STOCKS FOR DETECTING LINKAGE IN THE MOUSE, AND THE THEORY OF THEIR DESIGN

By T. C. CARTER

Medical Research Council, Institute of Animal Genetics, Edinburgh

AND D. S. FALCONER

Animal Breeding and Genetics Research Organization, Institute of Animal Genetics, Edinburgh

(With Three Text-figures)

1. Introduction

The value of an animal as material for genetic research increases as the knowledge of its chromosome map is extended. The house mouse (*Mus musculus* L.) is in this respect by far the best known of the mammals, but its value is nevertheless still seriously limited by the incomplete knowledge of its chromosome map and by the lack of suitable marker genes. This limitation is felt particularly in the study of chromosome rearrangements and of mutation, and it has prevented the development of inversion stocks comparable with those upon which many of the *Drosophila* techniques depend. The recent success of Slizynski (1949) in describing the cytological picture of the mouse's chromosomes gives promise of a fruitful synthesis of genetical and cytological studies, and makes the completion of the genetical map all the more urgent.

Recognition of the need to add to the chromosome map has often led the discoverer of a new mutant in the mouse to test it for linkage with as many of the previously known genes as possible. In this way considerable progress in mapping the twenty chromosomes has been made, and already twelve linkage groups are recognized, though not all have suitable marker genes on them and it is far from certain that the twelve recognized groups represent twelve different chromosomes. There are now, however, so many genes known that it has become a formidable task to test a newly discovered mutant fully for linkage, and most investigators have been content with tests against a limited number of genes, the choice of which was often decided more by convenience than by deliberate planning. The probability of locating the new mutant is then, of course, rather small, and—which is more important if the new mutant is suitable as a marker—independence of all the known linkage groups cannot be asserted. It has been recognized in the past that this difficulty would be greatly lessened by the use of special testing stocks, and such stocks have been constructed (e.g. by Snell, quoted by Cooper, 1939); but the list of marker genes in the mouse and our knowledge of its genetical map have increased so much in recent years that a new set of stocks is now needed.

The main problem in the design of a set of mouse linkage-testing stocks lies in the choice to be made where two or more linked markers are available. This particular aspect of planning linkage tests appears not to have been given theoretical consideration anywhere in the literature. The purpose of this paper, then, is first to restate some of the general principles which underlie the planning of linkage tests, and to develop out of these a new

theoretical concept, the 'swept radius', by means of which the problem of linked markers can be handled; and then to describe how this concept has been applied to the design of some special linkage-testing mouse stocks. The section on theoretical principles rests very largely on the applications of Fisher's (1925) maximum likelihood method of estimation developed by himself, Mather (1935, 1936), Finney (1943, 1949) and others.

The application of these principles has led to the conclusions that, if suitably arranged, all the known genes that are desirable as markers for testing linkage could be contained in five stocks; and that, if only 100 offspring were raised in tests with each of these five stocks, the fraction of the total genetical map that would be covered would be of the order of 0·5–0·6. In other words, there would be at least an even chance of success in assigning the new mutant under test to a linkage group. This gives so great an improvement over what is at present possible with a comparable expenditure of labour that we think these

Table 1. *Values appropriate to the* 12 *common types of linkage-testing mating, of the swept radius* (m_r) *and the corresponding recombination fraction* (p_r); *and of the length swept* (m_s) *around a marker gene, when corrected for the end effect. Values are given for progenies of* 50 ($N=50$) *and of* 100 ($N=100$). *A significance level of* $2\frac{1}{2}\%$, *corresponding to* $k=2$, *is assumed; the average genetic map length is assumed to be* 130 *cm.*

	Mates			$N=50$			$N=100$		
Mating type	Coupling	Repulsion	i	p_r	m_r	m_s	p_r	m_r	m_s
Backcross	**AZ/az** × **az/az**	**Az/aZ** × **Az/aZ**	4·0	0·359	0·451	0·745	0·400	0·549	0·867
Intercross, semi-dominance at both loci	**AZ/az** × **AZ/az**	**Az/aZ** × **Az/aZ**	4·0	0·359	0·451	0·745	0·400	0·549	0·867
Intercross, semi-dominance at one locus	**AZ/az** × **AZ/az**	**Az/aZ** × **Az/aZ**	2·6	0·327	0·391	0·664	0·378	0·487	0·792
Mixed cross, semi-dominance at one locus	**AZ/az** × **aZ/az**	**Az/aZ** × **aZ/az**	2·0	0·300	0·347	0·601	0·359	0·451	0·745
Intercross, dominance at both loci	**AZ/az** × **AZ/az**	**AZ/aZ** × **Az/aZ**	1·7	0·288	0·328	0·573	0·350	0·424	0·723
Mixed cross, dominance at both loci	**AZ/az** × **aZ/az**	**Az/aZ** × **aZ/az**	1·3	0·255	0·281	0·502	0·327	0·391	0·664

stocks, when built up, will be worthy of general adoption as standard linkage-testing stocks. Moreover, the addition of a few new markers, requiring perhaps two or three more stocks, could cover nearly the whole of the mouse's genetical map, and the identification of the linkage group to which new mutants belong could become almost a certainty.

It should be made clear at the outset that these new stocks are designed for the special purpose of assigning a new or hitherto untested mutant to its linkage group. The accurate localization of the mutant with reference to the other genes of the group is a different problem for which the stocks are not specially suited.

2. Theoretical background

Types of mating used for detecting linkage

Twelve types of mating are commonly used to test a new mutant (**Z, z**) for linkage with a marker gene (**A**, **a**). They are derived from the three basic types, backcross, intercross and mixed cross, by various combinations of phase and dominance relationships and are listed in Table 1. All show the essential features of a linkage-testing mating, namely, that the frequencies with which the progeny are expected to fall into two or more phenotypic

classes can be expressed in terms of the proportion of recombinant gametes produced by the segregating heterozygote.

Statistical procedures for analysing linkage data

Statistical procedures for analysing the results of linkage-testing matings have been described by many authors. All the techniques operate on the same raw data, namely, observations of the numbers of the progeny of a linkage-testing mating which fall into each phenotypic class; and all have the same objects, namely, the estimation of the fraction, p, of recombinant gametes produced by a heterozygote **Aa Zz** and the estimation of the reliability of this estimate. It is not within the scope of this paper to compare the various techniques, and in what follows it will be assumed that the data will be treated by a maximum likelihood method, e.g. the procedure described by Finney (1949).

