The Determination of the Arrangement of Genes From Tetrad Data

Ernest E. Shult and Carl C. Lindegren

Biological Research Laboratory, Southern Illinois University, Carbondale, Illinois, U.S.A.

Received March, 1, 1955

The procedure for establishing the linear order of three genes on a single chromosome was first enunciated by Sturtevant (1913): If the linkage of A to C is the sum of the linkages of A to B and B to C, then B is between A and C, but if the linkage of A to C is the difference between the linkages of A to B and B to C, then C is between A and B.

In tetrad analysis, another element enters the calculation since distance from the centromere can be determined even when considering only unordered tetradis (Lindegren, 1949).

For example, if one considers three genes linearly arranged as in Fig. 1, any two gene-pairs may segregate to form one of the three possible types of unordered ascis: viz., type I, AB, AB, ab, ab; type II, Ab, Ab, aB, aB; and type III, AB, Ab, aB, ab. If A/a and B/b segregate as type I, then the distal

---

1 This work has been supported by research grants from the U.S. Atomic Energy Commission Contract AT (11-1)-176 and the National Cancer Institute of the National Institutes of Health, Public Health Service C-1179.
markers, A/a and C/c will segregate to form the same type as B/b and C/c, since each B is representative of an A (Fig. 1, a, b, c, d, e, f). If A/a and B/b segregate as type II, and B/b and C/c also segregate as type II, then A/a and C/c will segregate as type I (Fig. 1, g). If A/a and B/b segregate as type II, and B/b and C/c segregate as type III, then A/a and C/c will segregate as type III (Fig. 1, h). If both A/a and B/b and B/b and C/c segregate as type III, then A/a and C/c may segregate as any of the three types (Fig. 1, i, j, k). The latter three cases correspond to 2-, 4-, and 3-strand double recombinations (as indicated by the arrows), and the frequency with which each occurs is denoted P, Q, and R, respectively. The above relationships between the segregation of intermediate and distal genes may be symbolized:

\[
\begin{align*}
I \cdot I &= I \\
I \cdot II &= II \\
I \cdot III &= III \\
II \cdot II &= I \\
II \cdot III &= III \\
III \cdot III &= P + QII + RIII
\end{align*}
\]

where the first numeral on the left represents the segregation of A/a and B/b; the second numeral on the left represents the segregation of B/b and C/c; the multiplication represents the combining of the two segregations. The numeral to which these are equated represents the segregation of the distal markers A/a and C/c. Inasmuch as these segregations may be read from either the right or left, each multiplication is commutative, i.e., \( II \cdot III = III \cdot II \). On the basis of these principles, order, as determined by tetrad analysis, may be defined.

*Three genes are said to be ordered if the segregation of the distal genes is the result of the simultaneous segregation of the intermediate genes.*

The frequency at which the genes A/a and B/b produce types I, II, and III will be denoted \( x_1, x_2 \) and \( x_3 \), respectively. If probability be regarded as the limit of a statistical frequency (H. Feigl, 1934), \( x_1, x_2 \) and \( x_3 \) may be treated as asymptotic approximations to the probability that the genes A/a and B/b form type I, type II, or type III. The probability distribution for these three types is symbolized as \( x_1I + x_2II + x_3III \), with the stipulation that \( x_1 + x_2 + x_3 = 1 \) and that no individual value is either negative or exceeds unity. The probability distribution for the gene combinations B/b, C/c and A/a, C/c are denoted: \( y_1I + y_2II + y_3III \), and \( z_1I + z_2II + z_3III \), respectively. If the formation of the three types for the genes A/a and B/b is independent of formation of types for the genes B/b and C/c, then the statement that the genes are in the order A–B–C (i.e., the segregation of A/a and C/c is the result of the simultaneous segregations of A/a and B/b and B/b and C/c) is logically equivalent to the symbolic statement:

\[
(x_1I + x_2II + x_3III)(y_1I + y_2II + y_3III) = z_1I + z_2II + z_3III.
\]

The following example shows that the multiplication of numerals is distributive. If the genes A/a and B/b segregate as type III in all cases considered and the genes B/b and C/c segregate as type I in 50 per cent of
the cases, and as type II in 50 per cent of the cases; then the segregation of the distal genes A/a and C/c will result in the combination of the segregations III and I (III·I) in half the cases, and from the combination of the segregations III and II (III·II) in the other half. That is, III (0.5I + 0.5II) = 0.5III·I + 0.5III·II = III. The validity of this example does not depend upon the particular numerals used, nor the value of the probabilities, indicating that multiplication is distributive for any such numbers used. For this reason, the product on the left-hand side of equation (1.2) may be expanded, and by use of the multiplication table (1.1), the coefficients of I, II, and III may be equated. This process yields the set of equations:

\[
\begin{align*}
X_1Y_1 + X_2Y_2 + PX_3Y_3 &= z_1 \\
X_1Y_2 + X_2Y_1 + QX_3Y_3 &= z_2 \\
(X_1 + X_2)Y_3 + X_3(Y_1 + Y_2) + RX_3Y_3 &= z_3
\end{align*}
\]

(1.3)

Since tetrad analysis makes it possible to detect centromere linkage and since the distances of B from its centromere and C from its centromere may be added (to give a distance of B to C), we may consider three genes distributed among two chromosomes in the manner illustrated in Fig. 2. The segregation of the genes A/a and C/c is the result of the segregations of the intermediate gene combinations, A/a, B/b and B/b, C/c and the genes are ordered A-B-C. The multiplication table (1.1) derived from the segregational arrangements at the bottom of Fig. 1, are dependent only on this notion of order. Consequently, the ordering principle is applicable to genes distributed among two chromosomes, as well as genes on the same chromosome, and equations (1.3) hold equally well in either case. The equations also hold if there is preferential segregation among the centromeres in Fig. 2, for the only necessary conditions for the validity of these equations are (a) an inter-regional independence in the number of recombinations (regardless of the peculiarities of any single region), and (b) an assigned order.

