GENETIC RECOMBINATION IN *ESCHERICHIA COLI*

II. CALCULATION OF INCORPORATION FREQUENCY AND RELATIVE MAP DISTANCE BY RECOMBINANT ANALYSIS

P. G. DE HAAN AND C. VERHOEF
Laboratory of Microbiology, The State University, Utrecht (The Netherlands)
(Received September 9th, 1965)

SUMMARY

In this paper a mathematical analysis based on the physical exchange of genetic material is presented for a four-factor cross. The incorporation frequency of donor markers and the relative map distances may be accurately estimated from the frequencies of the eight recombinant classes. The results obtained in K12 × K12 and K12 × B crosses are in good agreement with the theory.

INTRODUCTION

The genetic map of *Escherichia coli* is based on information obtained from transfer curves. The method is good for measuring the distance between markers which are more than one minute apart; at lesser distances its accuracy is restricted.

Mapping by recombinant analysis may be achieved by selecting for a distal donor marker and scoring for more proximal unselected markers. The cross-over frequency of an unselected marker is a measure of its distance from the selected marker. The cross-over frequencies observed in this kind of cross only represent the probability that a cross-over occurs in zygotes which give rise to a given type of (selected) recombinant, but they give no information about the probability of the same cross-over in all zygotes. The recombination frequencies or recombination units are not additive and no mathematical test on the results is available. In this paper a mathematical analysis based on the physical exchange of genetic material will be presented for a four-point cross. The parameters of a cross, *i.e.* the incorporation frequency of donor markers and the relative map distances may be accurately estimated from the observed numbers of recombinants

MATERIALS AND METHODS

The strains used in this paper and the method of crossing are given in our previous paper[5].

MATHEMATICAL ANALYSIS

In our previous paper[5] we have presented evidence that the linkage frequency of an unselected proximal marker is:

$$\beta = \alpha + (1 - \alpha)e^{-k\mu} \tag{1}$$

The incorporation frequency (α) and the average number of breakage events (k) per min transfer time were calculated from observed linkage frequencies and from transfer times determined in separate blender experiments. The error in the calculation of the parameters α and k is unfortunately dependent on the precision in the estimation of transfer times. A more precise method for the estimation of the incorporation frequency and the relative map distance may be obtained by the determination of frequencies of recombinant classes in a multi-factor cross.

The number of recombinant classes in a cross of the type given in Fig. 1 with n markers (one selected and $n - 1$ unselected) is 2^{n-1}. The number of unknowns is n (α and $n - 1$ lengths of segments). The number of degrees of freedom ($2^{n-1} - 1$) must be greater than the number of parameters (n) to be estimated in order to compare the observed frequencies of the unselected markers with the calculated frequencies. A four-point cross as given in Fig. 1 with seven degrees of freedom and four unknowns leaves us thus three degrees of freedom for the mathematical test.

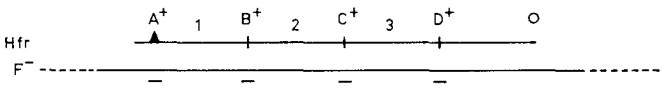

Fig. 1. Diagram of a four-point cross. ○, origin of the Hfr chromosome; ▲, selected marker.

In this cross A+ recombinants are selected and the markers B, C and D are scored as unselected markers. We will make the same assumptions as in our previous paper[5].

(1) The terminal ends of the Hfr fragment both initiate an obligatory breakage event.

(2) Further breakage events are randomly distributed.

(3) The breakage events in the segments which are proximal to D and distal to A have no influence on the incorporation of the markers.

(4) The incorporation frequency of a donor segment is independent of the incorporation events of adjacent segments. Each Hfr segment between two adjacent breakage events has thus a fixed probability for integration (incorporation frequency).

Let: l_1, l_2 and l_3 correspond to the lengths of the segments 1, 2 and 3 respectively (Fig. 1); k be the average number of breakage events per unit length; α be the incorporation frequency of a donor segment.

The average number of breakage events in a segment with length l is kl. The probability that no breakage event occurs in this segment is given by the first term (e^{-kl}) of the Poisson distribution:

$$P(n) = \frac{(kl)^n}{n!} e^{-kl}$$

Consequently the probability that (one or more) breakage events take place in this segment is $1 - e^{-kl}$.

