ISOLATION BY DISTANCE UNDER DIVERSE SYSTEMS OF MATING

SEWALL WRIGHT
The University of Chicago

Received August 3, 1945

INTRODUCTION

The effects of restricted dispersion on the genetic properties of a continuous population have been treated mathematically in a previous paper (Wright 1943a). The conclusions have been applied to the interpretation of observed local variability in a population of a plant, Linanthus Parryi (Wright 1943b) and in one of an animal, Drosophila pseudoobscura (Dobzhansky and Wright 1943). The mathematical treatment was based on the assumption of completely random union of gametes within each neighborhood and thus would rarely be strictly applicable to actual cases. The purposes of the present paper are to compare the effects of various systems of mating within continuous populations, and to present a more accurate method than before for estimating from data an important theoretical quantity, N, the effective size of population of a “neighborhood” in the sense discussed below.

It was postulated in the previous paper that a population of uniform density occupies either an indefinitely large area (area continuity) or a strip of indefinitely great length but of such narrow width that dispersion occurs across it within a single generation (linear continuity). Both uniparental and biparental reproduction were considered. In the former case, it was postulated that the locations of parents at some phase of the life cycle are distributed, relative to the corresponding locations of their progeny, according to a normal probability curve with a standard deviation \( \sigma \), if there is linear continuity, or in a bivariate normal distribution with standard deviation \( \sigma \), for both \( x \) and \( y \) coordinates of the parental locations (relative to their progeny) if there is area continuity. In the case of biparental reproduction the locations of the parents were assumed to be uncorrelated with each other. This is assumed to hold in the present paper except as qualified by self fertilization or brother-sister mating in excess of random.

A term is needed to designate the local population of which the parents may be considered as representative. Various terms were used in the preceding paper. “Panmictic unit” applies only if there is completely random mating locally and is thus not sufficiently general. “Parental group” is better but is somewhat awkward. An essential property of the population in question is that the individuals are neighbors in the sense that their gametes may come together. The term “neighborhood” is thus an appropriate one for this important unit.

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.

Genetics 31: 39 January 1946
THE POPULATION NUMBERS OF NEIGHBORHOODS

To obtain numerical results, it was assumed in the previous paper that
the mating system is equivalent to the random union of gametes produced by
a population of \( N \) monoecious individuals, thus involving the proportion \( r/N \)
of self fertilization. This raises the question of the relation of \( N \) to the density
of the population of mature individuals and the standard deviation of the
parental distribution. Consider first the situation in a population of uniform
density along a linear range. Assume that the location points of parents (\( x \))
relative to offspring may be represented sufficiently accurately by the normal
distribution \( y = (1/\sqrt{2\pi}) \exp(-x^2/2\sigma^2) \). Let \( n \) be the number of potential
parents in a strip of length \( 2\sigma \). The density per unit distance is then \( d = n/2\sigma \).
The average length of the territory occupied by each individual may be written
as follows

\[
\frac{1}{d} = \frac{2\sigma}{n} = \int_{x_i - \sigma/n}^{x_i + \sigma/n} \text{dx}
\]

The chance that a particular gamete comes from a particular member of the
parental generation at distance \( x_i \) is \( y_i/d \). The chance that two uniting gametes
came from the same individual is thus \( \Sigma (y_i/d)^2 \) where the summation applies
to all individuals in the parental distribution. This is the expression to be
equated to \( r/N \). Making use of (1):

\[
\frac{1}{N} = \frac{2\sigma}{n} \int_{-\infty}^{+\infty} \frac{y_i^2}{n} \int_{x_i - \sigma/n}^{x_i + \sigma/n} \text{dx} \approx \frac{2\sigma}{n} \int_{-\infty}^{+\infty} y^2 \text{dx}
\]

\[
= \frac{2\sigma}{n} \int_{-\infty}^{+\infty} \left[ e^{-x^2/2\sigma^2}/2\pi\sigma^2 \right] \text{dx}
\]

\[
= \frac{1}{n\sqrt{\pi}} \int_{-\infty}^{+\infty} \left[ e^{-x^2/(2\sqrt{2}\sigma)^2}/\sqrt{2\pi}(\sigma/\sqrt{2}) \right] \text{dx}
\]

\[
= 1/\sqrt{\pi} \quad \quad (2)
\]

\[
N = \sqrt{\pi} n = 2\sqrt{\pi} \sigma d = 3.545\sigma d
\]

Effective \( N \) is thus equivalent to the number of reproducing individuals
along a strip 3.5 \( \sigma \) long. About 92.4 per cent of the actual parents of individuals
should fall within the range \( \pm \sqrt{\pi} \sigma d \).

This method of relating \( N \) to \( \sigma \) and \( d \) can be extended to populations that
are continuous over an area. Let \( y = [1/2\pi\sigma^2] \exp[-(x_1^2+x_2^2)/2\sigma^2] \) be the distri-
bution of the locations of parents relative to those of their offspring. Let \( n \)
be the number of reproducing individuals in a square, \( 2\sigma \) on a side. The density
per unit area is \( d = n/4\sigma^2 \). The average area occupied by an individual (co-
dinates \( x_1, x_2 \)) may be written as follows:

\[
\frac{1}{d} = \frac{4\sigma^2}{n} \int_{x_1 - \sigma/n}^{x_1 + \sigma/n} \int_{x_2 - \sigma/n}^{x_2 + \sigma/n} \text{dx}_1\text{dx}_2
\]

\[
(4)
\]
The chance that two uniting gametes come from the same individual may be written as follows, using (4):

\[
\frac{1}{N} = \frac{4\sigma^2}{n} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y^2 dx_1 dx_2 = \frac{1}{n\pi} \quad (5)
\]

\[
N = \pi n = 4\pi \sigma^2 d = 12.566\sigma^2 d \quad (6)
\]

Effective N is equivalent to the number of reproducing individuals in a circle of radius \(2\sigma\). Such a circle would include 86.5 per cent of the parents of individuals at the center.

In an analysis of data from natural populations of *Drosophila pseudoobscura* (Dobzhansky and Wright 1943), effective N was taken as equivalent to the breeding population in a circle of such radius \(\sqrt{2}\sigma\) that the volume of a cylinder erected upon it would equal that of the bivariate normal distribution with \(\sigma_{x_1} = \sigma_{x_2} = \sigma\) and central ordinate equal to the height of the cylinder. This does not do justice to the dispersive effect of the more extreme parent-offspring distances. The estimate of effective N for a given parental standard deviation should be just twice that arrived at in this way. As the estimate in this application could at best be considered as giving only the order of magnitude, the effect of this correction is not very important.

The distribution of location points of grandparents relative to those of their grandchildren is compounded of their distributions about the location points of the parents and the similar distribution of these about the individuals in question. Thus the variance of the grandparental distribution is twice that of the parental distribution and its standard deviation is \(\sqrt{2}\sigma\). The standard deviation for ancestors of generation K is similarly \(\sqrt{K}\sigma\). The effective size of the population from which ancestors of generation K are taken is \(\sqrt{K}N\) if the range is linear but KN if it extends in all directions.

**CASE I. MONOECIOUS POPULATIONS WITH EQUAL DISPERSION FROM\nMALE AND FEMALE PARENTS**

This case was considered in the previous paper, subject to the postulate that union of gametes from the neighborhood is completely random (apart from the differential weighting of probabilities by distances of parental from offspring location). This implies self fertilization at the rate \(1/N\), a highly arbitrary postulate. It is desirable to determine the consequence under any specified percentage of self fertilization.

