A MODEL OF KIN-STRUCTURED MIGRATION

ALAN R. ROGERS

Department of Anthropology, University of Pittsburgh, Pittsburgh, PA 15260

Abstract. —When individuals disperse from one local group to another, they often do so in the company of relatives. This is known as "kin-structured migration," and its effect on genetic population structure is investigated here. It is shown that when migration is kin-structured, the ratio of between- to within-group variance is increased by a quantity that can be estimated either from behavioral or genetic data. Theoretical results indicate that kin-structured migration should be most important in populations with high mobility, and analysis of data for humans and lions suggests the kin-structured migration may have a substantial effect on genetic population structure in both species. Its effect seems to be small in a population of pine voles.

In many species, social interactions between individuals are influenced by their relatedness. Closely related individuals assist one another in various ways, and they may be antagonistic toward non-relatives. It is not surprising, therefore, that when individuals disperse from one local group to another, they often do so in the company of relatives. Alan Fix (1975, 1978) coined the term "kin-structured migration" (KSM) to refer to this phenomenon, and it seems to be quite widespread. Fix described it among the Semai Senoi, a human population in Malaysia, and it has also been reported in Rhesus macaques (Southwick et al., 1965; Chepko-Sade and Sade, 1979; Meikle and Vessy, 1981; Colvin, 1983), Japanese macaques (Furuya, 1968, 1969; Lee, 1983), baboons (Nash, 1976), lions (Bygott et al., 1979), vervet monkeys (Cheney, 1983a, 1983b), and pine voles (FitzGerald, unpubl.).

I will use the term "kin structure" to refer to the extent to which migration (or any other behavior) is structured by kinship, and will later discuss how the kin structure of migration can be measured. KSM is interesting because of the effect it has on genetic differences between local groups. In general, migration tends to reduce such differences, but kin structure diminishes this effect. Consequently, kin structure increases the genetic variation expected among local groups (Fix, 1978). Thus, KSM is a process through which social behavior affects genetic population structure.

KSM should be distinguished from a related phenomenon, "lineal fission." Both involve kin structure, but as I use the term, KSM refers to recurrent mobility among existing groups, whereas "lineal fission" refers to the process through which local groups split apart to form new groups. During fissions, close relatives often stay together and the split consequently occurs along lineage boundaries (Neel and Salzano, 1967). Lineal fissions can dramatically increase variation among local groups and have been documented for both primates and humans (Ward and Neel, 1970; Neel and Ward, 1972; Fix, 1975; Duggleby, 1977; Cheverud et al., 1978; Chepko-Sade and Sade, 1979; Smouse et al., 1981; O'Rourke and Bach Enciso, 1982; Cheverud and Dow, 1985). The relative importance of these phenomena depends both on the frequency with which fissions occur and on the speed with which their effects are destroyed by migration among groups. After a fission, the genetic structure of the population will converge gradually toward an equilibrium between the forces of migration and genetic drift, and as it does so, the effect of the fission will be eliminated. If this happens rapidly and fissions are infrequent, the effect of one fission will disappear long before the next occurs. In this case, the effects of lineal fissions will be transitory and of little evolutionary significance. Conversely, if convergence is slow and fissions occur frequently, their effects accumulate and are of profound importance, as the authors cited above have shown. Thus, KSM is likely to be important only if the process converges relatively rapidly toward its equilibrium.

The work of Rogers and Harpending (1986) suggests that this may often be the case, at least in human populations. They

studied the rate at which genetic structure converges toward its equilibrium in a model of recurrent migration and genetic drift, and estimated the half-life of convergence in several human populations. Their estimates are undoubtedly somewhat larger than they should be since the effect of systematic pressure was ignored. Nonetheless, in five of the 10 populations studied, the half-life was between two and four generations, implying that the effects of a fission would be eliminated fairly rapidly. These results suggest that in many human populations the effect of fissions on population structure may be less important than that of movement of individuals among existing groups.

KSM has been less thoroughly studied than lineal fission. Fix (1975, 1978, 1979, 1982) pointed out the discrepancy between the theories used to predict the effects of migration and the structure of migration in real populations. He argued that KSM must increase the variance among groups at equilibrium and verified this claim using a computer simulation. However, in the absence of a population genetics model, it has not been possible to predict the consequences of particular instances of KSM. Furthermore, although KSM appears to be common in natural populations, there have been no surveys of its extent. Indeed, no survey has been possible because no one has suggested a way to measure the extent to which migration is kin-structured. In this paper, I introduce a model intended to answer some of these questions and to facilitate efforts to answer the others.

Model

The model developed here is an extension of the "migration matrix" model of population structure, introduced by Malécot (1951), Bodmer and Cavalli-Sforza (1968), and Smith (1969). The set of individuals that disperse from local group i to j will be referred to as the "ijth migrant set." Let

$$m_{ij} = n_{ij} \bigg/ \sum_i n_{ij}, \text{ the fraction of the } j\text{th}$$

local group comprising immigrants from group i;

$\mathbf{M} = [m_{ij}]$, the migration matrix;

$n_{ij} = $ the size of the ijth migrant set;

$$n_{i\cdot} = \sum_j n_{ij}, \text{ the size of group } i \text{ after population regulation but before migration;}$$

$$n_{\cdot j} = \sum_i n_{ij}, \text{ the size of group } j \text{ after migration;}$$

$$N = \sum_{ij} n_{ij}, \text{ the total population size;}$$

$w_j = n_{\cdot j}/N$, the relative size of the jth local group.

Let us assume that the genotype has no influence on the propensity of an individual to migrate, and that a gene starting in any local group may eventually reach any other, though not necessarily in one generation. The latter assumption is essential, for if some local groups were completely isolated from the others, there could be no equilibrium between the effects of migration and drift, and the theory developed here would not apply to the population as a whole. It would, of course, apply to each of the isolated subdivisions separately, so this assumption is not restrictive. I also assume that some of the individuals born in each local group stay there to breed. This seems reasonable and guarantees that the system will have no cyclical tendencies. I also make the less reasonable assumption that the sizes of migrant sets, n_{ij}, are constant from generation to generation. The main results discussed below also assume that $n_{ij} = n_{ji}$, for all i and j. These last two assumptions will seldom hold exactly, but should often be fair approximations. They are discussed further by Rogers and Harpending (1986). The analysis in the appendix also assumes that the effective sizes of local groups equal their actual sizes. This assumption is less restrictive than it sounds, as will be discussed below.

