

# AN IMPROVED GENETICAL MAP OF PUNNETT'S 'B' CHROMOSOME IN THE SWEET PEA, *LATHYRUS ODORATUS* L.

By N. R. BHAT, *Department of Genetics, University of Cambridge*

## INTRODUCTION

It is common knowledge that statistics has been a close and helpful associate of genetics throughout its advancement. Bateson (1909, p. 318), attributing the success of Mendel primarily to the statistical nature of his investigations, alludes to Mendel's following remarks, 'Those who survey the work done in the department will arrive at the conviction that among all the numerous experiments made, not one has been carried out to such an extent and in such a way as to make it possible to determine the number of different forms under which the offspring of hybrids appear or to arrange these forms with certainty according to their separate generations or definitely to ascertain their statistical relations', and observes, 'It is to the clear conception of these three primary necessities that the whole success of Mendel's work is due. . . .' The fairly recent statistical methods, e.g. the proper use of  $\chi^2$  and the analysis of variance, have revolutionized the planning of controlled experiments and the interpretation of their results. More particularly for the present paper, the methods of applying the theory of estimation based on the 'Method of Maximum Likelihood' are likely also to be valuable in the field of genetics.

A characteristic feature of such statistics is the ease with which they enable different bodies of data, either wholly or partially alike, to be combined to make improved estimates of parameters. The combination is made through a statistic known as 'Information', which also serves as a measure of the success of experiments in terms of the effort, time and expenses involved; and this enables improvements to be devised if necessary. This point is cogently elucidated by Mather (1938). Such statistics are, therefore, a valuable aid to the modern experimenter who often works under pressure and without adequate facilities.

Soon after the rediscovery of Mendel's laws, Bateson & Punnett undertook a comprehensive programme of work on the genetics of the sweet pea (*Lathyrus odoratus* L.). Among several other characters they studied (1) dark-light axil, (2) fertile-sterile (in Darwin's terms 'contabescent') anthers and (3) normal-cretin flower, which, on being studied in pairs, led to the discovery of the phenomenon of linkage.

Bateson & Punnett (1911) tried to explain linkage by postulating what they called 'Reduplication of gametes' in the germ tract. They and Punnett (1913) have described the hypothesis in detail. Punnett (1913) made earnest attempts to interpret the abnormal segregations observed in associated characters including those referred to above on this basis; but it is evident from the numerous doubts and difficulties he encountered that he was not satisfied with the results. Later, both because the reduplication hypothesis had no factual basis, and also since the 'Chromosome theory' of heredity propounded by Morgan and his associates gave overwhelming evidence of the linear arrangement of genes and their linkage in chromosomes, he seems to have acquiesced in the genetical map produced by Bridges (1914) in respect of the characters in question. He named the chromosome as the 'B' chromosome (Punnett, 1923, 1932), and the genes responsible for the characters

as  $B_1$ ,  $B_2$  and  $B_3$  respectively. It must, however, be pointed out with regard to Bridges's map that it was, as he himself puts it, only 'illustrative', and this for two important reasons. First, the data, coming as they did from  $F_2$  segregations; were not very adequate; and secondly, the method used for analysis, viz. Yule's coefficient of association, was 'the least unsatisfactory' available at the time. Thirdly, in the light of present-day knowledge the recombination percentages may not directly represent linear distances between genes as shown in his map. Fortunately, there are now opportunities to remedy all these defects. Punnett has in his records a large collection of unused data (compiled up to 1928) on the characters, which he has kindly made available for use. Fisher's new method of estimating linkage by 'scoring' (1946*b*), admirably suited to such type of work, together with Kosambi's formula for conversion of recombination percentages into linear distances, are also at hand. It is, therefore, thought fit to present in this paper an improved map of Punnett's 'B' chromosome.

#### REVIEW OF METHODS OF ESTIMATION

Ever since linkage was discovered by Bateson & Punnett attempts have been made by various workers, both statisticians and geneticists, to evolve formulae and methods for estimating the degree of association between characters. Amongst these may be mentioned methods developed by Yule, Collins, Emerson, Haldane, Takezaki, Kappert, Owen, Fisher & Balmukand, and Immer. All of these, however, fall under one or the other of the four categories discussed by Fisher (1946*a*), the only method efficient for all types of problems of estimation being the Method of Maximum Likelihood. However, it is considered essential to review briefly the methods employed by Bateson & Punnett on the one hand and by Bridges on the other, and compare them with Fisher's scoring system, first, to justify the choice of the scoring method in this paper, and, secondly, to show why revision of the map with this new device has been considered worth while.

The practice that Bateson & Punnett followed was more or less empirical. They assumed the following types of gametic series:

	Double dominant	Single dominant	Double recessive
Coupling	$(n-1)^*$	1	$(n-1)$
Repulsion	1	$(n-1)$	1

generating zygotic series of the kind:

Coupling	$3n^2(2n-1)$	$(2n-1)$	$n^2 - (2n-1)$
Repulsion	$(2n^2+1)$	$(n^2-1)$	1

Theoretical expectations were derived on this basis for certain selected gametic ratios, e.g. 1:3:3:1, 1:7:7:1, 1:15:15:1, 1:31:31:1, etc. Frequencies observed in progenies were compared by eye with the series of such theoretical expectations to see to which of the gametic series they conformed.

This method, while it very roughly indicated the intensity of linkage, could not have been anything more than one of approximation for obvious reasons. No wonder, therefore, that the inefficient estimates arrived at created difficulties in reconciling the behaviour of different progenies even in the same cross.

The method adopted by Bridges was, as previously stated, Yule's coefficient of asso-

\*  $(n-1):1$  or  $n:1$  were supposed to be the gametic ratios brought about through alternating periclinal and anticlinal cell divisions in gametogenesis.

ciation,\* which belongs to the category of the 'Coefficient of correlation' advocated by Takezaki (1925). Immer (1931) has discussed at length the efficiency of the 'Coefficient of correlation'. He concludes: 'The correlation method is fairly efficient in the coupling phase and for loose linkage in repulsion. For close linkage in repulsion it is not efficient.' He also draws attention to the inaccuracy of 'the error formula as used by Takezaki, which is based on the incorrect method of treating the fourfold table of phenotypic frequencies as a normal frequency surface'. The coefficient of association being even slightly inferior to the coefficient of correlation (Yule, 1912) is not therefore very efficient for estimation.