A feature of this method which is of importance in connexion with planning linkage tests is the concept of statistical information, I. This is a quantity which gives a measure of the accuracy of the conclusion drawn from an experiment, since it is related to the standard error, s, of the estimate of p, by the equation

$$s = I^{-\frac{1}{2}}. \tag{1}$$

For any given mating type I is proportional to the number of progeny raised; the amount of information per individual progeny, i, is therefore a relative measure of the efficiencies of different types of mating. Moreover, the amounts of statistical information obtained in tests with different types of mating can be added, thereby leading to an estimate of the accuracy of the conclusion drawn from several bodies of data jointly.

In its simple form the method is suitable only for application to data which show good, Mendelian, single-factor segregations; it is therefore necessary, in general, that the marker genes and new mutants under test shall not show greatly disturbed single-factor segregations, and in particular that the penetrance shall be complete or nearly so. In what follows it will be assumed that the single-factor segregations have been tested and found to be Mendelian.

The concept of swept radius

Suppose two marker genes, known to be located in different, independent linkage groups, were tested one against the other for linkage. The value found for the recombination fraction, p, would probably not be exactly equal to the free-segregation value, $\frac{1}{2}$; but the values of p found in a series of similar experiments would be distributed round the mean value $\frac{1}{2}$. The spread of the distribution would be determined by the amount of statistical information, I, contained in each experiment and therefore by the type of mating used and the number of progeny raised.

Now suppose that one of the marker genes were replaced by a new mutant at an undetermined locus. Since the test would be aimed at detecting departures from free segregation, it would be assumed *ex hypothese* that free segregation existed; the value found for p would therefore be expected to be distributed round the value $\frac{1}{2}$, again with a standard error determined by the type of mating and number of progeny raised. If the recombination fraction found in the experiment deviated from the value $\frac{1}{2}$ by less than ks (where k is a constant which sets the significance level), it would be concluded that the experimental results did not contain any significant evidence of linkage; or, in other words

that the new mutant did not lie within a 'swept radius' m_r of the marker gene, where m_r is defined as the genetical map-length corresponding with the recombination fraction p_r such that

$$p_r = \tfrac{1}{2} \pm ks, \tag{2}$$

or, making use of equation (1), and ignoring the possibility of linkage significantly exceeding 50%,

$$p_r = \tfrac{1}{2} - kI^{-\frac{1}{4}}. \tag{3}$$

In order to obtain a measure of the swept radius, m_r, it is therefore necessary to know the relationship between genetical map-length, m, and recombination fraction, p. It is known from experiment that when the recombination fraction between two loci is very small, the linear relationship proposed by Morgan holds and m may be equated to p; it is also experimentally found that p tends to the value $\frac{1}{2}$ for large values of m; but there is no general agreement about the exact form of the intervening curve (which, in any case, is not necessarily the same for all linkage groups and all species). Various forms for this relationship, some purely empirical and some based on theoretical considerations, have been suggested by Haldane (1919), de Winton & Haldane (1935), Kosambi (1944), Fisher, Lyon & Owen (1947), Owen (1949) and Srinath (1949). We have found that our own mouse-linkage data give a good fit, where large recombination fractions are concerned, with yet another relationship, namely,

$$m = \tfrac{1}{4}\,(\tanh^{-1} 2p + \tan^{-1} 2p). \tag{4}$$

This is one of a family of curves to which the Morgan, Haldane and Kosambi relationships also belong, since they are all integrals of the differential equation

$$\frac{dp}{dm} = 1 - (2p)^n. \tag{5}$$

The differences between them lie in the value chosen for n; the Haldane relationship is obtained for $n=1$, the Kosambi relationship for $n=2$ and the Morgan relationship for $n=\infty$; the new relationship is based on $n=4$.

These relationships differ widely where large recombination fractions are under consideration, but they are closely similar when p is small; when values of p exceeding, say, 40% are excluded, the curves differ only slightly and the exact form of the expression chosen to relate recombination fraction to map-length is of relatively little importance. Therefore practical convenience may be considered, and for the lower p values we have adopted Kosambi's formula, because it is easier than the new fourth power relationship to handle mathematically. Kosambi's equation is

$$\begin{aligned} m &= \tfrac{1}{2} \tanh^{-1} 2p \\ &= \tfrac{1}{4}\,[\log_e (1+2p) - \log_e (1-2p)]. \end{aligned} \tag{6}$$

When combined with equation (3) it gives the expression for the swept radius, namely,

$$m_r = \tfrac{1}{4} \log_e \left[\frac{I^{\frac{1}{2}}}{k} - 1 \right]. \tag{7}$$

Planning linkage tests

A number of problems which arise when planning tests for linkage, especially those involving linked markers, can be handled by means of the concept of swept radius.

(i) *The length swept in a backcross test.* Provided that the locus of the marker gene does

not lie within a distance m_r of the end of the genetic map of the linkage group in which it lies, the distance swept in the course of a linkage test will be $2m_r$. For a backcross linkage-testing mating $i=4$; the 'swept length' for such a mating is therefore

$$m_s = 2 \cdot \tfrac{1}{4} \log_e \left[\frac{(4N)^{\frac{1}{2}}}{k} - 1 \right].$$

In a practical test with mice 100 progeny might be raised and deviations exceeding twice the standard error might be considered significant. (This corresponds with a significance level of about $2\frac{1}{2}\%$, since only one tail of the distribution of p is included because linkages giving values of p far in excess of $\frac{1}{2}$ are not expected.) Putting $N=100$ and $k=2$ gives the swept length

$$m_s = 109{\cdot}8 \text{ cM}.$$

(ii) *The length swept in tests of other types.* The method of calculating swept length for other types of linkage-testing mating is similar to that used for backcrosses; the only difference lies in the value of i, which must be the value appropriate to the mating type (see Appendix I, Table 2).

