THE EVOLUTIONARY DYNAMICS OF A POLYMORPHISM IN THE HOUSE MOUSE¹

R. C. LEWONTIN AND L. C. DUNN

*Department of Biology, University of Rochester, Rochester, New York and
Department of Zoology, Columbia University, New York, New York*

Received November 30, 1959

OVER the past several years DUNN and his collaborators have shown that most wild populations of *Mus musculus* in the United States are polymorphic for variant alleles at the *T* (Brachy, short tail) locus (DUNN and SUCKLING 1956; DUNN 1957). A mutant *t* allele at this locus has been identified in each of 16 different populations, and only one population thus far tested has been shown with any degree of certainty to be free of such alleles.

This widespread polymorphism has several remarkable features which make it unlike other polymorphic systems so far studied. First, of 16 alleles from wild populations, 13 are unconditional prenatal lethals when homozygous (*t*^{w1}, 3, 4, 5, 6, 10, 11, 12, 13, 14, 15, 16, 17). The remaining three (*t*^{w2}, 7, 8) although viable when homozygous, are completely male sterile. Thus, even if it is supposed that the viability of homozygotes for these latter alleles is as high as for normal homozygotes, and that the fertility of homozygous females is normal, the *t*^w alleles are at least semi-lethal in a genetic sense. In the ordinary course of events such alleles should be reduced to an extremely low frequency in a population unless some other mechanism counterbalanced to maintain them.

The second unusual aspect of this polymorphism is the abnormal ratio of + and *t* gametes in the effective sperm pool of heterozygous males (+/*t*^w). About 95 percent of the offspring of +/*t*^w males carry the *t*^w allele while only five percent carry the + allele from the father. Determinations of this ratio can be made by utilizing the dominant allele *T*. Heterozygotes *T*/+ have short ("Brachy") tails, heterozygotes *T*/*t*^w are completely tailless, and heterozygotes *t*^w/+ have normal tails (CHESLEY and DUNN 1936). Test matings of two sorts can be made.

A.	♂ <i>T/t</i> ^w tailless	×	♀ +/+ normal
B.	♂ +/ <i>t</i> ^w normal	×	♀ <i>T</i> /+ Brachy

Normal Mendelian segregation would result in equal numbers of Brachy (*T*/+) and normal (*t*^w/+) offspring from type A matings. From type B matings normal segregation would produce equal numbers of +/+ (normal), +/*t*^w (normal), *T*/+ (Brachy) and *T/t*^w (tailless) offspring. The ratio of Brachy to tailless offspring reveals the ratio of + to *t* in the effective sperm pool. Such matings

¹ This investigation, including the cost of extra pages, was supported by a research grant, RG-6223, from the Division of Research Grants, Public Health Service and by Atomic Energy Commission Contract AT(30-1) 1804.

of males heterozygous for a t allele from a wild population always produce a large excess of offspring carrying t^w . As the results of both type A and type B matings have been in complete accord with each other when tested, most of the characterization of segregation ratios has been carried out with the more efficient type A mating system (DUNN 1957). Table 1 shows the pooled results of all type A tests so far run on alleles of wild origin. The great excess of t^w gametes in the effective pool of sperm from heterozygous males is striking. These results cannot be ascribed to intrauterine selection against Brachy embryos for two reasons. First the results of type B matings are the same, and second, the cross of T/t^w females by normal males produces Brachy and normal offspring in equal numbers (Table 1), DUNN 1957.

Thus, the mutant alleles, although lethal or semilethal in homozygotes, are present in great excess in the effective sperm of heterozygous males and in normal proportions in the ova of heterozygous females.

Finally, there is some evidence of a greater fitness of heterozygous males relative to homozygous normal males (DUNN, BEASLEY and TINKER 1958). This is as yet preliminary, involving only a single allele, t^{w11} , and based upon a small number of animals. Taken at face value, the data show the heterozygous males $+/t^{w11}$ to have a fitness of 1.76 times that of homozygous normal males. This estimate is based upon only 53 individuals, however, and should not be given too much weight. In the same experiment only 19 females could be successfully tested and these showed no evidence of differential fitness. Nevertheless, the suspicion of

TABLE 1

Pooled segregation ratios of tailless males, heterozygous for t alleles, each derived from a different wild population

1 Males tested*	Offspring classified at birth			
	2 Normal (+/ t^w)	3 Brachy +/ T	4 Total	5 Ratio of t^w
T/t^{w1}	325	38	363	.895
T/t^{w2}	325	17	342	.950
T/t^{w3}	355	3	358	.992
T/t^{w4}	187	5	192	.974
T/t^{w5}	479	30	509	.941
T/t^{w6}	230	4	234	.983
T/t^{w7}	214	16	230	.930
T/t^{w8}	178	25	203	.876
T/t^{w10}	240	7	247	.972
T/t^{w11}	518	19	537	.965
T/t^{w12}	412	19	431	.956
T/t^{w13}	502	11	513	.979
T/t^{w14}	392	21	413	.949
T/t^{w15}	415	41	456	.922
T/t^{w16}	107	6	113	.947
T/t^{w17}	336	1	337	.998
All males	5,215	263	5,478	.952
All females T/t^{w17}	199	188	387	.517

* Fives males of each allelic type were tested by mating with wild type $+/+$ females.

heterosis from the male data suggests that an intensive study of this aspect of the problem would be worthwhile.