Fisher's (1946*b*) scoring method adopted in this paper is but the method of maximum likelihood in a form more convenient for use. Its efficiency is therefore assured. Its advantageous features, in addition to those of the Method of Maximum Likelihood, are:

(i) The scores are linear functions of the observed frequencies and additive; as such, 'the efficiently weighted combination of different lots of data is arrived at simply by adding their scores', thus making the combination an easy operation.

(ii) The aggregate score of a progeny is zero at the correct value of the estimate. As will further be seen, this property is very useful in detecting the type of mating.

(iii) The Information, which is the sampling variance of the score, while being essentially the same in nature as in the Maximum Likelihood method, serves, in this system, two useful objects, viz.:

(a) It indicates how far from or near to the trial value is the correct value of the parameter, the score at the trial value divided by its Information giving the correction factor leading towards the correct estimate.

(b) The square of the aggregate score of a progeny divided by its Information is its  $\chi^2$  for one degree of freedom. When several families or groups of families, say  $k$  in number, are pooled together, their homogeneity can be easily tested by adding their individual  $\chi^2$ 's derived in this way and subtracting therefrom the  $\chi^2$  for the totals, the remainder being  $\chi^2$  for  $k-1$  degrees of freedom for heterogeneity as in any ordinary case. When different kinds of groups, each with many sub-groups, are combined, the  $\chi^2$  can be analysed into components as in the case of hierarchical classification.

(iv) The method includes a novel feature, based on Kosambi's formula, for improving the estimate of a smaller segment by taking account of information from bigger segments containing it. This should interest anyone who wishes to take maximum advantage of all data at his disposal. In view of the fact that geneticists have been so far estimating recombination percentages of any segment from data collected on that segment alone, this would definitely constitute *an advance in the estimation of linkage*.

There is no other method known at present commanding such unique efficiency and manifold advantages. This fact, together with the peculiar nature of the data, which include even families of unknown mating types in different filial generations, have favoured its choice for estimation.

#### MATERIAL AND METHODS

As stated above, the material used is Punnett's data of segregation for three characters in the sweet pea, viz. (i) dark-light axil, (ii) fertile-sterile anthers and (iii) normal-cretin flowers. The estimate of the recombination percentage of a fourth character (iv) purple-

\* Yule's formulae (1900) for the coefficient of association and its error are

$$Q = \frac{(AB \times ab) - (Ab \times aB)}{(AB \times ab) + (Ab \times aB)} \quad \text{and} \quad B_Q = \frac{1 - Q^2}{2} \sqrt{\frac{1}{AB} + \frac{1}{Ab} + \frac{1}{aB} + \frac{1}{ab}}$$

maroon with no. (i) above from only one family of 155  $F_2$  plants (Punnett, 1923) has also been indicated at the end, though not made use of in locating the gene on the map. The first three characters have been repeatedly described by Bateson & Punnett. But, with a view to making this article self-contained and for facility in reading it, it has been thought proper briefly to reproduce their description below:

(i) *Dark-light axil.* 'Of the plants with white and coloured flowers in the sweet pea those with white (true whites have pale seed coats) have light (i.e. green) axils, whereas those with coloured flowers may have either dark or light axils.'

(ii) *Fertile-sterile (contabescent) anthers.* 'Contabescent anthers are sterile. Their chromosomes divide normally up to the reduction division when they form shapeless knots and fail to divide. Plants with sterile anthers are normal and fertile on the female side.'

(iii) *Normal-cretin flower.* 'The "Cretin" is a small deformed flower with varying structure of petals. In such a flower the standard fails to become elevated, the keel is cleft distally so that anthers are partially protruded, while the stigma sticks out far beyond the petals and is carried on in the line of the carpels instead of being abruptly bent at right angles to them as in the normal flower.'

(iv) *Purple-maroon flower colour.* The terms are self explanatory.

The data are presented in the Appendix, and are in all on 169 families consisting of 14,652 plants studied over a period of 25 years, from 1904 to 1928.\* On considering families observed for three characters separately for every pair, their number is increased to 225 comprising 19,306 plants. They consist of 138  $F_2$  and 87 later ( $F_3$  onwards) progenies. Classified according to character pairs they stand as given in Table 1.

Table 1. *Classification of plants according to character pairs, type of linkage, and families*

Character pair	Coupling		Repulsion	
	No. of families	No. of plants	No. of families	No. of plants
Dark-light axil	73	6635	35	4414
Fertile-sterile anthers	28	2239	21	1440
Normal-cretin flower	16	792	52	3786
Total	117	9666	108	9640

The number of plants observed (19,306) is nearly three times as large as that utilized by Bridges (6704), and should, on any consideration, be thought sufficient to give correct estimates of cross-over percentages.

It will, however, be seen that a large proportion of families belong to generations beyond  $F_2$ . Besides, even in the case of several  $F_2$  families, the nature of the mating could not be made out from the records. Since it is not ordinarily possible to detect in such generations the type of mating, coupling or repulsion, of parents from which a progeny has been derived when segregation is on the border-line, one may wonder as to how they could all be rightly grouped under the two categories with confidence. In fact, this was one of the main difficulties in Punnett's (1913) classification of certain dubious families. For example, in the  $dfN \times DFn$  mating (Table 11)† he put the four  $F_4$  progenies, viz. 113/13, 114/13, 117/13 and 123/13, in line with the two  $F_2$  progenies, 52/13 and 53/13. As can be seen from their present classification (vide Appendix) the progenies 114/13 and 117/13 had been misclassified so far as character pairs F-N and D-N were concerned. Instead of being grouped under 'coupling' they were put with the rest of the  $F_2$  and  $F_4$

\* These include all the families presented by Bateson, Saunders & Punnett (1905, 1908) and Punnett (1913). The later data were tabulated from Punnett's records kept at the Department of Genetics, Cambridge.

