Pingelap and Mokil Atolls: Migration

N. E. MORTON,¹ D. E. HARRIS,¹ SHIRLEY YEE,¹ AND RUTH LEW¹

The central problem in the study of population structure is to define a quantitative measure of kinship which is applicable to nearly all the data which might interest a geneticist, including migration, genealogy, anthropometrics, polymorphisms, and isonymy. This problem was solved by definition of the coefficient of kinship [1, 2] and its extension to various types of data [3, 4].

Next in importance is the problem posed by Lewontin [5, p. 59]: "It is unavoidable that effective population sizes and migration rates will be very difficult to estimate except for rare cases of completely isolated populations living in welldefined areas. For more continuous distribution, it may be that N and m are simply inappropriate parameters and that some other way of describing the breeding structure is preferable." Fortunately, this difficulty has now been overcome by redefinition of effective migration rate and effective population size so that they describe the increase of kinship from generation to generation.

For clarity we must distinguish the related quantities m and m_e . The long-range migration rate m does not depend on the pattern of short-range migration, while the effective migration rate m_e depends on both long- and short-range migration. In principle, m (and therefore m_e) includes the linearized effects of selection and mutation, but in practice the migrational component predominates. In this paper we shall assume that mutation and selection are negligible relative to migration, since the estimates of m_e are larger than all mutation rates and most selection coefficients.

We must also distinguish between N , the local effective population number determined by census size and fertility, and N_e , the effective number dependent on migration among localities. Thus N and m are taken as known quantities, while N_e and m_e must be calculated from the migration pattern, genealogies, or bioassay, using either a forward or backward solution.

In the forward solution, the migration matrix $[6, 7, 8]$ leads to predictions of kinship $\phi_{ii}(t)$ between gametes drawn from populations i and j in generation t, with only the values for $t = \infty$ being used to calculate m_e . In the backward solution, predictions or estimates of $\phi_{ij}(t)$ in successive generations are used to reduce the migration pattern and the vector of population sizes to a single value of m_e and N_e for each pair (i,j) , following the suggestion of Wright [9] that m_e be

Received August 6, 1970; revised December 1, 1970.

PGL paper no. 57. This work was supported by grant GM ¹⁷¹⁷³ from the U.S. National Institutes of Health.

¹ Population Genetics Laboratory, University of Hawaii, Honolulu, Hawaii 96822. o ¹⁹⁷¹ by the American Society of Human Genetics. All rights reserved.

MORTON ET AL.

so defined as to characterize the rate of increase of kinship from generation to generation. His proposal was restricted to gene frequencies in a special case. The same approach can be applied to genealogical data, or more generally to any set of estimates of the coefficient of kinship in successive generations (or half-generations). The island model (on which $m = m_e$ and $N = N_e$) gives for the coefficient of kinship the recurrence [10]

$$
\phi(t) = (1 - m)^2 \left[\frac{1}{2N} + \left(1 - \frac{1}{2N} \right) \phi(t - 1) \right], t = 1, 2, ...
$$

which may be written as

$$
\phi(t) - \phi(t-1) = \frac{(1-m)^2}{2N} + \phi(t-1) \left[(1-m)^2 \left(1 - \frac{1}{2N} \right) - 1 \right].
$$

Beginning with $\phi(0) = 0$, the successive values are

$$
\phi(1) - \phi(0) = \phi(1) = \frac{(1 - m)^2}{2N};
$$

$$
\phi(2) - \phi(1) = (1 - m)^4 \frac{(1 - 1/2N)}{2N};
$$

and, in general,

$$
\phi(t) - \phi(t-1) = (1-m)^{2t} \frac{(1-1/2N)^{t-1}}{2N} = \frac{1}{2N} e^{-(2m+1/2N)t}.
$$

In its integrated form,

$$
\phi(t) \doteq \frac{1}{2N} \int_0^t e^{-(2m+1/2N)t} dt = \Phi(1 - e^{-t/2N\Phi}), \qquad (1)
$$

where $\Phi = \phi(\infty) = 1/(4Nm + 1)$. This equation, which also holds with sexes separate, was almost derived by Malécot [2, p. 35] but he inadvertently omitted the factor Φ in the second term of $\phi(t)$.

Later he considered the mean number of chains between two individuals through a common ancestor t generations back [11]. We denote this mean by M(t). The kinship associated with each such chain is $(1/2)^{2t+1}$. Therefore $(1/2)^{2t+1}M(t) = \phi(t) - \phi(t-1)$, where $M(t) = (4^t/2N)(1-m)^{2t}(1-m)$ $1/2N$ ^{$t-1$}. In this island model, the migration rate plays a more important role and the population size a less important role than Malecot found for isolation by distance in the discontinuous and continuous cases.