Maintenance of the polymorphism

In the light of our present knowledge of the action of the t^w alleles it would seem that the polymorphism in nature is maintained by a balance between selective elimination of the mutant alleles in homozygous condition and the replenishment of these alleles by the extremely aberrant segregation ratio in heterozygous males. In addition, there is the possibility of heterosis as a balancing force.

There are two models that may be constructed for the evolution of the polymorphism, one deterministic and the other stochastic, i.e., involving chance processes. The deterministic model assumes that natural populations have very large breeding size, effectively infinite, and that the opposing forces of selection and abnormal segregation are entirely responsible for the state of the polymorphism in any population. The stochastic model, on the other hand, rejects the assumption of infinite population size and allows for the effect of random genetic drift and the probabilistic nature of selection. It is this latter model which is certainly closer to the real situation and, as we will presently show, more nearly explains what is observed in nature.

The general stochastic theory of population genetics is, as is well known, due to SEWALL WRIGHT. He has shown, in a series of papers too numerous to cite, that finite population size leads to fixation of alleles at rates which depend on population size, migration, mutation, selection, etc. The most general summary of this theory is to be found in WRIGHT (1949), and the results of the present paper are in accord with his general predictions.

The deterministic model

The first algebraic solution to the problem of the balance between selection and abnormal segregation ratio was given by PROUT (1943). This solution placed no restriction on the relative fitnesses of the genotypes but turns out to be incorrect when the segregation ratio is not equal in the two sexes. Because of this latter restriction, PROUT's solution is not applicable to the problem at hand.

BRUCK (1957) has given a series of expressions for the equilibrium gene frequency and frequency of heterozygotes for the case of a recessive lethal with abnormal segregation in one sex only. This solution is considerably closer to the situation in Mus and would be completely applicable except for the few nonlethal alleles and the fact that fitness values may be different in the two sexes. As a first approximation to the natural situation, however, BRUCK's model is quite satisfactory. Letting

m = proportion of t^w gametes in the effective sperm pool

$\hat{p}$ = equilibrium frequency of + alleles among adults

and

$\hat{h}$ = equilibrium frequency of heterozygotes among adults

then

$$(1) \quad \hat{p} = 1/2 - \frac{\sqrt{m(1-m)}}{2m}$$

and

$$(2) \quad \hat{h} = 2(1 - p) = 1 - \sqrt{\frac{1 - m}{m}}$$

Referring to Table 1, the lowest value of m is that for t^{ws} ($m = .876$), and the highest is for t^{w17} ($m = .998$). The average of all ratios is .952. Substituting these three values in equations (1) and (2) above we find

$$\begin{array}{ll} t^{ws}: \hat{p} = .788 & \hat{h} = .624 \\ t^{w17}: \hat{p} = .522 & \hat{h} = .956 \\ \text{average: } \hat{p} = .614 & \hat{h} = .772 \end{array}$$

Thus, between 60 and 95 percent of the adults in natural populations should be heterozygous for a mutant t allele. As we will discuss in a later section, this is much too high a frequency to correspond to what is actually found in nature. If, in addition, there is heterosis, the difficulty is exacerbated so that, in a sense, the deterministic solution proves too much.

The stochastic model

The deterministic model does not take into account the inbreeding which results from the restricted breeding size of *Mus* populations.

Although there are no direct observations from which the effective size of a breeding unit in this species can be estimated, the indications are that it is limited. The home ranges in a few populations studied appear to be small (review in BLAIR 1953); there are differences among local groups in the frequencies of different phenotypic characters (DUNN, BEASLEY and TINKER 1960; DEOL 1958); gene frequency observations show that different local groups are polymorphic for different alleles at one locus (t alleles) (DUNN 1957). The species population appears to consist of many partially separated breeding groups, each small in relation to the whole population. Although some local groups reach high densities and numbers (SOUTHWICK 1958), they probably pass through small population minima, and it is these which exercise a controlling effect on the evolution of gene frequencies. For these reasons, a useful approach in the construction of models is to test the effects on gene frequencies of small effective size of the breeding unit.

One method of dealing with the interaction of selection, segregation abnormality, and restricted population size is the construction of a stochastic model of population dynamics. Such a stochastic model may be analysed either by the algebraic theory of stochastic matrices (FELLER 1950) or by a *Monte Carlo* procedure. We have chosen the latter method as being the more efficient for our essentially exploratory study. In a Monte Carlo procedure, a mechanical or numerical analogue of a real population is formed and the rules by which the analogue population reproduces itself are made to conform with genetic rules of meiosis, fertilization, and selection. In addition chance plays a role in determining whether a given individual will survive and leave offspring. The mechanisms for Monte Carlo sampling may be beads mixed in an urn, numbers chosen from a table of random numbers, cards chosen from a shuffled pack, etc. The first sug-

gestion that such procedures might be useful for the t allele problem was made to one of us by PROF. GERT BONNIER in 1956. PROF. BONNIER saw clearly that finite population size must lead to eventual fixation of the normal alleles in contrast to the prediction of the deterministic model. He further suggested that "migrations of single but vigorous $+/t$ males" might be responsible for maintenance of the polymorphism. As will be shown later, we reach essentially the same conclusions.

