ISOLATION BY DISTANCE*

SEWALL WRIGHT
The University of Chicago[1]

Received November 9, 1942

STUDY of statistical differences among local populations is an important line of attack on the evolutionary problem. While such differences can only rarely represent first steps toward speciation in the sense of the splitting of the species, they are important for the evolution of the species as a whole. They provide a possible basis for intergroup selection of genetic systems, a process that provides a more effective mechanism for adaptive advance of the species as a whole than does the mass selection which is all that can occur under panmixia.

RANDOM DIFFERENTIATION UNDER THE ISLAND MODEL

Mathematical consideration requires the use of simple models of population structure. The simplest model is that in which the total population is assumed to be divided into subgroups, each breeding at random within itself, except for a certain proportion of migrants drawn at random from the whole. Since this situation is likely to be approximated in a group of islands, we shall refer to it as the island model.

The gene frequency (q) of a subgroup tends to vary about a certain equilibrium point ($\hat{q}$) in a distribution curve ($\phi(q)$) determined by the net systematic pressure (measured by Δq, the net rate of change of gene frequency per generation from recurrent mutation, immigration, and selection) in conjunction with the cumulative effects of accidents of sampling (random deviation δq, variance per generation $\sigma^2_{\delta q}$) (WRIGHT 1929, 1931, 1942).

$$\phi(q) = (C/\sigma^2_{\delta q}) \exp\left[2\int (\Delta q/\sigma^2_{\delta q})dq\right]. \tag{1}$$

Let N be the effective size of the subgroup, m the effective proportion of its population replaced in each generation by migrants, and q_t the gene frequency in the total population. The rate of change of gene frequency per generation in a subgroup, taking account only of immigration pressure, is $\Delta q = -m(q-q_t)$. In a random breeding population $\sigma^2_{\delta q} = q(1-q)/2N$. Substitution in (1) gives the following, choosing C so that $\int_0^1 \phi(q)dq = 1$ (WRIGHT 1931, 1942).

$$\phi(q) = \frac{\Gamma(4Nm)}{\Gamma(4Nmq_t)\Gamma[4Nm(1-q_t)]} q^{4Nmq_t-1}(1-q)^{4Nm(1-q_t)-1} \tag{2}$$

$$\bar{q} = \int_0^1 q\phi(q)dq = q_t \tag{3}$$

* A portion of the cost of composing the mathematical formulae is borne by the Galton and Mendel Memorial Fund.

[1] Acknowledgment is made to the DR. WALLACE C. and CLARA A. ABBOTT MEMORIAL FUND of the UNIVERSITY OF CHICAGO for assistance in connection with the calculations.

$$\sigma_q^2 = \int_0^1 (q - \bar{q})^2 \phi(q) dq = q_t(1 - q_t)/(4Nm + 1). \tag{4}$$

In the derivation of (1), it was assumed that Δq is sufficiently small that terms involving $(\Delta q)^2$ might be ignored. A more accurate value of σ_q^2 may be obtained directly. The deviation of a local gene frequency from the average, $(q - q_t)$, tends to be reduced to $(1 - m)(q - q_t)$ in the next generation. The mean sampling variance of $(q + \Delta q)$ is

$$\frac{1}{2N}\int_0^1 [q - m(q - q_t)][1 - q + m(q - q_t)]\phi(q)dq = [q_t(1 - q_t) - (1 - m)^2\sigma_q^2]/2N. \tag{5}$$

Thus with a steady balance between the effects of immigration and of the accidents of sampling

$$\sigma_q^2 = (1 - m)^2\sigma_q^2 + [q_t(1 - q_t) - (1 - m)^2\sigma_q^2]/2N \tag{6}$$

$$\sigma_q^2 = q_t(1 - q_t)/[2N - (2N - 1)(1 - m)^2]. \tag{7}$$

This is approximately the same as (4) for small values of m but becomes $q_t(1 - q_t)/2N$, the sampling variance, in the limiting case of no isolation whatever ($m = 1$). This is about twice as great as given by (4) in this extreme case.

The variance, excluding the immediate sampling variance may be obtained by multiplying (7) by $(1 - m)^2$ as indicated in (6). Formula (4) lies between the values with and without the immediate sampling variance.

Under exclusive uniparental reproduction, whether vegetative or by self-fertilization, the distribution of alternative genotypes may be treated by the same theory except for replacement of 2N by N. Immigration pressure is the same but the sampling variance is $q(1 - q)/N$.

THE INBREEDING COEFFICIENT

Departures from panmixia may be expressed in terms of the average inbreeding coefficient of individuals, relative to the total population under consideration. This coefficient has been defined as the correlation between uniting gametes with respect to the gene complex as an additive system. It has been shown that its value can be found for any pedigree by finding all paths by which one may trace back from the egg to a common ancestor (A) and thence forward to the sperm along a wholly different path. According to the theory of path coefficients, the correlation between uniting gametes is the sum of contributions from all such paths (WRIGHT 1921, 1922b).

$$F = \sum [(1/2)^{n_S + n_D + 1}(1 + F_A)] \tag{8}$$

where F and F_A are the inbreeding coefficients of the individual and of a common ancestor of sire and dam, respectively, and n_S and n_D are the numbers of generations from sire and dam, respectively, to this common ancestor. In a population in which the average inbreeding coefficient is F, the frequencies of genotypes (one pair of alleles) are as follows (WRIGHT 1921, 1922a).

(9)

Genotype	Frequency
AA	$x_t = q_t^2(1 - F) + q_tF$
Aa	$y_t = 2q_t(1 - q_t)(1 - F)$
aa	$z_t = (1 - q_t)^2(1 - F) + (1 - q_t)F$

The inbreeding, measured by F, may be of either of two extreme sorts: sporadic mating of close relatives with no tendency to break the population into subgroups, and division into partially isolated subgroups, within each of which there is random mating. The latter is the case in which we are primarily interested here. Assume that there are K subgroups each of size N. The proportion of heterozygotes within a subgroup is $2q'(1-q')$ where q' is the gene frequency in the parental generation, including immigrants.

$$(10) \qquad y_t = 2\sum_1^K q'(1 - q')/K = 2q_t - 2(\sum q'^2)/K.$$

The variance of the gene frequencies of the subgroups, not allowing for accidents of sampling in the last generation, is

$$(11) \qquad \sigma_{q'}^2 = \sum_1^K (q' - q_t)^2/K = (\sum q'^2)/K - q_t^2$$

$$(12) \qquad y_t = 2q_t(1 - q_t) - 2\sigma_{q'}^2 \quad \text{from (10) and (11)}$$

$$(13) \qquad \sigma_{q'}^2 = q_t(1 - q_t)F \quad \text{from (9) and (12).}$$

This formula does not allow for the contribution to variance due to accidents of sampling in the last generation. Thus it gives $\sigma_{q'}^2 = 0$ instead of $\sigma_q^2 = q_t(1-q_t)/2N$ for $F=0$. To compare with (7) it must be divided by $(1-m)^2$.

$$(14) \qquad \sigma_q^2 = q_t(1 - q_t)F/(1 - m)^2$$

$$(15) \qquad F = (1 - m)^2/[2N - (2N - 1)(1 - m)^2] \quad \text{from (14) and (7)}$$

$$(16) \qquad m = 1 - \sqrt{2NF/[(2N - 1)F + 1]}$$

$$(17) \qquad \sigma_q^2 = q_t(1 - q_t)[(2N - 1)F + 1]/2N.$$

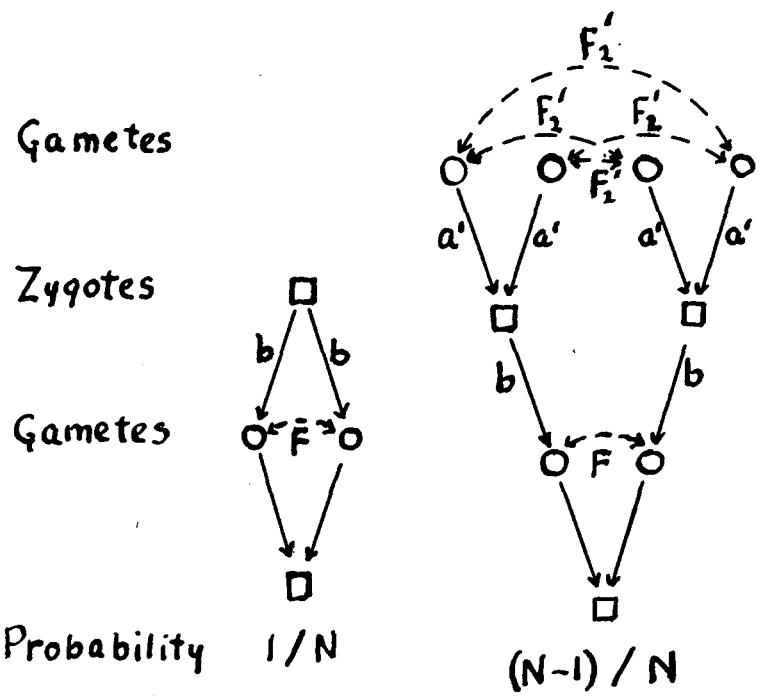
The formula $F = 1/[4Nm+1]$ given in a preceding paper (DOBZHANSKY and WRIGHT 1941) is a satisfactory approximation if m is small.

