AVERAGE RECOMBINATION PER CHROMOSOME¹

W. D. HANSON

Department of Genetics, North Carolina State College, Raleigh, North Carolina

Received November 6, 1961

THE problem to be examined can be simply stated: Given an individual heterozygous for n loci on a genetic linkage map, what statements can be made following a mating procedure concerning the probability that any two nonalleles of the gametic array will be in the same linkage phase as in the initial generation? The probability that any two nonalleles are not in the same linkage phase is the average recombination between loci for a mating system. Linkage intensity between two loci is based on the frequency of recombination per meiotic division. Recombination frequency for a mating system would be the probability that two nonalleles are not in an initial linkage phase following meiotic divisions for a mating system.

The conditions set for the development yield information with application to specific types of genetic material, such as in the progeny of the r th generation of mating (selfing or intermating) following the cross between two homozygous parents or in the progeny of the r th generation of mating following a cross between two populations. In the latter case, one would be concerned with the probability that two loci from population 1, say a cornbelt variety of corn, or from population 2, say an exotic corn variety from Latin America, will occur in a gamete. Linkage effects on population dynamics have been considered by GEIRINGER (1944) and can be expressed in terms of linkage disequilibrium (D). For a two-locus problem the relation between initial disequilibrium value (D_0) and that in the r th generation of random mating (D_r) involves only the linkage intensity between the two loci. For n linked loci, disequilibrium values would involve $n(n-1)/2$ linkage values. In this context, the average linkage intensity will have applicability. One assumes in the development that evolutionary pressures such as selection are absent.

GRIFFING (1960) has examined the average linkage intensity by establishing certain limits within which it should fall. The approach considered in this paper considers average recombination per chromosome in detail. The information contributes to our concepts of linkage in biological systems.

THEORETICAL DEVELOPMENT

When the products of one meiotic division are considered, the treatment of interference is difficult. The development of FISHER, LYON and OWENS (1947)

¹ Published as Journal Paper No. 1385 of the North Carolina Agricultural Experiment Station and supported in part by a grant from the Rockefeller Foundation.

related map distance to observable recombination. OWEN (1950) and PAYNE (1956) described the problem in terms of chi-square distributions. HANSON (1959a,b,c) described the distribution of linkage block lengths by first considering a partition of chromosome types in a gametic array with reference to the number of points of genetic recombinations (x). The approach offered some advantages. Firstly, it permitted a description of sets of chromosomes with a probability measure, P_x , for chromosomes with x points of recombination. All sets would have a probability measure however small this measure might be. The biological description for linkage dictates a concept of chromosome transmission. Thus, the mathematical description encompassed the biological description for linkage. Secondly, the introduction of mating systems involves the description of P_x only, and a within set evaluation could be made independent of mating system. However, a solution required an assumption of no interference.

HANSON (1959a) considered in detail the limitations of the assumption of no interference. The description of the parameter, average segment length per chromosome, was not essentially affected by the assumption. The complete description of the product of one meiotic division was limited; however, interference between two loci on different chromosome arms is absent and the number of points of recombination per chromosome per meiotic division is limited, $E[x] = \text{total map length}$. Certainly, the assumption of no interference is reasonable when the products of two or more meiotic divisions in a mating system are considered.

The probability that any two nonalleles on a chromosome are in an initial linkage phase was determined within the described set structure. Following a mating procedure, the homologues of the gametic array which have x points of effective recombination (recombinations involving heterozygous loci, HANSON 1959c) are identified as a set, A_x , and $P(A_x) = P_x$. The probability that two nonalleles are in an initial linkage phase can be evaluated for each set.

