Genet. Res., Camb. (1971), **18**, *pp.* 125–131
With 3 *text-figures*
Printed in Great Britain

Pattern of neutral polymorphism in a geographically structured population*

BY MOTOO KIMURA AND TAKEO MARUYAMA
National Institute of Genetics, Mishima, Japan

(*Received* 2 *December* 1970)

SUMMARY

In a two-dimensional stepping-stone model of finite size, if a pair of alleles happen to segregate in the whole population, marked local differentiation of gene frequencies can occur only if migration between colonies is sufficiently rare so that $Nm < 1$, where N is the effective size of each colony and m is the rate at which each colony exchanges individuals with four surrounding colonies each generation. On the other hand, if $Nm \geqslant 4$, the whole population behaves as if it were panmictic and the allelic frequencies become uniform over the entire distribution range unless mutation is unusually high. Tendency toward local differentiation is much weaker in two-dimensional than in one-dimensional habitats.

Very often, a mode of selection is inferred from an observed pattern of polymorphism in natural populations. Thus, if the same pair of alleles are found in uniform frequencies over wide distribution range of the species, it is claimed that natural selection is actively maintaining these frequencies. If, on the other hand, different alleles are fixed in different local populations, or if there is a cline, these are often considered to be the result of local adaptation of these alleles. Furthermore, if the frequencies of alleles are uniform within each locality but different among localities, this is also assumed to indicate the existence of some form of 'balancing selection'.

Actually, selection can be invoked to explain any pattern of polymorphism in natural populations. Often, such presumed selection is used to refute the neutral polymorphism theory. The purpose of the present paper is to show that all the above patterns of polymorphism can equally be explained in terms of migration and random frequency drift of selectively neutral mutations.

In his recent study on the stepping-stone models of finite size, Maruyama (1971*b*) has shown that for a two-dimensional habitat, marked local differentiation of gene frequencies can occur if migration is sufficiently rare between colonies so that

$$mN < 1, \tag{1}$$

where N is the effective size of each colony and m is the rate at which each colony exchanges individuals with four surrounding colonies each generation so that the migration rate between a pair of adjacent colonies is $\frac{1}{4}m$. Random mating is assumed

---

* Contribution no. 805 from the National Institute of Genetics, Mishima, Shizuoka-ken, Japan. Aided in part by a Grant-in-Aid from the Ministry of Education, Japan.

within each colony. In deriving this formula, it was assumed that the number of colonies (k) is large and they are arranged on a torus. The main reason for using torus-like space is that it facilitates exact mathematical analysis (see Maruyama, 1970*b*). For a more realistic but less tractable model of rectangular arrangement of colonies, however, results are similar (for details, see Maruyama, 1970*b*). On the other hand, if

$$mN > 1, \tag{2}$$

local differentiation is less pronounced, and, especially if

$$mN \geqslant 4, \tag{3}$$

the whole population tend to behave as if it were a single panmictic population. In such a population, the average length of time from occurrence until fixation of a neutral mutant is $4N_T$, where $N_T = kN$ (Maruyama, 1971*a*). A slight difference of gene frequencies may still occur among colonies and a detailed mathematical study on this point will be published elsewhere (Maruyama, 1972). Since Nm is the average number of individuals which each colony exchanges with four surrounding colonies each generation, condition (3) may almost always be satisfied by actively moving animals. In addition, there is a good possibility of occasional long-range migration in natural populations, making practical panmixis still easier to attain. It is possible, however, that condition (1) applies often to plant species with low migration ability.

These conditions are pertinent when a small number of alleles are segregating in the population. These conditions have been obtained by investigating the process of random genetic drift in the entire population. The crucial quantities describing this process are the dominant eigenvalue (λ) and the corresponding eigenvector of the relevant recurrence equation. The former gives the rate of steady decay of genetic variation while the latter gives a set of coefficients of kinship between colonies. It is known (Wright, 1931) that in a single panmictic population of effective size N_e the rate of steady decay of genetic variance is $1/(2N_e)$ per generation, so that, in our terminology (using the transition probability matrix method)

$$1-\lambda = 1/(2N_e).$$

This also gives the ultimate rate at which the population becomes homallelic if the number of coexisting alleles is two (although, more generally for n alleles the rate of losing one of the n alleles is $n(n-1)/2N_e$, see Kimura, 1955). Therefore, we should expect that if

$$1-\lambda = 1/(2N_T), \tag{4}$$

in which $N_T = kN$, then the entire population will form a panmictic population. It can be shown that under condition (2) this is valid as the first approximation.

