Genet. Res., Camb. (1963), **4**, *pp.* 416–426
With 9 text-figures
Printed in Great Britain

The interpretation of data from transduction experiments

By G. W. P. DAWSON
Department of Genetics, Trinity College, Dublin

(*Received* 4 *May* 1963)

The development of classical genetics largely depended on the building of algebraic and statistical models to explain the quantitative results of breeding experiments. The greater the precision of the model, the greater its predictive value. Further investigations stem from departures from what is predicted. The genetical study of bacteria by transduction has found so much of interest by qualitative means that the possibilities of quantitative analyses of the breeding data have been relatively unexplored. This is an attempt to interpret such data on the basis of our knowledge of transduction in *Salmonella typhimurium*. Data from a transduction involving two linked factors will be considered first and then data from a transduction in which there are three such factors. The discussion touches briefly on the difficulties of interpreting data from experiments in which a single factor is transduced.

1. TRANSDUCTIONS OF TWO LINKED FACTORS

Suppose that loci A and B are known from transduction studies to be linked then the transduction A^-B^- ($\times$) A^+B^+, with selection for B^+, will yield A^-B^+ and A^+B^+ transductants. Ozeki (1959) has argued that in *Salmonella typhimurium* (but

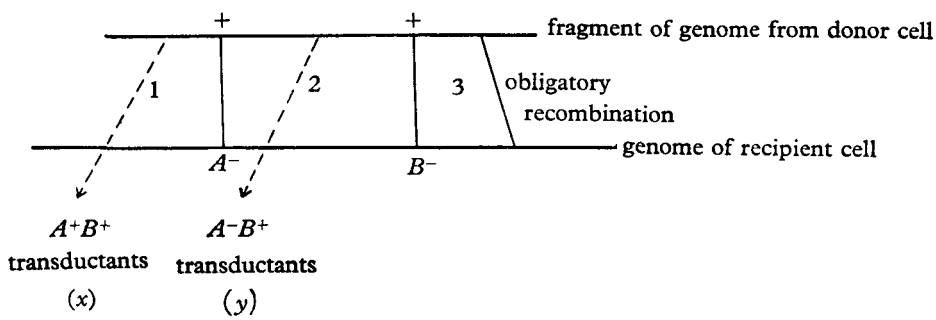


Fig. 1. The origin of A^+B^+ and A^-B^+ transductants from the transduction A^-B^- ($\times$) A^+B^+ with selection for B^+.

not in *Escherichia coli*) all transduced fragments arise by breakage points at constant positions in the genome of the donor bacteria. Thus all fragments that include the B locus include also the A locus and are equal in length. The relevant evidence persuades one strongly to accept this. If the relative frequencies of A^+B^+ and A^-B^+ transductants are x and y (Fig. 1) it has been assumed by Ozeki that x and y

are measures of the relative lengths of regions 1 and 2 in units of recombination frequency. I shall examine this proposition of proportionality assuming that the incorporation of parts of the transduced fragments into the genome of the recipient bacteria is formally similiar to recombination in higher organisms.

Imagine a population of paired fragments in all of which recombination has occurred in region 3. If the chance of recombination in region 1 is α and in region 2 is β the proportion of these paired fragments that yield A^+B^+ transductants is $\alpha(1-\beta)$. Thus

$$\frac{X}{Y} = \frac{\alpha(1-\beta)}{(1-\alpha)\beta}$$

where X and Y are the relative frequencies of A^+B^+ and A^-B^+ transductants in an infinite population of transductants. Our best estimate of X/Y is x/y and so we may write

$$\frac{x}{y} = \frac{a(1-b)}{(1-a)b}$$

where a and b are estimates of α and β.

Although a/b is not directly proportional to x/y we can nevertheless enquire what can be learnt about a/b and x/y. I shall discuss the three situations (1) where $x > y$, (2) where $y > x$, and (3) where $x = y$.

