THE GENETICAL STRUCTURE OF POPULATIONS

By SEWALL WRIGHT, *University of Chicago*

Galton Lecture at University College, London, 1950

First let me acknowledge the great honour that I feel in being asked to give the Galton Lecture. The genetics of populations has two main supports, one of which traces to Galton. A geneticist cannot read *Natural Inheritance* (1889) without feeling that Galton would have welcomed Mendel's paper if it had come to his attention during his active career, and that he would have seen that it and the statistical approach which he initiated were complementary rather than antithetic. Personally, I wish to testify to the great stimulus that I had in reading *Natural Inheritance* as a student.

Random mating and inbreeding

I propose to discuss certain genetic aspects of population structure. The term is used to include such matters as numbers, composition by age and sex, and state of subdivision. The best starting point is the consideration of the situation in a large random-breeding population, in which structure in the last sense is absent.

Gene frequencies and hence genetic variability tend to remain unchanged in such a population, generation after generation, because of the persistence of genes and the symmetry of the Mendelian mechanism. This contrasts with the expected loss of half the variability in each generation (under blending heredity) which Darwin and his contemporaries recognized to be the greatest difficulty in the theory of natural selection. This principle was first expressed in general form in 1908 in independent papers by Hardy and Weinberg.

In the same year, another deduction from Mendelian principles—that self-fertilization should result in random fixation of half the heterozygous loci in each generation—was used by Shull to interpret the loss of vigour and the fixation of diverse combinations of quantitatively varying characters which he found in his selfed lines of maize.

Table 1 shows at the left the composition of a random-breeding population with respect to a pair of alleles. The array of zygotic frequencies is the square of the array of gene frequencies.

Table 1. *The frequencies of zygotes from a pair of alleles under three conditions: panmixia, an intermediate degree of inbreeding and complete fixation without change of gene frequency. The intermediate condition is expressed in three equivalent ways in terms of the inbreeding coefficient (or fixation index) F and the panmictic index P $(=1-F)$*

	Panmixia ($r_{ss}=0$)	Intermediate ($r_{ss}=F$)			Complete fixation ($r_{ss}=1$)
		Deviation from panmixia	Panmictic and fixed components	Deviation from fixation	
Genotype	Frequency	Frequency	Frequency	Frequency	Frequency
$\mathbf{A_1A_1}$	q^2	$q^2+Fq(1-q)$	Pq^2+Fq	$q-Pq(1-q)$	q
$\mathbf{A_1a}$	$2q(1-q)$	$2q(1-q)-2Fq(1-q)$	$2Pq(1-q)$	$2Pq(1-q)$	—
$\mathbf{aa}$	$(1-q)^2$	$(1-q)^2+Fq(1-q)$	$P(1-q)^2+F(1-q)$	$(1-q)-Pq(1-q)$	$(1-q)$
	1	$1+0$	$P+F=1$	$1-0$	1

Long-continued self-fertilization without selection gives the array of fixed lines, with gene frequencies unchanged, shown at the right. The other columns show an intermediate situation according to three useful points of view indicated in the headings.

It is important to note that the same coefficient, F, that measures the degree of approach toward fixation is also the Galtonian correlation coefficient r_{es} for the alleles that come together at fertilization. This is zero under random mating, 1 under complete fixation and F as the weighted average in the intermediate population.

This sort of description can be extended at once to sets of three or more alleles, not subject to selective differences, by using the principle that any group of alleles may be treated as if one, and the system can therefore be treated in a variety of ways as if it consisted of only two alleles (Table 2). The coefficient F still measures the degree of approach toward fixation, and $P\ (=1-F)$ measures the relative amount of heterozygosis as compared with that in the random bred population. The correlation between alleles in uniting gametes is still F, irrespective of gene frequencies or of the values assigned the alleles (Appendices A, B).

Table 2. *The frequencies of zygotes from a set of three alleles under three conditions: panmixia, an intermediate degree of inbreeding and complete fixation. The intermediate condition is expressed in terms of the inbreeding coefficient F and the panmictic index P* $(=1-F)$

	Panmixia ($r_{es}=0$)	Intermediate ($r_{es}=F$)	Complete fixation ($r_{es}=1$)
Genotype	Frequency	Frequency	Frequency
$\mathbf{A_1A_1}$	q_1^2	$Pq_1^2+Fq_1$	q_1
$\mathbf{A_1A_2}$	$2q_1q_2$	$2Pq_1q_2$	—
$\mathbf{A_1A_3}$	$2q_1q_3$	$2Pq_1q_3$	—
$\mathbf{A_2A_2}$	q_2^2	$Pq_2^2+Fq_2$	q_2
$\mathbf{A_2A_3}$	$2q_2q_3$	$2Pq_2q_3$	—
$\mathbf{A_3A_3}$	q_3^2	$Pq_3^2+Fq_3$	q_3
	1	$P+F=1$	1

Statistical properties of populations

The statistical properties of a population change with changes in zygotic composition even though gene frequencies are not changed (Appendix C). Consider first the case of exact semi-dominance. The mean is not affected, but the genetic variance of the population as a whole increases from its value under random mating ($\sigma^2_{T(0)}$) by the proportion F becoming $(1+F)\sigma^2_{T(0)}$. If the system of mating is one of subdivision into strains with internal random mating, the average variance within these falls off by the proportion F, becoming $(1-F)\sigma^2_{T(0)}$ and the variance of strain means becomes $2F\sigma^2_{T(0)}$.

If, on the other hand, there is any departure from semi-dominance, the mean shifts from its value, $m_{T(0)}$, under random mating toward a value $m_{T(1)}$ characteristic of an array of completely fixed lines and the amount of change is proportional to F. The decrease in size, fecundity and viability that are usually observed on inbreeding seem to have their basis in an association of recessiveness with deleterious effect.

In this case the variance of the population as a whole is a quadratic instead of linear function of F. Nevertheless, the coefficient F is more useful than any other single coefficient in describing the properties of the population, relative to those in a random-bred stock. It was for this reason that it was suggested that it would be the most suitable inbreeding coefficient (Wright, 1922*a*).

The inbreeding coefficient F

The rate of decrease of heterozygosis in systems of mating more complicated than self-fertilization was first worked out from the recurrence relation between successive generations independently by Jennings (1914) and Fish (1914) for brother-sister mating and by Jennings (1916) for some others. The present writer, who had assisted Fish in his calculations, found a simpler way of finding this quantity, the method of path coefficients, based on the correlation between uniting gametes (Wright, 1921) (Appendix A). The results, following Jennings, were expressed in this paper in terms of gene frequency $\frac{1}{2}$, but it was soon shown that the formulae applied to any gene frequency (Wright, 1922*b*). The quantity F was therefore proposed as an inbreeding coefficient giving 'the departure from the amount of homozygosis under random mating toward complete homozygosis' (Wright, 1922*a*). It has been used since as a measure of such departure relative to a specified foundation stock, not necessarily random bred.

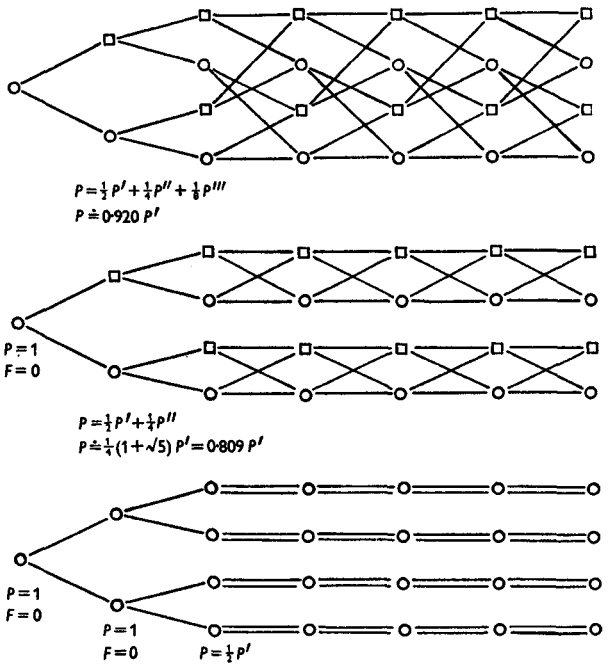


Fig. 1. Three systems of mating which agree in there being two parents and four in all more remote ancestral generations to the beginning.

The method used by Jennings and by Fish becomes too cumbersome to be practicable in dealing with systems much more complicated than mating of brother and sister. Bartlett & Haldane (1934) considerably extended its scope by using matrix algebra. The rates of decrease of heterozygosis indicated by this method have agreed in all comparable cases with those obtained by means of path coefficients (cf. Haldane, 1930, 1936, 1937, 1949; Wright, 1938). The methods are complementary. The matrix method gives a rather complete account of the history of the

population in all respects. The method of path coefficients yields only one property (F) but can obtain this readily from systems which would require matrices with enormous numbers of elements.

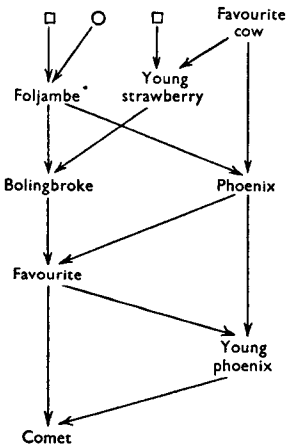
Malécot (1948) has recently shown how the general formula for F, given by the method of path coefficients, can also be demonstrated directly from the theory of probability (cf. also Haldane, 1949).

F is not the only inbreeding coefficient that has been suggested. Pearl (1917) attempted to devise one that was independent of any theory of heredity. Unfortunately, it may have the same value under systems of mating that give the most diverse results experimentally (Fig. 1) (cf. Wright, 1923*a*, *b*). Bernstein (1930) suggested a coefficient α to describe departure from panmixia. It is identical with F. R. A. Fisher (1949) considers that attempts to set up coefficients of inbreeding are unsatisfactory for reasons which do not apply to F. He himself proposes 'an absolute measure of the amount of progress made in inbreeding' based on the matrix method. It turns out to be identical with $[-\log_e(1-F)]$ and is thus not an essentially independent coefficient.

Hierarchic structure

The method of path coefficients leads to simple general formulae for F in terms of its value in preceding generations. These are different for autosomal, sex-linked and polysomic loci (Appendix B). In the case of regular systems of mating, capable of being represented in diagrams, recurrence relations can usually be found very easily without reference to the general formulae (Wright, 1921, 1933*a*, 1938) (Appendix E). The relations between linked loci can be worked out by allied methods (Wright, 1933*b*). The effects of enforced heterozygosis (unpublished) agree with those of the matrix method as far as this has been carried (Bartlett & Haldane, 1935) and can be extended to large inbreeding populations.

The general formulae can be applied to irregular pedigrees. Fig. 2 shows the calculations from the pedigrees (not quite complete) of the bull Favourite (born 1793), on which the Shorthorn breed is essentially based, and of his son Comet (Wright, 1922*a*). The method has been applied to all of the sixty-four cows of Bates's famous Duchess strain (Wright, 1923*b*), of which Darwin



$$F = \Sigma\left[\left(\tfrac{1}{2}\right)^n (1 + F_A)\right]$$

Favourite

A	n	$(\frac{1}{2})^n (1+F_A)$
Foljambe	3	0·125 × 1 = 0·125
Favourite cow	4	0·063 × 1 = 0·063
		F = 0·188

Comet

A	n	$(\frac{1}{2})^n (1+F_A)$
Favourite	2	0·250 × 1·19 = 0·298
Phoenix	3	0·125 × 1 = 0·125
Foljambe	5	0·031 × 1 = 0·031
Favourite cow	6	0·016 × 1 = 0·016
		F = 0·470

Fig. 2. Pedigree of the Shorthorn bull Comet (115) and his sire Favourite (252) and calculations of their inbreeding coefficients relative to the beginning of the Coates herd book (with minor exceptions that raise F to 0·192 for Favourite and to 0·471 for Comet).

(1868) stated: 'For thirteen years he bred most closely in-and-in; but during the next seventeen years, although he had the most exalted notion of his own stock, he thrice infused fresh blood into his herd. It is said that he did this not to improve the form of his animals, but on account of their lessened fertility.' Duchess 1 was a daughter of Comet and five generations in the straight female line from a foundation cow of the herd book. Duchesses 59 and 62 were eight generations later than Duchess 1 in the straight female line. This is about as far as one cares to go by analysis of full pedigrees. Fortunately, there is a simple sampling method (Appendix D), based on comparisons of single random lines back of sire and dam of each of a random selection of animals, which can be applied with appropriate standard errors to whole breeds (Wright & McPhee, 1925). Fig. 3 brings out the remarkably high values of F maintained by Bates throughout the whole history of the Duchess strain and the lower but still impressive figure for the breed as a whole at various periods (McPhee & Wright, 1925, 1926).

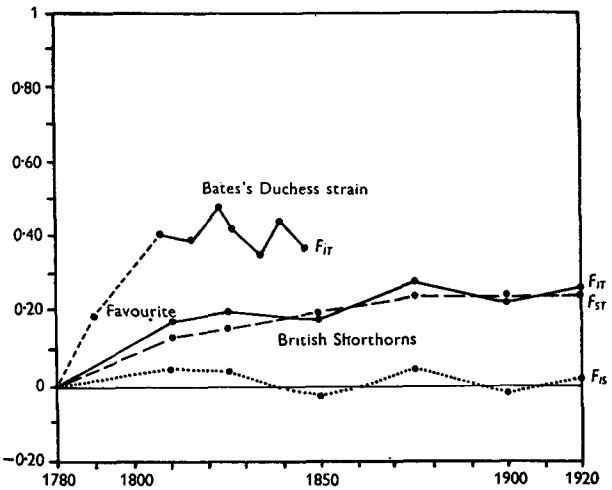


Fig. 3. The inbreeding coefficient (F_{IT}) of cows of Bates's Duchess strain and of the Shorthorn breed as a whole at various dates, relative to the beginning of the Coates herd book. The inbreeding (F_{ST}), relative to this foundation stock, that would persist if random mating were initiated, and the current inbreeding (F_{IS}) are also shown.

