THE EFFECTS OF OVERDOMINANCE ON LINKAGE IN A MULTILOCUS SYSTEM[1]

TSUNEYUKI YAMAZAKI[2]

National Institute of Genetics, Mishima, 411 Japan

AND

National Institute of Environmental Health Sciences

Research Triangle Park, North Carolina, U.S.A.

Manuscript received December 8, 1975
Revised copy received November 29, 1976

ABSTRACT

Computer simulations were performed with overdominant multiple alleles among tightly linked multiple loci under a multiplicative fitness model. The quantity $\chi^2/N(n-1)$ was introduced as a new measure of linkage disequilibrium which, unlike previously available measures, can be applied to multiple allele models, where N is the sample size, and n is the number of alleles at the locus possessing fewest alleles. Simulations showed that (1) With multiple (three or four) alleles, the approach to stable disequilibrium is slower and the amount of disequilibrium established is weaker than in a two allele system. (2) The number of complementary chromosomes is a function of number of alleles and of population size. (3) As population size increases, the rate of the approach to stable disequilibrium is slower. (4) There is an optimum selection coefficient which minimizes the transient fixation probability of alleles when linkage is present. (5) The absence of linkage disequilibrium is in most cases not a practical method of testing the hypothesis of balancing selection of genetic polymorphisms because it depends strongly on population size in determining linkage disequilibria.

MOST population genetics theory has considered a gene as an independent unit of selection. However, the effect of linkage has sometimes been taken into consideration in theoretical population genetics (*e.g.*, KIMURA 1956; LEWONTIN and KOJIMA 1960).

Recently it has been shown that initial genes may become strongly associated with each other in two complementary chromosomes by means of natural selection if many loci are tightly linked, if two alleles at each locus show overdominance, and if there is nonadditive interaction between loci (FRANKLIN and LEWONTIN 1970; WILLS, CRENSHAW, and VITALE 1970). An analytical treatment of this problem was carried out by SLATKIN (1972). A similar complementary organization of linked genes may also occur under frequency-dependent selection (YAMAZAKI 1974). In this context, the term "complementary chromosomes" refers to sets of chromosomes in which the great majority of alleles

[1] Contribution No. 1070 from the National Institute of Genetics, Mishima, Shizuoka-ken 411 Japan.

[2] Present address: Department of Biology, Kyushu University, Fukuoka-city 812 Japan.

are mutually exclusive. Complete complementarity will be attained only in extremely large populations, but the contribution of the type described here is nevertheless of great biological importance.

This paper examines by computer simulation the influence of linkage on the evolution of multilocus chromosomes where there are two or more overdominant alleles at each locus. A new measure of linkage disequilibrium (L) is introduced which allows the comparison of multi-allelic systems. This measure is superior to previously used measures, as these permit only the comparison of two-allele systems. The effects of population size on the establishment of linkage disequilibrium and fixation probability of alleles are also investigated. A preliminary report on this subject has been published (YAMAZAKI 1973).

METHODS

Most computer simulations were carried out using Tosbac 3400 at the National Institute of Genetics. Computations with population size $2N = 800$ were performed using PdP 1140 at the National Institute of Environmental Health Sciences in North Carolina. In all the simulations reported here, a chromosome is composed of 88 linked loci. Crossing over occurs independently without interference and the rate between two adjacent loci is, in most cases, 0.0022. An individual carries the diploid chromosome number, two. The fitness of an individual is the product of all the fitness effects at the 88 loci. The population size varies depending on the particular simulation. (A flow chart of this simulation can be obtained from the author.)

RESULTS

Measure of linkage disequilibrium

Considering two loci A, B, each having 2 alleles, say A_1, A_2, and B_1, B_2, there are four possible combinations of the gametes. Let us denote them by A_1B_1, A_1B_2, A_2B_1, and A_2B_2, and let N_{11}, N_{12}, N_{21} and N_{22} be the sample number of A_1B_1, A_1B_2, A_2B_1 and A_2B_2 respectively. Then,

$$D = \frac{1}{N^2}(N_{11}N_{22} - N_{12}N_{21}) = P_{11}P_{22} - P_{12}P_{21}$$

where N is the sample size, and P_{11}, P_{12}, P_{21}, P_{22} is the frequency of chromosome A_1B_1, A_1B_2, A_2B_1 and A_2B_2, respectively (KIMURA 1956; LEWONTIN and KOJIMA 1960).