† In Punnett (1913, Table XI) this has been misprinted as  $DFn \times DFn$ .

families under 'repulsion'. It is for this reason that his anticipation of repulsion between N and F and N and D was not realized for the whole group; indeed, the balancing effect of the mixture of 'coupling' and 'repulsion' families made him further suspect a normal segregation of 9 : 3 : 3 : 1. The fact that they were in coupling, and could not be otherwise, is proved by their very low  $\chi^2$ 's and homogeneity with the rest of the families in their present setting. Similarly, progenies 71/12 and 91/12 in the mating Nf  $\times$  nF (Table 9), though classed separately, were suspected to indicate, both individually and together, a normal segregation. Difficulty was also experienced in identifying with confidence the families 47/13 and 136/13 (Table 10), though they were put under 'coupling'. All these families have fitted in the present groupings so well that there is little doubt left about the types of mating they came from.

The correct allocation of the mating type to every family has been rendered possible by the feature of the scoring method whereby the aggregate score of a family is zero at the correct value of the estimate. As a corollary of this characteristic it follows that a body of data scored at 50% for the appropriate mating gives a positive score, whereas when scored for the wrong one gives a negative one. This serves as a very useful means to detect the mating when one is in doubt about it. This method has been invariably used for the entire classification of families in the third and subsequent filial generations, and surprisingly enough all the eighty-seven and more families so classified have accommodated themselves very well in the respective groups in which they were put, as can be seen from their low (non-significant)  $\chi^2$ 's (see Appendix).

After tabulation, totals were struck of all progenies falling under each category, and the scoring method was further employed to work out recombination percentages. The *modus operandi* of estimation may be briefly described as follows:

If  $c$  represents the recombination percentage and  $N$  the number in the progeny, the expectations, in terms of  $c$  in a dihybrid intercross, of the three classes, viz. (1) double dominants, (2) single dominants, and (3) double recessives are, respectively,

$$\frac{N}{4}(3-2c+c^2), \quad \frac{N}{4}(2c-c^2), \quad \frac{N}{4}(1-c)^2;$$

and the differentials of the expected frequencies are

$$-\frac{N}{2}(1-c), \quad +\frac{N}{2}(1-c), \quad -\frac{N}{2}(1-c).$$

The scores are obtained by dividing the differentials by expectations and changing sign, as below:

$$\frac{2(1-c)}{2+(1-c)^2}, \quad -\frac{2(1-c)}{2c-c^2}, \quad \frac{2}{(1-c)} \quad \text{for coupling;}$$

$$-\frac{2c}{2+c^2}, \quad \frac{2c}{1-c^2}, \quad -\frac{2}{c} \quad \text{for repulsion.}$$

The amount of Information per individual, which is the summation of squares of scores multiplied by expectations over the three classes is, in this case,

$$\frac{2+4(1-c)^2}{(2c-c^2)(3-2c+c^2)} \quad \text{for coupling,} \quad \text{and} \quad \frac{2+4c^2}{(1-c^2)(2+c^2)} \quad \text{for repulsion.}$$

To facilitate scoring Fisher has prepared tables of scores for various recombination percentages from 1 to 50 for intercross progenies. The scores are given at unit intervals and can be conveniently interpolated by the usual method (Fisher & Yates, 1938

'Introduction—Table XXXIII—Random Numbers—Interpolation'). Except in cases where extraordinary accuracy is aimed at a four-point interpolation is deemed satisfactory. For such an interpolated value the information is given by the formula

$$100D + \left\{1 - \frac{1}{6}(1 - 3\theta^2)\delta^2 u_{x+1}\right\} - \left\{1 - \frac{1}{6}(1 - 3\phi^2)\delta^2 u_x\right\},$$

where  $D$  is the positive difference between scores at  $x$  and  $x+1$ ,  $\theta$  = fraction beyond  $x$ ,  $\phi = 1 - \theta$ ,  $\delta^2$  = central second difference.

The estimation is done through a series of trials beginning with 50%. The total score of a family at any trial value divided by its Information gives the correction factor to arrive at a more correct value. This procedure is repeated until a satisfactory estimate giving a score almost zero\* is reached. With experience and proper judgement the number of such trials can be reduced to only about two or three. Primary estimates made in this way for the three segments  $B_1-B_2$ ,  $B_1-B_3$  and  $B_2-B_3$  and the scores and Information thereon are given below:

Segment	Recombination percentages	Score	Information
$B_1-B_2$	3.03	-85.5042894	
$B_1-B_3$	28.00	-113.5092077	
$B_2-B_3$	28.42	-0.0601295	7216.16941

At this stage, Kosambi's formula is employed to improve the estimates of the two smaller segments,  $B_1-B_2$  and  $B_1-B_3$ , with information obtained from the larger segment,  $B_2-B_3$ , which contains them both. Kosambi's formula enunciating the relationship between two smaller segments (say  $y_1$  and  $y_2$ ) with a bigger one containing them ( $y_{12}$ ) is

$$y_{12} = \frac{y_1 + y_2}{1 + 4y_1y_2}.$$

This relationship also provides factors with which to multiply the score of the larger segment to obtain its contribution of scores to the smaller ones. The factors are obtained from the following corollaries of the above formula:

$$\frac{\partial y_{12}}{\partial y_1} = \frac{1 - 4y_2^2}{(1 + 4y_1y_2)^2}, \quad \frac{\partial y_{12}}{\partial y_2} = \frac{1 - 4y_1^2}{(1 + 4y_1y_2)^2}.$$

The following are the actual calculations in this case:

Assuming  $B_1-B_2$  to be 3.03 and  $B_1-B_3$  to be 28.000,  $B_2-B_3$  by Kosambi's formula would be 30.0115288. The approximate score at this value of the estimate is +114.7872849, i.e.  $(-0.0601295 + 7216.16941 \times 0.015915288)$ , †

$$\frac{\partial y_{12}}{\partial y_1} = 0.6420812, \quad \frac{\partial y_{12}}{\partial y_2} = 0.9319977.$$

The scores and Information for the two segments are:

	Score	Information
$B_1B_2$		
From $B_1-B_2$	-85.5042894	222256.4513561
From $B_2-B_3$	+73.7027575	2954.8783446
Total	-11.8015318	225211.3297007
$B_1B_3$		
From $B_1-B_2$	-113.5092077	9259.3733681
From $B_2-B_3$	+106.9814855	6225.7170433
Total	-6.5277222	15485.0954114

\* When improvement of the estimate of a smaller segment with the help of information from a larger segment is aimed at, as is done in this paper, it may be preferable to bring the score approximately to such a figure as would serve to nullify the score that would be derived for the segment in question from the larger segment.

