Reproduced from the Philosophical Transactions of the Royal Society, B, 233: 55-87 (1947), with permission of the Royal Society

213

THE THEORY OF LINKAGE IN POLYSOMIC INHERITANCE

By R. A. FISHER, F.R.S.

(*Received* 15 *March* 1946)

CONTENTS

The paper is concerned with the combinatorial and statistical problems arising in the theoretical analysis of linkage in polysomic inheritance. The second section gives a method of enumerating the heterogenic genotypes capable of being composed with given numbers of genes. In the third section these are classified in isomorphic sets, which correspond with the partitions in two or more dimensions of the numbers 4 and 6. The members of an isomorphic set can be generated by systematic gene substitution from any member of the set, and in consequence yield populations of gametes with equivalent frequencies.

An aspect of the subject which has escaped attention is the multiplicity of the modes of gamete formation. Their enumeration is discussed in §4, and leads to the specification in §5 of the gametic matrices for the 16 sets or pairs of isomorphic sets arising in tetrasomics with two linked loci. In most cases the rank of the matrix is less than 11, so that the gametic series from the genotype concerned is deficient in respect of some of the information sought. Different sets of genotypes may be used to supplement one another's information. The final section discusses in outline the statistical problem of estimating the frequencies of modes of gamete formation from observed gametic series.

1. Parameters specifying the frequencies of modes of gametogenesis

The genetic constitution of an organism, derived from the fusion of two gametes, is specified by the genetic constitutions of the two constituent gametes. The laws of inheritance obtained by genetic studies are the rules whereby, given the constitution of an organism, the kinds of gametes it can produce, and their relative frequencies, can be predicted. As conceived by Mendel, no parameters entered into these rules; all possible combinations of elements drawn from the homologous pairs supplied by the parents were thought to appear in the gametes with equal frequency.

The discovery of the rules of gamete formation in cases of linkage in the early investigations with *Drosophila melanogaster* first altered this situation by showing that the gametic output from an organism doubly heterozygous for two linked factors could only be predicted in terms of the appropriate recombination fraction. Two modes of gametogenesis are conceived, one resulting in a gamete in which both of the two genes considered were derived from the same parent, whereas in the other they were derived from different parents. The latter is very generally the less frequent, and the proportion its frequency bears to the whole is known as the recombination fraction. It is often called the cross-over value

(or c.o.v.), a term which is inaccurate save for close linkage, since crossing-over, involving the interchange of germinal material, may occur in the intervening segment, without the recombination of the two genes specified.

With three linked genes, four modes of gamete formation are distinguishable, since recombination may have taken place in the first segment, in the second, in both, or in neither. There are thus three parameters to be determined empirically. With l linked genes, the number of distinguishable modes of gametogenesis becomes 2^{l-1}, and the number of parameters necessary to specify the gametic output is $2^{l-1}-1$. It is to be presumed that these may be expressible, at least approximately, in terms of the genetical lengths, or of the recombination fractions of the $l-1$ segments, but it should be emphasized that no satisfactory way of doing this has been put forward. Following early work by Haldane (1919), Kosambi (1944) has recently proposed that the relationship between the recombination fraction, y, and the map distance, x, may be approximately of the form

$$2y = \tanh(2x),$$

so that the recombination fraction corresponding with the sum of two segments, having recombination fractions y_1 and y_2, would be

$$y_{1+2} = \frac{y_1+y_2}{1+4y_1y_2}.$$

The frequencies of the four gametic types produced by a triple heterozygote could then be expressed in terms of y_1 and y_2 only:

		first segment		
		old combination	new combination	
second segment	old	$1-y_1-y_2+2y_1y_2(2-y_1-y_2)$	$y_1+2y_1y_2(y_1-y_2)$	$(1-y_2)(1+4y_1y_2)$
	new	$y_2-2y_1y_2(y_1-y_2)$	$2y_1y_2(y_1+y_2)$	$y_2(1+4y_1y_2)$
		$(1-y_1)(1+4y_1y_2)$	$y_1(1+4y_1y_2)$	$1+4y_1y_2$

Where tests are possible the formula is evidently sufficiently nearly correct to represent a step in the right direction. It does not, however, offer a complete theory of gamete formation for linked factors, since with four factors or more the frequencies of the different gametic types cannot by its means be specified in terms of single parameters for the different segments. With four loci, for example, it yields the recombination fractions between all six pairs, whereas seven relationships are needed among the eight frequencies of distinguishable modes of gametogenesis. A formally complete specification would express the series of frequencies

$$p_0, p_1, p_2, p_3, \ldots$$

of a segment in a gamete having experienced $0, 1, 2, \ldots$ breaks in the preceding meiosis, all in terms of a single parameter, in terms of which, therefore, the recombination fraction

$$y = p_1+p_3+p_5+\ldots,$$

and the map distance

$$x = p_1+2p_2+3p_3+\ldots,$$

could likewise be expressed.

A second extension of Mendel's ideas has come with the discovery of polysomic inheritance, in which organisms containing sets of four or six homologous loci are reproduced normally*

* I am not concerned in this paper with the abnormal gamete formation often observed in artificial polyploids.

by means of diploid or triploid gametes. This possibility greatly increases the number of genotypes which can be generated from one or a few gene substitutions. For, whereas with diploids one is concerned only with the partitions of 2,

(2) homozygous, (1^2) heterozygous,

with tetraploids, variation at a single locus will yield organisms of types corresponding with the partitions of 4, namely,

(4) homogenic,
(31) digenic, simplex or triplex,
(2^2) digenic, duplex,
(21^2) trigenic,
(1^4) tetragenic.

The terms simplex, duplex and triplex, indicating the number of dominant genes, are insufficient even with tetraploids to express the full classification by partitions, and this is still more so with the eleven partitions of 6 which represent the possible types of hexaploids. It is the distinction afforded by the partitions, which are themselves invariant for gene substitutions, which is needed for developing the laws of gametic output.

For a single locus, these laws, as has been shown by Fisher & Mather (1943), may be developed very simply from the consideration that in tetraploids and hexaploids only two modes of gamete formation are distinguishable. The two genes in the gamete may either be derived from different chromosomes of the zygote, or more rarely may be identical and originate from the same chromosome. The frequency of the latter event may be taken to be independent of the gene content of the locus in question, and has been designated by the letter α. Similarly, with hexaploids there might be representatives from three different chromosomes, or from two only, one having a double representation; the frequency of this latter event may be designated by β. It would, generally at least, be impossible for all three representatives to have been derived from a single chromosome, since in meiosis each gene is only duplicated. With hexaploids therefore only gametes of the modes of origin (1^3) and (21) are expected, while with octaploids the modes of origin (1^4), (21^2) and (2^2) are all possible.

Introducing only one parameter representing the frequency distribution of the two modes of gamete formation, the gametic output of the four heterogenic types of autotetraploids are as follows:

digenic types

partitional type of parent	representative genetic formula	gamete *aa*	gamete *Aa*	gamete *AA*	divisor
(31)	Aa_3	$2+\alpha$	$2-2\alpha$	α	$\div 4$
(2^2)	A_2a_2	$1+2\alpha$	$4-4\alpha$	$1+2\alpha$	$\div 6$

trigenic type

		gamete *aa*	gamete *Aa*	gamete *A'a*	gamete *AA*	gamete *AA'*	gamete *A'A'*	divisor
(21^2)	$AA'a_2$	$2+4\alpha$	$4-4\alpha$	$4-4\alpha$	3α	$2-2\alpha$	3α	$\div 12$

tetragenic type

		4 homogenic types of gametes	6 heterogenic types of gametes	divisor
(1^4)	$AA'aa'$	3α	$2-2\alpha$	$\div 12$

Similarly, the gametic output of the ten heterogenic types of hexasomic organisms can be expressed in terms of only a single parameter:

digenic types

parent		gamete a_3	a_2A	a_1A_2	A_3	divisor
(51)	Aa_5	$3+\beta$	$3-2\beta$	β	.	$\div 6$
(42)	A_2a_4	$3+3\beta$	$9-5\beta$	$3+\beta$	β	$\div 15$
(3^2)	A_3a_3	$1+3\beta$	$9-3\beta$	$9-3\beta$	$1+3\beta$	$\div 20$

trigenic types

parent		gametes a_3	Aa_2	$A'a_2$	A_2a	$AA'a$	A'_2a	A_3	A_2A'	AA'_2	A'_3	divisor
(41^2)	$AA'a_4$	$6+6\beta$	$9-5\beta$	$9-5\beta$	4β	$6-6\beta$	4β	.	β	β	.	$\div 30$
(321)	AA'_2a_3	$3+9\beta$	$9-3\beta$	$18-6\beta$	6β	$18-18\beta$	$9+3\beta$	.	4β	$3+\beta$	4β	$\div 60$
(2^3)	$A_2A'_2a_2$	2β	$3+\beta$	$3+\beta$	$3+\beta$	$12-12\beta$	$3+\beta$	2β	$3+\beta$	$3+\beta$	2β	$\div 30$

With four genes there will be twenty types of gamete possible, for five genes thirty-five and for six genes fifty-six. These are listed as follows:

tetragenic types

partition (31^3) formula $AA'A''a_3$			partition (2^21^2) formula $AA'a_2a'_2$		
type of gamete	number of kinds	frequency	type of gamete	number of kinds	frequency
a_3	1 kind	$3+9\beta$	a_3	2 kinds	4β
Aa_2	3 kinds	$9-3\beta$	a_2a'	2 kinds	$6+2\beta$
A_2a	3 kinds	6β	Aa_2	4 kinds	$3+\beta$
$AA'a$	3 kinds	$9-9\beta$	A_2a	4 kinds	4β
A_2A'	6 kinds	2β	Aaa'	2 kinds	$12-12\beta$
$AA'A''$	1 kind	$3-3\beta$	$AA'a$	2 kinds	$6-6\beta$
		60	A_2A'	2 kinds	2β
					60

pentagenic type

partition (21^4) formula $AA'A''A'''a_2$		
type of gamete	number of kinds	frequency
a_3	1 kind	4β
Aa_2	4 kinds	$3+\beta$
A_2a	4 kinds	4β
$AA'a$	6 kinds	$6-6\beta$
A_2A'	12 kinds	2β
$AA'A''$	4 kinds	$3-3\beta$
		60

hexagenic type

partition (1^6) formula $AA'A''aa'a''$		
type of gamete	number of kinds	frequency
a_2a'	30 kinds	2β
$aa'a''$	20 kinds	$3-3\beta$
		60

Although in the two instances in which Mendel's purely combinatorial scheme of inheritance has been modified by the introduction of parameters, which must be determined empirically, and which may be influenced by external circumstances such as temperature, and by genetic constitution, as by sex, yet there is no reason to suspect, save in quite extraordinary cases, that the frequencies of different modes of gametogenesis will be influenced by the gene content of the particular loci being studied. The combinatorial character of Mendel's law is maintained in the generalization that: *The gametic frequencies are invariant in respect of any gene substitution applied systematically to the genic content of an organism and of the gametes it produces.* With but few exceptions, moreover, though this is not universal, the genetic specification of the zygote takes no account of which half of its content was derived from either parent. On these two principles a rational approach can be made to the theoretical analysis of linkage in polysomic inheritance.

2. The enumeration of segregating genotypes

In disomic inheritance the number of different genotypes segregating for each of l linked factors, or heterogenic at each of l linked loci, is 2^{l-1}, and is equal to the number of distinguishable modes of gamete formation. In polysomic inheritance this equivalence no longer subsists. The number of genotypes in many cases is very large, and presents a

problem of enumeration. Moreover, whereas with disomic inheritance all the genotypes heterogenic at the same loci are isomorphic in the sense that they yield gametic series with the same frequencies, the series for any one genotype being derivable from that of any other by a mere permutation of the gene symbols in parent and gametes, the different genotypes of polysomic organisms fall into a number of classes, only members of the same class being isomorphic in their gametic output.

If there is a choice of ρ different genes at one locus and of σ at a second on the same chromosome, the chromosome can be made up in $\rho\sigma$ ways; if there is a choice of τ genes at a third locus, the number of combinations possible will be $\rho\sigma\tau$, and so on. If a chromosome can be made up in n ways, the number of selections of four such chromosomes possible will be

$$\frac{1}{4!}n(n+1)\,(n+2)\,(n+3);$$

the number of selections of six will likewise be

$$\frac{1}{6!}n(n+1)\,(n+2)\,(n+3)\,(n+4)\,(n+5).$$

Hence, giving ρ and σ the values of 1, 2, 3 and 4, the numbers of genotypes for tetrasomic organisms, classified in respect of two loci, will be the values in table 1, in which n is given the values of a multiplication table up to 4×4.

TABLE 1. $(\rho\sigma+3)!/(\rho\sigma-1)!\,4!$

σ \ ρ	1	2	3	4
1	1	5	15	35
2	5	35	126	330
3	15	126	495	1365
4	35	330	1365	3876

These are the total numbers of genotypes containing ρ or fewer different genes at one locus, and σ or fewer genes at the second. To find the numbers having exactly r and s genes respectively, one must take the rth advancing difference for columns, and the sth advancing difference for rows, from $\rho = 0$, $\sigma = 0$ respectively, i.e.

$$\Delta_\rho^r \Delta_\sigma^s \frac{1}{4!}\rho\sigma(\rho\sigma+1)\,(\rho\sigma+2)\,(\rho\sigma+3) \quad (\rho = 0,\ \sigma = 0),$$

giving the derived table 2. All differences beyond the fourth must vanish, since a tetrasomic organism cannot accommodate more than four allelomorphs of the same factor. Genotypes heterogenic at both loci are shown in the three last rows and columns. There are of these 266 genotypes in all. Before classifying these into sets of isomorphic genotypes, the results of applying the same method of enumeration to other cases may be given.

