YALE UNIVERSITY
OSBORN BOTANICAL LABORATORY
NEW HAVEN, CONNECTICUT

July 4, 1947.

Dear Mather:

In mulling over the statement of the problem which I had previously stated in too abbreviated and confusing fashion, I think I have achieved a partial solution; at least I hope it may serve to convey the nature of my difficulties. I hope you will have time to go over this and detect any fallacies that may have crept in.

Assume the parental configurations: ABCd and abcD. Given the relative frequencies of the only detectable recombination classes (A..D or prototrophs) to estimate the absolute map distance A..D, assuming linearity and no interference. The regions are as follows:

A B C D and for the sake of convenience, let us call
1 2 3 the three "single-interchange" classes,

AbcD; ABcD, and ABCD, j , and the "triple-interchange" class AbCD, k.
The grand-total accumulation of data in the case where ABCD are BM, Lac, V,TL respectively, is as follows:

1... 817	27.3%		27.8%
2...1389	46.3		47.3
3... 729	24.3	....	24.9
k... 63	2.1	2935..	00.00
2998	100.0%		

The k's are separated in the third column to give the relative distances.

In your analysis of the problem, you suggested that the absolute distances might be given by the expressions:
"1" = $p_1q_2q_3$; "2" = $p_2q_1q_3$ and "3" = $p_3q_1q_2$ and k = $p_1p_2p_3$.

where p_1, etc., was the chance of a crossover in region one, i.e., the distance from A to B. It has just occurred to me that the analysis which I had just completed is congruous with yours if a more accurate meaning is given to these symbols, namely that p_1 is the probability that there is an interchange in region 1, i.e., that there is an odd number of crossovers in this region. The map distance which can, of course exceed 1.00 while a true probability cannot, should be defined as:

more accurately, between A and B

$$\int_A^B p\,dx$$ For infinitesimal segments, of course, p can only refer to single crossovers.

According to this definition, the absolute distance A..D is the mean number of crossovers per chromatid pair.

As a first approximation, consider the case of two-strand crossing over, with the distances 1,2,and 3 equal to each other, and estimate what the total distance must be to allow the class k of given size (2.1%). There is assumed to be no interference. If the distance then is x, the proportions of n-crossover types will be given by the Poisson distribution: $e^{-x}(1, x, x^2/2!, \ldots x^n/n!)$. However, since the even-numbered crossover types will not be recovered, only the odd terms need be considered. If the total number of crossovers is odd, a j will be recovered. If there is an even number of crossovers i.e. no interchange in any of the regions 1, 2, 3. If, however, there is an odd number of crossovers in each region, a k-type proto-

YALE UNIVERSITY
OSBORN BOTANICAL LABORATORY
NEW HAVEN, CONNECTICUT

troph will be obtained. For purposes of orientation, it may be pointed out that the limiting probability of a k is 1/4: If the total (odd) number of crossovers is large, the chance that there will be an odd number in region 1 is = to the chance of an even number, i.e, = 1/2. Similarly for the chances of region 2. This determines region 3 as odd, so the overall probability is o.5 x o.5 = o.25. This is to be expected, since it means that with a lrge number of crossovers, each class has an equal expectation.

The chances of 1 crossover in each region, for x=3 can be estimated from the binomial distribution, and turns out to be 2/9 = .22. The limit of .25 is very nearly reached for n=5, but thecalculations are too space-consuming to be worthwhile writing down.

The value of x corresponding to k= .021 can now be determined.

Each single-crossover will yield only j. Triple-crossovers will yield .22 k and .78 j, while odd n-ple crossovers will yield even more closely .25 k and .75 j. The weight to be given to each crossover-class is given by the Poisson series. To simplify the calculations somewhat the yield for n=3 can be approximated as also .25. The odd terms of the Poisson distribution are, of course, sinh x. The first term is (n=1) e^{-x} (x). The proportion of single-crossovers to total odd crossovers is then x/sinh x . This will be the same as j-3k/j+k.= 1 - 4k/j+k.