Three genes of a prescribed order may be arranged in only three ways among the chromosomes. Since there are three orders for three genes (disregarding left and right) this amounts to nine arrangements. There exists one more possible arrangement: three genes, each on a different chromosome. Order cannot be defined for this arrangement. These ten possible arrangements are diagrammed in Fig. 3.
In proceeding to determine the order of three genes, the probability distributions for the formation of the three types are written out for each two of three selected gene-pairs. Each of the three possible orders is tested by equations (1.3). If, for a given order, any of the constants $P$, $Q$, and $R$ is negative or greater than one, then the order is impossible. If only one of the three orders is possible, the order is established. If no order is possible, the occurrence of recombinations in each of the regions is necessarily dependent. Further gene combinations may be used to check the initially established order.

If two of the genes segregate randomly (1/6I + 1/6II + 2/3III, Lindegren and Shult, in press), two of the orders are equally possible, since

\[(1/6I + 1/6II + 2/3III)(x_1I + x_2II + x_3III) = 1/6I + 1/6II + 2/3III\]

for any $x_1$, $x_2$, and $x_3$.

Example:

The segregation of the three gene-pairs $a/\alpha$, AN/an, HI/hi was calculated from data supplied by Mrs. Lindegren:

\[
\begin{align*}
\alpha - AN & : 0.235I + 0.118II + 0.647III \\
\alpha - HI & : 0.263I + 0.000II + 0.737III \\
AN - HI & : 0.706I + 0.059II + 0.235III
\end{align*}
\]

By equations (1.3)

\[
\begin{align*}
0.235 \cdot 0.706 + 0.118 \cdot 0.059 + P \cdot 0.647 \cdot 0.235 = 0.263 \\
P = 0.592
\end{align*}
\]

\[
\begin{align*}
0.235 \cdot 0.059 + 0.118 \cdot 0.706 + Q \cdot 0.647 \cdot 0.235 = 0.000 \\
Q \text{ must be negative}
\end{align*}
\]

Hence, this order is impossible.

By equations (1.3) the order $AN - \alpha - HI$ yield

\[
\begin{align*}
0.235 \cdot 0.263 + 0.000 \cdot 0.118 + P \cdot 0.647 \cdot 0.737 = 0.706 \\
P = 1.35
\end{align*}
\]

This is an impossible ordering. The only remaining order is $AN - HI - \alpha$. 
which gives \( P = .283, \ Q = .590, \) and \( R = .127. \) Hence the order is AN–HI–\( \alpha. \)

A second triplet (AN, HI, MA) is tested: The distribution of the three types for the new gene combinations are

\[
\begin{align*}
AN-MA & \ 0.235I + 0.118II + 0.647III \\
HI-MA & \ 0.190I + 0.190II + 0.619III.
\end{align*}
\]

The order AN–HI–MA, yields a negative value for \( Q. \) The order HI–AN–MA yields \( P = .112, \ Q = .612, \) and \( R = .276. \) This is a satisfactory order. The proposed order AN–MA–HI yields \( P = 1.598 \) and hence is impossible. Thus the only possible order is HI–AN–MA.

The integration of these two orders indicates that \( \alpha \) and MA are on opposite sides of the HI–AN linkage:

\[
MA-AN-HI-\alpha.
\]

It may be predicted, \textit{a priori} of a trial, that the order of (MA, AN, \( \alpha) \) must be MA–AN–\( \alpha, \) and that of (MA, HI, \( \alpha) \) must be MA–HI–\( \alpha. \) These predictions are tested below:

The new gene combination is \( \alpha-MA \ 0.227I + 0.182II + 0.591III. \) For the triplet \( \alpha, \ MA, \ AN, \) all orders are valid (which is to be expected for unlinked genes). The triplet \( \alpha, \ MA, \ HI, \) has two possible orders, MA–HI–\( \alpha \) and MA–\( \alpha–HI. \) It is inferred that the ordering method gives a consistent picture for the genes involved, and that the method may be extended indefinitely to other gene combinations.

The crucial requirement for the validity of the method is that (aside from chromatid interference) the occurrence of recombination in the intermediate regions is independent (i.e. no "chromosomal interference"). This may be tested by accumulating three-point tetrads and observing the number of strands involved in the recombinations (chromatid recombination) and the independence of their occurrence (chromosomal interference). Although it is customary to suppose that chromosomal interference usually occurs, critical analysis has shown that (a) this view has no sound basis (Shult and Lindegren, 1955) and (b) that chromosomal interference is not always demonstrable when adequate data are properly analyzed.

\textbf{References}

Fiegl, H. 1934. The logical character of the principle of induction. Philosophy of Science 1: 20–


Shult, Ernest E. and Lindegren, Carl C. 1955. The hypothesis of chromosomal interference. Nature : 