For convenience we will further use the substitutions:

$$x_1 = 1 - e^{-kl}1; \; 1 - x_1 = e^{-kl}1 \text{ etc.} \tag{2}$$

We may now write down the expectations for the various breakage events and the expectations for the incorporations necessary to obtain the eight recombinant classes. Table I gives these expectations. It may be seen from this table that $B^+C^+D^+$ recombinants may be obtained as the result of eight different "breakage and reunion" events. The total probability for this class is obtained by multiplication of the eight

TABLE I

THE PROBABILITIES FOR THE BREAKAGE AND INCORPORATION EVENTS IN A FOUR-POINT CROSS

The type of cross is given in Fig. 1. All recombinats are A^+; $+ - +$ represents a $B^+C^-D^+$ recombinant etc. The first column gives the possible breakage events and the second column gives the corresponding probabilities. In the third column the various incorporation probabilities for the eight recombinant classes are given. To obtain the total probability for a given recombinant, each incorporation probability must be multiplied by the corresponding probability for the breakage events (see text).

Breakage events in segment	*Probability*	*Probability of incorporation for a recombinant of type:*							
		+ + +	+ + −	+ − −	− − −	− − +	− + +	+ − +	− + −
—	$(1-x_1)(1-x_2)(1-x_3)$	α							
1	$x_1 \quad (1-x_2)(1-x_3)$	α^2			$\alpha(1-\alpha)$				
2	$(1-x_1) \; x_2 \quad (1-x_3)$	α^2		$\alpha(1-\alpha)$					
3	$(1-x_1)(1-x_2) \; x_3$	α^2	$\alpha(1-\alpha)$						
1 2	$x_1 \quad x_2 \quad (1-x_3)$	α^3		$\alpha^2(1-\alpha)$	$\alpha(1-\alpha)^2$		$\alpha^2(1-\alpha)$		
1 3	$x_1 \quad (1-x_2) \; x_3$	α^3	$\alpha^2(1-\alpha)$		$\alpha(1-\alpha)^2$	$\alpha^2(1-\alpha)$			
2 3	$(1-x_1) \; x_2 \quad x_3$	α^3	$\alpha^2(1-\alpha)$	$\alpha(1-\alpha)^2$				$\alpha^2(1-\alpha)$	
1 2 3	$x_1 \quad x_2 \quad x_3$	α^4	$\alpha^3(1-\alpha)$	$\alpha^2(1-\alpha)^2$	$\alpha(1-\alpha)^3$	$\alpha^2(1-\alpha)^2$	$\alpha^3(1-\alpha)$	$\alpha^3(1-\alpha)$	$\alpha^2(1-\alpha)^2$

probabilities for the breakage events (column 2) with the corresponding incorporation frequencies (column 3). The summation of the eight terms is the probability for a $B^+C^+D^+$ recombinant. It may be seen from Table I that one recombinant class is obtained as the result from eight different events; three recombinant classes as the result of four events, etc. The expected numbers of recombinants in the eight classes are obtained by multiplying the probabilities for the various recombinant classes by the unknown number of zygotes (Z) which were plated. The expected and observed numbers of recombinants obtained from Z zygotes are given in Table II.

The parameters may be estimated by means of the method of maximum likelihood (compare ref. 1). In practice one wants to have a reasonable good set of approximate estimates with which to start the iterative scoring procedure. Estimations based on the formulae:

$$\frac{\alpha}{1-\alpha} = \frac{(e+g)}{(b+h)} \times \frac{(a+b+f+h)}{(c+d+e+g)}$$

$$(1-\alpha)x_1 = \frac{d+e+f+h}{n}$$

$$(1-\alpha)x_2 = \frac{c+g}{a+b+c+g}$$

$$(1-\alpha)x_3 = \frac{b}{a+b}$$

may be used as initial estimates for the iterative calculation of the parameters. The expectations for the eight recombinant classes may then be calculated with the aid of the formulae which are given in Table II. The appropriate test for goodness-of-fit on calculated and observed data is a χ^2 test with three degrees of freedom. The relative map distance (kl) of a segment can be calculated from the estimated probability ($\hat{x}$) for one or more breakage events in the segment with the aid of Eqn. 2.