As an analogue to a natural population we have used the IBM 650 digital computer. The scheme of the Monte Carlo program is shown in Figure 1. A two-digit number represents each individual. Because of the nature of the computer, it was convenient to let 88 stand for $+/+$, 89 for $+/t$, and 99 for t/t genotypes. $N\delta$ males and $N\varnothing$ females are initially loaded into the machine in any desired genotypic combination. A separate sperm and ovum pool are calculated and stored. As the number of gametes is very large compared with the number of adults, the gametic pools are computed deterministically, using predetermined values of $m\delta$ and $m\varnothing$, the segregation ratios for males and females. What is actually stored are the proportions $p\delta$ and $p\varnothing$ of two gametes in each pool. Next, a random number is generated by the method of BOFINGER and BOFINGER (1958). This number which lies between zero and one determines whether an eight or a nine gamete is chosen from the sperm pool. If the number is less than or equal

FIGURE 1.—Schematic representation of the Monte Carlo program for the 650 computer. Each stage is more fully explained in the text.

to $p\delta$ an eight is chosen, while if it is greater than $p\delta$ a nine is chosen. The same procedure is repeated for the pool of ova with the result that a zygote is formed which may be an 88, 89, or 99. The zygote so formed is classified as to its genotype in preparation for the selection phase. A new number is generated and if this number is less than or equal to W , the fitness of the particular genotype of the zygote just formed, the individual is saved and stored as an adult of the next generation. Otherwise, it is discarded. The processes of gamete choice, fertilization, and selection are repeated until $N\delta$ males have been saved and stored. Obviously this means that some larger number of males must be generated since selection discards a certain proportion of generated individuals. If necessary, the male fitness values are now replaced by female fitnesses and the entire process is repeated until $N\eta$ females have been saved, whereupon the saved males and females become the adults of the next generation, replacing the input males and females. The entire cycle from adults to adults is repeated until a predetermined number of generations has passed, or until the population is completely homozygous, whichever occurs first. During the computation the following information is printed each generation:

G : the generation number
 $D\delta$: number of $+/+$ adult male parents
 $H\delta$: number of $+/t$ adult male parents
 $R\delta$: number of t/t adult male parents
 $D\eta$: number of $+/+$ adult female parents
 $H\eta$: number of $+/t$ adult female parents
 $R\eta$: number of t/t adult female parents
 p : frequency of $+$ allele among parental adults
 $p\delta$: frequency of $+$ allele in sperm produced by parents
 $p\eta$: frequency of $+$ allele in ova produced by parents.

The parameters that are assigned before the program is run are:

$N\delta$: number of adult males maintained
 $N\eta$: number of adult females maintained
 $W1\delta$: fitness of $+/+$ males
 $W2\delta$: fitness of $+/t$ males
 $W3\delta$: fitness of t/t males
 $W1\eta$: fitness of $+/+$ females
 $W2\eta$: fitness of $+/t$ females
 $W3\eta$: fitness of t/t females
 $m\delta$: proportion of t gametes from a heterozygous male
 $m\eta$: proportion of t gametes from a heterozygous female
 $G_{\max}$: maximum number of generations for run.

In addition, the exact genotypic composition of the initial generation must be specified. About 1.5 seconds are required to compute each generation for the parameter sets used in this study.

The general approach was to make a large number of runs with identical parameter sets and input populations in order to get an idea of the distribution of results. No two runs are identical because of the random effects introduced. Thus, making 100 runs with the same set of parameters is equivalent to setting up 100

experimental populations all with the same forces of selection, random drift and segregation. It is then possible to deal with means and variances just as in experiments with real populations.

Results of the Monte Carlo trials

Table 2 shows the various sets of parameters used and the number of runs for each parameter set. Because of the lack of any decisive information to the contrary, the fitnesses of males and females were assumed to be the same; the homozygous normal and heterozygotes were assumed to have the same fitness, and the female segregation ratio was assumed to be normal. The data so far available validate the last assumption.

A. *The large populations:* Sets 1 and 2 were not run more than a few times because it became evident that they were behaving in a virtually deterministic manner. The effective population sizes of 50 and 20, respectively, for these two sets were too high for any appreciable random effect to be felt. Examination of the output from each generation showed that it was extremely unlikely that populations with these parameters would reach fixation before many hundreds of generations. The pertinent statistics for parameter sets 1 and 2 are given in Table 3. These are based upon all generations except the first two in each run. The first two generations are eliminated in each case to avoid the effect of the initially imposed frequencies. As the table shows, the observed means are very close to the theoretical value of $p = .615$ from BRUCK's equation. It should be noted, however, that the confidence intervals for p do not quite overlap the theoretical value. Both observed values are smaller than theoretical, the small population deviating more than the large. As will be discussed below, this effect is even stronger in the still smaller populations. As expected, the variance and range of values for set 1 is less than for set 2, but even in set 2 the closest approach to homozygosis was six $+/+$ males, four $+/t$ males, and ten $+/+$ females. Because