This island model is not likely to be exactly realized in nature. In most cases, the actual immigrants to a population come from immediately surrounding localities in excess and thus are not a random sample of the species. This can be remedied to some extent by multiplying the proportion of replacement by an appropriate factor to obtain the effective immigration index. If q_m is the gene frequency in the actual immigrants (varying from group to group) the appropriate factor would be $(q-q_m)/(q-q_t)$. Unfortunately the values for effective m for different loci may be very different.

LOCAL INBREEDING IN A CONTINUOUS AREA

At the opposite extreme from the island model is that in which there is complete continuity of distribution, but interbreeding is restricted to small distances by the occurrence of only short range means of dispersal. Remote populations may become differentiated merely from *isolation by distance* (WRIGHT 1938, 1940).

Each individual has its origin at a particular place. Assume that its parents originated at distances from this place with a certain variance both in longitude



and in latitude. If the same condition held in preceding generations, the grandparents originated at distances with twice this variance in longitude and in latitude and the ancestors of generation K originated at distances with K times this variance in both directions. The parents may be considered as if drawn at random from a territory with a certain radius R and effective population size N. The ancestors of generation K may then be considered as drawn similarly from a territory of radius $\sqrt{K}$ R and effective population size KN.

We shall use the term parental group for the population (effective size N) from which the parents of an individual may be considered to be drawn; the term random breeding or panmictic unit will be used for any local population of the same effective size as the parental group.

The assumption of random union of gametes, including self fertilization (probability 1/N) can be made with sufficient accuracy even though there is actually no self fertilization. It has been shown that such unions in a population of constant size N lead to fixation at the rate $1/2N$ in comparison with the

rate $[(N+1)-\sqrt{(N^2+1)}]/2N$ either in a population of size N equally divided between males and females or in a population of N monoecious individuals in which self fertilization does not occur. As the latter formula may be written $[1-(1/2N)\cdots]/2N$ the difference is ordinarily negligible (WRIGHT 1931).

The inbreeding coefficient of individuals in such a population can be calculated from its definition as the correlation between uniting gametes. Let F_x be the correlation between random gametes drawn from a population of size xN and use primes to indicate preceding generations as in the text figure (p. 117). The inbreeding coefficient itself would be F_1 in this terminology. The values of these coefficients can be expressed in terms of coefficients for preceding generations by tracing all connecting paths and noting that the path coefficient b, relating gamete to parental zygote, has the value $\sqrt{(1+F')/2}$ and that the path coefficient, a, relating offspring zygote to one of the gametes that produced it, has the value $\sqrt{1/[2(1+F)]}$. The compound coefficient $ba'=\frac{1}{2}$ (WRIGHT 1921). It may easily be seen that (8) can be deduced at once from these considerations.

In the case of continuity

$$(18)\qquad \begin{cases} F = \dfrac{1}{N} b_2 + \dfrac{N-1}{N} 4b^2a'^2F_2' = \dfrac{1}{N}\left(\dfrac{1+F'}{2}\right) + \dfrac{N-1}{N} F_2' \\ F_2' = \dfrac{1}{2N}\left(\dfrac{1+F''}{2}\right) + \dfrac{2N-1}{2N} F_3'' \\ F_3'' = \dfrac{1}{3N}\left(\dfrac{1+F'''}{2}\right) + \dfrac{3N-1}{3N} F_4''' \text{ etc.} \end{cases}$$

(19) Thus

$$F = \frac{1+F'}{2N} + \frac{N-1}{N}\left\{\frac{1}{2N}\left(\frac{1+F'}{2}\right) + \frac{2N-1}{2N}\left[\frac{1}{3N}\left(\frac{1+F'''}{2}\right) + \cdots\right]\right\}.$$

If the same population structure has continued indefinitely, primes may be dropped.

$$(20)\qquad F = \left(\frac{1+F}{2N}\right)\left[1 + \frac{1}{2}\left(\frac{N-1}{N}\right) + \frac{1}{3}\left(\frac{N-1}{N}\right)\left(\frac{2N-1}{2N}\right) + \frac{1}{4}\left(\frac{N-1}{N}\right)\left(\frac{2N-1}{2N}\right)\left(\frac{3N-1}{3N}\right)\cdots\right].$$

This is an infinite series, but in practice the value of F that is of interest is that relative to some finite population. The correlation between random gametes in a population of size KN is F_K which may be taken as zero, thereby stopping the series at $(K-1)$ terms. Let t_x be the xth term in the series in brackets and $\sum_1^{K-1} t$ the sum of first $(K-1)$ such terms

$$F = \sum_{1}^{K-1} t \Big/ \left[2N - \sum_{1}^{K-1} t\right] \tag{21}$$

$$t_x = \frac{(x-1)N - 1}{xN}\, t_{(x-1)}. \tag{22}$$

Let $t_{(x-0.5)} = (t_x + t_{(x-1)})/2$ and $\Delta t_{(x-0.5)} = t_x - t_{(x-1)}$

$$\frac{\Delta t_{(x-0.5)}}{t_{(x-0.5)}} = -\frac{2(N+1)}{N(2x-1)-1}. \tag{23}$$

If the values of t are treated as ordinates of a curve with abscissas x, we may write t and x in place of $t_{(x-0.5)}$ and $(x-0.5)$, respectively. The following then hold approximately

$$\frac{dt}{tdx} = -\frac{2(N+1)}{2Nx-1} \tag{24}$$

$$t = C\left(x - \frac{1}{2N}\right)^{-(N+1)/N} \tag{25}$$

$$\sum_{K_1}^{K_2-1} t = \int_{K_1-0.5}^{K_2-0.5} tdx \quad \text{approximately} \tag{26}$$

$$\sum_{K_1}^{K_2-1} t = CN\left[\left(K_1 - \frac{1}{2} - \frac{1}{2N}\right)^{-1/N} - \left(K_2 - \frac{1}{2} - \frac{1}{2N}\right)^{-1/N}\right]. \tag{27}$$

The value of the constant C can be obtained by equating actual and estimated values of t. Estimates for all but the first few terms in the series are in close agreement. Thus if $N = 10$

Actual series $[1 + .45 + .285 + .206625 + \cdots]$
Estimated series $C[1.05805 + .47969 + .30423 + .22067 + \cdots]$

The estimated value of C from the first term is .9451, from the second term .9381, from the third term .9363. The limiting value is .935774. The value of C approaches 1 as N increases. Thus for $N = 100$, $C = .994157$.

Estimates of $\sum_1^{K-1} t$ directly from (27) are not good approximations, but most of the error is in the first few terms. Good estimates can be made by using the actual values from (22) for these terms and the estimates from (27) for the later terms. For $N = 10$

	Actual (22)	Estimate (27)	Error of Estimate
$\sum_1^3 t$	1.73500	1.86782	+.13282
$\sum_4^9 t$	.79002	.79250	+.00248
$\sum_{10}^{39} t$	.99511	.99541	+.00030
$\sum_{40}^{99} t$	.57228	.57228	+.00000

A priori, one would expect F to approach 1 as a limit as the size of population is increased without limit. This requires that $\sum_1^\infty t$ approach N. Trial for values of N from 10 to 10,000 indicates that this is actually the case and thus gives a good check on the theory. Following are examples:

	N = 10	N = 20	N = 50		N = 100
$\sum_1^{39} t$ from (22)	3.52013	3.86519	4.09266	$\sum_1^9 t$	2.797
$\sum_{40}^\infty t$ from (27)	6.47987	16.13481	45.90734	$\sum_{10}^\infty t$	97.203
	10.00000	20.00000	50.00000		100.000