Let \(N\) be the effective population number in neighborhoods.
Let \(q\) be the gene frequency in a neighborhood.
Let \(h\) be the proportion of self fertilization.

The following quantities are relative to the population from which ancestors of generation K were drawn.
Let \(N_K\) be the effective population number.
Let \(q_K\) be the average gene frequency.
Let \(S\) be the correlation between gametes from the same individual.
Let $D$ be the correlation between gametes from different individuals from the neighborhood, which unite or contribute to adjacent individuals. Let $D_X$ be that between gametes from different individuals from the $X$th ancestral generation—weighted according to their likelihood of contributing to the $(X-1)$st ancestral generation.

Let $E$ be the correlation between gametes which contribute to adjacent individuals. Let $E_X$ be the correlation between gametes from the $X$th generation which contribute to the $(X-1)$st ancestral generation.

Let $F$ be the correlation between uniting gametes. This is the inbreeding coefficient relative to the population of size $N_K$. It must be distinguished from the inbreeding coefficient relative to the neighborhood, which is $\frac{[h-(1/N)]}{[2-h-(1/N)]}$ from (16) below ($K=1$, $\Sigma t_A = 0$) or from (16) and (18) by the formula $(F-E)/(1-E)$, $\Sigma t_A$ with $(K-1)$ terms.

Let $a(=\sqrt{1/2(1+F)}$, Wright 1921) be the path coefficient relating zygote to one of the gametes which united in its production.

Let $b(=\sqrt{(1+F')/2}$, where $F'$ is the value of $F$ in the preceding generation, Wright 1921) be the path coefficient relating gamete to the zygote which produced it. The compound path coefficient relating a gamete to one of the two from which it traces a generation earlier is $ba'=1/2$ where $a'$ is the value of $a$ in the preceding generation.

It will be assumed that the same population structure has continued indefinitely. Under this assumption primes may be dropped.

The following equations may be written:

$$E = \frac{1}{N} S + \frac{N-1}{N} D \quad \text{by definition} \quad (7)$$

$$F = hS + (1-h)D \quad \text{by definition} \quad (8)$$

$$S = b^2 = (1 + F)/2 \quad \text{from fig. 1} \quad (9)$$

$$D = 4b^2a^2E_2 = E_2 \quad \text{from fig. 1} \quad (10)$$

Similarly

$$D_X = E_{X+1} \quad \text{by analogy with (10)} \quad (11)$$

Consider first the case of area continuity in which $N_K = KN$

$$E_X = \frac{1}{XN} S + \left[ 1 - \frac{1}{XN} \right] E_{X+1} \quad (12)$$
ISOLATION BY DISTANCE

Thus \( E = \frac{1}{N} S + \frac{N - 1}{N} E_2 \)

\[
E_2 = \frac{1}{2N} S + \frac{2N - 1}{2N} E_3
\]

\[
E_3 = \frac{1}{3N} S + \frac{3N - 1}{3N} E_4
\]

\( E_K = 0 \)

\[
E = \frac{S}{N} \left[ 1 + \frac{1}{2} \left( \frac{N - 1}{N} \right) + \frac{1}{3} \left( \frac{N - 1}{N} \right) \left( \frac{2N - 1}{2N} \right) \right] \quad (13)
\]

The series in brackets, with \( K-1 \) terms, will be designated \( \Sigma t_A \). Formulae for its approximate evaluation were given in the preceding paper (formulae 22 and 27). Its limiting value for \( K = \infty \) is \( N \).

\( E = S \sum t_A / N \) \quad from (13) (14)

\( D = S(\sum t_A - i)/(N - i) \) \quad from (7) and (14) (15)

\( F = [h(N - i) + (1 - h)(\sum t_A - i)] / [(N - i) + (1 - h)(N - \sum t_A)] \) \quad from (8), (9), (15) (16)

\( S = (N - i) / [(N - i) + (1 - h)(N - \sum t_A)] \) \quad from (9) and (16) (17)

\( E = (N - i) \sum t_A / N [(N - i) + (1 - h)(N - \sum t_A)] \) \quad from (14) and (17) (18)

\( D = (\sum t_A - i) / [(N - i) + (1 - h)(N - \sum t_A)] \) \quad from (15) and (17) (19)

The amount of differentiation among neighborhoods within the comprehensive population \( N_K \) can be found from a formula given in Wright 1943a. It was shown there that if there is random mating within each of a number of populations, the variance of the values of \( q \) of these populations is given by the formula

\[
\sigma_q^2 = q_K (1 - q_K) F
\]

using the symbols defined here. In the present paper, random mating within neighborhoods is not assumed. However, \( E \) is defined in such a way that it would be the coefficient of inbreeding (\( F \)) if there were random mating in the last generation.

\[
\sigma_q^2 = q_K (1 - q_K) E \quad (20)
\]

This formula may be derived in another way. The allele present in a gamete taken at random from a neighborhood may be considered to be determined jointly by the gene frequency \( q \) of the latter and an uncorrelated deviation. The degree of determination is the square of the correlation between gametic value and \( q \). It gives the portion of the variance of gametes, \( q_K (1 - q_K) F \), for which differentiation of neighborhoods, measured by \( \sigma_q^2 \), is responsible. But the correlation \( (E) \) between pairs of gametes taken at random from neighborhoods is
also given by this same squared correlation. Thus \( E = \sigma_a^2 / q_k (1 - q_k) \) in agreement with (20).

If self fertilization occurs wholly at random at all times (\( h = 1/N \)), we have the case treated in the preceding paper (formula 21, \textit{Wright} 1943a).

\[
E = F = \frac{\sum t_A}{2N - \sum t_A} \quad \text{from (18), } h = 1/N \quad (21)
\]

If there is no self fertilization (\( h = 0 \)), the result differs little, unless \( N \) is very small.

\[
E = (N - 1) \frac{\sum t_A}{N(2N - \sum t_A - 1)} \quad \text{from (18), } h = 0 \quad (22)
\]

Under exclusive self fertilization (\( h = 1 \)) the amount of differentiation of neighborhoods is the same as that relative to two alternatives under any form of uniparental reproduction. The result agrees with formula (37) of the preceding paper.

\[
E = \frac{\sum t_A}{N} \quad \text{from (18), } h = 1 \quad (23)
\]

The correlation between genotypes from a neighborhood (assuming additive effects) is independent of \( h \) in Case 1. The quantity, to be called \( G \), is equal to \( D/b^2 \) as may be seen from figure 1.

\[
G = D/b^2 = (\sum t_A - 1)/(N - 1) \quad (24)
\]

Analogous results can be obtained for the case of linear continuity. As brought out in the previous paper, it is necessary merely to substitute the following series \( \sum t_L \) for \( \sum t_A \). The approach to the limiting value \( N \) as the number of terms is increased is much more rapid.

\[
\sum t_L = \left[ 1 + \frac{1}{\sqrt{2}} \left( \frac{N - 1}{N} \right) + \frac{1}{\sqrt{3}} \left( \frac{N - 1}{N} \right) \left( \frac{\sqrt{2} N - 1}{\sqrt{2} N} \right) \cdots \right] \quad (25)
\]

The conclusions reached on the relation of effective \( N \) to density of population (\( d \)) and standard deviation of parental distances (\( \sigma \)), assuming random mating within neighborhoods, may seem to require reconsideration in this case in which a specified amount (\( h \)) of self fertilization is assumed. The quantity \( 1/N \), however, may be defined as the amount of self fertilization that there would be if there were random union of gametes in the neighborhood in the last generation and is independent of the amount of self fertilization that there actually is or has been. The chance that two gametes that enter into the production of two adjacent individuals (instead of two uniting gametes) come from the same individual may be written \( 1/N = \Sigma (y/d)^2 \), leading to formulae (3) and (6).

