THE THEORETICAL VARIANCE WITHIN AND AMONG SUBDIVISIONS OF A POPULATION THAT IS IN A STEADY STATE*[1]

SEWALL WRIGHT
Department of Zoology, The University of Chicago, Chicago, Illinois

Received October 26, 1951

IN a population that is subdivided into more or less isolated strains, the total genetic variance (σ_T^2) of a character that depends on multiple alleles at multiple loci with additive effects within and among the loci, can easily be analyzed into the variance of strain means (σ_m^2) and the average variance within strains (σ_w^2) in terms of the inbreeding coefficient F, and the variance, $\sigma_0^2 = 2\bar{q}(1-\bar{q})$, expected under panmixia with the same mean gene frequency, $\bar{q}$ (WRIGHT 1951).

$$\sigma_w^2 = (1 - F)\,\sigma_0^2 \tag{1}$$

$$\sigma_m^2 = 2F\,\sigma_0^2 \tag{2}$$

$$\sigma_T^2 = (1 + F)\,\sigma_0^2 \tag{3}$$

These simple relations hold whether the strains are completely isolated and drifting toward fixation in the absence of mutation or selection, or whether a steady state has been reached in which the tendency toward fixation is balanced by a certain amount of cross breeding, mutation or selection (acting alike on both sexes).

It should be noted that if the coefficient F is used for the purpose for which it was originally introduced, the description of population structure, it cannot take cognizance of rates of mutation or selection since these are specific for each locus. In this sense, F is related to heterozygosis, variability, correlation between relatives, etc., in only those respects in which the effects of recurrent mutation and selection are negligible. It is also desirable, however, to use F statistics that relate to specific loci and these must, of course, take account of the effects of all factors. The sort of use should be clear from the context. In either case, F can be defined as the proportional approach toward homozygosis from the situation under panmixia at the same gene frequency. The present discussion will be restricted to coefficients pertaining to disomic loci, and as random mating will always be assumed within strains, the only F coefficients considered are those of individuals relative to the total population.

* Part of the cost of the accompanying mathematical formulae has been paid by the GALTON and MENDEL MEMORIAL FUND.

[1] This investigation was aided by a grant from the Wallace C. and Clara A. Abbott Memorial Fund of The University of Chicago.

If there is dominance in any degree, it is still possible to express the total variance in terms of a function of F and gene frequency, or statistics derived from these. Letting m_T and σ_T^2 be the grand average and variance of the total population, m_0 and σ_0^2 those of a panmictic population with the same allelic frequencies and m_1 and σ_1^2 those of a random array of completely fixed strains (WRIGHT 1951)

$$m_T = (1-F)\, m_0 + F\, m_1 \tag{4}$$

$$\sigma_T^2 = (1-F)\, \sigma_0^2 + F\, \sigma_1^2 + F(1-F)\, (m_0 - m_1)^2 \tag{5}$$

The analysis of σ_T^2 into inter- and intra-strain components cannot, however, be made in these terms, if there is any degree of dominance. There are diverse possibilities with the same gene frequency and the same value of F, depending on the nature of the distribution of gene frequencies among the strains.

ALAN ROBERTSON (1952) has recently carried through the analysis in the important case of completely isolated strains of given size, tending toward fixation without interference from mutation or selection. The distribution of gene frequencies here takes a succession of forms depending wholly on the successive inbreeding coefficients, the theoretical values of which can easily be determined. The mathematical formulae for the distribution of gene frequencies are not indeed known, but as ROBERTSON shows, it is only necessary to find the law of change of the first four moments. This he has done by means of matrix algebra.

ROBERTSON shows that there is an almost qualitative difference from the results where dominance is lacking. Thus there is a considerable *reduction* in the total variance as inbreeding increases in case the gene frequency of the recessive allele is sufficiently high and a very considerable *increase* in the variance within strains with increasing F, up to a certain point, in the case in which the gene frequency of the recessive is sufficiently low.

SUBDIVIDED POPULATIONS IN A STEADY STATE

It may be of interest to make a comparison with the results from a very different situation, that in which a steady state has been reached in a population divided into partially isolated strains, the tendency toward fixation being balanced by occasional cross breeding. This case is of primary importance in the theory that evolution consists ordinarily of second order shifts in such states of balance. The analysis is simpler since the form of the distribution of gene frequencies is known in this case.