It has probably occurred to the reader that the coefficient of inbreeding may mean very different things in different cases. (1) There may be division of the population into completely isolated small strains, within each of which there is random mating. The inbreeding coefficient of individuals relative to the total is here due merely to the relationship of all members of the same strain and disappears at once with random mating among strains. (2) There may be frequent mating of close relatives but no permanent separation of strains. Here again random mating at once reduces the inbreeding coefficient to zero. (3) The sires used may be rather limited in number and derived from even more limited numbers of grandsires and great-grandsires. In this case there may be little apparent close inbreeding at any time, contrary to (2), and no division into strains, contrary to (1), but the value of F, relative to the foundation stock, keeps rising and cannot be much reduced by random mating. There are other possibilities. Clearly we need something more than a single value of F to give an adequate description of structure.

In the case of the pairs of random paternal and maternal lines of random Shorthorns, it was easy to find out what would happen if there were random mating merely by matching these lines

at random instead of matching those back of single actual animals. It came out that in some years the artificial F was slightly greater, in others less, than the actual F. We concluded that the breed as a whole in the later years was much inbred relative to the foundation stock of about 1780 (case 3). The amount of current inbreeding must have been low in the later years. Just what it was at each period can be deduced from a seemingly rather different situation.

Suppose that a population is divided on any basis into subpopulations. The average value of P for individuals relative to their substrains is given in principle by the ratio of average actual heterozygosis within strains to that expected from internal random mating. Represent this by P_{IS} and the panmictic index for individuals relative to the total by P_{IT}. The corresponding inbreeding coefficients are F_{IS} and F_{IT}. It has been shown (Wright, 1943*a*) that the correlation between random gametes, drawn from the same subpopulation, relative to the total, is given by the formula $F_{ST}=(F_{IT}-F_{IS})/(1-F_{IS})$. This can be expressed more conveniently in terms of P's (Wright, 1948*c*): $P_{IT}=P_{IS}P_{ST}$. If there are primary subpopulations (S_1) that are themselves subdivided (S_2), $P_{IT}=P_{IS_2}P_{S_2S_1}P_{S_1T}$. This sort of analysis can be continued to any degree of hierarchic subdivision.

Returning to the Shorthorns, we may look upon the whole breed of a later period as one of an indefinitely large number of isolated strains which might conceivably have been derived from the foundation stock of 1780 by the average mating system actually followed. The inbreeding coefficient of individuals relative to the foundation stock may thus be treated as of the type of F_{IT}. That from matching random lines of ancestry is of the type F_{ST}. The amount of current inbreeding (F_{IS}) can then be estimated as above. The formula actually used for this purpose in

Table 3. *The sets of values of F_{IS}, F_{ST} and F_{IT}; P_{IS}, P_{ST} and P_{IT} under extreme types of population structure*

Total population (T) Subpopulation (S) Individual (I) $P_{IT}=P_{IS}P_{ST}$	Fixation indices			Panmictic indices		
	F_{IS}	F_{ST}	F_{IT}	P_{IS}	P_{ST}	P_{IT}
(1) Each S composed of multiple fixed lines:						
(*a*) Only slight differentiation among S's	1	0+	1	0	1−	0
(*b*) Extreme differentiation among S's	1	1−	1	0	0+	0
(2) Each S panmictic:						
(*a*) Only slight differentiation among S's	0	0+	0+	1	1−	1−
(*b*) Extreme differentiation among S's	0	1−	1−	1	0+	0+
(3) Each S a single genotype:						
(*a*) Perfect negative correlation between alleles	−1	0	−1	2	1	2
(*b*) Random combinations of alleles	−1	½	0	2	½	1
(*c*) Strong positive correlation between alleles	−1	1−	1−	2	0+	0+
(*d*) Perfect positive correlation between alleles (all genotypes homozygous)	—	1	1	—	0	0

the 1925 paper was not correct, and the published results were slightly in error. The corrected values are shown in Fig. 3 in the dotted line.

There was in general a little current inbreeding (F_{IS}) responsible for the accumulation of considerable inbreeding of the breed as a whole (F_{ST}), but at certain periods (1850, 1900) there seems to have been a little tendency toward actual outcrossing within the breed.

Lush (1943) and co-workers and others have calculated the inbreeding coefficients of many breeds of cattle, horses, sheep and swine at different periods of their histories. None of these goes back to a recorded foundation stock in which there was as much inbreeding as in the foundation Shorthorns, but all show a gradual rise in the coefficient, which indicates that the effective number at any time was of the order of 100 or at most a few hundred, a very small proportion of the total numbers registered in each generation.

F_{ST} is necessarily positive but F_{IS} and F_{IT} can be negative. Table 3 shows sets of values that apply to extreme patterns. It is evident that specification of a set of F's or P's gives a more complete description of structure than possible from any one coefficient.

Natural populations

Realistic discussion of the structure of natural populations requires simultaneous consideration of all processes by which gene frequencies may change. These may be put into categories according to the degree of determinacy. There are, first, those which tend to bring about directed changes (Δq) according to some definite function of the gene frequencies. These can be classified exhaustively into recurrent mutation, immigration and selection. Collectively these lead either toward fixation of one allele or to a state of equilibrium between two or more alleles at which $\Delta q = 0$. Second are fluctuations (δq) which are indeterminate in direction but determinate in variance ($\sigma^2_{\delta q}$). These include fluctuations about the mean values of the coefficients for the steady processes and fluctuations due to accidents of sampling in the parentage of each generation. These variances tend to be cumulative with respect to the array of possible gene frequencies and thus to lead toward random fixation. This tendency is, however, balanced by the tendency toward equilibrium due to the steady processes. The resultant is a probability distribution of frequencies of gene frequencies which applies to any one strain in the long run, or to an array of strains, subject to the same conditions, at any one time. The formula for one pair of alleles is a relatively simple function of Δq and $\sigma^2_{\delta q}$ (Wright, 1938, 1939, 1949a):

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

It makes for clarity of thought to recognize a third category, changes in gene frequency that are best treated as indeterminate in both mean and variance; mutations that occur very rarely in the history of a species and similarly rare hybridization, unique selective incidents, etc.

Fig. 4 makes a comparison of distributions of gene frequencies that differ because of different amounts of fluctuation in a selection coefficient or because of different effective population numbers. We shall confine our attention to the latter here because of a close relation to the inbreeding coefficient F, but we note that there is a rough similarity in their effects that makes it possible to transfer general conclusions from one to the other (Wright, 1948a).

Fig. 5 shows the distribution of gene frequencies at different loci in local populations of a given effective size and subject to certain other conditions which are the same within each case.

Equilibrium at $q = 0{\cdot}50$ due to equal superiority of heterozygote over both homozygotes and equal mutation rates in both directions is assumed in one case. There are only slight departures from equilibrium if the selective advantage is relatively strong, much random drift into fixation if the selective advantage is sufficiently weak. In the other case, selection tends to drive one allele into fixation, while immigration tends to reduce its frequency to 0·10. Selection tends to dominate the situation at those loci in which the selective advantage is relatively strong, while low frequencies and loss are the rule where selection advantage is weak. The following discussion is largely restricted to type or near-type alleles among which selective differences are negligible. We are concerned here with the effect of population structure on the genetic background of alleles related to quantitative variability, rather than with major differences that are dominated by selection.

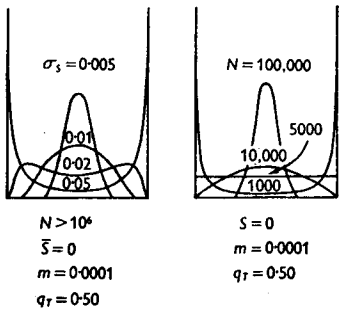


Fig. 4.

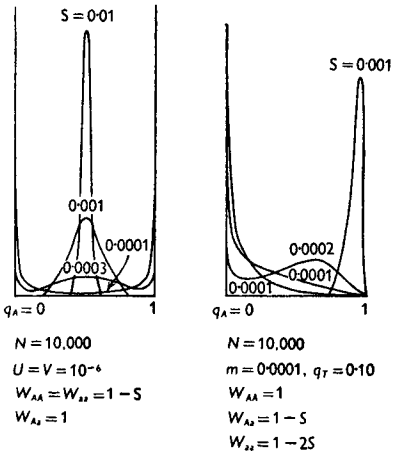


Fig. 5.

Fig. 4. Distribution of gene frequencies under two sets of conditions. In both, $\Delta q = -m(q - 0{\cdot}5)$, where m is the amount of replacement by immigrants with gene frequency $q_T = 0{\cdot}50$. In those to the left, random drift is due only to fluctuations in selection, σ_s,

$$\sigma^2_{\delta q} = \sigma^2_s q^2 (1-q)^2, \quad \phi(q) = Cq^{-2}(1-q)^{-2} \exp\{-m/[\sigma^2_s q(1-q)]\}.$$

In those to the right, fluctuations are due to small numbers, N,

$$\sigma^2_{\delta q} = q(1-q)/2N, \quad \phi(q) = C[q(1-q)]^{2Nm-1} \quad \text{(Wright, 1948}a\text{)}.$$

Fig. 5. Distributions of gene frequencies for loci which differ in the selective values of zygotes. In all cases random drift is due merely to small numbers, $N = 10{,}000$, $\sigma^2_{\delta q} = q(1-q)/20{,}000$. In those to the left, selection favours heterozygotes to varying extents (s), and there is mutation in both directions at equal rates ($u = v$), W is the selective value

$$\Delta q = -[2v + 2sq(1-q)][q - 0{\cdot}5], \quad \phi(q) = C[q(1-q)]^{-0{\cdot}96} \exp[40{,}000sq(1-q)].$$

In those to the right, selection at various rates is opposed by immigration, $m = 0{\cdot}0001$,

$$\Delta q = sq(1-q) - m(q - 0{\cdot}10), \quad \phi(q) = Cq^{-0{\cdot}6}(1-q)^{2{\cdot}6} \exp[40{,}000sq].$$

In the absence of selection, the mean gene frequency is the same, given a certain Δq, irrespective of the value of $\sigma^2_{\delta q}$. The variance of the distribution of gene frequencies is given by the formula $\sigma^2_q = q_T(1-q_T)F$ (Wright, 1943a) as well as by the formula $\sigma^2_q = \int_0^1 (q-\bar{q})^2 \phi(q)\,dq$. These formulae connect the two modes of attack on questions of population structure. The former gives additional significance to the coefficient F as a measure of the variance of a neutral gene frequency among

strains (σ_q^2), relative to the limiting value, $q_T(1-q_T)$ under complete fixation at the same gene frequency as in the total population.

The island model of structure

The mathematically simplest model for a heterogeneous natural population is that of an array of island populations of effective size N, largely isolated, but each replenished to the extent m from immigration, representative of the species as a whole (Wright, 1931, 1943*a*). By the method of path coefficients $F=(1-m)^2\left[\frac{1}{2N}+\left(1-\frac{1}{2N}\right)F'\right]$, where the prime indicates the preceding generation. In this case, F approaches a certain limiting value, other than 1. Putting $F=F'$

$$F=(1-m)^2/[2N-(2N-1)(1-m)^2],$$

$$F=\frac{1}{4Nm+1} \quad \text{approximately if } m \text{ is small.}$$

In the other mode of attack, we substitute $\Delta q=-m(q-q_T)$, the measure of the steady effect of immigration, and $\sigma_{\delta q}^2=q(1-q)/(2N)$, the sampling variance in one generation, in the general formula for $\phi(q)$:

$$\phi(q)=\frac{\Gamma(4Nm)}{\Gamma(4Nmq_T)\,\Gamma[4Nm(1-q_T)]}q^{4Nmq_T-1}(1-q)^{4Nm(1-q_T)-1},$$

$$\bar{q}=\int_0^1 q\phi(q)dq=q_T,$$

$$\sigma_q^2=\int_0^1(q-\bar{q})^2\phi(q)dq=\frac{q_T(1-q_T)}{4Nm+1},$$

$$F=\frac{1}{4Nm+1} \quad \text{in approximate agreement.}$$

Fig. 6 shows the distribution $\phi(q)$ for $q_T=\frac{1}{2}$ and various values of F. It is important to note that if F is as small as 0·05, the amount of differentiation in gene frequency with respect to neutral loci is considerable. Appreciable random fixation begins at $F=0{\cdot}33$.

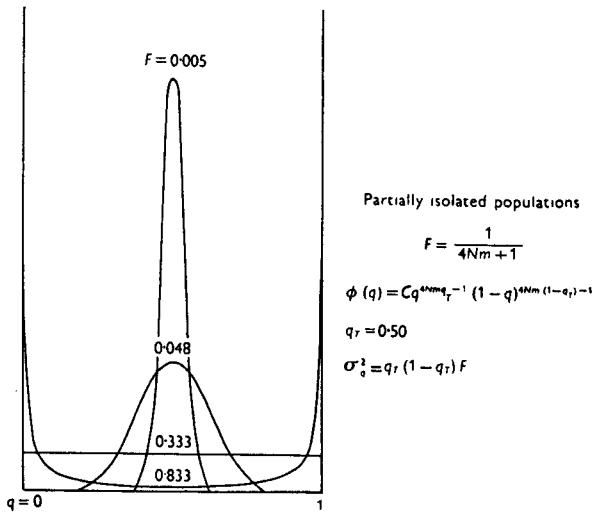


Fig. 6. The relation between variation of gene frequencies among island populations and the value of F in the total array due to the partial isolation of these. The mean gene frequency of the species, q_T, is assumed to be 0·50.

