

# THE NUMBER OF GENOTYPES WHICH CAN BE FORMED WITH A GIVEN NUMBER OF GENES

By J. B. S. HALDANE, *University College, London*

Since the time of Mendel it has been obvious that if we have a pair of allelomorphs at each of  $n$  unlinked loci in a diploid,  $3^n$  distinct genotypes are possible. For at one locus we can have  $AA$ ,  $Aa$ , or  $aa$ , at the next  $BB$ ,  $Bb$ , or  $bb$ , and so on. If, however, the two loci are linked, we have ten, not nine possible genotypes. For  $Aa Bb$  can exist in the two forms  $AB.ab$  and  $Ab.aB$ . While several more complicated cases than this have been considered, e.g. by Fisher (1947), the general case has not, I think, been discussed.

First consider a single locus with  $g$  allelomorphs,  $a^1$ ,  $a^2$ ,  $a^3$ , etc. In a diploid we can choose a like pair in  $g$  ways, an unlike pair in  $\frac{1}{2}g(g-1)$  ways, giving a total of  $\frac{1}{2}g(g+1)$ . Thus if there are  $g_1, g_2, g_3, \dots, g_n$  allelomorphs at the  $n$  loci, the total number of possible genotypes is the product of the various values of  $\frac{1}{2}g(g+1)$ , that is to say,

$$2^{-n} \Pi [g_r (g_r + 1)].$$

In an autotetraploid we can choose:

Four genes such as  $a^1 a^1 a^1 a^1$  in  $g$  ways

Four genes such as  $a^1 a^1 a^1 a^2$  in  $g(g-1)$  ways

Four genes such as  $a^1 a^1 a^2 a^2$  in  $\frac{1}{2}g(g-1)$  ways

Four genes such as  $a^1 a^1 a^2 a^3$  in  $\frac{1}{2}g(g-1)(g-2)$  ways

Four genes such as  $a^1 a^2 a^3 a^4$  in  $\frac{1}{24}g(g-1)(g-2)(g-3)$  ways.

Or altogether in  $\frac{1}{24}g(g+1)(g+2)(g+3)$  ways.

Thus in a tetraploid the number of possible genotypes with unlinked loci is

$$24^{-n} \Pi \{g_r (g_r + 1) (g_r + 2) (g_r + 3)\}.$$

In a  $m$ -ploid zygote a well-known result in the theory of combinations shows that at each locus there are  $\frac{(g+m-1)!}{(g-1)! m!}$  possible genotypes, and the number of genotypes which can be formed with  $n$  unlinked genes is the product of  $n$  such expressions.

Now let us consider  $n$  loci all in the same chromosome (or in the same linkage group if, as in *Oenothera*, several chromosomes are associated in one linkage group). Let the number of allelomorphs at the  $r$ th locus be  $g_r$ . Then the total number of types of chromosome which can be formed from them is  $C = g_1 g_2 g_3 \dots g_n = \Pi (g_r)$ .

It follows that the number of distinct diploid zygotic genotypes which can be made up is  $\frac{1}{2}C(C+1)$ , the number of distinct tetraploid genotypes  $\frac{C(C+1)(C+2)(C+3)}{4!}$ , and so on. The argument is exactly the same as in the case of the  $g$  genes. If we have a number of different chromosomes, the number of genotypes is the product of these expressions calculated for the various chromosomes, that is to say  $\Pi \left\{ \frac{(C_s + m - 1)!}{(C_s - 1)! m!} \right\}$  in an  $m$ -ploid, where  $C_s$  is the value of  $C$  for the  $s$ th chromosome.

Thus the general problem is solved. It seems, however, worth while to work out a few special cases.

Let  $D_2(n)$  be the number of diploid genotypes which can be formed from a pair of allelomorphs at each of  $n$  linked loci. This is the number of genotypes which could be produced from a single hermaphrodite individual apart from mutation, polyploidy, or aneuploidy. Owing to the rarity of multiple crossing over, they would not all be likely to appear in its immediate progeny even were this very large. Then  $D_2(n) = 2^{n-1}(2^n + 1)$ .

Let  $D_4(n)$  be the number of diploid genotypes which can be formed from four allelomorphs at each of  $n$  linked loci. This is the maximum number of genotypes which can occur in the descendants of a pair of individuals, if only these  $n$  loci are considered, apart from mutation, polyploidy or aneuploidy. For at each locus one parent can be  $a^1a^2$ , the other  $a^3a^4$ . It is to be noted that they cannot all occur in  $F_2$  in an organism such as *Drosophila* where crossing-over is limited to one sex. Nor in practice would they be obtained in the  $F_2$  of other organisms when  $n$  exceeds 3, owing to the rarity of multiple crossing-over. Then  $D_4(n) = 2^{n-1}(4^n + 1)$ .

