{Nt} over $\hat{q}$. The largest such increase in this study is found at generation 3600 in case 6 where Q_N has reached a value of 1.26 times the equilibrium value of $\hat{q} = .03162$. However, in those cases when x_t rapidly reaches the neighborhood of unity, the influence of Q_{Nt} is minimized by the noninversion chromosomes becoming rare. In such cases (cases 2, 6, 8 in Figure 1) the differential equation also shows x_t nearly reaching unity and so an appreciable degree of underestimation is only possible during the increasing stage of the curve.

It is in those intermediate situations where x_t increases rapidly to an intermediate value, such as in case 4, that the greatest discrepancy occurs even though Q_N does not show the greatest increase over $\hat{q}$. For example, in case 4, Q_N never is more than 1.0035 times the equilibrium value of $\hat{q} = .01$, and yet $x_{6000} = .1842$, while the numerical solution of the differential equation yields $x_{6000} = .1697$. This is the greatest difference between the results presented in Figure 1 and the results from the solution of the differential equation for the same cases.

DISCUSSION

In this paper partially recessive mutations and completely recessive mutations were treated separately in order to emphasize the different effects of these mutations on the increase of inversion chromosomes. However, the selection coefficient for heterozygotes (h_s) varies considerably for different mutations with the average between 0.02 and 0.05. Thus, in an actual genetic population the rate of increase in inversion frequency will be closer to the value described for partially recessive mutations than that for completely recessive mutations. This means that in general a newly arisen inversion does not become prevalent in an average population. Only when the inverted segment contains a large number of completely dominant favorable alleles do the inversion descendants become abundant. This conclusion is, however, not incompatible with the idea that occasionally there are new inversions which will spread through a population. VANN (1966)

has shown in *Drosophila melanogaster* that the frequency of radiation-induced inversions can increase considerably in artificial populations and the rate of increase of the frequency is highly correlated with the length of the inverted segment.

However, the genetic models being considered have some limitations in view of the present knowledge of the behavior of inversion chromosomes. Most inversion chromosomes maintained in an intermediate frequency in natural populations are kept at intermediate frequencies by selection pressure. However, under the present model, inversions which become common in populations tend to become selectively neutral in the long run with respect to noninversions. The present simple model depicts a possible process of inversions becoming abundant in a population, but does not necessarily agree to the observation on the end product of the process. There must be some other or additional factors of evolution involved in the process of obtaining a lasting balance between inversions and noninversions. Examples of such mechanisms are heterosis (e.g. DOBZHANSKY 1955) and frequency dependent selection (TEISSIER 1954; KOJIMA and YARBROUGH 1967).

There are several factors in the model which can be modified to meet the requirement of the observed balance between inversions and their standard chromosomes. For example, the inclusion of epistatic genes showing cumulative overdominance will produce superiority of inversion heterozygotes and stabilize the polymorphism in inversion chromosomes (cf. HALDANE 1957). The position effect of chromosome rearrangement is another possibility. As suggested by SPERLICH (1966), the inversion chromosome may have a position effect, and this position effect may be favorable in single dose but unfavorable in double dose. On the other hand, KOJIMA (1967) has shown in a Monte Carlo study that a balanced inversion polymorphism can be produced in a small population in which truncated selection is proceeding on dominant genes without epistasis or overdominance. The mechanism involved in KOJIMA's case was that inversion chromosomes with relatively few recessive deleterious alleles formed, with certain probability, heterotic polymorphisms with one kind of noninversion with recessive deleterious alleles which were different from those on the inversions.

SUMMARY

Some possibilities for an inversion to become abundant in a random mating population in mutation-selection equilibrium were theoretically investigated. The findings were that inversions with a large number of completely dominant favorable alleles could increase in frequency considerably when they were extremely rare at the beginning. Some of the difficulties in explaining observed inversion polymorphisms were discussed in terms of the models investigated.

LITERATURE CITED

- DOBZHANSKY, TH., 1955 A review of some fundamental concepts and problems of population genetics. Cold Spring Harbor Symp. Quant. Biol. **20**: 1-15.

- FISHER, R. A., 1930 *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.
- HALDANE, J. B. S., 1957 The conditions for coadaptation in polymorphism for inversions. *J. Genet.* **55**: 218-225.
- KOJIMA, K., 1967 Likelihood of establishing newly induced inversion chromosomes in small populations. *Ciencia e Cultura (São Paulo)* **19**: 67-77.
- KOJIMA, K., and T. M. KELLEHER, 1962 Survival of mutant genes. *Am. Naturalist* **96**: 329-345.
- KOJIMA, K., and K. M. YARBROUGH, 1967 Frequency-dependent selection at the Esterase 6 locus in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.* **57**: 645-649.
- MORTON, N. E., J. F. CROW, and H. J. MULLER, 1956 An estimate of the mutational damage in man from data on consanguineous marriages. *Proc. Natl. Acad. Sci. U.S.* **42**: 855-863.
- SPERLICH, D., 1966 Equilibria for inversions induced by X-rays in isogenic strains of *Drosophila pseudoobscura*. *Genetics* **53**: 835-842.
- TEISSIER, M. G., 1954 Sélection naturelle et fluctuation génétique. *Compt. Rend.* **238**: 1929-1931.
- VANN, E., 1966 The fate of X-ray induced chromosomal rearrangements introduced into laboratory populations of *Drosophila melanogaster*. *Am. Naturalist* **100**: 425-449.