Gene distribution within a genetic map: Information on the number of genes affecting a complex quantitative character and their distribution within a genetic map is extremely limited. As organisms are complex, so must be the genetic mechanisms controlling the phenotype of an organism. One can, in fact, argue that the majority of genes for organisms in some manner affect the phenotypic expression of a character as complex as grain yield. Certainly, the number of loci affecting a character such as grain yield must be large. For this development the assumption is made that the loci controlling a character are uniformly distributed with respect to the genetic map scale. This assumption is made for algebraic expediency but is not restricting if the number of loci is large. Evidence exists (LINDEGREN 1939) that points of recombination may have localized areas on the chromosome, and thus, loci may tend to be clustered with respect to the genetic map scale. It appears intuitively correct to say that such clustering will not unduly affect statements of average recombinations, since linked clusters within a linkage block are compensated by linked clusters between linkage blocks.

Genetic map distance for a segment is measured as the expected number of points of recombination per chromatid (FISHER 1948; HANSON 1959a). Thus, a

chromosome with genetic map length of $s = 1.0$ would have, as an average, one point of recombination per chromosome per gamete.

Consider n loci equally spaced with respect to the genetic map scale with a map distance s between the two terminal loci. The n loci divide the genetic map into $N = n - 1$ segments of length s/N , and it would be equally probable that a recombination would occur in any one of the N segments. Further, the assumption is made that two or more points of recombination occur independently with reference to the map scale (no interference). Ramifications for this assumption have been discussed. When the products of one meiotic division are considered, the assumption of independence undoubtedly introduces error; however, when the results of two or more meiotic divisions are considered, the assumption of no interference is realistic.

The probability ($P[B_i, B_j | x]$) that two nonalleles (B_i, B_j) on a chromosome map are in the initial linkage phase, given the set of chromosomes which contain x points of genetic recombination: For any pattern of points of recombination, there are $(N + 1)N/2$ possible pairs of nonalleles for n loci on a chromosome map. Consider that set of homologous chromosomes from a gametic array containing x points of genetic recombination (A_x). For x points of genetic recombination there are N^x possible patterns with respect to the $(N + 1)$ loci. The function, $\phi = \left(\sum_{i=1}^N \right) I_i^x$, will generate these patterns. For example, the term of the

multinomial, $\left(\frac{x!}{a!b!c!\dots} \right) (I_1^a I_2^b I_3^c \dots)$, $a + b + c + \dots = x$, gives the number of patterns with a points of recombination in segment 1, b in 2, etc. Let a_y be the number of pairs of nonalleles in the initial linkage phase. As an example, consider the case where $x = 4$ and $n = 7$ and the partition, $\left(\frac{4!}{1!0!2!0!0!1!} \right) (I_1^1 I_2^0 I_3^2 I_4^0 I_5^0 I_6^1)$. In terms of blocks of loci in the initial linkage phase, one has $(1, 2, 0, 3, 1)$. Thus, the number of pairs of loci in the initial linkage phase (a_y) would be $[(2 \times 1/2 + 3 \times 2/2) + (1 \times 1 + 2 \times 3)] = 11$ of a possible $7 \times 6/2$ pairs.

Each partition has a value, a_y , which can be generalized:

$$a_y = [2\gamma^2 - 2\gamma(N + 3) + (N^2 + 3N + 4)]/2$$

for x even, and

$$a_y = [2\gamma^2 - 2\gamma(N + 1) + N(N + 1)]/2$$

for x odd, $1 \leq \gamma \leq N$. The problem is to generalize the frequency of a_y for a set A_x .

The solution is algebraically tedious. As an example, consider A_3 (the set with three points of recombination per chromosome) and $N = 5$ segments (that is, 6 loci). Following the description given, the enumeration follows:

$$1(1, 0, 0, 5) = a_1$$

$$3(1, 0, 1, 4) = 3a_2$$

$$3(1, 0, 4, 1) = 3a_5$$

$$3(1, 1, 0, 4) = 3a_1$$

$$6(1, 1, 1, 3) = 6a_2$$

.

.

.

$$6(1, 1, 3, 1) = 6a_4$$

$$3(1, 2, 0, 3) = 3a_1$$

etc.

