On the fixation probability of mutant genes in a subdivided population*

By TAKEO MARUYAMA

National Institute of Genetics, Mishima, Japan

SUMMARY

Following Moran's (1962) method, it was shown that the fixation probability of a mutant gene is not altered by the subdivision of a population into partially isolated colonies, if the following conditions are met; fitness is additive, samplings and selection is done separately in each colony, and migration between colonies does not change the gene frequency in the whole population. This conclusion was checked by simulation experiments.

In evolutionary theory, and in animal and plant breeding, the fixation probability of genes plays a very important role. Kimura (1957 and 1962) has solved this problem for a randomly mating polpuation of finite size. He considers a locus with two alleles A_1 and A_2 segregating, and assumes that the relative fitnesses of the A_1A_1, A_1A_2 and A_2A_2 genotypes are 1, $1+2hs$ and $1+2s$ respectively. If the initial frequency of A_2 is p and the effective size of the population is N, Kimura has shown that the diffusion approximation to the ultimate fixation probability of allelle A_2 is

$$U(p) = \frac{\int_0^p G(x)\,dx}{\int_0^1 G(x)\,dx}, \qquad (1)$$

where $G(x) = \exp[-4Ns(2h-1)x(1-x) - 4Nsx]$.

Robertson (1960) used this fixation probability in his theory of selection limits. Kimura & Ohta (1969) used also this probability in the calculation of the time required for mutant genes to reach fixation. Hill & Robertson (1966) studied the fixation probability for two linked loci in a finite panmictic population. They investigated, in particular, the effect of one locus on the fixation probability of the other locus, and obtained interesting conclusions. Ohta (1968) also studied the problem for two linked loci, with special reference to the effect of initial linkage disequilibrium on the fixation probability. Kimura has solved the problem for a case of two independent loci with epistatic effects (see Ohta, 1968).

Since all natural populations are geographically distributed, it is an important

15-2

problem to consider the fixation probability in a population with geographical structure. Although they may be usually continuously distributed, the two-dimensional stepping stone model should be a good approximation to their structure. In this report I will consider a population consisting of a finite number of partially isolated colonies and obtain an approximation formula for the fixation probability for one locus with two alleles A_1 and A_2 which are additive in fitness.

Moran (1962) gave an alternative method in obtaining the fixation probability $U(p)$ in (1) for a special case of additive gene action. Let $y^{(t)}$ be the random variable describing the number of A_2 genes at the t-th generation in the population. Moran considers the quantity $E[\exp(zy^{(t)})]$, where E is the expectation operator and z is a real number. He shows that if $z = -2s$, then $E[\exp(-2sy^{(t)})]$ is almost invariant with time, i.e. $E[\exp(-2sy^{(t)})] \approx E[\exp(-2sy^{(t+1)})] \approx \dots$. Considering this quantity at time zero and at time infinity we have

$$\exp(-2s2Np) = U(p)\exp(-2s2N) + (1 - U(p)),$$

where p and $U(p)$ are the initial frequency and fixation probability of A_2. Therefore we have

$$U(p) = \frac{1 - \exp(-4Nsp)}{1 - \exp(-4Ns)},$$

which is a special case ($h = \frac{1}{2}$) of the expression for $U(p)$ given in (1).

This can be extended to more general situations, still assuming additive fitnesses ($h = \frac{1}{2}$). Suppose that we have a large population of size N and that the population is subdivided into colonies. Let N_i be the size of colony i, $Y_i^{(t)}$ be the number of A_2 alleles in colony i at the t-th generation and $Y^{(t)}$ be the total number of A_2 alleles at the t-th generation, i.e. $Y^{(t)} = \sum_i Y_i^{(t)}$ and $N = \sum_i N_i$. Assume that the samplings of gametes from one generation to the next is done independently in each colony. Then we have,

$$E[\exp(-2sY_1^{(t)})] \times E[\exp(-2sY_2^{(t)})] \times \dots = E[\exp(-2sY^{(t)})]. \tag{2}$$

Now suppose that selection is done independently in each colony. Then

$$E[\exp(-2sY_i^{(t)})] \approx E[\exp(-2sY_i^{(t+1)})] \tag{3}$$

for all i, as Moran has shown. Thus

$$E[\exp(-2sY_1^{(t+1)})] \times E[\exp(-2sY_2^{(t+1)})] \times \dots = E[\exp(-2s(\sum_i Y_i^{(t+1)}))]$$
$$\equiv E[\exp(-2sY^{(t+1)})] \approx E[\exp(-2sY^{(t)})]$$

follows from (2) and (3). Therefore $E[\exp(-2sY^{(t)})]$ is invariant over splitting of a population and selection within colonies. The invariance still holds over migration of individuals between colonies since this does not change the total number of A_2 alleles. If each colony can be reached from any other colony by migration in one