**CASE 2. PERMANENT PAIRS**

The simplest system of mating with separate sexes is that in which reproduction is wholly by permanent pairs and the amount of dispersion of males and females is the same. The analysis is somewhat similar to that in the preceding case if the pair is treated as the unit and located at the point at which
pairing first occurs. The treatment of individuals as located at the points at
which they began their careers as fertilized eggs (or at any later phase) would
be complicated by a correlation between the parental localizations relative to
those of offspring, beyond that which may be due to a tendency to brother-
sister mating. As before, correlations are relative to a population of specified
size \((KN_p)\).

Let \(N_P(=N/2)\) be the number of pairs in a neighborhood.
Let \(h\) here be the proportion of brother-sister mating.
Let \(C\) be the correlation between gametes produced by siblings.
Let \(D\) be the correlation between gametes of mated or adjacent non siblings,
and \(D_X\) that between gametes of non siblings of ancestral generation \((X - 1)\)
relative to pairs.
Let \(E\) be the correlation between gametes of adjacent individuals and \(E_X\) be
that between gametes of individuals belonging to ancestral generation \((X - 1)\)
relative to pairs.
Let \(F\) be the correlation between uniting gametes (the inbreeding coeffi-
cient).

\[
C = b^2a^2(2b^2 + 2F) = \frac{1 + 3F}{4} \\
D = 4b^2a^2E_2 = E_2 \\
D_X = 4b^2a^2E_{x+1} = E_{x+1} \\
E = C/N_p + E_2(N_p - 1)/N_p \quad \text{by definition and (27)} \\
E_X = C/XN_p + E_{x+1}(XN_p - 1)/N_p \quad \text{by analogy} \\
E_K = 0 \quad \text{by definition of } KN_p \text{ as the population of reference} \\
E = C \sum t_A/N_p \quad \text{cf. (13), (14)} \\
D = (N_pE - C)/(N_p - 1) = C(\sum t_A - 1)/(N_p - 1) \quad \text{from (29), (32)(33)} \\
F = C[h(N_p^* - 1) + (1 - h)(\sum t_A - 1)]/(N_p - 1)
\]

From inspection of figure 2

\[
C = b^2a^2(2b^2 + 2F) = \frac{1 + 3F}{4} \\
D = 4b^2a^2E_2 = E_2 \\
D_X = 4b^2a^2E_{x+1} = E_{x+1} \\
E = C/N_p + E_2(N_p - 1)/N_p \quad \text{by definition and (27)} \\
E_X = C/XN_p + E_{x+1}(XN_p - 1)/N_p \quad \text{by analogy} \\
E_K = 0 \quad \text{by definition of } KN_p \text{ as the population of reference} \\
E = C \sum t_A/N_p \quad \text{cf. (13), (14)} \\
D = (N_pE - C)/(N_p - 1) = C(\sum t_A - 1)/(N_p - 1) \quad \text{from (29), (32)(33)} \\
F = C[h(N_p^* - 1) + (1 - h)(\sum t_A - 1)]/(N_p - 1)
\]
by definition and (33) (34)

\[ F = \frac{h(N_p - 1) + (1 - h)(\sum t_A - 1)}{\left[(N_p - 1) + 3(1 - h)(N_p - \sum t_A)\right]} \quad \text{from (26), (34) (35)} \]

\[ E = (N_p - 1)\sum t_A/[N_p[(N_p - 1) + 3(1 - h)(N_p - \sum t_A)]] \quad \text{from (32, 26, 35) (36)} \]

The amount of variation of gene frequency among neighborhoods is the same as that given in (20), \( q_2^2 = q_R(1 - q_R)E. \)

If brother-sister mating occurs at random at all times within neighborhoods, \( h = 1/N_p \)

\[ E = \sum t_A/[4N_p - 3\sum t_A] \quad \text{from (36), } h = 1/N_p \] (37)

A case which approximates that in a continuous human population of low mobility is that in which \( h = 0 \) (no brother-sister mating).

\[ E = (N_p - 1)\sum t_A/[N_p(4N_p - 3\sum t_A - 1)] \quad \text{from (36), } h = 0 \] (38)

With exclusive brother-sister mating \( (h = 1) \) all lines become homallelic, and the differentiation among neighborhoods becomes the same as under uniparental reproduction, except that pairs, instead of individuals, must be taken as the units.

\[ E = \sum t_A/N_p \quad \text{from (36), } h = 1 \] (39)

The correlation between random zygotes from the same neighborhood, assuming no dominance or factor interaction, is given in this case by \( E/b^2. \)

\[ G = (N_p - 1)\sum t_A/[N_p[(N_p - 1) + (1 - h)(N_p - \sum t_A)]] \] (40)

In the case of a population with a linear range, it is necessary merely to substitute \( \Sigma t_L \) for \( \Sigma t_A \) in the preceding formulae.

The relation between effective number of pairs, standard deviation \( (\sigma_p) \) of distances between parent-offspring mating sites, and density of the pairs \( (d_p) \) can be found as in case 1.

\[ N_p = 2\sqrt{\pi} \sigma_p d_p \quad \text{with linear continuity (41)} \]

\[ N_p = 4\pi\sigma_p^2d_p \quad \text{with area continuity (42)} \]

Since \( N = 2N_p \) and \( d = 2d_p \) and \( \sigma = \sigma_p \) (if individuals are located at the site of first mating), formulae (41) and (42) apply to individuals by dropping all subscripts. They become identical with (3) and (6).

**CASE 3. SEPARATE SEXES AND RANDOM MATING**

Assume that males and females are both distributed uniformly but not necessarily with the same density over either an extensive linear range or an extensive area, that each offspring is produced by a separate random mating,
and that there is equal dispersion of sons and daughters. Let $N_m$ and $N_f$ be respectively the effective numbers of mature males and females per neighborhood. The diagrams below represent three of four possible relations between gametes of mated or adjacent individuals.

All correlations are assumed to be relative to a population of specified size $K(N_m + N_f)$.

Let $A$ be the correlation between gametes produced by full brother and sister.

Let $B$ and $C$ be the correlations between gametes produced by half brothers and sisters with common mother and father, respectively.

Let $D$ be the correlation between gametes of mated or adjacent individuals without a common parent, and $D_x$ that between gametes of non-siblings of ancestral generation $(x-1)$ relative to matings.

Let $E = F$ be the correlation between gametes of mated or adjacent individuals and $E_x$ that between gametes of individuals of ancestral generation $(x-1)$ relative to matings.

Let $F_c$ be the correlation between gametes of individuals that mate at some time with the same individual.