LOCAL INBREEDING ALONG A LINEAR RANGE

In a species with an essentially one dimensional range (parents drawn from the whole width) the extent along the range from which the ancestors of generation K are drawn is proportional to $\sqrt{K}$ as with area continuity, but the effective size of the corresponding population is $\sqrt{K}\,N$ instead of KN. By analogous reasoning

$$F = \sum t/(2N - \sum t) \tag{28}$$

where

$$\sum t = \left[1 + \frac{1}{\sqrt{2}}\left(\frac{N-1}{N}\right) + \frac{1}{\sqrt{3}}\left(\frac{N-1}{N}\right)\left(\frac{\sqrt{2}N-1}{\sqrt{2}N}\right)\cdots\right]$$

$$t_x = \frac{N\sqrt{(x-1)} - 1}{N\sqrt{x}}\, t_{(x-1)} \tag{29}$$

$$\frac{\Delta t_{(x-0.5)}}{t_{(x-0.5)}} = \frac{2N(\sqrt{x-1} - \sqrt{x}) - 2}{N(\sqrt{x-1} + \sqrt{x}) - 1}. \tag{30}$$

Treating this expression as the slope at the mid-interval and replacing (x−0.5) by x

$$\begin{aligned}\frac{dt}{tdx} &= \frac{2N(\sqrt{x-0.5} - \sqrt{x+0.5}) - 2}{N(\sqrt{x-0.5} + \sqrt{x+0.5}) - 1} \\ &= -\frac{N[1 + 1/(32x^2) + \cdots] + 2\sqrt{x}}{2Nx[1 - 1/(32x^2) + \cdots] - \sqrt{x}}.\end{aligned} \tag{31}$$

Ignoring $1/(32x^2)$ and smaller terms in the brackets, this yields

$$t = Ce^{-2\sqrt{x}/N}[\sqrt{x} - (1/2N)]^{-[1+(1/N)^2]}. \tag{32}$$

This seems to be as accurate an approximation as is warranted after replacement of Δt/t by dt/tdx.

Comparisons of actual and calculated values of t indicate that estimates of C approach stability after a few terms. For N = 10, C = 1.1529 (from 30th to

40th terms). For N = 100, C = 1.01465 (from 9th and 10th terms). For larger values of N, especially if x is 10 or more, it may be sufficiently accurate to take dt/tdx as $-(N+2\sqrt{x})/2Nx$, C = 1

$$t = e^{-2\sqrt{x}/N}/\sqrt{x} \quad \text{approximately.} \tag{33}$$

In this case

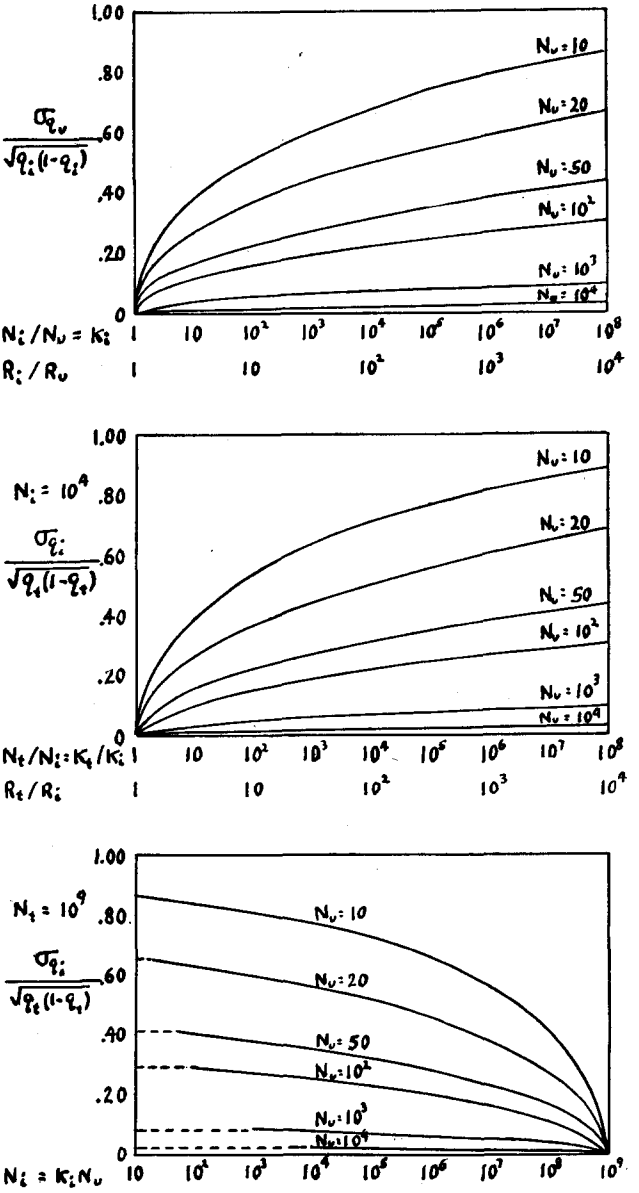
$$\sum_{K_1}^{K_2-1} t = N(e^{-2\sqrt{K_1}/N} - e^{-2\sqrt{K_2}/N}) \tag{34}$$

The value of $\sum_1^{K-1} t$ can be approximated by finding actual $\sum_1^9 t$ from (29), estimating $\sum_{10}^{K-1}$ from (34) and multiplying the latter by the mean ratio of t from (32) to that from (33). Calculation of $\sum_1^\infty t$, N = 10, by this method (by steps) gave 10.008 (instead of theoretical 10) and for N = 100 gave 100.07 instead of theoretical 100. These theoretical values are on the assumption that the limiting value of F is 1 which again is seen to be verified.

CORRELATION BETWEEN ADJACENT INDIVIDUALS UNDER UNIPARENTAL REPRODUCTION

The effect of isolation by distance on the frequencies of two alternative types in a population with exclusive uniparental reproduction can be treated similarly, again assuming that there are no complications from other factors. The treatment, however, cannot be in terms of the inbreeding coefficient. Let E be the correlation between adjacent individuals, and assume that there is short range dispersion in each generation such that individuals are derived from a parental group of effective size N. With area continuity, the ancestors of the Kth generation are drawn from a population of effective size KN. The correlation between adjacent individuals can be analyzed into two components, that due to the chance, 1/N, of derivation from the same parent and that due to the chance (N−1)/N, of derivation from different individuals of the group, the correlation between which may be represented by E_2' in analogy with F_2' in the case of biparental reproduction. This in turn can be analyzed into a component due to the chance 1/2N of derivation from the same individual of the second preceding generation and that due to the chance (2N−1)/2N of derivation from different individuals of this group, the correlation between which may be represented by E_3''.

$$\begin{cases} E = \dfrac{1}{N} + \dfrac{N-1}{N} E_2' \\ E_2' = \dfrac{1}{2N} + \dfrac{2N-1}{2N} E_3'' \\ E_3'' = \dfrac{1}{3N} + \dfrac{3N-1}{3N} E_4''' \text{ etc.} \end{cases} \tag{35}$$



Figures 1 to 3. Variability of gene frequencies of local populations within a continuously inhabited area that extends indefinitely in all directions. It is assumed that there is no appreciable long range dispersal or mutation. Each curve applies to a particular size (N_u) of random breeding unit and thus to a certain amount of short range dispersal. Variability is measured by $\sigma_x/\sqrt{q_y(1-q_y)}$ where q_x represents the gene frequencies of the subgroup in question and q_y that of the comprehensive population.

FIGURE 1 (top).—The variability of gene frequencies (q_u) of the random breeding units themselves, within areas up to 10^4 times their radius (R_i/R_u) or 10^8 times their population size (N_i/N_u).

Again we may drop primes if the same population structure has continued for a large number of generations.

$$E = \frac{1}{N}\left[1 + \frac{1}{2}\left(\frac{N-1}{N}\right) + \frac{1}{3}\left(\frac{N-1}{N}\right)\left(\frac{2N-1}{2N}\right) + \frac{1}{4}\left(\frac{N-1}{N}\right)\left(\frac{2N-1}{2N}\right)\left(\frac{3N-1}{3N}\right)\cdots\right]. \tag{36}$$

$$E = \sum t/N. \tag{37}$$

The series $\sum t$ is the same as encountered in the case of biparental reproduction, but the formula for E differs from that for F. It resembles it in approaching 1 as a limit, as is to be expected *a priori*, but for a given N, E is about twice as great as F for small values of $\sum t$, and the difference from the limit is only about half as great if $\sum t$ is close to 1. These relations are illustrated in figures 7 and 1 dealing with uniparental and biparental reproduction, respectively.

In the case of linear continuity and derivation of individuals from a parental population of N, the effective size of the population of the Kth ancestral generation is $\sqrt{K}\,N$, again as under biparental reproduction. By analogous reasoning $E = \sum t/N$ where $\sum t$ is the same series as in the biparental case. The relation of E to F for the same N is similar to that described above in the case of area continuity.