TABLE 2

Parameter sets for Monte Carlo trials

Set	$N\sigma$	$N\varphi$	$m\sigma$	G max.	Initial composition						Number of runs
					$+/+$	Male* $+/t$	t/t	$+/+$	Female† $+/t$	t/t	
1	25	25	.95	50	6	13	6	6	13	6	3
2	10	10	.95	200	2	6	2	2	6	2	3
3	4	4	.95	200	1	2	1	1	2	1	50
4	4	4	.90	200	1	2	1	1	2	1	50
5	2	6	.95	200	0	2	0	2	4	0	101
6	2	6	.95	100	1	1	0	6	0	0	119
7	2	6	.95	100	2	0	0	5	1	0	30
8	2	6	.98	200	0	2	0	2	4	0	56
9	2	6	.98	100	1	1	0	6	0	0	58
10	2	6	.90	200	0	2	0	2	4	0	56

* In all cases the fitness values for males and females are $W_1 = W_2 = 1$, $W_3 = 0$.

† Female segregation is normal ($m\varphi = .50$).

TABLE 3

Statistics calculated from parameter sets 1 and 2 of the Monte Carlo trials

Set	Total generations	$\bar{p}$	$\bar{q}$	σ_p^2	σ_p	σ_p^-	95 percent conf. int. for p
1	144	.6042	.3958	.00296	.0544	.00453	.5953-.6131
2	594	.5941	.4069	.00591	.0769	.00316	.5879-.6003

Set	Extremes						
	Female		Male		p	q	h
	+ / +	+ / t	+ / +	+ / t			
1	14	11	17	8	.81	.19	.38
	1	24	0	25	.51	.49	.98
2	6	4	10	0	.90	.10	.20
	0	10	0	10	.50	.50	1.00

of the segregation ratio in males, this high frequency of the + allele was immediately reduced in the next generation. It is this sort of result which led us to abandon these parameter sets and consider smaller population sizes.

B. *Small populations*: Because of the very large number of generations and runs, only a sample of the data has been analyzed for parameter sets 3-10. In all that follows, statistics are based on every tenth generation in each run.

For any given parameter set all populations (runs) were started with the same composition. After one generation, however, because of the random events in the model, all populations no longer have the same gene frequency but form a frequency distribution. A typical result is shown in Figure 2 for parameter set 5. The abscissa of each histogram is marked off in gene frequency, p , of the + allele. Since the total number of individuals is eight, there are only 16 genes in the population at the t locus so that p is always a multiple of $1/16$. The ordinate of each histogram shows the proportion of all populations (runs) having the given value of p . Results are plotted for generations 10, 20, 30, . . . 200. The trend in the observed frequencies (black bars) is obvious. As the number of generations increases, the proportion of populations reaching fixation of the + allele ($p = 1.0$) increases so that by generation 200, 84 percent of all populations are fixed. In the unfixed populations the most frequent class is $p = .50$ and the distribution is J-shaped. This J-shaped distribution is fairly stable in form over the entire set of 200 generations, the only change being a gradual decrease in the total frequency of unfixed populations. This stability is shown by the clear bars which represent the distribution of unfixed classes readjusted to give a total frequency of unity. That is, the clear bars show the proportion of *as yet unfixed* populations which fall in each class. The existence of such a stable distribution of unfixed classes is in accordance with the prediction of WRIGHT (1937).

Rates of fixation

The proportion of populations fixed for the normal allele in successive generations is given in Tables 4 and 5. Table 4 shows the proportion of previously unfixed populations which become newly fixed in each ten-generation period.

FIGURE 2.—Frequency distributions of gene frequencies in successive generations. The ordinate shows the proportion of all populations (runs) having a given gene frequency. The abscissa shows the values of p , the frequency of the normal allele. Black bars are the distributions of all runs. Clear bars are the conditional distributions for unfixed classes. Based on the data from parameter set 5.

They are then the *fixation rates* per ten generations. In Table 5 the cumulative proportions of fixed populations are given, and this latter information is plotted in Figure 3.

A number of interesting comparisons can be made by holding all parameters constant except one. The effect of a change in population size can be seen by comparing set 3 with set 5, and set 4 with set 10. Sets 3 and 5 both have a segregation ratio of .95 but set 3 with four males and four females has a higher effective population size ($N = 8$) than does set 5 ($N = 6$). The effective population size is calculated from the relation

$$N = \frac{4 N \delta N \varphi}{N \delta + N \varphi}$$

Similarly, sets 4 and 10 both have $m = .90$ but again have population sizes 8 and 6 respectively. As shown by the means at the bottom of Table 4, the fixation rates of the smaller populations are between one and a half and two times as large as for the larger populations. These mean fixation rates are roughly the equilibrium rates since generations 10 and 20 are not included in their calculation. The result of the difference in fixation rates between large and small populations is shown in Table 5 and Figure 3 as a difference in the proportion of populations fixed after n generations. In the case of set 10, for example, all populations are fixed by