"D" is often used to describe the gametic array in a population. The value of D lies between -0.25 and 0.25. Since the D value is strongly dependent on the gene frequency, it is misleading to compare D values between two different systems unless the gene frequencies are the same. Moreover, D cannot be defined for a multiple-allele system.

In this paper I will propose a different measure of linkage disequilibrium, "L", which is directly related to x^2 and has several desirable features (YAMAZAKI 1973). Let

p_i, p_j = frequency of alleles A_i, A_j $(i, j = 1, 2, \ldots n)$
q_k, q_l = frequency of alleles B_k, B_l $(k, l = 1, 2, \ldots m)$

Assume that symbols are assigned so that $m \geq n$ and the alleles at each locus are numbered in order of decreasing frequency. Let

$$P_{ik} = \text{frequency of chromosome } A_iB_k \ (p_i = \sum_k p_{ik})$$

With these definitions, we define linkage disequilibrium by

$$L = \frac{1}{n-1} \Sigma \frac{(p_{ik} - p_i q_k)^2}{p_i q_k} = \frac{1}{n-1} \Sigma \frac{D^2_{ik}}{p_i q_k} = \frac{\chi^2}{N(n-1)}$$

L takes its maximum value, 1, when

$$\begin{aligned} p_{ik} &= p_i = q_k \ , \quad & i &= k \\ p_{ik} &= 0 \qquad , \quad & i &\neq k \end{aligned}$$

The minimum, of course, is 0 when $p_{ik} = p_i q_k$. When $m > n$, the maximum value of L occurs when $q_k = 0$ for all $k > n$. For this reason $n - 1$ appears in the denominator as a normalizing factor instead of, for example, $m - 1$ or some kind of average.

For two alleles at each locus

$$L = \frac{\chi^2}{N(2-1)} = \frac{(p_{11}p_{22} - p_{12}p_{21})^2}{p_1 p_2 q_1 q_2} = \frac{D^2}{p_1 p_2 q_1 q_2}$$

This is equivalent to the measure σ_d^2 used by OHTA and KIMURA (1969), except that they considered expected values rather than observations. Recently HILL (1975) has provided a very similar analysis and suggested χ^2 as a natural measure of disequilibrium.

If the direction of linkage disequilibrium is meaningful, as it may be for two alleles, L is taken to have the same sign as D. The value L is equivalent to the squared correlation between gene frequencies of different loci in the case of two alleles at each of two loci. The amount of dependence of gene frequencies between two loci, or linkage disequilibrium, is indicated quite adequately by L, and the independence of gametic frequencies between two loci can be tested statistically by χ^2. It is an important property of L that it is independent of either the gene frequency or the number of alleles. L is, of course, not independent of gene frequency in the sense that the maximum value of L cannot be 1 when the gene frequencies of both loci in question are not the same. However, as the correlation of gene frequencies between two loci increases, the gene frequencies approach each other. L increases to 1 when the complete association of alleles is established. At this stage, the gene frequencies at both loci become equal. When the complete association of alleles is established, L is always 1 regardless of gene frequency. This measure therefore has several desirable properties, as it can be used to compare the degree of linkage disequilibrium between systems with different gene frequencies, with different number of alleles per locus, or with different numbers of loci. The ability of L to quantify linkage disequilibria between more than two loci will presumably be important in the future when data involving multilocus systems in natural populations become available.

Gametic arrays

When each locus has 2 overdominant alleles, the qualitative results are almost the same as those obtained by FRANKLIN and LEWONTIN (1970). In natural populations there are many loci with more than two alleles. The simulations were extended to multiallelic cases to test whether the same phenomenon is observed when more than two overdominant alleles are present at each locus. The fitness of all heterozygotes is assumed to be 1, and of homozygotes to be $1-S$.