RANDOM DIFFERENTIATION OF PANMICTIC UNITS IN A CONTINUUM

Returning to biparental reproduction, the situation in a random breeding unit imbedded in a continuous population of defined size may be compared in some respects with that in an "island" whose population is replaced to such an extent in each generation by migrants representative of the whole that the inbreeding coefficient of individuals is the same. There is the important difference that adjacent groups should be closely similar in the former but uncorrelated in the latter. Nevertheless the amount of differentiation among groups taken at *random* from the whole should be the same in both cases since equations (9) to (17) apply in both. It is most convenient to use $\sqrt{F} = \sigma_q/\sqrt{q_t(1-q_t)}$ (from (13)) to measure this differentiation. It should be noted that this excludes the variability due to the immediate effect of sampling.

The theoretical variabilities of random breeding units of various sizes (10 to 10,000) within populations up to 10^8 times the size of the units (or 10^4 times the radius), continuous in all directions, are compared in figure 1. In interpreting this variability, it may be noted that if $q_t = \frac{1}{2}$, a value of $\sqrt{F}$ (ordinate)

K_i is the average number of generations of separate ancestry of random individuals of the population N_i.

FIGURE 2 (middle).—The variability of gene frequencies (q_i) of populations of a given size, $N_i = 10^4$, within areas up to 10^4 times their radius (R_t/R_i) or 10^8 times their population size (N_t/N_i). Note the similarity to Figure 1.

FIGURE 3 (bottom).—The variability of gene frequencies (q_i) of populations of any size, N_i, within a region with a population of a given size, $N_t = 10^9$.

greater than .577 means a U-shaped distribution of gene frequencies and thus very great differentiation. The situation is similar to that found where Nm is less than 0.5 in the island model. There is important differentiation down to at least $\sqrt{F}=.22$ (equivalent to $Nm=5$). There is only slight differentiation if $\sqrt{F}$ is less than .07 (equivalent to $Nm=50$) (*cf.* fig. 1, WRIGHT 1940).

It is apparent from figure 1 (this paper) that there is a great deal of local differentiation if the random breeding unit is as small as 10, even within a territory the diameter of which is only ten times that of the unit. If the unit has an effective size of 100, differentiation becomes important only at much greater relative distances. If the effective size is 1000, there is only slight differentiation at enormous distances. If it is as large as 10,000 the situation is substantially the same as if there were panmixia throughout any conceivable range.

The situation is very different as may be seen from figure 4 in a species whose range is essentially one dimensional (for example, a shore line). Different alleles may approach fixation in different parts of a range only 100 times the length of the random breeding unit if the effective size of the latter is less than 100. The range must be about 1000 times the length of the unit if the latter has a size of 1000 and about 10,000 times its length if the size of the unit is 10,000 to give this result. This difference between area and linear continuity has been suggested on *a priori* grounds by THOMPSON (1931) in connection with a study of the correlation between water distance and amount of differentiation within species of fish.

RANDOM DIFFERENTIATION IN A HIERARCHY OF SUBDIVISIONS

The attempt to apply these conclusions to actual cases is hampered by the difficulty of determining what are the random breeding units and their effective sizes. To obviate this, we should find how groups of any arbitrary size vary within a more comprehensive population.

Consider a total population, size N_t, subdivided into H groups of intermediate size N_i and these in turn subdivided into K random breeding groups of size N_u. The inbreeding coefficient of individuals is zero relative to the unit groups, F_i relative to the intermediate groups and F_t relative to the total. Both H and K, in contrast with N_u, will be treated as large numbers.

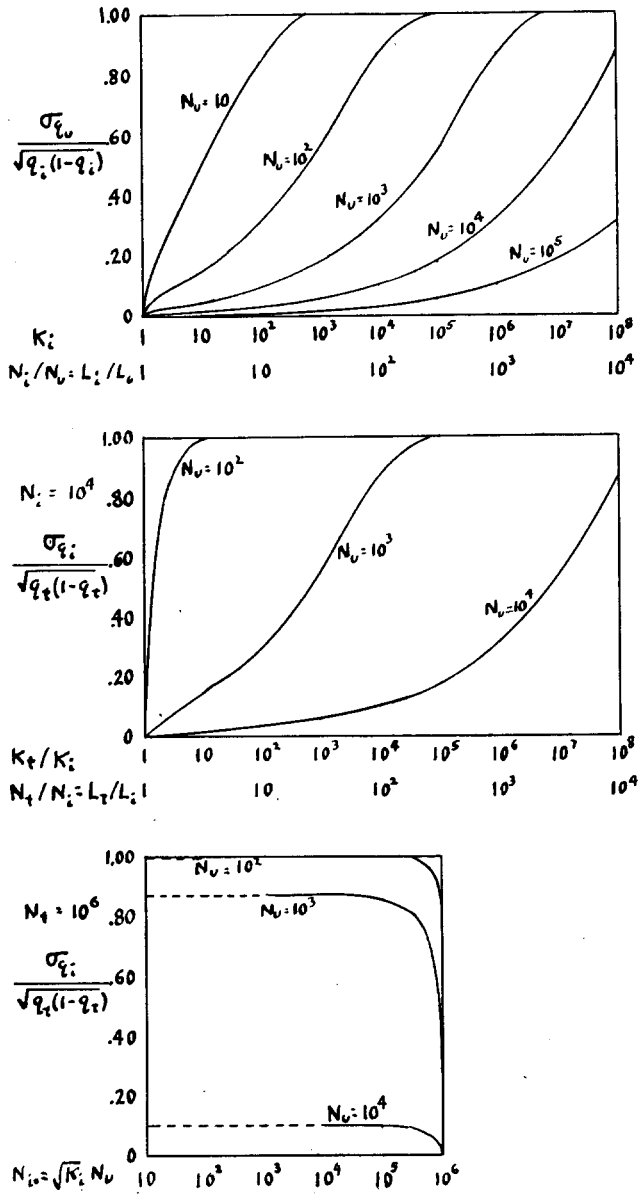
The variance of the gene frequency (q_u) of unit groups within the intermediate groups is given by (17) using the proper subscripts. The average value of this variance will be represented by $\sigma^2_{u\cdot i}$. The variance of the mean gene frequencies of the intermediate groups in the total will be represented by $\sigma^2_{i\cdot t}$ and that of q_u in the total by $\sigma^2_{u\cdot t}$.

(38) from (17) $$\sigma^2_{u\cdot i}=\sum_1^H [q_i(1-q_i)][(2N_u-1)F_i+1]/2HN_u$$

(39) $$\sigma^2_{u\cdot i}=[q_t(1-q_t)-\sigma^2_{i\cdot t}][(2N_u-1)F_i+1]/2N_u$$

(40) from (17) $$\sigma^2_{u\cdot t}=q_t(1-q_t)[(2N_u-1)F_t+1]/2N_u$$

(41) $$\sigma^2_{u\cdot t}=\sigma^2_{u\cdot i}+\sigma^2_{i\cdot t}$$



Figures 4 to 6. Similar to figures 1 to 3, respectively, except that a linear range (such as a shore line) is postulated.

FIGURE 4 (top).—The variability of gene frequencies (q_u) of the random breeding units themselves, within ranges up to 10^4 times their length (L_i/L_u) or population (N_i/N_u).

FIGURE 5 (middle).—The variability of gene frequencies (q_i) of populations of a given size, $N_i = 10^4$, within ranges up to 10^4 times their length (L_t/L_i) or population (N_t/N_i). Note the dissimilarity to figure 4 in contrast with the similarity of figures 1 and 2.

FIGURE 6 (bottom).—The variability of gene frequencies (q_i) of populations of any size (N_i) within a range with a population of a given size, $N_t = 10^6$.

(42) from (39) and (41) $$\sigma^2_{u\,t} = [q_t(1-q_t) - \sigma^2_{i\cdot t}][(2N_u - 1)F_i + 1]/2N_u + \sigma^2_{i\cdot t}.$$

(43) Equating (40) and (42) $$\sigma^2_{i\cdot t} = q_t(1-q_t)[F_t - F_i]/[1 - F_i].$$

This demonstration involves the assumption that there is inbreeding relative to the intermediate groups because these are subdivided. It may be noted that the same value of $\sigma^2_{i\cdot t}$ may be derived as follows without this assumption.

