A SIMULATION STUDY OF QUASI-FIXATION OF GENES DUE TO RANDOM FLUCTUATION OF SELECTION INTENSITIES

J. S. F. BARKER AND J. C. BUTCHER¹

*Departments of Animal Husbandry and Applied Mathematics,
University of Sydney, Sydney, Australia*

Received September 20, 1965

KIMURA (1954) investigated the process of change in the gene frequency distribution due to random fluctuation of selection intensities. Assuming a pair of alleles lacking dominance, and the selection coefficient (s) fluctuating from generation to generation around a mean value of zero, where s equals the selection coefficient against the heterozygote (i.e. heterozygote fitness = $1 - s$, disadvantageous homozygote fitness = $1 - 2s$), he showed that the gene frequency distribution will accumulate toward the terminal values of 0 or 1 with increasing time. However, the probability at these terminal values was always zero. That is, given sufficient generations, almost all populations will have the gene either almost fixed in the population or almost lost from it. To describe this phenomenon, KIMURA proposed the terms quasi-fixation and quasi-loss. KIMURA (1955) pointed out that only for $\bar{s} = 0$ could the equations be solved in closed form, but he discussed briefly other cases. In particular, he asked if $\bar{s}$ and V_s (the variance of the selection coefficient) are of the same order of magnitude, or $\bar{s}$ (> 0) is much smaller than V_s , is quasi-fixation of the disadvantageous allele possible? He argued that if the initial gene frequency of the advantageous allele is very low, this latter situation might allow some probability of quasi-fixation of the disadvantageous allele.

To answer these questions more definitively, a Monte Carlo type programme was developed for the University of Sydney automatic computer (SILLIAC), and results obtained using this programme are presented in this paper.

However, KIMURA (1962) has extended his earlier theoretical investigations in developing a more general formula for the probability of eventual fixation, which allows for random drift because of small population number, and also for that due to random fluctuations in selection intensity. Further, he could then solve for cases other than $\bar{s} = 0$, so that he was then able to answer his question relating to the probability of quasi-fixation of the disadvantageous allele. Results presented here are of interest in confirming KIMURA's theoretical analysis, and in providing information on the time required for the quasi-fixation process.

¹ Present address: Stanford Linear Accelerator Center, Stanford, California, U.S.A.

Operations of the Programme

Given the initial gene frequency (x_0) in each of 1000 populations (assumed infinitely large), the programme takes these through any specified number of generations, operating on each in each generation with a selection coefficient sampled uniformly in the interval $(\bar{s} - (3V_s)^{1/2}, \bar{s} + (3V_s)^{1/2})$, so that the mean and variance of the selection coefficient will agree with the given $\bar{s}$ and V_s . At any specified generation, the gene-frequency distribution can be printed out as the number of populations whose gene frequency falls within each of 190 specified gene-frequency ranges. As we are more interested in the distributions towards the termini, these ranges are not equally spaced, but become progressively narrower from the centre of the distribution towards each terminus. Thus the ranges are of width 10^{-1} from 0.1 to 0.9, 10^{-2} from 0.01 to 0.1 and from 0.9 to 0.99, 10^{-3} from 0.001 to 0.01 and from 0.99 to 0.999, and so on to a width of 10^{-11} ; with two terminal classes 0 to 10^{-11} and $1 - 10^{-11}$ to 1.

Together with this distribution, the programme output gives the numbers of populations with gene frequency greater than x_0 , and less than x_0 , and the mean gene frequencies of these, as well as the overall mean gene frequency.

Test of the Programme

To test the programme, parameters the same as those investigated by KIMURA (1954) were used, *viz.*: $x_0 = 0.5$, $\bar{s} = 0$, $V_s = 0.0483$, with the distribution printed out for generations 10, 28, 40, and 100. The numbers in each frequency range were tested against those expected from solution of Equation (5) of KIMURA (1954), using the chi-squared test. The distributions for each of the four generations were not significantly different.