Some form of linear systematic pressure is assumed to operate, preventing allele frequencies from drifting to fixation. This systematic pressure could be interpreted as mutation or weak selection, but will be interpreted here as long range migration from a "continent" with unchanging allele frequency π.

For simplicity, I assume that the fraction of external immigrants is the same in each local group and that all external immigrants

are drawn at random from the same population. The effect of violating these assumptions is negligible when local genetic structure is dominated by the effects of local migration, but may be important when external migration is relatively strong. Thus, the model is most appropriate for populations that are relatively isolated from the outside world.

Population regulation is assumed to occur prior to migration, implying that the numbers of individuals migrating are small. Consequently, stochastic changes in allele frequencies (genetic drift) occur during migration as well as during population regulation. This assumption is unusual; migration is usually assumed to precede population regulation, which implies that its effect on allele frequencies is deterministic. The assumption adopted here seems more appropriate for humans and other species in which mortality is lowest during the ages of migration and reproduction (Coale, 1972). For further discussion of this issue see Rogers and Harpending (1986).

The object of this investigation is to predict similarities and differences among local groups. These can be measured by the normalized covariance of group allele frequencies

$$r_{ij} = \frac{(p_i - \bar{p})(p_j - \bar{p})}{\bar{p}(1 - \bar{p})},$$

where

$p_i =$ the frequency of allele A_1 in group i after both local and continental migration;

$\bar{p} = \sum_j w_j p_j$, the allele frequency in the population as a whole.

In the appendix, a formula is obtained for the equilibrium value of the matrix $\mathbf{R}$, whose ijth entry is r_{ij}.

A useful summary of the information contained in $\mathbf{R}$ is

$$r_0 = \frac{\sum w_i(p_i - \bar{p}_i)^2}{\bar{p}(1 - \bar{p})}.$$

The expectation of r_0 will be denoted by ρ, and is equivalent to one of the several meanings that have been attached to Wright's (1951) F_{ST}. Wright's notation is avoided here

because of the variety of ways in which it has been used in the literature (see Wood, 1986).

In the appendix it is shown that if continental migration is weak compared with local migration, then at equilibrium

$$E[\mathbf{R}] = \mathbf{V}^T \mathbf{B} \mathbf{V}, \quad (1)$$

where $\mathbf{V}$ is a matrix containing the left eigenvectors of $\mathbf{M}$, and $\mathbf{B}$ is a diagonal matrix with diagonal entries $\mathbf{B}_{11} = 0$ and

$$\mathbf{B}_{ii} \simeq \frac{1 - \rho}{2N}\left(\frac{1}{1 - \lambda_i^2} + \theta\right), \quad (2)$$
$$\text{for } i > 1.$$

Here, the λ_i are eigenvalues of the migration matrix, $\mathbf{M}$, indexed in descending order, and θ is defined by $z_i = -\theta/(n_i. - 1)$, where z_i is the correlation of individuals in distinct migrant sets from group i. The parameter θ is a measure of the extent to which migration is kin-structured. When migration is not kin-structured, $\theta = 0$, and (2) reduces to the formula derived by Rogers and Harpending (1986). Rogers and Harpending show that ρ is equal to the sum of the $\mathbf{B}_{ii}$, so

$$\frac{\rho}{1 - \rho} \simeq \frac{1}{2\bar{n}}\left(\frac{1}{g - 1}\sum_{i=2}^{g}\frac{1}{1 - \lambda_i^2} + \theta\right), \quad (3)$$

where g is the number of groups and $\bar{n} = N/(g-1)$.

Rogers and Harpending (1986) define the effective migration rate, m_e, by

$$\frac{1}{2m_e} = \frac{1}{g - 1}\sum_{i=2}^{g}\frac{1}{1 - \lambda_i^2}, \quad (4)$$

reducing (3) to the form

$$\frac{\rho}{1 - \rho} \simeq \frac{1}{4\bar{n}m_e} + \frac{\theta}{2\bar{n}}. \quad (5)$$

The value of θ can be estimated from genetic data or from information about the extent to which migration is kin-structured. Let us refer to the groups of relatives that migrate together as "families," whether they are families in the usual sense or not. In the appendix it is shown that, if families migrate independently, if all are the same size, γ, and if the genetic correlation within families is κ, then $\theta = (\gamma - 1)\kappa$.

These results can be extended to the case of varying family size provided that many families are exchanged by each pair of local groups. Let $c^{(y)}$ denote the correlation within families of size y, and let $\bar{y}$ and Var [y] denote the mean and variance of family size. As shown in the appendix, the results above continue to hold if γ and κ are redefined as

$$\gamma = \frac{E\{y^2\}}{\bar{y}} = \bar{y} + \frac{Var\{y\}}{\bar{y}}, \qquad (6)$$

and

$$\kappa = \frac{E\{y(y-1)c^{(y)}\}}{E\{y(y-1)\}} . \qquad (7)$$

DISCUSSION

Let $G = 2\rho/(1 - \rho)$. This parameter can be interpreted as the ratio of the variance expected between groups to that expected within groups. Equation (5) states that the effect of KSM on G is additive, increasing it by $\theta/\bar{n}$, provided that some migration occurs. If families (i.e., groups of migrating relatives) are all the same size, or if each migrant set is composed of many families, θ does not depend on the amount of mobility among groups. Thus, the effect of KSM on genetic population structure is nearly independent of mobility.

The magnitude of θ is not necessarily trivial. For example, if all migration involves pairs of sibs, $\theta \simeq (2 - 1)0.5 = 0.5$. The value κ would actually be less than 0.5, since we are concerned with correlations relative to a small population rather than the infinite, time-invariant population of classical theory (Wright, 1969; Malécot, 1969). Correlations relative to small populations are somewhat reduced, particularly for distant relatives. For example, in a closed population of effective size 50, the correlation between cousins is 0.0683 rather than 0.1250 (Rogers, 1986). For first and second degree relatives, this distinction can be ignored provided that effective group size is greater than about 500.