(44) From (9) $$y_t = 2q_t(1 - q_t)(1 - F_t).$$

But y_t is also the average heterozygosis of the intermediate groups

(45) $$y_t = \sum_1^H [2q_i(1 - q_i)(1 - F_i)]/H = 2(1 - F_i)\left(q_t - \left(\sum_1^H q_i^2\right)/H\right)$$

(46) $$\sigma^2_{i\cdot t} = \sum_1^H (q_i - q_t)^2/H = \left(\sum_1^H q_i^2\right)/H - q_t^2.$$

(47) From (45) and (46) $$y_t = 2[1 - F_i][q_t(1 - q_t) - \sigma^2_{i\cdot t}].$$

(48) From (44) and (47) $$\sigma^2_{i\cdot t} = q_t(1 - q_t)[F_t - F_i]/[1 - F_i].$$

In neither demonstration is there any assumption as to the geographic distribution of the values of the mean gene frequencies, q_i, within the total. They may be distributed at random as implied in the island model or there may be gradients as expected with continuity.

The quantity

$$\sqrt{\frac{(F_t - F_i)}{(1 - F_i)}} = \frac{\sigma_{i\cdot t}}{\sqrt{q_t(1 - q_t)}}$$

may be used as an index of the amount of differentiation among populations of any size N_i within a more comprehensive population (N_t). The variabilities of populations of effective size $N_i = 10{,}000$ are considered in figure 2 (area continuity) and figure 5 (linear continuity). In the case of area continuity the curves are somewhat similar to those shown in figure 1 for unit groups. It appears that populations of 10,000 (or any other size) exhibit about the same amount of differentiation within a whole whose population is a certain multiple of their own as the unit groups exhibit in a population that is the same multiple of their size. Whatever the size of the subpopulations considered the variability depends on the size of the inbreeding unit. There is an important amount of differentiation among large regions if the unit group is as small as 10, appreciable differentiation if the unit group is as large as 100 but little if it is as large as 1000. It should be said that there are important qualifications if there are other factors (mutation, rare long range dispersal or selection) which will be considered later.

The situation differs considerably in the case of linear continuity. Groups of size $N_i = 10{,}000$ approach the limiting amount of differentiation within populations only three times their length of range if $N_u = 100$ or less. There must be virtually complete fixation of one allele or the other over long distances with only short regions of transition. If $N_u = 1000$ there is relatively little differentiation within 10-fold lengths (that are heterallelic at all) but an approach to 100 percent differentiation in 100-fold lengths. Thus transition regions are of the order of 10 lengths. If $N_u = 10{,}000$, the transition regions are of the order of 10^3 lengths, and such groups approach 100 percent differentiation within 10^4 lengths.

The interpretation of figures 1, 2, 4, and 5 is somewhat complicated by the fact that these do not measure variability on a constant scale. The denominators of the ordinates (namely, $\sqrt{q_t(1-q_t)}$ in 2 and 5) increase with the abscissas. The tendency toward fixation of large populations means that at the lower abscissas the average value of q_t must be close to 0 or 1, making $\sqrt{q_t(1-q_t)}$ small. The structure of a population is exhibited in perhaps the most easily interpreted form by considering a constant comprehensive population N_t and showing how much differentiation there is among subdivisions of all sizes from the random breeding units up to major subdivisions ($\sigma_{q_i}/\sqrt{q_t(1-q_t)}$ plotted against N_i). Here the denominator is constant so that variability is always on the same scale.

Figure 3 shows that with area continuity, the amount of differentiation falls off slowly with the size of the subdivision considered. If $N_u = 10$ and N_t is 10^9 (or any other size in the absence of other factors) there is marked differentiation among populations that are 10 percent of the total, although much less than among subdivisions of smaller sizes. If $N_u = 100$, there is only moderate differentiation among the smaller subdivisions and very little among ones that are as large as 10 per cent of the total. In the case of linear continuity (fig. 6) there is virtually complete fixation of all subdivisions up to 10 percent of the total if N_u is 100 or less. If, however, N_u is 1000 there is a considerable proportion of these unit groups that are not fixed ($\sigma_{q_u}/\sqrt{q_t(1-q_t)} = .87$). The differentiation among larger populations up to $N_i = 0.1\ N_t$ is not appreciably less than among the unit groups. If $N_u = 10{,}000$, $\sigma_{q_u}/\sqrt{q_t(1-q_t)}$ is only .10, but this index is practically as great among larger populations up to 10 percent of the total. Thus with linear continuity most of the differentiation is that among large subdivisions of the total (of the order of 10 percent of its size). With area continuity, differentiation is more uniformly distributed at all levels.

Area and linear continuity as well as the island model are ideal cases. There may be all grades of intermediacy between area and linear continuity as exhibited in branching and reticular distributions. Even with rather complete area continuity there are almost certain to be variations in density of population. The ancestry of individuals in the centers of high density would spread out less rapidly than under the ideal theory with the consequence that there would in general be more differentiation among such centers than indicated, unless this is interfered with by other factors, which must now be considered.

COMPLICATING EFFECTS OF MUTATION AND LONG RANGE DISPERSAL

The foregoing theory indicates the possibility of an approach to fixation of different alleles in large areas of the same continuous population without the help of any differential action of selection. It is obvious, however, that this very slow process would be greatly affected by other factors that change gene frequency. The very fact of persistence of more than one allele over a long period of time tends to indicate that such factors are present in some sort of balance. Thus there may be reversible mutation, selection opposed by mutation or selection against both of two homozygotes. Moreover, the short range means of dispersal that have been postulated are likely to be supplemented by occasional long range dispersal. All of these tend to prevent fixation of one type even locally. On the other hand, selection may favor one allele in some places and others in other places. This would tend to increase local differentiation. It is necessary to consider how such processes affect the situation.

It will be well to review first the joint effects of recurrent mutation and long range dispersal in the case of the island model (WRIGHT 1931). The rate of change of gene frequency under recurrent reversible mutation varies linearly with the gene frequency: $\Delta q = v(1-q) - uq = -(u+v)(q-\hat{q})$ where v is the mutation rate to the allele in question, u is the rate of mutation from it and $\hat{q}(=v/(u+v))$ is the value of q at equilibrium, which is the same in this case as $\bar{q}$ the mean value of q. This is similar in form to the expression for the effects of long range dispersal: $\Delta q = -m(q-q_t)$.

If both processes are occurring, the expressions merely need to be added:

$$\Delta q = v(1-q) - uq - m(q-q_t) = -(m+u+v)(q-\hat{q}) \tag{49}$$

where

$$\hat{q} = \bar{q} = (mq_t + v)/(m+u+v)$$

for a local population in which $\bar{q}$ is not necessarily the same as gene frequency for the whole species (q_t), since other factors may be at work in other localities. The long time distribution for such a population is approximately

$$\phi(q) = Cq^{4N(mq_t+v)-1}(1-q)^{4N[m(1-q_t)+u]-1} \tag{50}$$

$$\sigma_q^2 = \bar{q}(1-\bar{q})/[4N(m+u+v)+1]. \tag{51}$$

If conditions are the same in all islands, $\bar{q} = q_t = v/(u+v)$ and the variance $q_t(1-q_t)/[4N(m+u+v)+1]$ is not only the long time variance for a single island but also the variance of q, at any time, among the islands.

The variance of subpopulations (inbreeding coefficient F_i) in a total relative to which the inbreeding coefficient is F_t has been given (43, 48) as $q_t(1-q_t)[F_t-F_i]/[1-F_i]$ applicable to any case, including both the island model and that of a continuous population with only short range dispersal. The effective value of the immigration index in the latter may be obtained by equating with the expression for σ_q^2 given in (51).

$$m = [1-F_t]/[4N(F_t-F_i)]. \tag{52}$$

At first sight it might appear that the rate of change of gene frequency in cases in which there is long range dispersal and reversible mutation (joint coefficient m_1) in addition to predominant short range dispersal (coefficient m_2 from (52)) might be obtained by simply adding the contributions from these sources as calculated from their effects by themselves. This, however, overlooks the likelihood of an important interaction effect. It is necessary to go back to the formula for the correlation between uniting gametes (18) and determine how it is affected by mutation and long range dispersal.