The generalization for $x = 3$ becomes:

$$P[B_i, B_j | 3] = \frac{2}{(N+1)N^4} \sum_1^N a_y [-6y^2 + 6y(N+1) - (3N+2)], \quad (\text{ii})$$

where a_y is defined in (i) for x odd. The solutions for $P[B_i, B_j | x]$ were made for $x = 1, 2, 3, 4$, and 5 and are given in Table 1. A generalization for x is not apparent from these formulas.

TABLE 1
Formulas for calculating $P[B_i, B_j | x]$

x	N large
x even:	
0	1
2	$[2N/3 + 1/3 + 1/3N + 2/3N^2]/(N+1)$
4	$[3N/5 + 1/5 + 2/3N + 4/3N^2 - 4/15N^3 - 8/15N^4]/(N+1)$
.	
.	
x	$(x+2)/2(x+1)$
x odd:	
1	$[2N/3 - 2/3N]/(N+1)$
3	$[3N/5 - 1/3N - 4/15N^3]/(N+1)$
5	$[4N/7 - 8/6N^3 + 16/21N^5]/(N+1)$
.	
.	
x	$(x+3)/2(x+2)$

The algebraic manipulations required to obtain a solution for $P[B_i, B_j | x]$, $x > 5$, become extremely complex; however, certain limits can be established. For N large the dominant terms involve patterns where x points of recombination occur in different segments and where the coefficient for the partition is $x!$. The solutions for N large are:

$$P[B_i, B_j | x] \doteq 2(2!) \sum_1^N a_y (N-y)/N^4 \text{ for } x=2,$$

$$\doteq 2(3!) \sum_1^N a_y (N-y)y/N^5 \text{ for } x=3,$$

$$\begin{aligned} &\doteq 2(4!) \sum_1^N a_y (N - \gamma)^2 \gamma / 2N^6 \text{ for } x = 4, \\ &\doteq 2(5!) \sum_1^N a_y (N - \gamma)^2 \gamma^2 / 2^2 N^7 \text{ for } x = 5, \\ &\doteq 2(6!) \sum_1^N a_y (N - \gamma)^3 \gamma^2 / 2^2 3 N^8 \text{ for } x = 6, \end{aligned} \quad (\text{iii})$$

etc. The value for a_y follows from (i) for N large. The solutions for (iii) yield the limits in Table 1 for N large and permit a generalization for x by induction. Further, for two loci ($N = 1$), $P[B_i, B_j | x]$ is zero if x is odd and one if even. For three loci ($N = 2$), the probability that one of the two segments will have an odd number of recombinations equals the probability that it will have an even number of recombinations, and $P[B_i, B_j | x]$ is one third if x is odd and two thirds if even, for $x > 0$. By extension the $P[B_i, B_j | x]$ can be determined for x large and $N > 2$. These values together with those calculated from the formulas in Table 1 are given in Table 2 for selected values of N . When $N = 1$, $N = 2$, and $N \rightarrow \infty$ are substituted into the formulas in Table 1, the values for $P[B_i, B_j | x]$ correspond to those obtained by the alternative approach. For $N > 2$, the line, x large, in Table 2, gives the limits for the probability for values of $x > 5$. For example, if $N = 3$, $P[B_i, B_j | x] \rightarrow .583$ for large values of x which are even.

Since certain limiting values are known, the estimated values can be approximated graphically with reasonable accuracy. The values bracketed in Table 2 were estimated graphically. Further, these values are approximately .5. The

TABLE 2
 $P[B_i, B_j | x]$ for selected values of x and N calculated from formulas in Table 1
or estimated graphically