Assuming random mating, the probabilities of the four cases and their contributions to $E$ are readily determined.

$$E = [A + (N_m - 1)B + (N_f - 1)C] + (N_m - 1)(N_f - 1)D]/N_mN_f$$  \(43\)

$$A = b^2a^2(2b^2 + 2F) = (1 + 3F)/4$$  \(44\)

$$B = C = b^2a^2(b^2 + 2F + F_c) = (1 + 5F + 2F_c)/8$$  \(45\)

$$D = 4b^2a^2E_2 = E_2$$  \(46\)

$$D_x = 4b^2a^2E_{x+1} = E_{x+1}$$  \(47\)
Collecting terms and multiplying \( N_m \) and \( N_f \) by \( X \),

\[
E_X = \left( \frac{N_m + N_f}{X N_m N_f} \right) \left( \frac{1 + 5F + 2F_C}{8} \right) + \left[ 1 - \left( \frac{N_m + N_f}{X N_m N_f} \right) \right] E_{X+1} - \left( \frac{F + F_C - 2E_{X+1}}{2X^2 N_m N_f} \right)
\]

(48)

In the last term the numerator does not include population numbers and is twice the difference between two quantities that are both less than one and nearly the same if \( X \) is small; the denominator involves \( X^2 \). This term must always be much smaller than any of the others. It will be assumed that it can be ignored.

It will be convenient to write \( N' \) for \( N_m N_f / (N_m + N_f) \).

\[
E = \left[ (1 + 5F + 2F_C) / 8N' \right] + (N' - r) E_r / N' \text{ approximately}
\]

\[
E_2 = \left[ (1 + 5F + 2F_C) / 16N' \right] + (2N' - r) E_2 / 2N' \text{ approximately}
\]

\[
E_X = \left[ (1 + 5F + 2F_C) / 8XN' \right] + (XN' - r) E_{X+1} / XN'
\]

\[
E_K = 0
\]

\[
E = \left[ (1 + 5F + 2F_C) / 8N' \right] \sum t_A.
\]

(50)

The correlation \( F_C \) cannot differ appreciably from \( F \) (and \( E \)) although theoretically slightly smaller. Substituting \( E \) for both \( F \) and \( F_C \), the following approximate solution can be obtained.

\[
E = \sum t_A / (8N' - 7 \sum t_A).
\]

(51)

The amount of differentiation among neighborhoods is given as in other cases by

\[
E = \frac{1}{(\pi/\sigma)^2} \int_{-\infty}^{\infty} \frac{1}{x^2 + \sigma^2} \exp(-x^2/\sigma^2) \, dx
\]

(52)

Thus \( N_m = 2\sqrt{\pi} \sigma d_m = 3.5\sigma d_m \)

(53)

Similarly \( N_f = 2\sqrt{\pi} \sigma d_f = 3.5\sigma d_f \).
The result is the same per individual as in cases 1 and 2. This is also true for area continuity.

\[ N_m = 4\pi \sigma^2 d_m = 12.6 \sigma^2 d_m \]
\[ N_f = 4\pi \sigma^2 d_f = 12.6 \sigma^2 d_f. \]

**CASE 4. DISPERSION BY GAMETES OF ONLY ONE SEX**

The assumption that the distribution of parent-offspring distances is the same for both parents often does not apply. Thus among plants, dispersal of seed may be negligible in comparison with that of pollen. The assumption of different variances for male and female parents leads unfortunately to great algebraic complexity. It must suffice here to consider the limiting case in which all dispersal is through one parent (taken as the male). For simplicity, we shall assume all individuals to be hermaphrodites but with self fertilization occurring to any specified extent. The population will be assumed to be of uniform density, either along a strip of indefinitely great length, negligible width, or over an area of indefinite extent in all directions. In the former case, the distribution of pollen parent about progeny (and ovule parent) will be assumed to be univariate normal with standard deviation \( \sigma \), apart from excess (or defect) of the middle class due to excess (or defect) in amount of self fertilization relative to that under random fertilization. Letting \( n \) be the number of individuals in a strip \( 2\sigma \) long, the average density is \( d = n/2\sigma \) and the average interval between individuals is \( 2\sigma/n \). Let \( N \) be the effective number of individuals in the neighborhood that function as pollen parents, without taking cognizance of excess self fertilization. Then the chance of self fertilization under random union is \( (1/N) = (2\sigma/n) y_0 \) where \( y_0 \) is the midordinate of the normal distribution \( y = (1/\sigma \sqrt{2\pi}) \exp[-(x^2/2\sigma^2)]. \) It is assumed that \( N \) is sufficiently large that the middle class of the distribution can be represented reasonably well by the product of midordinate into class range. Thus

\[ 1/N = 2\sigma/n \sigma \sqrt{2\pi} = (1/n) \sqrt{2/\pi} \]
\[ N = n \sqrt{\pi/2} = \sqrt{2\pi} \sigma d = 2.50 \sigma d. \]

In the case of area continuity, we assume that the distribution of pollen parents is bivariate normal relative to ovule parents and progeny, \( y = (1/2\sigma^2) \exp[-(x_1^2 + x_2^2)/2\sigma^2]. \) Letting \( n \) be the number of individuals in a square \( 2\sigma \) on a side, the average density is \( d = n/4\sigma^2 \), and the average space per individual is \( 4\sigma^2/n \). Defining \( N \) as above, the chance of self fertilization under random mating of gametes in the neighborhood is \( 4\sigma^2 y_0/n \).

\[ 1/N = 4\sigma^2/(n(2\pi \sigma^2)) = 2/\pi n \]
\[ N = \pi n/2 = 2\pi \sigma^2 d = 6.28 \sigma^2 d. \]

\( N \) is thus equivalent to the number of individuals in a circle of radius \( \sqrt{2}\sigma \) in this case. It is half as great as in the case of an equally dense population of hermaphrodites with the same amount of dispersion of both male and female parents. However, it permits a much simpler mathematical analysis, and its use in this connection is recommended.
gametes as of male gametes in this case. If the variance of pollen dispersal is twice that of the preceding cases but \( d \) is the same, formulae (55) and (56) become the same as (3) and (6), respectively.

Let \( r \) be the proportion of the pollinations that may be considered as at random from the neighborhood and \( (1-r) \) the excess self fertilization. Assuming that \( N \) is at least 2, the variance of pollen parents in any direction is \( r\sigma^2 \) (area continuity), since self fertilization contributes nothing. The total proportion of self fertilization is \( h = (r/N) + (1-r) \).

The correlation \( (E) \) between ovules and pollen grains that contribute to adjacent zygotes and that \( (F) \) between ones that unite may be analyzed as indicated in figure 4.

\[
E = \frac{(C/N) + D(N - 1)}{N} \\
F = hC + (1 - h)D.
\]

In cases of cross pollination, the variance of father’s mother relative to mother’s mother is the same as that of father relative to mother \( (\sigma^2) \) because of the postulated absence of dispersion in the female line. The correlation between ovules from which fathers and mothers (where different) were derived is thus \( E \). The variance of father’s father about mother’s mother (father and mother different) is \((1+r)\sigma^2\), corresponding to an effective population of \((1+r)N\) if area continuity or \(\sqrt{1+r} N\) if linear continuity, assuming that \( N \) is at least 2. The same is true for father’s mother relative to mother’s father. The correlation between pollen and ovule produced by these grandparents will be called \( E_2 \). The variance of father’s father relative to mother’s father (father and mother different) is \((1+2r)\sigma^2\), corresponding to an effective population of \((1+2r)N\) if area continuity and \(\sqrt{1+2r} N\) if linear continuity. The correlation between pollen grains produced by these grandparents will be called \( E_2 \).