Replacing m, N by m_e, N_e to indicate reduction of a more complex migrational pattern to an island model, we carry out the analysis as follows. Let the input be a set of records bearing t and the corresponding estimate of $\phi(t)$, together with a weight $W(t)$ if it is desired to weight estimates by the number of observations.

Trial values of the basic parameters N_e and m_e are provided, either or both of which may be estimated iteratively by weighted least squares (see Appendix). The output consists of these estimates of N_e and m_e , together with the derived quantities Φ and $b = 1/2N_e \Phi = 2m_e + 1/2N_e$, with their standard errors.

ISOLATION BY DISTANCE

This method treats each of the reference populations (Pingelap and Mokil) separately. A mating is considered to belong to the population if either or both parents are from a sibship which belonged to the population in the preceding generation. This requires ascertainment through residents of Pingelap or Mokil and at least one chain of kinship back to a survivor of Typhoon Lengkieki which devastated the atolls around 1775 [12]. Matings between Pingelapese and Mokilese are allocated to the population of residence, or to the population of the father if resident in Kolonia on Ponape. The population of birth should be understood as a statement of parentage. Thus a child born on Kusaie of two Pingelapese parents would be considered Pingelapese, but hybridization more remote than F_1 is ignored.

Data on migration are given in table 1. Fathers are more commonly migrant than mothers, and Mokil has experienced more migration than Pingelap. Short-

POPULATION OF BIRTH	PINGELAP RESIDENTS				MOKIL RESIDENTS			
	Fathers	Mothers	Children of Fathers	Children of Mothers	Fathers	Mothers	Children of Fathers	Children of Mothers
Pingelap .	707	726	2,105	2,123	6	9	15	38
Mokil .	9	13	16	32	261	283	859	868
Gilberts .	2	4	10	13	4	6	18	13
Philippines	1	Ω	1	0	2	o	3	0
Kusaie .	\overline{c}	0	10	0	0		0	ı
Japan .	7	0	7	0	2	o	2	0
Mortlocks	6	1	15	1	\overline{c}	2	\overline{c}	3
Causasian .	4	0	5	0	12	Ω	18	0
Guam .	\overline{c}	0	4	0	o	Ω	0	0
New Guinea	\overline{c}	0	\overline{c}	0	0	0	Ω	0
Ponape .	3	5	3	13	6	2	14	3
Nukuoru \sim	\overline{c}	o	\overline{c}	0		Ω	1	Ω
Marshalls .	0	0	0	0	8	4	11	12
Tonga .	0	0	0	0		o	1	0
Ngatik .	2		2			2		3
Truk	0	0	0	0		Ω		0
Palau	0	o	Ω	o	2	Ω	\overline{c}	Ω
Marshallese-								
Caucasian $\overline{1}$	1	o	1	0		1	1	8
Nukuoru-Mokil	1	0	1	0	o	0	0	0
Kusaie-Ngatik \cdots	0	1	$\mathbf 0$	1	Ω	Ω	0	0
Total .	751	751	2,184	2,184	310	310	949	949

TABLE ¹ IMMIGRATION INTO PINGELAP AND MOKIL

range migrants come from these atolls and from Ponape, Kusaie, and the Mortlock Islands (fig. 1). Migration from large distances is exclusively male, and the mean family size for such migrants is small.

Parent-offspring distances have been computed both by including the Gilbert

FIG. 1.-Map of Micronesia

and Marshall Islands with the short-range migrants and by pooling them with the Caucasian, Japanese, and other long-range migrants (table 2). Ngatik and Nukuoro Atolls are considered long-range because of the Polynesian culture of Nukuoro, Caucasian admixture on Ngatik, and the small number of migrants from both atolls.

Let k be the proportion of short-range migrants after long-range migrants have been excluded. For example, treating the Gilberts and Marshalls as short-range populations, we estimate m for Pingelap as $m = 25.5/2(2184) = .00584$, where 25.5 is the number of children from long-range migrants (allowing for one child from a Marshallese-Caucasian father) among 2,184 children recorded in the Pingelapese population since Typhoon Lengkieki; the factor of 2 arises because only one of the two parents in an exogamous mating is an immigrant. Similarly, $k = 114.5/[2(2,184) - 25.5] = .02637$, where 114.5 is the number of children from short-range migrants into Pingelap. Malecot [2] provided the approximation

$$
m_e = \sqrt{m(m+2k)} = .0185\tag{2}
$$

where m_e is the effective migration rate into Pingelap. The corresponding figure for Mokil is .0561, or three times as great.