TABLE 4
Fixation rates per ten generations

Generation	Set						
	3	4	5	6	8	9	10
10	.000	.040	.020	.218	.018	.207	.054
20	.000	.083	.051	.140	.036	.065	.189
30	.060	.113	.117	.125	.075	.070	.209
40	.064	.128	.108	.086	.020	.000	.294
50	.023	.059	.054	.125	.042	.075	.291
60	.070	.031	.101	.107	.022	.027	.059
70	.000	.129	.097	.060	.022	.083	.313
80	.075	.074	.107	.064	.045	.000	.181
90	.000	.120	.100	.068	.000	.000	.333
100	.054	.136	.022	.073	.071	.061	.000
110	.029	.000	.067	...	.026	...	.000
120	.059	.034	.095	...	.000	...	.333
130	.031	.111	.000	...	.000	...	.500
140	.064	.250	.184	...	.000	...	.000
150	.069	.000	.065	...	.026	...	.000
160	.037	.167	.103	...	.054	...	.500
170	.077	.000	.038	...	.029	...	.000
180	.083	.000	.080	...	.000	...	1.000
190	.091	.100	.130	...	.058	...	...
200	.250	.222	.150	...	.031	...	...
Weighted mean	.055	.103	.090	.093	.030	.041	.234

TABLE 5
Cumulative fixation rates. Proportion of all runs of each set which are fixed by a given generation

Generation	Set						
	3	4	5	6	8	9	10
10	.00	.04	.02	.22	.02	.21	.05
20	.00	.12	.07	.33	.05	.26	.23
30	.06	.22	.18	.41	.12	.31	.39
40	.12	.32	.27	.45	.14	.31	.57
50	.14	.36	.31	.53	.18	.36	.70
60	.20	.38	.38	.58	.20	.38	.71
70	.20	.46	.44	.61	.21	.43	.80
80	.26	.50	.50	.63	.25	.43	.84
90	.26	.56	.54	.66	.25	.43	.89
100	.30	.62	.55	.68	.30	.47	.89
110	.32	.62	.58	..	.32	..	.89
120	.36	.64	.62	..	.32	..	.93
130	.38	.68	.62	..	.32	..	.96
140	.42	.76	.69	..	.32	..	.96
150	.46	.76	.71	..	.34	..	.96
160	.48	.80	.74	..	.37	..	.98
170	.52	.80	.75	..	.39	..	.98
180	.56	.80	.77	..	.39	..	1.00
190	.60	.82	.80	..	.43	..	1.00
200	.68	.86	.83	..	.45	..	1.00

FIGURE 3.—Proportions of populations fixed for the normal allele in successive generations. Triangles: set 3; circles: set 4; squares: set 5; crosses: set 6; vertical bars: set 8; inverted triangles: set 9; oblique circles: set 10.

generation 180, while only 80 percent of the populations in set 4 are fixed at that time.

The effect of changes in segregation ratio can be estimated by comparing set 3 with set 4, and set 5 with set 8 and with set 10. In each case the lower the segregation ratio the greater the fixation rate. Sets 10, 5 and 8 have segregation ratios respectively of .90, .95 and .98. Their mean fixation rates are .234, .090 and .030. One way of comparing fixation rates is from the "half-life" of each set, that is, the number of generations required for 50 percent of the populations to be fixed. For sets 10, 5 and 8, these half-lives are approximately 36, 80 and 220 generations, the last figure coming from an extrapolation.

Finally, the role of different initial compositions can be studied. Sets 5 and 6 are identical except that each population in set 6 began with all wild type individuals with the exception of a single heterozygous male. Table 4 shows that the *mean equilibrium* fixation rates are virtually identical for these two sets (.090 and .093 respectively). The very large difference in total fixation between the sets as given in Table 5 and Figure 4 is entirely due to the high fixation rate in the first few generations. If the single mutant allele from the heterozygous male succeeds in "infecting" the population, there will no longer be any effect of its originally low frequency. The same phenomenon can be seen in the comparison between sets 8 and 9. The probability of a successful infection of a pure normal population with a *t* allele is fairly high in both cases. For set 6 ($m = .95$) the introduced *t* allele lasts more than 20 generations in 67 percent of the trials. The comparable figure for set 9 ($m = .98$) is 74 percent. As is to be expected, the higher the segregation ratio the greater the probability of successful infection.

In one set, 7, ($m = .95$) which is not shown in any detail, a single heterozygous

FIGURE 4.—Stable distributions of unfixed populations for various parameter sets.

female was introduced into an otherwise normal population. In 30 runs, only two populations remained unfixed for more than 20 generations (31 and 98 generations to fixation), two were fixed between the tenth and twentieth generations, and the remaining 26 populations were fixed before the seventh generation. Thus, the lack of an abnormal segregation ratio in the female makes it extremely unlikely that a *t* allele brought in by a female will survive. The difference between the case of migrant females and migrant males lies essentially in the difference between the initial frequencies of the *t* allele in the gametic pool. Clearly, it is not the initial adult composition that matters but the proportions of *T* and *t* gametes which they produce.

The probability of successful introduction of a mutant allele into an otherwise normal population is an important consideration in the maintenance of the polymorphism, and we shall return to this point in discussing the role of breeding structure and migration.

The stable distribution of gene frequencies

In addition to comparing fixation rates from one parameter set to another, we may also examine the effects of parameter changes on the distribution of unfixed populations. Figure 4 shows the stable distributions of sets 3, 4, 5, 6, 8, 9 and 10. These are constructed by pooling the distributions of generations 30–200 (30–100

for sets 6 and 8) to give an average distribution of unfixed classes. The first two periods are not used so that the initial population composition will not affect the results.