In the present case, k/j+k = 2.1% or .021; x/sinh x = .916.

x= .73 or 73 morgans. This turns out not to be very different from your estimate. x is of course a very sensitive function of k, e.g. for k= .03, x= 90 morgans, so that the estimate can be regarded only as a rough measure. With x only .7, the class n=5 can be disregarded compared to n=3 (.012 : 1) so that the yield can be calculated on the basis of a comparison of the N=1 and n= 3 classes only. We can also modify the calculation to include the ~~found~~ relative distances, neglecting the possible slight modification ~~that~~ the triples among them might have introduced. The chance that three crossovers will be distributed one each in the relative distances: 27.8, 47.3 and 24.9 is: 6(.278x.473 x.249) =.196. The proportion of n=3 to n=1 is simply $x^2/6$. Thus we have

the proportions of single-exchange types in triple crossovers is slightly different from the proportional distances

$$\frac{.196\ x^2/6}{1+\ x^2/6} = k/j+k = .021.$$

x= .72 or 72 morgans as a more exact estimate. About 8% will be the actual frequency of n=3.

So much for the hypothetical two strand case.

If crossing over occurs at a four-strand stage, the calculations are much more involved. Let n be the number of crossovers per tetrad. If n=1 the situation is much as in two-strand, since only the prototroph will be recovered. If n=2, however, 3 tetrads out of 4 (the digressive and the progressive) will yield j prototrophs. If n=3, 4 tetrads out of 16 will contain k prototrophs, 2 will contain no prototrophs, while 11 will contain j's. (One overlap, with 1 j; 1 k). For n=4, the situation is clearly even more complicated, The n=3 case mentioned is for those where there is one crossover in each region.(.196 of total). In general, the following will have to be calculated:

a) for any n the chance that there will be crossovers in each region. also, the proportion of the disproportionate types which yield prototrophs (all j).
b) the proportions of j and k where there are crossovers in each region.

YALE UNIVERSITY
OSBORN BOTANICAL LABORATORY
NEW HAVEN, CONNECTICUT

K. Mather
7/4/47.

The problem seems so complex, that I have begun by making some simplifying approximations. Firstly, we will count the total number of recombinants, rather than zygotes, in making our estimations, to avoid the difficulty of enumerating zygotes containing two kinds of protrophs (digressive multiple exchanges). On this basis then, both single- and double- crossover types (n=1,2) will be counted as giving all j prototrophs, while the type where n=3, and there is one crossover in each region will be considered as yielding 3j:1k. I have not completed the analysis, but it seems very likely on this basis that for any value of n, there will be this ratio, for that 3:1 set of crossover classes in which there is one or more crossovers in each region. The difference between this and the two strand system is that the even-numbered types must be counted also.

Does it lead to any diff. in the long run -

Theoretically, it should not be difficult to find an expression for p=f(n) where p is the chance that if n marbles are thrown at random into three equal (or more precisely, somewhat unequal) boxes, that none of these boxes shall be empty. I have not been able to find it however, and have had to rely upon the binomial expansion, summing all the appropriate cases, to obtain values of p from n=3 to n=8. p must be known, since

$$k/j+k = \frac{\sum_{i=3} p/4 \cdot x^n/n!}{e^x - 1} = .021$$

The effect of increasing x, then, is to augment the proportions of higher values of n, which in turn increases the proportions of the types in which there is at least one exchange in each region.