These calculations are influenced by the division of migrants into short- and long-range groups. Since each population has two steppingstones in linear migration and four steppingstones for migration in a plane, the number of populations considered to yield short-range migrants should not be large. If populations on the Gilbert and Marshall Islands, which contribute many immigrants but lie at a considerable distance from Pingelap and Mokil, are treated as long-range migrants, the estimate of m_e becomes .0245 for Pingelap and .0820 for Mokil. Notice that these estimates were derived with no assumption about the effective population size and with only a qualitative consideration of parent-offspring distances.

The Forward Solution with a Migration Matrix

Our treatment of isolation by distance did not estimate the effective population size and therefore gave no prediction of the evolution of kinship from generation to generation; furthermore, the separation of short- and long-range migration seemed arbitrary. These deficiencies are avoided in a migration matrix, which we have constructed in table 3. The two columns corresponding to Pingelap and Mokil populations of residence are extracted from table 2. We cannot know how many Pingelapese and Mokilese had children in other populations, but it is mathematically convenient and seems plausible to assume that the numbers of migrants from i to j and from j to i are equal, permitting us to complete the first two rows of the matrix.

This leaves us with the submatrix for migration among the populations of Kusaie, the Mortlocks, Gilberts, Ponape, and the Marshalls, on which there is only anecdotal evidence [13]. It seems likely that migration between the Marshalls and Mortlocks, Gilberts and Mortlocks, and Gilberts and Ponape can be neglected. Since Kusaie and Ponape are large islands lying as close to the Mortlocks as Pingelap and Mokil, we assume that migration among them is equal to the sum of migrants to Pingelap and Mokil, and similarly for the Marshalls and Gilberts. Marginal totals were obtained as 6,204/1,208 times the census size, where 6,204 is the sum of the entries for Pingelap and Mokil in table 4 and 1,208 is the census size of the two atolls.

We see that it is possible to construct a migration matrix on reasonable assumptions, but only by making crude extrapolations from the data. If we had made as

TABLE 3

NUMBER OF MIGRANTS

ESTIMATES OF EFFECTIVE MIGRATION RATE m_{e} TABLE 4

MORTON ET AL.

thorough a migration census of the other populations over the last 200 years as we did for Pingelap and Mokil, we would have encountered migrants from additional populations, requiring either more extrapolation or truncation of the matrix. It does not seem to us that the errors introduced by extrapolation are likely to be serious for a study centered on the two reference atolls, but it should be stressed that methods based on migration matrices are inherently approximate, just as a model of isolation by distance is. Only by using these in conjunction with other methods can the genetic structure of populations be revealed.

Dividing each element by the column total, we obtain the column-stochastic exchange matrix P, in which the element p_{ij} is the probability that an individual having a child in j was born in i. Since table 3 is symmetrical, the transpose P' is the row-stochastic transition matrix T, in which the element t_{ii} is the probability that a parent born in i has a child in j . The computer program NUMIX [14] gives $\phi_{ii}(t)$ from a systematic pressure m, an exchange matrix P, and a vector n of effective population sizes [8]. Symmetry of table 3 guarantees that the effective population sizes are stable (apart from trivial effects of rounding errors in P).

Calculation of effective population size is fraught with difficulties, which we shall discuss later. At this point we take the effective size N as $N_0/3$, where N₀ is the current census size (table 2) and N corresponds roughly to the current breeding size, or the number of individuals between the ages of 15 and 45 years. To test whether estimates of effective migration rate m_e depend critically on population size, we have also considered effective sizes of $N/10$ and $10N$. In the island reduction for population i ,

$$
m_e = m + (1 - m) \sum_j p_{ji} (1 - \phi_{ji}/\phi_{ii}). \tag{3}
$$

In the steppingstone reduction,

$$
m_e = \sqrt{m'(m'+2k)}\tag{4}
$$

where $m' = m + (1 - m) \Sigma p_{ii} (1 = \phi_{ii}/\phi_{ii}), k = [(1 - m) \Sigma p_{ii}]/(1 - m')$, and the summation is over all j except steppingstones $[8, 14]$. We took either Pingelap or Mokil as a steppingstone for the other atoll. Table 4 shows that varying N by a hundredfold has virtually no effect on the estimates of m_e , which differ slightly between the island and steppingstone reductions.