TABLE 2

s \ r	1	2	3	4
1	1	3	3	1
2	3	19	30	14
3	3	30	63	36
4	1	14	36	24

For two linked loci of a hexasomic organism the value of

$$\frac{(n+5)!}{6!(n-1)!}$$

is required, when $n = \rho\sigma$,

and ρ and σ take values from 1 to 6. These are given in table 3, upon which, operating with

$$\Delta^r_\rho \Delta^s_\sigma \quad (\rho = 0, \sigma = 0),$$

when r and s take values from 1 to 6, one finds the results given in table 4, and of these 37,214 are heterogenic at both loci.

TABLE 3. $(\rho\sigma+5)!/(\rho\sigma-1)!\ 6!$

1	7	28	84	210	462
7	84	462	1,716	5,005	12,376
28	462	3,003	12,376	38,760	100,947
84	1,716	12,376	54,264	177,100	475,020
210	5,005	38,760	177,100	593,775	1,623,160
462	12,376	100,947	475,020	1,623,160	4,496,388

TABLE 4

1	5	10	10	5	1
5	60	211	324	230	62
10	211	1038	2022	1725	540
10	324	2022	4648	4500	1560
5	230	1725	4500	4800	1800
1	62	540	1560	1800	720

With three linked loci in a tetrasomic organism, in addition to the values of table 2, which may be thought of as generated by the operator

$$\Delta^r_\rho \Delta^s_\sigma \Delta^t_\tau,$$

acting on

$$\frac{(\rho\sigma\tau+3)!}{4!(\rho\sigma\tau-1)!},$$

table 5 shows the result when $t = 2$, table 6 when $t = 3$ and table 7 when $t = 4$. These tables give the numbers of genotypes for all classes of tetrasomic organisms distinguished by three linked loci. Of these 17,320 are heterogenic at all loci.

TABLE 5

		r = 2	r = 3	r = 4
s	2	172	348	196
	3	348	810	504
	4	196	504	336

TABLE 6

		r = 2	r = 3	r = 4
s	2	348	810	504
	3	810	1998	1296
	4	504	1296	864

TABLE 7

		r = 2	r = 3	r = 4
s	2	196	504	336
	3	504	1296	864
	4	336	864	576

3. The classification of segregating genotypes

Any genotype having r different genes at one locus, s at a second, t at a third, and so on, may be said to be isomorphic with any other specified by one of the $r!$ permutations of genes at the first locus, the $s!$ at the second, and so on. The maximum number of different genotypes in a set of isomorphic genotypes is thus $r!s!t!\ldots$, but there may not be so many as this. If the group of permutations for which the genotype is invariant is of order N, then the number of genotypes in the set is

$$r!s!t!\ldots/N.$$

For example, the tetrasomic genotype, having two different genes at each of two loci,

$$AB/Ab/aB/ab,$$

is invariant for the permutation (Aa), and also for (Bb). The group of permutations for which it is invariant is therefore

$$I, \quad (Aa), \quad (Bb), \quad (Aa)\,(Bb),$$

of order 4, since all these four operators leave the genotype unchanged. The number of isomorphic genotypes in the set is therefore

$$2!\,2!\div 4,$$

or unity, showing that the genotype is unique.

Since the gametic series produced by different isomorphic genotypes are equivalent, the experimental examination of the gametic series supplies equivalent information in respect to the frequency of different modes of gamete formation. The use of more than one genotype of an isomorphic set will be of value only in eliminating such disturbances of the apparent gametic output as arise from unequal viability. The parallel use of non-isomorphic genotypes may, on the contrary, be essential in experimental work, for each may supply information on points on which the other is either of low value of or none.

Sets may be related by permutations of loci. Permutations of loci for which a set is invariant can involve only permutations of loci having the same numbers of different genes. Thus with two loci having the same number of different genes, any set which is transformed into itself by interchange of loci may be said to be self-conjugate. Alternatively, it must belong to a pair of conjugate sets. Genotypes belonging to conjugate sets may supply supplementary information, for not all modes of gamete formation are invariant for permutations of loci.

When there are only two loci, the sets of isomorphic genotypes may be represented by bipartitions *in plano*, the rows and columns standing for the different genes available at the two loci, and each entry showing the number of chromosomes having the two genes specified. It is easy to recognize the symmetry of such bipartitions, and to put down the number of genotypes in the set. Thus the nineteen genotypes with two different genes at each of two linked loci of a tetrasomic organism fall into four self-conjugate sets, and one pair of conjugate sets. When s is not equal to t every set belongs to a pair of conjugate sets, falling in symmetrically placed entries of the enumeration table. The remaining genotypes of table 2 are thus briefly classified in table 8.

TABLE 8

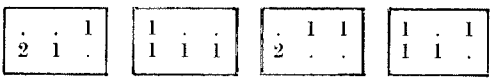
Four genotypes analogous to *AB/ab/ab/ab*, bisimplex coupling.
Self-conjugate, or invariant for interchange of loci.

Four genotypes analogous to *Ab/aB/ab/ab*, bisimplex repulsion.
Self-conjugate.

Two conjugate sets each of four genotypes analogous to *AB/Ab/ab/ab*, simplo-duplex.

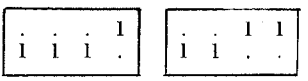
Two genotypes analogous to *AB/AB/ab/ab*, biduplex, coupling or repulsion.
Self-conjugate.

Unique genotype, *AB/Ab/aB/ab*, biduplex, neutral.
Self-conjugate.



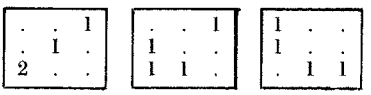
12 6 6 6

60 genotypes in four conjugate pairs of sets.



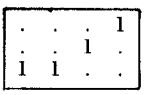
8 6

28 genotypes in two conjugate pairs of sets.



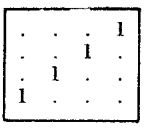
18 36 9

63 genotypes in three self-conjugate sets.



36

72 genotypes in one conjugate pair of sets.



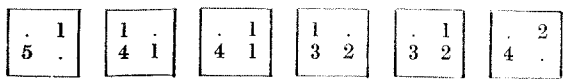
24

24 genotypes in one self-conjugate set.

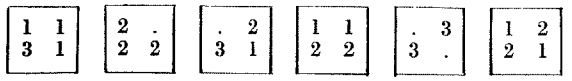
The whole of the genotypes of table 2, heterogenic at both loci, are thus comprised in eight self-conjugate sets of genotypes, and eight conjugate pairs of sets. There are thus sixteen cases requiring separate genetic investigation; these correspond with the bipartitions *in plano* of the number 4.

The bipartitions of the partible number 6, giving thirty-five self-conjugate sets and 121 conjugate pairs are set out, with some simple omissions, in table 9. These serve equally to classify the possible hexasomic genotypes involving two loci.

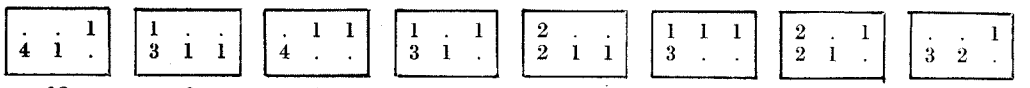
TABLE 9. BIPARTITIONS OF THE PARTIBLE NUMBER 6



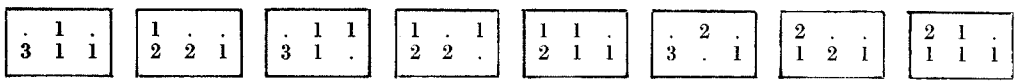
4 4 4 4 4 4 Total



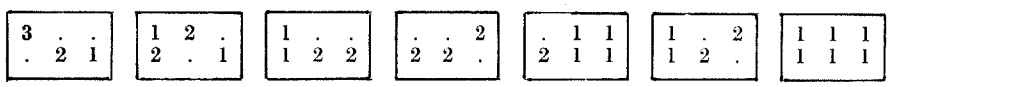
4 4 4 4 2 2 2 2 60



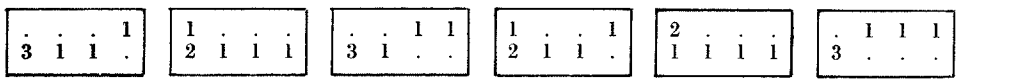
12 6 6 12 6 6 6 12



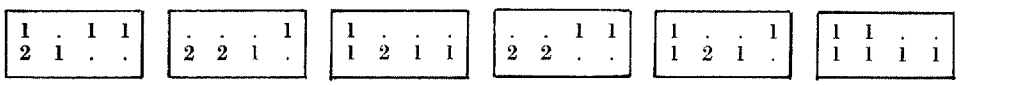
12 12 12 12 12 12 12 12



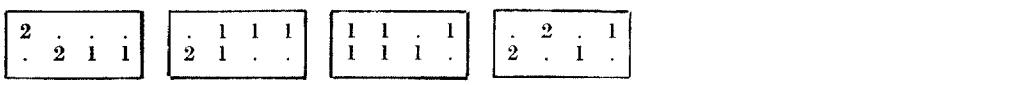
12 12 6 6 6 6 1 211



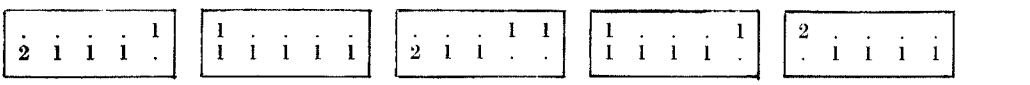
24 8 24 24 8 8



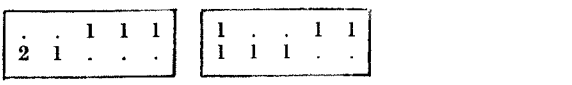
24 24 24 12 48 12



24 24 12 24 324



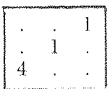
40 10 60 40 10



40 30 230

12 30 20 62

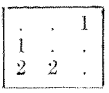
TABLE 9. (*continued*)

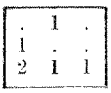

18

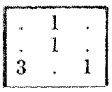

36

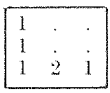

9

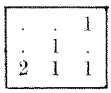

36 36

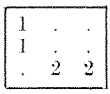

36 36

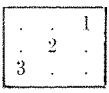

36 36


18 18

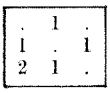

18 18

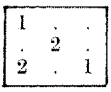

18 18

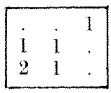

9 9

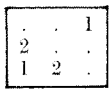

36

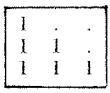

36

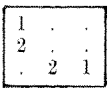

36

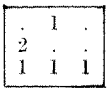

36

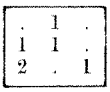

36

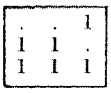

36

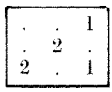

36

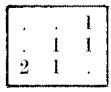

36

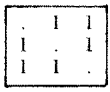

36 36

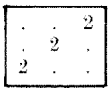

36 36

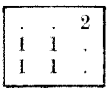

18 18


36 36

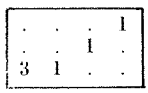

36 36

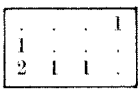

6

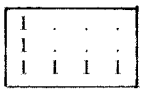

6

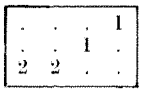

9

1038


72

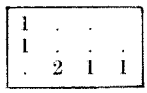

72


12

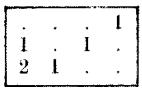

36

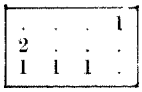

144

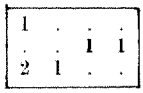

36


36

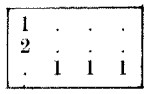

72


144


72


72


72


24


144

72

144

72

144

144

72

72

24

72

36

18

72

72

2022

180

120

15

360

360

120

180

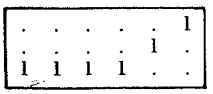
120

90

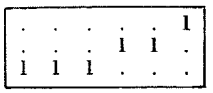
180

1725

TABLE 9. (*continued*)

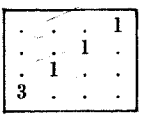


90

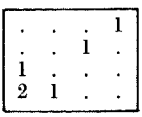


360

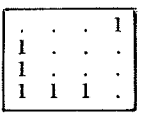
90 540



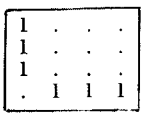
96



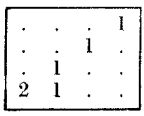
288



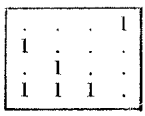
144



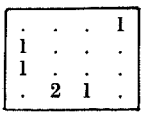
16



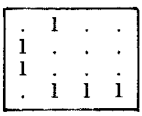
288 288



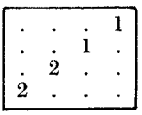
288 288



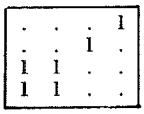
288 288



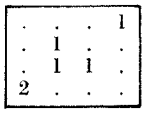
144 144



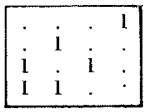
144



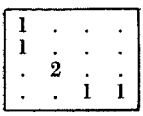
72



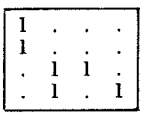
576



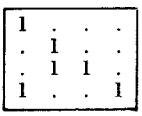
576



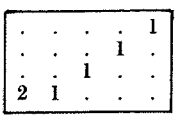
144



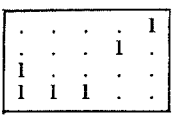
144 144



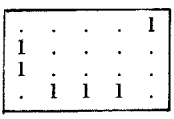
288 4648



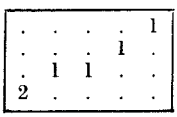
480



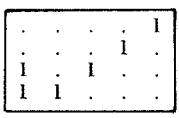
720



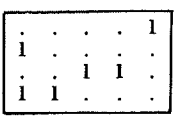
240



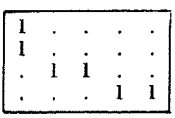
720



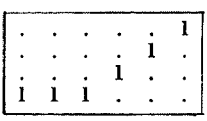
720



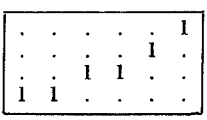
1440



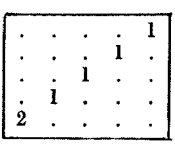
180 4500



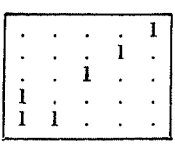
480



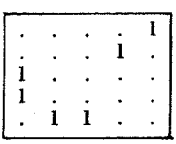
1080 1560



600



2400



1800 4800

With three linked loci in a tetrasomic organism, the isomorphic sets correspond with the tripartitions *in solido* of the number 4. These are not so easy to illustrate and enumerate. The following list accounts for the numbers enumerated in tables 5 to 7.