Assume that the proportion m_1 of the gametes represent a random sample from the whole species. The identity of the theories of long range dispersal and mutation make it possible to let m_1 here represent $(m+u+v)$ of preceding formulae. Cases in which one or both of the uniting gametes are included in this proportion make no contribution to the correlation between uniting gametes. The proportion which makes a contribution is $(1-m_1)^2$. Equations (18) are accordingly to be modified as follows:

$$(53)\qquad \begin{cases} F = (1-m_1)^2\left[\frac{1}{N}\left(\frac{1+F}{2}\right)+\left(\frac{N-1}{N}\right)F_2'\right] \\ F_2' = (1-m_1)^2\left[\frac{1}{2N}\left(\frac{1+F''}{2}\right)+\frac{2N-1}{2N}F_3''\right] \text{ etc.} \end{cases}$$

Again primes may be dropped, if the same situation has held for a long time.

$$(54)\qquad F = \left[\frac{1+F}{2N}\right]\left[(1-m_1)^2+\frac{(1-m_1)^4}{2}\left(\frac{N-1}{N}\right)+\frac{(1-m_1)^6}{3}\left(\frac{N-1}{N}\right)\left(\frac{2N-1}{2N}\right)\cdots\right]$$

$$(55)\qquad t = C(1-m_1)^{2x}[x-(1/2N)]^{-(N+1)/N}$$

$$(56)\qquad \sum_{K_1}^{K_t-1} t = C\int_{K_1-0.5}^{K_2-0.5}(1-m_1)^{2x}[x-(1/2N)]^{-(N+1)/N}dx \text{ approximately.}$$

This is a less convenient expression than obtained where $m_1=0$, but approximate values can be obtained by taking values of K at short enough intervals, finding

$$\left[\int_{K_1-0.5}^{K_2-0.5}(1-m_1)^{2x}dx\right]\left[C\int_{K_1-0.5}^{K_2-0.5}[x-(1/2N)]^{-(N+1)/N}dx\right]$$

and correcting according to the percentage error where both factors are of the form $[\int_0^1 e^{Kx}dx]$ with the K's chosen so as to give the same ratios of terminal ordinates.

$$(57)\qquad F = \sum_1^{K-1} t \Big/ \left[2N - \sum_1^{K-1} t\right].$$

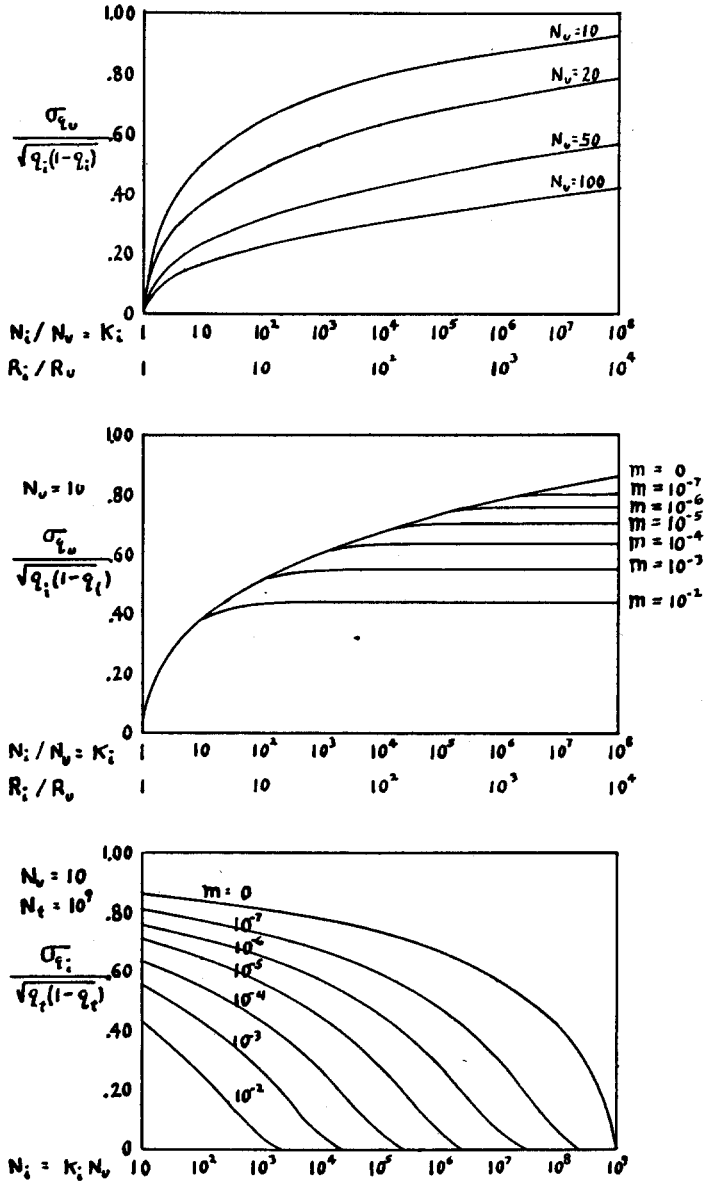


FIGURE 7 (top).—Similar to figure 1 except that exclusive uniparental reproduction is assumed. N_u is the population size of the group from which the parents of adjacent individuals are drawn at random and thus measures the extent of dispersal. The curves show the variability of gene frequencies (q_u) of such unit groups within areas up to 10^4 times their radius (R_i/R_u) or 10^8 times their population (N_i/N_u). Each curve applies to a particular extent of short range dispersal.

FIGURE 8 (middle).—The effect of occasional long range dispersal or mutation (rates up to $m = 10^{-2}$) on the variability of gene frequency of random breeding units of size $N_u = 10$ within areas

Figure 8 shows how $\sigma_{q_u}/\sqrt{q_i(1-q_i)}=\sqrt{F_i}$, for parental populations of size $N_u=10$, rises with the size of the population ($N_i=KN_u$) in the presence of random replacement (m_1) in the proportions 10^{-7} to 10^{-2}. The variability of the unit population is substantially the same as if there were no mutation or long range dispersal if N_i is less than $1/m_1$, but rather abruptly approaches a limit in larger populations. Instead of approaching 1 as when $m_1=0$, $\sqrt{F_i}$ approaches 0.81 if $m_1=10^{-7}$, 0.76 if $m_1=10^{-6}$, 0.70 if $m_1=10^{-5}$, 0.63 if $m_1=10^{-4}$, 0.55 if $m_1=10^{-3}$, and 0.44 if $m_1=10^{-2}$.

Figure 9 shows how the variability of subpopulations of any size N_i within a total population of size 10^9 is affected by the value of m_1. It is again assumed that short range dispersal is such as to give $N_u=10$. There is very little differentiation in this case of subpopulations larger than $30/m_1$. It is clear that it requires only a small amount of long range dispersal or mutation to prevent the differentiation of large populations.

The amount of differentiation of populations, that are a given multiple (K_i) of the unit population, falls off rapidly with increase of N_u. But the multiple beyond which differentiation virtually ceases is largely controlled by the factor $(1-m_1)^2$ and is thus nearly the same for all values of N_u under which there is any appreciable differentiation at any level. The value of $(1-m_1)^{2x}$ is reduced to approximately 10 per cent of its value each $\log_e 10/2m_1$ generations (assuming m_1 to be small).

Among populations of a given absolute size (N_i) there is, therefore, a certain range of dispersal (determining N_u) that is most favorable to differentiation in a continuous population. On the one hand, if the range of dispersal is such that N_u is larger than 1000, there is very little differentiation, but on the other hand, if N_u is so small that there are more than $3/m_1$ random breeding units in the population under consideration, there is also virtually no differentiation.

Linear continuity may be treated similarly, by multiplying the terms of (32) by $(1-m_1)^{2x}$.

Under exclusive uniparental reproduction, the chance that an individual is derived from the parental population without mutation is $(1-m_1)$, instead of $(1-m_1)^2$. Each term in the series $\sum t$ is accordingly to be multiplied by $(1-m_1)^x$.

The formula for the distribution of gene frequencies among subpopulations of a given size, N_i, in the total N_t, may be written approximately as follows:

$$(58) \qquad \phi(q) = Cq^{[(1-F_t)/(F_t-F_i)]q_t-1}(1-q)^{[(1-F_t)/(F_t-F_j)](1-q_t)-1}.$$

Here the F's incorporate the effects of mutation and long range dispersal as

---

up to 10^4 times their radius or 10^8 times their population size. The highest curve (m=0) is the same as the highest curve in figure 1.

FIGURE 9 (bottom).—The effect of occasional long range dispersal or mutation (rates up to 10^{-2}) on the variability of gene frequencies of populations of any size, N_i, within a region with a population of a given size, $N_t=10^9$. The random breeding unit is assumed to be $N_u=10$. The highest curve (m=0) is the same as that in figure 3.

well as of short range dispersal. This distribution has the mean q_t and the variance $q_t(1-q_t)[F_t-F_i]/[1-F_t]$ derived above. It differs considerably from the distribution

$$\phi(q) = Cq^{4N(m_1+m_2)q_t-1}(1-q)^{4N(m_1+m_2)(1-q_t)-1} \tag{59}$$

if m_2 is the estimate of effective m from (52) based on the value of F_i and F_t under short range dispersal in the absence of other factors. It is legitimate, however, if m_1 is known, to write $\phi(q)$ in the form of (59) with the understanding that m_2 measures the effect of short range dispersal in the presence of the other factors measured by m_1 with full allowance for the interaction effect. Indeed this seems to be the only practicable method to use in analyzing data from actual populations in view of the fact that no ideal model such as area or linear continuity is likely to be exactly realized.