Applying Robertson's (1964) result to the present model we have in general

$$1-\lambda = \frac{1}{2N_T}\frac{1-f_0}{1-\bar{f}}, \tag{5}$$

where f_0 is the coefficient of kinship of individuals within a colony and $\bar{f}$ is that of two different individuals randomly taken from the whole population. Note that Robertson considered a mating system in which each individual has the same number of progeny, whereas we are considering a system in which this number varies binomially so that the above formula has the factor 2 instead of 4 as in Robertson's original formula. Thus, condition (4) implies $f_0 = \bar{f}$. Formula (5) also shows that the ratio $(1-f_0)/(1-\bar{f})$ is an appropriate indicator of local differentiation by random drift.

On the other hand, it can be shown that, if the rate of migration is lower (condition (1)), we have

$$1-\lambda = m/(2k) \tag{6}$$

as the first approximation. In this case, local differentiation of gene frequencies is expected (for exact nature of the eigenvalue and eigenvector, see Maruyama, 1971*b*). The transition from practical panmixis to marked local differentiation is rather rapid as shown in Fig. 1, where exact numerical values of $1-\lambda$ are plotted with dots. The two lines given by equations (4) and (6) meet at $mN = 1$ on the abscissa. This point divides the set of values of mN into two regions, $0 \leqslant mN < 1$ and $mN \geqslant 1$; the approximation (6) is valid in the former while (4) is valid in the latter.

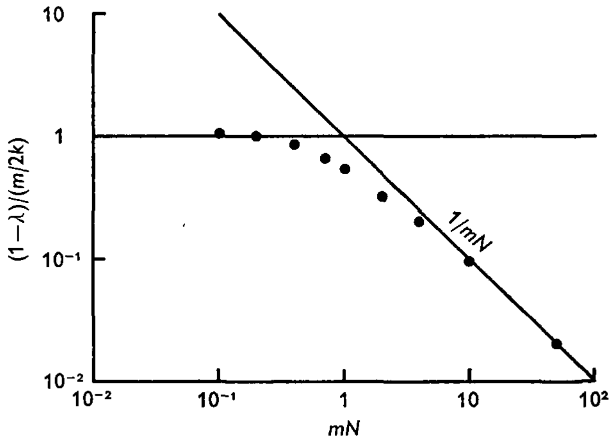

Fig. 1. Relationship between the rate of steady decay of genetic variation and the migration rate in a two-dimensional stepping-stone model. The two lines in the figure represent two approximations (4) and (6), while the dots represent exact numerical values computed by a matrix iteration method of Maruyama (1971*b*) applied to the case of 190 × 190 colonies arranged on a torus ($k = 190 \times 190$). N is the effective size of each colony, m is the rate at which each colony exchanges individuals with four surrounding colonies each generation, and λ is the dominant eigenvalue representing the steady decay.

Fig. 2 illustrates two examples from Monte Carlo experiments, one (A) illustrating the case of practical panmixis and the other (B) that of marked local differentiation. In these examples the total population consists of $20 \times 20 = 400$ colonies each with $N = 10$ breeding individuals. The simulation experiments (using TOSBAC 3400· computer) were continued over many generations and a new mutant allele is introduced whenever the whole population becomes homallelic. In the course of an experiment, if the average frequency, $\bar{q}$, of a mutant allele happened to become 0·1, its frequencies in all colonies were printed out to see the amount of local differentiation. Fig. 2A shows an example with $mN = 4$ while Fig. 2B that with $mN = 0{\cdot}25$. Note that in the latter case, in which local differentiation is evident, a cline is formed between regions where mutant frequencies are high and those where they are low. This pattern mimics a cline formed by selection.

