Fixation probabilities of mutant genes with artificial selection

William G. HILL

Institute of Animal Genetics, University of Edinburgh,
West Mains Road, Edinburgh EH9 3JN, U.K.

Summary

Fixation probabilities (u) of mutant genes, which are initially in single copy and have additive effects on a quantitative trait under truncation selection, are computed using Monte-Carlo simulation. A range of gene effects relative to the phenotypic standard deviation, heritabilities and numbers of parents and progeny are studied. The diffusion approximation is found to be an excellent predictor of u. Selection on individual performance, on within family deviation, on family mean and an index of individual and family mean performance are compared. Is is found that, particularly for genes of large effect, u is reduced if much weight is given to family mean.

Key words : Artificial selection, mutation, fixation, index.

Résumé

Probabilité de fixation de gènes mutants en sélection artificielle

Les probabilités de fixation (u) de gènes mutants présents originellement en une seule copie et ayant des effets additifs sur un caractère quantitatif soumis à une sélection par troncature ont été calculées grâce à une simulation de Monte-Carlo. Les valeurs de l'effet des gènes par rapport à l'écart-type phénotypique, celles de l'héritabilité et des nombres de parents et de descendants ont été paramétrées. La solution de l'équation de diffusion s'avère une excellente approximation de u. Diverses méthodes de sélection ont été comparées telles que des sélections sur performance individuelle, écart intra-famille, moyenne de famille ainsi qu'une sélection par indice combinant la valeur individuelle et la moyenne de la famille. Il apparaît, en particulier pour des gènes à effets importants, que la probabilité de fixation est réduite si l'on accorde une forte pondération à la moyenne des familles.

Mots clés : Sélection artificielle, mutation, fixation, indice.

I. Introduction

In previous analyses of the effects of mutations on long term response to artificial selection (HILL, 1982) fixation probabilities of the mutant gene have been calculated using the diffusion approximation (KIMURA, 1957). The approximation is likely to hold

best when population sizes are large, although numerical analyses by several workers have shown that it does well even for quite small populations (EWENS, 1979). For use in artificial selection the selective value (s) for additive genes has been computed as $s = i\alpha$, where i is the selection intensity and α the effect of the gene on the trait, measured as the difference between the homozygote in phenotypic standard deviation units. This relationship only holds closely when gene effects are small (LATTER, 1965). The diffusion approximation for fixation probability (u), is given by setting $s = i\alpha$ in the following equation for additive genes,

$$u = [1 - \exp(-2N_e sq)] / [1 - \exp(-2N_e s)], \tag{1}$$

where N_e is the effective population size and q the initial gene frequency. Numerical analysis has shown that (1) holds quite well, even for large α, but the approximation was tested only for genes at intermediate frequencies, $0.25 < q < 0.75$ (HILL, 1969).

A mutant gene is initially present only in single copy, so the previous numerical results do not necessarily apply. Even though tests on the diffusion approximation for mutants in single copy have been made, these have used the Wright-Fisher model of binomial distribution of frequencies. In the case of artificial selection, particularly when selection is intense and the mutant has substantial effect, its fate must usually be decided during the first selection process after it appears : either the individual carrying it is selected to become a parent and thus have the chance to have many progeny in the next generation, in which case the gene is likely to remain in the population, or the gene is immediately lost. It was therefore considered necessary to check the use of (1) for the case of the mutant gene with artificial selection. Monte-Carlo simulation was used, rather than the exact analysis of HILL (1969), so that two sexes and mating structures could be incorporated, complexities beyond the computational feasibility of the exact method.

In artificial selection programmes selection an index (I) of individual and family performance is often practised so as to increase the accuracy of selection, r_{IA}, the correlation of breeding value (A) and the index (LUSH, 1947). The selection limit using existing variation is a function of $N_e i r_{IA}$ assuming (1) applies (ROBERTSON, 1960), so maximising the initial rate of response by maximising i does not necessarily lead to the greatest selection limit because, for example, selection on family mean reduces effective population size. Also, as DEMPFLÉ (1975) pointed out, selection within families can be more efficient in the long term than predicted from these calculations, particularly with high heritabilities, because selection reduces the variation among families. For mutant genes this relationship of fixation probability to $N_e i r_{IA}$ might not be expected to apply exactly because of the critical nature of the first selection : the chance of a mutant of large effect surviving this process will be greater the more emphasis given to individual phenotype. Thus the influence of a mutant gene when it appears is likely to be reduced by emphasising family mean performance. In this article we shall therefore investigate the effects of different kinds of index on fixation probability.