Three loci with two genes each. Four self-conjugate sets, and seven trios

typical formula	number of conjugate sets	number of genotypes in each set	number of genotypes	invariant permutations: allelomorphs	invariant permutations: factors
ABC/abc/abc/abc	1	8	8	—	(*ABC*) (*abc*); (*AB*) (*ab*)
Abc/aBc/abC/abc	1	8	8	—	(*ABC*) (*abc*); (*AB*) (*ab*)
ABC/ABC/abc/abc	1	4	4	(*Aa*) (*Bb*) (*Cc*)	(*ABC*) (*abc*); (*AB*) (*ab*)
ABC/Abc/aBc/abC	1	2	2	(*Aa*) (*Bb*); (*Aa*) (*Cc*); (*Bb*) (*Cc*)	(*ABC*) (*abc*); (*AB*) (*ab*)
ABc/abC/abc/abc	3	8	24	—	(*AB*) (*ab*)
ABC/abC/abc/abc	3	8	24	—	(*AB*) (*ab*)
AbC/aBC/abc/abc	3	8	24	—	(*AB*) (*ab*)
Abc/aBC/abC/abc	3	8	24	—	(*AB*) (*ab*) (*Cc*)
ABC/Abc/aBc/abc	3	8	24	—	(*AB*) (*ab*)
ABC/ABc/abc/abc	3	8	24	—	(*AB*) (*ab*)
ABC/ABc/abC/abc	3	2	6	(*Aa*) (*Bb*); (*Cc*)	(*AB*) (*ab*)
			172		

Two loci with two genes, and the third with three genes. Seven trios and six sextets

typical formula	number of conjugate sets	number of genotypes in each set	number of genotypes	invariant permutations: allelomorphs	invariant permutations: factors
ABC/abC/abc/abc'	3	12	36	(*cc'*)	(*AB*) (*ab*)
ABc/abC/abC/abc'	3	24	72	—	(*AB*) (*ab*)
Abc/aBc'/abC/abC	3	24	72	—	(*AB*) (*ab*) (*cc'*)
AbC/aBC/abc/abc'	3	12	36	(*cc'*)	(*AB*) (*ab*)
ABC/ABC/abc/abc'	3	12	36	(*cc'*)	(*AB*) (*ab*)
ABC/ABc/abC/abc'	3	12	36	(*Aa*) (*Bb*) (*cc'*)	(*AB*) (*ab*)
ABc/AbC/aBC/abc'	3	12	36	(*Aa*) (*Bb*) (*cc'*)	(*AB*) (*ab*)
AbC/aBc/abC/abc'	6	24	144	—	—
ABC/aBC/Abc/abc'	6	12	72	(*Aa*) (*cc'*)	—
ABC/AbC/abc/abc'	6	12	72	(*cc'*)	—
ABC/Abc/abc'/abC	6	24	144	—	—
ABc/AbC/abc'/abC	6	24	144	—	—
ABc/Abc'/abC/abC	6	24	144	—	—
			1044		

Two loci with two genes, and the third with four. Four trios and one sextet

typical formula	number of conjugate sets	number of genotypes in each set	number of genotypes	invariant permutations: allelomorphs	invariant permutations: factors
ABC/abC'/abc/abc'	3	16	48	(*C'c'*); (*cc'C'*)	(*AB*) (*ab*)
AbC/aBC'/abc/abc'	3	48	144	(*cc'*)	(*AB*) (*ab*) (*CC'*)
ABC/ABC'/abc/abc'	3	12	36	(*cc'*); (*CC'*); (*Aa*) (*Bb*) (*Cc*) (*C'c'*)	(*AB*) (*ab*)
ABC/AbC'/aBc/abc'	3	24	72	(*Aa*) (*Cc*) (*C'c'*); (*Bb*) (*CC'*) (*cc'*)	(*AB*) (*ab*) (*C'c*)
ABC/AbC'/abc/abc'	6	48	288	(*cc'*)	—
			588		

Two loci with three genes, and the third with two. Nine trios and three sextets

typical formula	number of conjugate sets	number of genotypes in each set	number of genotypes	invariant permutations: allelomorphs	invariant permutations: factors
ABC/ABc/abc/a'b'c	3	36	108	(*aa'*) (*bb'*)	(*AB*) (*ab*) (*a'b'*)
ABc/ABc/abC/a'b'c	3	72	216	—	(*AB*) (*ab*) (*a'b'*)
ABC/ABc/abC/a'b'c	3	36	108	(*aa'*) (*bb'*) (*Cc*)	(*AB*) (*ab*) (*a'b'*)
ABC/ABC/abc/a'b'c	3	36	108	(*aa'*) (*bb'*)	(*AB*) (*ab*) (*a'b'*)
ABC/Abc/aBc/a'b'c	3	72	216	—	(*AB*) (*ab*) (*a'b'*)
ABc/Abc/aBc/a'b'C	3	72	216	—	(*AB*) (*ab*) (*a'b'*)
ABC/Abc/aBc/a'b'C	3	72	216	—	(*AB*) (*ab*) (*a'b'*)
AbC/Ab'C/aBc/a'Bc	3	18	54	(*aa'*); (*bb'*)	(*AB*) (*ab*) (*a'b'*) (*Cc*)
AbC/Ab'c/aBC/a'Bc	3	36	108	(*aa'*) (*bb'*) (*Cc*)	(*AB*) (*ab*) (*a'b'*)
ABc/AbC/aBc/a'b'c	6	72	432	—	—
ABC/AbC/aBc/a'b'c	6	72	432	—	—
AbC/Ab'c/aBc/a'Bc	6	36	216	(*aa'*)	—
			2430		

Loci with two, three or four genes respectively. Four sextets

typical formula	number of conjugate sets	number of genotypes in each set	number of genotypes	invariant permutations: allelomorphs	invariant permutations: factors
$Abc_1/a'bc_2/abc_3/abc_4$	6	144	864	(c_3c_4)	—
$Abc_1/a'bc_2/aBc_3/abc_4$	6	144	864	(*aa'*) (c_2c_4)	—
$ABc_1/a'Bc_2/abc_3/abc_4$	6	72	432	(c_3c_4); (*Aa'*) (c_1c_2)	—
$ABc_1/a'bc_2/aBc_3/abc_4$	6	144	864	(*Aa'*) (*Bb*) (c_1c_2) (c_3c_4)	—
			3024		

Two loci with four genes, and the third with two. Two trios

typical formula	number of conjugate sets	number of genotypes in each set	number of genotypes	invariant permutations: allelomorphs	invariant permutations: factors
$Ab_1c_1/ab_2c_2/ab_3c_3/ab_4c_4$	3	192	576	($b_2b_3b_4$) ($c_2c_3c_4$); (b_2b_3) (c_2c_3)	(b_1c_1) (b_2c_2) (b_3c_3) (b_4c_4)
$Ab_1c_1/Ab_2c_2/ab_3c_3/ab_4c_4$	3	144	432	(b_1b_2) (c_1c_2); (b_3b_4) (c_3c_4) (*Aa*) (b_1b_3) (b_2b_4) (c_1c_3) (c_2c_4)	(b_1c_1) (b_2c_2) (b_3c_3) (b_4c_4)
			1008		

Three different genes at each locus. Three self-conjugate sets and three trios

typical formula	number of conjugate sets	number of genotypes in each set	number of genotypes	invariant permutations: allelomorphs	invariant permutations: factors
ABC/ABC/abc/a'b'c'	1	108	108	(*aa'*) (*bb'*) (*cc'*)	(*AB*) (*ab*) (*a'b'*); (*ABC*) (*abc*) (*a'b'c'*)
ABC/Abc/aBc'/a'b'C	1	216	216	—	(*AB*) (*ab*) (*a'b'*) (*cc'*); (*ABC*) (*ab'c*) (*a'bc'*)
ABc/AbC/aBc/a'b'c'	1	216	216	—	(*AB*) (*ab*) (*a'b'*); (*ABC*) (*abc*) (*a'b'c'*)
ABC/AbC/aBc/a'b'c'	3	216	648	—	(*AC*) (*ac*) (*a'c'*)
ABc/Abc'/aBC/a'b'C	3	216	648	—	(*AC*) (*ac*) (*a'c'*) (*bb'*)
AbC/Ab'C/aBc/a'Bc'	3	54	162	(*bb'*); (*aa'*) (*cc'*)	(*AC*) (*ac*) (*a'c'*) (*bb'*)
			1998		

Three genes at two loci, and four at the third. Three trios

typical formula	number of conjugate sets	number of genotypes in each set	number of genotypes	invariant permutations: allelomorphs	invariant permutations: factors
$ABc_1/ABc_2/abc_3/a'b'c_4$	3	216	648	(c_1c_2); (aa') (bb') (c_3c_4)	(AB) (ab) $(a'b')$
$ABc_1/Abc_2/aBc_3/a'b'c_4$	3	864	2592	—	(AB) (ab) $(a'b')$ (c_2c_3)
$Abc_1/Ab'c_2/aBc_3/a'Bc_4$	3	216	648	(bb') (c_1c_2); (aa') (c_3c_4)	(AB) (ab) $(a'b')$ (c_1c_3) (c_2c_4)
			3888		

Four genes at two loci, and three at the third. One trio

typical formula	number of conjugate sets	number of genotypes in each set	number of genotypes	invariant permutations: allelomorphs	invariant permutations: factors
$Ab_1c_1/Ab_2c_2/ab_3c_3/a'b_4c_4$	3	864	2592	(b_1b_2) (c_1c_2); (b_3b_4) (c_3c_4) (aa')	(b_1c_1) (b_2c_2) (b_3c_3) (b_4c_4)

Four genes at each locus. One self-conjugate set

typical formula	number of conjugate sets	number of genotypes in each set	number of genotypes	invariant permutations: allelomorphs	invariant permutations: factors
$a_1b_1c_1/a_2b_2c_2/a_3b_3c_3/a_4b_4c_4$	1	576	576	all suffices	all loci

There are therefore 200 isomorphic sets in all, comprising 8 self-conjugate, 36 trios and 14 sextets, or 58 essentially different kinds of genotype.

Addendum on certain enumerations of isomorphic sets and genera

For numbers of loci exceeding three, the numbers of genotypes are large and their relationships complex. The enumeration of the isomorphic sets presents a problem somewhat analogous to that of the enumeration of the modes of gamete formation, discussed in § 4. The following results for l loci in tetrasomic organisms may be noted:

	number of isomorphic sets	
all simplex (3 1)	$(4^l+6.2^l+8)/24$	$l>0$
all duplex (2^2)	$(3^l+3)/6$	$l>0$
all digenic	$(7^l+9.3^l+14)/24$	
all trigenic (2 1^2)	$(6^l+9.2^l)/24$	$l>0$
digenic or trigenic	$(13^l+9.5^l+14)/24$	
all heterogenic	$(14^l+9.6^l+14.2^l)/24$	

Numerically these give, so far as six loci:

$l=$	1	2	3	4	5	6
all simplex (3 1)	1	2	5	15	51	187
all duplex (2^2)	1	2	5	14	41	122
all digenic	2	6	25	131	792	5,176
all trigenic (2 1^2)	1	3	12	60	336	1,968
digenic or trigenic	3	17	139	1,425	16,643	206,977
all partitions	4	24	200	2,096	25,344	331,264

It is curious that the number using all partitions (other than (4)) should be just 2^l times the number using digenic partitions only.

The isomorphic sets enumerated above are distinct in that no one can be transformed into another by gene substitutions at the same locus; different sets may, however, be related by permutation of loci, and so many as $l!$ different sets may thus belong to the same genus.

The enumeration of such genera involves expressions analogous to those of the Eulerian generating functions for the enumeration of partitions. Indeed, for factors all simplex the number is merely the number of partitions of not more than four parts, while for factors all duplex it is the number of partitions of not more than three parts. The others are more complex and need somewhat intricate arguments to build them up.

