EARLY GENERATION ANALYSIS OF LENGTHS OF HETEROZYGOUS CHROMOSOME SEGMENTS AROUND A LOCUS HELD HETEROZYGOUS WITH BACKCROSSING OR SELFING

W. D. HANSON

*Crops Research Division, Agricultural Research Service,
U.S. Department of Agriculture, Beltsville, Maryland*

Received February 2, 1959

WHEN a locus is held heterozygous for n generations of backcrossing, the average length of heterozygous segments surrounding the selected locus asymptotically approaches $2/n$ (FISHER 1949). With the equivalence of eight to ten generations of backcrossing this criterion is adequate. Experiments which involve early generation tests with genetically similar or "isogenic lines", chromosome substitution studies, factor transfer or the development of genetic stocks require a more exact criterion for heterozygous segment lengths for the early generations of inbreeding. The objective of this paper is to develop a distribution of the segment lengths heterozygous about a locus held heterozygous for any number of generations of backcrossing or selfing.

Derivation of expected length and variance of lengths heterozygous after selected generations of inbreeding

Consider a chromosome of length s measured with reference to the true genetic map scale (FISHER 1948) and the conditions (i) that crossovers occur independently and (ii) that the locus Aa which is being held heterozygous is in the center of the chromosome. The restriction (ii) is not essential but simplifies appreciably the derivation. The resulting calculations will serve as an adequate guide for the interpretation of the segment analysis. Let c be any length measured with reference to the true genetic map scale ($0 \leq c \leq s$). Then the transformation, $t = c/s$, reduces the measurement of a segment to a proportion of the chromosome. The length of any chromosome is 1.0 and the locus of Aa is 0.5, with reference to the proportionate scale. The length of segment t will be measured only from the left of the Aa locus. Thus, $0 \leq t \leq \frac{1}{2}$, since only the proportion of a segment occurring in this half of the chromosome will be considered. For further discussion, these lengths will be defined as half lengths. Our first problem will be to determine the cumulative distribution for t , $F(t)$, in the chromosomes arising from a meiotic division, considering only half lengths as defined.

The true genetic map scale is the expected number of breaks per unit length. Chromosomes resulting from a meiotic division may contain $0, 1, \dots, x, \dots$ breaks or points of genetic recombination. The distribution of x is defined by some discrete distribution. With the assumption of independence the probability P_x , that a chromosome will contain x breaks may be defined by the Poisson dis-

tribution, $P_x = (e^{-s}s^x/x!)$, where $s = E[x]$ is the length of the chromosome (HANSON 1959). The probability density for the x breaks within a chromosome will be uniform.

The derivation of $F(t)$ can best be illustrated by considering specific cases. Consider as a set, A_3 , all possible chromosomes which involve three breaks. The probability that any chromosome will occur in this set can be shown to be P_3 . Each chromosome of this set will contain four segments identified by sequence, 0 to γ_1 , γ_1 to γ_2 , γ_2 to γ_3 , and γ_3 to 1.0, where γ_i represents the locus of a break. These segments are identified by subsets, A_{31} , A_{32} , A_{33} , and A_{34} , respectively. A segment in subset A_{31} will contain the Aa locus only if $1/2 < \gamma_1 < \gamma_2 < \gamma_3 < 1.0$, which constitutes $1/8$ of the elements in the subset A_{31} . However, the length of t is $1/2$ for all elements of this restricted subset since only half-lengths are considered. A segment in subset A_{32} will contain the Aa locus only if $0 < \gamma_1 < 1/2$, $1/2 < \gamma_2 < \gamma_3 < 1.0$ which constitutes $3/8$ of the elements of subset A_{32} . The probability that a randomly selected half length in this restricted sample will be less than or equal to t can be shown to be $1 - 2(1/2 - t)$. Similarly, it can be shown that $3/8$ of the elements in subset A_{33} will contain the Aa locus with a corresponding probability, $1 - 2^2(1/2 - t)^2$. The corresponding values for A_{34} are $1/8$ and $1 - 2^3(1/2 - t)^3$.