When family size varies, γ is given by (6) and will be larger than the mean family size, but smaller than the maximum family size. A few large families can greatly inflate the value of γ. For example, if 90% of families are of size 2 and 10% are of size 15, mean

family size is 3.3, whereas $\gamma = 7.9$. Thus, variation in family size can greatly increase the effect of KSM. Variation in family size also affects κ, which becomes an average of correlations within families weighted by a function of family size. Because of this weighting, κ is more strongly influenced by correlations within large families than within small ones. Any tendency for large families to have lower correlations would reduce the effect of KSM.

The absolute increase in G is probably of less general interest than the proportional increase, which is $2m_e\theta$. Since this is proportional to the effective migration rate, KSM should be most important in populations with high mobility. The effective migration rate ranges from zero, when there is no mobility among groups, to ½, when group of residence is statistically independent of group of origin. If, for example, migration involves groups of three sibs ($\theta = 1$) in a population with very high mobility ($2m_e = 1$), KSM would double G. In a population with less mobility, for example $2m_e = 0.1$, the same level of KSM would increase G by only 10%.

Although the model analyzed here assumes that effective and actual group sizes are equal, these conclusions do not depend on this assumption. Rogers and Jorde (1987) have studied the effects of factors that, like KSM, modify migration's effect on group differences, and they find that the proportional increase in G does not depend on effective population size. Since their model is general enough to encompass the one used here as a special case, the conclusions of the preceding paragraph should hold regardless of the relationship between effective and actual population size.

What if Sibships Disperse as Units?

The most extreme form of KSM that seems plausible is that in which entire sibships migrate together. Fix (1978) reports that this is not unusual among the Semai Senoi. The variance of sibship size in human populations is usually between 1.5 and 3 times the mean (Cavalli-Sforza and Bodmer, 1972). Therefore, if mean sibship size were four, γ would fall somewhere between 5.5 and 7. Since the correlation between full sibs is 0.5, θ would be between 2.75 and 3.5.

TABLE 1. The movement of male lions among prides. The second row shows the number of male groups of each size that succeeded in moving from one pride to another. These data are from Bygott et al. (1979).

No. of males in group	1	2	3	4	5	6	7
No. of successful groups	4	23	11	4	1	3	1

Rogers and Harpending (1986) have estimated m_e in 11 human populations, and their estimates range from 0.051 for Oxfordshire villages (Hiorns et al., 1977) to 0.288 for the !Kung San (Harpending and Jenkins, 1974). The latter estimate implies that $2m_e\theta$, the proportional increase in G due to migration of entire sibships, could be as large as 2.02. Thus, this form of KSM could triple G. These numbers would be larger still where the average size of sibships is greater than four.

Dispersal of Male Lions

Bygott et al. (1979) report the results of an extended study of the social behavior of lions, and some of their results are shown in Table 1. In lion prides, females generally stay in their natal pride but males leave, frequently in groups. Table 1 contains the frequency distribution of the sizes of groups of males that have been successful in migrating from one pride to another. The estimate of γ obtained from these data is

$$\hat{\gamma} = \frac{\Sigma y^2}{\Sigma y} = \frac{441}{129} = 3.42.$$

According to Bygott et al. (1979), the genetic correlation of individuals within families is 0.22, and I will take this as an estimate of κ.

To predict the effect of this level of KSM, we must also assume something about the pattern of migration among prides. Suppose that all females remain in their natal prides and that all males disperse at random. The female migration matrix is then the identity matrix ($\mathbf{I}$) and the male migration matrix is $\mathbf{1}w^T$. The overall migration matrix is the average of these, $\mathbf{M}^T = (\frac{1}{2})(\mathbf{I} + \mathbf{1}w^T)$. This matrix has one eigenvalue equal to unity and the rest are equal to $\frac{1}{2}$. Substituting this value into (4) yields $m_e = \frac{3}{8}$. The proportional increase in G due to KSM is therefore

$2m_e\theta = 0.4$. Thus, the ratio of between- to within-group variance may be inflated by a factor of 1.4 in lion prides.

Dispersal of Pine Voles

FitzGerald (1984, unpubl.) has studied dispersal of pine voles (*Microtus pinatorum*) among breeding territories. Of 32 episodes of dispersal, 29 involved single individuals, one involved a pair of full brothers, and two involved a father and son. With these data, $\hat{\gamma} = 1.1714$ and κ is 0.5, since that is the correlation both between full sibs and between fathers and sons.

Let us assume that a proportion $1 - m$ of the individuals breed in the territory in which they were born, and that a proportion m disperse at random so that the territory in which they breed is statistically independent of their birthplace. The migration matrix implied by this assumption is $\mathbf{M}^T = (1 - m)\mathbf{I} + m\mathbf{1}w^T$. This matrix has one eigenvalue of unity and all the others are equal to $(1 - m)$. Substituting this into Equation (4) produces $m_e = m - m^2/2$. Thus, the proportional increase in G is $2m_e(\gamma - 1)\kappa = (2m - m^2)(0.1714)(0.5) = (2m - m^2)(0.0857)$.

Only 9% of the marked animals in the population were successful in establishing themselves in a territory other than their natal territory. However, this is probably an underestimate of m since not all individuals that disperse are likely to be recaptured, and since the breeding success of those that disperse is higher than that of those that do not (FitzGerald, pers. comm.). If m were 0.09, the proportional increase in G would be 0.0147, a very small increase. Even if m were 0.9, the proportional increase in G would be only 0.0848. Thus, it is unlikely that KSM has a large effect on the genetic structure of this population of voles.

The assumption that large numbers of families are exchanged by each pair of groups is not satisfied by any of these examples, so the effect of KSM may have been exaggerated. All that can be said with confidence about these particular populations is that the effect of KSM is larger in some than in others, and that in some it is so small as to be negligible. The quantitative results should be interpreted as relating to hypothetical populations in which m_e is the same and

migration is kin-structured to the same extent, but the migrant sets and group sizes are large enough to satisfy the assumptions.