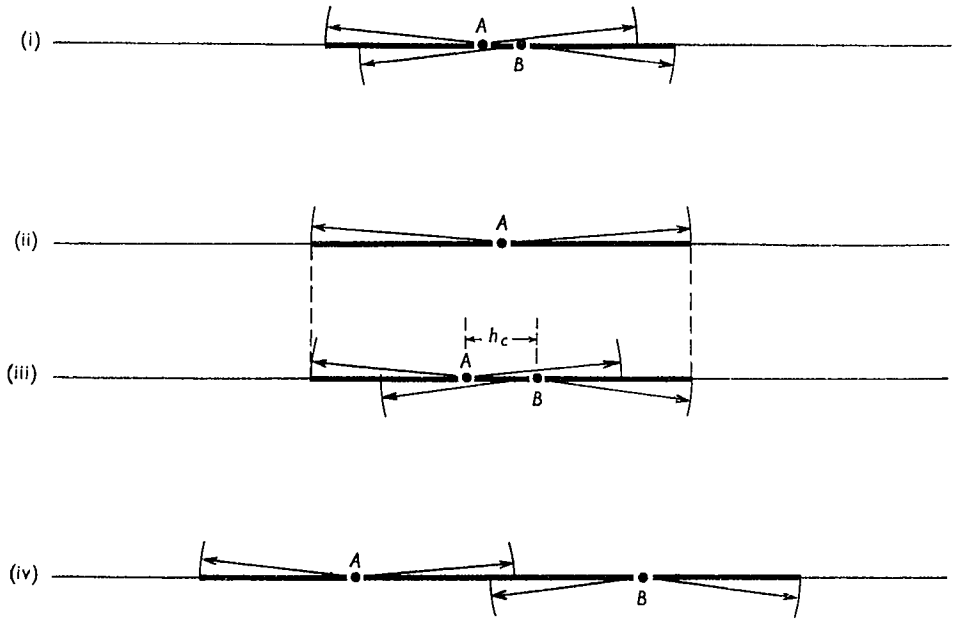

Fig. 1. For explanation see text.

(iii) *Alternative linked markers, both recessive or both dominant.* Sometimes a doubly marked chromosome, otherwise desirable, may not be available; or there may be strong interactions between two linked markers (e.g. **p** and **c**) which make their use in the same test undesirable. It may then be necessary to decide whether it is better to expend all the available effort on tests with one marker, or to divide the effort equally between separate tests with the two markers. It is clear that the former course would be better when the markers are closely linked, as **se** and **d** (Fig. 1(i), (ii)); and that the latter course would be better when they are loosely linked, as **s** and **lx** (Fig. 1(ii), (iv)). There must, therefore, be some critical separation of the two linked markers for which the two courses are equally good, and this critical separation, h_c, must be known before the correct decision can be

made. It can be found by equating the swept lengths attained by the two procedures (Fig. 1(ii),(iii)); the first will yield $2I$ units of information in tests with one locus; the second will yield I units of information in tests with each of two loci; therefore the critical separation is given by

$$2 \cdot \tfrac{1}{4} \log_e \left[\frac{(2I)^{\frac{1}{2}}}{k} - 1 \right] = h_c + 2 \cdot \tfrac{1}{4} \log_e \left[\frac{I^{\frac{1}{2}}}{k} - 1 \right],$$

whence

$$h_c = \tfrac{1}{4} \log_e 2 + \tfrac{1}{2} \log_e \left[\frac{I^{\frac{1}{2}} - k\sqrt{\frac{1}{2}}}{I^{\frac{1}{2}} - k} \right].$$

In all practical tests $I^{\frac{1}{2}}$ will be much bigger than k, so the last term will tend to vanish, since $\log_e 1 = 0$. Hence

$$h_c \simeq \tfrac{1}{4} \log_e 2$$
$$= 0{\cdot}173 \text{ or } 17{\cdot}3 \text{ cM}.$$

As I and k do not appear in this expression, this result holds good for all practical significance levels and however large the scale of the experiment. The available experimental effort should therefore be concentrated on tests with only one of the linked markers unless the marker separation exceeds 17 cM.

(iv) *Linked recessive and dominant markers in separate stocks.* Some linkage groups have a dominant and a recessive marker situated fairly close to each other, e.g. **Ca** and **bt** in linkage group VI or **Wh** and **wa-1** in linkage group XI. When a new mutant to be

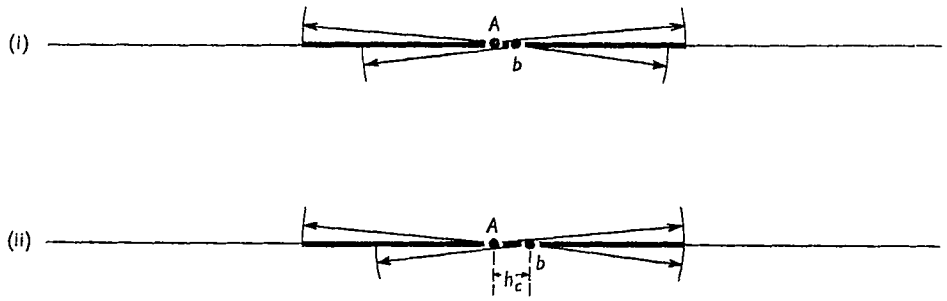

Fig. 2. For explanation see text.

tested is recessive, a backcross would normally be used for the test with the dominant marker, whereas an intercross must be used for the test with the recessive marker. The two markers would therefore require to be tested separately and would normally be kept in different stocks. If equal numbers of progeny were raised in each test, the radius swept round the dominant marker would be greater than the radius swept round the recessive marker, by virtue of the greater efficiency of backcrosses than intercrosses; and if the length of the segment between the markers were sufficiently small, it might happen that the part of the map swept in the test with the recessive marker lay entirely within the part already swept in the test with the dominant marker (Fig. 2(i)). The test with the recessive marker would then be superfluous. The critical marker separation, above which the test with the recessive marker would first begin to sweep a new part of the genetic map, is that separation for which the radius swept round the dominant marker is equal to the sum of the marker separation and the radius swept round the recessive marker

(Fig. 2(ii)). For a backcross $i=4$ and for an intercross with complete dominance $i=\frac{16}{9}$; the critical separation is therefore given by

$$\tfrac{1}{4}\log_e\left[\frac{(4N)^{\frac{1}{2}}}{k}-1\right]=h_c+\tfrac{1}{4}\log_e\left[\frac{(16N/9)^{\frac{1}{2}}}{k}-1\right],$$

whence

$$h_c=\tfrac{1}{4}\log_e\tfrac{3}{2}+\tfrac{1}{4}\log_e\left[\frac{4N^{\frac{1}{2}}-2k}{4N^{\frac{1}{2}}-3k}\right].$$

Here also the last term tends to vanish for all practical values of k and N, leaving

$$h_c\simeq\tfrac{1}{4}\log_e\tfrac{3}{2}$$
$$=0{\cdot}101 \text{ or } 10{\cdot}1 \text{ cM.}$$

Hence a test with the recessive marker would be superfluous unless its locus were more than 10 cm. from that of the dominant marker.