Gene frequency is represented by q with distribution, $\phi(q)$ among strains, characterized by mean $\bar{q}$, variance σ_q^2 and higher moments. The strains are assumed to be alike in effective size (N) and other conditions. The distribution $\phi(q)$ is properly discontinuous, with values at steps of $1/(2N)$ in q. The integrals below may be considered as Stieltjes integrals applicable to step functions as well as continuous ones. The results are theoretically exact if $\phi(q)$ is an exact discontinuous distribution but are merely close approximations if continuous. Practically, of course, the irregularities in size and in other

respects in actual cases make both merely approximate models. The following four equations are merely definitions, that for F being here based on proportional approach to homozygosis.

$$\int_0^1 \phi(q)dq = 1 \tag{6}$$

$$\int_0^1 q\phi(q)dq = \bar{q} \tag{7}$$

$$\int_0^1 (q-\bar{q})^2\phi(q)dq = \sigma_q^2 \tag{8}$$

$$2\int_0^1 q(1-q)\phi(q)dq = 2\bar{q}(1-\bar{q})(1-F) \tag{9}$$

From these

$$\int_0^1 q^2\phi(q)dq = \bar{q}^2 + \sigma_q^2 = \bar{q}^2 + \bar{q}(1-\bar{q})F \tag{10}$$

The mode of analysis may be illustrated most simply by the case of no dominance, with character values 0, a and $2a$ assigned to genotypes aa, Aa and AA respectively. For a strain with zygotic array $[(1-q)a+qA]^2$, the character mean, m, is $2qa$ and the variance of the character, σ_w^2, is $2q(1-q)a^2$. For the total population, we are led at once to the results already referred to (in application here, however, merely to a pair of alleles). It is sometimes convenient to use p for $(1-q)$ for brevity.

$$\bar{m} = \int_0^1 m\phi(q)dq = 2a\int_0^1 q\phi(q)dq = 2\bar{q}a \tag{11}$$

$$\overline{\sigma_w^2} = \int_0^1 \sigma_w^2\phi(q)dq = 2a^2\int_0^1 q(1-q)\phi(q)dq = 2\bar{p}\bar{q}(1-F)a^2 \tag{12}$$

$$\sigma_m^2 = \int_0^1 (m-\bar{m})^2\phi(q)dq = 4a^2\int_0^1 q^2\phi(q)dq - \bar{m}^2 = 4\bar{p}\bar{q}Fa^2 \tag{13}$$

$$\sigma_T^2 = \overline{\sigma_w^2} + \sigma_m^2 = 2\bar{p}\bar{q}(1+F)a^2 \tag{14}$$

Consider now the case of complete dominance, letting q be the frequency of the recessive allele in a strain and a the differential effect of aa. For the mean and variance in the phenotypic array $[(1-q^2)A-+q^2aa]$ in a strain we have $m = q^2a$, $\sigma_w^2 = q^2(1-q^2)a^2$. In the total population

$$\bar{m} = a\int_0^1 q^2\phi(q)dq = (\bar{q}^2 + \bar{p}\bar{q}F)a \tag{15}$$

$$\overline{\sigma_w^2} = a^2\left[\int_0^1 q^2\phi(q)dq - \int_0^1 q^4\phi(q)dq\right] \tag{16}$$

$$\sigma_m^2 = a^2\int_0^1 q^4\phi(q)dq - \bar{m}^2 \tag{17}$$

$$\sigma_T^2 = a^2\int_0^1 q^2\phi(q)dq - \bar{m}^2 = \left[\bar{q}^2 + \bar{p}\bar{q}F\right]\left[1 - (\bar{q}^2 + \bar{p}\bar{q}F)\right]a^2 \tag{18}$$

Thus the total variance can be evaluated in terms of $\bar{q}$, F and a irrespective of the form of $\phi(q)$. The value agrees with that obtained by substituting $m_0 = \bar{q}^2a$, $\sigma_0^2 = \bar{q}^2(1-\bar{q}^2)a^2$, $m_1 = \bar{q}a$ and $\sigma_1^2 = \bar{q}(1-\bar{q})a^2$ in the more general formula cited earlier. In this case it reduces to $\bar{m}(a-\bar{m})$. but this is not true in general.