When compared with the two-dimensional habitat, the tendency toward local differentiation is much more pronounced in one-dimensional habitat. This is illustrated in Fig. 3 using four examples from Monte Carlo experiments. In these examples the total population consists of 10 colonies arranged on a circle and each colony consists of $N = 50$ breeding individuals ($\bar{q} = 0{\cdot}2$).

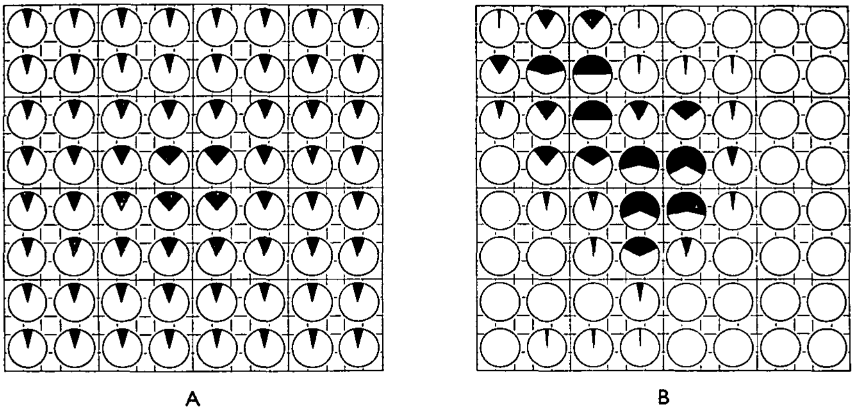

Fig. 2. Two examples from Monte Carlo experiments simulating the genetic change in a two-dimensional stepping-stone model with 20 × 20 colonies arranged on a torus. The product of the migration rate (m) and the colony size (N) is 4·0 in A but 0·25 in B. In both examples the average frequency of one of the alleles is 0·1 ($\bar{q} = 0{\cdot}1$). The dark sector in each circle represents the average frequency of this allele in 6·25 neighbouring colonies.

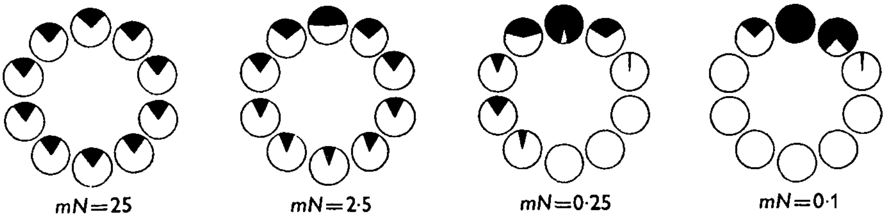

Fig. 3. Four examples from Monte Carlo experiments similar to Fig. 2 but assuming ten colonies arranged on a circle (a circular stepping-stone model).

It was shown by Maruyama (1970*a*) that, in circular stepping-stone models, a marked local differentiation occurs if

$$2mN < k/\pi^2. \tag{7}$$

This corresponds to condition (1) in the two-dimensional case. An important point to note here is that in the one-dimensional case a marked local differentiation is possible whenever k (number of colonies) is large.

Let us now consider a situation in which mutations occur at a constant rate each generation. Following Kimura & Crow (1964) we shall assume that the number of possible allelic states is so large that whenever a mutant appears it represents a new, not pre-existing allele. When equilibrium is reached between mutational production of alleles and their random extinction, a certain number of alleles are maintained in the population. Therefore, comparison between the effective number of alleles in the whole population and that in each colony will reveal the amount of local differentiation. The ratio between these two effective numbers of alleles turns out to be $f_0/\bar{f}$. This quantity is called by Crow & Maruyama (1971) the 'effective number of colonies' (k_e). Table 1 lists $\bar{f}$ and f_0 for several examples of the stepping-stone model of finite size. The table has been calculated by using formulae (3–8), (3–9), (5–10) and (5–12) of Maruyama (1970*b*). The table shows that for $4N_T u$ up to 0·1 (cases 1–10), local differentiation is slight if $mN \geqslant 4$, but is pronounced if $mN < 1$. For higher rate of production of mutants such that $4N_T u \geqslant 2$ (cases 11–15), marked local differentiation can occur as will be discussed later. Case 16 is an example of one-dimensional arrangement of colonies where tendency for local differentiation is much stronger than in the cases of two-dimensional arrangements.