Generating functions for numbers of genera

all simplex	$1/(1-x)\,(1-x^2)\,(1-x^3)\,(1-x^4)$
all duplex	$1/(1-x)\,(1-x^2)\,(1-x^3)$
all digenic	$(1+x^3+x^4+x^5+x^6+x^9)/(1-x)^2\,(1-x^2)^2\,(1-x^3)^2\,(1-x^4)$
all trigenic	$(1+x^3+x^4+x^5+x^6+x^9)/(1-x)\,(1-x^2)^2\,(1-x^3)^2\,(1-x^4)$
digenic or trigenic	$\left.\begin{matrix}1+2x^2+10x^3+19x^4+34x^5+66x^6+96x^7\\ +128x^8+152x^9\end{matrix}\right\}/(1-x)^3\,(1-x^2)^4\,(1-x^3)^4\,(1-x^4)^2$
all heterogenic	$\left.\begin{matrix}+128x^{10}+96x^{11}+66x^{12}+34x^{13}+19x^{14}\\ +10x^{15}+2x^{16}+x^{18}\end{matrix}\right\}/(1-x)^4\,(1-x^2)^4\,(1-x^3)^4\,(1-x^4)^2$

The numerical values are:

Numbers of genera for different numbers of loci

$l=$	1	2	3	4	5	6
all simplex	1	2	3	5	6	9
all duplex	1	2	3	4	5	7
all digenic	2	5	11	22	40	72
all trigenic	1	3	6	11	18	32
digenic or trigenic	3	12	42	132	380	1030
all heterogenic	4	16	58	190	570	1600

4. The enumeration of modes of gamete formation

The enumeration of the modes of gamete formation is essential for the study of linkage in polysomic organisms, for, given the mode of formation, any one of the multiply-heterogenic genotypes enumerated in the previous section yields a known gametic series merely by applying Mendel's law. Moreover, the frequency of different modes of formation is very generally independent of the genotype, so that if determined experimentally for chosen genotypes the frequencies ascertained give the gametic series applicable to all. In disomic inheritance, the number of modes of formation distinguishable by the use of l linked loci is 2^{l-1}, requiring $2^{l-1}-1$ independent parameters for the specification of all frequencies. With polysomic inheritance at a single locus only one parameter is required for tetrasomic and hexasomic organisms, while octosomic and decasomic organisms would require two.

The recognition of the modes of gamete formation may be illustrated with the case of two linked loci in a tetrasomic organism. If all four genes at each locus are different, the number of different possible chromosomes is 16, and the number of different possible gametes is

$$\tfrac{1}{2}.16.17 = 136.$$

Every possible gamete contains a selection of portions from one or more of the four parent chromosomes. Permutations of the four chromosomes from which it arises will generate a set of gametes having the same mode of origin. All possible gametes may thus be accounted for. Table 10 shows the results from the parent $a_1b_1/a_2b_2/a_3b_3/a_4b_4$.

TABLE 10. MODES OF GAMETE FORMATION FOR TWO LOCI IN A TETRASOMIC ORGANISM

mode of formation	typical gamete	number of types of gamete	
1	a_1b_1/a_2b_2	6	
2	a_1b_1/a_1b_1	4	
3	a_1b_1/a_2b_3	24	
4	a_1b_1/a_1b_2	12	conjugate pair
5	a_1b_1/a_2b_1	12	
6	a_1b_2/a_3b_4	12	
7	a_2b_1/a_3b_1	12	conjugate pair
8	a_1b_2/a_1b_3	12	
9	a_1b_2/a_2b_3	24	
10	a_1b_2/a_1b_2	12	
11	a_1b_2/a_2b_1	6	
		136	

There are therefore eleven modes of gamete formation to be distinguished, of which seven are invariant for interchange of loci, while four are in two conjugate pairs.

The number of modes of formation in tetrasomic organisms distinguishable by means of l linked loci may be found as follows. In the two gametes to be selected there are $2l$ places to be filled. The number of ways of dividing these $2l$ places into exactly s divisions is

$$\frac{1}{s!}\Delta^s\sigma^{2l} \quad (\sigma = 0).$$

The sum of these values for s equal to 1, 2, 3 and 4 will therefore enumerate all possible gametic types, but will count twice those which are altered by interchange of the two chromosomes. The true number may therefore be obtained by adding the number unaltered by interchange of the two chromosomes, and dividing the whole by 2.

In any gamete which can be transformed into itself by interchange of chromosomes, and subsequent permutation of the sources from which its portions are drawn, the two chromosomes must be divided alike. Since each contains l loci the number of ways in which they can be divided alike into s divisions is

$$\frac{1}{s!}\Delta^s\sigma^l \quad (\sigma = 0).$$

Each such subdivision will yield a number of different types according to the value of s.

If $s = 1$, there will be two symmetrical types, with chromosomes alike and unlike respectively, as in (1) and (2) above.

If $s = 2$, there will be five symmetrical types, corresponding with numbers (6), (7), (8), (10) and (11).

If $s = 3$, there will be four having three sources only, and six having four sources, or ten in all.

If $s = 4$, there will be ten having four sources, this being the number of elements, including the identity, of the permutation group of four objects, of which the squares reduce to the identity.

The number of symmetrical gametes which can be selected is therefore

$$(2\Delta+\tfrac{5}{2}\Delta^2+\tfrac{10}{6}\Delta^3+\tfrac{10}{24}\Delta^4)\,\sigma^l \quad (\sigma=0).$$

Half the sum of the two expressions now obtained is readily found to be

$$\tfrac{1}{48}.16^l+\tfrac{1}{3}.4^l+\tfrac{1}{3}.$$

When l takes the values 1, 2, 3, 4 the numbers of modes of formation are thus

$$2,\quad 11,\quad 107,\quad 1451,$$

increasing thereafter nearly 16-fold for each additional locus.

The 107 modes of formation for three linked genes consist of five self-conjugate, twenty-two trios and six sextets. They may be put on record as shown in table 11.

TABLE 11. MODES OF FORMATION FOR THREE LINKED LOCI, IN A TETRASOMIC ORGANISM

type of gamete	number of sets	kinds of gamete per set	kinds of gamete
$a_1b_1c_1/a_1b_1c_1$	1	4	4
$a_1b_1c_1/a_2b_2c_2$	1	6	6
$a_1b_2c_3/a_1b_2c_3$	1	24	24
$a_1b_2c_3/a_2b_3c_1$	1	24	24
$a_1b_1c_1/a_2b_3c_4$	1	24	24
$a_1b_1c_1/a_1b_1c_2$	3	12	36
$a_1b_1c_1/a_1b_2c_2$	3	12	36
$a_1b_1c_2/a_1b_1c_2$	3	12	36
$a_1b_1c_2/a_1b_2c_1$	3	12	36
$a_1b_2c_1/a_2b_1c_2$	3	6	18
$a_1b_1c_2/a_1b_2c_2$	3	12	36
$a_1b_1c_1/a_1b_2c_3$	3	24	72
$a_1b_1c_2/a_1b_1c_3$	3	12	36
$a_1b_2c_2/a_2b_2c_3$	3	24	72
$a_1b_1c_1/a_2b_2c_3$	3	24	72
$a_1b_1c_2/a_3b_3c_1$	3	24	72
$a_1b_2c_2/a_1b_3c_3$	3	12	36
$a_1b_2c_3/a_1b_3c_2$	3	12	36
$a_1b_1c_2/a_2b_3c_3$	3	24	72
$a_1b_2c_3/a_3b_3c_4$	3	24	72
$a_1b_2c_2/a_1b_3c_4$	3	24	72
$a_1b_1c_2/a_3b_3c_4$	3	12	36
$a_1b_2c_3/a_2b_1c_4$	3	12	36
$a_1b_2c_3/a_1b_2c_4$	3	12	36
$a_1b_2c_2/a_3b_3c_4$	3	24	72
$a_1b_2c_3/a_4b_2c_1$	3	24	72
$a_1b_2c_3/a_2b_3c_4$	3	24	72
$a_1b_1c_2/a_1b_3c_3$	6	24	144
$a_1b_2c_2/a_1b_3c_2$	6	24	144
$a_1b_1c_2/a_1b_2c_3$	6	24	144
$a_1b_2c_1/a_3b_1c_2$	6	24	144
$a_1b_2c_3/a_1b_1c_4$	6	24	144
$a_1b_1c_2/a_2b_3c_4$	6	24	144
	107		2080

$$107=\tfrac{1}{48}.16^3+\tfrac{1}{3}.4^3+\tfrac{1}{3},\quad 2080=\tfrac{1}{2}.4^3(4^3+1).$$

It is not so easy to calculate the number of modes of gamete formation for l linked loci in hexaploids. The expression

* $$\frac{1}{6!3!}.210^l+\tfrac{1}{288}.60^l+\tfrac{1}{1440}.30^l+\tfrac{1}{108}.24^l+\tfrac{1}{96}.20^l+\tfrac{1}{32}.10^l+\tfrac{67}{864}.6^l+\tfrac{2}{27}.3^l+\tfrac{7}{48}.2^l$$

is integral for all values of l, and correct for $l=1$, $l=2$; for $l=3$ it gives 3175 modes, which may perhaps be checked by direct enumeration.

* This has been checked and found correct by an independent method. - R.A.F., February, 1949. (see Paper 237).

If three appearances of the same gene were admissible, the number of types of gamete produced by a hexaploid organism completely heterogenic at each of l linked loci would be

$$6^l(6^l+1)\,(6^l+2) \div 6.$$

Of these the number using three identical genes at each of r loci is

$$6^r \frac{l!}{(l-r)!\,r!}\, 6^{l-r}(6^{l-r}+1)\,(6^{l-r}+2) \div 6.$$

Hence the number of gametic types in which triplicate genes are never used is

$$\begin{aligned} & 6^l(6^l+1)\,(6^l+2) \div 6, \\ -6l\,.\,& 6^{l-1}(6^{l-1}+1)\,(6^{l-1}+2) \div 6, \\ +6^2\,.\,\tfrac{1}{2}l(l-1)\,.\,& 6^{l-2}(6^{l-2}+1)\,(6^{l-2}+2) \div 6, \\ & \text{etc.}, \end{aligned}$$

or, summing the series, $6^{l-1}\{(6^2-1)^l+3(6-1)^l\} = 6^{l-1}5^l(7^l+3)$.

If $l = 1$, this is 50, namely, the twenty types made of the selection of three genes out of six, and the thirty in which one gene is chosen to appear twice in conjunction with a single representative of each of the others. When $l = 2$ the number is 7800; this is distributed among forty modes of gamete formation, of which twenty are invariant for interchange of loci, while there are ten conjugate pairs, as shown in table 12.

TABLE 12 MODES OF GAMETE FORMATION FOR TWO LOCI IN A HEXASOMIC ORGANISM

type of gamete	number of sets	kinds of gamete per set	kinds of gamete
$a_1b_1/a_2b_2/a_3b_3$	1	20	20
$a_1b_1/a_1b_1/a_2b_2$	1	30	30
$a_1b_1/a_2b_2/a_3b_4$	1	180	180
$a_1b_1/a_2b_1/a_2b_2$	1	30	30
$a_1b_1/a_1b_1/a_2b_3$	1	120	120
$a_1b_1/a_2b_3/a_4b_5$	1	360	360
$a_1b_1/a_2b_3/a_2b_3$	1	120	120
$a_1b_1/a_2b_3/a_3b_2$	1	60	60
$a_1b_1/a_2b_3/a_3b_4$	1	360	360
$a_1b_2/a_2b_2/a_2b_3$	1	120	120
$a_1b_1/a_1b_2/a_2b_1$	1	30	30
$a_1b_2/a_3b_4/a_5b_6$	1	120	120
$a_1b_2/a_1b_2/a_3b_4$	1	360	360
$a_1b_2/a_2b_1/a_3b_4$	1	180	180
$a_1b_2/a_3b_2/a_3b_4$	1	360	360
$a_1b_2/a_3b_1/a_3b_2$	1	120	120
$a_1b_2/a_2b_1/a_2b_1$	1	30	30
$a_1b_2/a_2b_3/a_4b_5$	1	720	720
$a_1b_2/a_2b_3/a_3b_4$	1	360	360
$a_1b_2/a_2b_3/a_3b_1$	1	40	40
$a_1b_1/a_2b_2/a_2b_3$	2	120	240
$a_1b_1/a_2b_3/a_2b_4$	2	180	360
$a_1b_1/a_1b_2/a_3b_4$	2	360	720
$a_1b_1/a_1b_2/a_3b_2$	2	120	240
$a_1b_1/a_1b_2/a_2b_3$	2	120	240
$a_1b_2/a_1b_3/a_4b_5$	2	360	720
$a_1b_2/a_1b_3/a_3b_4$	2	360	720
$a_1b_2/a_1b_3/a_4b_1$	2	180	360
$a_1b_2/a_1b_3/a_2b_1$	2	120	240
$a_1b_2/a_1b_2/a_2b_3$	2	120	240
	40		7800

5. The gametic matrices

The chief practical obstacle to work with polysomic species lies in the recognition of the genotype of the gametes produced. With disomic species a backcross to a multiple recessive will produce offspring from the phenotype of which the genotype of the gamete responsible can be recognized. With polysomic reproduction, however, a gamete may contain one dominant gene, or perhaps more, and the phenotype of the offspring, unless dominance is incomplete, does not specify the gamete which produced it. Moreover, even if dominance were completely absent, and simplex, duplex, triplex and quadruplex organisms could all be recognized, still no distinction could be made between the offspring from such gametes as *Ab/aB* and *AB/ab*, which contain the same genes differently associated.