Recombinations per chromosome	Segments per chromosome (N)										
(x)	1	2	3	4	6	8	10	15	20	50	Large
0	1	1	1	1	1	1	1	1	1	1	1
2	1	.667	.630	.625	.630	.635	.640	.647	.652	.660	.667
4	1	.667	.588	.569	.564	.567	.571	.578	.583	.592	.600
6	1	.667	.584	.555	.543	.538	.54	.54	.55	.56	.571
8	1	.667	.583	.551	.532	.525	.53	.53	.53	.54	.556
10	1	.667	.583	.550	.528	.519	.52	.52	.52	.52	.545
12	1	.667	.583	.550	.524	.515	.51	.51	.52	.52	.538
Large	1	.667	.583	.550	.524	.514	.51	.51	.50	.50	.500
1	0	.333	.444	.500	.555	.583	.600	.622	.633	.653	.667
3	0	.333	.420	.462	.506	.529	.542	.561	.571	.588	.600
5	0	.333	.417	.453	.489	.508	.519	.536	.536	.560	.571
7	0	.333	.417	.451	.482	.499	.51	.52	.53	.54	.556
9	0	.333	.417	.450	.479	.495	.50	.51	.52	.53	.545
11	0	.333	.417	.450	.477	.492	.50	.50	.51	.53	.538
13	0	.333	.417	.450	.476	.490	.50	.50	.51	.52	.533
Large	0	.333	.417	.450	.476	.486	.49	.49	.50	.50	.500

cases for x and N are given in considerable detail in Table 2 since $P[B_i, B_j | x]$ does not depend upon mating system or genetic map length. Rather, the distribution for x depends upon these two factors.

The probability that any two nonalleles on a chromosome map are in an initial linkage phase is

$$P[B_i, B_j] = \sum_{x=1}^{\infty} P_x P[B_i, B_j | x]. \quad (\text{iv})$$

HANSON (1959c) derived P_x as

$$P_x = e^{-s'} (s')^x / x!$$

where s' was defined as equivalent map length which involves the genetic map length of the chromosome (s) and a parameter due to mating system. The term, equivalent map length, was used since s' would be equivalent to a map length which would generate the distribution of points of genetic recombination in one meiotic division. Let us consider s' for a few cases which are of interest: (a) For a gametic distribution from an F_1 individual $s' = s$, and average recombination value per chromosome is the average linkage intensity between loci on a genetic map of length s , (b) for the r th generation gametes from an intermated population of say exotic *vs.* adapted corn, $s' = s(2 + r)/2$, where linkage refers to the exotic or adapted genes following r generations of intermating, and (c) for the gametes from a population of inbred lines from an F_1 , $s' = s(1 + \frac{1}{2} + \frac{1}{4} + \dots) = 2s$ and average recombination between loci is considered with reference to homozygous lines. Given an s' and the conditional probabilities in Table 2, one can calculate $P[B_i, B_j]$ for selected values of N .

Average recombination ($\bar{p}$) per chromosome: Since $P[B_i, B_j]$ is the probability that any two nonalleles on a chromosome are in an initial linkage phase, then the average recombination per chromosome for a mating system is

$$\bar{p} = 1 - P[B_i, B_j],$$

noting that $P[B_i, B_j]$ depends upon a selected mating system. The average recombination for a mating system will depend upon the gene density per chromosome, the genetic map length between terminal loci, and the mating system. Let us first consider a range of segment numbers per chromosome map (N) and

TABLE 3

*Average recombination per chromosome for selected number of segments (N)
per chromosome and effective genetic map lengths (s')*

Genetic map length s'	Segments per chromosome (N)								
	1	2	4	8	10	15	20	50	Large
.5	.32	.24	.19	.16	.16	.15	.14	.14	.13
1.0	.43	.36	.29	.26	.25	.24	.23	.22	.22
1.5	.48	.42	.36	.32	.31	.30	.29	.28	.27
2.0	.49	.45	.40	.37	.35	.34	.33	.32	.31
3.0	.50	.48	.45	.41	.40	.39	.38	.37	.36
4.0	.50	.49	.47	.44	.43	.42	.41	.40	.39

effective genetic map length (s') (Table 3). As the number of loci per chromosome ($N + 1$) is increased, the average recombination becomes essentially constant. In fact, if one assumes at least ten loci per chromosome map for a character under consideration, loci density does not essentially modify one's concept of average linkage per chromosome. The critical feature in establishing the limits of average recombination per chromosome is the range of genetic map lengths (s) expected for a species. Tabulations are given (Table 3) for s' as large as 4.0. Additional tabulations for $s' > 4.0$ would not contribute to one's concept of average recombination frequency since, on the average, the loci can be considered essentially independent.