If for two computer runs, the ratio $\bar{s} : V_s$ were constant, but both were some fraction (α) in the second run of their value in the first, one would expect (KIMURA 1955) that the gene-frequency distribution after t generations in the first would be the same as in the second after t/α generations. The programme was tested to see that it satisfied this condition as follows:

Run 1: $x_0 = 0.5$, $\bar{s} = 0.05$, $V_s = 0.3125$, $t = 1, 2, 3, 4, 5$, and 10 generations;

Run 2: $x_0 = 0.5$, $\bar{s} = 0.025$, $V_s = 0.15625$, $t = 2, 4, 6, 8, 10$, and 20 generations.

The distributions for the consecutive pairs of generations were compared using the chi-squared test, and were not significantly different, except for the first pair, where $\chi^2_{(7)} = 36.84$, $P < 0.001$. In this first pair, although the distributions in the two runs were in fact different, both correspond to the same Fokker-Planck equation, which is an approximation to the genetic model in which we are interested. The agreement as the number of generations increased is apparently due to the fact that the approximations are better in these cases. As we are interested in the form of the distributions after many generations, this difference was not considered important.

RESULTS AND DISCUSSION

KIMURA (1962) presents an equation for the probability of quasi-fixation of the desirable allele, in terms of its initial frequency (x_0), and a parameter k ($= 2\bar{s}/V_s$). These probabilities are given in Table 1 for a range of values of x_0 and k . For a particular value of k (< 1.0), the probability of quasi-fixation of the desirable allele increases as its initial frequency increases. For a particular value of x_0 and constant $\bar{s}$, the probability of quasi-fixation increases as k increases (i.e. as V_s decreases).

We have studied in detail the case where $x_0 = 0.9$. That is, the desirable allele is already at a high frequency in the population, so that one is concerned with the probability of quasi-fixation of the undesirable allele, or alternatively, the probability of quasi-loss of the desirable allele. A constant value of $\bar{s}$ (0.01) has been

TABLE 1

Probabilities of fixation of the desirable allele for various values of x_0 (its initial frequency) and k ($2s/V_s$) (Solution of equation 19 of Kimura 1962)

k	x_0				
	0.01	0.1	0.5	0.9	0.99
0.1	.01731	.13692	.56908	.92858	.99435
0.2	.02935	.18364	.63674	.95025	.99689
0.3	.04876	.24158	.70163	.96632	.99834
0.4	.07947	.31200	.76251	.97795	.99914
0.5	.12711	.39582	.81831	.98615	.99957
0.6	.19957	.49345	.86818	.99177	.99980
0.7	.30753	.60458	.91151	.99548	.99991
0.8	.46491	.72793	.94794	.99783	.99997
0.9	.68902	.86102	.97739	.99923	.99999
1.0	1.00000	1.00000	1.00000	1.00000	1.00000

used in a series of "experiments," k being varied by varying V_s (values given in Table 2). Each experiment was continued at least until the gene frequency in all populations was either less than 0.001 or greater than 0.999. Two experiments were run for a further 200 generations, but as no population with frequency less than 0.001 increased above this, these values were taken as convenient end points.

The numbers of populations showing such quasi-loss, together with the numbers theoretically expected, are given in Table 2. Generally, the observed number was greater than that expected. The distributions of gene frequencies were output every 20 generations to generation 420, then for generation 450, and thereafter, every 25 generations, so that changes in the number tending to quasi-loss can be followed (Table 3). It is difficult to account for the observed quasi-loss being higher than that expected. Over all generations, it is unlikely that the actual mean selection coefficient would be less than that specified. However, if

TABLE 2

Expected probabilities of loss of the desirable allele (from Table 1), and expected and observed numbers of populations (of 1000 in each experiment) showing such loss, where $x_0 = 0.9$ and $\bar{s} = 0.01$

Experiment number	$k(2s/V_s)$	V_s	Probability of loss	Expected number showing loss	Observed number showing loss
0	0.1	0.2	0.07142	71.42	79
1	0.2	0.1	0.04975	49.75	93
2	0.3	0.06667	0.03368	33.68	64
3	0.4	0.05	0.02205	22.05	45
4	0.5	0.04	0.01385	13.85	11
5	0.6	0.03333	0.00823	8.23	14
6	0.7	0.02857	0.00452	4.52	18
7	0.8	0.025	0.00217	2.17	6
8	0.9	0.02222	0.00077	0.77	3
9	1.0	0.02	0	0	3

