Decay of genetic variability in geographically structured populations[†]

(population genetics/random drift/population structure/migration)

THOMAS NAGYLAKI

Department of Biophysics and Theoretical Biology, University of Chicago, 920 East 58th Street, Chicago, Illinois 60637

ABSTRACT   The ultimate rate and pattern of approach to equilibrium of a diploid, monoecious population subdivided into a finite number of equal, large, panmictic colonies are calculated. The analysis is restricted to a single locus in the absence of selection, and every mutant is assumed to be new to the population. It is supposed that either the time-independent backward migration pattern is symmetric in the sense that the probability that an individual at position x migrated from y equals the probability that one at y migrated from x, or it depends only on displacements and not on initial and final positions. Generations are discrete and nonoverlapping. Asymptotically, the rate of convergence is approximately $(1 - u)^{2t}[1 - (2N_T)^{-1}]^t$, where u, N_T, and t denote the mutation rate, total population size, and time in generations, respectively; the transient part of the probability that two homologous genes are the same allele is approximately independent of their spatial separation. Thus, in this respect the population behaves as if it were panmictic.

Much effort has been devoted to investigating the evolution of populations under the joint action of mutation, migration, and random genetic drift. The amount and spatial pattern of genetic variability at equilibrium have been thoroughly explored for both finite (1–7) and infinite (7–17) habitats. Knowledge of the rate of convergence to equilibrium is essential for understanding when the equilibrium might actually be observed. The transient problem is particularly significant because, in the absence of mutation, not only finite, but even geographically structured infinite populations in one and two dimensions tend to complete genetic uniformity. The approach to equilibrium is studied in refs. 5–7, 13, and 18–21 for finite habitats and in refs. 7, 13, 15, 16, 22, and 23 for infinite ones.

The only general result known on the convergence to equilibrium in finite habitats was derived in a continuous space-continuous time model (13). If the migration function has no infinitesimal first moments and the population density is very high, it was shown that the ultimate rate of convergence is close to $\exp\{-[2u + (2N_T)^{-1}]t\}$, where u, N_T, and t are the mutation rate, total population size, and time in generations, respectively. It was further demonstrated that asymptotically the transient part of the probability of allelic identity is constant. However, in more than one dimension the continuous time-continuous space model has an unacceptable singularity (5, 13). Furthermore, some populations are distributed in discrete colonies and have nonoverlapping generations. Therefore, it is important to study this problem in the stepping-stone model (24, 25).

Let us suppose that there are N diploid monoecious individuals in each panmictic colony. We use the vector x to locate the various colonies. The number of spatial dimensions is arbitrary. The n colonies exchange migrants so that the probability that an individual at x migrated from y is $m(y,x)$, independent of time. Thus,

$$\sum_y m(y,x) = 1, \qquad [1]$$

where the sum is over all the colonies, as are all sums below. We posit that there is no selection, and all alleles at the locus under consideration mutate to types not preexisting in the population at rate u per generation. Let $f(t,x,y)$ represent the probability that two homologous genes chosen at random in generation t from subpopulations at x and y are the same allele. Identity in allelic state is directly measurable and pertinent to the amount of genetic diversity in an evolving natural population. With discrete nonoverlapping generations, f satisfies (9, 16, 22)

$$f(t + 1,x,y) = (1 - u)^2 \left\{ \sum_{zw} m(z,x)m(w,y)f(t,z,w) \right.$$
$$\left. + (2N)^{-1} \sum_z m(z,x)m(z,y)[1 - f(t,z,z)] \right\}. \quad [2]$$

We decompose f into its equilibrium and transient components according to

$$f(t,x,y) = f_\infty(x,y) + F(t,x,y), \qquad [3]$$

where $F(t,x,y) \to 0$ as $t \to \infty$. Substituting [3] into [2], we obtain

$$f_\infty(x,y) = (1 - u)^2 \left\{ \sum_{zw} m(z,x)m(w,y)f_\infty(z,w) \right.$$
$$\left. + (2N)^{-1} \sum_z m(z,x)m(z,y)[1 - f_\infty(z,z)] \right\}, \quad [4a]$$

$$F(t + 1,x,y) = (1 - u)^2 \left\{ \sum_{zw} m(z,x)m(w,y)F(t,z,w) \right.$$
$$\left. - (2N)^{-1} \sum_z m(z,x)m(z,y)F(t,z,z) \right\}. \quad [4b]$$

To find the rate of convergence, we set

$$F(t,x,y) = (1 - u)^{2t}\lambda^t\psi(x,y)$$

and deduce from [4b]

$$(L + Q)\psi = \lambda\psi, \qquad [5a]$$

with

$$L\psi(x,y) = \sum_{zw} m(z,x)m(w,y)\psi(z,w), \qquad [5b]$$

$$Q\psi(x,y) = -(2N)^{-1} \sum_z m(z,x)m(z,y)\psi(z,z). \qquad [5c]$$

As the subpopulation sizes become very large, $N \to \infty$, we expect the dominant eigenvalue, λ_0, to tend to unity. Indeed,