The spores produced by the \( X \)th ancestral generation, where the parents are different, are drawn from populations with variances \([1+(X-2)\sigma^2, [1+(X-1)^2]\sigma^2\) and \([1+X\sigma^2]\sigma^2\) according as ovule-ovule, ovule-pollen or pollen-pollen. The effective population numbers are proportional,
[1+(X-z)r]N, etc., if area continuity and proportional to the square roots, \sqrt{1+(X-z)r} N, etc., if linear continuity. Thus if \( D_X \) is the correlation between spores from different individuals of this generation,

\[
D_X = \frac{(E_X + 2E_{X+1} + E_{X+2})}{4} \quad (59)
\]

\[
C = b^2 = \frac{(i + F)}{2}. \quad (60)
\]

With area continuity,

\[
E = \frac{C}{N} + \frac{N - 1}{4N} (E + 2E_2 + E_3)
\]

\[
E_2 = \frac{C}{(i + r)N} + \frac{(i + r)N - i}{4(i + r)N} (E_2 + 2E_3 + E_4)
\]

\[
E_X = \frac{C}{[i + (X - r)r]N} + \frac{i + (X - r)N - i}{4[i + (X - r)r]N} (E_X + 2E_{X+1} + E_{X+2})
\]

\[E_K = 0.\]

The equations above (except the last) are obviously satisfied if all \( E \)'s equal \( i \), which, as in the preceding cases, is the limiting value approached as the area of reference is made indefinitely large. The attempt to solve for \( E \) with finite \( N_K \) leads, as in case 3, to an unmanageable series. An approximate solution may be obtained as follows:

Let

\[
(E_X + 2E_{X+1} + E_{X+2})/4 = E_{X+1} + \Delta_{X+1} \quad (62)
\]

\[
E = \frac{C}{N} \left\{ i + \left( \frac{i}{i + r} \right) \left( \frac{N - 1}{N} \right) \right. \\
\left. + \left( \frac{i}{i + 2r} \right) \left( \frac{N - 1}{N} \right) \left[ \frac{(i + r)N - i}{(i + r)N} \right] \ldots \\
+ \Delta_2 \left( \frac{N - 1}{N} \right) + \Delta_3 \left( \frac{N - 1}{N} \right) \left[ \frac{(i + r)N - i}{(i + r)N} \right] \ldots \right\} \quad (63)
\]

As \( \Delta_{X+1} = \frac{1}{4} [(E_X - E_{X+1}) - (E_{X+1} - E_{X+2})] \), the \( \Delta \)'s are equal to fourths of the second differences of the \( E \)'s. They can easily be shown to be negligibly small, and a first approximation may be obtained by ignoring them.

The series that is left in brackets differs from those considered before and may be called \( 2t_\alpha' \).

\[
E = C \sum t_{\alpha'}/N \quad \text{approximately} \quad (64)
\]

It can easily be found from the equations given above, including the approximation \( D = E_2 = (NE - C)/(N - 1) \) from (57), the first equation of the set (61), (58) and (60) that

\[
C = \frac{(i + rE)}{(i + r)} = \frac{(i + F)}{2} \quad (65)
\]
\[ F = \frac{[r + r(2E - r)]}{(1 + r)} \]  
(66)

Thus \[ E = \left( \frac{r + rE}{1 + r} \right) \frac{\sum t_A'}{N} = \frac{\sum t_A'}{N + r(N - \sum t_A')} \approx \text{approximately} \]  
(67)

The series \( \Sigma t_A' \) can be evaluated in the same way as \( \Sigma t_A \) (Wright 1943a). If self fertilization occurs only at random in the neighborhood (\( r = 1 \)), \( \Sigma t_A' \) becomes in fact \( \Sigma t_A \) and

\[ E = \sum t_A/\left[ 2N - \sum t_A \right] \approx \text{approximately} \]  
(68)

This is the same formula as that obtained earlier for hermaphrodites with equal dispersion of both parents relative to offspring. However, there is greater differentiation for a given density of population and given dispersion of pollen, since effective \( N \) is only half as great if there is no dispersion of ovules as with dispersion of ovules equal to that of pollen.

If \( r \) is an integer (\( H \)), the series \( \Sigma t_A' \) can be expressed in terms of \( \Sigma t_A \). Let \( N' = rN \) and assume, as before, that this is at least 2. \( \Sigma t_A \) below is defined in terms of \( N' \) instead of \( N \).

\[ \sum t_A' = H \left[ \frac{1}{H + z} + \left( \frac{1}{H + z} \right) \left( 1 - \frac{1}{HN'} \right) \right] \]  
(69)

\[ E = 2(\sum t_A - 1) / [3N' - \sum t_A - 2] \]  
(71)

COMPARISON OF CASES

The effects of different systems of mating on populations with uniform distributions over large areas are compared in tables 1 to 4. Consider first the variance \((\sigma_q^2)\) of gene frequencies of neighborhoods with populations of effective size \( 200 \) (table 1) (excluding in this and all other cases the variance, \((q(1-q)/2N))\), due immediately to accidents of sampling). In an indefinitely extended population of hermaphrodites in which the gametes combine at random in the neighborhood and are dispersed equally along both parental lines, \( \sigma_q^2 \) reaches only 2.45 per cent of its maximum within an area that includes two million individuals and only 4.75 per cent of the maximum if 20 billion are included. Complete exclusion of self fertilization makes no appreciable difference. There is also no appreciable difference if the neighborhoods consist of separate sexes that mate at random to form 100 permanent pairs, assuming equal dispersion of males and females \((\sigma_q^2 \text{ for neighborhoods rises to } 2.51 \text{ per cent of the maximum within an area including one million pairs and to } 4.97 \text{ per cent of the maximum within an area including ten billion pairs})\). If each offspring is produced by a separate random mating but otherwise conditions are the same, differentiation of neighborhoods is slightly greater \((\sigma_q^2 = 2.64 \text{ per})\)
The correlation (E) between gametes of adjacent individuals belonging to neighborhoods of size \( N(=z0) \), relative to populations of size \( N_K(=KN) \), which in turn are portions of a population distributed with uniform density (d) over an indefinitely large area. The effective population of a neighborhood is determined by the density and the variance (\( s^2 \)) of parent-offspring distances. \( N = 4s^2d \). 

E also measures the variance, \( s^2 \), of gene frequencies of neighborhoods relative to the limiting value \( q_K(r - q_K) \) of the population of size \( N_K \). 

\[
E = uz/qK(1-uz) 
\]

Case 1 is that of a population of hermaphrodites with various amounts (h) of self-fertilization. Case 2 is that of permanent pairs with various amounts (h) of brother-sister mating. Case 3 is that of a separate random mating for each offspring. The results would be the same for \( N_m=50, N_f=\infty \) as for \( N_m=N_f=100 \). Equal dispersion in the male and female lines is postulated in cases 1, 2 and 3.

Case 4 is that of hermaphrodites with no dispersion in the female line, dispersion of double variance in the male line.