What is a typical map length for a species? This information is extremely limited. In corn (RHOADES 1955) the shortest map length could be approximately .5 while the longest map length would be at least, say, 1.6–1.7 with an average of 1.0 or greater. Average linkage per chromosome in the gametes of an F_1 individual could be as low as .15 or as high as .30 with an average linkage intensity of about .22–.25 for a species such as corn ($s' = s$). In a random set of homozygous lines ($s' = 2s$) average recombination per chromosome would range from about .23–.38 with an average recombination of about (.31–.34). If an exotic and an adapted corn varieties are crossed and intermated for r generations ($s' = s(2 + r)/2$), the average recombination with reference to exotic and adapted linkage phase should be at least .36 or near .4 after four cycles of intermating, providing the recombination frequency for exotic-adapted hybrids is the same as that found with normal homologies.

DISCUSSION

The restrictions imposed for the description of average recombination per chromosome were that the n loci were equally spaced on the genetic map and that the events (recombinations) were independent. As the number of loci ($n = N + 1$) becomes large, the assumption of equal spacing on the chromosome map imposes trivial limitations. The assumption of independence could yield modified recombination values when the products of one meiotic division are considered; however, the assumption is not limiting where the products of two or more divisions are considered. Average linkage intensity for a genetic map would involve average recombination between loci per meiotic division. Average recombination for a mating system would be the probability that any two nonalleles were not in an initial linkage phase following the meiotic divisions for the mating system.

A relatively high average recombination per chromosome does not contradict the observations of HANSON (1959a,b,c) on average linkage block length per chromosome. Genes within a linkage block or between linkage blocks removed by even integers have a perfect positive correlation with respect to the initial linkage phase; however, genes between linkage blocks removed by odd integers have a perfect negative correlation with respect to initial linkage phase.

Average recombination per chromosome may be more of academic than practical interest since the fraction of intrachromosome pairs of nonalleles is of the order $1/m$ of the total possible pairs when the number of loci per chromosome is

large. $2m$ is the chromosome number of the species. For a species such as corn ($2m = 20$), one would not expect the average recombination for the genotype to be less than .45. Observations of this nature have been made by GRIFFING (1960). Except for species where the loci are distributed among a limited number of chromosomes, average recombination per chromosome would have limited value when considering the joint effects of all loci affecting a character. It has value in describing the uniqueness of linkage.

The recombination values for N large (Table 3) would correspond to the case considered by GRIFFING. To compare these values with the ranges bracketed by his formulation requires first a translation of scales. GRIFFING assumes that a chiasmata results from a point of recombination and considers chiasma per chromatid. The expected chiasma per segment is the genetic map length (s) as used in this paper. Taking $m = 1$ yields:

GRIFFING	$\bar{p}$	
$.24 < \bar{p} < .33$	.31	for $s = 2.0$,
$.20 < \bar{p} < .30$	.27	for $s = 1.5$, and
$.10 < \bar{p} < .26$	.23	for $s = 1.1$

Although GRIFFING's approach yields a rough approximation for average recombination per chromosome, it would be adequate when considering average recombination for a genotype where m becomes the dominant feature in the formulation.

Two points should be raised on the interpretation of chiasmata frequency. Firstly, a chiasmata is a cytological observation based on spore mother cells in a meiotic division. If one assumes that a chiasmata results from a genetic recombination, then the chiasmata frequency *per spore mother cell* is twice the genetic map length, since the observations are made at the four strand stage, and the chiasmata values considered by GRIFFING must be doubled to make them coincide with cytological observations. Secondly, the correlation between chiasmata frequency and genetic map length does not necessarily infer that chiasmata frequency per spore mother cell is twice the genetic map length. The assumption of equivalency for chiasmata and recombination frequencies is subject to criticism.