The results also apply where a single individual is introduced into a population carrying a gene previously absent from it, provided this gene's fixation probability is not influenced by its association (linkage disequilibrium) with other genes carried by the individual.

II. Simulation procedure

Each generation, the N_m selected males were each mated to N_f/N_m (integral) females. If family (sibship) sizes were random, a total of T_m and T_f male and female progeny were sampled with equal probability from each full-sib family, to give a multinomial distribution of family size. If family sizes were set to be equal, then in each family T_m/N_f and T_f/N_f male and female progeny were sampled, again to make T_m and T_f males and females in all.

Variability was of two kinds. The first was due to the additive effects of all other genes, apart from the mutant, and environmental effects. Each was normally distributed and the simulation was conducted such that the within-family genetic component of variance was constant, and that between-families depended on the selection (BULMER, 1970). With a heritability of 0.4, for example, the effects of selection were such that slightly more than a fraction 0.2 of the total variance was genetic within full sib families and the same amount less than 0.2 was genetic between families. The second source of variation, additive to the first, was due to the mutant gene, which was assumed to be additive with effect α phenotypic standard deviations difference between the homozygotes. The mutant was randomly assigned to one individual and subsequently truncation selection continued until it was fixed or lost. All genes were assumed to be unlinked.

For mass selection individuals were ranked on their own performance (X) and the best N_m males and N_f females selected. For other schemes, means $(\overline{X})$ of the performance of the $(T_m + T_f)/N_f$ individuals in each full sib family were computed and the best N_m and N_f selected over all families on the basis of family mean $(\overline{X})$, within family deviation $(X - \overline{X})$ or index $(X + \overline{X})$. For within family selection (W), the best male and female in each full sib family were selected.

Effective population sizes were computed by assigning a gene with no effect on the trait an initial frequency of 0.5 and estimating the rate of decline in heterozygosity. No mutant gene affecting the trait was included in the runs which were used to check the effect of selection on effective population size (ROBERTSON, 1961).

III. Results and discussion

Fixation probabilities computed by Monte-Carlo simulation for mass selection are given in Table 1. For comparison, values computed from (1), by substituting $q = 1/(2T)$ and $s = \bar{i}\alpha$, namely

$$u = [1 - \exp(-N_e\bar{i}\alpha/T)] / [1 - \exp(-2N_e\bar{i}\alpha)] \qquad (2)$$

where $T = T_m + T_f$ and $\bar{i}$ is the mean selection intensity for males and females computed for selecting N_m/T_m males and N_f/T_f females (FALCONER, 1981), and $N_e = 4N_mN_f/(N_m + N_f)$. The agreement between simulated and diffusion results is very good indeed over the whole range of values of gene effect and population size — further evidence of the remarkable power of the diffusion approximation, here applied to a very special process. The fixation probability is lower for a heritability of 0.4 than of 0.0. This is presumably associated with, but as results discussed later show not

precisely described by, the reduction in effective population size. The reduction in effective size is rather less than predicted by ROBERTSON (1961), as JONES (1969) in experiments with *Drosophila* and L. DEMPFLÉ (personal communication) using simulation have previously found. Providing $N_e \bar{i}\alpha$ exceeds 1.0 eq. (2) reduces to

$$u = \bar{i}\alpha N_e/T \tag{3}$$

approximately, which is seen to agree well with the simulation results except when $\bar{i}\alpha$ (the selective value) is very large (> 1.0).

TABLE 1

Fixation probability[+] and effective population size for mass selection with random family size.
S/.0 and S/.4 denote simulation with heritability .0 and .4, D denotes the diffusion approximation.

Probabilité de fixation et taille efficace en sélection individuelle
avec une taille efficace de population aléatoire.
S/.0 et S/.4 signifient : « simulation avec héritabilité .0 et .4 », D « approximation par diffusion ».