A summary of these results is shown in Figures 1 and 2 (in terms of L) and in Table 1 (in terms of gametic arrays). With one exception, simulations were

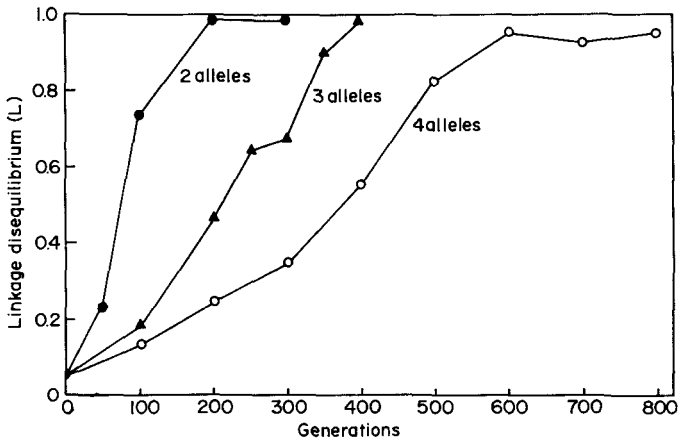

FIGURE 1.—Simulation of linkage disequilibrium (L) for the two-allele (●), three-allele (▲) and four-allele (○) case. (See text for details).

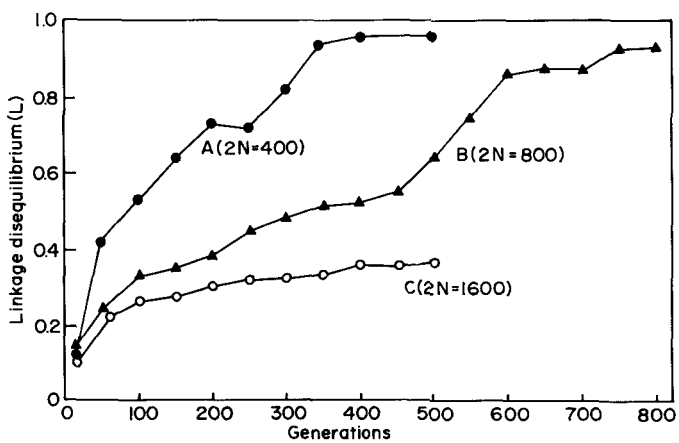

FIGURE 2.—Effect of population size on linkage disequilibrium (L).

A: Population size $2N = 400$ (●).

B: Population size $2N = 800$ (▲).

C: Population size $2N = 1600$ (○). Simulation was terminated after 500 generations due to excessive computation time.

repeated at least four times under the same conditions, although Figures 1 and 2 and Table 1 show only one example of each simulation. In the case where $2N = 1600$ (Figure 2) only one simulation was carried out for practical reasons of computer time. As the number of alleles per locus increases, the rate of increase of linkage disequilibrium as measured by L decreases. In the case of two alleles, the maximum value of L is attained at approximately 200 generations, but under the three-allele model, about 400 generations are needed, and it takes approximately 600 generations for L to attain its maximum value under the four-allele model. Moreover, the amount of linkage disequilibrium obtained after equilibrium is established by selection for several hundred generations is weaker under the four-allele case than under the two-allele model (see Figure 1 and Table 1). Two complementary chromosomes were established by selection in the case of the two-allele model (FRANKLIN and LEWONTIN 1970). I expected to find three and four complementary chromosomes at equilibrium under the three- and four-allele models, respectively. In the actual simulations, however, only two complementary chromosomes were established under the three-allele model and only three under the four-allele model. The gametic arrays of the two- and four-allele cases are shown in Table 1. At least four runs were performed under the same condition in each case, but the same result was always obtained: one of the complementary chromosomes was lost before equilibrium was reached.

Under what conditions can we obtain all possible numbers of complementary chromosomes? The effects of population size on the formation of linkage disequilibrium were investigated by performing simulations with an initial number of four alleles per locus, under different population sizes. The results are shown in Figure 2. When the population size was small ($2N = 400$), the linkage disequilibrium was established quickly and two types of chromosomes complementary to each other remained in the population. When the size was increased to $2N =$

TABLE 1

The gametic arrays of two- and four-allele cases

Chromosome constitution					Chromosome frequency
2 alleles (after 180 generations)					
*		* *		*	
21222	12222	11112	22221	11212	42.5%
12121	21111	12211	11112	22111	43.0%
				others	14.5%
4 alleles (after 800 generations)					
*	*	* **	*		
33012	20001	30022	11323	01223	22.4%
22113	22322	20322	02300	13311	25.0%
11213	23113	00122	33311	32232	30.5%
				others	22.1%

* Indicates loci fixed after selection. Numbers 1, 2, 3, and 4 indicate different alleles.