(v) *Linked recessive and dominant or semi-dominant markers in the same stock.* A problem related to the last arises when a single stock is available carrying both a recessive and

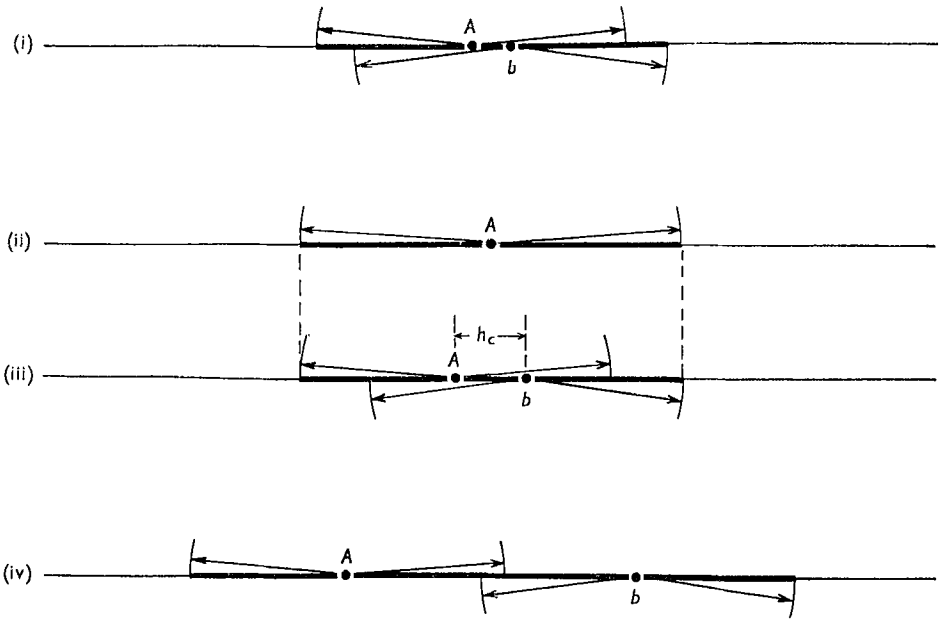

Fig. 3 For explanation see text.

a dominant or semi-dominant marker linked to it, e.g. **pa** and $\mathbf{A^w}$ or **hr** and $\mathbf{W^v}$. When such a stock is used for tests with a recessive new mutant, two courses are open, namely, (*a*) the triply-heterozygous F_1 could be mated to an animal from the new mutant stock, thereby constituting a backcross test involving only the new mutant and the dominant or semi-dominant marker; (*b*) triply-heterozygous F_1 could be mated *inter se*, thereby constituting a triple intercross testing the new mutant against both markers. The former procedure is clearly desirable when the marker separation is very small, since a backcross is more efficient than an intercross (Fig. 3(i),(ii)); the latter procedure is better when the separation is large (Fig. 3(ii),(iv)). The critical marker separation will be that for which the same map length is swept by either method (Fig. 3(ii),(iii)).

When one marker is fully dominant, so that $i=\frac{16}{9}$ for the intercross, the critical separation will be given by

$$2.\tfrac{1}{4}\log_e\left[\frac{(4N)^{\frac{1}{2}}}{k}-1\right]=h_c+2.\tfrac{1}{4}\log_e\left[\frac{(16N/9)^{\frac{1}{2}}}{k}-1\right],$$

whence

$$h_c\simeq\tfrac{1}{2}\log_e\tfrac{3}{2}$$
$$=0{\cdot}203 \text{ or } 20{\cdot}3 \text{ cM.}$$

Therefore it is better to backcross the triply heterozygous F_1 to the new mutant stock, thereby neglecting the recessive marker, unless the marker separation exceeds 20 cm.

When the two markers are a recessive and a semi-dominant, the analysis is slightly different; for an intercross with semi-dominance, i is $\frac{8}{3}$ instead of $\frac{16}{9}$, and the critical separation is found to be

$$h_c\simeq\tfrac{1}{8}\log_e\tfrac{27}{8}$$
$$=0{\cdot}152 \text{ or } 15{\cdot}2 \text{ cM.}$$

Therefore it is better to use the triple intercross when the marker separation exceeds 15 cM.

Corrections for finite genetic map lengths

In the previous section it has been assumed that the locus of the marker gene is well away from the end of the genetic map of the linkage group in which it lies. When the marker gene locus is near the end of the genetic map the swept radius may sometimes extend beyond the end of the map and part of it will then contribute nothing to the probability of success in detecting a linkage. Allowance for this 'end-effect' may be made by supposing the marker gene to occupy all positions in the map with equal probability, and then calculating the length of map that would, on the average, be swept by tests with such a marker. Consider, for example, what would be the average swept length in a test involving a marker lying in a linkage group of map-length L. If the marker locus were at one end of the map, the actual length swept would be m_r; if it were at a distance m_r from the end of the map, the length swept would be $2m_r$; and so over the intervening range the average length swept would be $\frac{1}{2}(m_r+2m_r)$. Similarly for the other end of the map. For the middle portion of the map, of length $(L-2m_r)$, the length swept would be $2m_r$. Hence the overall average is given by

$$m_s=\frac{\frac{1}{2}m_r(m_r+2m_r)+2m_r(L-2m_r)+\frac{1}{2}m_r(m+2m_r)}{L},$$

i.e.

$$m_s=2m_r-\frac{m_r^2}{L}. \tag{8}$$

A similar argument can be applied to the more general case when there are two linked markers, separated by a map-distance h, and the swept radii are f round one marker and g round the other. Provided that the separation is not so great that the swept radii fail to overlap, this leads to

$$m_s=(f+g+h)-\frac{(f^2+g^2)}{2(L-h)}. \tag{9}$$

When f and g are both equal to m_r, this reduces to

$$m_s=(2m_r+h)-\frac{m_r^2}{L-h}. \tag{10}$$

In each case the expression giving the swept length consists of two parts, namely, (*a*) the sum of the swept radii and the marker separation; (*b*) an end-correction involving L, which expresses the loss of swept length that occurs when the marker loci are near the end of the genetic map. L occurs only in the denominator of this end-correction and it therefore tends to vanish when L is large compared with m_r^2.