(1) If $x > y$ then $a > b$. If a and b are each numerically small $(1-a)$ and $(1-b)$ each approach 1 and $x/y \to a/b$.

At the other extreme when a and b are numerically large a and $(1-a)$ each approach 0·5 and $x/y \to (1-b)/b$, whence:

$$b = \frac{1}{(x/y)+1}$$

and

$$\frac{a}{b} = 0{\cdot}5\left(\frac{x}{y}+1\right) = \frac{x+y}{2y}$$

We conclude that when x/y is greater than 1 the value of a/b lies between x/y (maximum) and $(x+y)/2y$ (minimum). That there is no necessary expectation that the value of a/b is more likely to be in one region of this span than another can be demonstrated by:

$$\frac{x}{y} = \frac{a(1-b)}{(1-a)b}$$

whence

$$\frac{1}{b} = \frac{x}{ay} - \frac{x}{y} + 1$$

$$\frac{a}{b} = \frac{x}{y} - \left(\frac{x}{y} - 1\right)a$$

Thus the value of a/b is linearly related to the magnitude of a for any particular value of x/y.

https://doi.org/10.1017/S0016672300003803 Published online by Cambridge University Press

(2) If $y > x$ then $b > a$. If a and b are each numerically small $(1-a)$ and $(1-b)$ each approach 1 and $x/y \to a/b$.

At the other extreme when a and b are numerically large b and $(1-b)$ each approach 0·5 and $x/y \to a/(1-a)$, whence:

$$a = \frac{x/y}{1+(x/y)}$$

and

$$\frac{a}{b} = \frac{1}{0{\cdot}5} \times \frac{x}{y}\left(\frac{1}{1+(x/y)}\right) = \frac{2x}{x+y}$$

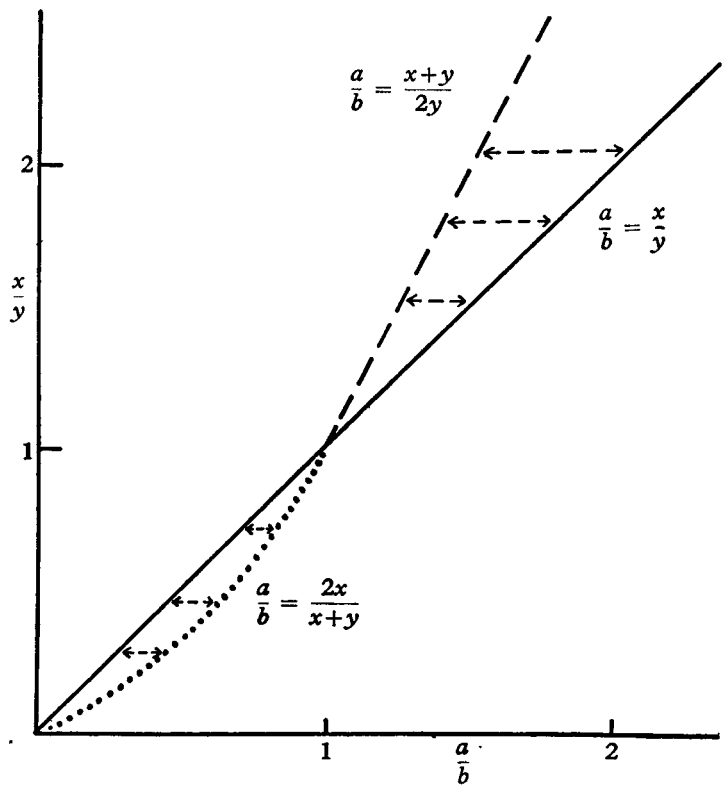


Fig. 2. The maximum and minimum values of a/b (the relative frequencies of recombination in regions 1 and 2 of Fig. 1) that are consistent with x/y (the relative frequencies of the two classes of transductants in Fig. 1).