† This is the difference between the provisional (0.2842) and the calculated (0.300115288) estimate of  $y_{12}$ .

The product term of Information for the two segments is obtained by multiplying the information for  $B_2-B_3$  at 30-0115288 by the product of the two factors derived above:

$$\begin{aligned} &= 7690.047928 \times 0.642812 \times 0.9319977 \\ &= 4289.0834065. \end{aligned}$$

The adjustments of estimates for the two segments  $p$  for  $B_1-B_2$  and  $q$  for  $B_1-B_3$  are now calculated by solving the following simultaneous equations:

$$\begin{aligned} 225211.3297007p + 4289.0834065q &= +11.8015318,* \\ 4289.0834065p + 15485.0954114q &= + 6.5277222,* \end{aligned}$$

whereby  $p=0.00004$  and  $q=0.0041$ .

The new values obtained by applying the adjustments are

$$B_1B_2=3.034; B_1B_3=28.041; B_2B_3=30.052 \text{ (newly calculated).}$$

On repeating the procedure twice over the following scores and Information were obtained for the two smaller segments:

	Score	Information
	$B_1B_2$	
From $B_1B_2$	- 75.5293950	221830.969681
From $B_2B_3$	+ 75.5292207	2945.253238
Total	- 0.0001743	224776.222919
	$B_1B_3$	
From $B_1B_3$	- 109.7725205	9248.513214
From $B_2B_3$	- + 109.7780603	6221.910222
Total	+ 0.005398	15470.423436

Here the scores are very nearly reduced to zero. It is, therefore, not worth while to work out further adjustments. Standard errors of estimates are calculated by the following formulæ:

$$\begin{aligned} V(y_1) &= \frac{I_{22}}{\Delta}, & V(y_2) &= \frac{I_{11}}{\Delta}. \\ V(y_{12}) &= \left(\frac{\partial y_{12}}{\partial y_1}\right)^2 \frac{I_{22}}{\Delta} - 2 \frac{\partial y_{12}}{\partial y_1} \frac{\partial y_{12}}{\partial y_2} \frac{P}{\Delta} + \left(\frac{\partial y_{12}}{\partial y_2}\right)^2 \frac{I_{11}}{\Delta}. \\ I_{11} &= \text{total Information for } y_1. \\ I_{22} &= \text{total Information for } y_2. \\ P &= \text{product term of Information to be calculated as above.} \\ \Delta &= I_{11}I_{22} - P^2. \end{aligned}$$

On arriving at final estimates, map lengths in 'centimorgans' were found by referring to Fisher & Yates (1938, Table No. XII, Transformation of 'r' to 'z' Interclass).

### THE RESULTS

The entire results obtained are set out in Table 2.

Table 2. *Recombination percentages, scores, Information and map lengths for the three segments in Punnett's 'B' chromosome*

Segment	Recombination percentage	Map length (cm.)	Score	Information
$B_1B_2$	3.0359 ± 0.21	3.04 ± 0.21	- 0.0001743	224776.222919
$B_1B_3$	28.0410 ± 0.81	31.70 ± 1.18	- 0.005398	15470.423436
$B_2B_3$	30.0535 ± 0.75	34.74 ± 1.17	+ 117.8159978	7166.402000

The standard error (s.e.) of map length ( $x$ ) corresponding to the r.f. ( $y$ ) is calculated by using the formula

$$\text{s.e. } (x) = \text{s.e. } (y) / (1 - 4y^2).$$

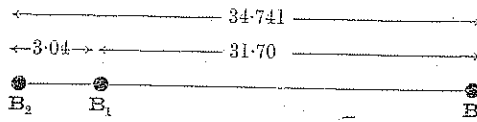
\* Written after reversing the signs.

## Genetical map of Punnett's 'B' chromosome in the sweet pea

Table 3. Estimates of recombination percentages arrived at by Punnett, Bridges and the author

Segment		Punnett	Bridges	Author
B <sub>1</sub> -B <sub>2</sub>	C	6.25	4.4	3.04
	R	Not ascertained	4.9	
B <sub>1</sub> -B <sub>3</sub>	C	25.00	28.0	28.04
	R		27.0	
B <sub>2</sub> -B <sub>3</sub>	C	25.00	33.0	30.05
	R		28.0	

It will be seen from the table that even Bridges's values are somewhat different from those estimated by the writer, though the differences are not very striking to a casual observer. The value of such differences can be appreciated only when one considers the effect they would have in genetical experiments. For example, in a repulsion intercross the expectation of a double recessive with 4.4% R.P. would be 1 in 2056, while that with 3.04% R.P. would be 1 in 4328, not an inconsiderable difference when one is confronted with the question of obtaining such a plant. Apart from its utility or consequences a genuine difference, however small, should be welcome as a scientific fact, especially when it is brought out with the support of enormous convincing evidence. Looking at the figures from this point of view the new estimations appear justified. The map drawn on the basis of the new estimates is given below:



After obtaining the estimates,  $\chi^2$  tests were made, first, to judge whether it was appropriate to use Kosambi's formula in this case, and secondly, to test the homogeneity of the data used to derive the estimates.

The first test is essential since Kosambi's formula may not be applicable to all types of cases, and unless it is proved that the estimates are in agreement with the formula they would not be reliable. The  $\chi^2$ 's for this test were calculated by squaring the score of each segment and dividing by the amount of Information at the final estimated value. Their total is the  $\chi^2$  to judge the conformity of the estimates to the formula. The figures are produced in Table 4. The value of  $\chi^2$ , though high for one degree of freedom, is not sufficiently high (being less than 3.84, the five per cent value) to contradict the formula.