Considering all possible subsets, one can construct the following recurrent relationship for the cumulative distribution of half lengths (t):

$$F(t) = \sum_{x=1}^{\infty} \sum_{r=2}^{x+1} (1/2)^x P_x \binom{x}{r-1} [1 - 2^{r-1}(1/2 - t)^{r-1}], \quad 0 \leq t < 1/2,$$

$$Pr(t = 1/2) = \sum_{x=0}^{\infty} (1/2)^x P_x.$$

The distributions can be simplified as follows:

$$F(t) = 1 - \sum_{x=0}^{\infty} P_x (1/2)^x (1 + 2Z)^x, \quad 0 \leq t < 1/2$$

$$Pr(t = 1/2) = \sum_{x=0}^{\infty} P_x (1/2)^x,$$

where $Z = (1/2 - t)$. With the assumption of independence, P_x can be defined by the Poisson distribution, and the following simplified expressions will result:

$$F(t) = 1 - e^{-s(1/2 - Z)}, \quad 0 \leq t < 1/2,$$

$$Pr(t = 1/2) = e^{-s/2}.$$

$F(t)$ is discontinuous at $t = 1/2$. $F(t)$ represents the cumulative distribution of half lengths, as measured from the locus of Aa (0.5), in the chromatids generated by a meiotic division. The probability that a crossover will not occur within a region from the Aa locus is $1 - F(t)$, t being the length of the segment.

The procedure for determining the cumulative distribution, $G(t_n)$, for half lengths which are heterozygous after n generations of backcrossing follows the procedure given by FISHER (1949, p. 50),

$$G(t_n) = 1 - (1 - F(t))^n, \quad 0 \leq t < \frac{1}{2},$$

which reduces to

$$G(t_n) = 1 - e^{-snt}, \quad 0 \leq t < \frac{1}{2}.$$

The expected value of t^m after n generations of backcrossing is:

$$E[t^m] = \int_0^{\frac{1}{2}} t^m dG(t_n) + (\frac{1}{2})^m (e^{-ns/2}),$$

noting that $F(t)$ is discontinuous at $t = 1/2$. Since the exact cumulative distribution of $F(t)$ is known and the limits of integration are defined, the segment lengths can be evaluated for all generations of backcrossing. Since t represents a proportion of a chromosome, the expectations can be expressed in terms of the genetic map units by the conversion $s^m E[t^m]$.

It should be noted that for selfing, $n = 2n' - 2$, where n' represents the filial generation, since crossing over in either male or female gametes will effectively reduce the length of heterozygous segment around the Aa locus. The expected half lengths together with the variance of half lengths for chromosome of lengths $1/2$, 1.0 , and 2.0 are given in Table 1. As n becomes large, the expected length and variance expressed in genetic map units for half lengths which are heterozygous approach $1/n$ and $1/n^2$, respectively. The entries in Table 1 must be doubled to include the lengths on both sides of the Aa locus. The initial increase in $E[V(c)]$ is attributed to the high proportion of the long lengths, $t = 1/2$, in the original chromatid array.

TABLE 1

The expected half length (centimorgans/100) and variance of heterozygous chromosome segments associated with a locus held heterozygous in backcrossing or selfing for selected chromosome lengths and generations of inbreeding

Length chromosome	Stat.*	Generation									
		BC ₁	BC ₂ F ₂	BC ₃	BC ₄ F ₃	BC ₅	BC ₆ F ₄	BC ₈ F ₅	BC ₁₀ F ₆	BC ₁₂ F ₇	BC ₂₀ F ₁₁
.5	$E[c]$	.221	.197	.176	.158	.143	.129	.108	.092	.079	.050
	$E[V(c)]$	.0041	.0064	.0076	.0081	.0081	.0078	.0068	.0058	.0048	.0023
1.0	$E[c]$	.393	.316	.259	.216	.184	.158	.123	.099	.083	.050
	$E[V(c)]$	.0256	.0322	.0312	.0275	.0233	.0194	.0133	.0093	.0067	.0025
2.0	$E[c]$	.632	.432	.317	.245	.199	.166	.125	.100	.083	.050
	$E[V(c)]$	.1289	.1101	.0776	.0533	.0373	.0270	.0155	.0100	.0069	.0025
Limiting Function	$E[c]$	1.000	.500	.333	.250	.200	.167	.125	.100	.083	.050
	$E[V(c)]$	1.0000	.2500	.1111	.0625	.0400	.0278	.0156	.0100	.0069	.0025