Table 1. *Exact numerical values of several statistics* (f_0, $\bar{f}$, *etc.*) *relating to the amount of local differentiation for various combination of k* (*number of colonies*), *N* (*colony size*), *u* (*mutation rate*) *and m* (*migration rate*)

(N_T stands for Nk. Those statistics were computed by using the analytical solutions of Maruyama (1970*b*). Of the 16 cases listed, only the last one is one-dimensional, all others being two-dimensional.)

Case	k	N	$4N_T u$	mN	f_0	$\bar{f}$	$f_0/\bar{f}$	$\frac{(1-f_0)}{(1-\bar{f})}$
1	200 × 200	125	0·05	0·2	0·9605	0·7891	1·2171	0·1874
2	200 × 200	125	0·05	0·5	0·9558	0·8823	1·0834	0·3752
3	100 × 100	100	0·05	1·0	0·9540	0·9193	1·0378	0·5698
4	100 × 100	100	0·05	4·0	0·9528	0·9447	1·0084	0·8555
5	200 × 200	125	0·10	0·1	0·9496	0·5037	1·8854	0·1015
6	200 × 200	125	0·10	0·2	0·9351	0·6485	1·4421	0·1845
7	200 × 200	125	0·10	0·5	0·9215	0·7842	1·1752	0·3635
8	100 × 100	100	0·10	1·0	0·9150	0·8495	1·0772	0·5643
9	100 × 100	100	0·10	4·0	0·9106	0·8940	1·0186	0·8430
10	100 × 100	100	0·10	10·0	0·9096	0·9034	1·0069	0·9354
11	100 × 100	100	2·00	1·0	0·5597	0·2202	2·5413	0·5647
12	500 × 500	40	2·00	10·0	0·3733	0·3133	1·1914	0·9127
13	1000 × 1000	1000	100·00	10·0	0·1081	0·0089	12·1239	0·9000
14	1000 × 1000	1000	100·00	100·0	0·0207	0·0098	2·1112	0·9890
15	1000 × 1000	1000	100·00	400·0	0·0124	0·0099	1·2561	0·9974
16	1000 × 1	10000	0·40	4·0	0·8985	0·2537	3·5416	0·1361

Note that $(1-f_0)/(1-\bar{f})$ is a good indicator of local differentiation when $4N_T u$ is small, say much less than unity. In such a case, both f_0 and $\bar{f}$ are close to unity and the ratio $f_0/\bar{f}$ is not very informative as an indicator of local differentiation. On the other hand, when $4N_T u$ is much larger, say larger than 2, both f_0 and $\bar{f}$ may be small and $(1-f_0)/(1-\bar{f})$ cannot be very informative. In such a case, $f_0/\bar{f}$ serves as a much better indicator of local differentiation.

Great difference between one- and two-dimensional habitats regarding their tendency toward local differentiation can be shown by the following treatment where mutation is also taken into account. First, consider two-dimensional distribution of k colonies. Let u be the rate per gamete per generation for neutral mutation. Suppose a mutant allele appears in one of the colonies. Let σ^2 be the mean squared distance of individual migration for one generation, taking one step as the unit length. Then the mean squared distance between an individual having this mutant and its descendants t generations later is $t\sigma^2$. In terms of standard deviation in one direction (either horizontally or vertically), this is $\sigma\sqrt{t}$. Since $1/u$ is the average length of time between two consecutive substitutions of neutral mutants (cf. Crow & Kimura (1970), p. 369), we should expect a marked local differentiation to occur by random drift if the standard deviation of migration in this interval of time (i.e. ($\sigma\sqrt{t}$ with $t = 1/u$) is much smaller than the total length of the habitat in one direction; that is, $\sqrt{k}$. In such a situation, a new 'lucky' mutant always starts to increase its frequency well before the previous mutant reaches complete fixation.

Noting that $\sigma^2 = m$ in the stepping-stone model, we have

$$\sqrt{(m/u)} \ll \sqrt{k}. \tag{8}$$

Squaring both sides of this inequality, multiplying through N and noting $Nk = N_T$, this gives

$$Nm \ll N_T u \tag{9}$$

as a condition for marked local differentiation.