Table 4. Values of  $\chi^2$  to test agreement with Kosambi's formula

Segment	Score	Information	$\chi^2$
B <sub>1</sub> B <sub>2</sub>	-75.529395	221830.969681	0.026
B <sub>1</sub> B <sub>3</sub>	-109.7725205	9248.513214	1.303
B <sub>2</sub> B <sub>3</sub>	117.8159978	7166.402000	1.937
Total			3.266

The test of homogeneity was done in the usual manner,  $\chi^2$  for each family at the final estimated value having been calculated exactly as in the above case. The figures of  $\chi^2$  for individual families and probabilities are given in the Appendix. The  $\chi^2$  of heterogeneity between mating types, and between families in each type are set out in Table 5.



Table 5.  $\chi^2$ 's and probabilities for heterogeneity between mating types, and between families within mating types

Factor pair	Between mating types		Mating type	Between families within mating types		
	$\chi^2$	P		D.F.	$\chi^2$	P
$B_1-B_2$	0.00	> 0.99	Coupling	72	82.55	0.19*
			Repulsion	34	27.09	0.80*
$B_1-B_3$	0.37	0.5~0.7	Coupling	27	27.77	0.3~0.5
			Repulsion	20	15.59	0.7~0.8
$B_2-B_3$	0.09	0.7~0.8	Coupling	15	8.18	0.9~0.95
			Repulsion	51	48.13	0.60*

\* Probability in terms of the normal deviate obtained by the formula  $\sqrt{2\chi^2} - \sqrt{(2n-1)}$  for one degree of freedom.

The test of homogeneity is very important in this case, as the number of progenies handled is very large, and they come from different filial generations and mating types. It is much more so because, as previously stated, the mating types in a majority of cases have been fixed on the basis of their total scores at 50% level. Since the method is totally unconventional it is possible that its efficacy may be doubted, particularly with the data involving families which Punnett (1913) had found very heterogeneous and irregular, unless proved by this test.

It is very striking, however, that every mating type, even with its large number of families, is clearly homogeneous. The same is also the case with the two mating types for each pair of characters. Further, a glance at the  $\chi^2$ 's of individual families would show that they are almost unanimous in indicating the cross-over values estimated in the respective cases. Only one family, viz. 22/15 ( $B_1B_2$  repulsion) shows a  $\chi^2$  beyond the 1% level of probability. The reasons for including it are:

*Family no. 22/15.* The high  $\chi^2$  in this case is the result of one plant having appeared in the double recessive class instead of the expected (0.006) fraction of a plant. There are in all 4414 plants in the  $B_1B_2$  repulsion group and, with estimated r.p. of 3.03%, just one double recessive plant is expected to occur. Since the group consists of thirty-four families in all it has to appear in some one of these and, when it does, it is *but to be expected* that it will upset the balance of that particular family. When it is borne in mind that the double recessive plant holds the balance between an accurate and inaccurate estimate in such a case of close linkage in repulsion it would be seen that it would have been inadmissible to leave out the family.

Out of all the progenies found in the records five  $B_1-B_2$  progenies appeared to be irregular. All these belong to the coupling group as can be easily made out from the distribution of their frequencies, and as confirmed by their scores at 50%. Yet they were far too different from other sister progenies entered in the Appendix, as is indicated by their very high  $\chi^2$ 's at the recombination percentage estimated from the rest of the families, i.e. 3.10%. Their particulars are set out in Table 6.

Table 6. Particulars of five  $B_1-B_2$  (coupling) families which were originally found to be irregular

Serial no.	Family ref. no.	Frequencies				$\chi^2$ at 3.10%
		$B_1B_2$	$B_1b_2$	$b_1B_2$	$b_1b_2$	
1	351/6G	14	—	4	7	13.29
2	96/22	50	1	11	10	46.35
3	57/27	185	3	49	66	199.31
4	58/27	314	7	100	152	449.54
5	75/27	69	1	18	40	54.91

None of these five came from the lot that Punnett found difficult to classify. On the other hand, every one of those that were not amenable to grouping and interpretation is, with grouping altered where necessary, in perfect harmony with the rest in the particular class wherein it is placed, in indicating the estimated recombination value.

In the case of the last four progenies the segregation for the 'fertile-sterile' character was more or less in conformity with the expected 3:1 ratio, but that for 'dark-light' was unsatisfactory. Closer examination of the classification showed that an additional factor 'self-marbled', located in the 'P' chromosome, was also involved. It was further noticed that all 'marbled' plants had been classified only as 'light' and that the segregation between 'self-dark', 'self-light' and 'marbled-light' bore a relation of 9:3:4. It was clear, therefore, that some of the 'marbled' plants contained the gene for 'Dark axil', and that it could not manifest itself due to the absence of the gene for 'self' colour. On such a hypothesis, after excluding the 'marbled' plants, the remaining ones should indicate a recombination of about 3%, which they did very well. These families were, therefore, included in the B<sub>1</sub>-B<sub>2</sub> coupling group, and now show very low  $\chi^2$ 's (vide Appendix) at the joint estimate of 3.04%.

In the case of the first family, viz. 351/6G, the difficulty has remained unsolved. The segregation for the two factors under study is as follows:

dark-fertile	dark-sterile	light-fertile	light-sterile
14	0	4	7

The ratio for dark-light is 9:7, while that for fertile-sterile is 3:1. Clearly there is an additional gene involved in the segregation of the dark and light characters. The estimate made on the basis of these individual segregations is 16.74%, which is very different from the estimate of 3.0% arrived at from other progenies. The actual classification by Punnett is as below:

	fertile	sterile
Full purple: dark	7	—
light	1	4
Light purple: dark	—	—
light	1	1
Full red: dark	4	—
light	2	2
Pale red: dark	3	—
light	—	—

It can be seen from the distribution that there is no consistent absence of either 'dark' or 'light' plants in either 'full' or 'pale (or light)' colours—purple or red, which might serve to eliminate the effect of the second gene on the dark-light segregation. The progeny also being very small, its exclusion will not involve appreciable loss of Information; so it has been omitted from consideration.