generation or several generations, either A_1 or A_2 will be eventually lost from the whole population. For such a situation, as above, we have

$$\exp\left(-2s\,Y^{(0)}\right) \approx U(\bar{p})\exp\left(-2s\,2N\right)+(1-U(\bar{p}))$$

in which $U(\bar{p})$ is the fixation probability of A_2 in the population with the initial frequency,

$$\bar{p} = \frac{1}{N}\sum N_i(Y_i^{(0)}/2N_i) = \frac{1}{N}\sum_i p_i N_i,$$

where p_i is the initial frequency in colony i. Therefore we have

$$U(\bar{p}) = \frac{1-\exp\left(-4Ns\bar{p}\right)}{1-\exp\left(-4Ns\right)}. \tag{4}$$

This is equal to the fixation probability in a panmictic population of size $N = \sum\limits_i N_i$. However it should be noted that if the subdivision of the population prolongs the time of fixation very much formula (4) may not be valid.

In order to check the validity of formula (4), I have performed a number of Monte Carlo simulations by computer. The scheme of the experiment is as follows. A population is divided into colonies of equal size, N_i. At the beginning of each generation each colony produces infinite number of A_1 and A_2 gametes whose relative frequencies are equal to those of A_1 and A_2 alleles in the colony at the end of the
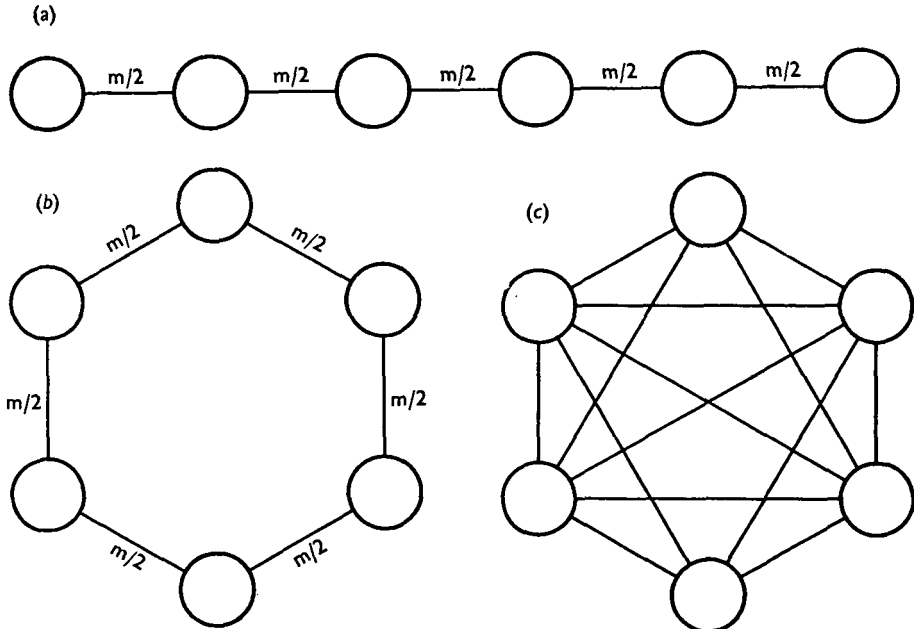

Fig. 1. (*a*) Linear stepping stone model, (*b*) circular stepping stone model and (*c*) Island model. Open circles indicate colonies and lines indicate migration. In the stepping stone models migration is only between geographically adjacent colonies. In the island model migration occurs between any pair of colonies and the rates are equal for every pair.

Table 1. *Comparison of the theoretical value with the value obtained by simulation experiments*