800, three complementary types were eventually established, but it took much longer, approximately 600 generations, to reach equilibrium. When the size of the population increased to $2N = 1600$, the rate of increase of linkage disequilibrium became very slow. Even after 500 generations of selection the value of L was about 0.4. At generation 500 there were still many different types of chromosomes in the population. With such large population sizes, computation time becomes excessive, and the computation was terminated at generation 500. The results shown in Figure 2 reveal an interesting phenomenon: the amount of linkage disequilibrium, or the formation of complementary chromosomes, depends on population size. This result differs from that of FRANKLIN and LEWONTIN (1970). However, it is also possible that the results shown in Figure 2 were obtained not because of the difference in population size, but because of the difference in the number of complementary chromosomes established after selection. The effects of population size alone were examined by performing simulations starting from two overdominant alleles at each locus, instead of four. The results are essentially the same as those in Figure 2. Large population size tends to retard the formation of linkage disequilibrium; L values greater than 0.90 were not established for 140 generations when $2N = 80$ and 240 generations when $2N = 200$. Approximately 600 generations of selection were necessary to reach an L value greater than 0.6 when $2N = 800$. All these results clearly indicate that the formation of complementary chromosome is not only delayed, but also weakened as population size increased.

The fixation probability of alleles

The effects of linkage on the probability of fixation or loss of allele in a 2-allele system when the selection coefficient varies was investigated by comparing results from linked systems with those from an unlinked system whose behavior is well known (*i.e.*, CROW and KIMURA 1970). The effect of linkage on the ultimate fixation probability of alleles was also studied by HILL and ROBERTSON (1966). Figure 3 shows the fixation probability of alleles as a function of the number of generations at a selection intensity s. The expected values of the corresponding fixation probability of unlinked genes are also shown.

For single genes, it is possible to compute the fixation probability $\mu(t, x)$ where t is time measured in generations and x is the initial frequency of the allele in question. Then it satisfies the following Kolmogorov backward equation:

$$\frac{\partial u(t,x)}{\partial t} = \frac{x(1-x)}{4N}\frac{\partial^2 u(t,x)}{\partial x^2} + s(1-2x)x(1-x)\frac{\partial u(t,x)}{\partial x}$$

with boundary conditions

$$\begin{aligned} u(0,x) &= 1 \quad \text{for } x = 1 \\ &= 0 \quad \text{for } 0 < x < 1 \\ u(t,1) &= 1 \text{ and } u(t,0) = 0 \quad \text{for } t > 0 \end{aligned}$$

The values of $u(t, x)$ in Figure 3 were obtained by solving numerically the above differential equation. Figure 3 presents the number of fixed loci among the total

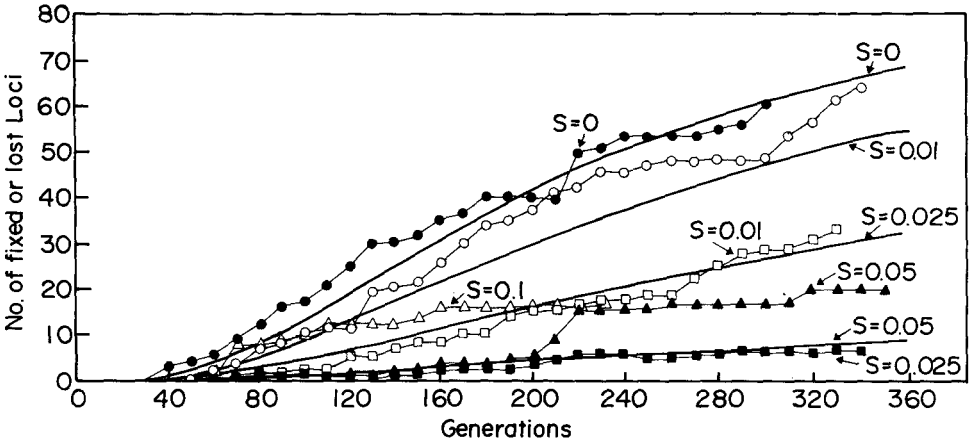

FIGURE 3.—Fixation probability of alleles as a function of number of generations at a selection intensity s. The expected values of the corresponding fixation probability of unlinked genes are also shown.

of 88 loci. The expected number of fixations was obtained assuming an initial allele frequency of 0.5.