Before a corrected swept length can be calculated, using equations (8)–(10), it is necessary to know the value of L, the map-length of the linkage group in question. The average value of L can be obtained from cytological evidence, since it is equal to half the average number of chiasmata per bivalent. The latter has been found to be approximately 2·6 at diplotene in *Mus musculus* (Crew & Koller, 1932); hence the average value of L is 1·3, or 130 cM.

Sex is at present the only marker whose position relative to the end of the genetic map is known, its effective locus being presumed to lie at one end of the map (Fisher, Lyon & Owen, 1947). Thus the part of the map swept when sex is used as a marker lies wholly to one side of its locus, and the swept length is consequently equal to the swept radius.

3. THE NEW LINKAGE-TESTING STOCKS

The first step in planning the new linkage testing stocks was to decide which of the many known genes would be suitable for use as markers. The list of suitable markers was then reduced as much as possible by the elimination of redundant genes according to the principles outlined in the previous section. Finally, the genes chosen for use as marker were grouped together for the construction of the stocks in such a manner that all would be contained in the smallest possible number of stocks.

Genes suitable for use as markers

Three properties were considered to be essential for a gene to qualify for use as a marker in the stocks. The most important, without which the statistical calculations would become unduly laborious, was that the gene should give good Mendelian single-factor segregation. Genes with incomplete or irregular penetrance such as **Fu**, **py** and **si**, were therefore excluded. The second essential property was that the gene should be easy to classify at or before the age of weaning and without the need for special methods. On this ground genes such as rodless-retina, **r**, and the histocompatibility genes were excluded. Siderocyte anaemia (**f**), on the other hand, was considered to be suitable, in spite of the slight extra trouble caused by the necessity for classifying within forty-eight hours of birth. Finally it was considered essential that the gene should not seriously impair viability or fertility. Thus lethals such as grey lethal (**gl**) and steriles such as pituitary dwarf (**dw**) were clearly unsuitable. The inclusion of genes of the waltzer-shaker group seemed to be unavoidable, even though some of them reduce fertility quite considerably, because there are so many of them and some occupy important positions in known linkage groups. Every opportunity was, however, taken in later stages of the planning to eliminate as many as possible of these undesirable genes.

The genes that were considered by the above three criteria to be suitable for use as markers are listed in Table 2. They were selected from the list of genes of the mouse published by Law (1948), and they are arranged here by their linkage group. The genes shown in heavy type are those that were finally chosen, for the reasons described below, to be included in the stocks.

Choice of markers for stocks

Though all the genes* listed in Table 2 would be suitable for use as marker genes, not all merited a place in the linkage testing stocks. It was, of course, undesirable to have any gene in the stocks that would not contribute materially to the total swept length attainable with the stocks. On the grounds of redundancy, therefore, the list of markers was substantially reduced. The following genes were eliminated because they lie between two other markers whose swept radii together would cover the whole intervening segment: **W**; **pa**†; **pi**; **un**; **we**; **sh-2**. The following recessives were excluded because their swept lengths would be overlapped by those of adjacent dominant markers in the stocks: **bt**; **wa-1**. The following lie so close to another marker of the same type that the gain in swept length did not seem to be worth the cost of another gene in the stocks, and they were therefore excluded: **d**; **Ca**; **m**; **Ki**. The reasons for choosing one rather than the other of the adjacent pairs were the following: **se** rather than **d** because it does not mimic or interact with any other known gene; **N** rather than **Ca** because no mimics are known; **b** rather than **m** because classifica-

Table 2. *Genes of the mouse that would be suitable for use as markers, arranged according to their linkage groups. Those chosen for the stocks are shown in heavy type*

Linkage group	Gene symbol
I	c^{ch}, $\mathbf{c^e}$, *c*; **p**
II	**se**; *d*
III	**s**; *W*, W^v; *pi*; **lx**
IV	(None)
V	*kr*; A^y, A^w, **A**, $\mathbf{a^t}$, *a*; *un*; *we*; *pa*; **fi**
VI	*bt*; *Ca*; **N**
VII	**wa-2**; *sh*-2; **Re**
VIII	**b**; *m*
IX	**T**; *Ki*
X	**v**
XI	*wa*-1; **Wh**
XII	**ru**; **je**
'Independent'	**f**; **fz**; **ln**; **Sd**; **Va**

tion is easier; **T** rather than **Ki** because it is better known genetically and was available in our laboratory.

The existence of multiple alleles provided a choice of genes at two loci. At the albino locus $\mathbf{c^e}$ was chosen because **c** masks all other colour genes and $\mathbf{c^{ch}}$ was found to be difficult to classify in presence of **ru** which had to be allocated to the same stock. At the agouti locus two alleles, **A** and $\mathbf{a^t}$, were chosen for inclusion in the stocks. The presence of two alleles was necessary in order to ensure that the linkage testing stocks would always contain an allele different from that associated with the new mutant to be tested. **A** and $\mathbf{a^t}$ were chosen because each exhibits partial dominance over the other and both are dominant to **a**. $\mathbf{A^w}$ was rejected because it has not this advantage of incomplete dominance, and $\mathbf{A^y}$ was rejected because it impairs fertility and is lethal when homozygous.

The basic principle underlying the elimination of markers described above was the existence of a more suitable marker more or less closely linked to the marker eliminated. Several of the markers finally chosen, however, have been very inadequately tested for linkage with the others, and it therefore became imperative to survey the published data

* A glossary of the gene symbols used in this paper will be found in Appendix II.

† **pa** has recently been shown to lie between **a** and **fi** (Carter, T. C., *J. Genet.*, in the press).

upon which rests the assumption that the known linkage groups and the supposedly 'independent' genes are really independent of each other. The results of this survey are contained in another paper. They show that the early known genes have been extensively tested, though there are several pairs among them that have not been tested against each other or that have been tested by only small numbers of progeny. Of the more recently discovered genes, **fi** and **wa-2** have been tested against only a few other genes, while **ru**, **fz**, **Sd** and **Va** are almost completely untested. The retention of these genes in the linkage-testing stocks will therefore be conditional on their remaining independent of the others after they have been adequately tested.