In conclusion it may be stated that every one of the dubious families is satisfactorily classified in the present analysis. The striking homogeneity found in the large amount of data studied speaks highly both for the remarkable accuracy of the data and for the efficiency of the scoring system of estimation.

*The purple-maroon (B<sub>4</sub>) character*

It may be stated in passing that the family 50/22, which Punnett studied (1923) and found difficult to interpret, indicates a recombination value of 19.3% between the characters, dark-light axil and purple-maroon flower colour. For want of more data it is not possible to give a definite estimate of it. Neither can the position of the gene be located, as the character has not been studied with others in the chromosome.

## SUMMARY

An improved genetical map of Punnett's 'B' chromosome in the sweet pea is presented on the basis of Punnett's data recorded from 1904 to 1928 on the characters (1) dark-light axil, (2) fertile-sterile anthers and (3) normal-cretin flowers. The data have been classified and analysed by Fisher's scoring method estimating linkage. The procedure of estimation has been briefly illustrated.

Prof. R. C. Punnett, F.R.S., has very generously left his original notebooks at the Department of Genetics, Cambridge, for the use of students. I am very thankful to him for the data drawn from his books and also for his encouragement in this work. It is difficult to express my deep gratitude to Prof. R. A. Fisher, Sc.D., F.R.S., at whose instance the study was undertaken and under whose guidance it was accomplished. I am also thankful to Dr D. S. Falconer and Mr T. C. Carter, M.A., for kindly going through the paper and making useful suggestions.

## REFERENCES

- BATESON, W. (1909). *Mendel's Principles of Heredity*. Camb. Univ. Press.
- BATESON, W. & PUNNETT, R. C. (1911). On the gametic series involving reduplication of certain terms. *J. Genet.* 1, 293-302.
- BATESON, W., SAUNDERS, E. R. & PUNNETT, R. C. (1905). Experimental studies in the physiology of heredity. *2nd Rep. Evol. Comm. Roy. Soc.* pp. 80-99.
- BATESON, W., SAUNDERS, E. R. & PUNNETT, R. C. (1908). Experimental studies in the physiology of heredity. *4th Rep. Evol. Comm. Roy. Soc.* pp. 2-16.
- BRIDGES, C. B. (1914). The chromosome hypothesis of linkage applied to cases in sweet peas and *Primula*. *Amer. Nat.* 48, 524-34.
- IMMER, F. R. (1931). The efficiency of the correlation coefficient for estimating linkage intensities. *Amer. Nat.* 65, 567-72.
- FISHER, R. A. (1946 a). *Statistical Methods for Research Workers*. 10th ed. Edinburgh: Oliver and Boyd.
- FISHER, R. A. (1946 b). A system of scoring linkage data with special reference to pied factors in mice. *Amer. Nat.* 80, 568-78.
- FISHER, R. A. & YATES, F. (1938). *Statistical Tables*. Edinburgh: Oliver and Boyd.
- MATHER, K. (1938). *The Measurement of Linkage in Heredity*. Methuen.
- PUNNETT, R. C. (1913). Reduplication series in sweet peas. *J. Genet.* 3, 77-103.
- PUNNETT, R. C. (1923). Linkage in the sweet pea. *J. Genet.* 13, 101-23.
- PUNNETT, R. C. (1925). *Lathyrus odoratus*. *Bibliogr. genet.* 1, 69-82.
- PUNNETT, R. C. (1927). Linkage groups and chromosome number in *Lathyrus*. *Proc. Roy. Soc. B.* 102, 236-8.
- PUNNETT, R. C. (1932). Further studies of linkage in the sweet pea. *J. Genet.* 26, 97-112.
- TAKEZAKI, Y. (1925). On a new method of computation of the cross-over value from  $F_2$  zygotic series. *J. Genet.* 3, 105-13.
- YULE, G. U. (1900). On the association of attributes in statistics. *Philos. Trans. A*, 194, 257-319.
- YULE, G. U. (1912). On the methods of measuring association between two attributes. *J. Roy. Statist. Soc.* 75, 579-652.

## APPENDIX

Table showing the segregation of the characters, dark-light axil ( $B_1$ )  
and fertile-sterile anthers ( $B_2$ )

Family no. From the <i>Rep. Evol.</i> <i>Comm. Roy. Soc.</i> 1905, p. 91	Coupling				Total	$\chi^2$ (at 3-0359)
	$B_1B_2$	$B_1b_2$	$b_1B_2$	$b_1b_2$		
141/4F	121	5	2	39	167	0.82
142/4F	25	2	—	9	36	0.75
143/4F	29	—	—	11	40	1.33
144/4F	43	—	2	17	62	0.00
181/4G	23	—	2	12	37	0.51
120/5F	40	1	3	9	53	4.14
316/5G	41	1	2	20	64	0.40
317/5G	53	1	1	14	69	0.00
342/6G	50	1	—	11	62	0.27
348/6G	20	1	1	5	27	1.93
349/6G	22	1	2	6	31	4.88
350/6G	57	3	1	21	82	0.89
352/6G	28	2	1	6	37	3.64
353/6G	7	1	0	6	14	0.52
354/6G	22	1	2	5	30	5.34
358/6G	17	—	—	8	25	0.94
1/9	15	—	1	8	24	0.05
2/9	49	2	2	11	64	2.66
62/10	62	1	1	18	82	0.06
87/11	114	2	3	30	149	0.13
102/12	69	3	3	24	99	3.06
104/12	24	—	—	5	29	0.72
108/12	32	—	—	11	43	1.36
109/12	7	—	—	1	8	0.17
52/13	14	—	—	7	21	0.81
53/13	19	1	1	7	28	1.57
113/13*	23	3	—	15	41	1.95
114/13*	40	—	2	10	52	0.19
117/13*	38	1	1	10	50	0.22
123/13*	25	—	—	7	32	0.91
71/21	27	1	—	9	37	0.01
72/21	318	5	2	110	435	3.08
73/21	100	2	1	28	131	0.16
96/21	71	—	1	22	94	1.16
99/21	38	—	—	14	52	1.71
100/21	46	—	2	5	53	0.34
73/22	15	—	1	6	22	0.17
75/22	4	1	—	9	14	0.27
76/22	25	1	—	9	35	0.01
77/22	34	3	—	15	52	1.13
92/22*	41	1	4	17	63	4.81
93/22*	58	2	1	21	82	0.09
96/22*	19	1	—	7	27	0.03
97/22*	50	1	—	9	60	0.19
99/22*	52	2	—	10	64	0.05
60/24*	24	—	—	9	33	1.09
64/24*	22	—	—	8	30*	0.98
65/24	31	1	3	10	45	5.33
43/25	27	3	—	6	36	3.82
44/25	110	—	2	18	130	0.46
45/25	53	2	—	16	71	0.00
46/25	101	1	2	26	130	0.12
47/25	106	1	4	39	150	0.03
48/25	86	1	1	32	120	0.85
49/25	127	2	2	40	171	0.24
50/25	147	—	2	44	193	2.39
96/25	131	2	5	32	170	1.00
	70	—	—	20	90	2.59