In an actual species, there may be a large central core of population in which gene frequency remains substantially constant, surrounded by small almost isolated populations with a great deal of differentiation in each element of genetic background.

Isolation by distance

Another useful model is that of a population of uniform density and a uniform and highly restricted amount of dispersal from each locality. The properties turn out to depend primarily on the size of neighbourhood, i.e. on the effective population number in an area from which the parents may be assumed to be drawn at random (Wright, 1940, 1943*a*, 1946). It appears that this is approximately the effective number in a circle of radius twice the standard deviation of the distribution of parent relative to offspring in one direction, largely irrespective of the form of this distribution curve (which seems usually to be leptokurtic; Dobzhansky & Wright, 1943, 1947; Bateman, 1947). We assume here that the variance of the distribution of grandparents is twice that of parents, and in general that the variance of ancestors of generations X is X times that of parents. These ancestors are thus drawn from an effective population of size XN_N.

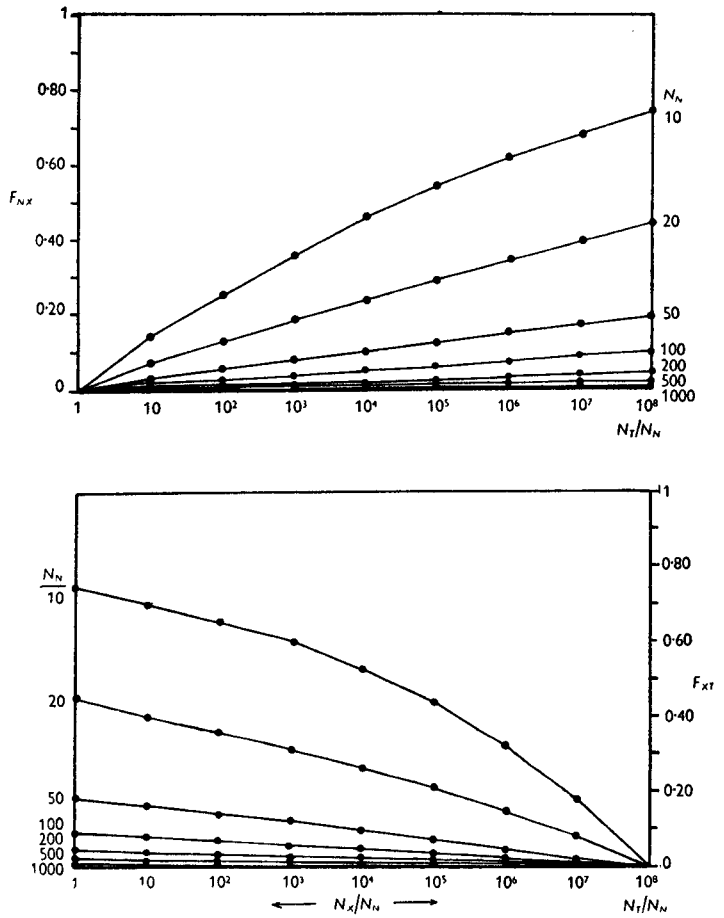


Fig. 7. Above, inbreeding coefficients (F_{NX}) of neighbourhoods of various sizes (N_N), relative to more comprehensive populations (N_X), and below, those (F_{XT}) of populations of various sizes (N_X), relative to a given total ($N_T = 10^8 N_N$) where there is continuity over an indefinitely large area. Abscissae in both: N_X/N_N.

The value of F_{NX}, measuring the inbreeding of neighbourhoods relative to a population of specified size N_X, is readily found by the method of path coefficients (Appendix F). Fig. 7 (top) shows how F_{NX} rises as N_X is taken larger. This may be interpreted, as noted, as measuring the amount of differentiation (at neutral loci) of neighbourhoods with respect to more comprehensive populations.

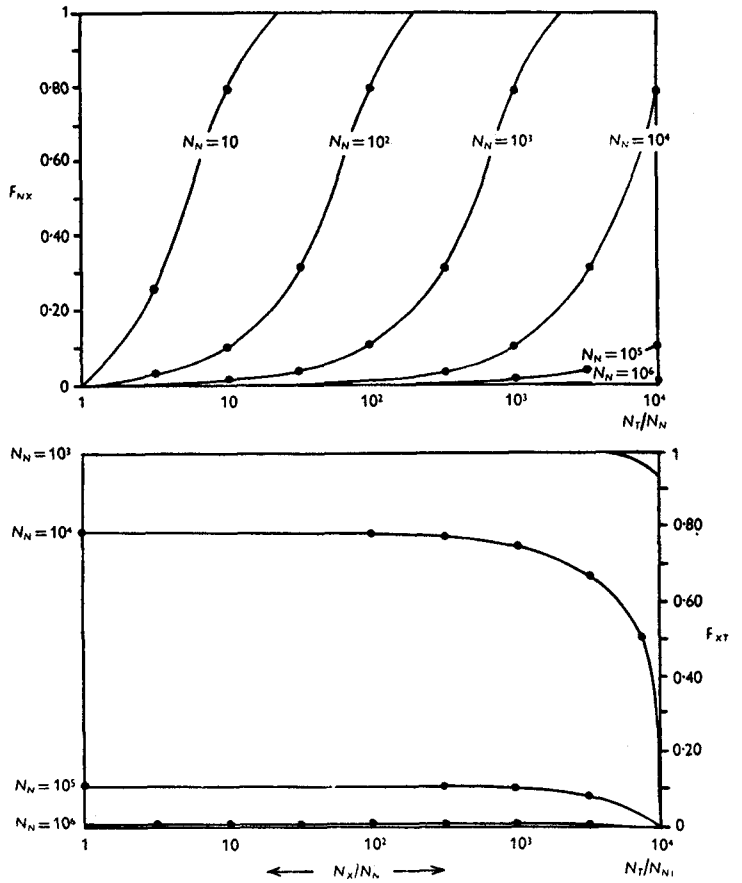


Fig. 8. Similar to Fig. 7 except that the range is along a single line instead of over an area. The total population considered, N_T, is merely $10^4 N_N$.

A better description of population structure can, however, be obtained by finding the amount of differentiation of populations of various specified sizes (N_X) within a total that is a given multiple of the neighbourhood. This depends on calculation of F_{XT} from F_{NX} and F_{NT} by the rule for hierarchic coefficients. Values are shown for increasing N_X relative to $N_T = 10^8 N_N$ in the lower part of Fig. 7.

It may be noted that differentiation of neighbourhoods carries with it considerable differentiation of much larger areas. This implies that there is a correlation between gene frequencies of neighbourhoods which falls off as their distance apart increases. Malécot (1948) has approached the problem of isolation by distance in a continuum by a different mathematical method and has made the determinations of the above correlation his primary objective. Returning to the method of attack by path coefficients, we may note that this correlation is given approximately by the ratio F_{XT}/F_{NT} for a distance of the order of $\sqrt{(N_X/N_N)}$ times the radius of a neighbourhood, and is thus the ratio of the ordinate at N_X/N_N to that at 1 in the lower part of Fig. 7.

One of the most important points that is brought out is that random local differentiation is slight under this model unless the effective number in neighbourhoods is decidedly small. Differentiation is very great if N_N is of the order of 20, not negligible if of the order of 200, but there is almost the equivalent of universal panmixia if it is as large as 1000.

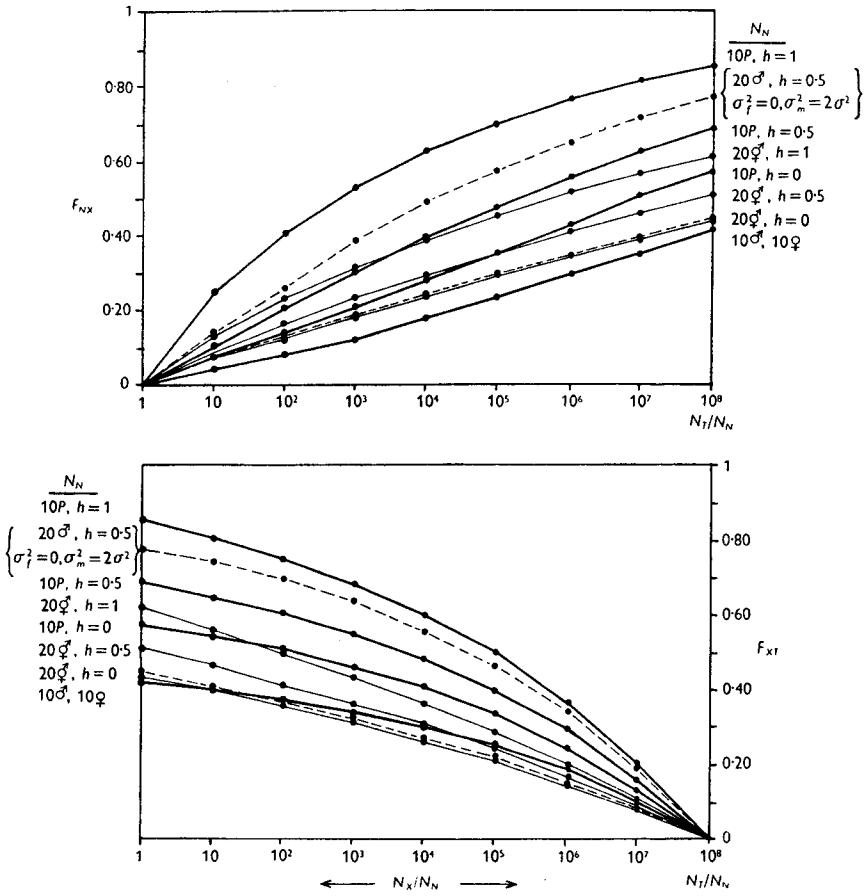


Fig. 9. Inbreeding coefficients F_{NX} and F_{XT} as in Fig. 7 but all with $N_N = 20$. Various systems of mating are compared (10♂, 10♀ mating at random, 20 monoecious individuals (20♀)) with varying amounts of self-fertilization ($h = 0$, 0·50, 1), ten permanent pairs (10P) with varying amount of brother-sister mating ($h = 0$, 0·50, 1) and 20♀ with no dispersal of ovules, double dispersion variance of pollen, either with merely random self-fertilization (broken line nearly same as 20♀ $h = 0$) or with a little more than 50% self-fertilization ($r = 0{\cdot}5$, $h = 0{\cdot}525$).

In the case of differentiation along a linear range, on the other hand, the ancestors of generation X came from a population of effective size $\sqrt{X N_N}$ instead of $X N_N$. Marked differentiation occurs with relatively large neighbourhoods (Fig. 8).

The hypothesis on which all of the preceding figures are based is the arbitrary but mathematically simple one of random union of gametes from N_N monoecious individuals. There is no appreciable difference if self-fertilization is excluded ($h = 0$). The effects of other systems of mating are compared in Fig. 9 (area continuity, $N_N = 20$). The amount of differentiation increases considerably with increase in the amount of self-fertilization ($h = 0{\cdot}5$, $h = 1$). In the preceding cases equal dispersion of the sexes is assumed. If there is no appreciable dispersion of one sex, but dispersion of double the variance in the preceding cases, in the case of the other sex and if

self-fertilization is no more than random, the result is practically as above, but with increased self-fertilization there is disproportionate increase in differentiation. With separate sexes and random mating (10♂, 10♀ per neighbourhood) the effect is again rather similar to that from random union of gametes of twenty monoecious organisms. If this same number of individuals is in permanent pairs (10P) instead of mating at random, differentiation is greater.

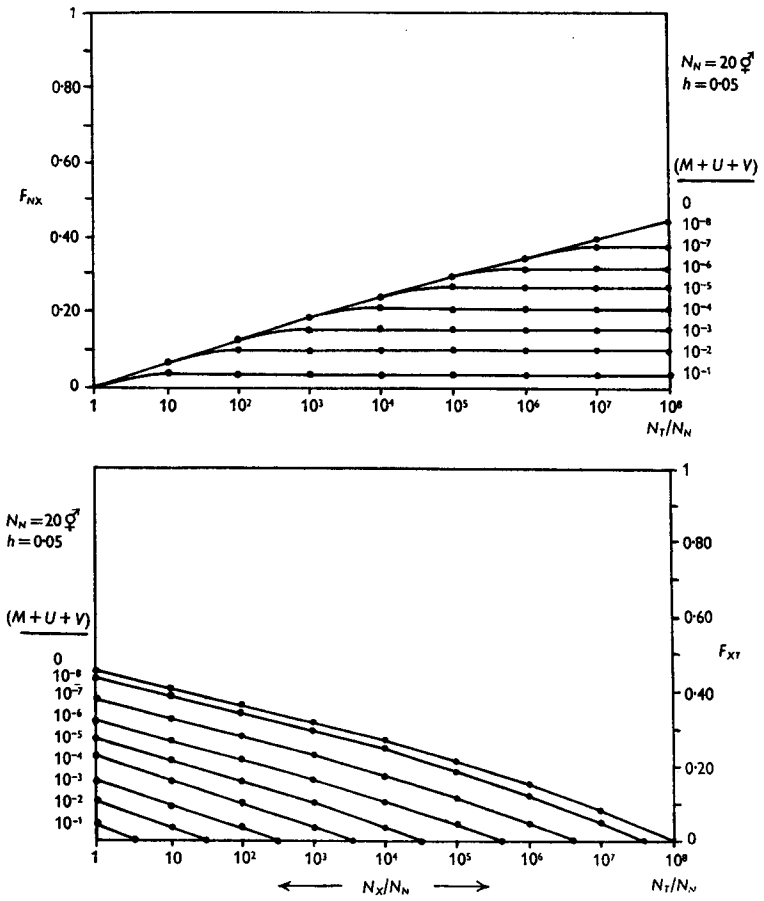


Fig. 10. Inbreeding coefficients F_{NX} and F_{XT} with $N_N = 20$, random union of gametes, but with universal dispersion (m) or reversible mutation (u, v) preventing complete fixation.