Apportionment of this total variance into $\overline{\sigma_w^2}$ and σ_m^2 requires evaluation of $\int_0^1 q^4\phi(q)dq$. ROBERTSON, as noted, found the law of change of this quantity under the cumulative effect of accidents of sampling among strains of a given

size. To deal with a population in a steady state it is necessary to find the formula of $\phi(q)$ under the postulated conditions.

If a gene frequency is subject to systematic change at the rate Δq per generation, and to random fluctuations (δq) including the effects of inbreeding, the resultant probability distribution has the formula

$$\phi(q) = (C/\sigma_{\delta q}^2)e^{2\int(\Delta q/\sigma_{\delta q}^2)dq}, \qquad \int_0^1 \phi(q)dq = 1 \tag{19}$$

This can be derived as solution of the FOCKER-PLANCK equation of physics for the case of a steady state (cf. KOLMOGOROFF 1935; WRIGHT 1945; MALÉCOT 1948). The present author (1938) derived it independently from the condition that mean and variance remain unchanged. This is not a complete proof but, as noted several times since, this mode of demonstration can be extended to cover the conditions that *all* moments remain unchanged. This form of the demonstration is given in the appendix.

Letting m represent the effective amount of replacement of each strain by immigrants representative of the total, the systematic tendency toward change of gene frequency is $\Delta q = -m(q - \bar{q})$. The sampling variance of 2N gametes in the array $[(1-q)A + qa]^{2N}$ is $\sigma_{\delta q}^2 = q(1-q)/2N$. Substitution in the general equation for $\phi(q)$ leads to a Beta distribution, a formula originally derived for this case by still another method (WRIGHT 1931).

$$\phi(q) = \frac{\Gamma(a)}{\Gamma(a\bar{q})\,\Gamma[a(1-\bar{q})]}\, q^{a\bar{q}-1}(1-q)^{a(1-\bar{q})-1}, \qquad a = 4Nm \tag{20}$$

The moments about zero can be easily evaluated by use of the formulae

$$\int_0^1 x^{m-1}(1-x)^{n-1}dx = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}, \text{ and } \Gamma(x+1) = x\Gamma(x)$$

$$\int_0^1 \phi(q)dq = 1 \tag{21}$$

$$\mu_1' = \int_0^1 q\phi(q) = \bar{q} \tag{22}$$

$$\mu_2' = \int_0^1 q^2\phi(q) = \frac{\bar{q}(a\bar{q}+1)}{a+1} \quad (=(\bar{q}^2 + \bar{p}\bar{q}F) \text{ by } (10)) \tag{23}$$

$$\text{Thus } a = (1-F)/F \tag{24}$$

$$F = 1/(a+1) = 1/(4Nm+1) \tag{25}$$

$$\mu_3' = \int_0^1 q^3\phi(q)dq = \frac{\bar{q}(a\bar{q}+1)(a\bar{q}+2)}{(a+1)(a+2)} = (\bar{q}^2 + \bar{p}\bar{q}F)\left(\bar{q} + \frac{2F\bar{p}}{1+F}\right) \tag{26}$$

(27)
$$\mu_4' = \int_0^1 q^4 \phi(q)\,dq = \frac{\bar{q}(a\bar{q}+1)(a\bar{q}+2)(a\bar{q}+3)}{(a+1)(a+2)(a+3)}$$
$$= (\bar{q}^2 + \bar{p}\bar{q}F)\left(\bar{q} + \frac{2F\bar{p}}{1+F}\right)\left(\bar{q} + \frac{3F\bar{p}}{1+2F}\right)$$

From (16), (17), (23), and (27)

(28) $$\overline{\sigma_w^2} = (\bar{q}^2 + \bar{p}\bar{q}F)\left[1 - \left(\bar{q} + \frac{2F\bar{p}}{1+F}\right)\left(\bar{q} + \frac{3F\bar{p}}{1+2F}\right)\right]\alpha^2$$

(29) $$\sigma_m^2 = (\bar{q}^2 + \bar{p}\bar{q}F)\left[\left(\bar{q} + \frac{2F\bar{p}}{1+F}\right)\left(\bar{q} + \frac{3F\bar{p}}{1+2F}\right) - (\bar{q}^2 + \bar{p}\bar{q}F)\right]\alpha^2$$