Thus, when x/y is less than 1 the value of a/b lies between x/y (minimum) and $2x/(x+y)$ (maximum) and, by the demonstration set out in (1) above, there is no necessary expectation that the value of a/b is more likely to lie in one region of this span than another.

(3) If $x = y$ the only possible conclusion is that $a = b$.

These results can be represented graphically by plotting x/y against a/b (Fig. 2). The maximum values of a/b when $x/y > 1$ and the minimum values when $x/y < 1$

https://doi.org/10.1017/S0016672300003803 Published online by Cambridge University Press

describe a straight line $a/b = x/y$. The minimum values of a/b when $x/y > 1$ describe the straight line:

$$\frac{a}{b} = \frac{x+y}{2y}$$

$$\frac{a}{b} = 0{\cdot}5\frac{x}{y} + 0{\cdot}5$$

The maximum values of a/b when $x/y < 1$ describe the curve:

$$\frac{a}{b} = \frac{2x}{x+y}$$

$$\frac{a}{b} = \frac{1}{0{\cdot}5 + 0{\cdot}5(x/y)}$$

In the absence of any further information about the magnitudes of a and b it is possible to deduce from the experimentally determined x/y only a range of possible values of a/b. There are, however, occasions when a and b are known to be of the same order of magnitude in different transduction experiments although the magnitudes themselves are unknown. For example, in the transductions $A_1^- B^-(\times)A^+B^+$ and $A_2^- B^-(\times)A^+B^+$, where A_1^- and A_2^- are known to be closely linked site mutations within the A locus, the values of a and b will be similar in the two crosses. Although the two values of x/y may imply ranges of a/b values which overlap it is valid to conclude that the lower x/y ,providing it is significantly lower, implies a lower a/b.

We have enquired what we can learn of the value of a/b from the transduction $A^-B^-(\times)A^+B^+$ when B^+ transductants are selected. If, instead of selecting for B^+, we select for A^+ transductants we can obtain similar information about the relative frequencies of recombination in regions 2 and 3.

For convenience of description I have assumed that the observed x/y is based on scoring a very large number of transductants. In many transduction experiments, however, only a relatively small number is scored and it is then important to calculate the maximum and minimum values of X/Y with which x/y is consistent. We can state these limiting values as

$$\frac{x+2d}{y-2d} \text{ (maximum) and } \frac{x-2d'}{y+2d'} \text{ (minimum),}$$

where

$$d = \sqrt{\left[\frac{(x+2d)(y-2d)}{x+y}\right]}$$

$$= \frac{(x-y)+\sqrt{[(x-y)^2+(x+y+4)xy]}}{x+y+4}$$

and

$$d' = \sqrt{\left[\frac{(x-2d')(y+2d')}{x+y}\right]}$$

$$= \frac{(y-x)+\sqrt{[(y-x)^2+(x+y+4)xy]}}{x+y+4}$$

https://doi.org/10.1017/S0016672300003803 Published online by Cambridge University Press

In the earlier discussion it is of no account whether the numbers of transductants themselves or their reduction to the simplest possible ratio is used but in these equations for estimating d and d' the total numbers of transductants scored must be used for x and y. When the maximum value of X/Y is calculated in this way there is about one chance in twenty of any sample of $x+y$ transductants showing as great or greater departure from this value than the observed x/y. Similarly for the minimum value. When these values have been calculated they should be used to determine the range of consistent a/b values by using the minimum value of X/Y for calculating the minimum value of a/b and the maximum value of X/Y for calculating the maximum value of a/b.

As an illustration we can analyse the data from the transduction *leu-39 araB*$^+$ (×) *leu*$^+$*araB-9*. Of 405 *leu*$^+$ transductants 145 were *araB*$^+$ and 260 were *araB-9* (Dawson & Smith-Keary, 1960). This transduction is set out in Fig. 3.