To assess the importance of KSM, it will be necessary to estimate the value of θ in a wide variety of natural populations. This can be done with behavioral data, as in the lion and vole examples, or with genetic data. KSM inflates the variance of the allele frequencies of migrant sets. When both birthplace and adult residence of each individual are known, these allele frequencies can be estimated and their variance used to estimate the extent to which migration is kin-structured (Rogers, unpubl.).

Conclusions

Kin-structured migration (KSM) occurs when individuals disperse in the company of relatives. The extent to which migration is kin-structured is measured by a parameter θ which is defined in terms of the correlation between individuals that migrate independently. KSM adds a term to G, the ratio of between- to within-group variance, that is proportional to θ and nearly independent of the level of mobility.

The proportional increase in G is $2m_e\theta$, where m_e is the effective migration rate (Rogers and Harpending, 1986). In populations with low mobility (small m_e), the increase in variance attributable to KSM will be small compared with the variance expected in its absence. Consequently, KSM should be most important in populations with high mobility among groups.

The parameter θ can be estimated with either genetic or behavioral data. If "families" (i.e., groups of migrating relatives) are all of size γ, and the correlation within families is κ, then $\theta = (\gamma - 1)\kappa$. This result generalizes to the case of varying family size provided that many families are exchanged by each pair of groups. In that case, the effect of KSM is sensitive to the variance in family size as well as to its mean.

An extreme form of kin-structured migration is that in which entire sibships migrate as units. This form of migration would inflate G by a factor of three in some human populations, assuming that the average sibship size is four. Thus, the effect of KSM is potentially large. Bygott et al. (1979) showed that dispersal among lion prides is kin-structured, and analysis of their data indicates that G may be inflated by kin structure by a factor of 1.4. On the other hand, the kin structure apparent in FitzGerald's (1984, unpubl.) data on pine voles would increase G by no more than a few percent. These examples indicate that there are appreciable differences between populations in the importance of KSM, and that its effect may be substantial in some.

Literature Cited

BODMER, W. F., AND L. L. CAVALLI-SFORZA. 1968. A migration matrix model for the study of random genetic drift. Genetics 59:565–592.

BYGOTT, G. D., B. C. R. BERTRAM, AND J. P. HANBY. 1979. Male lions in large coalitions gain reproductive advantages. Nature 282:839–841.

CARMELLI, D., AND L. L. CAVALLI-SFORZA. 1976. Some models of population structure and evolution. Theoret. Popul. Biol. 9:329–359.

CAVALLI-SFORZA, L. L., AND W. F. BODMER. 1972. The Genetics of Human Populations. Freeman, San Francisco, CA.

CHENEY, D. L. 1983a. Intergroup encounters among Old World monkeys, pp. 233–241. In R. A. Hinde (ed.), Primate Social Relationships: An Integrated Approach. Sinauer, Sunderland, MA.

———. 1983b. Proximate and ultimate factors related to the distribution of male migration, pp. 241–249. In R. A. Hinde (ed.), Primate Social Relationships: An Integrated Approach. Sinauer, Sunderland, MA.

CHEPKO-SADE, D., AND D. S. SADE. 1979. Patterns of group splitting within matrilineal kinship groups: A study of social group structure in Macaca mulatta (Cercopithecidae: Primates). Behav. Ecol. Sociobiol. 5:67–86.

CHEVERUD, J., J. BUETTNER-JANUSCH, AND D. SADE. 1978. Social group fission and the origin of intergroup genetic differentiation among the rhesus monkeys of Cayo Santiago. Amer. J. Phys. Anthropol. 49:449–456.

CHEVERUD, J. M., AND M. M. DOW. 1985. An autocorrelation analysis of genetic variation due to lineal fission in social groups of Rhesus Macaques. Amer. J. Phys. Anthropol. 67:113–122.

COALE, A. 1972. The growth and structure of human populations: A mathematical investigation. Princeton Univ. Press, Princeton, NJ.

COLVIN, J. 1983. Influences of the social situation on male emigration, pp. 160–171. In R. A. Hinde (ed.), Primate Social Relationships: An Integrated Approach. Sinauer, Sunderland, MA.

DUGGLEBY, C. 1977. Blood group antigens and the population genetics of *Macaca mulatta* adults on Cayo Santiago. II. Effects of social group division. Yearbook Phys. Anthropol. 20:263–271.

FITZGERALD, R. W. 1984. Population ecology and social biology of a free-ranging population of pine voles, *Microtus pinatorum*. Ph.D. Diss., SUNY Binghamton, Binghamton, NY.

FIX, A. G. 1975. Fission-fusion and lineal effect: Aspects of the population structure of the Semai Senoi of Malaysia. Amer. J. Phys. Anthropol. 43:295–302.

———. 1978. The role of kin-structured migration in genetic microdifferentiation. Ann. Hum. Genet. 41:329–339.

———. 1979. Anthropological genetics of small populations. Ann. Rev. Anthropol. 8:207–230.

———. 1982. Genetic structure of the Semai, pp. 179–204. *In* M. Crawford and J. Mielke (eds.), Current Developments in Anthropology. Plenum, N.Y.

FURUYA, Y. 1968. On the fission of troops of Japanese monkeys. Primates 9:323–349.

———. 1969. On the fission of troops of Japanese monkeys (Part II). Primates 10:47–60.

HARPENDING, H. C., AND T. JENKINS. 1974. !Kung population structure, pp. 137–165. *In* J. F. Crow and C. Denniston (eds.), Genetic Distance. Plenum, N.Y.

HARPENDING, H. C., AND R. H. WARD. 1982. Chemical systematics and human populations, pp. 213–256. *In* M. H. Nitecki (ed.), Biochemical Aspects of Evolutionary Biology. Univ. Chicago Press, Chicago, IL.

HIORNS, R. W., G. A. HARRISON, AND J. B. GIBSON. 1977. Genetic variation in some Oxfordshire villages. Ann. Hum. Biol. 4:197–210.