TABLE II

EXPECTATIONS AND OBSERVATIONS FOR A FOUR-POINT CROSS

The type of zygotes are given in Fig. 1. All recombinants are A^+. Z represents the number of zygotes which have received the Hfr fragment origin-selected marker (A^+). Note that αZ is equal to the number of A^+ recombinants.

Recombinant class			*Observed numbers*	*Expected numbers*
B	*C*	*D*		
+	+	+	a	$\alpha Z[1-(1-\alpha)x_1][1-(1-\alpha)x_2][1-(1-\alpha)x_3]$
+	+	−	b	$\alpha Z[1-(1-\alpha)x_1][1-(1-\alpha)x_2](1-\alpha)x_3$
+	−	−	c	$\alpha Z[1-(1-\alpha)x_1](1-\alpha)x_2(1-\alpha x_3)$
−	−	−	d	$\alpha Z(1-\alpha)x_1(1-\alpha x_2)(1-\alpha x_3)$
−	−	+	e	$\alpha Z(1-\alpha)x_1(1-\alpha x_2)\alpha x_3$
−	+	+	f	$\alpha Z(1-\alpha)x_1\alpha x_2[1-(1-\alpha)x_3]$
+	−	+	g	$\alpha Z[1-(1-\alpha)x_1](1-\alpha)x_2\alpha x_3$
−	+	−	h	$\alpha Z(1-\alpha)x_1\alpha x_2(1-\alpha)x_3$
Total			n	αZ

RESULTS

The crosses presented in this paper were all performed with the technique described in our previous paper[5]. The formation of mating pairs was restricted to a period of 5 min and 50 min were allowed for chromosome transfer. The zygotes were then incubated under aeration in broth for an additional period of 60 min. The recombinants were then plated on selective medium. The goodness-of-fit was unsatisfactory when the zygotes were plated without post-aeration. The explanation is that the eight types of recombinants have different viabilities when *zygotes* are plated on the selective plates. Leucine, in particular, suppresses the formation of recombinants. The effect of leucine disappeared when the zygotes were incubated for a period of 60 min in broth. The conclusion is thus that the post-aeration period of zygotes is essential to obtain reliable results.

The complete results of two crosses are given in Tables III and IV. The parameters ($\hat{\alpha}$ and $\hat{x}$) and the variances were estimated with the method of maximum likelihood with the aid of a computer. Table III gives the results of a Hfr $R_4 \times K12$ *ade*$_K$-*thr*-*leu*-*pro*$_A$ cross. The incorporation frequency in this cross was 0.505 ± 0.017 and this value is very close to the value (0.5) found in the same cross but estimated from linkage frequencies and distances in time units[5]. The average number of breakage

TABLE III

OBSERVED AND EXPECTED NUMBERS OF ade_K^+ RECOMBINANTS IN A Hfr R_4 × K12 $ade_K^- thr^- leu^- pro_A$ CROSS

The estimated parameters in this cross are: $\hat{\alpha} = 0.505 \pm 0.017$; $\hat{x}_1 = 0.544 \pm 0.030$; $\hat{x}_2 = 0.191 \pm 0.015$; $\hat{x}_3 = 0.687 \pm 0.026$; $k\hat{l}_1 = 0.79$; $k\hat{l}_2 = 0.21$; $k\hat{l}_3 = 1.16$; $k\hat{l}_{1+2+3} = 2.16$.