N_m, N_f	5, 5			5, 5			10, 10		
T_m, T_f ($\bar{i}$)	10, 10 (0.74)			20, 20 (1.21)			20, 20 (0.77)		
	S/.0	S/.4	D	S/.0	S/.4	D	S/.0	S/.4	D
α	Fixation probability								
0.125	.050	.038	.054	.050	.030	.039	.022	.038	.048
0.25	.096	.096	.091	.068	.056	.073	.080	.056	.091
0.5	.164	.148	.169	.126	.110	.140	.160	.162	.175
1.0	.280	.266	.309	.276	.252	.261	.322	.266	.320
2.0	.510	.462	.523	.526	.492	.453	.524	.522	.537
	Effective population size								
	11.1	10.3	10.0	10.7	9.5	10.0	22.4	20.2	20.0

N_m, N_f	5, 10			10, 20			5, 20		
T_m, T_f ($\bar{i}$)	20, 20 (0.99)			40, 40 (1.01)			40, 40 (1.19)		
	S/.0	S/.4	D	S/.0	S/.4	D	S/.0	S/.4	D
α	Fixation probability								
0.25	.066	.076	.079	.080	.042	.081	.050	.044	.058
1.0	.264	.292	.281	.272	.268	.286	.222	.188	.212
	Effective population size								
	14.7	13.7	13.3	—	—	26.7	—	—	16.0

− Simulation too computer time consuming.
+ Standard errors of fixation probability with 500 replicates.

u	.025	.05	.1	.15	.2	.3	.4	.5
SE = $[u(1-u)/500]^{1/2}$	.007	.010	.013	.016	.018	.020	.022	.022

TABLE 2

Fixation probability and effective population size for selection on combinations of individual performance (X) and full sib family mean (X̄) and within family selection (W), each with equal family sizes. (Standard errors as for Table 1).

Probabilité de fixation et taille efficace de population dans le cas de sélections portant sur la performance individuelle (X) et la moyenne de famille de germains (X̄) ou d'une sélection intra-famille (W), les tailles de famille étant égales. (Ecarts-types identiques à ceux du tableau 1).

h^2	Individual X		Index X + X̄		Family X̄		Deviation X − X̄		Within W	
	.0	.4	.0	.4	.0	.4	.0	.4	.0	.4
α	$N_m = N_f = 5$ $T_m = T_f = 10$ $i = 0.74$									
	Fixation probability									
0.25	.108	.076	.104	.066	.050	.036	.072	.080	.066	.080
.5	.210	.164	.210	.138	.110	.088	.130	.146	.134	.134
1.0	.382	.322	.328	.332	.242	.172	.306	.296	.244	.282
2.0	.602	.586	.590	.562	.354	.318	.530	.556	.508	.486
	Effective population size									
	15.6	12.9	12.1	10.5	8.5	7.4	16.4	15.8	19.3	19.1
α	$N_m = N_f = 5$ $T_m = T_f = 20$ $i = 1.21$									
	Fixation probability									
0.25	.044	.062	.074	.058	.036	.030	.054	.052	.048	.062
1.0	.320	.244	.246	.232	.146	.068	.246	.232	.248	.278
	Effective population size									
	12.3	11.1	8.9	7.3	4.3	3.8	14.7	13.0	18.2	17.8
α	$N_m = N_f = 10$ $T_m = T_f = 20$ $i = 0.77$									
	Fixation probability									
0.25	.110	.076	.100	.084	.102	.042	.090	.048	.050	.082
1.0	.366	.342	.312	.254	.270	.176	.306	.278	.250	.270
	Effective population size									
	28.4	24.3	19.0	19.9	16.4	14.5	33.7	31.1	41.1	—