When $s = 0$, fixed loci begin to appear from generations 40 to 60, and they increase linearly in later generations. In both of the replicates, fixation occurred at about half the loci by about generation 220. These results are in agreement with the theoretical expectations for a single locus. When s is 0.1 ($Ns = 10$), fixed loci begin to appear after 60 generations and continue to increase until around generation 170. However, once complementary chromosomes are established few further loci become fixed. When loci are not linked, the expected number of fixed loci is very small even after several hundred generations. The number is only 0.25 after selection for 400 generations and 0.5 after 1000 generations. In other words, when there is strong selection at each locus ($s = 0.1$ or 0.05), linkage of loci actually accelerates the fixation of alleles rather than retarding it, as least as far as the initial several thousand generations are concerned. When s is small, the situation is different; only 6.8% of the linked loci are fixed by generation 340 when s is 0.025, although the expected number for single loci is as much as 34%. When s is 0.01, both the expected and experimental values of fixation with linkage approach those expected on the hypothesis of neutrality. However, this case is similar to the case of $s = 0.025$, in the sense that fixation rate is lower with linkage than without it.

DISCUSSION

These simulations have revealed several interesting phenomena. Under conditions of overdominance and recombination that are not extremely restrictive and may well represent those found in nature, complementary chromosomes are produced by natural selection. Complementary chromosomes can also be established when frequency-dependent selection of the minority-advantage type is

operating (YAMAZAKI 1974). In this type of selection, gametic pools become divided into two groups, one with common alleles, the other with rare alleles, and their relative frequency changes every several generations. The characteristics in both cases are that only super-fit individuals and near-lethals are produced in the population with the elimination of individuals of intermediate fitness.

The number of possible completely complementary chromosomes is, of course, in theory equal to the number of alleles per locus. However, these simulations have shown that the number of chromosomal types established in the population depends on population size. Less than the theoretically expected number of chromosomal types were always found using a multiple-allele model with a population size of several hundred. Moreover, the time to the establishment of equilibrium becomes longer and the amount of linkage disequilibrium established becomes less as the number of alleles per locus or population size increases.

The proportion of homozygotes decreases as the number of alleles per locus increases. The effective selection coefficients therefore decrease, and the time to reach equilibrium is longer as the number of alleles increases. The delay in reaching equilibrium with large population sizes may arise for several different reasons. One plausible explanation for this delay is that some chromosomes with disadvantageous allele combinations are easily eliminated in small populations, and linkage disequilibrium or complementary chromosomes are therefore relatively easily established.

In these simulations only a multiplicative fitness model was used, since the use of a different model (such as truncation selection) is unlikely to have an important effect on the results. Thus, linkage disequilibrium was established under several different models, as long as there were nonadditive interactions between loci (FRANKLIN and LEWONTIN 1970; WILLS, CRENSHAW and VITALE 1970).

Substantial amounts of data about linkage disequilibrium have been accumulated within the last few years in natural populations of several organisms, most notably *Drosophila melanogaster*. There is some evidence of the existence of linkage disequilibrium between enzyme loci in this species (*e.g* CHARLESWORTH and CHARLESWORTH 1973; ZOUROS and KRIMBAS 1973). However, extensive studies by MUKAI, WATANABE and YAMAGUCHI (1974) and LANGLEY, TOBARI and KOJIMA (1974) could demonstrate little linkage disequilibrium between enzyme loci in natural populations of *D. melanogaster*. Considering all the available evidence it seems that there is, in general, rather little linkage disequilibrium between enzyme loci in natural populations (with the exception of a few special chromosomal rearrangements or in selffertilizing organisms).

Of course, nature is more complex than are the simulations described here. Therefore, an absence of linkage disequilibrium between loci coding for protein polymorphisms in natural populations does not prove the absence of balancing selection such as overdominance or frequency-dependent selection at such loci, but it may suggest that simple overdominance or frequency-dependent selection with epistasis between these loci is not a general occurrence. If balancing selection is common, however, it is possible that strong linkage disequilibrium is not

found, as it is less likely to be established in large populations. One of the most important conclusions of this paper is therefore that the presence or absence of linkage disequilibrium without information on population size and complete number of alleles (SINGH, HUBBY and THROCKMORTON 1975) is not an efficient test for distinguishing between selection and neutral hypotheses. Of course, little linkage disequilibrium is expected under any form of balancing selection if selection is additive between loci.