(30) $$\sigma_T^2 = (\bar{q}^2 + \bar{p}\bar{q}F)[1 - (\bar{q}^2 + \bar{p}\bar{q}F)]\alpha^2 \quad (=(18))$$

Table 1 shows values of σ_T^2 for various values of $\bar{q}$ and F letting $\alpha = 1$. These apply to the case treated by ROBERTSON as well as to the steady state treated here, or to any other case of complete dominance of one of a pair of alleles. Table 2 shows the corresponding value of $\overline{\sigma_w^2}$, applicable only to a steady state in which the inbreeding effect is balanced by a linear systematic process (and dominance of one of the pair of alleles is complete). Table 3 shows the ratio of intra-strain variance under unimpeded inbreeding, as calculated by ROBERTSON's formula, to the values in table 2. ROBERTSON's formula is as follows in the terminology of this paper.

(31) $$\overline{\sigma_w^2} = \bar{p}\bar{q}\left[\binom{4}{5}(1-F) - (1-2\bar{q})(1-F)^3 + \left(\binom{1}{5} - \bar{p}\bar{q}\right)(1-F)^6\right]\alpha^2$$

Table 3 brings out the point that there is not very much difference in the results in the two cases. The values of $\overline{\sigma_w^2}$ differ by less than 2 percent for all values of q if F is as small as .10 and for values of q in a diagonal across the table from about q = .50 for small F to q = .10 for very large F. Intra-strain variance is, however, some 10 percent greater for small q, and F in the neighborhood of .50, in the case of progressive fixation, than in that of a steady state and the reverse is true for large q, and F in the neighborhood of .70. The reason is that for a given variance of gene frequencies and hence F, the distribution is more compact (platykurtic) where deviations from $\bar{q}$ tend to be reduced in proportion to their magnitude by crossbreeding (steady state) than where unimpeded. The variance of the character is maximum in strains in which q = .707. In a total population with low $\bar{q}$ and intermediate F, the proportion of the strains with gene frequencies that yield a high variance may be expected to be greater in the less compact distribution. The opposite situation holds where $\bar{q}$ is high and F rather high.

TABLE 1

The variance (σ_T^2) of a total population characterized by inbreeding coefficient F in the case of a dominant-recessive pair of alleles with unit difference in grade.

	F										
q	0	.10	.20	.30	.40	.50	.60	.70	.80	.90	1.00
.05	.0025	.0072	.0119	.0165	.0210	.0256	.0300	.0345	.0398	.0432	.0475
.10	.0099	.0186	.0272	.0356	.0439	.0520	.0599	.0677	.0753	.0827	.0900
.20	.0384	.0529	.0668	.0803	.0932	.1056	.1175	.1289	.1398	.1501	.1600
.30	.0819	.0957	.1146	.1296	.1437	.1570	.1693	.1808	.1914	.2012	.2100
.40	.1344	.1501	.1647	.1782	.1905	.2016	.2116	.2204	.2281	.2346	.2400
.50	.1875	.1994	.2100	.2194	.2275	.2344	.2400	.2444	.2475	.2494	.2500
.60	.2304	.2365	.2415	.2454	.2481	.2496	.2500	.2492	.2473	.2442	.2400
.70	.2499	.2499	.2490	.2472	.2445	.2410	.2365	.2312	.2250	.2180	.2100
.80	.2304	.2257	.2204	.2147	.2084	.2016	.1943	.1865	.1782	.1693	.1600
.90	.1539	.1482	.1424	.1364	.1303	.1240	.1176	.1109	.1041	.0971	.0900
.95	.0880	.0841	.0803	.0763	.0723	.0683	.0642	.0601	.0560	.0518	.0475
1.00	0	0	0	0	0	0	0	0	0	0	0

The decrease in total variance with increased F where the recessive allele is relatively abundant, referred to earlier, is shown in table 1. This holds for $q > (\sqrt{2(1-F)+F^2}-F)/2(1-F)$ and thus for $q > .707$ if F is close to 0, and for $q > .50$ if F is close to 1.