LEE, P. C. 1983. Home range, territory and intergroup encounters, pp. 231–233. *In* R. A. Hinde (ed.), Primate Social Relationships: An Integrated Approach. Sinauer, Sunderland, MA.

MALÉCOT, G. 1951. Un traitement stochastique des problèmes linéaires en génétique de population. Ann. Univ. Lyon Sect. A 14:79–118.

———. 1969. The Mathematics of Heredity. Freeman, San Francisco, CA.

———. 1973. Isolation by distance, pp. 72–75. *In* N. E. Morton (ed.), Genetic Structure of Populations. Univ. Press of Hawaii, Honolulu.

MEIKLE, D. B. AND S. H. VESSY. 1981. Nepotism among rhesus monkey brothers. Nature 294:106–161.

NASH, L. T. 1976. Troop fission in free-ranging baboons in the Gombe Stream National Park, Tanzania. Amer. J. Phys. Anthropol. 48:63–77.

NEEL, J. V., AND F. SALZANO. 1967. Further studies on the Xavante Indians. X. Some hypotheses-generalizations resulting from these studies. Amer. J. Hum. Genet. 19:554–574.

NEEL, J. V., AND R. H. WARD. 1972. Genetic structure of a tribal population, the Yanomama Indians. VI. Analysis by F-statistics, including a comparison with the Makiritare and Xavante. Genetics 72:639–666.

O'ROURKE, D., AND V. BACH ENCISO. 1982. Primate social organization, ecology, and genetic variation, pp. 1–28. *In* M. Crawford and J. Mielke (eds.), Current Developments in Anthropological Genetics, Vol. 2. Ecology and Population Structure. Plenum, N.Y.

ROGERS, A. 1986. Correlations between relatives in small populations. Amer. J. Phys. Anthropol. 71: 377–380.

ROGERS, A. R., AND H. C. HARPENDING. 1986. Migration and genetic drift in human populations. Evolution 40:1312–1327.

ROGERS, A. R., L. B. JORDE. 1987. The effect of nonrandom migration on genetic differences between populations. Ann. Hum. Genet. 51. *In press.*

SMITH, C. A. B. 1969. Local fluctuations in gene frequencies. Ann. Hum. Genet. 32:251–260.

SMOUSE, P. E., V. J. VITZTHUM, AND J. V. NEEL. 1981. The impact of random and lineal fission of the genetic divergence of small human groups: A case study among the Yanomama. Genetics 98:179–197.

SOUTHWICK, C. H., M. A. BEG, AND M. R. SEDDIQI. 1965. Rhesus monkeys in northern India, pp. 111–159. *In* I. DeVore (ed.), Primate Behavior. Holt, Rinehart & Winston, N.Y.

WAHLUND, S. 1928. Zusammensetzung von populationen und korrelationserscheinungen vom standpunkt der vererbungslehre aus betrachtet. Hereditas 11:65–106. (English transl. 1975. *In* K. M. Weiss and P. A. Ballanoff [eds.], Demographic Genetics, pp. 224–263. Wiley, N.Y.)

WARD, R. H., AND J. V. NEEL. 1970. Gene frequencies and microdifferentiation among the Makiritare Indians. IV. A comparison of a genetic network with ethnohistory and migration matrices; a new index of genetic isolation. Amer. J. Hum. Genet. 22:538–561.

WOOD, J. 1986. Convergence of genetic distances in a migration matrix model. Amer. J. Phys. Anthropol. 71:209–220.

WRIGHT, S. 1951. The genetical structure of populations. Ann. Eugen. 15:323–354.

———. 1969. Evolution and the genetics of populations. II. The theory of gene frequencies. Univ. Chicago Press, Chicago, IL.

Corresponding Editor: M. K. Uyenoyama

APPENDIX

This appendix extends the model of Rogers and Harpending (1986) to incorporate the effects of kin-structured migration (KSM). The point of departure is equation (8), which was derived by Rogers and Harpending under the assumptions discussed above, and involves no assumptions about kin structure.

The Matrices R and L

Rogers and Harpending (1986) show that, at equilibrium, the expectation of **R** is approximately

$$E\{\mathbf{R}\} = \sum_{i=0}^{\infty} (1 - s)^{2i}(\mathbf{L}^i)^{\mathrm{T}}\mathbf{CL}^i, \qquad (8)$$

which is analogous to the similar formulas derived for slightly different models by Malécot (1973), Carmelli and Cavalli-Sforza (1976), and Smith (1969). In Rogers

and Harpending's formulation, the matrix $\mathbf{C}$ contains normalized variances and covariances of the changes in allele frequency produced by genetic drift in the various groups, and $\mathbf{L}$ is the "reduced migration matrix," defined by $\mathbf{L}^t = \mathbf{M}^t(\mathbf{I} - w\mathbf{1}^T)$, where $\mathbf{M}$ is the backwards migration matrix, w is a column vector of relative population sizes (w_i), $\mathbf{I}$ is the identity matrix, and $\mathbf{1}$ is a column vector, each element of which is equal to unity. Since $\mathbf{M}^0 = \mathbf{I}$, the zero'th power of $\mathbf{L}$ is $\mathbf{L}^0 = \mathbf{I} - w\mathbf{1}^T$.

The matrix $\mathbf{L}$ depends on the numbers of individuals exchanged by each pair of groups, but does not depend on their relatedness, and is therefore unaffected by kin structure. Thus, the influence of kin-structure on $\mathbf{R}$ derives solely from its effect on $\mathbf{C}$. To investigate the effect of KSM, we first study its effect on $\mathbf{C}$, the matrix of normalized variances and covariances of the effects of a single generation of genetic drift.