The effect of population size is shown by comparison of the distributions for set 3 with set 5, and set 4 with set 10. In both cases the smaller populations show a higher frequency in the extreme left-hand class, $p = .50$, and consequently lower frequencies in the other classes. Table 6 contains the probabilities associated with χ^2 tests for the differences between the distributions of different parameter sets. As these probabilities show, the effects of changing the population size are highly significant. The piling up of populations in the left-hand class is a result of the asymmetry of natural selection. A smaller population size causes greater genetic drift both toward lower and higher frequencies of the + allele with a consequent depletion of intermediate classes. Since t/t homozygotes are lethal, however, the frequency of the + allele can never fall below 0.50. The populations that would ordinarily go to fixation at $p = 0$ are then piled up at the barrier imposed by selection at $p = .50$.

Another view of the same effect can be obtained from the means and variances of the distributions as listed in Table 7. The results of t tests for differences between these means are given in Table 6. In discussing the large populations it was noted that the mean values of p were significantly lower than the theoretical

TABLE 6

Probabilities resulting from tests for the difference between parameter sets that differ in one parameter

Set	3	4	5	6	8	9	10
3		<.001	<.001	...	...	...	...
4	<.001		...	...	...	...	.001
5	<.001	...		.84	<.001	...	<.001
6	...	...	.60		...	...	...
8	...	...	<.001	...		.27	<.001
9	...	...	...	...	.12		...
10	...	.80	<.001	...	<.001	...	

Above the diagonal: χ^2 tests for the difference between the stable distributions of Figure 4.
Below the diagonal: t tests for the difference between means of Table 7.

TABLE 7

Means and variances of the stable distributions of unfixed classes of Figure 4

Set	$\bar{p}$	$\bar{q}$	σ_p	σ_q	σ_p^-	95 percent conf. int. for p	$P_{\text{theor.}}$
3	.5935	.4065	.0130	.1139	.0047	.5843-.6027	.6147
4	.6388	.3612	.0117	.1079	.0058	.6275-.6501	.6667
5	.5656	.4344	.0101	.1003	.0036	.5586-.5726	.6147
6	.5692	.4308	.0109	.1043	.0052	.5591-.5793	.6147
8	.5290	.4710	.0056	.0746	.0028	.5235-.5345	.5714
9	.5380	.4620	.0067	.0818	.0049	.5285-.5475	.5714
10	.6356	.3644	.0186	.1363	.0115	.6131-.6581	.6667

values derived from BRUCK's equation. This is even more true of the small populations, and those with effective size 6 deviate more from the theoretical values than those with effective size 8, although the difference between sets 4 and 10 is clearly not significant. This lowering of the mean of unfixed classes is a result of the greater genetic drift with a piling up of populations at the lower limit of $p = .50$.

The consequence of a change in segregation ratio is reflected in the difference among sets 10, 5 and 8. As Figure 4 shows, an increase in m from .90 to .95 to .98 causes a piling up of the distribution at the left end. These differences are significant (Table 6) and are perfectly paralleled by a drop in the mean value of p from .6356 to .5656 to .5290 respectively. The concentration of the unfixed classes around $p = .50$ for high segregation ratios is also in accord with the slower rate of fixation observed for these parameter sets.

Finally, the lack of any detectable effect of initial composition on the equilibrium rate of fixation is paralleled by a lack of difference in the stable distributions of unfixed populations. The distribution for set 5 is virtually identical with that for set 6, and an almost equal degree of similarity exists between the distributions for sets 8 and 9. The χ^2 tests for differences between 5 and 6 and between 8 and 9 are nonsignificant as contrasted with the other comparisons in Table 6. The similarity between the distributions for sets 5 and 6 and for sets 8 and 9 is also shown by their means in Table 7.

Comparisons with populations in nature

The observations on the geographic distribution of t alleles were obtained before the models were derived and were primarily designed to detect and accumulate for genetic study a variety of t alleles from different sources. Gene frequencies were thus obtained only as by-products of this general sampling process, and in nearly all cases are based on very small samples. While they are adequate to give some clues to the state of this locus in wild populations, they cannot serve as strict tests of the applicability of the models to conditions in nature.

Mice were obtained from traps set in 31 localities distributed over the United States with one site sampled in Nova Scotia; and in addition samples were obtained from three enclosed populations descended from wild mice captured in three additional localities. The methods of trapping varied. Usually one or a few traps were set to test whether *Mus* was present. At first this was done in buildings; later eight feral populations were sampled. Since the object was mainly to capture mice from the locality, it is quite possible that mice caught at different times in the locality came from different breeding units or families, and there is some suggestion in the testing records that this occurred. It is evident therefore that mice from different trapping stations in the same locality may or may not represent samples from the same breeding unit. Although the mice in the localities of capture have been referred to as "populations," this has only a general geographic, not a strict biological meaning.