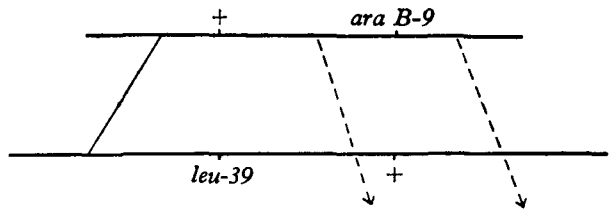


Fig. 3. The origin of 145 *leu*$^+$ *ara*$^+$ and 260 *leu*$^+$ *ara*$^-$ transductants from the transduction *leu-39 araB*$^+$ (×) *leu*$^+$ *araB-9*.

The calculations proceed as follows:

$$\frac{x}{y} = \frac{260}{145}$$

Maximum value of

$$\frac{X}{Y} = \frac{260+2d}{145-2d}$$

where
$$d = \frac{(260-145)+\sqrt{[(260-145)^2+(260+145+4)\times 260\times 145]}}{260+145+4}$$
$$= 9{\cdot}9$$

The maximum value of $\frac{X}{Y} = \frac{260+2\times 9{\cdot}9}{145-2\times 9{\cdot}9} = \frac{279{\cdot}8}{125{\cdot}2} = 2{\cdot}23$

As X/Y is greater than 1 the maximum value of

$$\frac{a}{b} = \frac{X}{Y} = 2{\cdot}23$$

Minimum value of

$$\frac{X}{Y} = \frac{260-2d'}{145+2d'}$$

where
$$d' = \frac{(145-260)+\sqrt{[(145-260)^2+(260+145+4)\times 260\times 145]}}{260+145+4}$$
$$= 9{\cdot}3$$

https://doi.org/10.1017/S0016672300003803 Published online by Cambridge University Press

The minimum value of $\frac{X}{Y} = \frac{260-2\times 9{\cdot}3}{145+2\times 9{\cdot}3} = \frac{241{\cdot}4}{163{\cdot}6}$

As X/Y is greater than 1 the minimum value of

$$\frac{a}{b} = \frac{X+Y}{2Y} = \frac{241{\cdot}4+163{\cdot}6}{2\times 163{\cdot}6} = 1{\cdot}24$$

We conclude that the data is consistent with any value of a/b between 2·23 and 1·24 in the absence of any knowledge of the magnitudes of a and b.

2. TRANSDUCTIONS OF THREE LINKED FACTORS

A transduction $A^-B^-C^-$ ($\times$) $A^+B^+C^+$, with selection for C^+, is represented in Fig. 4.

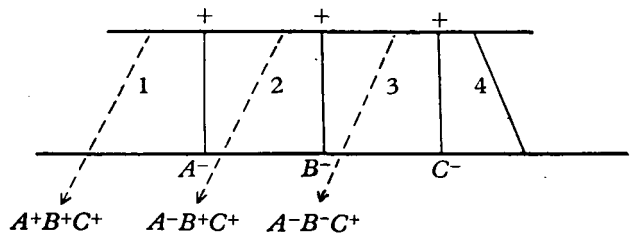


Fig. 4. The origin of $A^+B^+C^+$, $A^-B^+C^+$ and $A^-B^-C^+$ transductants from the transduction $A^-B^-C^-$ ($\times$) $A^+B^+C^+$ with selection for C^+. In addition $A^+B^-C^+$ transductants will be formed when recombination occurs in each of the four regions.

The discussion proceeds as for two linked factors by imagining a population of paired fragments with recombination in region 4, letting α, β and γ be the chances of recombination in regions 1, 2 and 3 and then substituting a, b and c as our best estimates of α, β and γ. The transductants are of four genotypes and these, with the proportion of paired fragments that is expected to give rise to each, are listed in Fig. 5.

Genotype of transductant	Regions of recombination	Expected relative frequencies
$A^+B^+C^+$	1, 4	$a(1-b)(1-c)$
$A^-B^+C^+$	2, 4	$(1-a)b(1-c)$
$A^-B^-C^+$	3, 4	$(1-a)(1-b)c$
$A^+B^-C^+$	1, 2, 3, 4	abc

Fig. 5. The expected relative frequencies of the four classes from the transduction set out in Fig. 4, if the chances of recombination in regions 1, 2, 3 and 4 are a, b, c and d.