Arrangement of the marker genes in the stocks

After the number of possible markers had been reduced to the minimum in the manner described above, twenty-one were left and the next task was to arrange these in groups suitable for maintenance in the smallest possible number of stocks. Two important principles governed the manner of grouping the genes. The first was that dominant and recessive markers should be contained in different stocks, because these call for different types of mating when tests for linkage with a new recessive mutant are to be made. The second was that all the genes in a stock should be easily classifiable in all combinations with each other.

The observance of these two principles left some latitude of choice in the allocation of the genes. This was utilized first by taking advantage of any combinations of genes that happened to be already available in the laboratory, and second by allocating to different stocks genes with symbols having the same initial letter. This latter, somewhat trivial, consideration was intended to reduce the probability of clerical errors in recording.

By the application of the above principles the genes chosen as markers were allocated to five stocks in the manner shown in Table 3. The construction of these stocks was immediately started, and the markers that have been combined in each stock at the time of going to press are shown in the right-hand column of Table 3. No serious trouble has arisen from interactions or impaired viability, and no reason has yet arisen for doubting that the stocks will be capable of construction in the form planned.

Assessment of total swept length

The potential value of these stocks for the detection of linkage may be seen from an examination of the total swept length attainable by their use. The separate lengths swept by each marker or pair of linked markers can be calculated without difficulty in the manner described in § 2. These calculations have been made with the assumption that the stocks will be used in the way outlined in Appendix I; that is to say, when the new mutant under test is dominant all the data will come from backcross matings, and when the new mutant is recessive, the tests against dominant markers will be backcrosses, those against recessive markers will be intercrosses and those against sex will be mixed crosses. The swept lengths calculated for each marker or pair of linked markers are shown in Table 4. Values appropriate to dominant or recessive new mutants are given for tests based on 50 and on 100 classified progeny.

The map distances between the markers on doubly marked chromosomes have been calculated from the recombination fractions noted at the foot of the Table. Kosambi's formula was used for the short linkage groups and the new fourth power formula given on p. 310 where linkages exceed 40%. For linkage group VII, however, the map proposed

by Fisher, Lyon & Owen (1947) has been adopted; the effective locus of sex has accordingly been supposed to lie at a distance of 65 cM. from **Re**. On this supposition the length swept by the three loci will cover the whole of a map 130 cM. long. The greater part of the swept length from sex will, however, be overlapped by that from **Re**, and sex will therefore contribute little to the total swept length.

Table 3. *Constitution of the linkage-testing stocks*

Stock no.	Markers to be present	Combinations available at time of going to press
I	**Sd Va Re A**	All
II	**T Wh N** a^t	All
III	**p lx v se fz** a^t	(**p se fz**), (**v lx** a^t)
IV	**b ln fi s wa-2** a^t	(**b ln s wa-2** a^t), (**fi** a^t)
V	**ru** c^e **je f** a^t	All

Table 4. *Map lengths in cM. swept by the separate markers or pairs of linked markers in the stocks. N is the number of progeny classified*

			Swept lengths in cM.			
			New mutant dominant		New mutant recessive	
Linkage group	Markers	References	$N=50$	$N=100$	$N=50$	$N=100$
I	c^e **p**	(1)	87·1	98·2	70·8	83·7
II	**se**		74·5	86·7	57·3	72·3
III	**s lx**	(2)	127·2	129·9	100·9	125·9
V	**A fi**	(3)	104·8	112·1	98·5	107·2
VI	**N**		74·5	86·7	74·5	86·7
VII	**wa-2 Re** sex	(4)	130·0	130·0	130·0	130·0
VIII	**b**		74·5	86·7	57·3	72·3
IX	**T**		74·5	86·7	74·5	86·7
X	**v**		74·5	86·7	57·3	72·3
XI	**Wh**		74·5	86·7	74·5	86·7
XII	**ru je**	(5)	118·1	124·7	106·3	115·9
—	**f**		74·5	86·7	57·3	72·3
—	**fz**		74·5	86·7	57·3	72·3
—	**ln**		74·5	86·7	57·3	72·3
—	**Sd**		74·5	86·7	74·5	86·7
—	**Va**		74·5	86·7	74·5	86·7
Total swept length			1386·7	1548·6	1222·8	1430·0
Fraction of 2600·0			0·53	0·60	0·47	0·55

References: (1) 14% recombination (average of sexes) (Grüneberg, 1936).
(2) 8% recombination **s** to **hr** (Snell, 1931).
42% recombination **hr** to **W** (Gates & Pullig, 1945).
16% recombination $\mathbf{W}^{\mathbf{v}}$ to **lx** (Carter, 1949).
(3) 33% recombination (Carter & Grüneberg, 1950).
(4) Map distances: sex to **Re** = 65 cM., **Re** to **wa-2** = 55 cM. (Fisher, Lyon & Owen, 1947).
(5) 45% recombination (Fisher & Snell, 1948).

The summation of the separate swept lengths to give an estimate of the total for the five stocks is shown at the foot of Table 4. If 100 progeny are classified in tests with each of the five stocks the total swept length will be 1549 cM. when the new mutant under test is dominant, and 1430 cM. when the new mutant is recessive. Since the mouse has 20 chromosomes with average map lengths of 130 cM., its total map length is 2600 cM. Therefore by raising 500 test progeny a dominant new mutant will be tested against 0·60 of the total genetic map, and a recessive new mutant against 0·55. These fractions therefore express the probability of success in assigning a new mutant to its linkage group by the use of the new stocks.

It must, however, be emphasized that the total swept length attained by the stocks cannot yet be estimated with precision, and the figures given above for the probability of success in detecting a linkage should be regarded as rough estimates. The most important source of uncertainty in the estimate is the possible existence of linkage between some of the markers now assumed to be independent. At least four of the markers in the stocks have been very inadequately tested for linkage, and it is possible that one or more of these will eventually be found to be linked, perhaps closely, with another marker. The total swept length would then be somewhat reduced. Other sources of uncertainty in the estimate of the total swept length and of the probability of detecting a linkage, which do not seem to be important enough to merit a detailed discussion, are: (i) the assumption that all the linkage-group maps are of equal length; (ii) the use of a possibly inaccurate relationship between recombination value and map-length; (iii) the adoption of a $2\frac{1}{2}\%$ significance level for the calculation of swept radii, and the assumption that a sub-significant indication of linkage, if found, would not be followed up by additional breeding tests.