The increase in average intra-strain variance with increased F, where the recessive allele is relatively rare and F not too large (shown in table 2), depends largely on the increase in the mean with increase in F. There is no increase in mean in the absence of dominance, but a very pronounced one in the case of dominance and small $\bar{q}$. For very small $\bar{q}$, the intra-strain variance approaches $\bar{q}F(1-F)(1+4F)/(1+F)(1+2F)$ in the case of steady state,

TABLE 2

The average variance ($\overline{\sigma_w^2}$) within partially isolated subdivisions of a population in which a steady state (constant F) has been reached between the tendency toward fixation due to inbreeding and the opposed effect of occasional crossbreeding, for the same character as in table 1. The variance of the means of the subdivisions is the difference between corresponding entries in table 1 and 2.

	F (steady state)										
q	0	.10	.20	.30	.40	.50	.60	.70	.80	.90	1.00
.05	.0025	.0068	.0100	.0120	.0128	.0126	.0114	.0095	.0069	.0037	0
.10	.0099	.0174	.0226	.0254	.0262	.0252	.0225	.0185	.0133	.0071	0
.20	.0384	.0483	.0538	.0554	.0539	.0496	.0430	.0345	.0244	.0128	0
.30	.0819	.0885	.0898	.0869	.0806	.0717	.0605	.0475	.0329	.0170	0
.40	.1344	.1325	.1260	.1162	.1039	.0896	.0738	.0567	.0386	.0197	0
.50	.1875	.1734	.1571	.1395	.1208	.1016	.0818	.0617	.0413	.0208	0
.60	.2304	.2032	.1772	.1524	.1285	.1056	.0834	.0618	.0408	.0202	0
.70	.2499	.2122	.1793	.1502	.1238	.0997	.0773	.0563	.0366	.0179	0
.80	.2304	.1896	.1562	.1279	.1033	.0816	.0622	.0447	.0286	.0138	0
.90	.1539	.1234	.0994	.0797	.0632	.0492	.0369	.0261	.0165	.0079	0
.95	.0880	.0697	.0556	.0442	.0348	.0268	.0200	.0141	.0088	.0042	0
1.00	0	0	0	0	0	0	0	0	0	0	0

TABLE 3

The ratio of the average variance within completely isolated strains with unimpeded progress toward complete fixation, measured by successive values of F (ROBERTSON'S case) to the average variance within strains under conditions which yield the same values of F in steady states (of table 2).

	F									
$\bar{q}$	0	.10	.20	.30	.40	.50	.60	.70	.80	.90
.05	1.000	1.015	1.042	1.069	1.087	1.095	1.092	1.076	1.049	1.010
.10	1.000	1.012	1.034	1.056	1.072	1.079	1.076	1.063	1.039	1.004
.20	1.000	1.007	1.022	1.036	1.046	1.050	1.048	1.038	1.020	.994
.30	1.000	1.004	1.012	1.019	1.024	1.025	1.022	1.014	1.002	.984
.40	1.000	1.002	1.004	1.005	1.005	1.003	.999	.993	.984	.973
.50	1.000	1.000	.997	.993	.988	.983	.977	.972	.967	.963
.60	1.000	.998	.991	.983	.973	.965	.957	.953	.951	.954
.70	1.000	.996	.986	.973	.960	.948	.939	.935	.936	.944
.80	1.000	.994	.981	.964	.948	.933	.922	.918	.921	.935
.90	1.000	.993	.977	.957	.936	.919	.906	.902	.907	.926
.95	1.000	.992	.975	.953	.931	.912	.899	.894	.900	.921

with maximum at .46 which agrees with ROBERTSON's result for progressive inbreeding, although his formula $q[(4/5)(1-F)-(1-F)^3+(1/5)(1-F)^6]$ is quite different in appearance. The largest value of q at which there is an increase in $\overline{\sigma_w^2}$ with increase in F is .41 $(=\sqrt{1/6})$ in both cases, since in both, $\overline{\sigma_w^2}$ approaches $\bar{q}^2(1-\bar{q}^2)+F\overline{pq}(1-6\bar{q}^2)$ for very small F.