All of the cases discussed so far apply only when mutation rates are very low (less than 10^{-8} with $N_T/N_N = 10^8$), and there is no mode of occasional dispersal over the whole range of the species in one generation. It is important to consider what happens if these conditions do not hold. Fig. 10 shows calculations from random union of gametes in neighbourhoods of size 20, but mutation or universal dispersion displacing all populations to extents ranging from 10^{-1} to 10^{-8}. It may be seen that F_{NX} no longer approaches 1 asymptotically, but rather abruptly approaches a certain upper limit in each case. Correspondingly there is a definite limit to the size of population which shows any differentiation as a result of the differentiation of neighbourhoods. These are the results which can be most directly compared with those of Malécot. His final formula for the correlation as a function of distance is in a different mathematical form (involving Bessel functions) and does not involve size of neighbourhood. Empirically this correlation, as the ratio

F_{XT}/F_{NT}, seems to be almost independent of N_N in my results. I am indebted to Mr W. S. Russell for making calculations at three values of m (10^{-2}, 10^{-4}, 10^{-6}) from Malécot's formula. The results are in substantial agreement with those from F_{XT}/F_{NT}.

We may note here that if the values of F are small, systematic pressures (Δq) involving selection may be treated approximately in the same way as those involving m by using the best

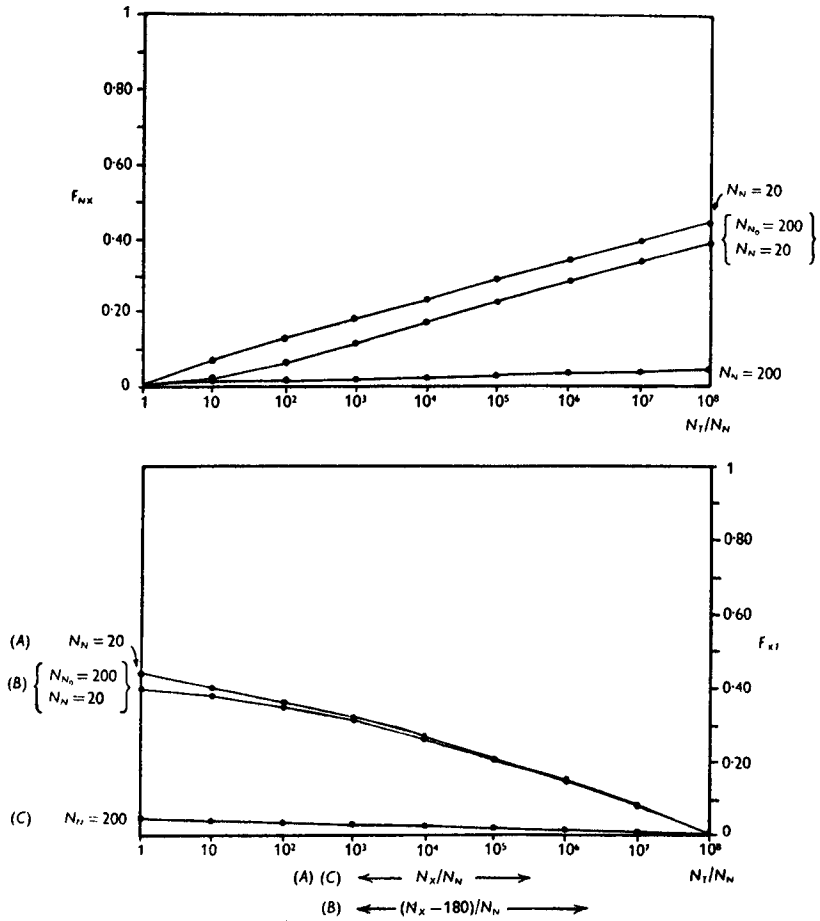


Fig. 11. Inbreeding coefficients F_{NX} and F_{XT} with random union of gametes but with three conditions with respect to numbers in neighbourhoods and to ancestral population numbers. In the highest curve in each case there is homogeneity with respect to density and dispersion and the number in a neighbourhood is 20. In the lowest curve in each case there is also homogeneity with respect to density and dispersion but N_N is 200. In the intermediate curve in each case, the neighbourhoods considered have the number 200 but because of heterogeneity in the conditions, the numbers in ancestral generations increase only by 20's.

linear expression for Δq near the equilibrium point (Wright, 1943*a*). Diversity in degree and direction of selection among localities is of course a wholly different matter and may bring about great differentiation if not overbalanced by migration (Wright, 1931, 1943*a*).

Actual populations are likely to show great heterogeneity in density and amount of dispersion. Some appreciation of the effects can be obtained by modifying the mathematical model. Assume that there are scattered relatively populous neighbourhoods with effective number MN, but that the ancestors of generation X are drawn from a population of $(M+X-1)N$ instead of

XMN or XN. The parents in such localities are thus drawn from relatively large populations, and the populations from which ancestors are drawn increase linearly with number of generations but at a low rate. Fig. 11 makes comparisons for $MN = 200$ but N only 20, with those for $N_N = 200$ and $N_N = 20$. The amount of differentiation in the heterogeneous case is much like that with N_N equal to 20 except for a lag.

It will probably be difficult to find actual cases in nature that correspond closely to these ideal cases. The primary difficulty is that data can only be obtained for cases in which there are conspicuous differences in the effects of alleles, which are cases in which selection is likely to be the dominating factor. Fig. 12, however, presents a possible case based on observed variation in flower colour (blue and white) of a small plant, *Linanthus Parryae* in an area of some 80 × 10 miles along the piedmont of the San Gabriel and San Bernardino mountains, which it occupies rather uniformly. 1258 samples of 100 plants each, distributed in a systematic way in this area, were examined by Epling & Dobzhansky (1942). The mode of inheritance is unfortunately not known, but estimates of effective size of neighbourhood on the hypothesis of no selection differed little (about 15–25) whether blue is recessive (assumed in the curves shown) or dominant or dependent on multiple factors and a threshold. Whatever the mechanism of differentiation it appears that there was marked differentiation at all levels, including not only that among six primary divisions but also of secondary and smaller divisions within the primary divisions. The curves indicate at least roughly the type of population structure discussed here (Wright, 1943*b*).

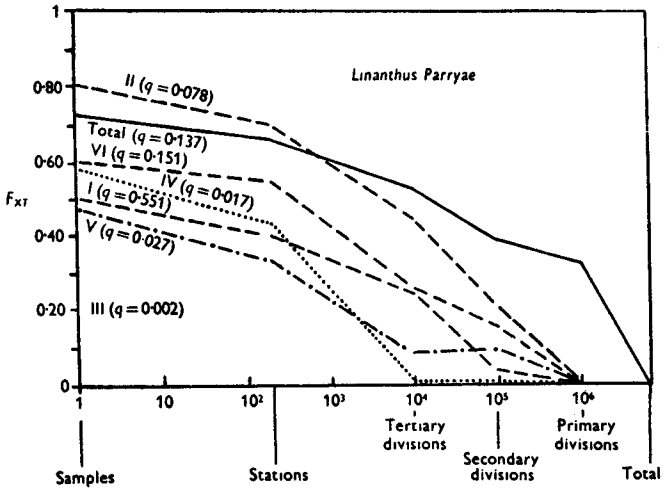


Fig. 12. Inbreeding coefficients (F_{XT}) of areas of various sizes in the total population and in the six primary subdivisions of this in *Linanthus Parryae* under the assumptions in the text.

Population structure in evolution

In considering the significance of population structure in evolution, we must distinguish two aspects of the process: multiplication of species and transformation. Partial isolation obviously tends toward the splitting off of a new species if the environmental conditions are markedly different. The thesis which I wish to stress, however, is that a fine-scaled structure of partial isolation, without marked environmental differences, presents the most favourable condition for transformation as a single species (Wright, 1931, 1932 and later).

The advantages over a panmictic population of comparable total size are of several sorts. In the first place, the subdivided population maintains more alleles at each locus and more at

moderately high frequencies. In a panmictic species there is a tendency toward the establishment of one type allele at each locus. Possible favourable mutations at two or more removes (i.e. mutations of mutations) have little chance of occurrence. There are, to be sure, ways in which two or more alleles may be kept at subequal frequencies (selective advantage of heterozygotes, or advantage of mere rarity), but these are poor compared with those in a subdivided species. In the latter, different alleles may easily be maintained as the types in different places by selective advantages in relation to different conditions, as already indicated, and up to a certain point this favours adaptability of the species as a whole rather than splitting. Of more importance for transformation as a whole is local differentiation with respect to genotypes that give adaptation to the same conditions. The best adapted form in a species is usually one that is close to the average in all quantitatively varying characters. There are certain to be a great many genotypes that are of approximately this same optimal type except for pleiotropic genic effects that are of less selective significance at the time. In a panmictic population one of these genotypes inevitably gets a lead over the others and tends to become fixed. In a subdivided population, local fluctuations in conditions that shift slightly the position of the optimum insure the predominance of different genotypes of essentially the same phenotype in different places (Wright, 1935 *a*, *b*). This may indeed be brought about in sufficiently small local populations by the accumulation of sampling accidents without any fluctuations in conditions. This process, moreover, brings about differentiation in all other loci in which there are two or more type or near-type alleles with differences in selective value that are below a certain critical ratio to the reciprocal of the effective local population number (Wright, 1931). It may be well to note here that the effective population number may be expected to be much less than the apparent one for a variety of reasons (Wright, 1931, 1940).

The kaleidoscopic shiftings in the sets of gene frequencies of neighbourhoods may be expected to lead occasionally to predominance of combinations with selective advantages under the prevailing conditions that are far in excess of that indicated by the sum of the net effects of the components. Control by selection thereupon supersedes control by stochastic processes and immigration. Here we come to the second great advantage of a subdivided population over a panmictic one. Favourable combinations may spread through the species by a selective process that operates on the whole instead of merely on the separate net effects of the components.

The conditions for such a spread are more favourable in a continuum with isolation merely by distance than under the island model. In the former, centres in which a favourable combination happens to have appeared increase in population and contribute more than their share of migrants to adjacent regions. Because of the strong correlation in all gene frequencies, these adjacent regions do not need much immigration of this sort to carry them beyond the critical point at which selective increase of the combination becomes autonomous. The combination thus spreads through the species in concentric circles without any appreciable mass displacement of populations. Favourable combinations in other respects may arise at other centres and spread similarly, and the two circles of spread may overlap and cross without much interference. On the other hand, combinations that increase adaptation to the same conditions, but which differ genotypically, interfere, and the superior one ultimately displaces the other unless some aspects of both lead to a combination that is better than either, which thereupon spreads concentrically from its point of origin. With fine scaled structure, new differentiation may be expected to occur concomitantly with the spreading of combinations.

Any number of genes may contribute to a favourable new combination. Some may play an essential role and others that of minor modifiers. An important case is that of a single major mutation that has the potentiality of great advantage if it can be combined with an array of modifiers that bring its effects into harmonious adjustment with other characters. The importance of homoeotic mutations in evolution has been much stressed by Goldschmidt (1940). Such mutations usually have low penetrance and may thus be carried by a species and diffused through it at appreciable gene frequencies in spite of severe antagonistic selection when manifested in their pristine raw form. There is little chance of reaching adjustment in the face of this selection in a panmictic species. In a finely subdivided one, there is a good chance that somewhere, at some time, a combination of modifiers will be encountered that gives sufficient adjustment to reverse the direction of selection locally and thus create a centre from which the system, homoeotic mutation and modifiers can spread as a unit through the species.

Ecologic opportunity

It must be recognized that most species are held to such restricted ecologic niches by the pressure of other species that only an increase in specialization for this niche is probable even with the most favourable population structure. If, however, a more extensive ecologic opportunity is ever presented, a finely divided structure is that which is most favourable for its rapid exploitation (Wright, 1949*a*,*b*). Such an opportunity arises when migrants reach territory in which there are many unoccupied niches in which they can live. It also arises when a permanent change in conditions eliminates many species but leaves the relatively pre-adapted survivors to take over the vacant niches. Most important of all perhaps are the occasional cases in which a specialization turns out to be usable in a hitherto unoccupied way of life or to give a general advantage in many ways of life. The origin of the higher categories seems to have its basis in such ecological opportunities (Wright, 1941, 1948*b*, 1949*a*,*b*).

Evolution in general

Organic evolution is not the only sort of evolution in the sense of a process of cumulative change. When a level of intelligence was reached in an anthropoid line that made symbolic speech possible, a new evolutionary process emerged, enormously more rapid than organic evolution. The line of persistence and accumulation was that from speaker to listener, and later from writer to reader, instead of the germ line. The principles of Mendelian heredity do not apply to the evolution of culture. Nevertheless, the general qualitative conclusion would still seem to hold that this or any other evolutionary process depends on a continually shifting but never obliterated state of balance between factors of persistence and change, and that the most favourable condition for this occurs where there is a finely subdivided structure in which isolation and cross-communication keep in proper balance.

APPENDIX A

The method of path coefficients

As most of this discussion rests on the values taken by the inbreeding coefficient, F, under diverse population structures, and these values have been found by the method of path coefficients, we begin with a brief sketch of this method.

It is a method of dealing with linear systems of variables that are closed in the sense that each variable is either represented as linearly and completely determined by others in the system or is

one of the ultimate factors, with indicated correlations with all other ultimate factors. The use of the method is greatly facilitated by diagrams in which arrows run to each variable that is represented as dependent from those represented as affecting it directly, and the system is completed by double-headed arrows to indicate correlations between ultimate factors, due to unknown common factors, in all cases in which a zero correlation cannot be safely postulated.