Genotype of recombinant			*Observed number of recombinants*	*Expected number of recombinants*	
thr	*leu*	*pro*			
+	+	+	664	660.6	
+	−	−	337	339.8	
+	−	−	66	68.3	
−	−	−	242	240.0	
−	−	+	125	127.6	$\chi_3^2 = 1.59$
−	−	−	23	26.0	$P = 0.75$
+	−	+	38	36.3	
−	+	−	17	13.4	
Total			1512	1512.0	

TABLE IV

OBSERVED AND EXPECTED NUMBERS OF ade_K^+ RECOMBINANTS IN A Hfr R_4 × B $ade_K^- thr^- leu^- pro_B^-$ CROSS

The estimated parameters in this cross are: $\hat{\alpha} = 0.019 \pm 0.003$; $\hat{x}_1 = 0.757 \pm 0.010$; $\hat{x}_2 = 0.312 \pm 0.019$; $\hat{x}_3 = 0.958 \pm 0.012$; $k\hat{l}_1 = 1.41$; $k\hat{l}_2 = 0.38$; $k\hat{l}_3 = 3.17$; $k\hat{l}_{1+2+3} = 4.96$.

Genotype of recombinant			*Observed number of recombinants*	*Expected number of recombinants*	
thr	*leu*	pro_B			
+	+	+	23	24.1	
+	+	−	379	373.5	
+	−	−	167	172.1	
−	−	−	1615	1614.1	$\chi_2^2 = 5.65$
−	−	+	23	30.1	$0.1 > P > 0.05$
−	+	+	2	0.6	
+	−	+	4	3.2	
−	+	−	14	9.3	
Total			2227	2227.0	

events over the segment ade_K–pro_A is 2.16, giving an average number of 0.216 events per min transfer time. This value is in good agreement with that found in the same cross with the method described earlier (0.211). The expected numbers in the eight recombinant classes give a very satisfactory goodness-of-fit ($\chi_3^2 = 1.59$). Table IV gives the data from a Hfr R_4 × B $ade_K^- thr^- leu^- pro_B^-$ cross. The incorporation frequency in this cross is very low (0.019). The average number of breakage events over the segment ade_K–pro_B is 4.96, giving an average number of 0.404 events per min transfer time. The estimated values for both parameters are again in good agreement with the values calculated previously (0.019 and 0.378 respectively). The number of recombinants in two classes (− + + and + − +) is very low and these classes were therefore taken together for the calculation of the goodness-of-fit. A quite satisfactory goodness-of-fit χ_2^2 of 5.64 was obtained.

The relative map distances ade_K–*thr*, *thr*–*leu* and *leu*–pro_A in the K12 × K12 cross are 36%, 10% and 54% of the segment ade_K–pro_A. The relative map distances

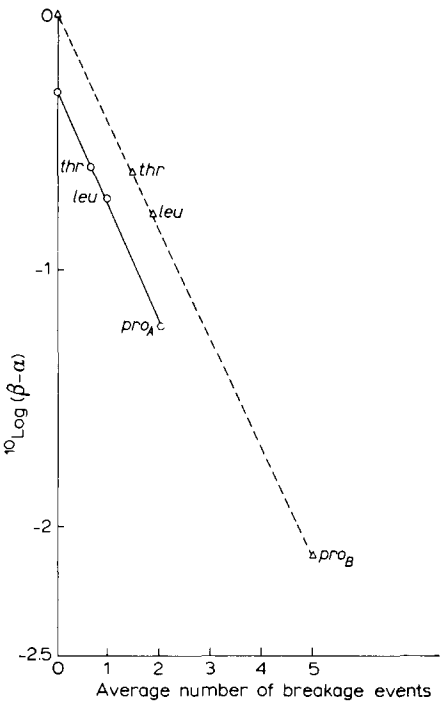

Fig. 2. Plot of $^{10}\log(\beta - \alpha)$ against relative map distance. For values of relative map distances (*kl*) see Tables III and IV. ○·····○, Hfr $R_4 \times$ K12 $ade_K^- thr^- leu^- pro_A^-$; △----△, Hfr $R_4 \times$ B $ade_K^- thr^- leu^- pro_B^-$.

in the K12 × B cross, based on a relative distance of the *ade$_K$–thr* segment of 36%, are 36%, 10.3% and 82% for the segments *ade$_K$–thr*, *thr–leu* and *leu–pro$_B$*, respectively. The relative map distances of the segment *ade$_K$–pro$_B$* is thus 128% of the segment *ade$_K$–pro$_A$*. The segment *thr–pro$_B$* is 44% larger than the segment *thr–pro$_A$*. On Taylor's map the difference is about 30%.