In analysing data from an experiment in which two factors are transduced we saw that it was only possible to derive estimates of the *relative* lengths of the three regions. The data from experiments involving three factors yield estimates of the *numerical values* of the frequencies of recombination between the markers by the following procedure.

https://doi.org/10.1017/S0016672300003803 Published online by Cambridge University Press

The expected ratio of $A^-B^-C^+$ to $A^+B^-C^+$ transductants can be equated to that observed:

$$\frac{(1-a)(1-b)c}{abc} = \frac{W}{Z}$$

Similarly,

$$\frac{(1-a)b(1-c)}{a(1-b)(1-c)} = \frac{X}{Y}$$

These equations can be solved for a as follows:

$$\frac{(1-a)(1-b)c}{abc} \times \frac{(1-a)b(1-c)}{a(1-b)(1-c)} = \frac{W}{Z} \times \frac{X}{Y}$$

$$\frac{(1-a)^2}{a^2} = \frac{WX}{ZY}$$

$$a = \frac{1}{1+\sqrt{\left(\frac{WX}{ZY}\right)}}$$

Similarly

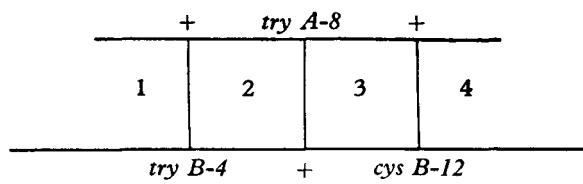
$$b = \frac{1}{1+\sqrt{\left(\frac{WY}{XZ}\right)}}$$

and

$$c = \frac{1}{1+\sqrt{\left(\frac{XY}{ZW}\right)}}$$

Values for b and c can also be obtained when A^+ transductants are selected. In addition this selection will yield an estimate of recombination in region 4.

As an illustration we can analyse data from the transductions *tryB-4 tryA-8*$^+$ *cysB-12* (×) *tryB-4*$^+$ *tryA-8 cysB-12*$^+$ when *tryB-4*$^+$ transductants are selected and when *cysB-12*$^+$ transductants are selected. These data (E.V. Glanville, personal communication) are set out in Fig. 6. In the first experiment 53 transductant



(a) *tryB-4*$^+$ selected		(b) *cysB-12*$^+$ selected	
Regions of recombination	Transductants observed	Regions of recombination	Transductants observed
1, 2	45	4, 3	225
1, 3	297	4, 2	21
1, 4	246	4, 1	84
1, 2, 3, 4	15	4, 3, 2, 1	7

Fig. 6. Data from the transduction *tryB-4 tryA-8*$^+$ *cysB-12* (×) *tryB-4*$^+$ *tryA-8 cysB-12*$^+$ when *tryB-4*$^+$ transductants are selected (a) and when *cysB-12*$^+$ transductants are selected (b).

https://doi.org/10.1017/S0016672300003803 Published online by Cambridge University Press

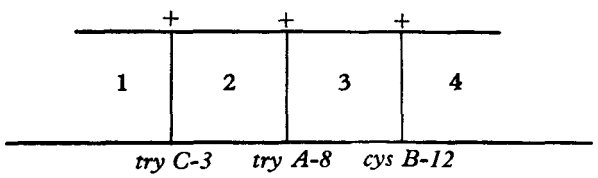
colonies, and in the second 35, were found to be a mixture of two or three recombinant genotypes. As the mechanism by which these mixed colonies arise is not known they cannot be referred to one of the four classes of transductants and so we have no alternative but to exclude them from the analysis.