Let V_0, V_1, V_2, etc., be the variables. Express each in standard form ($X_0 = (V_0 - \overline{V}_0)/\sigma_0$, etc.). Assume that V_0 is represented as linearly and completely determined by $V_1, V_2, \ldots, V_k$:

$$X_0 = \sum_{i=1}^{k} p_{0i} X_i.$$

A coefficient p_{0i} pertaining to the indicated path of influence $V_i \rightarrow V_0$ is known as an elementary path coefficient.

If V_q is any variable in the system and there are n sets of observations:

$$r_{0q} = \frac{1}{n} \sum^{n} X_0 X_q$$

$$= \sum_{i=1}^{k} p_{0i} r_{iq} = \sum_{i=1}^{k} p_{0\overline{iq}}.$$

Any of the correlations, r_{iq}, may itself be capable of such analysis through variables by which either V_i or V_q is represented as determined. On carrying such analysis back as far as the system permits, we arrive at the basic principle that the correlation between any two variables in the system is equal to the sum of contributions (compound path coefficients such as $p_{0\overline{iq}}$ above) pertaining to the paths by which one may go from one to the other in the diagram without going forward and then back and without passing through any variable twice in the same path. The contribution of such a connecting path is the product of the path coefficients pertaining to the elementary paths along its course. Thus, if $p_{ac\dot{d}b}$ indicates a compound path coefficient relating to a path of the type $a \leftarrow c \leftarrow d \rightarrow b$, its value is $p_{ac} p_{cd} p_{bd}$. The symbol $p_{a\overline{cd}b}$ indicates a connexion of the type $a \leftarrow c \longleftrightarrow d \rightarrow b$ with value $p_{ac} p_{cd} p_{bd}$ in which $p_{cd} = r_{cd}$.

If we express the correlation of a variable with itself (necessarily 1 in a closed system), we obtain a useful equation expressing complete determination:

$$r_{00} = \sum_{i=1}^{k} p_{0i} r_{0i} = \sum^{k} p_{0i}^2 + 2\Sigma p_{0i} p_{0j} r_{ij} = 1 \quad (i < j).$$

It may be noted that if a variable (e.g. V_u) is omitted, $\Sigma p_{0i} r_{0i} (= 1 - r_{0u}^2)$ is the squared multiple correlation between V_0 and the determining variables that are included, and r_{0u}^2 is the degree of determination by the residual variables. The method is identical with those of multiple regression and of factor analysis in the appropriate closed systems but is designed for algebraic use in irregular systems with intermingled known and hypothetical variables, known and unknown path coefficients and correlation coefficients, in which these methods are not applicable in the conventional ways.

The applicability of the method is greatly extended by use of the principle that a variable that is derived from another by a process of random sampling may be represented as linearly and completely determined by the latter and a hypothetical variable, 'accidents of sampling', that has a zero correlation with all variables in the system that do not involve the same actual process of sampling.

Finally, it may be noted that it makes no difference in principle, in a closed system, which variables are treated as determined by which, and which are represented as ultimate factors, but there are usually certain arrangements that are more interesting than others from the standpoint of interpretation.

APPENDIX B

General coefficients of inbreeding (Fig. 14)

In applying the method to zygotes (autosomal diploid locus) and gametes, in a diagram a system of mating (Fig. 14, left), we note first that a zygote may be considered to be linearly, completely and equally determined by the gametes that unite at fertilization. Let a be the path coefficient relating a zygote to one of the determining gametes and let F be the correlation between the latter

$$2a^2 + 2a^2F = 1, \quad a = \sqrt{\frac{1}{2(1+F)}}.$$

The path coefficient, b, relating a gamete to the zygote that produces it, is also the correlation coefficient as there is only one connecting path, and is equal to the correlation $(a' + a'F')$ between this gamete and a gamete that contributed to it in the preceding generation (indicated by a prime). It is assumed that no systematic changes in gene frequency due to selection, mutation or immigration intervene. Then

$$b = \tfrac{1}{2}(1 + F').$$

The compound path coefficient, ba', relating gamete in one generation to one in the preceding, has the value $\frac{1}{2}$ irrespective of the system of mating, under the postulated conditions.

Application of the general principle stated in Appendix A gives at once the general formula (for diploid autosomal loci)

$$F = \Sigma[(\tfrac{1}{2})^n(1 + F_A)],$$

in which n is the number of zygotes in a connecting path (including sire and dam) and F_A is the inbreeding coefficient of the common ancestor (A) to which sire and dam trace in this particular path. The summation relates to all paths by which sire and dam are connected, without tracing forward to descendants and then back and without passing through any individual twice in the same path.

This formula has been given previously (Wright, 1922*a* and in many later papers) in the form

$$F = \Sigma[(\tfrac{1}{2})^{n_s + n_d + 1}(1 + F_A)]$$

in which n_s and n_d are the number of *generations* from sire and dam respectively to the ancestor A. The present form leads to a simpler analogous form in the case of sex-linked inheritance than does the older one.

In the case of sex-linkage (males treated as XO, females XX), the constitution of males is completely determined by that of the egg and completely determines that of their own gametes (Fig. 14, middle). There is no contribution from a connecting path that includes males in succession:

$$F = \Sigma[(\tfrac{1}{2})^{n_f}(1 + F_A)].$$

In this case, n_f is the number of females in a connecting path and the summation relates to paths which have no males in succession. F_A is treated as 0 if the connecting ancestor is a male.

The contributions of ancestral connexions in three typical cases are shown in Fig. 13.

The method has been extended to polysomic loci. In the diagram (Fig. 14, right) the many paths connecting the ultimate factors are omitted for simplicity. We define F here as the correlation between genes of the same zygote, F_D as that between genes that enter the zygote from different

gametes, and F_S as that between genes that enter the zygote in the same gamete. F_S would equal F' except for the small chance, e, that two genes of the zygote trace to one in the preceding generation. In a $2k$-somic case

$$F_S = (1-e)\,F' + e,$$

$$F_D = \Sigma\left\{\left(\frac{1}{2k}\right)^n [1+(2k-1)\,F_A]\right\},$$

$$F = [(k-1)\,F_S + kF_D]/(2k-1).$$

It is to be noted that none of these general formulae for F involves gene frequency. Thus, a single coefficient characterizes the effects of a given population structure on all autosomal disomic loci, irrespective of gene frequencies in the absence of differences due to selection or mutation. The effects of immigration may usually be incorporated into the representation of population structure. Other coefficients characterize the effects on sex-linked and polysomic loci.

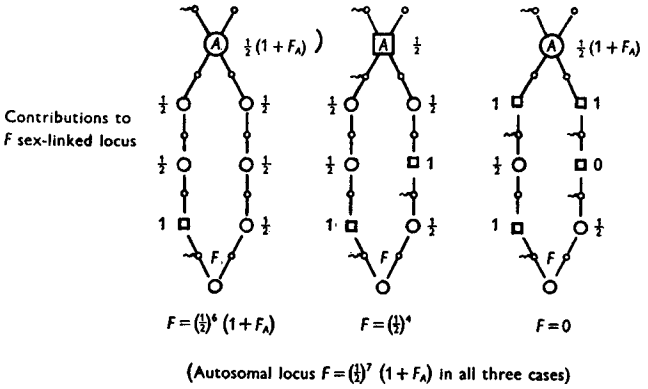


Fig. 13. Contribution of ancestral connexions to the value of F in three typical cases.

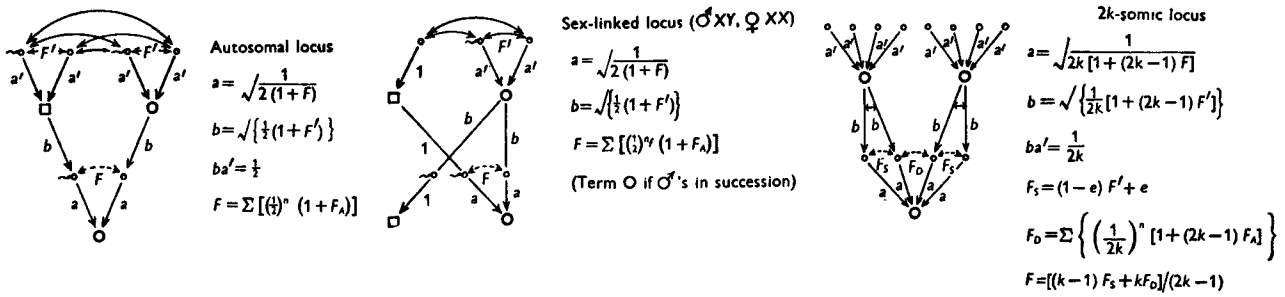


Fig. 14. Diagrams for analysis of inbreeding in the cases of autosomal loci, sex-linked loci and $2k$-somic loci. In the last case, the numerous two-headed arrows connecting the ultimate variables (gametes) are omitted for simplicity. The dotted two-headed arrows indicate correlations to be analysed in terms of connecting paths.

A coefficient F, as the correlation between alleles, could be found in principle for each locus subject to significant mutation rates of selective differentials. In the latter case, it would depend not only on the selection coefficient but also on gene frequency, except in so far as selection pressure can be treated as linearly related to gene frequency in the range of gene frequencies involved.

It is obvious from Table 1 that the degree of approach towards complete fixation in the case of two alleles under a system of mating that does not alter gene frequency in the total (diploid) population is measured by the same coefficient F that gives the correlation between alleles of the same zygote. The derivation of the general formulae for F in the latter sense applies as well to

multiple alleles as to pairs of alleles, and it is easy to see that the other interpretation also applies to multiple alleles. The Mendelian mechanism is such that any group of alleles may be treated formally as if a single allele. Thus, if A_i and A_j in a multiple series are treated as if a single allele A_{ij} with frequency $q_{ij} = q_i + q_j$, the frequency of 'homozygotes' $A_{ij}A_{ij}$ must be $(1-F)\,q_{ij}^2 + Fq_{ij}$. If we subtract from this the frequencies of A_iA_i and A_jA_j which must be $(1-F)\,q_i^2 + Fq_i$ and $(1-F)\,q_j^2 + Fq_j$ respectively, recalling that F is independent of gene frequency, we find that the frequency of A_iA_j is $2(1-F)\,q_iq_j$ and that of all heterozygotes must be of this type. The array of zygotic frequencies can accordingly be analysed into a panmictic component $(1-F)\,[\sum^n q_iA_i]^2$ and a fixed component $F[\sum^n q_iA_i^2]$ as in Table 2. The correlation between alleles of the same zygote is the weighted average of the correlations within these portions (0 and 1 respectively), and is thus F irrespective of gene frequencies and of the values assigned to those alleles, as expected.

APPENDIX C

Properties of populations as related to F

The relation of the mean and variance of characters to the inbreeding coefficient, noted in the text, apply to multiple alleles as well as to pairs of alleles. We assume here that contributions of loci and of environmental factors are additive. The array of zygotic frequencies of n alleles of type A_i gene frequency q_i is

$$(1-F)\,[\sum^n q_iA_i]^2 + F[\Sigma q_iA_iA_i].$$

Let c_{ij} represent the contribution of A_iA_j to the character. Let $m_{T(F)}$ be the mean and $\sigma^2_{T(F)}$ the genetic component of the variance under a system of mating characterized by coefficient F. Then $m_{T(0)}$ and $\sigma^2_{T(0)}$ are the corresponding statistics under random mating ($F = 0$), and $m_{T(1)}$ and $\sigma^2_{T(1)}$ those for an array of completely fixed lines ($F = 1$), arrived at without selection:

$$\begin{aligned}
m_{T(F)} &= (1-F)\sum_{j=1}^{n}\sum_{i=1}^{n}(c_{ij}q_iq_j) + F\sum_{i=1}^{n}(c_{ii}q_i)\\
&= (1-F)\,m_{T(0)} + Fm_{T(1)}\\
&= m_{T(0)} + F[m_{T(1)} - m_{T(0)}],\\
\sigma^2_{T(F)} &= (1-F)\,\Sigma\Sigma\,(c_{ij}^2q_iq_j) + F\Sigma(c_{ii}^2q_i) - m^2_{T(F)}\\
&= (1-F)\,(\sigma^2_{T(0)} + m^2_{T(0)}) + F(\sigma^2_{T(1)} + m^2_{T(1)}) - [(1-F)\,m_{T(0)} + Fm_{T(1)}]^2\\
&= (1-F)\,\sigma^2_{T(0)} + F\sigma^2_{T(1)} + F(1-F)\,[m_{T(1)} - m_{T(0)}]^2.
\end{aligned}$$

Thus the change in mean due to the system of mating is proportional to F, but that in the total variance is a quadratic fraction of F.