Malécot's formula [2] for isolation by distance, $\phi(d) = ae^{-bd}$, may be fitted to $\phi_{ii}(\infty)$. Measuring distance from Pingelap and taking $m = .00584$, we find $b =$.0091 \pm .0009. Measuring distance from Mokil and taking $m = .01923$ gives $b =$.0124 \pm .0016. These estimates (which do not depend on the population size) are greater than the Malécot approximation for large distances, $b = \sqrt{2m/\sigma^2}$, where σ^2 is the mean value of d^2 from parent-offspring pairs. Even with the Marshalls and Gilberts treated as long-distance populations, the data of table 2 predict $b = .0028$ for Pingelap and .0070 for Mokil. Imaizumi and Morton [15] obtained $b = .0014 \pm .0005$ from bioassay of blood groups and .0005 \pm .0001 from cognate frequencies in Micronesia, but observed that b was significantly greater at small distances.

The Backward Solution with a Migration Matrix

The forward solution gave no estimates of the effective migration rate between Pingelap and Mokil, and there were small differences between the island and steppingstone reductions. These problems are eliminated in the backward solution, which we believe gives the best estimate of m_e . From predictions of $\phi_{ii}(t)$ by the forward solution, we fitted the values for $t = 1$, 10, and ∞ to equation (1), iterating N_e and m_e simultaneously. Again, the estimates of m_e are consistent at different multiples of N (table 4). Agreement with the forward solution is closer for the island than the steppingstone reduction. The average of reciprocal values of *m* for Pingelap \times Mokil is .0170, indicating (in conjunction with large estimates of N) a slow approach to equilibrium.

DISCUSSION

We have seen that different methods lead to consistent estimates of the effective migration rate m_e even when the vector of effective population size is varied by a factor of 100. The effective population size, on the other hand, is sensitive to the migration matrix as well as to other demographic parameters. Table ⁵ shows that

TABLE ⁵

ESTIMATES OF EFFECTIVE POPULATION SIZE N_e in Backward Solution WITH m_e ITERATED SIMULTANEOUSLY

Assumed Size	Pingelap $m = .00584$	Mokil $m = .01923$	$m = .00584$	Pingelap \times Mokil Mokil \times Pingelap $m = .01923$
N	293	158	1,665	1,817
$N/10$	-29	16	173	187
$10N$	2.926	1,584	16,581	18,304
Effective size/assumed size	1.08	1.21	\cdots	\cdots

the estimates of N_e are discrepant and tend to be larger than N. This is because each local population participates in the gene pool of its neighbors, as demonstrated for artificial populations [16]. Population growth, avoidance of incest, and the mean and variance of fertility also influence the effective population size.

The best estimation procedure would therefore seem to require that m_e be calculated from migration data, as above, and then N_e be inferred at this value of m_e from genealogical data or from bioassay of polymorphisms, anthropometrics, or isonymy. This avoids convergence problems when N_e and m_e are estimated simultaneously in small samples.

SUMMARY

A theory is derived which estimates the effective population size N_e and the effective migration rate m_e from coefficients of kinship in successive generations, and thereby predicts the equilibrium value of kinship from incomplete genealogies. The effective migration rate determined in this way from the exchange matrix is

MORTON ET AL.

.0267 for Pingelap, .0763 for Mokil, and .0170 for Pingelap \times Mokil. Selection and mutation seem to be negligible stabilizing pressures compared to migration in these populations.

APPENDIX

DEMOGEN. A program for demographic analysis of genealogical data. Let $y(t)$ be an estimate of a coefficient of kinship $\phi(t)$ with weight $W(t)$ at generation $t = 1, 1.5, 2$, etc., where *n* values of *t* have $W(t) > 0$ and *k* parameters are to be estimated. We take $W(t) = 1$ unless defined on input. To minimize the variance

$$
\sigma^2 = \frac{\sum W(t) \left[y(t) - \phi(t) \right]^2}{n - k}
$$

where

$$
\phi(t) = \left[\frac{1}{4Nm + 1}\right] \left[1 - e^{-(2m+1/2N)t}\right],
$$

we use the iteration

$$
\left[\begin{array}{c}N\\m\end{array}\right]=\left[\begin{array}{c}N_0\\m_0\end{array}\right]+UK^{-1}.
$$