The genotype of the gamete can, however, usually be recognized by performing a second backcross, i.e. by crossing to a multiple recessive in two successive generations, so that each gamete is recognized, not by the appearance of a single individual, but by the frequency distribution observed in a family of 50 or 100 of its offspring. This method of the second backcross, though laborious and time-consuming, is extremely powerful in unravelling the complex situations of polysomic inheritance. It will therefore be supposed that the gametic output of chosen genotypes is tested in this way.

For tetrasomic inheritance the number of gametic genotypes produced by any organism is easily calculated. For a tetrasomic organism having r genes at one locus and s at a second it is

$$\tfrac{1}{2}rs(rs+1),$$

if there were also a choice of t genes at a third locus, the number would be

$$\tfrac{1}{2}rst(rst+1),$$

and so on. For hexasomic inheritance the number will depend also on how many genes are only singly represented.

Putting r, s equal to 2, 3 and 4, and $t = 1$, then the numbers of gametic genotypes will be

		r		
		2	3	4
	2	10	21	36
s	3	21	45	78
	4	36	78	136

Since the gametic output is specified for each mode of gamete formation in question, of which in this case there are 11, any given genotype is associated with a matrix of 11 rows, each of which gives the frequency distribution of the different possible gametes. The matrix has thus as many columns as there are gametic genotypes.

The frequency distributions for different modes of formation are not, however, entirely independent, for, at each locus, the average number of genes of each kind in the gametes must be half the number in the parent zygote. The numbers of restrictions due to this cause are shown in similar form below:

2	3	4
3	4	5
4	5	6

The rank of the gametic matrix cannot exceed the difference between the corresponding entries in the first table and the second. Thus the nineteen genotypes having only two genes

at each of two loci have each a gametic matrix of rank no greater than 8. Since only a matrix of rank 11 could determine the frequencies of the eleven modes of gamete formation, it is clear that no one of these used alone could suffice for the purpose. The information supplied by each of these genotypes must be somewhat deficient. An examination of the matrices is needed to ascertain to what extent they suffer from deficiencies common to all, and to what extent some may supplement the deficiencies of others.

A gametic matrix need be prepared only for each set of isomorphic genotypes, for permutation of the genes will permute all rows of the matrix alike, and therefore permute whole columns. Hence for different isomorphic genotypes the same series of frequencies will be observed, though these will refer to different gametic genotypes. There is, of course, no such relationship between the frequencies of phenotypes of the first backcross generation.

The gametic matrix of a simplo-duplex tetrasomic organism is shown in table 13. There are 11 rows and 10 columns, but on examination the rank of the matrix is found to be only 7. It must therefore be possible to find a 4×11 matrix of rank 4 which will premultiply it to zero. This is shown in table 14. This may be called the deficiency matrix, as it specifies the extent and nature of the deficiency of the information available from this group of genotypes for the evaluation of the frequencies of the eleven modes of gamete formation. The sum of the products of the elements of each row of the deficiency matrix with each column of the gametic matrix vanishes. In other words, any multiple of any row of the deficiency matrix may be added to the frequencies of the modes of gamete formation without altering the gametic output.

TABLE 13. GAMETIC MATRIX FOR SIMPLO-DUPLEX GENOTYPES

mode of formation	composition of gamete: $\frac{AB}{AB}$	$\frac{aB}{aB}$	$\frac{ab}{ab}$	$\frac{Ab}{Ab}$	$\frac{AB}{Ab}$	$\frac{aB}{AB}$	$\frac{ab}{aB}$	$\frac{Ab}{ab}$	$\frac{AB}{ab}$	$\frac{Ab}{aB}$	divisor
1. a_1b_1/a_2b_2	.	.	2	.	2	.	.	4	4	.	12
2. a_1b_1/a_1b_1	3	.	6	3	.	.	.	.	.	.	12
3. a_1b_1/a_2b_3	.	.	1	1	1	.	1	4	3	1	12
4. a_1b_1/a_1b_2	.	.	4	2	4	.	2	.	.	.	12
5. a_1b_1/a_2b_1	1	.	2	1	.	2	.	6	.	.	12
6. a_1b_2/a_3b_4	.	.	.	2	.	.	2	4	2	2	12
7. a_2b_1/a_3b_1	.	1	1	2	.	2	.	6	.	.	12
8. a_1b_2/a_1b_3	.	.	2	4	2	.	4	.	.	.	12
9. a_1b_2/a_2b_3	.	.	1	1	1	.	1	4	1	3	12
10. a_1b_2/a_1b_2	1	2	4	5	.	.	.	.	.	.	12
11. a_1b_2/a_2b_1	.	.	2	.	2	.	.	4	.	4	12

Parental genotype is taken to be *AB/Ab/ab/ab*.
The interchanges (*Aa*) and (*Bb*) generate a set of four equivalent genotypes. A second set of four is obtained by interchanging the loci, and making the row interchange (45) (78).

TABLE 14. DEFICIENCY MATRIX FOR SIMPLO-DUPLEX GENOTYPES

mode of gametic formation: $\frac{a_1b_1}{a_2b_2}$	$\frac{a_1b_1}{a_1b_1}$	$\frac{a_1b_1}{a_3b_3}$	$\frac{a_1b_1}{a_1b_2}$	$\frac{a_1b_1}{a_2b_1}$	$\frac{a_1b_2}{a_3b_4}$	$\frac{a_2b_1}{a_3b_1}$	$\frac{a_1b_2}{a_1b_3}$	$\frac{a_1b_2}{a_2b_3}$	$\frac{a_1b_2}{a_1b_2}$	$\frac{a_1b_2}{a_2b_1}$
1	.	−2	.	.	1	.	.	.	.	.
.	.	.	.	.	1	.	.	−2	.	1
.	1	.	.	−2	.	2	.	.	−1	.
1	.	.	−2	.	−2	.	2	.	.	1

As arranged in table 14 the first two components of the deficiency matrix are seen not to involve the two pairs 4 and 5, and 7 and 8 of modes of formation, which are interchanged by interchange of loci. These two components are therefore common to the deficiency matrices of the two conjugate sets of genotypes. Using genotypes from both these conjugate sets, however, the deficiency is reduced to these two elements.

The gametic matrices of the two other sets of four isomorphic genotypes, shown in table 15, typified by the genotypes bisimplex coupling, and bisimplex repulsion, are both of rank 8,

TABLE 15. GAMETIC MATRIX FOR BISIMPLEX COUPLING, AND SIMILAR GENOTYPES

composition of gamete

mode of formation	$\frac{AB}{AB}$	$\frac{aB}{aB}$	$\frac{ab}{ab}$	$\frac{Ab}{Ab}$	$\frac{AB}{Ab}$	$\frac{aB}{AB}$	$\frac{ab}{aB}$	$\frac{Ab}{ab}$	$\frac{AB}{ab}$	$\frac{Ab}{aB}$	divisor
1	.	.	6	.	.	.	.	.	6	.	12
2	3	.	9	.	.	.	.	.	.	.	12
3	.	.	3	.	.	.	3	3	3	.	12
4	.	.	6	.	3	.	3	.	.	.	12
5	.	.	6	.	.	3	.	3	.	.	12
6	.	.	.	.	.	.	6	6	.	.	12
7	.	3	3	.	.	.	.	6	.	.	12
8	.	.	3	3	.	.	6	.	.	.	12
9	.	.	3	.	.	.	3	3	.	3	12
10	.	3	6	3	.	.	.	.	.	.	12
11	.	.	6	.	.	.	.	.	.	6	12

Parental genotype *AB/ab/ab/ab*, belonging to a self-conjugate set of 4.

Deficiency matrix

mode of gamete formation

1	2	3	4	5	6	7	8	9	10	11
1	.	−2	.	.	1	.	.	.	.	.
.	.	.	.	.	1	.	.	−2	.	1
.	.	.	.	.	1	−1	−1	.	1	.

Gametic matrix for bisimplex repulsion, and similar genotypes

composition of gamete

mode of formation	$\frac{AB}{AB}$	$\frac{aB}{aB}$	$\frac{ab}{ab}$	$\frac{Ab}{Ab}$	$\frac{AB}{Ab}$	$\frac{aB}{AB}$	$\frac{ab}{aB}$	$\frac{Ab}{ab}$	$\frac{AB}{ab}$	$\frac{Ab}{aB}$	divisor
1	.	.	2	.	.	.	4	4	.	2	12
2	.	3	6	3	.	.	.	.	.	.	12
3	.	.	3	.	.	.	3	3	1	2	12
4	.	.	4	2	1	.	5	.	.	.	12
5	.	2	4	.	.	1	.	5	.	.	12
6	.	.	4	.	.	.	2	2	2	2	12
7	.	1	5	.	.	2	.	4	.	.	12
8	.	.	5	1	2	.	4	.	.	.	12
9	.	.	3	.	.	.	3	3	2	1	12
10	1	2	7	2	.	.	.	.	.	.	12
11	.	.	2	.	.	.	4	4	2	.	12

Parental genotype *Ab/aB/ab/ab*, belonging to a self-conjugate set of 4.

Deficiency matrix

mode of gamete formation

1	2	3	4	5	6	7	8	9	10	11
1	.	−2	.	.	1	.	.	.	.	.
.	.	.	.	.	1	.	.	−2	.	1
−1	−1	.	2	2	1	−1	−1	.	.	−1

and have 3×11 deficiency matrices. It will be seen that all four sets have two elements of deficiency in common, and no two of them have any other common element. Consequently, genotypes belonging to any two of these four sets may be used to reduce the deficiency to two elements, or to supply eight equations for the ten unknowns.

The two remaining sets (table 16 below), which comprise the three biduplex genotypes, supply one but not both of the missing equations. The set of two, biduplex coupling and repulsion, has a matrix of rank 6, while the unique genotype biduplex neutral has one only of rank 4. The former is the more useful, especially as its two genotypes may be made up at will by crossing homogenic organisms. Of these doubly digenic genotypes a set of three belonging to different isomorphic sets, of which one is biduplex, will leave only one degree of indeterminacy. Moreover, the information lacking from such a test is just that which is not needed in predicting the gametic output of any of these nineteen genotypes.

If, from a simplo-duplex organism, instead of breeding a second generation by the double backcross, all four genotypes of the set had been used, and the first generation classified in four phenotypes, an 11×16 phenotypic matrix would be obtained. This, of course, must have the same elements of deficiency as the gametic matrix of the same parents, but it may have others in addition. On examination it is found to be of rank 6, and so to have one additional element of deficiency. This is easily seen to be the element

$$1 \ldots\ldots\ldots - 1,$$

for the first and last rows of the gametic matrix are equivalent save for the interchange of the gametes *Ab*/*aB* and *AB*/*ab*. Since this is true of the gametic matrices of all six sets, it is clear that the phenotypic classification of the first backcross, however thoroughly carried out, will always leave a second element of deficiency, which will leave one ignorant of the frequencies with which the different kinds of gamete are produced.

Since the first two modes of gamete formation involve no recombination, while the next three involve one recombinant chromosome, and the last six two recombinant chromosomes each, the recombination fraction is equal to the sum of the frequencies of modes 6 to 11, together with half the frequency of modes 3 to 5. It may be noticed that the vector

$$0\ 0\ \tfrac{1}{2}\ \tfrac{1}{2}\ \tfrac{1}{2}\ 1\ 1\ 1\ 1\ 1\ 1,$$

regarded as a column matrix, is premultiplied to zero by all the deficiency matrices obtained above, except that for the unique genotype of biduplex neutral. Consequently, the recombination fraction can be ascertained by a double backcross from all except one of the doubly heterogenic genotypes. Moreover, all genotypes will suffice to ascertain the frequencies of double reduction at the two loci severally, given by the vectors

$$0\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ 0\ 1\ 0 \quad \text{and} \quad 0\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ 0.$$

These three ratios are therefore easily ascertainable by double backcrossing. With single backcrossing, however, the frequency of recombination must always be indeterminate, owing to the equivalence of the first and eleventh modes of formation, having 0 and 2 recombinant chromosomes respectively. The method of determining the recombination fraction, which first presents itself, by making up biduplex organisms in coupling and repulsion, and backcrossing these to double recessives (or intercrossing them) as practised by

TABLE 16. GAMETIC MATRIX FOR BIDUPLEX COUPLING, AND REPULSION

mode of formation	composition of gamete: $\frac{AB}{AB}$	$\frac{aB}{aB}$	$\frac{ab}{ab}$	$\frac{Ab}{Ab}$	$\frac{AB}{Ab}$	$\frac{aB}{AB}$	$\frac{ab}{aB}$	$\frac{Ab}{ab}$	$\frac{AB}{ab}$	$\frac{Ab}{aB}$	divisor
1	2	.	2	.	.	.	.	.	8	.	12
2	6	.	6	.	.	.	.	.	.	.	12
3	.	.	.	.	2	2	2	2	4	.	12
4	2	.	2	.	4	.	4	.	.	.	12
5	2	.	2	.	.	4	.	4	.	.	12
6	.	2	.	2	.	.	.	.	4	4	12
7	.	2	.	2	.	4	.	4	.	.	12
8	.	2	.	2	4	.	4	.	.	.	12
9	.	.	.	.	2	2	2	2	.	4	12
10	2	4	2	4	.	.	.	.	.	.	12
11	2	.	2	.	.	.	.	.	.	8	12

Parental genotype *AB/AB/ab/ab*, belonging to a self-conjugate set of 2.