The calculations from these data yield the values of *a*, *b*, *c* and *d* (the estimates of recombination in regions 1, 2, 3 and 4) which are shown in Fig. 7. Although there is a suggestion of negative interference between each region of obligatory recombination and the adjacent region the estimates from the two experiments are very close. This agreement between the estimates obtained when *tryB-4*$^+$ and *cysB-12*$^+$ are selected is further evidence that the fragments that include these markers are equal in length.

	(a) *tryB-4*$^+$ selected	(b) *cysB-12*$^+$ selected
a	—	0·261
b	0·088	0·081
c	0·405	0·483
d	0·348	—

Fig. 7. Estimates of *a*, *b*, *c* and *d* (the frequencies of recombination in regions 1, 2, 3 and 4) from the data presented in Fig. 6.

Data from a second pair of transductions (E. V. Glanville, personal communication) are set out in Fig. 8 and the calculated estimates of *a*, *b*, *c* and *d* are in Fig. 9.



(a) *tryC-3*$^+$ selected		(b) *cysB-12*$^+$ selected	
Regions of recombination	Transductants observed	Regions of recombination	Transductants observed
1, 2	90	4, 3	281
1, 3	333	4, 2	39
1, 4	303	4, 1	85
1, 2, 3, 4	22	4, 3, 2, 1	18

Fig. 8. Data from the transductions *tryC-3 tryA-8 cysB-12* (×) *tryC-3*$^+$ *tryA-8*$^+$ *cysB-12*$^+$ when *tryC-3*$^+$ transductants are selected (a) and when *cysB-12*$^+$ transductants are selected (b).

	(a) *tryC-3*$^+$ selected	(b) *cysB-12*$^+$ selected
a	—	0·272
b	0·123	0·147
c	0·342	0·552[1]
d	0·322	—

[1] Not significantly different from 0·5.

Fig. 9. Estimates of *a*, *b*, *c* and *d* (the frequencies of recombination in regions 1, 2, 3 and 4) from the data presented in Fig. 8.

There is no suggestion of negative interference between regions 1 and 2 but a stronger indication than before of negative interference between regions 3 and 4. Bearing in mind that *tryC-3* is known on other evidence to be located a little to the left of *tryB-4* the agreement between the estimates from these two pairs of transduction experiments is an encouraging indication of the usefulness of this method of analysis.

3. DISCUSSION

Any attempt to interpret the data from transduction experiments with precision must face the problem of excluding from the arguments as many as possible of those stages in a transduction experiment which are influenced by factors other than the location of the genetic markers. These factors include the ratio of the number of phage particles to the number of recipient cells, the proportion of phage particles which carry the relevant genetic fragment, the chance of this fragment pairing with the homologous region of the genome of the recipient cells and the proportion of recipient cells that survive infection. This has been achieved in the analyses I have described by always starting by considering a population of paired fragments with recombination in a particular region. This recombination is the guarantee of pairing.

The above procedure is impossible for single factor transductions (i.e. A^- ($\times$) A^+), where the only data is the number of wild-type colonies *per* number of phage particles or *per* number of recipient bacteria or *per* number of survivors. If the number of transductants is determined only by the position of the marker relative to the ends of the fragment we would expect that the number of transductants would be related to the known linkage order of the markers. Neither the extensive data of P. E. Hartman for histidine mutants (Hartman, 1956) nor the data of Demerec & Z. Hartman (1956) for tryptophane mutants of *Salmonella typhimurium* show this. These data are of the total number of transductants that are obtained when constant numbers of recipient cells and phage particles are used. We shall be able to interpret the results of single factor transductions in terms of the relative positions of the markers only if we can control the experimental conditions, and devise a way of expressing the data, so that we are effectively comparing the results of recombination in approximately equal populations of paired fragments. Smith-Keary has pointed out to me that this might be achieved if the experiments were carried out with a constant ratio of phage particles to recipient cells and the number of transductants expressed per number of surviving bacteria.