THE EFFECTIVE SIZE OF INBRED POPULATIONS

The effect of inbreeding on the effective size of populations is a matter that requires some consideration. Size of population enters into the formulae for the distribution of gene frequencies principally through the sampling variance which is $q(1-q)/2N$ in a random breeding diploid population. Assume that individuals have an inbreeding coefficient F_i relative to an island population. It makes a difference in the sampling variance whether this is due to mating of relatives, not resulting in any territorial subdivision, or whether it is due to partial isolation of subdivisions that breed at random within themselves. In the former case, the increased frequency of homozygotes causes an increased sampling variance of the whole island. If there were nothing but homozygotes, $(q_iAA+(1-q_i)aa)$, as under long continued self-fertilization, the sampling variance would be $q_i(1-q_i)/N_i$, twice that under random mating. With random bred and inbred components in the array of equations (9) in the proportions $(1-F_i)$ to F_i, the sampling variance would be the weighted average.

$$\begin{aligned}\sigma^2_{\delta q_i} &= (1-F_i)q_i(1-q_i)/2N_i + F_iq_i(1-q_i)/N_i \\ &= q_i(1-q_i)(1+F_i)/2N_i.\end{aligned} \tag{60}$$

If on the other hand, the island population is subdivided into partially isolated groups that breed at random within themselves and if each group tends to maintain its numbers (that is, there is no intergroup selection) the sampling variance of the total island population is *less* than if there were random mating throughout. In each subgroup, the sampling variance is $q_u'(1-q_u')/2N_u$, average $\sigma^2_{\delta q_u}=\sum_1^K q_u'(1-q_u')/2N_uK$. The sampling variance for the mean gene frequency of the island would be $\sigma^2_{\delta q_i}=\sigma^2_{\delta q_u}/K=\sum_1^K q_u'(1-q_u')/2N_iK$ if $N_i=KN_u$. But from (10) $y_i=2\sum_1^K q_u'(1-q_u')/K$. Thus $\sigma^2_{\delta q_i}=y_i/4N_i$. From (9) $y_i=2q_i(1-q_i)(1-F_i)$ giving

$$\sigma^2_{\delta q_i} = q_i(1-q_i)(1-F_i)/2N_i. \tag{61}$$

The situation in an arbitrarily delimited region in a continuum resembles the second. Effective N in such a formula as (59) is thus $KN_u/(1-F_i)$.

COMPLICATING EFFECTS OF SELECTION

Consider next the complications introduced by selection. The effects of various kinds of selection on gene frequency (contributions to Δq) and the form taken by $\phi(q)$ on substitution in (1) have been discussed in previous papers (WRIGHT 1931, 1942). These are applicable directly to the island model. The case of arbitrarily delimited portions of a continuum can be treated in the same way, but if so, m_2 of formula (59) includes the interaction effect of selection as well as of mutation (and of long range dispersal if this can be distinguished from the short range dispersal). The index m_2 is to be interpreted as the effective amount of replacement of the subpopulations in question by representatives of the species as a whole under the conditions of mutation and selection that actually hold. As noted in connection with the complications introduced by mutation and long range dispersal, this seems to be the most practicable method of dealing with concrete data. It is important, however, to determine the theoretical relations between the values of m among subdivisions of different sizes under various ideal population structures.

For such theoretical consideration of the interaction of selection with the effects of short range dispersal, it is necessary to return to the derivation of F by path coefficients (18) in analogy with the treatment of the complications due to mutation and long range dispersal (53). But in attempting to carry out the analogy we encounter a serious difficulty.

Long range dispersal (by definition) and mutation may be treated as introducing a random admixture into the local population in constant proportion m_1. Selection may also be treated as introducing a certain random admixture, but it is not in constant proportion. The amount of such admixture in the case of mutation and long range dispersal may be represented as

$$(62) \qquad [-\Delta q/(q-\hat{q})] = (m+u+v) = m_4.$$

This formula may be applied where Δq also involves selection pressure. Consider the case of a balance between opposing pressures of mutation and selection in the simplest case, that of no dominance, and assume that the same situation holds throughout the species.

$$(63) \quad \Delta q = v(1-q) - sq(1-q) = -s(1-q)(q-\hat{q}) \quad \text{where} \quad \hat{q} = v/s$$

$$(64) \quad m_1 = [-\Delta q/(q-\hat{q}] = s(1-q).$$

The joint effect of mutation and selection in this case is equivalent to immigration of a random sample, but to an extent that is a function of the local gene frequency. A rough idea of the effect may be obtained by substituting $\hat{q}$ for q and treating $m_1 = s(1-\hat{q}) = s-v$ as a constant. If s is much larger than v we may indeed simply take $m_1 = s$ and use $(1-s)^2$ in place of $(1-m_1)^2$ in the theory developed for mutation and long range dispersal. Inspection of figures 8 and 9 shows how selection of this sort interferes with the differentiation that

would occur within the continuous population under the specified conditions if there were no complication of this sort.

As another example consider the case of selection against both of two homozygotes. Representing the relative selective values of AA, Aa and aa by $1-s_{AA}$, 1 and $1-s_{aa}$ respectively

$$\Delta q = -(s_{AA} + s_{aa})q(1 - q)(q - \hat{q}) \tag{65}$$

where $$\hat{q} = s_{aa}/(s_{AA} + s_{aa})$$

$$m_1 = (s_{AA} + s_{aa})q(1 - q). \tag{66}$$

While selection does nothing to local populations that have become fixed and the equivalent immigration index m_1 is accordingly 0 if q is either 0 or 1, the average value may well be such as to severely restrict differentiation of even rather small subdivisions of a continuous population. Again a rough idea of the effect may be obtained by substituting $\hat{q}$ for q. It should be noted that if there are numerous alleles and selection for heterosis is general, selection tends to increase differentiation.

In a recent paper (WRIGHT, DOBZHANSKY and HOVANITZ 1942) an attempt was made to interpret the frequencies of lethals in a continuous population of *Drosophila pseudoobscura* on Mt. San Jacinto. The following formula was arrived at for the rate of change of the frequency of a typical lethal gene.

$$\Delta q = \bar{v}(1 - q) - m(q - \bar{q}) - q(\bar{s} + F) - q^2(1 - 3\bar{s} - 2F) \tag{67}$$

where $\bar{v}$ is the mean mutation rate per generation, $\bar{s}$ the mean selective disadvantage of heterozygotes, $\bar{q}$ the mean gene frequency, F the inbreeding coefficient, and m the effective immigration coefficient of the territory under consideration. It was shown that approximately the same variance of gene frequencies was reached by replacing the above expression by one in which the component of Δq, measuring the tendency toward increase of gene frequency—namely, $(\bar{v}+m\bar{q})\ (1-q)$, is balanced by the linear expression that gives the same mean as the correct expression namely, $-(\bar{v}+m\bar{q})\ (1-\bar{q})q/\bar{q}$

$$\Delta q = -(m + \bar{v}/\bar{q})(q - \bar{q}) \text{ approximately} \tag{68}$$

$$m_1 = (m + \bar{v}/\bar{q}) \text{ approximately.} \tag{69}$$

DIFFERENTIATION OF SUBDIVISIONS BY SELECTION

If selection acts differently in different regions, it is obvious that none of the formulae given here apply to the distribution of values q among these regions, but only to the long term distribution within single ones. As a basis for discussion consider the following simple case, which refers to rate of change of gene frequencies in an island as affected by the local conditions of selection measured by s (assuming no dominance) and the amount of immigration measured by m (WRIGHT 1931, 1940).

$$\Delta q = sq(1 - q) - m(q - q_t). \tag{70}$$

In a local population in which s (whether plus or minus) is smaller in absolute value than m, gene frequency can depart only slightly from the average of the species ($\dot{q}=q_t+(s/m)q_t(1-q_t)$) approximately. Crossbreeding here swamps the tendency toward selective differentiation. On the other hand, local gene frequency tends to be dominated by the local conditions of selection in populations in which s is larger than m in absolute value $\dot{q}=1-(m/s)(1-q_t)$ or $\dot{q}=(-m/s)q_t$ approximately, depending on whether s is positive or negative.

The effectiveness of selection here is not related directly to the size of the island population. However, there is likely to be indirect relationship. This may be illustrated by considering three situations.