Let  $T_2(n)$  be the total number of autotetraploid genotypes which can be formed from two allelomorphs at each of  $n$  linked loci. Then

$$T_2(n) = \frac{1}{3} \cdot 2^{n-3}(2^n + 1)(2^n + 2)(2^n + 3).$$

Let  $T_4(n)$  be the total number of autotetraploid genotypes which can be formed from four allelomorphs at each of  $n$  linked loci, and therefore the maximum possible in the progeny of a single individual heterozygous at  $n$  loci.

Then  $T_4(n) = \frac{1}{3} \cdot 2^{n-3}(4^n + 1)(4^n + 2)(4^n + 3)$ .

Let  $T_8(n)$  be the total number of autotetraploid genotypes which can be formed from eight allelomorphs at each of  $n$  loci. This is the total number possible in the descendants of a single pair, though since double reduction does not seem to occur with appreciable frequency near the centromere, they could not usually all be formed in  $F_2$  in practice, even if  $n$  were small and linkage loose:

$$T_8(n) = \frac{1}{3} \cdot 8^{n-1}(8^n + 1)(8^n + 2)(8^n + 3).$$

For comparison we give  $d_2(n) = 3^n$ , the number of genotypes possible in a diploid with two allelomorphs at  $n$  unlinked loci, and  $d_4(n) = 10^n$ ,  $t_2(n) = 5^n$ ,  $t_4(n) = 35^n$ ,  $t_8(n) = 330^n$ , the numbers corresponding to  $D_4$ ,  $T_2$  and  $T_8$  in the absence of linkage. All these are tabulated in Table 1.

Table 1

$n$	1	2	3	4	5	6
$d_2$	3	9	27	81	243	729
$D_2$	3	10	36	136	528	2,080
$d_4$	10	100	1,000	10,000	100,000	1,000,000
$D_4$	10	136	2,080	32,896	524,800	8,390,656
$t_2$	5	25	125	625	3,125	15,625
$T_2$	5	35	330	3,876	52,360	766,480
$t_4$	35	1,225	42,875	1,500,625	—	—
$T_4$	35	3,876	766,480	18,009,376	—	—
$t_8$	330	108,900	35,937,000	—	—	—
$T_8$	330	766,480	1,202,842,240	—	—	—

With the aid of this table we can generally calculate the number of genotypes in any desired case. Thus if we have three unlinked gene pairs, a group of three linked gene pairs, and another linked group of four, the number of possible zygotic genotypes is  $3^3 \times 36 \times 136$ , or 132,192 in a diploid, and  $5^3 \times 330 \times 3876$ , or 159,885,000 in an autotetraploid.

Where the numbers of allelomorphs considered at different loci are different, the formulae given above must be used. Thus Winge (1936) finds eleven loci with two allelomorphs and two with three allelomorphs in his first linkage group in *Pisum*.  $C = 2^{11} \cdot 3^2 = 181,432$ , and the total possible number of genotypes is  $\frac{1}{2}C(C+1) = 169,878,528$ , as compared with  $3^{11} \cdot 6^2$ , or 5,377,292 were these genes not linked.

Two special cases may be noted. The number of possible genotypes in a single X chromosome is simply  $C = \Pi(g_r)$ , which can be applied to the heterogametic sex of any animal. The maximum number of genotypes for a given chromosome possible in the  $F_2$  of such an animal as *Drosophila* with no crossing-over in the male is  $2^{n+1} - 1$  if the cross has been made between homozygotes differing in respect of  $n$  pairs of genes. For suppose the  $F_1$  to be  $\frac{A B C D \dots H}{a b c d \dots h}$ , then the female can form  $2^n$  types of gametes (supposing all multiple cross-overs to be possible) and the male only two types, namely  $A B C D \dots H$  and  $a b c d \dots h$ . From the union of these two types of spermatozoa with the  $2^n$  types of eggs we get  $2^{n+1}$  combinations, all different, except for the  $F_1$  type, which is formed in two different ways.

#### SUMMARY

Formulae and a table are given for the numbers of genotypes which can be obtained from a given assortment of genes where allowance is made for linkage.

#### REFERENCES

- FISHER, R. A. (1947). The theory of linkage in polysomic inheritance. *Philos. Trans. B*, **233**, 55.  
 WINGE, O. (1936). Linkage in *Pisum*. *C.R. Lab. Carlsberg, Sec. Physiol.* **21**, 15.