Using this formula, and the calculated values of p, the following trials were made as enumerated in the table:

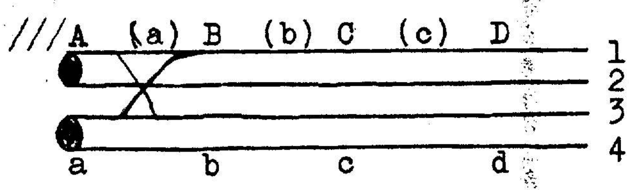
This is based on the found distances, not each region equal, which would give .222

			n = 1		n = 2		n = 1.6	
n	p	F(x)	F(1)	p/4.F(1)	F(2)	p/4.F(2)	F(1.6)	p/4.F(1.6)
1	0	x	1	0	2	0	1.6	0
2	0	$x^2/2$	.5	0	2	0	1.28	0
3	.196	$x^3/6$	.17	.008	1.33	.065	.68	.033
4	~~.445~~	$x^4/24$	.04	.001	.75	.083	.27	.030
5	.617	$x^5/120$	.01	-	.27	.041	.09	.004
6	.740	$x^6/720$	-	-	.09	.017	.02	.004
7	.890	$x^7/5040$	-	-	.02	.004	-	-
8	c.1.0	$x^8/40320$	-	-	.005	-	-	-
		$e^x - 1$	1.72	.009	6.46	.210	3.93	.081
				r= .005		r=.032		r=.021

Thus, n=1.6 gives r=.021 in agreement with the expected j/j+k= 2.1%. 1.6 crossovers per tetrad is, of course equivalent to 80 morgans between A..D, which is surprisingly close to the 2-strand estimate. Is this more than an accident??

In the course of all these calculations, I became rather tired of drawing four chromatids and spotting innumerable combinations of crossovers on them to determine the proportions of various types. Therefore I set myself to formulating an operational treatment of multiple crossovers; I should be interested to hear whether it would of any use to anyone, and whether anything similar has already been published.

The method is:



A crossover is written as a_{13} for the exchange indicated, in region a involving strands 1 and 3. a_{13} is an operator, the operands of which are the 4 strands- s_1, s_2, s_3, and s_4 = S

For the purpose of developing the arithmetic of these operators, there are two parts to the operation: $a_{13}.s$ which have to be considered:

a) on s_1 it substitutes a for A . This may be written $(\frac{a}{A})$ or more simply, (a). on s_3 it substitutes A for a. This has the same notation.
b) it converts s_1 from 'rank' 1 to rank 3.
c) it acts, of course, in similar fashion on s_3.
d) an operator has no effect, e.g. (1)x , on operands of different rank.
e) the operators are written from left to right, in the sequence of the crossovers; the order of action is from right to left.
f) a.a = (1)
g) the combinations of operators are:

a = (a)	ab = (b) (* (a)(ab))	
b = (ab)	ac = (bc)	
c = (abc)	bc = (c)	abc = (a.bc) = (ac)

order is general

To combine operators, e.g. $a_{13}.b_{23}.c_{14}.s$, write the operands in separate columns:

s_1	s_2	s_3	s_4
c_4	(1)	(1)	c_1
(1)	b_3	b_2	(1)
(1)	a_1	(1)	a_3

, and perform the operation indicated by the righthand-most term. (c_{14}). The change in rank is very readily symbolized by using the subscript.

The rank of a complex is its left-hand most subscript

Then perform the operation indicated by the penultimate symbol, keeping in mind the shift in ranks of the previous operator. And so forth. The tetrad then consists of: c_4s_1; $a_1b_3\ s_2$; b_2s_3; $a_3c_1s_4$. or ~~ABcD, abCD, aBcd~~ abcD; AbCD; ABcd; aBCd, which, as you can verify is the correct answer. The generalization of this method to any number of crossovers or regions is, of course, obvious. I am going, now, to work on derived rules for the production of the various crossover types, based on the relationships between subscripts. E.g., a triple excahnge must be abc or (ac). This can only be produced if: $a_{1m}.b_{mn}.c_{np}.s$ and would only be a prototroph in half the cases ~~if~~. From such rules, the relative frequencies of various types should not be too difficult to calculate, even in more complex instances. The method is much easier to use than to explain.

i.e. AbCd or aBcD with these parents

I hope I have made this letter somewhat clearer (if at the expense of brevity.) Please accept my gratitude for your continued interest.

Yours sincerely,

Joshua Lederberg

Joshua Lederberg