When one considers that the average recombination per chromosome would likely be about .25 (or larger if the mating system permitted additional generations for recombination) and that loci between chromosomes segregate independently, he certainly would question whether linkage biases could be detected through variance estimates. That it is present in the estimates of genetic covariances, however, has been clearly defined (SCHNELL 1961). Conditions exist where linkage biases could be estimated by variance techniques. Although one would normally assume that the joint expression of two (or more) nonallelic genes would be independent of the position of the loci in the genome, coadaptive balance of genes within the chromosome could have evolved through selection, especially in self-pollinated species. Further, a population in extreme disequilibrium should show linkage bias in covariance estimates, especially when the chromosome number of the species is small or the number of loci affecting the character is small. That linkage biases may be difficult to demonstrate with second

degree statistics does not infer that linkage is not important in quantitative genetics. For example, if one makes a cross between two populations and intermates the material for four generations, then the probability that any two non-alleles are not in an initial linkage phase with respect to the two populations is at least .36 for a map length of $s = 1.0$ (Table 3, $s' = 3.0$). However, the average segment length per chromosome (HANSON 1959c) is approximately .32. Linked blocks of genes and difficulties in identifying the superior recombinant types represent a critical block in breeding programs.

SUMMARY

Average recombination per chromosome was formulated and evaluated for different gene densities per chromosome and chromosome map lengths. With at least ten loci per genetic map length, loci density could essentially be ignored when describing average recombination per chromosome. Average recombination per chromosome depended primarily upon the genetic map length between the terminal loci or the equivalent map length when considering recombination following a mating system. For typical chromosome lengths, the average linkage intensity per chromosome should be approximately .25. With intermating, average recombination per chromosome for the mating system approaches that for independence even though the average linkage block length per chromosome may be relatively large.

LITERATURE CITED

- FISHER, R. A., 1948 A quantitative theory of genetic recombination and chiasma formation. *Biometrics* **4**: 1-13.
- FISHER, R. A., M. F. LYON, and A. R. G. OWENS, 1947 The sex chromosome in the mouse. *Heredity* **1**: 355-365.
- GEIRINGER, H., 1944 On the probability theory of linkage in Mendelian heredity. *Ann. Math. Stat.* **15**: 25-57.
- GRIFFING, B., 1960 Accommodation of linkage in mass selection theory. *Australian J. Biol. Sci.* **13**: 501-526.
- HANSON, W. D., 1959a The theoretical distribution of lengths of parental gene blocks in the gametes of an F_1 individual. *Genetics* **44**: 197-209.
- 1959b Theoretical distribution of the initial linkage block lengths intact in the gametes of a population intermated for n generations. *Genetics* **44**: 839-846.
- 1959c The breakup of initial linkage blocks under selected mating systems. *Genetics* **44**: 857-868.
- LINDEGREN, CARL C., and GERTRUDE LINDEGREN, 1939 Nonrandom crossing-over in the second chromosome of *Neurospora crassa*. *Genetics* **24**: 1-7.
- OWEN, A. R. G., 1950 The theory of genetical recombination. *Advances in Genet.* **3**: 117-157.
- PAYNE, L. C., 1956 The theory of genetical recombination: a general formulation for a certain class of intercept length distributions appropriate to the discussion of multiple linkage. *Proc. Roy. Soc. London B* **144**: 528-544.
- RHOADES, M. M., 1955 The cytogenetics of maize. Chapter four. *Corn and Corn Improvement*. Academic Press, Inc. New York, New York.
- SCHNELL, F. W., 1961 Some general formulations of linkage effects in inbreeding. *Genetics* **46**: 947-957.