Deficiency matrix

mode of gamete formation

1	2	3	4	5	6	7	8	9	10	11
1	.	−2	.	.	.	.	.	2	.	−1
.	1	.	−2	.	.	.	2	.	−1	.
.	1	.	.	−2	.	2	.	.	−1	.
.	.	.	.	.	.	1	1	−2	−1	1
1	.	.	.	−2	−2	2	.	.	.	1

Gametic matrix for biduplex, neutral

mode of formation	composition of gamete: $\frac{AB}{AB}$	$\frac{aB}{aB}$	$\frac{ab}{ab}$	$\frac{Ab}{Ab}$	$\frac{AB}{Ab}$	$\frac{aB}{AB}$	$\frac{ab}{aB}$	$\frac{Ab}{ab}$	$\frac{AB}{ab}$	$\frac{Ab}{aB}$	divisor
1	.	.	.	.	4	4	4	4	4	4	24
2	6	6	6	6	.	.	.	.	.	.	24
3	1	1	1	1	2	2	2	2	6	6	24
4	2	2	2	2	8	.	8	.	.	.	24
5	2	2	2	2	.	8	.	8	.	.	24
6	.	.	.	.	4	4	4	4	4	4	24
7	2	2	2	2	.	8	.	8	.	.	24
8	2	2	2	2	8	.	8	.	.	.	24
9	1	1	1	1	2	2	2	2	6	6	24
10	6	6	6	6	.	.	.	.	.	.	24
11	.	.	.	.	4	4	4	4	4	4	24

Parental genotype *AB/Ab/aB/ab*, unique genotype.

Deficiency matrix

mode of gamete formation

1	2	3	4	5	6	7	8	9	10	11
1	.	.	.	.	.	.	.	.	.	−1
.	1	.	.	.	.	.	.	.	−1	.
.	.	1	.	.	.	.	.	−1	.	.
.	.	.	1	.	.	.	−1	.	.	.
.	.	.	.	1	.	−1	.	.	.	.
1	.	.	.	.	−2	.	.	.	.	1
3	1	−2	−1	−1	.	.	.	.	.	.

TABLE 17. GENOTYPES TRIGENIC AT ONE LOCUS, DIGENIC AT THE OTHER

Genotype *Ab/Ab/aB/a'b*, representing a set of 12

composition of gamete

mode of formation	AB/AB	Ab/Ab	aB/aB	ab/ab	a'B/a'B	a'b/a'b	AB/Ab	aB'/ab	a'B/a'b	AB/aB	Ab/ab	AB/a'B	Ab/a'b	aB/a'B	ab/a'b	AB/ab	Ab/aB	AB/a'b	Ab/a'B	aB/a'b	ab/a'B	divisor	deficiency matrix
1	.	2	.	.	.	.	.	.	.	.	.	.	4	.	.	.	4	.	.	2	.	12	1 1
2	.	6	3	.	.	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	12	. .
3	.	1	.	.	.	.	1	.	.	.	2	.	2	.	1	.	2	1	1	1	.	12	−2 −2
4	.	4	.	.	.	2	2	3	1	.	.	.	.	.	.	.	.	.	.	.	.	12	. .
5	.	2	.	.	.	.	.	.	.	2	2	.	4	1	1	.	.	.	.	.	.	12	. .
6	.	.	.	.	.	.	2	.	.	.	4	.	.	.	2	.	.	2	2	.	.	12	. 1
7	1	1	.	.	.	.	.	.	.	.	4	2	2	.	2	.	.	.	.	.	.	12	. .
8	.	2	.	3	.	1	4	.	2	.	.	.	.	.	.	.	.	.	.	.	.	12	. .
9	.	1	.	.	.	.	1	.	.	.	2	.	2	.	1	2	.	1	1	.	1	12	2 .
10	2	4	.	3	1	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	12	. .
11	.	2	.	.	.	.	.	.	.	.	.	.	4	.	.	4	.	.	.	.	2	12	−1 .

Genotype *AB/Ab/ab/a'b*, representing a set of 6

composition of gamete

mode of formation	AB/AB	Ab/Ab	aB/aB	ab/ab	a'B/a'B	a'b/a'b	AB/Ab	aB'/ab	a'B/a'b	AB/aB	Ab/ab	AB/a'B	Ab/a'b	aB/a'B	ab/a'b	AB/ab	Ab/aB	AB/a'b	Ab/a'B	aB/a'b	ab/a'B	divisor	deficiency matrix
1	.	.	.	.	.	.	4	.	.	.	4	.	4	.	4	4	.	4	.	.	.	24	1 1
2	6	6	.	6	.	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	. .
3	.	2	.	.	.	.	2	.	.	.	4	.	4	.	2	3	1	3	1	1	1	24	−2 −2
4	.	4	.	4	.	4	8	2	2	.	.	.	.	.	.	.	.	.	.	.	.	24	. .
5	2	2	.	.	.	.	.	.	.	2	6	2	6	.	4	.	.	.	.	.	.	24	. .
6	.	4	.	.	.	.	.	.	.	.	4	.	4	.	.	2	2	2	2	2	2	24	. 1
7	.	4	.	.	.	.	.	.	.	2	6	2	6	2	2	.	.	.	.	.	.	24	. .
8	.	8	.	2	.	2	4	4	4	.	.	.	.	.	.	.	.	.	.	.	.	24	. .
9	.	2	.	.	.	.	2	.	.	.	4	.	4	.	2	1	3	1	3	1	1	24	2 .
10	2	10	2	4	2	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	. .
11	.	.	.	.	.	.	4	.	.	.	4	.	4	.	4	.	4	.	4	.	.	24	−1 .

Genotype *AB/AB/ab/a'b*, representing a set of 6

composition of gamete

mode of formation	AB/AB	Ab/Ab	aB/aB	ab/ab	a'B/a'B	a'b/a'b	AB/Ab	aB'/ab	a'B/a'b	AB/aB	Ab/ab	AB/a'B	Ab/a'b	aB/a'B	ab/a'b	AB/ab	Ab/aB	AB/a'b	Ab/a'B	aB/a'b	ab/a'B	divisor	deficiency matrix
1	2	.	.	.	.	.	.	.	.	.	.	.	.	.	2	4	.	4	.	.	.	12	1 . 1
2	6	.	.	3	.	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	12	. 1 .
3	.	.	.	.	.	.	2	.	.	1	1	1	1	.	.	2	.	2	.	1	1	12	−2 . .
4	2	.	.	1	.	1	4	2	2	.	.	.	.	.	.	.	.	.	.	.	.	12	. −2 .
5	2	.	.	.	.	.	.	.	.	2	2	2	2	.	2	.	.	.	.	.	.	12	. . −2
6	.	2	.	.	.	.	.	.	.	.	.	.	.	2	.	2	2	2	2	.	.	12	. . −2
7	.	2	.	.	.	.	.	.	.	2	2	2	2	2	.	.	.	.	.	.	.	12	. . 2
8	.	2	1	.	1	.	4	2	2	.	.	.	.	.	.	.	.	.	.	.	.	12	. 2 .
9	.	.	.	.	.	.	2	.	.	1	1	1	1	.	.	.	2	.	2	1	1	12	2 . .
10	2	4	2	1	2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	12	. −1 .
11	2	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	4	.	4	.	.	12	−1 . 1

Genotype *AB/Ab/aB/a'b*, representing a set of 6

composition of gamete

mode of formation	AB/AB	Ab/Ab	aB/aB	ab/ab	a'B/a'B	a'b/a'b	AB/Ab	aB'/ab	a'B/a'b	AB/aB	Ab/ab	AB/a'B	Ab/a'b	aB/a'B	ab/a'b	AB/ab	Ab/aB	AB/a'b	Ab/a'B	aB/a'b	ab/a'B	divisor	deficiency matrix
1	.	.	.	.	.	.	4	.	.	4	.	.	4	.	.	.	4	4	.	4	.	24	1 . 1
2	6	6	6	.	.	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	. 1 .
3	1	1	.	.	.	.	2	.	.	1	1	1	1	1	1	2	4	4	2	2	.	24	−2 . .
4	2	2	2	.	.	2	8	4	4	.	.	.	.	.	.	.	.	.	.	.	.	24	. −2 .
5	2	2	.	.	.	.	.	.	.	6	2	2	6	2	2	.	.	.	.	.	.	24	. . −2
6	.	.	.	.	.	.	4	.	.	.	4	4	.	.	.	2	2	2	2	2	2	24	. . −2
7	2	2	.	.	.	.	.	.	.	2	6	6	2	2	2	.	.	.	.	.	.	24	. . 2
8	2	2	.	2	2	.	8	4	4	.	.	.	.	.	.	.	.	.	.	.	.	24	. 2 .
9	1	1	.	.	.	.	2	.	.	1	1	1	1	1	1	4	2	2	4	.	2	24	2 . .
10	6	6	2	4	4	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	. −1 .
11	.	.	.	.	.	.	4	.	.	4	.	.	4	.	.	4	.	.	4	.	4	24	−1 . 1

Haldane & de Winton (1931), is bound for this reason to leave the recombination fraction indeterminate.

The four conjugate pairs of sets of genotypes having three genes at one locus and two at the other (table 17), all have matrices deficient in the same element which is common to all genotypes digenic at both loci. Experimentation with these genotypes will not therefore supply the missing information. Equally, the gametic series for these sixty genotypes can be determined by experimentation with digenic organisms only.

The same is true of the two conjugate pairs of sets of genotypes having four genes at one locus and two at the other. In the genotype used for the first of these, the formula is invariant for all permutations of the three genes a_2, a_3 and a_4. Consequently, all gametes which can be generated from any one by such permutations must appear with equal frequency for all modes of gamete formation, and so can be accommodated in the same column of the matrix. The thirty-six different kinds of gamete then require only thirteen columns. In the genotype used for the second set the corresponding group of permutations is that generated from (a_1a_2), (a_3a_4) and $(a_1a_3)(a_2a_4)(Bb)$. Table 18 gives the numbers of distinguishable gametes represented in each column, and the total frequency of these for each mode of gamete formation.

TABLE 18. GENOTYPES TETRAGENIC AT ONE LOCUS, AND DIGENIC AT THE OTHER.

Genotype $a_1B/a_2b/a_3b/a_4b$, representing a set of 8

number of distinguishable gametes

mode of formation	gametic formula ...	1 $\frac{a_1B}{a_1b}$	1 $\frac{a_1b}{a_1b}$	3 $\frac{a_2B}{a_2b}$	3 $\frac{a_2b}{a_2b}$	1 $\frac{a_1B}{a_2b}$	3 $\frac{a_2B}{a_1b}$	3 $\frac{a_1B}{a_2B}$	3 $\frac{a_1b}{a_2b}$	3 $\frac{a_2B}{a_3B}$	3 $\frac{a_2b}{a_3b}$	3 $\frac{a_1B}{a_2b}$	3 $\frac{a_1b}{a_2B}$	6 $\frac{a_2B}{a_3b}$	divisor	deficiency matrix
1		.	.	.	.	.	.	.	.	.	2	2	.	.	4	1 .
2		1	.	.	3	.	.	.	.	.	.	.	.	.	4	. .
3		.	.	.	.	.	.	.	1	.	1	1	.	1	4	−2 .
4		.	.	.	2	1	1	.	.	.	.	.	.	.	4	. .
5		.	.	.	.	.	.	1	1	.	2	.	.	.	4	. .
6		.	.	.	.	.	.	.	2	.	.	.	.	2	4	1 1
7		.	.	.	.	.	.	.	2	1	1	.	.	.	4	. .
8		.	1	.	1	.	2	.	.	.	.	.	.	.	4	. .
9		.	.	.	.	.	.	.	1	.	1	.	1	1	4	. −2
10		.	1	1	2	.	.	.	.	.	.	.	.	.	4	. .
11		.	.	.	.	.	.	.	.	.	2	.	2	.	4	. 1

Genotype $a_1B/a_2B/a_3b/a_4l$, representing a set of 6

number of distinguishable gametes

mode of formation	gametic formula ...	4 $\frac{a_1B}{a_1b}$	4 $\frac{a_1b}{a_1b}$	4 $\frac{a_1B}{a_1b}$	2 $\frac{a_1B}{a_2B}$	8 $\frac{a_1B}{a_3B}$	2 $\frac{a_1b}{a_2b}$	4 $\frac{a_1B}{a_2b}$	4 $\frac{a_1B}{a_3b}$	4 $\frac{a_1b}{a_3B}$	divisor	deficiency matrix
1		.	.	.	1	.	.	.	2	.	3	1 . 1
2		3	.	.	.	.	.	.	.	.	3	. 1 .
3		.	.	.	.	1	.	1	1	.	3	−2 . .
4		1	.	2	.	.	.	.	.	.	3	. −2 .
5		.	.	.	1	2	.	.	.	.	3	. . −2
6		.	.	.	.	.	1	.	1	1	3	. . −2
7		.	.	.	.	2	1	.	.	.	3	. . 2
8		.	1	2	.	.	.	.	.	.	3	. 2 .
9		.	.	.	.	1	.	1	.	1	3	2 . .
10		1	2	.	.	.	.	.	.	.	3	. −1 .
11		.	.	.	1	.	.	.	.	2	3	−1 . 1