Another interesting discovery in the present study is that, as is shown in Figure 3, intermediate selection coefficients are most efficient in maintaining polymorphisms with strong linkage, although this effect is strongly dependent on the degree of recombination. When selection is weak, linkage tends to retard the fixation of alleles. When selection is intense, few loci are fixed without linkage, but, with linkage, a substantial number of loci are fixed during the establishment of complementary chromosomes. This effect may be important in understanding the behavior of alleles in small populations.

I am very grateful to DR. M. KIMURA and DR. T. MARUYAMA for the help and valuable discussion during the whole course of the present work. DR. MARUYAMA computed the expected values of fixation probability of alleles without linkage. I also wish to thank DR. J. F. CROW, DR. J. S. JONES, DR. C. H. LANGLEY and DR. M. SLATKIN for their valuable comments on the manuscript. The discussion with DR. R. C. LEWONTIN and DR. T. PROUT was also valuable for the completion of the manuscript.

LITERATURE CITED

CHARLESWORTH, B. and D. CHARLESWORTH, 1973 A study of linkage disequilibrium in population of *Drosophila melanogaster*. Genetics **73**: 351–359.

CROW, J. F. and M. KIMURA, 1970 *An introduction to population genetics theory*. Harper, New York.

FRANKLIN, I. and R. C. LEWONTIN, 1970 Is the gene the unit of selection? Genetics **65**: 707–734.

HILL, W. G., 1975 Linkage disequilibrium among multiple neutral alleles produced by mutation in a finite population. Theor. Pop. Biol. **8**: 117–126.

HILL, W. G. and A. ROBERTSON, 1966 The effect of linkage on limits to artificial selection. Genet. Res. **8**: 269–287.

KIMURA, M., 1956 A model of a genetic system which leads to closer linkage by natural selection. Evolution **10**: 278–287.

KIMURA, M., 1962 On the probability of fixation of mutant genes in a population. Genetics **47**: 713–719.

KIMURA, M. and T. OHTA, 1971 *Theoretical aspects of population genetics*. Princeton University Press, Princeton, New Jersey.

LANGLEY, C. H., Y. N. TOBARI and K. KOJIMA, 1974 Linkage disequilibrium in natural populations of *Drosophila melanogaster*. Genetics **78**: 921–936.

LEWONTIN, R. C. and K. KOJIMA, 1960 The evolutionary dynamics of complex polymorphisms. Evolution **14**: 458–472.

MUKAI, T., T. K. WATANABE and O. YAMAGUCHI, 1974 The genetic structure of natural populations of *Drosophila melanogaster*. XII. Linkage disequilibrium in a large local population. Genetics **77**: 771–793.

SINGH, R. S., J. L. HUBBY and L. H. THROCKMORTON, 1975 The study of genic variation by electrophoretic and heat denaturation techniques at the octanol dehydrogenase locus in members of the *Drosophila virilis* group. Genetics **80**: 637–650.

SLATKIN, M., 1972 On treating the chromosome as the unit of selection. Genetics **72**: 157–168.

WILLS, C., J. CRENSHAW and J. VITALE, 1970 A computer model allowing maintenance of large amounts of genetic variability in Mendelian population. I. Assumptions and results for large populations. Genetics **64**: 107–123.

YAMAZAKI, T., 1973 Computer uses in Biology (in Japanese). The Heredity (Iden) **27**: 102–111.

YAMAZAKI, T., 1974 Organization of linked genes under frequency-dependent selection of minority advantage. Japan J. Genet. **49**: 33–36.

ZOUROS, E. and C. B. KRIMBAS, 1973 Evidence for linkage disequilibrium maintained by selection in two natural populations of *Drosophila subobscura*. Genetics **73**: 659–674.

Corresponding editor: R. C. LEWONTIN

Corrigendum

Volume 85, January 1977

In FRANKHAM, R., 1977 The nature of quantitative genetic variation in Drosophila. III. Mechanism of dosage compensation for sex-linked abdominal bristle polygenes. Genetics **85**: 185–191.

On page 186 line 16, the word "half" should read "quarter." On page 191 line 17, the word "Biometry" should read "Biometrics."