TABLE 3

Number of populations with gene frequency less than initial frequency ($x_0 = 0.9$), and final number tending to quasi-loss, with generation after which this number remains stable

Generation	Experiment number								
	0	1	2	3	4	5	6	7	8
20	222	267	288	279	299	299	323	297	283
40	138	186	198	202	226	188	235	227	223
60	111	171	155	172	159	149	173	195	166
80	93	138	125	121	124	128	146	139	146
100	88	155	107	106	87	98	142	101	120
120	80	130	94	99	74	79	133	106	115
140	80	121	88	85	61	72	110	96	111
160	79	109	89	75	55	72	105	87	97
180	.	104	80	76	49	68	92	75	77
200	.	108	78	69	42	65	91	65	68
220	.	104	77	67	40	55	96	55	58
240	.	101	71	66	38	46	78	47	49
260	.	99	72	63	34	42	69	42	46
280	.	99	70	64	29	38	63	39	47
300	.	95	71	63	25	36	59	33	42
320	.	94	69	60	20	32	57	32	27
									(325)
340	.	94	69	58	21	29	52	28	26
									(350)
360	.	94	69	56	18	29	46	25	21
									(375)
380	.	93	67	53	19	28	45	23	19
400	.	.	67	53	17	26	39	22	19
650	.	.	64	.	.	.	.	.	.
825	.	.	.	45	.	.	.	.	.
900	.	.	.	.	.	.	.	6	.
1025	.	.	.	.	11	.	.	.	.
1275	.	.	.	.	.	.	18	.	.
1400	.	.	.	.	.	.	.	.	3
1800	.	.	.	.	.	.	.	3	.
2500	.	.	.	.	.	14	.	.	.

over a small number of generations a greater than expected number of frequencies drift to less than 0.5, they will tend to continue towards zero. That this may partly account for the observed excess of quasi-loss is shown in Figure 1, which gives the numbers of frequencies less than x_0 for two experiments where observed and expected were similar (Nos. 0 and 4), and two where they were more discrepant (Nos. 1 and 6). It will be seen that for the former there was generally a steady decline to the final stable value. However, for the latter there were periods where the number remained fairly steady or actually increased (*e.g.* Experiment 1—generations 80 to 100, Experiment 6—generations 200–220). A further computer experiment was then run, using the same parameters as for Experiment 1, but a different sequence of random numbers. In this run, 58 populations showed quasi-loss; still greater than expected, but not as discrepant as Ex-

FIGURE 1.—Changes in observed numbers of frequencies less than x_0 in Experiments 0, 1, 4, and 6.

periment 1. If x_0 were less than 0.5, the probability of fixation of the undesirable allele then might be less than that predicted from KIMURA's equation. One experiment was done, using the same parameters as for Experiment 1, but with $x_0 = 0.1$. The expected number of populations showing fixation of the undersirable allele (from Table 1) is then 816, while the number observed was 815. In itself, this is hardly critical evidence, but it seems likely that the discrepancies between observed and expected probabilities of quasi-loss are the result of stochastic effects in the populations when the gene frequency is around 0.5.

The number of generations to final stability of quasi-loss is given in Table 3. As might be expected, this number of generations generally increased as V_s decreased (i.e., as k increased). For Experiments 1 to 4 inclusive, the relationship between number of generations and k is linear, but the number of generations to stability in the remaining experiments deviate markedly from this linear relationship. Some of the exceptions to the trend, such as Experiment 5, are of interest. Here, the number of frequencies less than x_0 was 17 in Generation 925, three of which were between 0.01 and 0.9. These three eventually tended to quasi-fixation, one becoming greater than 0.9 by Generation 1250, one by Generation 1800, and the last by Generation 2500. Thus, with randomly fluctuating selection intensity, the gene frequency may drift over intermediate values for long periods of time, producing an apparent, but unstable polymorphism.