All animals were tested individually by matings to *Brachys* ($T/+$) at Nevis

Biological Station. At least one heterozygote ($+/t^w$) was found in the sample from each of 14 localities and from each of three enclosed populations. Of 66 mice tested from 14 localities, 23 were $+/t^w$. Sample sizes per locality ranged from one to ten. It is obvious that because of small and variable sample sizes and the method of trapping, no reliable gene frequency can be based on the above. It does show, however, that heterozygotes for these alleles are rather common (minimum frequency about 35 percent) in the populations in which they occur.

Better estimates can be obtained from extensive tests of enclosed populations bred in captivity. One of these has been maintained for 13 years at the Rockefeller Institute for Medical Research in New York, having originated from a few wild mice captures in New York City and Philadelphia in 1946. The breeding colony is kept at 125–250 individuals in the winter, reduced to about a dozen pairs each summer, so considerable inbreeding has occurred. Four samples from this population kindly supplied by DR. HOWARD SCHNEIDER, have been tested (Jan. '52, June '52, June '55, June '59). Heterozygote frequencies have remained at 50–60 percent. Up to now, of 197 animals tested 110 (or 56 percent) have been $+/t^w$.

One population in Rumford, Virginia, which was sampled by A. B. BEASLEY yielded in 1955 one $+/t^w$ out of eight tested. In the spring of 1959 an attempt was made by JOHN BEASLEY to trap all the animals in this barn. Of 41 animals captured 14 have completed tests. Eight of these (57 percent) are $+/t^w$. A similar attempt was made to capture all mice in a feed room of an isolated barn near Storrs, Connecticut. Of seven which bred in captivity, five were $+/t^w$. These experiences suggest that where t alleles occur, heterozygotes reach frequencies of 50–60 percent; the scattered small samples yield estimates below this value.

Testing of successive samples from Clinton, Montana, supplied by Miss VIRGINIA VINCENT suggest a possible reason for such discrepancies. A sample of 11 taken from this station (in a small barn) in September 1957 all proved to be $+/+$; a sample of three taken in October 1958 all proved to be $+/t^w$. Some of the latter may have come from the nearby house. The combined estimate is 3/14, but this may include family groups with high and with low frequencies.

Samples from 17 localities failed to yield a $+/t^w$ animal. The sample sizes tested were in general small, in 13 cases five or less. From only one locality (Great Gull Island in Long Island Sound) were sufficient mice tested (40) to give a high probability that t alleles were not present.

The general picture that emerges from this survey is that local groups of *Mus* free from t alleles are probably rare in the continental area. The only one established as t free is on a well isolated coastal island. Where t alleles occur they seldom or never reach the frequencies predicted by the deterministic model based on the average male segregation ratios actually found.

An effect of applying the stochastic model to small breeding groups is to suggest a reason for the failure of the deterministic model to fit the actual results. The equilibrium predicted by the latter model is not attained because of the chance loss of such alleles in small breeding groups. The general trend in the small groups is clearly toward fixation of wild type alleles. This process is, to be sure, retarded

by very high values of the segregation ratio, but the direction is clear; an appreciable portion become fixed with the passage of time.

This process would produce a dichotomy amongst family groups in the same locality, some fixed, others unfixed, and the average gene frequencies for the locality would be determined largely by the proportions between such groups in the locality.

The sampling methods used heretofore are inadequate to test whether something like the above situation obtains in nature, but at least we know now what kind of observations must be made. Family and household structure must be known, and methods designed for the capture of households or extended families.

In the few cases in which collections of families have been studied (Rockefeller and Virginia samples) the average gene frequencies are not far from those predicted on the stochastic model and far below those predicted on the deterministic model. This cannot be taken as validation of the model until other variations in parameters have been shown to correspond to those assumed in the model.

A new and useful view of the methods by which *t* alleles may spread to new populations is also given by the analog calculations. It appears that alleles must travel in males in order to have a good chance of persisting in the newly "infected" population. This gives new impetus to studies of relative mobilities of the two sexes as factors in gene dispersal. This may have special importance in this case, since an implication of the stochastic model is that *t* lethals are at best "genes of passage" in small breeding groups because of the probability of chance loss. Those which are found, therefore, may have been acquired rather recently from migrants. This is more probable than acquisition by mutation since, as yet, we know of no instance of mutation of the wild type allele to *t*.

A final point which emerges from the stochastic model concerns selection among the various *t* alleles. Table 1 shows that all the *t* alleles found in natural populations exhibit very high segregation ratio abnormalities. This is not true, however, of all "mutant" alleles at this locus. In addition to the alleles of wild origin there are 19 alleles at this locus which have arisen in the laboratory. These arose from balanced lethal lines of the type *T/tⁿ*, in the form of single normal-tailed individuals which later proved to be *tⁿ/t^x*. When the segregation ratios of the newly arisen laboratory "mutants" are compared with alleles from wild populations, a radical difference, shown in Table 8, is found. Newly arisen alleles have segregation ratios ranging from very high to very low with a distinct mode at .50. Clearly the wild alleles are a selected sample of all the possible "mutations"

TABLE 8

Distribution of male segregation ratios for t alleles extracted from wild populations and for t alleles arisen in the laboratory by mutation from other t alleles