The case of semi-dominance at all pertinent loci is of interest. Let c_i be contribution of allele A_i and $c_{ij} = c_i + c_j$. The mean is unaffected by the system of mating in this case:

$$\begin{aligned}
m_{T(F)} &= m_{T(0)} = m_{T(1)} = 2\sum^n c_iq_i,\\
\sigma^2_{T(0)} &= \sum_{j=1}^{n}\sum_{i=1}^{n}(c_i + c_j)^2\,q_iq_j - 4(\sum^n c_iq_i)^2\\
&= 2\sum^n c_i^2q_i - 2(\sum^n c_iq_i)^2 = 2\sigma^2_{c_i},\\
\sigma^2_{T(1)} &= 4\sum^n c_i^2q_i - 4(\Sigma c_iq_i)^2 = 4\sigma^2_{c_i},\\
\sigma^2_{T(F)} &= (1-F)\,\sigma^2_{T(0)} + F\sigma^2_{T(1)}\\
&= (1+F)\,\sigma^2_{T(0)}.
\end{aligned}$$

If the inbreeding is due to isolation of a number (n) of strains (s) breeding at random within themselves, the variance within one is as follows, letting q_{is} be the frequency of A_i in this strain:

$$\sigma^2_{S(0)} = 2[\sum^n c_i^2 q_{is} - (\sum^n c_i q_{is})^2]$$

$$= 2[\sum^n c_i^2 q_{is}(1-q_{is}) - 2\Sigma c_i c_j q_{is} q_{js}] \quad (i<j).$$

But $2q_{is}(1-q_{is})$ is the amount of heterozygosis of A_i with all other alleles collectively, within the strain. The average for all strains is $2q_i(1-q_i)\,(1-F)$. Similarly, $2q_{is}q_{js}$ is the proportion of the specific heterozygote A_iA_j within the strain, the average of which for all strains is $2q_iq_j(1-F)$. Thus, the average intra-strain variance is

$$\overline{\sigma^2_{S(0)}} = 2(1-F)\,[\sum^n c_i^2 q_i(1-q_i) - 2\Sigma c_i c_j q_i q_j], \quad (i<j)$$

$$= 2(1-F)\,[\sum^n c_i^2 q_i - (\Sigma c_i q_i)^2]$$

$$= (1-F)\,\sigma^2_{T(0)}.$$

The variance of strain means is

$$\sigma^2_{ms}(= \sigma^2_{T(F)} - \sigma^2_{S(0)}),$$

$$\sigma^2_{ms} = 2F\sigma^2_{T(0)}.$$

The genetic correlation between any relatives is readily found by the method of path coefficients assuming additive effects of genes.

In the case of autosomal diploids, the correlation between propositi Z_1 and Z_2 is given by

$$r_{Z_1Z_2} = \frac{\Sigma[(\frac{1}{2})^{n-1}(1+F_A)]}{\sqrt{\{(1+F_{Z_1})(1+F_{Z_2})\}}} \quad \text{(Wright, 1922}a\text{)}.$$

For sex-linked loci

$$r_{Z_1Z_2} = \frac{\Sigma[(\frac{1}{2})^{n_f-\frac{1}{2}n_{Zf}}(1+F_A)]}{\sqrt{\{(1+F_{Z_1})(1+F_{Z_2})\}}},$$

where n_{Zf} is the number of female propositi (0, 1 or 2). These formulae may be modified in the usual ways to take care of dominance, types of factor interaction and environmental effects.

APPENDIX D

The inbreeding coefficient of breeds

The estimation of the inbreeding coefficient of large populations on the basis of pedigrees requires use of sampling methods. The best method seems to be to take an adequate random sample of individuals and trace single *random* lines back of sire and dam (Wright & McPhee, 1925). These either show a common ancestor or they do not. No attention need be paid to the generation in which a tie occurs (autosomal F), since remoteness is exactly compensated for by increase in the number of possible ties ($2^{n_s+n_d}$) of which the one observed is representative. In m such two-line pedigrees with k observed ties,

$$F = \frac{1}{m}\sum^k [2^{n_s+n_d}(\tfrac{1}{2})^{n_s+n_d+1}(1+F_A)]$$

$$= \frac{k}{2m}(1+\bar{F}_A).$$

The value of F for the sixty-four cows of Bates's Duchess strain of Shorthorns was 0·409 from tracing of all pedigrees to the beginning of the Coates herd book. The estimate from random two-line pedigrees was $0{\cdot}422 \pm 0{\cdot}011$.

The situation is more complicated in the case of sex-linked F. F applies here only to females (XX). It is necessary to calculate the number of possible ties between lines with no successive males.

Possible contributory lines back of sire

	Ancestral generations								
Ancestor	0	1	2	3	4	5	6	7	n_s
Male	1	0	1	1	2	3	5	8	$f(n_s-1)$
Female	—	1	1	2	3	5	8	13	$f(n_s)$
Total	1	1	2	3	5	8	13	21	$f(n_s+1)$

Possible contributory lines back of dam

	Ancestral generations								
Ancestor	0	1	2	3	4	5	6	7	n_d
Male	—	1	1	2	3	5	8	13	$f(n_d)$
Female	1	1	2	3	5	8	13	21	$f(n_d+1)$
Total	1	2	3	5	8	13	21	34	$f(n_d+2)$

Of the 2^{n_s} lines tracing to ancestors of the n_sth ancestral generation back of the sire, only $f(n_s+1)$ involve no successive males, where $f(n_s)$ is the n_sth Fibonacci number starting from $f(0) = 0, f(1) = 1$, and following the rule that each is the sum of the two preceding

$$f(n) = f(n-1)+f(n-2).$$

In the case of females, the number of lines tracing to ancestors of the n_dth ancestral generation that involve no successive males is $f(n_d+2)$. If m two-line pedigrees of females show k ties and n_f is the number of females, including the dam, in a typical tie,

$$F = \frac{1}{m}\sum^{k}\left[\frac{f(n_s+1)f(n_d+2)\,(1+F_A)}{2^{n_f}}\right].$$

APPENDIX E

Regular systems of mating

In the case of regular systems of mating, the diagrammatic representation may usually be closed after a small number of ancestral generations. Consider the case of a population of monoecious individuals of constant effective number N in which fertilization is at random except for a known proportion, h, of self-fertilization, not necessarily $1/N$. Let E be the correlation between random gametes and F that between uniting gametes (autosomal diploid loci):

$$E = \frac{1}{N}b^2+\left(1-\frac{1}{N}\right)E',$$

$$F = hb^2+(1-h)E',$$

$$b^2 = \tfrac{1}{2}(1+F'),$$

$$E' = \left(\frac{F-\frac{1}{2}h(1+F')}{(1-h)}\right) = \frac{1}{N}\left(\frac{1+F''}{2}\right)+\left(1-\frac{1}{N}\right)\left(\frac{2F'-h(1+F'')}{2(1-h)}\right)$$

Solving for F:

$$F = \left(1+\frac{h}{2}-\frac{1}{N}\right)F' + \left(\frac{1}{2N}-\frac{h}{2}\right)F'' + \frac{1}{2N},$$

$$P = \left(1+\frac{h}{2}-\frac{1}{N}\right)P' + \left(\frac{1}{2N}-\frac{h}{2}\right)P''.$$

If $h = 0$, $P = \left(1-\frac{1}{N}\right)P' + \frac{1}{2N}P''$.

If $h = \frac{1}{N}$, $P = \left(1-\frac{1}{2N}\right)P'$.

If $h = 1$, $(P-\frac{1}{2}P') = \left(1-\frac{1}{N}\right)(P'-\frac{1}{2}P'')$, $P = \frac{1}{2}P'$.

The rate at which heterozygosis falls off in the absence of self-fertilization can be found by equating P/P' to P'/P'' (rapidly approached in a population of constant size):

$$P = \frac{1}{2}\left[1-\frac{1}{N}+\sqrt{\left(1+\frac{1}{N^2}\right)}\right]P'$$

$$= \left[1-\frac{1}{2N}\left(1-\frac{1}{2N}\right)\right]P' \quad \text{approximately.}$$

It is evident that it makes no appreciable difference whether self-fertilization occurs at random or is excluded unless N is very small. In the other limiting case, $h = 1$, $P = \frac{1}{2}P'$ as expected.

The case of a population with separate sexes, N_p permanent pairs, and the proportion h of brother-sister mating can be worked out similarly:

$$E = \frac{1}{N_p}(ba')^2(2F'+2b'^2) + 4\left(1-\frac{1}{N_p}\right)(ba')^2E'$$

$$= \frac{1}{4N_p}(1+2F'+F'') + \left(1-\frac{1}{N_p}\right)E',$$

$$F = \tfrac{1}{4}h(1+2F'+F'') + (1-h)E',$$

$$E' = \frac{F-\frac{1}{4}h(1+2F'+F'')}{1-h} = \frac{1}{4N_p}(1+2F''+F''') + \left(1-\frac{1}{N_p}\right)\left(\frac{F'-\frac{1}{4}h(1+2F''+F''')}{1-h}\right),$$

$$F = \left(1+\frac{h}{2}-\frac{1}{N_p}\right)F' + \left(\frac{1}{2N_p}-\frac{h}{4}\right)F'' + \left(\frac{1}{4N_p}-\frac{h}{4}\right)F''' + \frac{1}{4N_p},$$

$$P = \left(1+\frac{h}{2}-\frac{1}{N_p}\right)P' + \left(\frac{1}{2N_p}-\frac{h}{4}\right)P'' + \left(\frac{1}{4N_p}-\frac{h}{4}\right)P'''.$$

If $h = 0$, $P = \left(1-\frac{1}{N_p}\right)P' + \frac{1}{2N_p}P'' + \frac{1}{4N_p}P'''$.

If $h = \frac{1}{N_p}$, $P = \left(1-\frac{1}{2N_p}\right)P' + \frac{1}{4N_p}P''$.

If $h = 1$, $(P-\frac{1}{2}P'-\frac{1}{4}P'') = \left(1-\frac{1}{N_p}\right)(P'-\frac{1}{2}P''-\frac{1}{4}P''')$, $P = \frac{1}{2}P'+\frac{1}{4}P''$.

It makes no appreciable difference whether brother-sister mating occurs at random or is excluded, unless N_p is very small. The rate in the former case is exactly the same as that for the same number ($N = 2N_p$) of monoecious individuals with self-fertilization excluded.

The limiting case, $h = 1$, is that of brother-sister mating with $P = \frac{1}{2}P' + \frac{1}{4}P''$, as given by Jennings (1914) (cf. Wright, 1921):

$$P \doteq \tfrac{1}{4}(1+\sqrt{5})P' = 0{\cdot}809P' \quad \text{(Wright, 1931).}$$

The simplest system, compatible with $h = 0$, is that of double-first cousin mating ($N_p = 2$)

$$P = \tfrac{1}{2}P' + \tfrac{1}{4}P'' + \tfrac{1}{8}P'''$$ (Wright, 1921).

On putting $P/P' = P'/P'' = P''/P'''$ to find the limiting ratio in successive generations we arrive at the equation $8X^3 - 4X^2 - 2X - 1 = 0$:

$$P = 0{\cdot}920P'$$ (Wright, 1933*a*).

The matrix method yields an equation of the 12th degree from which Fisher (1949) factors out the above cubic and thus obtains a solution ($X = 0{\cdot}91964$) in agreement with that yielded by path coefficients.

The series of results for small populations in which consanguineous mating is avoided as far as possible within each generation, make an interesting comparison (Wright, 1921):

N	P (exact)	Approximate P
1	$\tfrac{1}{2}P'$	$0{\cdot}500P'$
2	$\tfrac{1}{2}P' + \tfrac{1}{4}P''$	$0{\cdot}809P'$
4	$\tfrac{1}{2}P' + \tfrac{1}{4}P'' + \tfrac{1}{8}P'''$	$0{\cdot}920P'$
8	$\tfrac{1}{2}P' + \tfrac{1}{4}P'' + \tfrac{1}{8}P''' + \tfrac{1}{16}P^{IV}$	$0{\cdot}965P'$
16	$\tfrac{1}{2}P' + \tfrac{1}{4}P'' + \tfrac{1}{8}P''' + \tfrac{1}{16}P^{IV} + \tfrac{1}{32}P^{V}$	$0{\cdot}983P'$
N		$\left(1 - \frac{1}{4N}\right)P'$

With random mating in a population of effective size N, heterozygosis falls off by approximately $1/(2N)$ of its previous value per generation (exactly in a monoecious population). With maximum avoidance of inbreeding within each generation this rate of falling off is approximately halved.

Another important case is that of random mating in a population of N_m males, N_f females (Wright, 1931):

$$P = P' - \left(\frac{N_m + N_f}{8N_mN_f}\right)(2P' - P''),$$

$$\frac{P - P'}{P'} = -\frac{1}{2}\left(1 + \frac{N_m + N_f}{4N_mN_f}\right) + \frac{1}{2}\sqrt{\left\{1 + \left(\frac{N_m + N_f}{4N_mN_f}\right)^2\right\}}$$

$$= -\left(\frac{1}{8N_m} + \frac{1}{8N_f}\right)\left(1 - \frac{1}{8N_m} - \frac{1}{8N_f}\right) \quad \text{approximately.}$$

In the preceding cases, it has been assumed that generations are wholly distinct. Cases of overlapping generations can however be worked out. The simplest is that of alternate parent-offspring mating which was shown by Jennings (1916) to give exactly the same result as brother-sister mating. The method of path coefficients also gives the same result (Wright, 1921) as does the matrix method (Fisher, 1949).

In the case of sex-linked inheritance, N_m males N_f females (Wright, 1933*a*):

$$P = P' - \left(\frac{N_f + 1}{8N_f}\right)(2P' - P'') + \frac{(N_f - 1)(N_m - 1)}{8N_mN_f}(2P'' - P''').$$

Putting $P/P' = P'/P'' = P''/P'''$ and $y = \dfrac{(P - P')}{P'}$,

$$y^3 + y^2(2 + 2C_1) + y(1 + 3C_1 - 2C_2) + (C_1 - C_2) = 0,$$

where $C_1 = (N_f + 1)/8N_f$ and $C_2 = (N_f - 1)(N_m - 1)/8N_mN_f$,

$$y = -\left(\frac{2N_m + N_f}{9N_mN_f}\right), \quad \text{approximately.}$$

Following are examples of the effects of inbreeding on $2k$-somic loci (Wright, 1938). These can easily be extended. The chance that a gene may be represented twice in a gamete is ignored here for simplicity, though easily introduced if desired.