Results for various combinations of index selection are given in Table 2, in which family sizes are fixed. Thus for mass selection (criterion is individual performance, X), fixation probabilities are higher than with random family sizes shown in Table 1. These differences reflect, but seem less than proportional to, the differences in effective population size — for example with 5 male and female parents, and 10 male and

female progeny, the effective size is a little over 10 for random size and approximately 15 for fixed size. The pattern of results is not too clear, but several points emerge : (i) family selection $(\overline{X})$ almost always leads to the lowest fixation probabilities ; (ii) the differences in fixation probability between selection on individual performance (X), deviation from family mean $(X - \overline{X})$ and the simple index $(X + \overline{X})$ are usually small, but generally the index $(X + \overline{X})$ and deviation $(X - \overline{X})$ gave results intermediate between those for mass selection (X) and family selection $(\overline{X})$; (iii) at high heritability, within family selection (W) generally gives higher fixation probabilities than at low heritability, while the reverse is the case for schemes $(X, X + \overline{X}$ and $\overline{X})$; for low heritability within family selection (W) gives substantially lower fixation probabilities than mass selection (X), for high heritability differences are less predictable.

Short term response, which utilises existing variation, is proportional to the accuracy, r_{IA} and selection intensity. The accuracies of the alternative schemes, expressed relative to h, are as follows, where the intraclass correlation is taken as $h^2/2$ and there are 4 progeny per family, i.e. $(T_m + T_f)/N_f = 4$

	r_{IA}/h	
	$h^2 \rightarrow 0$	$h^2 = 0.4$
Individual (X)	1	1
Index $(X + \overline{X})$	1.23	1.10
Family $(\overline{X})$	1.25	0.99
Deviation $(X - \overline{X})$	0.43	0.48

For within family selection, the reduced selection intensity (i_w) has to be taken into account, so the relative response is proportional to $0.48i_w/i$ for $h^2 = 0.4$ or $0.43i_w/i$ for $h^2 \rightarrow 0$, where for example, with $N_m = N_f = 5$, $T_m = T_f = 10$, $i_w/i = 0.763$. In Table 3 the fixation probabilities computed in Table 2 are expressed relative to $N_e \bar{i}\alpha r_{IA}/(Th)$, so if the most simple formulation were applied i.e. $u = N_e \bar{i}\alpha r_{IA}/(Th)$ by extension of eq. (3), all values would equal 1.0. It is seen that for mass selection and indices in which family selection is given positive weight, lower values are obtained, while for deviations from family mean and within family selection the fit is good. Therefore the relative efficiencies of the alternative criteria at fixing additive genes, and in generating response to very long term selection, differ from their efficiencies in obtaining short term response by utilising existing variation, particularly for genes of large effect.

In the examples given in Table 2 correlation of family members $(h^2/2)$ is introduced solely through additive genetic variation. A limited amount of simulation has confirmed that the general relation among the different selection schemes is not affected if this correlation is environmental (c^2) : namely the fixation probabilities for selection schemes giving positive weight to family mean tend to decrease and those for within family selection tend to increase.

When genes are neutral $(\alpha = 0)$, the fixation probability is simply $1/(2T)$. Simulation has not been conducted for very small values of $N\bar{i}\alpha$, when fixation probabilities are low, because sufficient precision can not be obtained in reasonable computing time. (SE (u) decreases as u decreases, but the coefficient of variation increases). However, if $N\bar{i}\alpha$ is small, the fate of the mutant is not « decided » just in the first selection cycle, for even if the mutant survives that, it may well be lost subsequently ; therefore the pattern must then follow $N\bar{i}r_{IA}$, as for initially segregating genes. Similarly, analysis has not been done using recessive mutants, because fixation probabilities are so low. For mutants that are completely dominant or have substantial effect in the heterozygote,

the general pattern is likely to be similar to that for additive genes, because most of the « decisions » are made in the first generation or two. Thus within family selection is likely to be efficient.

These results illustrate the conflict between short and long term response in any breeding programme with limited resources. In the short term the product of selection intensity and accuracy $(\bar{\imath}r_{IA})$ has to be maximised. In the very long term, and especially when mutational variation has to be taken into account, selection intensities need to be reduced to increase effective population size to maximize $N_e\bar{\imath}r_{IA}$, as noted by ROBERT-SON (1960). Further, endorsing the conclusions of DEMPFLÉ (1975), within family selection of low accuracy is relatively more efficient than schemes involving use of family information, even when account is taken of the difference in accuracy and effective population size. Perhaps the breeder should maintain both a highly selected line and a large, less intensely selected, line as a reserve.