TABLE 19

Genotype $Ab/Ab'/aB/a'B$, representing a set of 9

number of distinguishable gametes

mode of formation	gametic formula ...	1 $\frac{AB}{AB}$	2 $\frac{Ab}{Ab}$	2 $\frac{aB}{aB}$	4 $\frac{ab}{ab}$	2 $\frac{AB}{Ab}$	1 $\frac{Ab}{Ab'}$	4 $\frac{aB}{ab}$	2 $\frac{ab}{ab'}$	2 $\frac{AB}{aB}$	1 $\frac{aB}{a'B}$	4 $\frac{Ab}{ab}$	2 $\frac{ab}{a'b}$	4 $\frac{AB}{ab}$	4 $\frac{Ab}{aB}$	4 $\frac{Ab}{ab'}$	4 $\frac{aB}{a'b}$	2 $\frac{ab}{a'b'}$	divisor	deficiency matrix
1		.	.	.	.	.	2	.	.	.	2	.	.	.	8	.	.	.	12	1
2		.	6	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	12	.
3		.	.	.	.	2	.	.	.	2	.	.	.	.	4	2	2	.	12	−2
4		.	.	2	.	4	2	4	.	.	.	.	.	.	.	.	.	.	12	.
5		.	2	.	.	.	.	.	.	4	2	4	.	.	.	.	.	.	12	.
6		2	.	.	.	.	.	.	.	.	.	.	.	4	4	.	.	2	12	.
7		2	.	.	.	.	.	.	.	4	.	4	2	.	.	.	.	.	12	.
8		2	.	.	.	4	.	4	2	.	.	.	.	.	.	.	.	.	12	.
9		.	.	.	.	2	.	.	.	2	.	.	.	4	.	2	2	.	12	2
10		4	2	2	4	.	.	.	.	.	.	.	.	.	.	.	.	.	12	.
11		.	.	.	.	.	2	.	.	.	2	.	.	8	.	.	.	.	12	−1

Genotype $AB/AB/ab/a'b'$, representing a set of 18

number of distinguishable gametes

mode of formation	gametic formula ...	1 $\frac{AB}{AB}$	2 $\frac{Ab}{Ab}$	2 $\frac{aB}{aB}$	2 $\frac{ab}{ab}$	2 $\frac{Ab'}{ab'}$	2 $\frac{Ab}{Ab'}$	1 $\frac{AB}{Ab}$	2 $\frac{aB}{ab}$	2 $\frac{aB}{ab'}$	2 $\frac{ab}{ab'}$	1 $\frac{aB}{a'B}$	2 $\frac{AB}{aB}$	2 $\frac{Ab}{ab}$	2 $\frac{Ab}{a'b}$	2 $\frac{ab}{a'b}$	1 $\frac{ab}{a'b'}$	1 $\frac{ab'}{a'b}$	2 $\frac{AB}{ab}$	2 $\frac{AB}{ab'}$	2 $\frac{Ab}{aB}$	2 $\frac{Ab}{ab'}$	2 $\frac{Ab}{a'B}$	2 $\frac{Ab}{a'b'}$	2 $\frac{aB}{a'b}$	2 $\frac{aB}{a'b'}$	divisor
1		1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	4	.	.	.	.	.	.	.	6
2		3	.	.	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6
3		.	.	.	.	.	.	1	.	.	.	.	1	.	.	.	.	.	1	1	.	.	.	1	.	1	6
4		1	.	.	.	.	.	2	2	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6
5		1	.	.	.	.	.	.	.	.	.	.	2	2	.	1	.	.	.	.	.	.	.	.	.	.	6
6		.	.	.	.	1	.	.	.	.	.	1	.	.	.	.	.	.	.	2	.	.	2	.	.	.	6
7		.	1	.	.	.	.	.	.	.	.	1	2	.	2	.	.	.	.	.	.	.	.	.	.	.	6
8		.	.	1	.	.	1	2	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6
9		.	.	.	.	.	.	1	.	.	.	.	1	.	.	.	.	.	.	1	1	1	.	1	.		6
10		1	2	2	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6
11		1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	4	.	.	.	.	.	6

Genotype $AB/Ab/aB/a'b'$, representing a set of 36

gametic formula

mode of formation	$\frac{AB}{AB}$	$\frac{Ab}{Ab}$	$\frac{Ab'}{Ab'}$	$\frac{aB}{aB}$	$\frac{a'B}{a'B}$	$\frac{ab}{ab}$	$\frac{a'b'}{a'b'}$	$\frac{ab'}{ab'}$	$\frac{a'b}{a'b}$	$\frac{Ab}{Ab'}$	$\frac{AB}{Ab}$	$\frac{AB}{Ab'}$	$\frac{aB}{ab}$	$\frac{a'B}{a'b'}$	$\frac{aB}{ab'}$	$\frac{a'B}{a'b}$	$\frac{ab}{ab'}$	$\frac{a'b}{a'b'}$
1	.	.	.	.	.	.	.	.	.	.	4	.	.	.	.	.	.	.
2	6	6	.	6	.	.	6	.	.	.	.	.	.	.	.	.	.	.
3	1	.	.	.	.	.	.	.	.	1	1	1	.	.	.	.	.	.
4	2	.	.	2	.	.	.	.	.	2	6	2	2	4	2	.	.	2
5	2	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
6	.	.	.	.	.	.	.	.	.	.	.	4	.	.	.	.	.	.
7	2	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
8	2	.	.	.	2	.	.	.	.	2	2	6	2	.	2	4	2	.
9	1	.	.	.	.	.	.	.	.	1	1	1	.	.	.	.	.	.
10	6	2	4	2	4	2	.	2	2	.	.	.	.	.	.	.	.	.
11	.	.	.	.	.	.	.	.	.	.	4	.	.	.	.	.	.	.

gametic formula

mode of formation	$\frac{aB}{a'B}$	$\frac{AB}{aB}$	$\frac{AB}{a'B}$	$\frac{Ab}{ab}$	$\frac{Ab'}{a'b'}$	$\frac{Ab}{a'b}$	$\frac{Ab'}{ab'}$	$\frac{ab}{a'b}$	$\frac{ab'}{a'b'}$	$\frac{ab}{a'b'}$	$\frac{ab'}{a'b}$	$\frac{AB}{ab}$	$\frac{AB}{a'b'}$	$\frac{AB}{ab'}$	$\frac{AB}{a'b}$	$\frac{Ab}{aB}$	$\frac{Ab'}{a'B}$	$\frac{Ab}{ab'}$	$\frac{Ab}{a'b}$	$\frac{Ab}{a'B}$	$\frac{Ab'}{aB}$	$\frac{Ab'}{a'b'}$	$\frac{Ab}{ab}$	$\frac{Ab'}{ab}$	$\frac{aB}{a'b}$	$\frac{a'B}{ab'}$	$\frac{aB}{a'b'}$	$\frac{a'B}{ab}$	divisor
1	.	4	.	.	.	.	.	.	.	.	.	.	4	.	.	4	.	.	.	.	.	.	4	.	.	.	4	.	24
2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24
3	1	1	1	.	.	.	.	.	.	1	.	1	3	1	1	2	.	1	.	2	2	1	.	1	.	1	.	24
4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24
5	2	6	2	2	4	2	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24
6	.	.	4	.	.	.	.	.	.	.	.	.	.	2	2	.	.	2	.	2	2	.	2	2	.	.	2	24
7	2	2	6	2	.	2	4	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24
8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24
9	1	1	1	.	.	.	.	.	.	.	1	2	.	2	2	1	3	.	1	1	1	.	1	.	1	.	1	24
10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24
11	.	4	.	.	.	.	.	.	.	.	.	4	.	.	.	.	4	.	4	.	.	.	.	.	4	.	.	24

When three genes are present at each locus, the only possible component of deficiency is that common to the bisimplex and the biduplex digenic types and, in fact, the smallest of the three sets trigenic at both loci in table 19 has a matrix deficient in this component. The two remaining sets, of eighteen and thirty-six genotypes respectively, are not in any way deficient, having gametic matrices of rank 11.

This is true also for the conjugate pair of sets of thirty-six genotypes having three genes at one locus and four at the other (table 20). There are in this case seventy-eight different kinds of gametes, but in respect of mode of formation these fall into only twenty-nine classes, ten homogenic at the locus with four genes, and nineteen heterogenic. The first class is produced only by modes of gamete formation 2, 4, 8, 10, and the second class by the remaining seven modes of gamete formation. The gametic matrix thus falls into two portions. These gametic matrices are written out (table 20) for the genotype $a_1B/a_2B/a_3b/a_4b'$; by permutation of the four a-genes and the three b-genes, the equivalent matrices may be found for each of the thirty-six genotypes of the set.

TABLE 20

Genotype $a_1B/a_2B/a_3b/a_4b'$, representing a set of 36

number of distinguishable gametes

mode of formation	gametic formula ...	2	2	4	2	2	4	2	2	2	2	divisor
		a_1B	a_3B	a_1b	a_3b	a_3b'	a_1B	a_1b	a_3B	a_3B	a_3b	
		a_1B	a_3B	a_1b	a_3b	a_3b'	a_1b	a_1b'	a_3b	a_3b'	a_3b'	
2		3	.	.	3	.	.	.	.	.	.	6
4		1	.	.	.	.	2	.	2	.	1	6
8		.	1	.	.	.	2	1	.	2	.	6
10		1	2	2	.	1	.	.	.	.	.	6

number of distinguishable gametes

mode of formation	gametic formula ...	1	4	1	2	4	4	2	4	4	4	4	4	2	2	2	4	4	1	1	divisor
		a_1B	a_1B	a_3B	a_1b	a_1b	a_1b	a_3b	a_1B	a_1B	a_3B	a_1B	a_4B	a_3B	a_4B	a_1b	a_1b	a_3b	a_3b	a_4b	
		a_2B	a_3B	a_4B	a_2b	a_3b	a_4b	a_4b	a_2b	a_3b	a_1b	a_4b	a_1b	a_4b	a_3b	a_2b'	a_3b'	a_1b'	a_4b'	a_3b'	
1		1	.	.	.	.	.	.	.	4	.	.	.	.	.	.	.	.	1	.	6
3		.	1	.	.	.	.	.	1	1	.	1	.	.	1	.	.	1	.	.	6
5		1	2	.	.	2	.	1	.	.	.	.	.	.	.	.	.	.	.	.	6
6		.	.	1	.	.	.	.	.	.	.	2	2	.	.	1	.	.	.	.	6
7		.	2	1	1	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	6
9		.	1	.	.	.	.	.	1	.	1	.	1	1	.	.	1	.	.	.	6
11		1	.	.	.	.	.	.	.	.	4	.	.	.	.	.	.	.	.	1	6

In the case of the set of twenty-four genotypes having four genes at each locus, all eleven modes of gamete formation are separate, so that the classification of the gametes supplies a direct enumeration of the frequency of the corresponding modes of formation. It is, however, exceptional for the experimenter to possess more than two allelomorphs of any given factor, so that the opportunity of making up genotypes of this set must be for a long while altogether unexpected.

6. THE ESTIMATION OF FREQUENCIES OF MODES OF GAMETE FORMATION

In disomic inheritance the backcross of a multiple heterozygote to a multiple recessive allows the frequencies of the 2^l gametic types to be estimated, subject to errors of random sampling, in the first generation. Further, the 2^{l-1} pairs of complementary gametic types

recognized in this way arise by the 2^{i-1} possible modes of gamete formation. With polysomic species it was seen that a second backcross generation suffices for the recognition of the different gametic types, and therefore for the estimation of their frequencies, but only if all the genes in the parent organism, at the loci tested, are distinct will it be possible to assign these gametic types each uniquely to its own particular mode of formation. This would require the use of four different allelomorphs at each locus with tetrasomic organisms, and of six with hexasomics.

A good many polysomic species are now known, and in these an increasing number of factors are available for experimentation. It is, however, exceptional to possess more than two genes at any one locus. In many cases perhaps no more than two genes exist. Information as to the frequency of modes of gamete formation will in such cases be indirect, and sometimes in some measure deficient. The statistical problem of utilizing such information deserves consideration.

The general nature of the statistical problem presented may be stated as follows: If

$$F = (f_1, f_2, f_3, \ldots)$$

be a row matrix representing the unknown frequencies of the different modes of gamete formation, having, for example, eleven components for tetrasomic organisms tested at two loci, then

$$f_1 + f_2 + f_3 + \ldots = 1.$$

If, further, $$g = \{g_{ik}\}$$

is the gametic matrix, specifying for each mode of formation, i, the frequency of the gametic type, k, so that

$$g_{i1} + g_{i2} + g_{i3} + \ldots = 1$$

for all values of i, then if N gametes have been tested for the genotype in question, it follows that

$$G = Ng,$$

with the understanding that when several genotypes have been used, with different numbers of gametes tested N, N', N'', ..., the matrix G will have columns, each with identification symbol k, corresponding with all combinations of parental genotypes and gametic types which need to be distinguished, and rows corresponding with the modes of gamete formation, so that the sum of the elements in each row is

$$N + N' + N'' + \ldots.$$

Then the matrix equation $$FG = M$$

defines a row matrix $$M = (m_k),$$

specifying the expected numbers of all relevantly distinguished types of gamete.

The data provided by any experimental investigation will consist in the observed frequencies a_k of the same types of gamete, and the principle of maximum likelihood requires that

$$\sum_k a_k \log m_k$$

should be maximized for variation of the components of F.

A general method for solving equations of estimation which are multiple and non-linear is to replace them by linear equations chosen so as to be consistent everywhere, and efficient in the neighbourhood of a trial solution. Successive approximations will then usually give adjustments of increasing relative precision as in Newton's method of approximation. The second approximation arrived at in this way will usually be fully efficient in the sense that the efficiency tends to 100 % as the sample tested is increased.