N monoecious individuals, random union of gametes:

$$P = \frac{1}{2N(2k-1)}[(6Nk-4N-2k+1)P' - (2N-2)(k-1)P''].$$

The simplest special case is that in which $N = 1$ (exclusive self-fertilization):

$$P = \left(\frac{4k-3}{4k-2}\right)P',$$

Diploid ($k = 1$) $\quad P = \frac{1}{2}P'$,

Tetraploid ($k = 2$) $\quad P = \frac{5}{6}P'$,

Hexaploid ($k = 3$) $\quad P = \frac{9}{10}P'$,

Octoploid ($k = 4$) $\quad P = \frac{13}{14}P'$.

The results for tetraploids and hexaploids agree with those arrived at by Haldane (1930) by the matrix method.

If self-fertilization is excluded we obtained

$$P = \frac{1}{2N(2k-1)}[(6Nk-4N-4k+2)P' - (2Nk-2N-4k+3)P''].$$

The simplest special case is that of sibling mating ($N = 2$):

$$P = P' - \frac{1}{8k-4}(2P'-P''),$$

$$P = \left(\frac{4k-3+\sqrt{(16k^2-16k+5)}}{8k-4}\right)P' \quad \text{approximately.}$$

	P (exact)	Approximate P
Diploid	$\frac{1}{2}P' + \frac{1}{4}P''$	$\frac{1}{4}(1+\sqrt{5})\,P' = 0{\cdot}80902P'$
Tetraploid	$(5/6)\,P' + (1/12)\,P''$	$(1/12)\,(5+\sqrt{(37)})\,P' = 0{\cdot}92356P'$
Hexaploid	$(9/10)\,P' + (1/20)\,P''$	$(1/20)\,(9+\sqrt{(101)})\,P' = 0{\cdot}95249P'$
Octoploid	$(13/14)\,P' + (1/28)\,P''$	$(1/28)\,(13+\sqrt{(197)})\,P' = 0{\cdot}96556P'$

In the case of tetraploids the matrix yields an octic equation from which Bartlett & Haldane (1934) had given 0·92356 as the pertinent root. Fisher (1949) obtains the same equation and factors out a quadratic which yields $(1/12)\,(5+\sqrt{(37)})$ as the pertinent root in agreement with these results.

We will not review here the somewhat more complicated cases in which the interference of inbreeding with the recombination of the linked genes has been studied (Wright, 1933*b*). General formulae for populations of N_m males, N_f females, and for the effects of restrictions on mating were considered. The results in population of 1 and 2 agreed with those from the matrix method, (Robbins, 1918*a*, *b*; Haldane & Waddington, 1931), with a few qualifications. In this case an additional correlational term must be recognized between the constitution of a gamete with respect to one locus and that with respect to the other, since the same sampling process is involved.

Bartlett & Haldane (1935) have dealt with the effects of enforced heterozygosis in a locus linked with one under consideration in populations of 1, 2 or 3 by the matrix method. These can readily be generalized for larger populations by the method of path coefficients.

APPENDIX F

Isolation by distance

Continuity over an area. Assume that there is uniform density over a large area but highly restricted dispersal in each generation, equal for male and female gametes. The variances of the distances between birthplaces of ancestors and individuals should be proportional to the number of intervening generations. Let F_{XK} be the correlation between random gametes drawn at random from a population of size XN relative to those drawn at random from a population of size KN. By the method of path coefficients

$$F_{1K} = \frac{1}{N}b^2 + 4\left(1-\frac{1}{N}\right)(ba')^2 F'_{2K} = \frac{1}{N}\left(\frac{1+F'_{1K}}{2}\right) + \left(1-\frac{1}{N}\right)F'_{2K},$$

$$F'_{2K} = \frac{1}{2N}\left(\frac{1+F''_{1K}}{2}\right) + \left(1-\frac{1}{2N}\right)F''_{3K},$$

$$F''_{3K} = \frac{1}{3N}\left(\frac{1+F'''_{1K}}{2}\right) + \left(1-\frac{1}{3N}\right)F'''_{4K}, \quad \text{etc.}$$

If a stationary state has been reached with respect to the F's, primes may be dropped. Solving for F_{1K}

$$F_{1K} = \sum_{X=1}^{K-1} t_X \Big/ \left[2N - \sum_{X=1}^{K-1} t_X\right],$$

where

$$\sum_{X=1}^{K-1} t_X = \left[1 + \frac{1}{2}\left(1-\frac{1}{N}\right) + \frac{1}{3}\left(1-\frac{1}{N}\right)\left(1-\frac{1}{2N}\right)\cdots\frac{1}{K}\prod_{X=1}^{K-1}\left(1-\frac{1}{XN}\right)\right],$$

and

$$t_X = \left[\frac{(X-1)N-1}{XN}\right]t_{X-1}.$$

The inbreeding coefficient F_{1K} should approach 1 as K is increased without limit and Σt should therefore approach N. An approximate formula for Σt, given in the 1943*a* paper, was shown to do this with high precision in all cases worked out, but no general proof was given. I am indebted to Mr Alan Robertson and to Mr D. J. Hooton for two different demonstrations which lead in different cases to great simplification.

Mr Hooton's demonstration rests on the following development of the recurrence equation for t_X:

$$KNt_K = (K-1)Nt_{K-1} - t_{K-1},$$

$$(K-1)Nt_{K-1} = (K-2)Nt_{K-2} - t_{K-2},$$

.....

$$3Nt_3 = 2Nt_2 - t_2,$$

$$2Nt_2 = N-1.$$

Adding

$$\sum_{X=1}^{K-1} t_X = N(1-Kt_K).$$

But

$$Kt_K = \prod_{X=1}^{K-1}\left(1-\frac{1}{XN}\right) < \frac{1}{1+\frac{1}{N}\sum_{X=1}^{K-1}\frac{1}{X}}.$$

As $\Sigma\frac{1}{X}$ is divergent, $Kt_K \to 0$ and $\Sigma t \to N$.

This also follows from the expansion of $\log Kt_K$ which gives a simpler method of calculating the inbreeding coefficients than that used previously. The results agree:

$$\begin{aligned}\log Kt_K &= \log \prod_{X=1}^{K-1}\left(1-\frac{1}{XN}\right)\\ &= \sum_{X=1}^{K-1}\log\left(1-\frac{1}{XN}\right)\\ &= -\left[\frac{1}{N}\Sigma\frac{1}{X}+\frac{1}{2N^2}\Sigma\frac{1}{X^2}+\frac{1}{3N^3}\Sigma\frac{1}{X^3}+\ldots\right]\\ &= -\left[\frac{1}{N}[\log(K-0{\cdot}5)+0{\cdot}5772]+\frac{1}{2N^2}\left(1{\cdot}6449-\frac{2}{2K-1}\right)\right.\\ &\qquad\left.+\frac{1}{3N^3}\left(1{\cdot}202-\frac{2}{(2K-1)^2}\right)+\frac{1}{4N^4}\left(1{\cdot}082-\frac{2}{(2K-1)^3}\right)\ldots\right].\end{aligned}$$

It is convenient to express this in terms of common logarithms for purposes of calculation:

$$\begin{aligned}-\log_{10}(Kt_K) &= \frac{1}{N}[\log_{10}(K-0{\cdot}5)+0{\cdot}2507]+\frac{1}{N^2}\left[0{\cdot}3572-\frac{0{\cdot}4343}{2K-1}\right]\\ &\qquad+\frac{1}{N^3}\left[0{\cdot}174-\frac{0{\cdot}29}{(2K-1)^2}\right]+\frac{0{\cdot}12}{N^4}+\ldots.\end{aligned}$$

The inbreeding coefficient may now be expressed in terms of Kt_K instead of Σt_X

$$F_{1K}=\frac{\sum^{K-1} t_X}{2N-\sum^{K-1} t_X}=\frac{1-Kt_K}{1+Kt_K}.$$

In Figs. 7–11 F_{NX} is used for the inbreeding coefficient of neighbourhoods (N_N) relative to large populations, N_X, ranging up to $10^8\,N_N$ in size. $F_{NX}=F_{1X}$ under random union of gametes. It is of greater interest to find the coefficient for populations of any size N_X relative to a specified total N_T. This was previously found from the relation $F_{XT}=(F_{NT}-F_{NX})/(1-F_{NX})$ or its equivalent $P_{XT}=P_{NT}/P_{NX}$. It can also be found directly from the initial set of equations above. Returning to the symbolism used there and using S for a given value of X,

$$\begin{aligned}F_{SK} &= \left[\frac{1+F_{1K}}{2}\right]\left[\frac{1}{SN}+\frac{1}{(S+1)N}\left(1-\frac{1}{SN}\right)\cdots\frac{1}{K}\prod_{X=S}^{K-1}\left(1-\frac{1}{XN}\right)\right]\\ &= \left[\frac{1+F_{1K}}{2}\right]\left[\sum_{X=1}^{K-1}t_X-\sum_{X=1}^{S}t_X\right]\Big/\prod_{X=1}^{S-1}\left(1-\frac{1}{XN}\right)\\ &= \left(\frac{1}{1+Kt_K}\right)\left(\frac{St_S-Kt_K}{St_S}\right),\\ P_{SK} &= \frac{Kt_K(1+St_S)}{St_S(1+Kt_K)}.\end{aligned}$$

Since

$$P_{1K}=\frac{2Kt_K}{1+Kt_K}\quad\text{and}\quad P_{1S}=\frac{2St_S}{1+St_S},$$

$$P_{1K}=P_{1S}P_{SK}$$

as expected.

In Figs. 7–11 F_{XT} is used for the inbreeding coefficient of populations ranging from N_N to $10^8\,N_N$ relative to that of $10^8\,N_N$.

F_{XT} was interpreted in the papers 1943*a* and 1946 as the ratio of the variance (σ_q^2) of gene frequencies at neutral loci of populations of size N_X within areas of size N_T, to the limiting variance $q_T(1-q_T)$ under complete local fixation. As noted in the text, an interpretation in terms of the

correlation between gene frequencies of neighbourhoods, the statistic discussed by Malécot (1948) can be given.

Let q_{N1} and q_{N2} be the gene frequencies of neighbourhoods drawn at random from the range of ancestral generations X. Their average distance apart is of the order of $\sqrt{(N_X/N_N)}\,\sigma$, where σ is the standard deviation of dispersion in one direction. Let $\sigma^2_{q(NX)}$ be the variance of the gene frequencies of such neighbourhoods.

The variances of differences between pairs is $2\sigma^2_{q(NX)}$, there being no correlation. If, however, such differences (within populations of size N_X) are averaged over a more comprehensive population of size N_T and this average is expressed in terms of the variance of neighbourhoods within this total, the correlation between the pairs, $r_{N_1N_2(XT)}$, must be taken into account:

$$2\sigma^2_{q(NX)} = 2\sigma^2_{q(NT)}(1-r_{N_1N_2(XT)}),$$

$$r_{N_1N_2(XT)} = 1-\frac{\sigma^2_{q(NX)}}{\sigma^2_{q(NT)}} = \frac{\sigma^2_{q(XT)}}{\sigma^2_{q(NT)}} = \frac{q_T(1-q_T)\,F_{XT}}{q_T(1-q_T)\,F_{NT}}$$

$$= \frac{F_{XT}}{F_{NT}}.$$

Continuity along a single line. In the case of continuity along a single line, the ancestors of generation X should be drawn from populations of $\sqrt{X\,N_N}$ instead of XN:

$$F_{1K} = \sum_1^{K-1} t_X\left[2N - \sum_1^{K-1} t_X\right],$$

$$\sum_1^{K-1} t_X = \left[1+\frac{1}{\sqrt{2}}\left(1-\frac{1}{N}\right)+\frac{1}{\sqrt{3}}\left(1-\frac{1}{N}\right)\left(1-\frac{1}{\sqrt{2N}}\right)\cdots\frac{1}{\sqrt{K}}\prod_{X=1}^{K-1}\left(1-\frac{1}{\sqrt{XN}}\right)\right],$$

$$t_X = \frac{N\sqrt{(X-1)}-1}{N\sqrt{X}}\,t_{X-1} \quad \text{(Wright, 1943}a\text{)}.$$

Mr Hooton notes that it can be shown in the same way as in the case of area continuity that

$$\sum_{X=1}^{K-1} t_X = N(1-\sqrt{K\,t_K}),$$

and that $\sqrt{K\,t_K}\to 0$, $\sum_{X=1}^{K-1} t_X \to N$ as K increases

$$\log(\sqrt{K\,t_K}) = \sum_{X=1}^{K-1}\log\left(1-\frac{1}{\sqrt{X\,N}}\right)$$

$$= \frac{1}{N}\Sigma\frac{1}{\sqrt{X}}+\frac{1}{2N^2}\Sigma\frac{1}{X}+\frac{1}{3N^3}\Sigma\frac{1}{X^{\frac{3}{2}}}+\frac{1}{4N^4}\Sigma\frac{1}{X^2}+\ldots.$$

Diverse systems of mating. The effects of various systems of mating where there is continuity over an area have been considered (Wright, 1946). The inbreeding coefficients for individuals, F_{1K} and for neighbourhoods, F_{NT}, are given below in terms of Kt_K instead of $\sum_1^{K-1} t_X$, the form used in the above paper.

Case 1. Monoecious individuals with the proportion h of self-fertilization. Equal dispersion of male and female gametes:

	General	$h=1/N$	$h=1$
$F_{1K}=$	$\dfrac{(N-1)-N(1-h)Kt_K}{(N-1)+N(1-h)Kt_K}$	$\dfrac{1-Kt_K}{1+Kt_K}$	1
$F_{NK}=$	$\dfrac{(N-1)(1-Kt_K)}{(N-1)+N(1-h)Kt_K}$	$\dfrac{1-Kt_K}{1+Kt_K}$	$1-Kt$

Case 2. Separate sexes with equal dispersion. N_p permanent pairs. Proportion h of brother-sister mating. Kt_K based on N_p.