TABLE 3

Fixation probability for selection schemes as in Table 2 expressed as $u/[N_e\bar{\imath}\alpha r_{IA}/(Th)]$.

Probabilité de fixation pour les schémas de sélection du tableau 2
exprimée par le rapport $u/[N_e\bar{\imath}\alpha r_{IA}/(Th)]$.

h^2	Individual X		Index X + $\overline{X}$		Family $\overline{X}$		Deviation X − $\overline{X}$		Within W	
	.0	.4	.0	.4	.0	.4	.0	.4	.0	.4
α				$N_m = N_f = 5$		$T_m = T_f = 10$				
0.25	.75	.64	.76	.62	.51	.53	1.10	1.13	1.12	1.24
0.5	.73	.69	.76	.65	.56	.65	.99	1.03	1.14	1.03
1.0	.66	.68	.60	.78	.62	.64	1.17	1.05	1.04	1.08
2.0	.52	.61	.54	.66	.45	.59	1.01	.98	1.08	.87
α				$N_m = N_f = 5$		$T_m = T_f = 20$				
0.25	.47	.74	.82	.92	.69	1.01	1.03	1.01	.88	1.04
1.0	.86	.72	.68	.92	.70	.57	1.18	1.12	1.13	1.16
α				$N_m = N_f = 10$		$T_m = T_f = 20$				
0.25	.81	.65	.89	.80	1.04	.61	1.29	.67	.80	1.23
1.0	.67	.73	.70	.61	.69	.64	1.09	.96	1.00	1.01
			Mean over population size and gene effect combinations							
	.68	.68	.72	.74	.66	.65	1.11	.99	1.02	1.08

The diffusion equation predicts and the results presented here show that the fixation probabilities of mutant genes are approximately independent of the size of populations having the same selection intensity and selection scheme (e.g. $N_m = N_f = 5$, $T_m = T_f = 10$ vs $N_m = N_f = 10$, $T_m = T_f = 20$).

However, the number of mutants per generation and thus the long term rate of response from mutations is proportional to population size (HILL, 1982). Although, for the same number of parents, an increase in the number of progeny recorded (e.g. $N_m = N_f = 5$, $T_m = T_f = 10$ vs $N_m = N_f = 5$, $T_m = T_f = 20$) mean that a mutant's initial frequency and thus fixation probability is decreased, the corresponding increase in number of mutants each generation more than compensates for this, so rates of response are expected to be higher.

Acknowledgements

References

BULMER M.G., 1970. — The effect of selection on genetic variability. *Am. Natur.,* **105**, 201-211.

DEMPFLÉ L., 1975. — A note on increasing the limit of selection through selection within families. *Genet. Res.,* **24**, 127-135.

EWENS W.J., 1979. — *Mathematical Population Genetics.* 325 pp., Springer-Verlag, Berlin.

FALCONER D.S., 1981. — *Introduction to Quantitative Genetics.* 2e ed., 340 pp., Longman, London.

HILL W.G., 1969. — On the theory of artificial selection in finite populations. *Genet. Res.,* **13**, 143-163.

HILL W.G., 1982. — Rates of change in quantitative traits from fixation of new mutations. *Proc. Natl. Acad. Sci. (U.S.A.),* **79**, 142-145.

JONES L.P., 1969. — Effects of artificial selection on rates of inbreeding in populations of *Drosophila melanogaster. Austr. J. Biol. Sci.,* **22**, 143-145.

KIMURA M., 1957. — Some problems of stochastic processes in genetics. *Ann. Math. Statist.,* **28**, 882-901.

LATTER B.D.H., 1965. — The response to artificial selection due to autosomal genes of large effect. I. Changes in gene frequency at an additive locus. *Austr. J. Biol. Sci.,* **18**, 585-598.

LUSH J.L., 1947. — Family merit and individual merit as bases for selection. *Am. Natur.,* **81**, 241-261, 362-379.

ROBERTSON A., 1960. — A theory of limits in artificial selection. *Proc. Roy. Soc. Lond.,* **B 153**, 234-249.

ROBERTSON A., 1961. — Inbreeding in artificial selection programmes. *Genet. Res.,* **2**, 189-194.