Some Definitions

For a set of g local groups, let

$p_j^{(t)}$ = the frequency of allele A_1 in group j in generation t after both local and continental migration;

$\epsilon_j^{(t)}$ = $p_j^{(t)} - \mathrm{E}\{p_j^{(t)}|p_k^{(t-1)}: k = 1, 2, \ldots, g\}$, the effect of genetic drift on the allele frequency of group j during generation t;

$\mathbf{C}^*$ = the expectation of $\epsilon\epsilon^T$, that is, the matrix of incremental variances and covariances produced by genetic drift;

$\mathbf{C}$ = the expectation of $\epsilon\epsilon^T/\bar{p}(1 - \bar{p})$, a matrix of normalized variances and covariances;

q_{ij} = the frequency of allele A_1 in the ijth migrant set;

$q_{i\cdot}$ = $n_{i\cdot}^{-1} \sum_j n_{ij} q_{ij}$, the frequency of A_1 in the ith local group after population regulation but before migration;

$q_{\cdot j}$ = $\sum_i m_{ij} q_{ij}$, the frequency of A_1 in the jth local group among adults after local but not continental migration;

s = the fraction of each group exchanged with the continent each generation;

π = the continental allele frequency, which is assumed constant;

u_j = the frequency of allele A_1 among continental migrants to group j.

Where necessary, the generations to which these quantities refer will be indicated with superscripts as in $p_j^{(t)}$.

The Form of ϵ_j

Our definitions and assumptions imply that

$$p_j^{(t+1)} = (1 - s) \sum_i m_{ij} q_{ij}^{(t+1)} + s u_j^{(t+1)}. \quad (9)$$

If selection is absent and if migratory propensities are independent of genotype, then the conditional expectation, given $p_i^{(t)}$, of $q_{ij}^{(t+1)}$ is $p_i^{(t)}$, and the conditional expectation of u_j is π. Consequently the difference between $p_j^{(t+1)}$ and its conditional expectation is

$$\epsilon_j^{(t+1)} = (1 - s) \sum_i m_{ij}(q_{ij}^{(t+1)} - p_i^{(t)})$$
$$+ s(u_j^{(t+1)} - \pi). \quad (10)$$

Clearly, the moments of ϵ depend on those of q_{ij}.

Genotypic Values, their Means and Variances

The allele frequency of the ijth migrant set can be written as

$$q_{ij} = \frac{1}{n_{ij}} \sum_k x_{ijk}, \quad (11)$$

where x_{ijk} is the "genotypic value" of the kth individual in the ijth migrant set, and is equal to 1, ½, or 0 for genotypes A_1A_1, A_1A_2, and A_2A_2, respectively. So far, we have expressed the equilibrium value of $\mathbf{R}$ in terms of $\mathbf{C}$, $\mathbf{C}$ in terms of ϵ, ϵ in terms of q_{ij}, and q_{ij} in terms of x_{ijk}. We can now proceed in the opposite direction, first finding the moments of the x_{ijk}, then using those to find the moments of q_{ij}, and eventually obtaining an expression for the expectation of $\mathbf{R}$.

The conditional expectation of $x_{ijk}^{(t+1)}$, given the array of allele frequencies of adults in the previous generation, is $p_i^{(t)}$, and if mating within groups is at random its conditional variance is

$$v_i^{(t+1)} = \mathrm{Var}\{x_{ijk}^{(t+1)}\} = p_i^{(t)}(1 - p_i^{(t)})/2. \quad (12)$$

Correlations between Genotypic Values

To the extent that migration is kin-structured, correlations of the genotypic values of individuals in the same migrant set will exceed those between individuals in different migrant sets. Correlations measure similarity relative to some reference population, and the algebra below will be simpler if correlations are taken as relative to the local group in the previous generation. The correlation between the genotypic values of two individuals born in group i in generation $t+1$ is defined by

$$\mathrm{corr}\{x_{ijk}^{(t+1)}x_{ij'k'}^{(t+1)}\} = \frac{\mathrm{E}\{x_{ijk}x_{ij'k'}\} - (p_i^{(t)})^2}{v_i^{(t+1)}}.$$

The groups of relatives that migrate together will be referred to as "families," though they need not be families in the usual sense, and I assume that families migrate independently. Let

z_i = the correlation of "non-relatives," i.e., individuals in different families;

h_{ij} = the correlation of distinct individuals in the ijth migrant set.

Since families migrate independently, the correlation of non-relatives is the same whether they are in the same migrant set or not. Consequently, z_i is also the correlation between individuals in different migrant sets. For simplicity, I assume that effective and actual population sizes are equal, which implies that, in the absence of KSM, z_i would be zero. The effect of KSM will be measured by a parameter, θ, defined by

$$z_i = \frac{-\theta}{n_{i\cdot} - 1} \quad (13)$$

The connection between θ, family size, and the cor-

relations within families will be discussed later. The two correlations just defined are related by

$$h_{ij} = -z_i \frac{n_{i\cdot} - n_{ij}}{n_{ij} - 1}, \qquad \text{if } n_{ij} > 1, \qquad (14)$$

and I define $h_{ij} = 0$, if $n_{ij} < 1$.

Proof.—Since effective and actual population sizes are equal, the correlation between distinct random individuals in the ijth migrant set is zero. On the other hand, we can also write this correlation as the weighted average of correlations of individuals within migrant sets and those in different migrant sets. Thus, if $n_{ij} > 1$,

$$0 = \frac{n_{ij} - 1}{n_{i\cdot} - 1} h_{ij} + \frac{n_{i\cdot} - n_{ij}}{n_{i\cdot} - 1} z_i,$$

which is satisfied if and only if h_{ij} is as given in (14), as claimed.

The Moments of q_{ij}

In view of (11), if $n_{ij} > 1$, the variance of q_{ij} is given by

$$\text{Var}\{q_{ij}\} = \frac{1}{n_{ij}^2} \sum_{kl} \text{Cov}\{x_{ijk}, x_{ijl}\}$$

$$= \frac{v_i}{n_{ii}} [1 + (n_{ij} - 1) h_{ij}], \qquad (15)$$

and, for $j \neq k$, and $n_{ij} > 0$,

$$\text{Cov}\{q_{ij}, q_{ik}\} = \frac{1}{n_{ij} n_{ik}} \sum_{lm} \text{Cov}\{x_{ijl}, x_{ikm}\}$$

$$= v_i z_i, \qquad (16)$$

since each of the terms in the summation is equal to $v_i z_i$.