Next, consider the opposite situation in which $\sqrt{(m/u)}$ is much larger than $\sqrt{k}$. Then, the whole population should behave as if it were panmictic. It is likely that if the standard deviation $\sqrt{(m/u)}$ is twice or more of the entire length $\sqrt{k}$, practical panmixis is attained. Thus, we have approximately

$$Nm > 4N_T u \tag{10}$$

as a condition for practical panmixis for a two-dimensional habitat.

So far, we have obtained two set of conditions, (1), (3) and (9), (10), using different methods of treatment for the two-dimensional case. They are complementary and their validity depends on the magnitude of $N_T u$. Namely, if $N_T u$ is much smaller than unity, conditions (1) and (3) are valid, for a segregating population contains at most a pair of alleles and the type of analysis made by Maruyama (1971*b*) is pertinent. The majority of observed protein polymorphisms may belong to this case. If, on the other hand, $N_T u$ is much larger than unity, (9) and (10) are valid. A marked local differentiation can occur if (9) is satisfied even if $mN \geqslant 4$. In addition practical panmixis is possible under (10) even if $N_T u$ is very large. Table 1 lists several examples to show these points.

A similar treatment as used to derive (9) can be applied to one-dimensional arrangement of colonies. In this case the total length of habitat is k rather than $\sqrt{k}$, so that the condition for marked local differentiation becomes

$$\sqrt{(m/u)} \ll k$$

or

$$Nm \ll N_T uk. \tag{11}$$

Note that the right-hand side now contains factor k, so that a marked local differentiation occurs whenever k is large, analogous to condition (7). These conditions show that tendency toward local differentiation is much stronger in one-dimensional than in two-dimensional habitats. The same conclusion has also been obtained by Wright (1951) in his analysis of 'isolation by distance'.

Returning to condition (3), we should like to point out that observed pattern and frequencies of protein polymorphisms in *Drosophila*, mouse and man may be explained by neutral mutation-random drift theory by assuming that in these species $4N_T u$ is about 0·1 on the average per cistron and that there is sufficient migration between colonies so that condition (3) is usually satisfied. If this is correct, the majority of protein polymorphisms are a transient phase of molecular evolution as recently claimed by Kimura & Ohta (1971).

We would like to thank Dr Kazutoshi Mayeda for reading the manuscript and correcting the English. Thanks are also due to the referee for constructive criticism.

REFERENCES

CROW, J. F. & KIMURA, M. (1970). *An Introduction to Population Genetics Theory.* New York: Harper and Row.

CROW, J. F. & MARUYAMA, T. (1971). The number of neutral alleles maintained in a finite, geographically structured population. *Theoretical Population Biology* (in the Press).

KIMURA, M. (1955). Random genetic drift in multi-allelic locus. *Evolution* **9**, 419–435.

KIMURA, M. & CROW, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.

KIMURA, M. & OHTA, T. (1971). Protein polymorphism as a phase of molecular evolution. *Nature* **229**, 467–469.

MARUYAMA, T. (1970*a*). On the rate of decrease of heterozygosity in circular stepping stone models of populations. *Theoretical Population Biology* **1**, 101–119.

MARUYAMA, T. (1970*b*). Effective number of alleles in a subdivided population. *Theoretical Population Biology* **1**, 273–306.

MARUYAMA, T. (1971*a*). Speed of gene substitution in a geographically structured population. *American Naturalist* **105**, 253–265.

MARUYAMA, T. (1971*b*). Asymptotic rate of decrease of genetic variability in subdivided populations. Submitted to *Theoretical Population Biology.*

MARUYAMA, T. (1972). On the local differentiation of gene frequencies in a geographically structured population. *Theoretical Population Biology.* (In preparation.)

ROBERTSON, A. (1964). The effect of non-random mating within inbred lines on the rate of inbreeding. *Genetical Research* **5**, 164–167.

WRIGHT, S. (1931). Evolution in Mendelian populations. *Genetics* **16**, 97–159.

WRIGHT, S. (1951). The genetical structure of populations. *Annals of Eugenics* **15**, 323–354.