* $E[c] = sE[t] = \frac{1}{n} [1 - e^{-sn/2}]$ (centimorgans/100).
 $E[c^2] = s^2 E[t^2] = \frac{1}{n^2} [2 - (sn + 2)e^{-sn/2}]$ (centimorgans/100)².

DISCUSSION

The exact distribution for the heterozygous half lengths measured from a gene pair held heterozygous during a backcrossing or selfing program has been presented. Half lengths have been defined as lengths measured in only one direction from the fixed locus. The locus of the factor pair was selected as the center of the genetic map, and the derivations were made on the assumption of no interference. The results are applicable to the early generations of backcrossing or selfing but are limited by the two restrictions stated. As the number of generations of backcrossing or selfing becomes large, the limitations imposed by the two restrictions are inconsequential. The statistics given in Table 1 must be doubled to define the distribution of lengths surrounding the fixed locus which is still heterozygous after selected levels of inbreeding.

The relationships developed by FISHER (1949) and by BARTLETT and HALDANE (1935) are limiting functions. The limiting functions for the expected half lengths ($1/n$) and the variance of half lengths ($1/n^2$) have been included in Table 1 for comparative purposes. As might have been expected, the shorter the genetic map length for a chromosome the greater will be the error in using the limiting values to estimate the mean and variance of half lengths in the early generations of inbreeding. For a chromosome of map length 1.0, which could represent an average genetic map length of a chromosome, the limiting values appear adequate after the equivalence of about eight generations of backcrossing.

The statistics given in Table 1 represent the expected mean and variance of the half lengths occurring in independent lines where the locus has been held heterozygous through selected generations of backcrossing or inbreeding. It is evident that the heterozygous chromosome segments after eight to ten generations of backcrossing are of considerable length. This information can be used in the designing of experiments to study the association of quantitative or qualitative characters with the fixed locus (Aa). Further, since in any one line considerable sampling variation may occur in heterozygous segment lengths associated with the Aa locus, the variances given in Table 1 can be used to determine the number of independent lines which must be carried to the selected level of inbreeding so that a desired reproducibility is obtained in the segment surrounding the fixed locus. Table 1 can thus be used as a guide to determine the adequacy of sampling in the early generation testing of "isogenic lines".

SUMMARY

The cumulative distribution for the heterozygous half lengths from a gene pair held heterozygous during a backcrossing or a selfing program was developed. Since the development involved the cumulative distribution for heterozygous half lengths following a meiotic division, the average heterozygous segment length and the expected variance for heterozygous segment lengths were available for the early generations of backcrossing or selfing. The statistics were evaluated for selected chromosome map lengths, and the restrictions required for

the development were discussed. The information presented would be pertinent for the designing of experiments involving early generation testing of genetically similar or "isogenic lines".

LITERATURE CITED

- BARTLETT, M. S., and J. B. S. HALDANE, 1935 The theory of inbreeding with forced heterozygosis. *J. Genet.* **31**: 327-340.
- FISHER, R. A., 1948 A quantitative theory of genetic recombination and chiasma formation. *Biometrics* **4**: 1-13.
- 1949 *The Theory of Inbreeding*. HAFNER PUB. CO. INC. New York, N.Y.
- HANSON, W. D., 1959 The theoretical distribution of lengths of parental gene blocks in the gametes of an F_1 individual. *Genetics* **44**: 197-209.