**Table 1**

<table>
<thead>
<tr>
<th>( K )</th>
<th>( N_K )</th>
<th>( N_m=200 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CASE 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma^2=\sigma_m^2=\sigma^2 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( h=0 )</td>
</tr>
<tr>
<td>1</td>
<td>200</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>200</td>
<td>0.0071</td>
</tr>
<tr>
<td>10^2</td>
<td>200</td>
<td>0.0129</td>
</tr>
<tr>
<td>10^3</td>
<td>200</td>
<td>0.0187</td>
</tr>
<tr>
<td>10^4</td>
<td>200</td>
<td>0.0244</td>
</tr>
<tr>
<td>10^5</td>
<td>200</td>
<td>0.0302</td>
</tr>
<tr>
<td>10^6</td>
<td>200</td>
<td>0.0359</td>
</tr>
<tr>
<td>10^7</td>
<td>200</td>
<td>0.0416</td>
</tr>
<tr>
<td>10^8</td>
<td>200</td>
<td>0.0474</td>
</tr>
</tbody>
</table>

**Separate Sexes \( N_m=100; N_f=100 \)**

| \( K \) | \( N_K \) |
|-------|-------|-------------|
|       |       | CASE 2      | CASE 3     |
|       |       | \( \sigma^2=\sigma_m^2=\sigma^2 \) | \( \sigma^2=\sigma_m^2=\sigma^2 \) |
|       |       | \( h=0 \)   | \( h=0.01 \) | \( h=0.50 \) | \( h=1 \) |
| 1     | 200   | 0           | 0           | 0           | 0           | 0.0073 |
| 10    | 200   | 0.0071       | 0.0071      | 0.0113      | 0.0280      | 0.0073 |
| 10^2  | 200   | 0.0130       | 0.0131      | 0.0207      | 0.0505      | 0.0135 |
| 10^3  | 200   | 0.0189       | 0.0191      | 0.0300      | 0.0722      | 0.0198 |
| 10^4  | 200   | 0.0249       | 0.0251      | 0.0393      | 0.0933      | 0.0264 |
| 10^5  | 200   | 0.0309       | 0.0312      | 0.0486      | 0.1140      | 0.0331 |
| 10^6  | 200   | 0.0370       | 0.0373      | 0.0586      | 0.1341      | 0.0401 |
| 10^7  | 200   | 0.0432       | 0.0435      | 0.0674      | 0.1538      | 0.0473 |
| 10^8  | 200   | 0.0494       | 0.0497      | 0.0768      | 0.1731      | 0.0547 |

percent of the maximum in areas including one million of each sex and 5.47 percent in areas including ten billion of each). If, in a population of monoecious plants, there is no appreciable dispersion of seed but dispersion of pollen of
The standard deviations $\sigma_{g(i)}$ of gene frequencies of populations of size $N_i$ relative to that of size $2 \times 10^8$ where the effective size of neighborhood is $200$. Cases as in table 1

$$\sigma_{g(i)}/\sigma_{g(E)} = \sqrt{(E_k - E_i)/(1 - E_i)}$$

### Table 2

**Hermaphrodites $N = 200$**

<table>
<thead>
<tr>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_i^2 = \sigma^2$</td>
<td>$\sigma_i^2 = \sigma^2$</td>
<td>$\sigma_i^2 = 2\sigma^2$</td>
</tr>
<tr>
<td>$h = 0$</td>
<td>$h = 0$</td>
<td>$h = 0$</td>
</tr>
<tr>
<td>$h = .005$</td>
<td>$h = .005$</td>
<td>$h = .005$</td>
</tr>
<tr>
<td>$h = .500$</td>
<td>$h = .500$</td>
<td>$h = .500$</td>
</tr>
<tr>
<td>$h = 1$</td>
<td>$h = 1$</td>
<td>$h = 1$</td>
</tr>
</tbody>
</table>

- **TABLE**: The standard deviations $\sigma_{g(i)}$ of gene frequencies of populations of size $N_i$ relative to that of size $2 \times 10^8$ where the effective size of neighborhood is 200. Cases as in table 1

<table>
<thead>
<tr>
<th>$K$</th>
<th>$N_i$</th>
<th>$h = 0$</th>
<th>$h = .005$</th>
<th>$h = .500$</th>
<th>$h = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>200</td>
<td>.2176</td>
<td>.2179</td>
<td>.2494</td>
<td>.3011</td>
</tr>
<tr>
<td>10</td>
<td>$2 \times 10^3$</td>
<td>.2014</td>
<td>.2017</td>
<td>.2309</td>
<td>.2787</td>
</tr>
<tr>
<td>10$^2$</td>
<td>$2 \times 10^4$</td>
<td>.1868</td>
<td>.1870</td>
<td>.2141</td>
<td>.2584</td>
</tr>
<tr>
<td>10$^3$</td>
<td>$2 \times 10^5$</td>
<td>.1710</td>
<td>.1712</td>
<td>.1959</td>
<td>.2365</td>
</tr>
<tr>
<td>10$^4$</td>
<td>$2 \times 10^6$</td>
<td>.1533</td>
<td>.1535</td>
<td>.1758</td>
<td>.2121</td>
</tr>
<tr>
<td>10$^5$</td>
<td>$2 \times 10^7$</td>
<td>.1332</td>
<td>.1333</td>
<td>.1526</td>
<td>.1843</td>
</tr>
<tr>
<td>10$^6$</td>
<td>$2 \times 10^8$</td>
<td>.1001</td>
<td>.1002</td>
<td>.1250</td>
<td>.1590</td>
</tr>
<tr>
<td>10$^7$</td>
<td>$2 \times 10^9$</td>
<td>.0773</td>
<td>.0774</td>
<td>.0886</td>
<td>.1070</td>
</tr>
<tr>
<td>10$^8$</td>
<td>$2 \times 10^{10}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- **SEPARATE SEXES**: $N_f = 100$, $N_m = 100$

<table>
<thead>
<tr>
<th>Case 3</th>
<th>Case 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_i^2 = \sigma^2$</td>
<td>$\sigma_i^2 = \sigma^2$</td>
</tr>
<tr>
<td>$h = 0$</td>
<td>$h = 0$</td>
</tr>
<tr>
<td>$h = .01$</td>
<td>$h = .01$</td>
</tr>
<tr>
<td>$h = .50$</td>
<td>$h = .50$</td>
</tr>
<tr>
<td>$h = 1$</td>
<td>$h = 1$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$K$</th>
<th>$N_i$</th>
<th>$h = 0$</th>
<th>$h = .01$</th>
<th>$h = .50$</th>
<th>$h = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>200</td>
<td>.2222</td>
<td>.2300</td>
<td>.2772</td>
<td>.4160</td>
</tr>
<tr>
<td>10</td>
<td>$2 \times 10^3$</td>
<td>.2064</td>
<td>.2071</td>
<td>.2574</td>
<td>.3864</td>
</tr>
<tr>
<td>10$^2$</td>
<td>$2 \times 10^4$</td>
<td>.1919</td>
<td>.1926</td>
<td>.2304</td>
<td>.3593</td>
</tr>
<tr>
<td>10$^3$</td>
<td>$2 \times 10^5$</td>
<td>.1762</td>
<td>.1768</td>
<td>.2197</td>
<td>.3298</td>
</tr>
<tr>
<td>10$^4$</td>
<td>$2 \times 10^6$</td>
<td>.1584</td>
<td>.1590</td>
<td>.1976</td>
<td>.2966</td>
</tr>
<tr>
<td>10$^5$</td>
<td>$2 \times 10^7$</td>
<td>.1380</td>
<td>.1385</td>
<td>.1721</td>
<td>.2584</td>
</tr>
<tr>
<td>10$^6$</td>
<td>$2 \times 10^8$</td>
<td>.1133</td>
<td>.1137</td>
<td>.1413</td>
<td>.2171</td>
</tr>
<tr>
<td>10$^7$</td>
<td>$2 \times 10^9$</td>
<td>.0806</td>
<td>.0809</td>
<td>.1005</td>
<td>.1509</td>
</tr>
<tr>
<td>10$^8$</td>
<td>$2 \times 10^{10}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- **Note**: twice the variance postulated in the preceding cases (in populations of the same density), the amount of differentiation is practically the same as in the case of hermaphrodites with equal dispersion along both parental lines. This rule applies in random breeding populations, irrespective of density, and applies under linear as well as area continuity. It may be surmised that in such populations the effects with unequal dispersion are substantially the same as if there were equal dispersion to the extent of the average variance along the two parental lines. Returning to the case of neighborhoods of size 200 in a population continuous over an area, it may be added that with dispersion of pollen equal to that in case 1 and no dispersion of seed, $\sigma_i^2$ is 5.24 per cent instead of
The correlation ($E$) between gametes of adjacent individuals, belonging to neighborhoods of size $N = 20$, relative to populations of size $N_K$, which in turn are portions of a population distributed with uniform density over an indefinitely large area. The cases are as in table 1. In case 3 the results would be the same for neighborhoods containing five males and an indefinite number of females as with ten of both.