The Moments of $q_{\cdot j}$

The section will demonstrate that if the allele frequencies of the various groups do not vary too greatly, and if the system is close to its equilibrium, then the conditional variances and covariances of $q_{\cdot j}$ and $q_{\cdot k}$, given parental allele frequencies, are approximately

$$\text{Var}\{q_{\cdot j}\} \simeq \frac{(1 - \rho)\bar{p}(1 - \bar{p})}{2 n_{\cdot j}}$$

$$\cdot \left(1 + \theta \sum_i m_{ij} \frac{n_{i\cdot} - n_{ij}}{n_{i\cdot} - 1}\right), \qquad (17)$$

and

$$\text{Cov}\{q_{\cdot j}, q_{\cdot k}\} \simeq -(1 - \rho)\theta \bar{p}(1 - \bar{p})$$

$$\cdot \left(\sum_i \frac{m_{ij} m_{ik}}{2(n_{i\cdot} - 1)}\right). \qquad (18)$$

Proof.—The variance of $q_{\cdot j}$ is

$$\text{Var}\{q_{\cdot j}^{(t+1)}\} = \sum_i m_{ij}^2 \text{Var}\{q_{ij}^{(t+1)}\}$$

$$= \frac{1}{n_{\cdot j}} \sum_i p_i^{(t)}(1 - p_i^{(t)}) m_{ij}$$

$$\cdot [1 + (n_{ij} - 1) h_{ij}], \qquad (19)$$

since $m_{ij} = n_{ij}/n_{\cdot j}$. If the allele frequencies of the various groups do not differ too greatly, this equation can be approximated closely by replacing each of the $p_i(1 -$

$p_i)$ by their weighted average, which is given by an identity due to Wahlund (1928): $\bar{p}(1 - \bar{p})(1 - r_0) = \sum_i w_i p_i (1 - p_i)$. We are interested in the equilibrium, at which $r_0 = \rho$. Equation (17) is obtained by making these substitutions and also substituting Equation (14).

For $j \neq k$,

$$\text{Cov}\{q_{\cdot j}^{(t+1)}, q_{\cdot k}^{(t+1)}\} = \sum_i m_{ij} m_{ik} \text{Cov}\{q_{ij}^{(t+1)}, q_{ik}^{(t+1)}\},$$

$$= \frac{1}{2} \sum_i m_{ij} m_{ik} p_i^{(t)}(1 - p_i^{(t)}) z_i.$$

Equation (18) is obtained by substituting the formula for z_i and using the approximation discussed above.

The Form of C

This section demonstrates that, if continental migration is weak, family size is small compared to the sizes of local populations, and $s \ll \rho$, then

$$\mathbf{C} \simeq \frac{(1 - \rho)(1 + \theta)}{2N} \mathbf{W}^{-1} - \frac{(1 - \rho)\theta}{2N} \mathbf{M}^\mathsf{T} \mathbf{W}^{-1} \mathbf{M}, \qquad (20)$$

where $\mathbf{W}$ is a diagonal matrix of relative group sizes, $\mathbf{W}_{ii} = w_i = n_i/N$.

Proof.—Equation (10) implies that

$$\mathbf{C}_{jj}^* = (1 - s)^2 \text{Var}\{q_{\cdot j}\} + s^2 E\{(u_j - \pi)^2\},$$

where u_j is the allele frequency of continental migrants to group j. The effect of kin structure on continental migration is ignored so that $E\{(u_j - \pi)^2\} = \pi(1 - \pi)/(2 s n_{\cdot j})$. Provided that s is small, this term affects the result only slightly so this assumption is not too restrictive. For the same reason we can also substitute $s\bar{p}(1 - \bar{p})$ for $s\pi(1 - \pi)$. Thus,

$$\mathbf{C}_{jj}^* \simeq (1 - s)^2 \text{Var}\{q_{\cdot j}\} + \frac{s\bar{p}(1 - \bar{p})}{2 n_{\cdot j}}. \qquad (21)$$

Assuming u_j and u_k to be independent, we also have

$$\mathbf{C}_{jk}^* = (1 - s)^2 \text{Cov}\{q_{\cdot j}, q_{\cdot k}\}. \qquad (22)$$

$\mathbf{C}$ is obtained from $\mathbf{C}^*$ as follows. By definition,

$$\mathbf{C} = E\{\epsilon \epsilon^\mathsf{T} / \bar{p}(1 - p)\},$$

Conditioning this expectation on $\bar{p}$ produces

$$\mathbf{C} = E\{E\{\epsilon \epsilon^\mathsf{T} \mid \bar{p}\}/\bar{p}(1 - \bar{p})\} = E\{\mathbf{C}^*/\bar{p}(1 - \bar{p})\}.$$

But as equations (17) through (22) indicate, $\mathbf{C}^*$ is proportional to $\bar{p}(1 - \bar{p})$, so $\mathbf{C}^*/\bar{p}(1 - \bar{p})$ is a constant. Thus,

$$\mathbf{C} = \mathbf{C}^*/\bar{p}(1 - \bar{p}).$$

We have

$$\mathbf{C}_{jj} = \frac{(1 - s)^2(1 - \rho) + s}{2 n_{\cdot j}} + \left(\frac{(1 - s)^2(1 - \rho)\theta}{2 n_{\cdot j}}\right)$$

$$\cdot \left(\sum_i m_{ij} \frac{n_{i\cdot} - n_{ij}}{n_{i\cdot} - 1}\right),$$

$$\mathbf{C}_{jk} = -(1 - s)^2(1 - \rho)\theta \sum_i \frac{m_{ij} m_{ik}}{2(n_{i\cdot} - 1)}.$$

If migration is nil, both summations above are zero, since $m_{ij} = 0$ and $n_{ii} = n_{i\cdot}$ in that case. $\mathbf{C}$ is then in-

dependent of θ. Otherwise if $1 \ll n_{i\cdot}$ and $s \ll \rho$ these equations are equivalent to (20), as claimed.