	.99 .90	.89 .80	.79 .70	.69 .60	.59 .55	.5 .50	.49 .45	.44 .40	.39 .30	.29 .20	<i>n</i>
Wild alleles	15	1	0	0	0	0	0	0	0	0	16
Newly arisen alleles	3	1	0	0	2	8	2	1	2	0	19

* Ratios within $\pm 2 \sigma$ of .5, i.e., normal ratios.

at this locus, selected for high segregation ratio in favor of the mutant. The deterministic model would not predict such selection since any mutant allele with a segregation ratio above .50 would be maintained in a stable equilibrium. The stochastic model, however, does predict the selection of mutants with higher segregation ratios. The Monte Carlo results showed the rate of loss of *t* alleles due to random drift to be very sensitive to the value of *m*. Alleles with relatively small segregation advantages would be lost quite rapidly with the result that the only alleles still seen in natural populations are those with high values of *m*. As time goes on this concentration of alleles at the high end of the range should become even sharper. These considerations once more emphasize that the polymorphism for *t* alleles is a transient one and very probably subject to violent oscillations. New alleles arise occasionally from old ones by a process which is not understood but which may involve recombination. Such new alleles as have high segregation ratios have a high probability of spreading in the population and of "infecting" other populations by the migration of heterozygous males. The alleles then spread in the species but are constantly being lost by drift and revived by migration. Eventually they must be lost entirely only to be replaced by yet another rare mutant of high segregation advantage.

SUMMARY

Wild populations of *Mus musculus* are polymorphic for mutant alleles at the *t* locus. When homozygous, these alleles are usually unconditional prenatal lethals, but a few are viable and male-sterile. Heterozygous males have an average of 95 percent *t*-bearing sperm in their effective sperm pool. This abnormal segregation ratio opposes the loss of the *t* alleles due to their lethal effect. Deterministic models for the evaluation of this polymorphic system predict (1) that any allele with a segregation ratio above .50 should be maintained in a stable equilibrium, and (2) that alleles with segregation ratios of the order of 95 percent should be in such high frequency as to make between 60 and 95 percent of the population heterozygous. Neither of these predictions is in accord with the observations from nature since: (1) only alleles with very high segregation ratio are found in contrast to the range known in laboratory mutants and (2) populations contain about 35 percent to 50 percent heterozygotes.

A stochastic model has been constructed, and results of this model were obtained by Monte Carlo procedures on a digital computer. These results show that for populations composed of small family groups there will be a loss of the *t* allele due to random genetic drift. The rate of loss is sensitive to differences in segregation ratio so that only alleles with a high ratio will remain in the population for appreciable periods. A geographical population will be composed mainly of family groups fixed for the wild type allele and of family groups with a high frequency of the mutant *t* allele. A normal population is easily "infected" with a mutant allele if this allele is carried into the population by a heterozygous male. These results are all in good accord with observations from nature insofar as the sampling methods thus far used permit a comparison.

LITERATURE CITED

- BLAIR, W. F., 1953 Population dynamics of rodents and other small animals. *Advances in Genet.* **5**: 1-41.
- BOFINGER, E., and V. J. BOFINGER, 1958 On a periodic property of pseudorandom sequences. *J. Assoc. for Computing Machinery* **5**: 261-265.
- BRUCK, D., 1957 Male segregation ratio advantage as a factor in maintaining lethal alleles in wild populations of house mice. *Proc. Natl. Acad. Sci. U.S.* **43**: 152-158.
- CHESLEY, P., and L. C. DUNN, 1936 The inheritance of taillessness (anury) in the house mouse. *Genetics* **21**: 525-536.
- DEOL, M. S., 1958 Genetical studies on the skeleton of the mouse. XXIV. Further data on skeletal variation in wild populations. *J. Embryol. and Exp. Morphol.* **6**: 569-574.
- DUNN, L. C., A. B. BEASLEY, and H. TINKER, 1958 Relative fitness of wild house mice heterozygous for a lethal allele. *Am. Naturalist* **92**: 215-220.
- 1960 Polymorphism in wild populations of *Mus musculus*. *J. Mammal.* (in press).
- DUNN, L. C., 1957 Studies of genetic variability in populations of wild house mice. II. Analysis of eight additional alleles at locus *T*. *Genetics* **42**: 299-311.
- DUNN, L. C., and J. SUCKLING, 1956 Studies of genetic variability in populations of wild house mice. I. Analysis of seven alleles at locus *T*. *Genetics* **41**: 344-352.
- FELLER, W., 1950 *An Introduction to Probability Theory and Its Applications*. John Wiley and Sons. New York.
- PROUT, T., 1953 Some effects of the variations in segregation ratio and of selection on the frequency of alleles under random mating. *Acta Genet. et Statist. Med.* **4**: 148-151.
- SOUTHWICK, C. H., 1958 Population characteristics of house mice living in English corn ricks: density relationships. *Proc. Zool. Soc. London* **131** (Pt. 2): 163-175.
- WRIGHT, S., 1937 The distribution of gene frequencies in populations. *Proc. Natl. Acad. Sci. U.S.* **23**: 307-320.
- 1949 Adaptation and selection. *Genetics, Paleontology and Evolution*. Edited by G. L. JEPSON, G. G. SIMPSON, and E. MAYR. Princeton University Press. Princeton, N.J.