	General	$h = 1/N$	$h = 1$
$F_{1K} =$	$\dfrac{(N_p-1)-N_p(1-h)Kt_K}{(N_p-1)+3N_p(1-h)Kt_K}$	$\dfrac{1-Kt_K}{1+3Kt_K}$	1
$F_{NK} =$	$\dfrac{(N_p-1)(1-Kt_K)}{(N_p-1)+3N_p(1-h)Kt_K}$	$\dfrac{1-Kt_K}{1+3Kt_K}$	$1-Kt_K$

Case 3. N_m males, N_f females. Equal dispersion and random mating. Kt_K based on

$$N' = 2N_mN_f/(N_m+N_f).$$

A correction is required in this case. In the formulae on pp. 47 and 48 of the 1946 paper, N_m and N_f should be multiplied by 2 to give the numbers in the parental generation, and in Tables 1 and 2 the figures for case 3 apply to $N_m = N_f = 50$, and similarly in Tables 3 and 4 they apply to $N_m = N_f = 5$.

$$F_{1K} = N_{NK} = \frac{1-Kt_K}{1+7Kt_K} \quad \text{approximately.}$$

Case 4. No dispersion of ovules, dispersion of pollen in normal distribution relative to distance in one direction from monoecious individuals. Size of neighbourhood N, apart from self-fertilization in excess of random. Proportion r of pollinations at random and thus $(1-r)$ excess self-fertilization and $h = (r/N)+(1-r)$ the total proportion of self-fertilization:

$$F_{NK} = \frac{\sum_{X=1}^{K-1} t_X}{N+r\left(N-\sum_{X=1}^{K-1} t_X\right)},$$

where

$$\sum_{X=1}^{K-1} t_X = \left[1+\left(\frac{1}{1+r}\right)\left(1-\frac{1}{N}\right)+\left(\frac{1}{1+2r}\right)\left(1-\frac{1}{N}\right)\left(1-\frac{1}{(1+r)N}\right)+\dots \right.$$
$$\left. +\left(\frac{1}{1+(k-1)r}\right)\prod_{X=1}^{K-1}\left(1-\frac{1}{[1+(X-1)r]N}\right)\right.,$$

if $r = 1$ (self-fertilization random)

$$F_{NK} = \frac{1-Kt_K}{1+Kt_K}.$$

The result in this case is the same as with equal dispersion of ovules and pollens, provided that the variance of pollen is twice that in the latter. With excess self-fertilization, F_{NK} rises disproportionately as shown in Fig. 9.

If there is a certain amount of universal dispersion (m) or of reversible mutation at rates u and v per generation at the locus under consideration, the total amount of displacement by a random sample from the species is $(m+u+v)$. For simplicity we will use merely m. The formulae for Σt_X is as follows (Wright, 1943*a*):

$$\sum_{X=1}^{K-1} t_X = (1-m)^2+\frac{(1-m)^4}{2}\left(1-\frac{1}{N}\right)+\frac{(1-m)^6}{3}\left(1-\frac{1}{N}\right)\left(1-\frac{1}{2N}\right)+\dots\frac{(1-m)^{2K}}{K}\prod_{X=1}^{K-1}\left(1-\frac{1}{XN}\right).$$

The factor $(1-m)^{2X}$ is approximately $(1-2mX)$ if $2mX$ is small. The effect on $\sum_{X=1}^{K-1} t_X$ and hence on F_{NX} is negligible for most purposes if K is less than $1/(10m)$. If, on the other hand, X is greater than $3/m$, $(1-m)^{2X}$ becomes less than 0·0025 and contributions to $\sum^{K-1} t_X$ become negligible for most purposes. Thus the curve representing the value of F_{NX} in relation to log (N_X/N_N) follows that with $m = 0$ up to about $N_X/N_N = 1/(10m)$ and then rapidly approaches an asymptote which is less

than 1. This asymptote was estimated by a rather cumbersome method, in the 1943*a* paper. Mr Robertson has arrived at a very simple formula for finding its value in the case of area continuity. This depends on writing $\frac{1}{N}\sum_1^\infty t$ as follows:

$$\frac{1}{N}\sum_1^\infty t = \left[(1-m)^2\frac{1}{N} - \frac{(1-m)^4}{2!}\frac{1}{N}\left(\frac{1}{N}-1\right) + \frac{(1-m)^6}{3!}\frac{1}{N}\left(\frac{1}{N}-1\right)\left(\frac{1}{N}-2\right)\cdots \right.$$
$$\left. = 1-[1-(1-m)^2]^{1/N}\right].$$

The numerical results in the 1943*a* paper are all in agreement.

Uniform density and amount of dispersion is a limiting case, not likely to be realized in nature. A rough model for a more typical situation may be obtained by considering individuals whose parents are drawn from a relatively large population (MN) and whose more remote ancestors are drawn from populations that increase linearly with the number of generations but by increments (N) that are only a fraction of that of the parental population:

$$F_{1K} = \frac{1}{MN}\left(\frac{1+F_{1K}}{2}\right) + \left(1-\frac{1}{MN}\right)F_{2K},$$

$$F_{2K} = \frac{1}{(M+1)N}\left(\frac{1+F_{2K}}{2}\right) + \left(1-\frac{1}{(M+1)N}\right)F_{3K}, \quad \text{etc.},$$

$$F_{1K} = \left(\frac{1+F_{1K}}{2N}\right)\left[\frac{1}{M} + \left(\frac{1}{M+1}\right)\left(1-\frac{1}{MN}\right) + \left(\frac{1}{M+2}\right)\left(1-\frac{1}{MN}\right)\left(1-\frac{1}{(M+1)N}\right)\cdots \right.$$
$$\left. + \left(\frac{1}{M+K-1}\right)\prod_M^{M+K-2}\left(1-\frac{1}{XN}\right)\right],$$

$$F_{1K} = \frac{\Sigma t'}{2N-\Sigma t'},$$

$$\Sigma t' = \left[\sum_1^{M+K-2} t - \sum_1^{M-1} t\right] \Big/ \prod_{X=1}^{M-1}\left(1-\frac{1}{XN}\right)$$
$$= N[Mt_M - (M+K-1)t_{(M+K-1)}]/Mt_M,$$

$$F_{1K} = \frac{Mt_M - (M+K-1)t_{(M+K-1)}}{Mt_M + (M+K-1)t_{(M+K-1)}}.$$

REFERENCES

Bartlett, M. S. & Haldane, J. B. S. (1934). The theory of inbreeding in autotetraploids. *J. Genet.* **29**, 175–80.

Bartlett, M. S. & Haldane, J. B. S. (1935). The theory of inbreeding with forced heterozygosis. *J. Genet.* **31**, 327–40.

Bateman, A. J. (1947). Contamination in seed crops. III. Relation with isolation distance. *Heredity*, **1**, 303–36.

Bernstein, F. (1930). Fortgesetzte Untersuchungen aus der Theorie der Blutgruppen. *Z. indukt. Abstamm.- u. VererbLehre*, **56**, 233–73.

Darwin, C. (1868). *The Variation of Animals and Plants under Domestication.* London.

Dobzhansky, Th. & Wright, S. (1943). Genetics of natural populations. X. Dispersion rates in *Drosophila pseudoobscura*. *Genetics*, **28**, 304–40.

Dobzhansky, Th. & Wright, S. (1947). Genetics of natural populations. XV. Rate of diffusion of a mutant gene through a population of *Drosophila pseudoobscura*. *Genetics*, **32**, 303–24.

Epling, C. & Dobzhansky, Th. (1942). Genetics of natural populations. VI. Microgeographical races in *Linanthus Parryae*. *Genetics*, **27**, 317–32.

FISH, H. D. (1914). On the progressive increase of homozygosis in brother-sister matings. *Amer. Nat.* **48**, 759–61.
FISHER, R. A. (1949). *The Theory of Inbreeding*, pp. 120. Edinburgh: Oliver and Boyd.
GALTON, F. (1889). *Natural Inheritance*. London: Macmillan & Co.
GOLDSCHMIDT, R. (1940). *The Material Basis of Evolution*, pp. 436. New Haven: Yale University Press.
HALDANE, J. B. S. (1930). Theoretical genetics of autopolyploids. *J. Genet.* **22**, 359–72.
HALDANE, J. B. S. (1936). The amount of heterozygosis to be expected in an approximately pure line. *J. Genet.* **32**, 375–91.
HALDANE, J. B. S. (1937). Some theoretical results of continued brother-sister mating. *J. Genet.* **34**, 265–74.
HALDANE, J. B. S. (1949). The association of characters as a result of inbreeding and linkage. *Ann. Eugen., Lond.*, **15**, 15–23.
HALDANE, J. B. S. & WADDINGTON, C. H. (1931). Inbreeding and linkage. *Genetics*, 357–74.
HARDY, G. H. (1908). Mendelian proportions in a mixed population. *Science*, **28**, 49–50.
JENNINGS, H. S. (1914). Formulae for the results of inbreeding. *Amer. Nat.* **48**, 693–6.
JENNINGS, H. S. (1916). The numerical results of diverse systems of breeding. *Genetics*, **1**, 53–89.
LUSH, J. L. (1943). *Animal Breeding Plans*, pp. 437. Ames, Iowa: Iowa State College Press.
MCPHEE, H. C. & WRIGHT, S. (1925). Mendelian analysis of the pure breeds of livestock. III. The Shorthorns. *J. Hered.* **16**, 205–15.
MCPHEE, H. C. & WRIGHT, S. (1926). Mendelian analysis of the pure breeds of livestock. IV. The British Dairy Shorthorns. *J. Hered.* **17**, 397–401.
MALÉCOT, G. (1948). *Les mathématiques de l'hérédité*, pp. 63. Paris: Masson et Cie.
PEARL, R. (1917). Studies on inbreeding. *Amer. Nat.* **51**, 545–59, 636–9.
ROBBINS, R. B. (1918*a*). Applications of mathematics to breeding problems. II. *Genetics*, **3**, 73–92.
ROBBINS, R. B. (1918*b*). Some applications of mathematics to breeding problems. III. *Genetics*, **3**, 375–89.
SHULL, G. H. (1908). The composition of a field of maize. *Rep. Amer. Breed. Ass.* **4**, 296–301.
WEINBERG, W. (1908). Über den Nachweis der Vererbung beim Menschen. *Jahreshefte d. Ver. f. vaterländische Naturkunde in Württemberg*, **64**, 368–82.
WRIGHT, S. (1921). Systems of mating. *Genetics*, **6**, 111–78.
WRIGHT, S. (1922*a*). Coefficients of inbreeding and relationship. *Amer. Nat.* **56**, 330–8.
WRIGHT, S. (1922*b*). Effects of inbreeding and crossbreeding on guinea pigs. *Bull. U.S. Dep. Agric.* no. 1121, pp. 59.
WRIGHT, S. (1923*a*). Mendelian analysis of the pure breeds of livestock. I. The measurement of inbreeding and relationship. *J. Hered.* **14**, 339–48.
WRIGHT, S. (1923*b*). Mendelian analysis of the pure breeds of livestock. II. The Duchess family of Shorthorns as bred by Thomas Bates. *J. Hered.* **14**, 405–22.
WRIGHT, S. (1931). Evolution in Mendelian population. *Genetics*, **16**, 97–159.
WRIGHT, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. 26th Int. Congr. Genet.*, pp. 356–66.
WRIGHT, S. (1933*a*). Inbreeding and homozygosis. *Proc. Nat. Acad. Sci., Wash.*, **19**, 411–20.
WRIGHT, S. (1933*b*). Inbreeding and recombination. *Proc. Nat. Acad. Sci., Wash.*, **19**, 420–33.
WRIGHT, S. (1935*a*). The analysis of variance and the correlations between relatives with respect to deviations from an optimum. *J. Genet.* **30**, 253–6.
WRIGHT, S. (1935*b*). Evolution in populations in approximate equilibrium. *J. Genet.* **30**, 257–66.
WRIGHT, S. (1938). The distribution of gene frequencies in populations of polyploids. *Proc. Nat. Acad. Sci., Wash.*, **24**, 372–7.
WRIGHT, S. (1939). *Statistical genetics in relation to evolution. Exposés de Biométrie et de statistique biologique*, no. 802, pp. 63. Hermann et Cie, Editeurs.
WRIGHT, S. (1940). Breeding structure of populations in relation to speciation. *Amer. Nat.* **74**, 232–48.
WRIGHT, S. (1941). *The Material Basis of Evolution*, by R. Goldschmidt (Review). *Sci. Mon.* **53**, 165–70.
WRIGHT, S. (1943*a*). Isolation by distance. *Genetics*, **28**, 114–38.
WRIGHT, S. (1943*b*). An analysis of local variability of flower color in *Linanthus Parryae*. *Genetics*, **28**, 139–56.
WRIGHT, S. (1946). Isolation by distance under diverse systems of mating. *Genetics*, **31**, 39–59.
WRIGHT, S. (1948*a*). On the roles of directed and random changes in gene frequency in the genetics of populations. *Evolution*, **2**, 279–94.
WRIGHT, S. (1948*b*). Evolution, organic. *Encycl. Brit.* **8**, 915–29.
WRIGHT, S. (1948*c*). Genetics of populations. *Encycl. Brit.* **10**, 111, 111A–D, 112.
WRIGHT, S. (1949*a*). *Adaptation and Selection*. Chapter 20. Genetics, Paleontology and Evolution, pp. 365–89. Edited by G. L. Jepson, G. G. Simpson and E. Mayr.
WRIGHT, S. (1949*b*). Population structure in evolution. *Proc. Amer. Phil. Soc.* **93**, 471–8.
WRIGHT, S. & MCPHEE, H. C. (1925). An approximate method of calculating coefficients of inbreeding and relationship from livestock pedigrees. *J. Agric. Res.* **31**, 377–83.