### Table 3

**Hermaphrodites $N = 20$**

<table>
<thead>
<tr>
<th>$K$</th>
<th>$N_K$</th>
<th>$\sigma^2 = \sigma_m^2 = \sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$h = 0$</td>
</tr>
<tr>
<td>1</td>
<td>$10^1$</td>
<td>0.0669</td>
</tr>
<tr>
<td>10</td>
<td>$2 \times 10^2$</td>
<td>0.1209</td>
</tr>
<tr>
<td>100</td>
<td>$2 \times 10^4$</td>
<td>0.1821</td>
</tr>
<tr>
<td>1000</td>
<td>$2 \times 10^6$</td>
<td>0.2362</td>
</tr>
<tr>
<td>10000</td>
<td>$2 \times 10^8$</td>
<td>0.2890</td>
</tr>
<tr>
<td>100000</td>
<td>$2 \times 10^10$</td>
<td>0.3401</td>
</tr>
<tr>
<td>1000000</td>
<td>$2 \times 10^12$</td>
<td>0.3804</td>
</tr>
</tbody>
</table>

**Separate Sexes $N_t = 10$, $N_m = 10$**

<table>
<thead>
<tr>
<th>$K$</th>
<th>$N_K$</th>
<th>$\sigma^2 = \sigma_m^2 = \sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$h = 0$</td>
</tr>
<tr>
<td>1</td>
<td>$10^1$</td>
<td>0.0723</td>
</tr>
<tr>
<td>10</td>
<td>$2 \times 10^2$</td>
<td>0.1378</td>
</tr>
<tr>
<td>100</td>
<td>$2 \times 10^4$</td>
<td>0.2071</td>
</tr>
<tr>
<td>1000</td>
<td>$2 \times 10^6$</td>
<td>0.2799</td>
</tr>
<tr>
<td>10000</td>
<td>$2 \times 10^8$</td>
<td>0.3545</td>
</tr>
<tr>
<td>100000</td>
<td>$2 \times 10^{10}$</td>
<td>0.4289</td>
</tr>
<tr>
<td>1000000</td>
<td>$2 \times 10^{12}$</td>
<td>0.5013</td>
</tr>
</tbody>
</table>

2.45 per cent of the maximum in areas including two million individuals and is 9.82 per cent instead of 4.75 per cent in areas including twenty billion.

A tendency toward self fertilization or toward brother-sister mating somewhat increases the amount of differentiation of neighborhoods. But even 50 per cent self fertilization in the case of hermaphrodites and 50 per cent brother-sister mating in the case of permanent pairs raise $\sigma^2$ as a percentage of the maximum within a population of 20 billion only from 4.75 per cent to 6.22 per cent in the former and from 4.97 per cent to 7.68 per cent in the latter. The ef-
The standard deviations of gene frequencies of populations of size \( N_t \) relative to that of a population of size \( 2 \times 10^8 \), where the effective size of neighborhood is 20. Cases as in table 1.

### Table 4

**Hermaphrodites \( N_t = 20 \)**

<table>
<thead>
<tr>
<th>( K )</th>
<th>( N_t )</th>
<th>( \sigma^2 = \sigma_{m}^2 = \sigma^2 )</th>
<th>( \sigma^2 = 0 )</th>
<th>( \sigma_{m}^2 = 20 \sigma^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( h = 0 )</td>
<td>( h = 0.05 )</td>
<td>( h = 0.50 )</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>0.6585</td>
<td>0.6657</td>
<td>0.7144</td>
</tr>
<tr>
<td>10</td>
<td>( 2 \times 10^2 )</td>
<td>0.6279</td>
<td>0.6375</td>
<td>0.6789</td>
</tr>
<tr>
<td>10^2</td>
<td>( 2 \times 10^3 )</td>
<td>0.5926</td>
<td>0.6000</td>
<td>0.6440</td>
</tr>
<tr>
<td>10^3</td>
<td>( 2 \times 10^4 )</td>
<td>0.5579</td>
<td>0.5620</td>
<td>0.6031</td>
</tr>
<tr>
<td>10^4</td>
<td>( 2 \times 10^5 )</td>
<td>0.5123</td>
<td>0.5160</td>
<td>0.5538</td>
</tr>
<tr>
<td>10^5</td>
<td>( 2 \times 10^6 )</td>
<td>0.4457</td>
<td>0.4590</td>
<td>0.4927</td>
</tr>
<tr>
<td>10^6</td>
<td>( 2 \times 10^7 )</td>
<td>0.3844</td>
<td>0.3832</td>
<td>0.4135</td>
</tr>
<tr>
<td>10^7</td>
<td>( 2 \times 10^8 )</td>
<td>0.2781</td>
<td>0.2801</td>
<td>0.3007</td>
</tr>
<tr>
<td>10^8</td>
<td>( 2 \times 10^9 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### Separate Sexes \( N_t = 10 \), \( N_m = 10 \)

<table>
<thead>
<tr>
<th>( K )</th>
<th>( N_t )</th>
<th>( \sigma^2 = \sigma_{m}^2 = \sigma^2 )</th>
<th>( \sigma^2 = 0 )</th>
<th>( \sigma_{m}^2 = 20 \sigma^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( h = 0 )</td>
<td>( h = 0.10 )</td>
<td>( h = 0.50 )</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>0.7549</td>
<td>0.7677</td>
<td>0.8264</td>
</tr>
<tr>
<td>10</td>
<td>( 2 \times 10^2 )</td>
<td>0.7324</td>
<td>0.7448</td>
<td>0.8017</td>
</tr>
<tr>
<td>10^2</td>
<td>( 2 \times 10^3 )</td>
<td>0.7079</td>
<td>0.7200</td>
<td>0.7749</td>
</tr>
<tr>
<td>10^3</td>
<td>( 2 \times 10^4 )</td>
<td>0.6764</td>
<td>0.6879</td>
<td>0.7404</td>
</tr>
<tr>
<td>10^4</td>
<td>( 2 \times 10^5 )</td>
<td>0.6246</td>
<td>0.6454</td>
<td>0.6947</td>
</tr>
<tr>
<td>10^5</td>
<td>( 2 \times 10^6 )</td>
<td>0.5777</td>
<td>0.5875</td>
<td>0.6324</td>
</tr>
<tr>
<td>10^6</td>
<td>( 2 \times 10^7 )</td>
<td>0.4969</td>
<td>0.5054</td>
<td>0.5440</td>
</tr>
<tr>
<td>10^7</td>
<td>( 2 \times 10^8 )</td>
<td>0.3710</td>
<td>0.3773</td>
<td>0.4061</td>
</tr>
<tr>
<td>10^8</td>
<td>( 2 \times 10^9 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
tiplication differs in not leading to homozygosis. The differentiation of neighborhoods is the same as in the above cases, taking \( q \) as the frequency of one of two alternative genotypes.