---

[†] This is the third paper of a series. The second paper is ref. 15.

owing to [1], the constant ϕ_0 is an eigenfunction of L with eigenvalue 1: $L\phi_0 = \phi_0$. Introducing for any functions $\phi(x,y)$, $\psi(x,y)$ the inner product

$$(\phi,\psi) = \sum_{xy} \phi^*(x,y)\psi(x,y) \qquad [6]$$

where the asterisk indicates complex conjugation, we normalize ϕ_0 so that $(\phi_0,\phi_0) = 1$, which implies $\phi_0 = 1/n$. To apply standard perturbation theory to estimate λ_0, we need to prove that L is self-adjoint. In that case (26),

$$\lambda_0 \approx 1 + (\phi_0,Q\phi_0). \qquad [7]$$

We shall establish self-adjointness under two conditions. We exclude pathologies with the biologically reasonable assumption that the migration pattern is irreducible and ergodic (27).

Symmetric migration

Let us assume that the probability that an individual at y originated at x equals the probability that one at x migrated from y:

$$m(x,y) = m(y,x). \qquad [8]$$

Exchange between nearest neighbors provides the simplest example. From [5b], [6], and [8] we have

$$(\phi,L\psi) = \sum_{xyzw} \phi^*(x,y)m(z,x)m(w,y)\psi(z,w)$$

$$= \sum_{xyzw} [m(x,z)m(y,w)\phi(x,y)]^*\psi(z,w)$$

$$= (L\phi,\psi),$$

as required. Now [1], [5c], [7], and [8] yield

$$\lambda_0 \approx 1 - (2N)^{-1} \sum_{xyz} \phi_0 m(z,x)m(z,y)\phi_0$$

$$= 1 - (2Nn^2)^{-1} \sum_{z} \sum_{x} m(x,z) \sum_{y} m(y,z)$$

$$= 1 - (2N_T)^{-1}, \qquad [9]$$

where $N_T = Nn$ is the total population size.

Maruyama has derived [9] and the approximate uniformity of the probability of identity for nearest neighbor exchange between subpopulations arranged in a circle (18) and in a line (19).

Homogeneous migration

We assume now that the migration pattern depends only on displacement and not on initial and final positions. With a finite number of colonies, this can happen if the habitat has a natural periodicity. Thus, we take

$$m(x,y) = \mu(y - x| \text{ mod } b), \qquad [10]$$

where the notation means that $y_i - x_i$ is calculated mod b_i for all components. For example, for 13 colonies in a circle, x and y are just integers, $y - x$ being evaluated mod 13. In two dimensions, we have the torus-like model (1, 2, 7, 21). Clearly,

$$\psi(x,y) = \Psi(x - y| \text{ mod } b). \qquad [11]$$

From [5b] and [6] we have

$$(\phi,L\psi) = \sum_{xyzw} \Phi^*(x - y| \text{ mod } b)\mu(x - z| \text{ mod } b)$$

$$\times \mu(y - w| \text{ mod } b)\Psi(z - w| \text{ mod } b)$$

$$= \sum_{xx'y'z'} \Phi^*(x'| \text{ mod } b)\mu(z'| \text{ mod } b)$$

$$\times \mu(z' + y' - x'| \text{ mod } b)\Psi(y'| \text{ mod } b),$$

where $x' = x - y$, $y' = z - w$, and $z' = x - z$. Therefore,

$$(\phi,L\psi) = n \sum_{x'y'} \Phi^*(x'| \text{ mod } b)r(y' - x'| \text{ mod } b)$$

$$\times \Psi(y'| \text{ mod } b), \qquad [12]$$

in which

$$r(x| \text{ mod } b) = \sum_{y} \mu(y| \text{ mod } b)\mu(y + x| \text{ mod } b). \qquad [13]$$

Now, r is even, for

$$r(-x| \text{ mod } b) = \sum_{y} \mu(y| \text{ mod } b)\mu(y - x| \text{ mod } b)$$

$$= \sum_{z} \mu(z + x| \text{ mod } b)\mu(z| \text{ mod } b)$$

$$= r(x| \text{ mod } b). \qquad [14]$$

Hence, [12] yields

$$(L\phi,\psi) = n \sum_{xy} r(y - x| \text{ mod } b)\Phi^*(y| \text{ mod } b)\Psi(x| \text{ mod } b)$$

$$= n \sum_{xy} \Phi^*(y| \text{ mod } b)r(x - y| \text{ mod } b)\Psi(x| \text{ mod } b)$$

$$= (\phi,L\psi),$$

as required for self-adjointness.

Returning to [5c] and [7], we derive

$$\lambda_0 \approx 1 - (2Nn^2)^{-1} \sum_{xyz} \mu(x - z| \text{ mod } b)\mu(y - z| \text{ mod } b)$$

$$= 1 - (2N_T)^{-1},$$

as for symmetric migration.