\* Families in generations later than  $F_2$ .

Family no.	B <sub>1</sub> B <sub>2</sub>	B <sub>1</sub> b <sub>2</sub>	b <sub>1</sub> B <sub>2</sub>	b <sub>1</sub> b <sub>2</sub>	Total	$\chi^2$ (at 3-0359)
97/25	103	1	6	61	171	0.20
98/25	67	—	—	17	84	2.28
51/27	148	2	7	45	202	1.55
52/27	136	5	1	53	195	0.00
53/27	108	3	1	35	147	0.04
54/27	114	2	3	39	158	0.00
57/27	185	3	3	46	237	0.06
58/27	314	7	7	106	434	0.06
63/27	113	—	2	33	148	1.27
64/27	103	1	—	33	137	2.39
65/27	157	1	3	50	211	0.86
66/27	39	1	—	11	51	0.15
71/27	87	2	1	31	121	0.15
73/27	60	—	4	23	87	0.64
75/27	69	1	2	32	104	0.59
73	4865	95	106	1569	6635	0.025
					Deviation $\chi^2$ (1)	0.025
					Heterogeneity $\chi^2$ (72)	82.545
					Total	82.570

*Repulsion*

Family no.	B <sub>1</sub> B <sub>2</sub>	B <sub>1</sub> b <sub>2</sub>	b <sub>1</sub> B <sub>2</sub>	b <sub>1</sub> b <sub>2</sub>	Total	$\chi^2$ (at 3-0359)
84/8	95	45	44	—	184	0.03
95/8	31	12	21	—	64	0.02
97/8	37	29	31	—	97	0.07
106/8	53	17	22	—	92	0.01
107/8	214	95	102	—	411	0.07
108/8	43	22	29	—	94	0.03
109/8	81	50	58	—	189	0.09
110/8	72	27	39	—	138	0.02
111/8	110	55	49	—	214	0.04
112/8	177	73	99	—	349	0.07
56/10	27	17	18	—	62	0.03
60/10	126	59	69	—	254	0.06
48/12	49	28	27	—	104	0.03
48/13	129	55	55	—†	239	0.03
49/13	24	17	15	—	56	0.03
50/13	17	14	11	—	42	0.02
51/13	50	28	25	—	103	0.03
30/14	27	13	19	—	59	0.02
31/14	62	24	35	—	121	0.02
32/14	58	27	18	—	103	0.01
33/14	40	17	20	—	77	0.01
22/15*	85	41	30	1	157	26.14
79/15	104	50	46	—	200	0.04
80/15	121	50	60	—	231	0.04
54/16	15	12	6	—	33	0.01
55/16	59	20	19	—	98	0.03
59/16	31	7	11	—	49	0.00
127/19	19	8	13	—	40	0.01
57/20*	55	20	20	—	95	0.06
58/20*	63	19	34	—	116	0.01
75/20*	6	7	3	—	16	0.01
79/20*	17	5	6	—	28	0.00
87/20*	71	26	50	—	147	0.04
92/21*	42	31	27	—	100	0.05
93/21*	31	10	11	—	52	0.03
35	2241	1030	1142	1	4414	0.00
					Deviation $\chi^2$ (1)	0.00
					Heterogeneity $\chi^2$ (34)	27.09
					Total	27.09

\* Families in generations later than  $F_2$ .

† Two plants in this class have been eliminated on discovering a remark by Punnett in the record book that they were 'not good plants', as without them the family conforms to the degree of linkage indicated by the rest of the families.

## Genetical map of Punnett's 'B' chromosome in the sweet pea

Table showing the segregation of the characters, dark-light axil ( $B_1$ ) and normal-cretin flower ( $B_3$ )

Family no.	Coupling				Total	$\chi^2$ (at 28-041)
	$B_1B_3$	$B_1b_3$	$b_1B_3$	$b_1b_3$		
48/13	151	33	31	26	241	1.30
49/13	37	4	5	10	46	2.60
50/13	27	4	5	6	42	0.20
51/13	67	11	8	17	103	2.41
114/13*	32	7	7	4	50	0.91
117/13*	22	3	4	3	32	0.00
30/14	32	8	12	7	59	2.43
31/14	76	10	21	14	121	0.25
32/14	71	14	8	10	103	0.02
33/14	52	5	9	11	77	1.33
22/15*	104	22	14	17	157	0.00
36/15*	44	7	5	6	62	0.17
39/15*	15	4	1	1	21	0.18
55/15*	20	3	8	2	33	2.27
57/15*	14	1	9	3	27	2.01
59/15*	30	—	4	3	37	1.48
60/15*	14	3	4	1	22	1.39
79/15	127	27	23	23	200	0.24
80/15	140	31	30	30	231	0.45
54/16	26	1	5	1	33	0.00
55/16	69	10	14	5	98	1.02
59/16	32	6	6	5	49	0.08
127/19	24	3	7	6	40	0.00
57/20*	61	14	7	13	95	0.21
79/20*	21	1	3	3	28	0.75
87/20*	82	15	23	27	147	0.15
60/24*	19	3	2	6	30	1.58
64/24*	22	10	8	5	45	4.92
28	1431	260	283	265	2239	0.58
						0.58
						27.77
						28.35

\* Families in generations later than  $F_3$ .