Fig. 2 gives the plot of $^{10}\log(\beta - \alpha)$ against the average number of breakage events (relative map distance[5]) in the various segments. Parallel straight lines (with a slope of $^{10}\log$ e) were obtained.

DISCUSSION

In this paper the expectations for the eight recombinant classes in a four-point cross have been developed. The results from two crosses given in Tables III and IV show that the observed and expected numbers in the eight recombinant classes are in good agreement. In a previous paper[4] evidence was presented that the mathematical model of BAILEY[1] gave satisfactory results. This method, however, is based on the assumption that odd numbers of cross-over events between two donor markers separate the markers. The evidence presented in our previous paper[5] shows that this assumption is incorrect. The explanation for the satisfactory results obtained with BAILEY's model is that the expectations for the various recombinant classes as presented in this paper are basically identical with BAILEY's expectations when the

incorporation frequency is exactly 0.5. The probability for a cross-over in the segment between the most proximal marker and the origin in BAILEY's model then becomes exactly 0.5. It is thus expected that crosses in which the incorporation frequency is about 0.5 will give satisfactory results with both methods, whereas crosses in which the incorporation frequency is low will only give satisfactory results with our method. This was indeed found in the K12 × B cross in which the incorporation frequency was very low (0.019). A satisfactory goodness-of-fit was obtained with our method ($\chi_2^2 = 5.49$) against a goodness-of-fit $\chi_2^2 = 60$ with BAILEY's method.

The incorporation frequency in the K12 × K12 cross is 0.505 ± 0.017. The probability for incorporation of a donor marker is thus equal to the probability for incorporation of an acceptor marker. This means that the segments are randomly incorporated; there is no preference for the incorporation of donor or acceptor segments.

In most K12 × K12 crosses a value of about 0.5 for the incorporation frequency was found (C. VERHOEF, unpublished) but in some crosses the incorporation frequency is significantly lower than 0.5. The lowest value found in K12 × K12 cross was 0.23 (ref. 5). No explanation for the low incorporation frequency was found. The goodness-of-fit in K12 × K12 crosses is, however, less satisfactory when the incorporation frequency is significantly lower than 0.5. A possible explanation is that the low incorporation frequency is caused by viability effects.

The low incorporation frequency in the K12 × B cross may be explained on the basis of the genetic inhomology of the parental DNA strands. It seems that a preference for the incorporation of the acceptor DNA exists. The low incorporation frequency in this system explains the low linkage observed in K12 × B crosses. The greater probability for a breakage event per unit length observed in the K12 × B cross has a less pronounced influence on linkage than the very low incorporation frequency. The restriction for K12 DNA in *E. coli* B (refs. 2 and 3) is thus mainly due to the preferential incorporation of acceptor DNA into the recombinant, probably as the result of the preferential breakdown of the donor segments after the breakage events.

The average number (k) of breakage events per unit length differs from cross to cross, the estimated kl therefore represents the relative map distance. These relative map distances are additive as long as interference is absent.

ACKNOWLEDGEMENT

Professor Dr. G. J. LEPPINK and Drs. A. TH. VAN DER BURGT are gratefully acknowledged for valuable aid in the mathematical treatment of our results.

REFERENCES

1 BAILEY, N. T. J., Bacterial genetics, Chapter 8 in *Introduction to the Mathematical Theory of Genetic Linkage*. Clarendon Press, Oxford, 1961 p. 119–136.
2 BOYER, H., Genetic control of restriction and modification in *Escherichia coli*. *J. Bacteriol.*, 88 (1964) 1652–1660.
3 HOEKSTRA, W. P. M. AND P. G. DE HAAN, The location of the restriction locus for λ · K in *Escherichia coli* B. *Mutation Res.*, 2 (1965) 204–212.
4 STOUTHAMER, A. H., P. G. DE HAAN AND H. J. J. NIJKAMP, Mapping of purine markers in *Escherichia coli* K_{12}. *Genet. Res.*, 6 (1965) 442–453.
5 VERHOEF, C. AND P. G. DE HAAN, Genetic recombination in *Escherichia coli*. I. Relation between linkage of unselected markers and map distance. *Mutation Res.*, 3 (1966) 101–110.