Despite the uncertainty attendant on the attempt to estimate the total swept length, it is claimed that these stocks, when completed, will offer the best material for routine linkage tests that can at present be constructed. The discovery of new linkages and of new marker genes will, however, necessitate changes in the stocks, perhaps even before they are in fact fully constructed. But the necessary changes are unlikely to arise so fast that the stocks cannot form a useful basis for the construction and maintenance of a set of standard linkage-testing mouse stocks.

SUMMARY

1. The importance of mapping the chromosomes of the mouse is stressed, and the need for specially designed stocks for routine linkage testing is pointed out.

2. The theory of planning linkage tests is outlined, and the new concept of 'swept radius' is introduced. This makes it possible (*a*) to decide, on the basis of efficiency, between a number of alternative procedures that present themselves in the planning of linkage tests; and (*b*) to design special linkage testing stocks.

3. The composition of five specially designed linkage testing stocks is described. By the use of these stocks a gene can be tested against a little over half of the total genetic map, at the cost of raising about 500 test progeny, which is a considerable improvement over what is at present possible.

4. An outline is given in Appendix I of the procedure to be adopted for the use of the stocks.

REFERENCES

CARTER, T. C. (1949). *Nature, Lond.*, **164**, 1138
CARTER, T. C. & GRÜNEBERG, H. (1950). *Heredity* (in the Press).
COOPER, C. B. (1939). *J. Hered.* **30**, 212.
CREW, F. A. E. & KOLLER, P. C. (1932). *J. Genet.* **26**, 359–83.
FINNEY, D. (1943). *Ann. Eugen., Lond.*, **12**, 31–43.
FINNEY, D. (1949). *J. Genet.* **49**, 159–76.
FISHER, R. A. (1925). *Proc. Camb. Phil. Soc.* **22**, 700–25.
FISHER, R. A. (1946). *Amer. Nat.* **80**, 568–78.
FISHER, R. A., LYON, M. F. & OWEN, A. R. G. (1947). *Heredity*, **1**, 355–65.
FISHER, R. A. & SNELL, G. D. (1948). *Heredity*, **2**, 271–3.

GATES, W. H. & PULLIG, T. (1945). *Proc. La. Acad. Sci.* **9**, 57–60.
GRÜNEBERG, H. (1936). *J. Genet.* **33**, 255–65.
HALDANE, J. B. S. (1919). *J. Genet.* **8**, 299–309.
KOSAMBI, D. D. (1944). *Ann. Eugen., Lond.*, **12**, 172–5.
LAW, L. W. (1948). *J. Hered.* **39**, 300–8.
MATHER, K. (1935). *Ann. Eugen., Lond.*, **6**, 399–410.
MATHER, K. (1936). *Ann. Eugen., Lond.*, **7**, 251–64.
OWEN, A. R. G. (1949). *Proc. Roy. Soc.* B, **136**, 67–94.
SLIZYNSKI, B. M. (1949). *J. Genet.* **49**, 242–5.
SNELL, G. D. (1931). *Genetics*, **16**, 42–74.
SRINATH, K. V. (1949). *Proc. Roy. Soc.* B, **136**, 126–30.
DE WINTON, D. & HALDANE, J. B. S. (1935). *J. Genet.* **31**, 67–100.

APPENDIX I

Procedure for the use of the linkage-testing stocks

Since the value of the stocks as a means of detecting linkage depends on their being correctly used, and the method of use differs according to the nature of the new mutant under test, a brief description of the methods intended is given here. In every case the test requires two generations. The first is always an outcross of an animal carrying the new mutant under test to an animal from each of the stocks. The animals from the stocks should have the mutant phenotype of all the markers in the stocks from which they come. The multiple heterozygotes, whose gamete formation is to be examined for evidence of linkage, occur in this F_1 generation. The second generation is obtained from the mating of a multiple heterozygote selected from the F_1 generation: but the type of animal to which it is to be mated depends on the character of the new mutant under test and on the character of the markers against which it is being tested. The type of animal to be mated to the multiple heterozygote under the different conditions is shown in Table 1. Three categories of new mutant require different treatment, namely, (i) dominants, (ii) recessives and (iii) recessives with sterile homozygotes. (Lethal recessives with homozygotes which die before other characters can be classified are not specifically considered, though the stocks can be used to test them for linkage.) The treatment of the first two categories is fully enough described by the Table, but the treatment of new mutants in the third category needs some more explanation. In the first place, the animal used for the outcross will, of course, have to be a heterozygote, and therefore only half the F_1 will carry the new mutant.

When a new mutant in this category is being tested against a recessive-carrying stock two courses are open; F_1 animals may be tested for the new mutant by mating to known heterozygotes, and suitable F_1 animals then mated *inter se*; or F_1 animals may be mated *inter se* at random, in which case the new mutant will be expected to segregate from only one mating in four. When a new mutant in the third category is being tested against a dominant-carrying stock two courses are again open. Multiply-mutant F_1 animals may either be mated to known heterozygotes from the new mutant stock or they may be mated *inter se*. The former course will usually be better, because the new mutant will be expected to segregate from one mating in two, compared with one in four intercross matings; only when the marker gene is semi-dominant, and the amount of information available per individual from an intercross is consequently twice the amount available from a mixed cross, will the two courses be equally good.

All the stocks test for linkage with sex if a male F_1 animal is used. The type of mating testing sex will be the same as that testing the marker genes in the stock except where this is an intercross: sex will then be tested by a mixed cross. The phase will depend on the sexes of the animals used in the first or outcross generation. If the animal carrying the

Appendix Table 1. *Type of animal to be mated to the multiply heterozygous F_1; and type of the resulting linkage-testing mating*

	Stock against which new mutant is to be tested	
Category of new mutant under test	Dominant-carrying (I, II)	Recessive-carrying (III, IV, V)
Dominant	Multiple recessive: i.e. wild-type for all markers in the stock, except **A**, $\mathbf{a}^t$ *Repulsion backcross*	Multiple recessive from the stock. *Coupling backcross*
Recessive	Homozygote of new mutant *Coupling backcross*	Multiple heterozygote: i.e. F_1 mated *inter se* *Repulsion intercross*
Recessive, with sterile homozygote	Heterozygote of new mutant *Coupling mixed cross*	Multiple heterozygote: i.e. F_1 mated *inter se* *Repulsion intercross*

Appendix Table 2. *Scoring coefficients and amount of information per individual for different types of linkage-testing mating.* **A** *and* **B** *represent the dominant alleles of the two genes in the test,* **a** *and* **b** *their recessive alleles, irrespective of which are the mutant alleles. The genes are in coupling when* **A** *and* **B** *entered the test together from the same parent*