Family no.	Repulsion				Total	$\chi^2$ (at 28-041)
	$B_1B_3$	$B_1b_3$	$b_1B_3$	$b_1b_3$		
52/13	13	7	7	1	28	0.13
53/13	19	7	12	3	41	4.52
113/13*	25	15	11	1	52	0.06
119/13*	48	27	22	3	100	0.17
123/13*	20	8	8	1	37	0.17
34/14	54	21	14	3	92	1.93
35/14	22	7	9	1	39	0.24
36/14	20	5	3	—	28	0.01
37/14	52	24	21	1	98	0.33
38/14	199	63	79	6	347	0.24
25/15*	31	16	20	1	68	0.50
29/15*	31	11	19	3	64	1.71
30/15*	22	9	10	—	41	0.64
38/15*	15	6	6	—	27	0.33
40/15*	21	7	3	1	32	1.16
41/15*	41	12	8	1	62	0.48
56/15*	39	24	19	1	83	0.72
58/15*	20	7	16	—	43	1.42
61/15*	11	4	10	1	26	0.08
58/20*	71	11	33	1	116	0.00
75/20*	9	4	2	1	16	1.81
21	783	295	332	30	1440	1.09
						1.09
						15.59
						16.68

\* Families in generations later than  $F_3$ .

Table showing the segregation of the characters fertile-sterile anthers ( $B_2$ )  
and normal-cretin flower ( $B_3$ )

<i>Coupling</i>						
Family no.	$B_2B_3$	$B_2b_3$	$b_2B_3$	$b_2b_3$	Total	$\chi^2$ (at 30-0535)
76/12*	90	30	20	34	174	0.22
91/12*	20	5	4	3	32	0.23
38/13*	16	9	8	12	45	0.08
39/13*	27	9	17	14	67	1.23
40/13*	20	4	7	9	40	0.38
41/13*	12	6	6	9	33	0.00
47/13*	23	5	4	4	36	0.00
114/13*	32	7	7	4	50	0.51
117/13*	22	3	4	3	32	0.03
136/13*	6	1	1	1	9	0.02
16/15*	18	2	3	2	25	0.06
58/20*	91	6	13	6	116	0.73
75/20*	7	2	4	3	16	0.33
60/24*	19	3	2	6	30	2.14
64/24*	25	9	5	6	45	0.36
17/28*	28	1	6	7	42	2.10
16	456	102	111	123	792	0.24
					Deviation $\chi^2$ (1)	0.24
					Heterogeneity $\chi^2$ (15)	8.18
					Total	8.42

\* Families in generations later than  $F_2$ .

<i>Repulsion</i>						
Family no.	$B_2B_3$	$B_2b_3$	$b_2B_3$	$b_2b_3$	Total	$\chi^2$ (at 30-0535)
5/9	51	33	30	1	115	2.89
72/11	26	10	14	1	51	0.05
73/11	21	12	12	1	46	0.18
74/11	24	8	9	—	41	0.37
75/11	22	4	4	2	32	5.15
76/11	30	5	12	1	48	0.27
77/11	78	32	43	3	156	0.26
78/11	59	24	15	—	98	0.64
79/11	25	15	12	2	54	0.11
70/12*	37	34	22	1	94	3.33
71/12*	22	9	9	2	42	1.15
80/12*	8	1	6	—	15	0.29
82/12*	12	8	9	—	29	1.68
86/12*	17	6	10	—	33	0.79
87/12*	49	16	11	1	77	0.09
88/12*	11	6	7	—	24	0.99
90/12*	24	13	11	—	48	1.35
92/12*	11	5	7	—	23	0.80
99/13*	12	7	3	—	22	0.36
43/13*	20	4	7	1	32	0.66
45/13*	37	10	9	2	58	1.80
46/13*	18	7	7	2	34	2.02
48/13	133	49	51	8	241	2.16
49/13	25	17	14	—	56	2.54
50/13	19	13	9	1	42	0.11
51/13	48	27	27	1	103	1.81
52/13	13	7	7	1	28	0.04
53/13	17	6	14	4	41	6.57
113/13*	27	15	9	1	52	0.03
123/13*	19	8	9	1	37	0.02
30/14	32	12	14	1	59	0.02
31/14	74	23	23	1	121	0.04
32/14	52	24	27	—	103	2.75
33/14	44	16	17	—	77	0.88
22/15*	81	34	37	5	157	0.48
79/15	105	45	45	5	200	0.06
80/15	123	58	47	3	231	0.69
54/16	20	1	11	1	33	0.49
55/16	65	13	18	2	98	1.09

*Repulsion (cont.)*

Family no.	$B_1B_3$	$B_2b_3$	$b_2B_3$	$b_3b_3$	Total	$\chi^2$ (at 30-0535)
57/16	13	4	5	—	22	0.18
58/16	20	7	7	—	34	0.29
59/16	31	11	7	—	49	0.15
127/19	24	8	7	1	40	0.27
128/19	49	22	23	2	96	0.05
57/20	51	24	17	3	95	0.48
79/20	19	4	5	—	28	0.00
80/20	37	14	17	1	69	0.12
81/20*	43	24	24	1	92	1.42
82/20*	26	15	13	1	55	0.28
83/20*	58	25	31	3	117	0.00
87/20*	80	41	25	1	147	1.07
121/21*	47	17	26	1	91	0.60
52	2009	853	855	69	3786	1.79
				Deviation $\chi^2$ (1)		1.79
				Heterogeneity $\chi^2$ (51)		48.13
				Total		49.92

\* Families in generations later than  $F_2$ .