Even the largest of the values of \( \sigma_q^2 \) in table 1 may appear rather small, suggesting that, if the effective size of neighborhoods is as large as 200, there is no important differentiation due to the cumulative effects of accidents of sampling even within enormous areas. However, the square roots of these figures give perhaps a fairer idea of amount of differentiation. Thus values of \( \sigma_q \) of 22 per cent to 24 per cent of the possible maximum, occurring independently at all heterallelic loci, are large enough to give the basis for great diversity among the gene systems of neighborhoods and thus a basis for intergroup selection. But groups of only 200 individuals are rather small, and it is desirable to consider the amount of differentiation among larger groups. Table 2 shows the standard deviation \( \sigma_{q(i)} \) for populations of various effective sizes \( N_i \) within a comprehensive population of 20 billion, in relation to the maximum standard deviation. For populations consisting of 10 per cent of the total, mean gene frequency varies about one-third as much as it does for neighborhoods. It varies about half as much among groups that include 1 per cent of the total and about two-thirds as much among ones that include 0.1 per cent of the total (that is, populations of 20 million within one of 20 billion). There is therefore significant, if not very great, chance differentiation of large groups, if the effective size of the neighborhood is as large as 200. As noted in the previous paper, there is virtual equivalence to panmixia with regard to chance differentiation where neighborhoods are larger than 1000 (Wright 1943a). All of this discussion, it should be noted, applies only to area continuity. Under linear continuity there is enormously more differentiation.

Tables 3 and 4 present the same cases for \( N = 20 \) that are presented in tables 1 and 2 for \( N = 200 \). They show that there is very much greater differentiation where the neighborhood is small. There is also much more difference in the results of diverse systems of mating. Thus for neighborhoods of 20 random breeding hermaphrodites the variance of \( q \) within areas including the equivalent of 100 million neighborhoods is 44.3 per cent of its maximum, for neighborhoods of ten permanent pairs within the same total it is 58.9 per cent of maximum, and for neighborhoods of ten males and ten females, mating at random to produce each offspring, it is 85.0 per cent of maximum, assuming equal dispersion along both lines in all these cases. Table 4 gives the standard deviation of \( q \) among populations of various sizes \( N_i \) within a constant total of two billion individuals. We note that populations including 1 per cent of this total are strongly differentiated where neighborhoods are as small as 20. With neighborhoods of 20 hermaphrodites with equal dispersion along both lines, the standard deviation of mean gene frequency of populations of 20 million within 2 billion is 38.5 per cent of maximum, for neighborhoods of ten permanent pairs the corresponding figure is 50.5 per cent, and for neighborhoods in which each offspring is the product of a separate random mating among ten males and ten females, the figure is 72.3 per cent.
It was noted in the preceding paper (Wright 1943a) that either reversible mutation or occasional long range dispersion (rate $m_1$) puts a sharp upper limit on the amount of differentiation among neighborhoods and prevents differentiation of groups larger than about $3/m_1$ neighborhoods. A correction should be made to a statement on p. 131, referred to again in the summary.

"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 is accordingly to be multiplied by $(1-m_1)^X$ (where $X$ is the rank of the term in question)." This is correct if $q$ refers to the frequency of one of only two alternative genotypes, which are multiplying vegetatively. In reference to gene frequencies of a diploid organism the factor to be applied to terms in $\Sigma t$ is $(1-m_1)^2X$, irrespective of system of mating.

It should be added that the occurrence of mutations in an indefinitely extended series of multiple alleles, far from limiting the amount of local differentiation, would greatly increase it. The slow process of diffusion of new alleles from small neighborhoods may not keep up with further mutation and thus each allele may have only a local distribution.

Uniform selection pressure limits differentiation, while local differences in the conditions of selection may enormously increase it, provided that dispersion is sufficiently slow to prevent swamping by panmixia. But even with no differences in the external conditions in different localities, mere chance differentiation at all loci of the sort discussed here may be expected to bring about differences in the direction of selection pressure on individual loci in regions in which different gene systems have been arrived at. The differences in selective trends may be expected to be cumulative, with intergroup selection by means of differential rates of dispersal as a further consequence.

The variability of frequencies of two alleles, due to the cumulative effects of accidents of sampling is thus merely the foundation for much more significant evolutionary processes.

**SUMMARY**

The properties of large, uniformly distributed populations depend on the system of mating and the effective population number ($N$) of the random breeding "neighborhoods." It is shown that with density $d$, of breeding individuals and with standard deviation $\sigma$, of the coordinates of location of parents at some phase of their life cycles relative to the corresponding phase for offspring, $N = 2\sqrt{\pi d}$ if the population has an essentially one dimensional distribution (for example, shore line, river), and $N = 4\pi \sigma^2 d$ if there is continuity over an extensive area. Random mating and equal parental dispersions are assumed here. If there is no dispersion along one parental line (for example, no dispersion of seed) but dispersion measured by variance $2\sigma^2$ along the other line (for example, of pollen), the formulae are the same.

The differentiation of mean gene frequencies among neighborhoods or larger areas, within still larger areas, is investigated in four cases. (1) Populations of hermaphrodites, derived from equally dispersed gametes derived at random
from the neighborhood except for a specified tendency toward self fertilization.  
(2) Populations consisting of permanent pairs, derived by random mating from the neighborhood, except for a specified tendency toward brother-sister mating, with equal dispersion of the sexes. (3) Populations in which each individual is produced by a separate random mating from the neighborhood. The densities of the populations of the two sexes may differ in this case, but dispersion is assumed to be the same. (4) Populations of hermaphrodites, derived from union of ♀ gametes that are not dispersed to any appreciable extent with ♂ gametes derived at random from the neighborhood, except for a specified tendency toward self fertilization.

If the effective breeding population of neighborhoods is 200, there is a moderate amount of differentiation among large subgroups as well as among neighborhoods, within still larger groups inhabiting an indefinitely large area. There are only slight differences among the four cases. Even 100 per cent self fertilization in case 1 or 100 per cent brother-sister mating in case 2 increases differentiation only to a rather slight extent.

If the effective breeding population of neighborhoods is only 20, there is great differentiation among large subgroups as well as among neighborhoods. The amount of differentiation is considerably greater in case (2) than in case (1) or (4) and considerably greater in case (3) than in case (2).

It is noted that this differentiation, due to cumulative effects of accidents of sampling, may be expected in actual cases to be complicated by the effects of occasional long range dispersal, mutation, and selection, but that in combination with these it gives the foundation for much more significant evolutionary processes than these factors can provide separately.

LITERATURE CITED