The usefulness of the analyses of two- and three-point transduction data is that they enable us to construct maps that are equivalent to the linkage maps for higher organisms. From two-point transductions we learn the relative frequency of recombination in the regions and from three-point transductions we obtain a numerical estimate of the frequency of recombination in each region. In addition we can make an attempt to estimate the lengths of the fragments that are transduced. For example, the analysis (B above) of the tryptophane-cystine fragment revealed 8·45% recombination between *tryA-8* and *tryB-4* and 13·5% between *tryA-8* and *tryC-3*. *tryA*, *B* and *C* are known to be adjacent loci. The position of

C-3 within the *C* locus is not known; *B-4* is near the centre of the known mutants of the *B* locus and *A-8* is known to be further from the *B* locus than most *A* mutants. Thus the average length of these loci is of the order of 6% recombination—by comparison with fungi and higher organisms this is a remarkably high value but is comparable with findings at the *r*II locus in bacteriophage T4 (Chase & Doermann, 1958). By simple summation of the recombination percentages (Figs. 7 and 9) we obtain estimates of 114 and 118 for the length of the fragment. These will be somewhat low estimates as they take no account of multiple crossing-over in the longer regions. If the fragment is a sequence of loci similar in lengths to the tryptophane loci it will consist of a minimum of 19 loci. From his two-point transduction data Smith-Keary has calculated that the minimum length of the leucine-arabinose fragment is 13 times the average length of the four leucine complementation groups (Smith-Keary & Dawson, 1963).

Hartman, Loper & Serman (1960) have discussed the possibility that there is an 'action of specific alleles on recombination frequency'. This effect is not unexpected if the mutation is a structural alteration of the gene. It is not yet clear whether apparent site mutations generally affect the frequency of recombination in adjacent regions or whether this frequency is affected by only a small minority of such mutations. The interpretations of data that I have discussed would seem to be a necessary basis for the further study of this phenomenon.

4. SUMMARY

1. Data from transduction experiments with *Salmonella typhimurium* can be interpreted in terms of frequencies of recombination by accepting the arguments of Ozeki that all transduced fragments that include a particular locus are equal in length.

2. From a two-point transduction the ratio of the two classes of transductants is not a measure of the ratio of the frequencies of recombination in the two relevant regions but only provides maximum and minimum values for this ratio.

3. From three-point transductions the frequencies of recombination in the regions between the markers, and between the terminal markers and the ends of the fragment, can be estimated.

4. The difficulties of interpreting data from experiments in which a single factor is transduced are briefly discussed.

I wish to thank Dr P. F. Smith-Keary for so helpfully discussing many of the arguments presented in this paper. The research reported herein has been sponsored by the Office, Chief of Research and Development, U.S. Department of the Army, through its European Research Office.

REFERENCES

Chase, M. & Doermann, A. H. (1958). High negative interference over short segments of the genetic structure of bacteriophage T4. *Genetics*, **43**, 332–353.

Dawson, G. W. P. & Smith-Keary, P. F. (1960). Analysis of the *su-leuA* locus in *Salmonella typhimurium*. *Heredity*, **15**, 339–350.

Demerec, M. & Hartmann, Z. (1956). Tryptophan mutants in *Salmonella typhimurium*. *Genetic Studies with Bacteria, Carnegie Institution of Washington publication*, **612**, 5–33.

Hartman, P. E. (1956). Linked loci in the control of consecutive steps in the primary pathway of histidine synthesis in *Salmonella typhimurium*. *Genetic Studies with Bacteria, Carnegie Institution of Washington publication*, **612**, 35–61.

Hartman, P. E., Loper, J. C. & Serman, D. (1960). Fine structure mapping by complete transduction between histidine-requiring *Salmonella* mutants. *J. gen. Microbiol.* **22**, 323–353.

Ozeki, H. (1959). Chromosome fragments participating in transduction. *Genetics*, **44**, 457–470.

Smith-Keary, P. F. & Dawson, G. W. P. (1963). Transduction analysis using leucine requiring mutants of *Salmonella typhimurium*. *Genet. Res.* **4**, 427–440.