The case to be considered presents certain peculiarities which may be a source of confusion:

(i) the relative frequencies to be estimated are not independent,

(ii) the gametic matrix is usually such that an arbitrary frequency matrix, such as for tetrasomic organisms

$$\epsilon,\ 0,\ -2\epsilon,\ 0,\ 0,\ 0,\ 0,\ 0,\ 2\epsilon,\ 0,\ -\epsilon$$

may be added to F, without affecting any expectations.

For any method which may involve heavy computation it is most desirable that the different components should be treated symmetrically; for unknowns, however, connected by a linear equation, the information matrix will contain infinite elements. This difficulty may be overcome by introducing new symmetrical variables θ_i, defined by the relations

$$\frac{\partial}{\partial\theta_i} = f_i\left\{\frac{\partial}{\partial f_i} - \sum_i f_i \frac{\partial}{\partial f_i}\right\},$$

from which it can be seen that identically,

$$\sum_i \frac{\partial}{\partial\theta_i} = 0,$$

throughout the region $\Sigma f_i = 1.$

Any value θ may now be varied, keeping all others constant; the information matrix for θ will be finite, though degenerate, as is the covariance matrix for f, which as will be seen may be derived from it.

The simplest treatment of the second difficulty is to choose one of the frequencies involved, such as f_{11}, to be zero in the trial solution and to ignore this variable. Confirmation that the choice has been judicious is obtained if no other variate tends to zero on fitting to the data. The arbitrary frequency matrix may then be subtracted from the solution for positive values of ϵ, up to that value for which either f_1 or f_9 vanishes.

Now, in the present problem,

$$\frac{\partial}{\partial\theta_i}\log m_k = \frac{f_i}{m_k}\left\{\frac{\partial m_k}{\partial f_i} - \sum_i f_i \frac{\partial m_k}{\partial f_i}\right\},$$

but

$$\frac{\partial m_k}{\partial f_i} = G_{ik},$$

and

$$\sum_i f_i \frac{\partial m_k}{\partial f_i} = \sum_i f_i G_{ik} = m_k.$$

Hence the equations of maximum likelihood take the form

$$\sum_k a_k f_i \frac{G_{ik}}{m_k} - N f_i = 0,$$

and for any trial values the left-hand side represents the discrepancy from which a correction can be obtained.

Further, the typical term of the information matrix

$$\sum_k \frac{1}{m_k}\frac{\partial m_k}{\partial \theta_i}\frac{\partial m_k}{\partial \theta_j}$$

is equal to

$$f_i f_j \sum_k \frac{1}{m_k} G_{ik} G_{jk} - N f_i f_j,$$

and these elements may be found from a table of $f_i G_{ik}$.

If any corresponding row and column of this information matrix be omitted the remaining matrix is in general no longer degenerate and may be inverted to form the covariance matrix of the remaining parameters θ_i. Multiplying each column of the covariance matrix by the discrepancies corresponding with the rows, the adjustments $\Delta\theta_i$ are obtained for an improved solution. Using the relations

$$\Delta f_i = f_i\left(\Delta\theta_i - \sum_i f_i \Delta\theta_i\right)$$

with the adjustments $\Delta\theta_i$ so obtained, the corresponding adjustments Δf_i follow, adding of necessity to zero, for all values of i. Thus the values $\Delta\theta_i$ are each multiplied by f_i and added for all values for which $\Delta\theta_i$ has been obtained. This general value is subtracted from $\Delta\theta_i$, or from zero, and the difference multiplied by f_i. Exactly the same process applied in succession to the rows and columns of the covariance matrix for θ_i produces that for f_i, all rows and columns in this latter case necessarily add to zero.

The procedure may now be exemplified using the simple case provided by a tetrasomic plant having four genes at one locus and three at the other. Double reduction at the first locus is now recognizable, and occurs in only four modes of gamete formation, these being responsible for ten types of gamete.

TABLE 21. GAMETIC MATRIX FOR GAMETES SHOWING DOUBLE REDUCTION AT THE FIRST LOCUS

Genotype $a_1B/a_2B/a_3b/a_4b'$. Invariant permutation group (a_1a_2); (a_3a_4) (bb'); (a_1a_2) (a_3a_4) (bb')

		number of distinguishable gametes										
		2	2	4	2	2	4	2	2	2	2	
gametic formula ...		$\frac{a_1B}{a_1B}$	$\frac{a_3B}{a_3B}$	$\frac{a_1b}{a_1b}$	$\frac{a_3b}{a_3b}$	$\frac{a_3b'}{a_3b'}$	$\frac{a_1B}{a_1b}$	$\frac{a_1b}{a_1b'}$	$\frac{a_3B}{a_3b}$	$\frac{a_3B}{a_3b'}$	$\frac{a_3b}{a_3b'}$	divisor
mode of formation	2	3	.	.	3	.	.	.	.	.	.	6
	4	1	.	.	.	.	2	.	2	.	1	6
	8	.	1	.	.	.	2	1	.	2	.	6
	10	1	2	2	.	1	.	.	.	.	.	6

Suppose the frequencies observed in the ten classes showing double reduction at the a locus to be

$$18\ 9\ 6\ 16\ 1\ 7\ 2\ 3\ 2\ 2, \qquad A;\ S(a) = 66.$$

Let the four modes of gametic formation be (erroneously) guessed to have frequencies

$$4\ 2\ 2\ 3 \div 11, \qquad F_0.$$

Then the expected frequency for mode of formation i and gametic class k is $f_i G_{ik}$ as shown in table 22, the row of totals being the expectation vector **M** given by

$$m_k = \sum_i f_i G_{ik}.$$

TABLE 22. EXPECTATION FOR EACH MODE OF FORMATION

2	12	.	.	12	.	.	.	.	.	.	24
4	2	.	.	.	.	4	.	4	.	2	12
8	.	2	.	.	.	4	2	.	4	.	12
10	3	6	6	.	3	.	.	.	.	.	18
total	17	8	6	12	3	8	2	4	4	2	66

Summing the squares and products of the rows, divided by the column totals, one obtains the symmetrical 4×4 matrix:

20·47059	1·41176	.	2·11765	24
1·41176	8·23529	2·0	0·35294	12
.	2·0	8·5	1·5	12
2·11765	0·35294	1·5	14·02941	18

From each term is now deducted the product of the margins divided by 66, giving the simply degenerate matrix, which is the information matrix of the θ_i:

$$I_\theta \quad \begin{matrix} 11{\cdot}74332 & -2{\cdot}95188 & -4{\cdot}36364 & -4{\cdot}42780 \\ -2{\cdot}95188 & 6{\cdot}05348 & -0{\cdot}18182 & -2{\cdot}91978 \\ -4{\cdot}36364 & -0{\cdot}18182 & 6{\cdot}31819 & -1{\cdot}77273 \\ -4{\cdot}42780 & -2{\cdot}91978 & -1{\cdot}77273 & 9{\cdot}12031 \end{matrix}$$

Omitting any corresponding row and column, this may be read as the information matrix for the remaining adjustments $\Delta\theta_i$. Thus, omitting the fourth row and column and inverting the matrix, it is seen that V_θ is

$$V_\theta \quad \begin{matrix} 0{\cdot}1395922 & 0{\cdot}0710268 & 0{\cdot}0984529 \\ 0{\cdot}0710268 & 0{\cdot}2014768 & 0{\cdot}0548524 \\ 0{\cdot}0984529 & 0{\cdot}0548524 & 0{\cdot}2278479 \end{matrix}$$

the covariance matrix of these variables.

To find the discrepancies between the values proposed and those required by the observations, the expected frequencies are expressed as fractions of the totals m_k, as in the following table, and are then multiplied by a_k and summed for columns, with a marginal deduction of $66f_i$, as shown in the final column:

										f_i	discrepancy
0·7058823	.	.	1·0	.	.	.	.	.	.	0·3̇6̇	4·705882
0·1176471	.	.	.	.	0·5	.	1·0	.	1·0	0·1̇8̇	−1·382353
.	0·25	.	.	.	0·5	1·0	.	1·0	.	0·1̇8̇	−2·250000
0·1764706	0·75	1·0	.	1·0	.	.	.	.	.	0·2̇7̇	−1·073529
										1·00	0

The first three of these discrepancies, multiplied by the rows of the covariance matrix V_θ, give the required adjustments

$$\Delta\theta_0 \quad \begin{cases} \Delta\theta_2 & +0{\cdot}3372031 \\ \Delta\theta_4 & -0{\cdot}0676862 \\ \Delta\theta_8 & -0{\cdot}1251754 \\ \Delta\theta_{10} & 0 \end{cases}$$

whence may be calculated $\Delta f_i = f_i(\Delta\theta_i - \Sigma f_i \Delta\theta_i)$,

so that

$$\Delta F_0 \quad \begin{cases} \Sigma f_i \Delta\theta_i & 0{\cdot}0875529 \\ \Delta f_2 & 0{\cdot}0907812 \\ \Delta f_4 & -0{\cdot}0282253 \\ \Delta f_8 & -0{\cdot}0386779 \\ \Delta f_{10} & -0{\cdot}0238781 \end{cases}$$

adding necessarily to zero.

Since the corrections are somewhat large the process should be repeated, using the improved expectations

$$F_1 = F_0 + \Delta F_0.$$

Then the second expectations are

14·995	.	.	14·995	.	.	.	.	.	.	29·990
1·690	.	.	.	.	3·379	.	3·379	.	1·690	10·138
.	1·575	.	.	.	3·149	1·575	.	3·149	.	9·448
2·737	5·475	5·475	.	2·737	.	.	.	.	.	16·424
19·422	7·050	5·475	14·995	2·737	6·528	1·575	3·379	3·149	1·690	66·000

yielding the discrepancies

										f_i	discrepancy
0·7720626	.	.	1·0	.	.	.	.	.	.	0·4543939	−0·092873
0·0870147	.	.	.	.	0·5176164	.	1·0	.	1·0	0·1536061	0·051579
.	0·2234043	.	.	.	0·4823836	1·0	.	1·0	.	0·1431515	−0·060676
0·1409227	0·7765957	1·0	.	1·0	.	.	.	.	.	0·2488485	0·101970

of which the largest is little more than one-fiftieth of the largest of the first series. The approximation is now sufficiently good for the information matrix to indicate the true precision of each estimate. The new information matrix is

12·94480	−3·30186	−4·29311	−5·34983
−3·30186	5·40783	0·17870	−2·28467
−4·29311	0·17870	5·24239	−1·10798
−5·34983	−2·28467	−1·10798	8·76248

Omitting the first row and column and inverting, the covariance matrix of θ_2, θ_3 and θ_4 is found to be

$$V_\theta \quad \begin{matrix} 0{\cdot}2079206 & 0{\cdot}0047176 & 0{\cdot}0548178 \\ 0{\cdot}0047073 & 0{\cdot}1962931 & 0{\cdot}0264958 \\ 0{\cdot}0548178 & 0{\cdot}0264958 & 0{\cdot}1318265 \end{matrix}$$

and, by the transformation explained above, the covariance matrix for the estimated components of F

$$V_F \quad \begin{matrix} 0{\cdot}0048268 & -0{\cdot}0015967 & -0{\cdot}0007831 & -0{\cdot}0024470 \\ -0{\cdot}0015967 & 0{\cdot}0032748 & -0{\cdot}0011783 & -0{\cdot}0004998 \\ -0{\cdot}0007831 & -0{\cdot}0011783 & 0{\cdot}0030501 & -0{\cdot}0010887 \\ -0{\cdot}0024470 & -0{\cdot}0004998 & -0{\cdot}0010877 & 0{\cdot}0040355 \end{matrix}$$

Similarly, the new V_θ, applied to the discrepancies, yields the second adjustments ΔF_1

$$\Delta F_1 \quad \begin{matrix} -0{\cdot}0021935 \\ 0{\cdot}0017206 \\ -0{\cdot}0019745 \\ 0{\cdot}0024474 \end{matrix}$$

leading to the final estimates, to which may be appended standard errors from V_F:

$$F_2 \quad \begin{matrix} f_2 & 0{\cdot}4522 \pm 0{\cdot}0695 \\ f_4 & 0{\cdot}1553 \pm 0{\cdot}0572 \\ f_8 & 0{\cdot}1412 \pm 0{\cdot}0552 \\ f_{10} & 0{\cdot}2513 \pm 0{\cdot}0635 \end{matrix}$$

If, however, one is concerned to estimate the absolute frequencies of these four modes of gamete formation, and if N gametes in all had been tested, the estimates obtained above

would be multiplied by $66/N$. The covariance matrix V_F should then be multiplied by $(66/N)^2$, but before doing this a second component

$$\left(\frac{1}{66}-\frac{1}{N}\right)F^2$$

should be added to take account of the sampling variance of the number of gametes out of N, ascribed to these four modes of formation. More frequently, however, the precision of all modes of formation will be obtained in a single covariance matrix.

References

Fisher, R. A. & Mather, K. 1943 The inheritance of style length in *Lythrum salicaria*. *Ann. Eugen., Camb.*, **12**, 1–23.

Haldane, J. B. S. 1919 The combination of linkage values, and the calculation of distance between the loci of linked factors. *J. Genet.* **8**, 299–309.

Haldane, J. B. S. & de Winton, D. 1931 Linkage in the tetraploid *Primula sinensis*. *J. Genet.* **24**, 121–144.

Kosambi, D. D. 1944 The estimation of map distances from recombination values. *Ann. Eugen., Camb.*, **12**, 172–175.