Taking both the expectation of quasi-loss and the time to reach stability as V_s increases, the rate of decline in the number of frequencies less than x_0 increases, but the final probability of quasi-loss also increases, giving a cross-over pattern as shown in Figure 2. This relationship may be clarified by considering the probability of the desirable allele being within $1/(2N)$ of fixation at large popula-

FIGURE 2.—Expected pattern of changes in the proportion of populations with frequencies less than x_0 for varying values of V_s .

tion sizes. Since random sampling of gametes is important near the terminal classes even for large population sizes, this does not necessarily reflect the probability of fixation, but is a reasonable definition of quasi-fixation. Thus, in the computer simulation experiments, gene frequencies greater than 0.9999 and 0.99999 respectively imply quasi-fixation for population sizes of 5,000 and 50,000. These probabilities of quasi-fixation can be obtained from the gene-frequency distribution output, and results for Experiments 1, 3, and 7 are given in Figure 3. Progress towards quasi-fixation was more rapid, but its ultimate probability was less, as V_s increased.

Changes in the mean gene frequency of the desirable allele for those populations with frequency less than x_0 , and for all populations, are shown in Figures

FIGURE 3.—Changes in the probability of quasi-fixation with time, assuming each population to be of size 5,000 or 50,000, for Experiments 1, 3, and 7.

FIGURE 4.—Changes, with time, in the mean gene frequency of the desirable allele for those populations with frequency less than x_0 .

4 and 5. In Figure 4, a difference in the trend of change to quasi-loss is apparent between the experiments where V_s was highest (0, 1, 2, and 3), and the remaining experiments, whereas one might expect something resembling a family of

FIGURE 5.—Changes, with time, in the mean gene frequency of the desirable allele for all populations. The smooth continuous curve is the expected change in gene frequency for a constant selection coefficient of 0.01. Designation of experiments—as in Figure 4.

negative exponential curves. No explanation for this apparent difference has been found, but if it is real, and not simply a product of stochastic variation in the present series of experiments, it would repay further theoretical investigations. A further difference between these two sets of experiments is seen in Figure 5. Whereas mean gene frequency in Experiments 4 to 9 shows a relatively steady increase from x_0 to the final value, in Experiments 0, 1 and 2 it tends to increase to a maximum, and then decrease to the final value. Experiment 3 is intermediate, in that it increases to somewhat less than the final value and then decreases before steadily increasing to the final value. Clearly, the difference in trend is mainly due to the rate of movement of the distribution towards the termini (Figure 4), but there is again the apparent break at a value of k of about 0.4.

This study was supported by a University of Sydney Research Grant to J.S.F.B.

SUMMARY

A Monte Carlo computer programme has been developed to investigate quasi-fixation of genes due to random fluctuation of selection intensities. Starting with the gene frequency of the desirable allele = 0.9, and a constant mean selection coefficient = 0.01, ten simulation experiments have been done with the variance of the selection coefficient (V_s) ranging from 0.02 to 0.2. In terms of the probability of quasi-loss of the desirable allele, the results confirm the theoretical expectations of KIMURA (1962), although the observed probabilities were generally somewhat higher than the expected. The number of generations to final stability of quasi-loss tended to increase as k ($= 2s/V_s$) increased, and could be expected to be at least 1000 for $0.5 \leq k < 1.0$. A difference in the trend of change to quasi-loss is apparent between those experiments where $k < 0.5$, and those where $k \geq 0.5$. If real, this would repay further theoretical investigation.

LITERATURE CITED

- KIMURA, M., 1954 Process leading to quasi-fixation of genes in natural populations due to random fluctuation in selection intensities. *Genetics* **39**: 280-295. — 1955 Stochastic processes and distribution of gene frequencies under natural selection. Cold Spring Harbor Symp. Quant. Biol. **20**: 33-53. — 1962 On the probability of fixation of mutant genes in a population. *Genetics* **47**: 713-719.