First, consider islands with various populations but the same absolute amount of immigration (as might well be the case if the areas are the same but population densities differ). Among such islands, Nm is constant. All have the same amount of nonadaptive differentiation (measured by $1/(4Nm+1)$ but a given selection pressure is more effective on the islands with larger population (and hence smaller m) than among those with smaller populations.

A second situation is that in which size of population is proportional to area and the number of immigrants is proportional to the extent of boundary ($Nm \propto \sqrt{N}$). Here there is more nonadaptive differentiation on the smaller islands and more adaptive differentiation of the larger ones, although the latter effect is less marked than in the preceding case.

Finally, if both size of population and amount of immigration are proportional to the area (m constant), there is markedly more nonadaptive differentiation on the smaller islands but no relationship between adaptive differentiation and size of population.

Summing up, any sort of differentiation is favored by small m, but the large populations tend on the whole to exhibit predominant adaptive differentiation, while the smaller ones exhibit predominantly nonadaptive differentiation.

The situation in a continuous population is similar in that nonadaptive differentiation should be most conspicuous locally and adaptive differentiation among larger subdivisions. The most significant thing, however, given a certain amount of differential action of selection, is the size of the random breeding unit. If this is large—for example, over 1000, very little nonadaptive differentiation is to be expected and only rather strong differences in the action of selection avoid swamping. If on the other hand, there is only short range dispersal—for example, $N_u=10$, large regions tend to become adaptively differentiated under the influence of slight differences in selection, and superimposed on this should be a large amount of nonadaptive differentiation of small regions. The maximum amount of nonadaptive differentiation among populations of a given size however, is not found with the smallest N_u, but at a certain optimum value.

If a population spreads over a large territory in which the environmental conditions are substantially uniform, there would primarily be only nonadaptive differentiation, the amount depending on the value of m or of N_u depend-

ing on the model that is most appropriate. With such differentiation occurring simultaneously but more or less independently in all series of alleles, each locality would have a slightly different genetic system from every other locality. These systems may be expected to differ in their success in meeting the environmental conditions. Among those which are relatively successful, adaptation is likely to have a slightly different basis in each case. The populations with such systems tend to become denser and to send out more than their share of migrants and thus enlarge in extent. Each would tend to perfect the line of adaptation on which it had started. Thus permanent differential action of selection would soon be brought into play in spite of the postulated uniformity of the conditions.

The expansion of centers of population characterized by certain genetic systems and contraction of those characterized by other systems is the process of intergroup selection referred to in the opening paragraph. The genetic system, including its state of heterogeneity as well as its central type, is the basis of selection instead of merely the net favorable or unfavorable effect of each single gene, which is the only basis for selection under panmixia; or the single genotype, which is the most probable basis under self-fertilization or vegetative multiplication. The present analysis indicates that this most favorable basis for evolutionary advance of the species as a whole may be present under certain conditions in a continuous population as well as in one consisting of partially isolated groups.

SUMMARY

Formulae are derived relating the variance of the gene frequencies of subgroups (σ_q^2) to the effective population number of these (N), the effective proportion of replacement per generation by immigrants (m), the inbreeding coefficient of individuals relative to the total population (F), and the mean gene frequency in the latter (q_t). Thus $\sigma_q^2 = q_t(1-q_t)/[2N-(2N-1)(1-m^2)] = q_t(1-q_t)F/(1-m)^2$ including the immediate sampling variance, but $\sigma_q^2 = q_t(1-q_t)F$ excluding this.

The effect of isolation by distance in a continuous population in which there is only short range dispersal in each generation is worked out on the hypothesis that the parents of any individual may be treated as if they were taken at random from a group of a certain size (N). It is shown that the inbreeding coefficient of individuals in such a population relative to a population of size KN can be expressed in the form $F = \sum_1^{K-1} t/[2N - \sum_1^{K-1} t]$ where $\sum t$ is the sum of a series of terms in which $t_1 = 1$ and $t_x = t_{(x-1)}[(x-1)N-1]/xN$ or approximately $C[x-(1/2N)]^{-(N+1)/N}$ where C is a constant close to 1. The value of $\sum_1^{K-1} t$ can be obtained sufficiently accurately by actual calculation of the first few terms, supplemented by the approximate formula

$$\sum_{K_1}^{K_2-1} t = CN\left[\left(K_1 - \frac{1}{2} - \frac{1}{2N}\right)^{-1/N} - \left(K_2 - \frac{1}{2} - \frac{1}{2N}\right)^{-1/N}\right]$$

for later terms. The limiting value $\sum_1^{\infty} t$ is N. Thus F approaches 1 in an indefinitely large continuous population.

The preceding results apply to area continuity. With continuity in a linear range (for example, shore line), $F=\sum t/[2N-\sum t]$ as above, $t_1=1$ but $t_x=t_{(x-1)}[N\sqrt{x-1}-1]/N\sqrt{x}$ or approximately $Ce^{-2\sqrt{x}/N}[\sqrt{x}-(1/2N)]^{-(N^2+1)/N^2}$.

In a continuous population with exclusive uniparental reproduction, the correlation between adjacent individuals is of the form $E=\sum t/N$ where $\sum t$ is the same as above for area or for linear continuity as the case may be.

The variance of gene frequencies in subdivisions of any size, N_i, within a more comprehensive population N_t is given by the formula $\sigma^2_{i.t}=q_t(1-q_t)[F_t-F_i]/[1-F_i]$ where F_i and F_t are the inbreeding coefficients relative to the populations of size N_i and N_t, respectively.

It is shown that in the absence of disturbing factors, short range dispersal (N less than 100 in the case of area continuity) leads to considerable differentiation not only among small subdivisions but also of large ones. Values of N greater than 10,000 give results substantially equivalent to panmixia throughout a range of any conceivable size. With linear continuity, there is enormously more differentiation than with area continuity. There is somewhat more differentiation under uniparental than under biparental reproduction.

Recurrent mutation, long range dispersal and selection are factors that restrict greatly the amount of random differentiation of large (but not small) subdivisions of a continuous population. A term $(1-m_1)^{2x}$ under biparental, $(1-m_1)^x$ under uniparental, reproduction is introduced into the expressions for t referred to above. In this $m_1=[-\Delta q/(q-q_t)]$ where Δq is the rate of change of gene frequency (q) which such factors tend to bring about.

The effective size of a population characterized by the inbreeding coefficient F depends on whether F is due to a tendency toward mating of relatives not associated with territorial subdivision, or to such subdivision. In the former case the sampling variance is $\sigma^2_{\delta q}=q(1-q)(1+F)/2N$, in the latter, $q(1-q)(1-F)/2N$, in contrast with $q(1-q)/2N$ in a random bred population.

If different regions are subject to different conditions of selection, the amounts of both adaptive and nonadaptive differentiation depend on the smallness of m (if subdivision into partially isolated "islands") or of N, size of the random breeding unit (if a continuous distribution). If these are sufficiently large there is no appreciable differentiation of either sort; if sufficiently small there is predominantly adaptive differentiation of the larger subdivisions with predominantly nonadaptive differentiation of smaller subdivisions superimposed on this. Even under uniform environmental conditions, random differentiation tends to create different adaptive trends in different regions and a process of intergroup selection, based on gene systems as wholes, that presents the most favorable conditions for adaptive advance of the species.

LITERATURE CITED

DOBZHANSKY, TH., and S. WRIGHT, 1941 Genetics of natural populations. V. Relations between mutation rate and accumulation of lethals in populations of *Drosophila pseudoobscura*. Genetics **26**: 23–51.

THOMPSON, D. H., 1931 Variation in fishes as a function of distance. Trans. Illinois Acad. Sci. **23**: 276–281.

WRIGHT, S., 1921 Systems of mating. Genetics **6**: 111–178.

1922a The effects of inbreeding and crossbreeding on guinea pigs. III. Crosses between highly inbred families. Bull. U. S. Dept. Agric. No. **1121**.

1922b Coefficients of inbreeding and relationship. Amer. Nat. **56**: 330–338.

1929 The evolution of dominance. Amer. Nat. **58**: 1–5.

1931 Evolution in mendelian populations. Genetics **16**: 97–159.

1938 Size of population and breeding structure in relation to evolution. Science **87**: 430–431.

1940 Breeding structure of populations in relation to speciation. Amer. Nat. **74**: 232–248.

1942 Statistical genetics and evolution. Bull. Amer. Math. Soc. **48**: 223–246.

WRIGHT, S., TH. DOBZHANSKY, and W. HOVANITZ, 1942 Genetics of natural populations. VII. The allelism of lethals in the third chromosome of *Drosophila pseudoobscura*. Genetics **27**: 363–394.