	Scoring coefficients: Phenotypic class						
	AA B	**Aa B**	**AA b**	**Aa b**	**a B**	**a b**	Information per individual
Sign for Coupling	−	−	+	+	+	−	
Sign for Repulsion	+	+	−	−	−	+	
A semi-dominant:							
Intercross	4/3	0	4	0	4/3	4	8/3
Mixed cross (**A** intercrossed)	2	0	2	0	2	2	2
AA masks **B,b** distinction:							
Intercross	0	0	0	0	4/3	4	4/3
Mixed cross (**A** intercrossed)	0	0	0	0	2	2	1
AA inviable:							
Intercross	0		0		4/3	4	16/9
Mixed cross (**A** intercrossed)	0		0		2	2	4/3
A and **B** fully dominant:							
Backcross	2		2		2	2	4
Intercross	4/9		4/3		4/3	4	16/9
Mixed cross (**A** intercrossed)	2/3		2/3		2	2	4/3
aa masks **B,b** distinction:							
Backcross	2		2		0		2
Intercross	4/9		4/3		0		4/9
Mixed cross (**A** intercrossed)	2/3		2/3		0		1/3
Mixed cross (**B** intercrossed)	2/3		2		0		2/3
A indistinguishable from **B**:							
Backcross			±2/3			∓2	4/3
aa indistinguishable from **bb**:							
Intercross	∓4/9				±4/7		16/63

new mutant was a male, the phase will be coupling when the new mutant is dominant and repulsion when it is recessive. If the animal carrying the new mutant was a female the phases will be reversed.

When the breeding data have been obtained, it remains to examine the segregation of the

new mutant with each marker in order to detect deviations from free segregation; an estimate of the apparent recombination fraction and its standard error will also usually be required. We have found that this analysis can most conveniently be done by the procedure described below; it adds nothing new to the procedures given by Mather (1935), Finney (1943) and Fisher (1946), but does not follow any one in all details.

(i) The observed number of progeny in each of the recognizable phenotypic classes is multiplied by the 'scoring coefficient' appropriate to the class and to the type of mating. This product is the contribution of the class to the score for that type of mating. Scoring coefficients are listed in Table 2 for all the types of mating likely to be encountered in the tests. The signs of the coefficients, which depend on the phase of the multiple heterozygote under test, are shown at the heads of the columns in the Table: they are applicable only on condition that when the dominant alleles of the two genes enter the heterozygote together from the same parent the phase is said to be coupling, and when they enter separately from different parents it is said to be repulsion, irrespective of whether the mutant allele is dominant or recessive. The classes representing grand-parental combinations thus have negative coefficients, and a negative score favours linkage.

(ii) The score, D, for each type of mating is obtained by summing the contributions from the classes, due attention being paid to sign. Scores are summed over the mating types, if more than one type has been used, to give a total score, ΣD.

(iii) The number of progeny from each type of mating is multiplied by the amount of information per individual, i, appropriate to the type of mating; this is also listed in Table 2. The product is the amount of information, I, for the type of mating. If more than one type has been used, the amounts of information are summed to give a total ΣI.

(iv) The total score is divided by the total information to obtain a correction, which when added to $\frac{1}{2}$ (with due attention to sign) gives a first estimate, p_1, of the recombination fraction. Thus

$$p_1 = \tfrac{1}{2} + \frac{\Sigma D}{\Sigma I}.$$

(v) The standard error of p_1 is obtained by taking the square root of the reciprocal of the total amount of information

$$s_{p_1} = 1/\sqrt{(\Sigma I)}.$$

(vi) The significance of the deviation from free segregation is tested by comparing the correction $\Sigma D/\Sigma I$ with its standard error, which is the same as s_{p_1}; it is therefore a c-test:

$$c = \frac{1}{s_{p_1}} \frac{\Sigma D}{\Sigma I} = \frac{\Sigma D}{\sqrt{(\Sigma I)}}.$$

Alternatively the squared form may be used:

$$\chi_1^2 = \frac{(\Sigma D)^2}{\Sigma I}.$$

If linkage is found, the first estimate of the recombination fraction, p_1, may not be sufficiently accurate. A second estimate, based on revised scoring coefficients, will then be necessary; this may be found by the method and tables given by Finney (1949).

APPENDIX II

Glossary of gene symbols used in this paper and a later one, with a brief indication of the characters affected. References to detailed descriptions may be found in Law, L.W., 1948, Mouse Genetics News, No. 2, *J. Hered.* **39**, 300–8.

A	Agouti	coat colour
a	non-agouti	coat colour
a^t	tan	coat colour
A^w	White-bellied agouti	coat colour
A^y	Yellow	coat colour
b	brown	coat colour
bt	belted	white spotting
c	albino	coat and eye colour
c^e	extreme dilution	coat colour
c^{ch}	chinchilla	coat colour
Ca	Caracul	waved coat
d	dilute	coat colour
dw	dwarf	pituitary defect
f	flexed tail (siderocyte anaemia)	tail deformity and anaemia
fi	fidget	behaviour
Fu	Fused	tail deformity
fz	fuzzy	coat texture
gl	grey-lethal	bone deformity and coat colour
hr	hairless	loss of hair
j	jittery	behaviour
je	jerker	behaviour
Ki	Kinky tail	tail deformity
kr	kreisler	behaviour
ln	leaden	coat colour
lx	luxate	absence of tibia
m	misty	coat colour
N	Naked	loss of hair
p	pink-eye	coat and eye colour
pa	pallid	coat and eye colour
pi	pirouette	behaviour
r	rodless	defect of retina
Re	Rex	waved coat
ru	ruby	coat and eye colour
s	piebald	white spotting
Sd	Danforth's short tail	tail deformity
se	short-ear	reduced pinna
sh-1	shaker–1	behaviour
sh-2	shaker–2	behaviour
si	silver	coat colour
T	Brachyury	tail deformity
un	undulated	tail deformity
v	waltzer	behaviour
Va	Variant waddler	coat colour, spotting, behaviour
W	Dominant spotting (macrocytic anaemia)	white spotting
W^v	Dominant spotting, viable allele	white spotting and coat colour
wa-1	waved–1	waved coat
wa-2	waved–2	waved coat
we	wellhaarig	waved coat
Wh	White	coat colour