No. of colonies (n)	Colony size (N_i)	Total population size ($N = nN_i$)	Migration* rate (m)	Selection coefficient (s)	Initial frequency† p_1 / p_6	p_2 / p_7	p_3 / p_8	p_4 / p_9	p_5 / p_{10}	Average initial frequency (p)	Fixation probability Formula (4) $U(p)$	Fixation probability Simulation	No. of simulation experiments performed
Linear stepping stone model													
5	10	50	0.05	0.05	0.00 / 0.00	0.00 / 0.00	0.10 / 0.00	0.00 / 0.00	0.00 / 0.00	0.02	0.181	0.186	1000
5	10	50	0.10	0.05	0.00 / 0.00	0.00 / 0.00	0.10 / 0.00	0.00 / 0.00	0.00 / 0.00	0.02	0.181	0.178	1000
10	2.5	25	0.10	0.02	1.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.10	0.209	0.193	963
10	4	40	0.05	0.01	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.40 / 0.00	0.04	0.0777	0.0846	910
10	5	50	0.05	0.01	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.40 / 0.00	0.04	0.0952	0.0889	714
Circular stepping stone model													
5	5	25	0.20	0.10	0.10 / 0.00	0.10 / 0.00	0.10 / 0.00	0.10 / 0.00	0.10 / 0.00	0.10	0.632	0.580	100
5	10	50	0.05	0.10	0.05 / 0.00	0.10 / 0.00	0.05 / 0.00	0.00 / 0.00	0.00 / 0.00	0.04	0.551	0.564	1000
10	5	50	0.10	0.02	0.10 / 0.00	0.20 / 0.00	0.10 / 0.20	0.30 / 0.00	0.10 / 0.00	0.14	0.437	0.474	471
10	5	50	0.10	0.05	0.10 / 0.00	0.40 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.10	0.632	0.626	561
10	10	100	0.10	0.10	0.10 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.01	0.330	0.304	950
Island model													
3	15	45	0.01	0.02	0.05 / 0.00	0.15 / 0.00	0.10 / 0.00	—	—	0.10	0.311	0.310	510
3	10	30	0.05	0.10	0.05 / 0.00	0.10 / 0.00	0.05 / 0.00	—	—	0.066	0.550	0.538	2393
3	10	30	0.01	0.10	0.05 / 0.00	0.30 / 0.00	0.10 / 0.00	—	—	0.15	0.835	0.814	1020
3	10	30	0.10	0.10	0.05 / 0.00	0.50 / 0.00	0.10 / 0.00	—	—	0.217	0.926	0.925	1760
3	10	30	0.005	0.10	0.05 / 0.00	0.15 / 0.00	0.10 / 0.00	—	—	0.10	0.699	0.692	1208
3	10	30	0.10	0.10	0.10 / 0.00	0.50 / 0.00	0.10 / 0.00	—	—	0.233	0.939	0.921	882
3	50	50	0.10	0.10	0.01 / 0.00	0.07 / 0.00	0.04 / 0.00	—	—	0.04	0.909	0.907	224
3	10	30	0.10	0.02	0.20 / 0.00	0.00 / 0.00	0.025 / 0.00	—	—	0.075	0.835	0.810	241
5	5	25	0.10	0.10	0.10 / 0.00	0.10 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.04	0.330	0.334	1000
10	5	50	0.05	0.10	0.50 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.05	0.632	0.607	915
10	5	50	0.10	0.05	0.10 / 0.00	0.40 / 0.00	0.30 / 0.00	0.20 / 0.00	0.00 / 0.00	0.10	0.632	0.626	720
10	10	100	0.10	0.10	0.10 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.01	0.330	0.301	1000

* For stepping stone models the rate between adjacent colonies is equal to $m/2$, and for the island model the rate between any pair of colonies is equal to m/n.

† p_i = Initial frequency of A_2 gene in colony i. For linear stepping stone model colony 1 is the one locating at one end of the habitat, colony 2 is the next colony to it and so on. For circular stepping stone model colony 1 is set arbitrary but colonies 2, 3, ..., follows to it in one

previous generation. After production of gametes, migration of gametes takes place at certain rates. After the migration, the gametes mate at random in each colony, and selection occurs separately in each colony. After the selection, $2N_i$ gametes are chosen for the next generation by binomial sampling in each colony separately. The actual sampling of $2N_i$ gametes in colony i is done as follows: We draw $2N_i$ pseudo-random numbers uniformly distributed in $(0, 1)$ and compare them with the frequency, x_i, of A_2 gene after selection in colony i. The number of A_2 alleles in colony i at the beginning of the next generation is set equal to the number of the pseudo-random numbers drawn which are less than x_i. This procedure is repeated for each colony for each generation. The population structures used in the simulations are illustrated in Fig. 1. The results of the simulation experiments are presented in Table 1. The numerical results agree well with the theoretical values, and in fact there is no case in which the experimental result deviates significantly from the theoretical value.

REFERENCES

HILL, W. G. & ROBERTSON, A. (1966). The effect of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294.

KIMURA, M. (1957). Some problems of stochastic processes in genetics. *Ann. Math. Statist.* **28**, 882–901.

KIMURA, M. (1962). On the probability of fixation of mutant genes in a population. *Genetics, Princeton* **47**, 713–719.

KIMURA, M. & OHTA, T. (1969). The average number of generations until fixation of a mutant gene in a finite population. *Genetics, Princeton* **61**, 763–771.

MORAN, P. A. P. (1962). *The Statistical Process of Evolutionary Theory*. Oxford: Clarendon Press.

OHTA, T. (1968). Effect of initial linkage disequilibrium and epistasis on fixation probability in small population, with two segregating loci. *Theor. and Appl. Genetics* **38**, 243–248.

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