Here N_0 and m_0 are trial values and

$$
U = [\Sigma W x_N y / \sigma^2 \quad \Sigma W x_m y / \sigma^2], K = \left[\begin{array}{cc} \Sigma W x_N^2 / \sigma^2 & \Sigma W x_N x_m / \sigma^2 \\ & \Sigma W x_m^2 / \sigma^2 \end{array} \right]
$$

where

$$
x_N = \frac{\partial \phi(t)}{\partial N} = -\left[\frac{1}{4Nm + 1}\right] \left[\frac{t}{2N^2}\right] e^{-(2m + 1/2N)t}
$$

$$
- \left[1 - e^{-(2m + 1/2N)t}\right] \left[\frac{4m}{(4Nm + 1)^2}\right],
$$

$$
x_m = \frac{\partial \phi(t)}{\partial m} = \left[\frac{2t}{4Nm + 1}\right] e^{-(2m + 1/2N)t}
$$

$$
- \left[1 - e^{-(2m + 1/2N)t}\right] \left[\frac{4N}{(4Nm + 1)^2}\right],
$$

and $y = y(t) - \phi(t)$. If either parameter is not iterated, the corresponding elements are deleted from U and K . The standard errors are

$$
\sigma_N = \sqrt{1/K_{NN}}
$$
 if *m* is not iterated, $= \sqrt{K^{NN}}$ if *m* is iterated;

and

$$
\sigma_m = \sqrt{1/K_{mm}}
$$
 if N is not iterated, $= \sqrt{K_{mm}}$ if N is iterated.

Derived quantities are

$$
\Phi = \frac{1}{4Nm + 1}
$$

$$
b = 2m + 1/2 N
$$

with standard errors $\sigma_{\Phi} = \sqrt{\partial K^{-1} \partial \theta}$, where

$$
\partial = \left[\frac{\partial \Phi}{\partial N} \frac{\partial \Phi}{\partial m} \right] = \left[- \frac{4m}{(4Nm + 1)^2} - \frac{4N}{(4Nm + 1)^2} \right],
$$

and $\sigma_b = \sqrt{\Delta K^{-1} \Delta'}$ where

$$
\Delta = \left[\begin{array}{cc} \frac{\partial b}{\partial N} & \frac{\partial b}{\partial m} \end{array} \right] = \left[\begin{array}{cc} - & 1 \\ - & \frac{2N^2}{N^2} & 2 \end{array} \right].
$$

Again the corresponding elements of ∂ , Δ , and K are omitted if a parameter is not estimated.

REFERENCES

- 1. WRIGHT S: Systems of mating. I. The biometric relations between parent and offspring. Genetics 6:111-123, 1921
- 2. MALÉCOT G: Les mathématiques de l'hérédité. Paris, Masson, 1948
- 3. MORTON NE, YASUDA N: The genetical structure of human populations, in Les deplacements humains, edited by SUTTER J, Entretiens de Monaco en Sciences Humaines, Paris, Hachette, 1962, pp 185-203
- 4. CROW JF, MANGE AP: Measurements of inbreeding from the frequency of marriages between persons of the same surname. Eugen Quart 12:199-203, 1965
- 5. LEWONTIN RC: Population genetics. Ann Rev Genet 1:37-70, 1967
- 6. MALE'COT G: Quelques schemas probabilistes sur la variabilite des populations naturelles. Ann Univ Lyon (Sec A) 13:37-60, 1950
- 7. BODMER W, CAVALLI-SFORZA LL: A migration matrix model for the study of random genetic drift. Genetics 59:565-592, 1968
- 8. IMAIZUMI Y, MORTON NE, HARRIS DE: Isolation by distance in artificial populations. Genetics 66:569-582, 1970
- 9. WRIGHT S: Stochastic processes in evolution, in Stochastic Models in Medicine and Biology, edited by GURLAND J, Madison, Univ. Wisconsin Press, 1964, pp 199-244
- 10. WRIGHT S: Evolution in Mendelian populations. Genetics 16:97-157, 1931
- 11. MALÉCOT G: Conséquences statistiques de la parenté. Mathematical Statistical Institute, 36th Sess., Sydney, Bloxham & Chambers, 1969, pp 651-669
- 12. MORTON NE, HuRD J, LITTLE GF: Pingelap and Mokil atolls: ^a problem in population structure. In preparation
- 13. RIESENBERG SH: Table of voyages affecting Micronesian islands. Oceania 36:155- 170, 1965
- 14. HARRIS DE: NUMIX, in A Genetics Program Library, edited by MORTON NE, Honolulu, Univ. Hawaii Press, 1969, p 41
- 15. IMAIZUMI Y, MORTON NE: Isolation by distance in New Guinea and Micronesia. Archeology Phys Anthrop Oceania 5:218-235, 1970
- 16. HARRIS DE, YEE S, MORTON NE: Evolution of kinship in artificial populations. In preparation