PROGRESSIVE INBREEDING: LIMITING CASE

While the form of the distribution of gene frequencies continually changes in the case considered by ROBERTSON, it approaches an almost rectangular distribution, $\phi(q) = 1$ between the limits 0 and 1, as F increases. The proportion of the strains that are heterallelic continually decreases at the rate $1/(2N)$ per generation as new strains drift into fixation at $q = 0$ or $q = 1$ (WRIGHT 1931). Let x and y be the proportional frequencies at $q = 0$ and $q = 1$, respectively, at a given time, leaving $1-x-y$ as the proportion still unfixed. The values of x and y must always be such that the mean gene frequency $\bar{q}$ is constant and the second moment about zero is related to F by the formula, $\mu_2' = q^2 + \overline{pq}F$

$$(32) \qquad \phi(q) = \begin{cases} x & \text{for } q = 0 \\ 1-x-y & \text{for } 0 < q < 1 \\ y & \text{for } q = 1 \end{cases}$$

The first and second moments about zero are as follows:

$$(33) \qquad \mu_1' = y + (1-x-y)\int_0^1 q\,dq = y + (1-x-y)/2 = \bar{q}$$

$$(34) \qquad \mu_2' = y + (1-x-y)\int_0^1 q^2\,dq = y + (1-x-y)/3 = \bar{q}^2 + \bar{p}\bar{q}F$$

From these

$$(35)\qquad \left\{\begin{aligned} x &= \bar{p} - 3\bar{p}\bar{q}(1-F) \\ 1-x-y &= 6\bar{p}\bar{q}(1-F) \\ y &= \bar{q} - 3\bar{p}\bar{q}(1-F) \end{aligned}\right.$$

$$(36)\quad \text{Thus } \mu_4' = y + (1-x-y)\int_0^1 q^4 dq = y + (1-x-y)/5$$

$$= \bar{q} - (9/5)\bar{p}\bar{q}(1-F)$$

$$(37)\qquad \overline{\sigma_w^2} = \mu_2' - \mu_4' = (4/5)\bar{p}\bar{q}(1-F)$$

This is the first term of the general expression arrived at by ROBERTSON by his wholly different mode of attack, and is the limiting value as F increases. It is 4 percent smaller than the corresponding limiting value, $\overline{\sigma_w^2} = (5/6)\overline{pq}(1-F)$ which the formula for the steady state takes when F is very close to 1.

OTHER COMPLICATIONS

The effects of incomplete dominance and overdominance in a system of partially isolated strains in a steady state can be analyzed similarly by use of the first four moments of $\phi(q)$. The formulae are in general more cumbersome than with complete dominance.

Recurrent mutation can be introduced into the concept of F without difficulty since its effect on gene frequency is like that of immigration. Letting u and v be the rates of mutation from and to the gene in question $\Delta q = v(1-q) - uq - m(q-\bar{q})$. It is merely necessary to let $a = 4N(m+u+v)$ to arrive at the same formulae as before in terms of $\bar{q}$ and F.

The introduction of selection into the concept is more difficult since selection tends to produce changes in gene frequency that are quadratic even with no dominance, and cubic with dominance in any degree, without considering the complications from the fact that selection operates on the genotype as a whole rather than on the separate loci. The formulae for $\phi(q)$ became unintegrable except by empirical means even in the simplest cases. If the state of balance is such that the standard deviation of $\phi(q)$ is small, an approximation can be obtained by using the best linear expression for Δq in the neighborhood of $\bar{q}$.

SUMMARY

The variance of a character, dependent on a completely recessive gene, in a population with partially isolated strains in which the tendency toward fixation due to inbreeding is balanced at a certain level of inbreeding by occasional cross breeding, is analyzed into the variance of strain means and the variance within strains. The formulae are in terms of gene frequency, inbreeding coefficient and gene effect.

The results are compared with those obtained by ROBERTSON in the case of subdivision into completely isolated strains that are tending toward fixation.

LITERATURE CITED

KOLMOGOROV, A., 1935 Deviations from Hardy's Formula in partial isolation. C. R. de l'Acad. des Sciences de l'U.R.S.S. 3(7) : 129–132.