We conclude that with respect to the transient component of the probability of allelic identity the population behaves asymptotically as if it were panmictic. The range of subpopulation sizes for which this result holds depends on the arrangement of the colonies and the migration pattern (5–7, 13, 18–21). Our analysis shows that even if L is not self-adjoint, ultimately the transient component of the probability of identity will be approximately constant. Provided L can be diagonalized, as is generically the case, [5] informs us that the rate of convergence is $(1 - u)^{2t}[1 - k(2N_T)^{-1}]^t$, where k depends on the configuration and number of colonies and the migration structure.

1. Maruyama, T. (1970) "Effective number of alleles in a subdivided population," *Theor. Pop. Biol.* 1, 273–306.
2. Maruyama, T. (1970) "Stepping-stone models of finite length," *Adv. Appl. Prob.* 2, 229–258.
3. Maruyama, T. (1970) "Analysis of population structure. I. One-dimensional stepping-stone models of finite length," *Ann. Hum. Genet.* 34, 201–219.
4. Maruyama, T. (1971) "Analysis of population structure. II. Two-dimensional stepping-stone models of finite length and other geographically structured populations," *Ann. Hum. Genet.* 35, 179–196.
5. Fleming, W. H. & Su, C.-H. (1974) "One-dimensional migration

Genetics: Nagylaki

Proc. Natl. Acad. Sci. USA 74 (1977) 2525

models in population genetics theory," *Theor. Pop. Biol.* **5**, 431–449.

6. Nagylaki, T. (1974) "Genetic structure of a population occupying a circular habitat," *Genetics* **78**, 777–790.

7. Malécot, G. (1975) "Heterozygosity and relationship in regularly subdivided populations," *Theor. Pop. Biol.* **8**, 212–241.

8. Malécot, G. (1967) "Identical loci and relationship," *Proc. Fifth Berkeley Symp. Math. Stat. Prob.* **4**, 317–332.

9. Malécot, G. (1969) *The Mathematics of Heredity* (W. H. Freeman and Co., San Francisco).

10. Kimura, M. & Weiss, G. H. (1964) "The stepping-stone model of population structure and the decrease of genetic correlation with distance," *Genetics* **49**, 561–576.

11. Weiss, G. H. & Kimura, M. (1965) "A mathematical analysis of the stepping-stone model of genetic correlations," *J. Appl. Prob.* **2**, 129–149.

12. Maruyama, T. (1969) "Genetic correlation in the stepping-stone model with non-symmetrical migration rates," *J. Appl. Prob.* **6**, 463–477.

13. Nagylaki, T. (1974) "The decay of genetic variability in geographically structured populations," *Proc. Natl. Acad. Sci. USA* **71**, 2932–2936.

14. Nagylaki, T. (1975) "The relation between distant individuals in geographically structured populations," *Math. Biosci.* **28**, 73–80.

15. Nagylaki, T. (1976) "The decay of genetic variability in geographically structured populations. II," *Theor. Pop. Biol.* **10**, 70–82.

16. Nagylaki, T. (1977) "The geographical structure of populations," in *Studies in Mathematical Biology*, ed. Levin, S. (Mathematical Association of America, Washington, DC), in press.

17. Sawyer, S. (1977) "Asymptotic properties of the probability of identity in a geographically structured population," *Adv. Appl. Prob.*, in press.

18. Maruyama, T. (1970) "On the rate of decrease of heterozygosity in circular stepping-stone models of populations," *Theor. Pop. Biol.* **1**, 101–119.

19. Maruyama, T. (1970) "Rate of decrease of genetic variability in a subdivided population," *Biometrika* **57**, 299–311.

20. Maruyama, T. (1971) "The rate of decrease of heterozygosity in a population occupying a circular or a linear habitat," *Genetics* **67**, 437–454.

21. Maruyama, T. (1972) "Rate of decrease of genetic variability in a two-dimensional continuous population of finite size," *Genetics* **70**, 639–651.

22. Sawyer, S. (1976) "Results for the stepping-stone model for migration in population genetics," *Ann. Prob.* **4**, 699–728.

23. Sawyer, S. (1977) "Rates of consolidation in a selectively neutral migration model," *Ann. Prob.*, in press.

24. Malécot, G. (1950) "Quelques schémas probabilistes sur la variabilité de populations naturelles," *Ann. Univ. Lyon Sci. Sec. A* **13**, 37–60.

25. Kimura, M. (1953) "Stepping-stone model of population," *Ann. Rept. Natl. Inst. Genet. Jpn* **3**, 62–63.

26. Mathews, J. & Walker, R. L. (1964) *Mathematical Methods of Physics* (W. A. Benjamin, Inc., New York).

27. Feller, W. (1968) *An Introduction to Probability Theory and Its Applications* (John Wiley & Sons, Inc., New York).