The Equilibrium Dispersion Matrix

The value of $E\{\mathbf{R}\}$ at equilibrium can be predicted by beginning with an initial guess about its value and iterating (8) in a computer program until the result ceases to change. With suitable assumptions, however, it is also possible to derive an analytical expression. I assume that for each j and k, the number of migrants from group j to group k equals that from k to j. Harpending and Ward (1982) and Rogers and Harpending (1986) show, using this assumption, that

$$\sum_{i=0}^{\infty} (1 - s)^{2i}(\mathbf{L}^i)^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{L}^i = \mathbf{V}^{\mathrm{T}}\mathbf{D}\mathbf{V},$$

where $\mathbf{V}$ is a matrix containing the left eigenvectors of $\mathbf{M}$ (and also of $\mathbf{L}$), and $\mathbf{D}$ is a diagonal matrix with zero in the first diagonal position and, in the ith diagonal position, $\mathbf{D}_{ii} = [1 - (1 - s)^2\lambda_i^2]^{-1}$.

As $t \to \infty$, (8) becomes

$$E\{\mathbf{R}\} = \mathbf{V}^{\mathrm{T}}\mathbf{B}\mathbf{V}, \qquad (23)$$

where $\mathbf{B}$ is diagonal with $\mathbf{B}_{11} = 0$, and, for $i \neq 1$ and $s \ll \rho$,

$$\mathbf{B}_{ii} \simeq \frac{1 - \rho}{2N} \frac{1}{1 - \lambda_i^2}(1 + \theta(1 - \lambda_i^2)),$$

$$\simeq \frac{1 - \rho}{2N}\left(\frac{1}{1 - \lambda_i^2} + \theta\right). \qquad (24)$$

Here, λ_i is the ith eigenvalue of the migration matrix, $\mathbf{M}$. When migration is not kin-structured, $\theta = 0$, and (2) reduces to the formula derived by Rogers and Harpending (1986). It can be shown that ρ is equal to the sum of the $\mathbf{B}_{ii}$ (see Rogers and Harpending, 1986). Hence,

$$\frac{\rho}{1 - \rho} \simeq \frac{1}{2n}\left(\frac{1}{g - 1} \sum_{i=2}^{g} \frac{1}{1 - \lambda_i^2} + \theta\right), \qquad (25)$$

where g is the number of local groups, and $\bar{n} = N/(g - 1)$ is a measure of the effective group size when variances are taken about the current population mean, $\bar{p}$, rather than about π, the continental gene frequency (Rogers and Harpending, 1986).

Kin-structured Migration

Before this theory can be used, θ must be expressed in terms of quantities that can be observed, such as the sizes of families and the genetic correlations of individuals within them. If all families are of size γ, and the genetic correlation of individuals within families is κ, then

$$\theta \simeq (\gamma - 1)\kappa. \qquad (26)$$

Proof.—Within the ijth migrant set, each individual has $\gamma - 1$ family members and $n_{ij} - \gamma$ non-relatives. The average correlation between distinct individuals in this migrant set is therefore

$$h_{ij} = \frac{\gamma - 1}{n_{ij} - 1} \kappa + \frac{n_{ij} - \gamma}{n_{ij} - 1}z_i.$$

Setting this equal to equation (14) gives

$$(\gamma - 1)\kappa = -z_i\frac{n_{i\cdot} - \gamma}{n_{ij} - 1} = \theta\frac{n_{i\cdot} - \gamma}{n_{i\cdot} - 1}, \qquad (27)$$

which is approximately θ, as claimed.

The Case of Varying Family Size

Let $f^{(y)}$ denote the frequency of families of size y, and $c^{(y)}$ the correlation of individuals within such families. The results derived above also hold for the case of varying family size provided that many families are exchanged by each pair of local groups and the parameters γ and κ are redefined as

$$\gamma = \frac{E\{y^2\}}{\bar{y}} = \bar{y} + \frac{\mathrm{Var}\{y\}}{\bar{y}}, \qquad (28)$$

and

$$\kappa = \frac{E\{y(y - 1)c^{(y)}\}}{E\{y(y - 1)\}}, \qquad (29)$$

where $\bar{y}$ is the expectation of family size, and $\mathrm{Var}\{y\}$ is its variance.

Proof.—The variance of $q_{i\cdot}$ is

$$\mathrm{Var}\{q_{i\cdot}\} = \frac{1}{n_{i\cdot}^2} \sum_{jk} \sum_{j'k'} \mathrm{Cov}\{x_{ijk}, x_{ij'k'}\} \qquad (30)$$

$$= \frac{v_i}{n_{i\cdot}^2}(n_{i\cdot} + A + B), \qquad (31)$$

where A is the sum of correlations between "relatives," i.e., individuals of the same family, and B refers to individuals of different families. Since each migrant set comprises many families, so too does each local group and the relative frequency of families of size y within migrant sets or local groups is very close to its probability, $f^{(y)}$. Within group i prior to migration there are $n_{i\cdot}/\bar{y}$ families, of which a fraction $f^{(y)}$ are of size y, and each individual in these families has $y - 1$ relatives. Thus, the number of pairs of relatives from families of size y is $f^{(y)}y(y - 1)n_{i\cdot}/\bar{y}$. Using (28) and (29),

$$A = \frac{n_{i\cdot}}{\bar{y}} \sum_y f^{(y)}y(y - 1)c^{(y)}$$

$$= n_{i\cdot}(\gamma - 1)\kappa. \qquad (32)$$

The sum in (30) contains $n_{i\cdot}^2$ terms, of which $n_{i\cdot}$ refer to individuals taken singly, and $E\{y(y - 1)\}n_{i\cdot}/\bar{y}$ refer to distinct individuals in the same family. Therefore,

$$\mathbf{B} = n_{i\cdot}(n_{i\cdot} - 1 - E\{y(y - 1)\}/\bar{y})z_i = n_{i\cdot}(n_{i\cdot} - \gamma)z_i.$$

On the other hand, our assumption about effective population size implies that $\mathrm{Var}\{q_{i\cdot}\} = v_i/n_{i\cdot}$, and this implies that $A = -B$. Setting these quantities equal and solving for z_i produces equation (27), just as in the case of fixed family size.