MALÉCOT, G., 1948 Les mathématiques de l'hérédité. 63 p. Paris: Masson & Cie.
ROBERTSON, A., 1952 The effect of inbreeding on the variation due to recessive genes. Genetics **37**: 189–207.
WRIGHT, S., 1931 Evolution in Mendelian populations. Genetics **16**: 97–159.
1938 The distribution of gene frequencies in populations of polyploids. Proc. Nat. Acad. Sci. **24**: 372–377.
1945 The differential equation of the distribution of gene frequencies. Proc. Nat. Acad. Sci. **31**: 382–389.
1951 The genetical structure of populations. Ann Eug. **15**: 323–354.

APPENDIX

The conditions for a steady state with respect to the distribution of gene frequencies can be represented by a series of equations representing the persistence unchanged of each moment, after the occurrence of a systematic change and a random change. All moments obviously exist because of the finite range, 0 to 1. If all remain unchanged, the frequency curve does not change. The equation representing persistence of the n'th moment is as follows. The frequencies of q and δq are represented by $f(q)$ and $f(\delta q)$ respectively.

$$(38) \qquad \sum_{q=0}^{1} \sum_{\delta q=-q}^{1-q} [(q-\bar{q}) + (\Delta q + \delta q)]^n f(\delta q) f(q) = \sum_{q=0}^{1} (q-\bar{q})^n f(q)$$

Expanding the left hand member in powers of $(q-\bar{q})$ and $(\Delta q + \delta q)$, we note that the first term cancels the right hand member. Moreover, the following must hold for the *random* deviations: $\Sigma \delta q f(\delta q) = 0$, $\Sigma (\delta q)^2 f(\delta q) = \sigma_{\delta q}^2$, $\Sigma\Sigma [\delta q (q-\bar{q}) f(\delta q) f(q)] = 0$, $\Sigma\Sigma [\delta q \Delta q f(\delta q) f(q)] = 0$.

Unless $\sigma_{\delta q}^2$ is of the order of Δq or greater, the latter dominates so much that the distribution is practically restricted to the equilibrium value of q. The case of interest is that in which terms in $(\Delta q)^2$, $(\delta q)^3$, $(\Delta q)(\delta q)^2$ and higher powers may be treated as negligible. With this assumption the equations reduce to the following.

$$(39) \qquad \sum_{q=0}^{1} (q-\bar{q})^{n-1} \Delta q f(q) + \frac{n-1}{2} \sum_{q=0}^{1} [(q-\bar{q})^{n-2} \sigma_{\delta q}^2 f(q)] = 0$$

It is convenient at this point to substitute integration for summation, and $\phi(q)dq$ for $f(q)$, and represent $\Delta q\, \phi(q)dq$ by $d\chi(q)$

$$(40) \qquad \int_0^1 (q-\bar{q})^{n-1} d\chi(q) + \frac{n-1}{2} \int_0^1 (q-\bar{q})^{n-2} \sigma_{\delta q}^2 \phi(q) dq = 0$$

Integrating the first term by parts

$$(41)\quad \left[\chi(q)(q-\bar{q})^{n-1}\right]_0^1 - (n-1)\int_0^1 \chi(q)(q-\bar{q})^{n-2}dq + \frac{n-1}{2}\int_0^1 (q-\bar{q})^{n-2}\sigma_{\delta_q}^2\phi(q)dq = 0$$

$$(42)\quad \text{If } n=1,\ [\chi(q)]_0^1 = 0,\ \chi(1)=\chi(0)$$

$$(43)\quad [\chi(q)(q-\bar{q})^{n-1}]_0^1 = \chi(1)[(1-\bar{q})^{n-1} - (-\bar{q})^{n-1}] = (n-1)\chi(1)\int_0^1 (q-\bar{q})^{n-2}dq$$

Equation (41) becomes

$$(44)\quad \int_0^1 (q-\bar{q})^{n-2}[\chi(1) - \chi(q) + (1/2)\sigma_{\delta_q}^2\phi(q)]dq = 0$$

Thus *all* moments are the same before and after the occurrence of systematic and random processes if the following holds.

$$(45)\quad \chi(q) - \chi(1) = (1/2)\sigma_{\delta_q}^2\phi(q)$$

$$(46)\quad \frac{d}{dq}\log[\chi(q)-\chi(1)] = \left[\frac{d}{dq}\chi(q)\right] \Big/ [\chi(q)-\chi(1)] = \frac{\Delta q\phi(q)}{(1/2)\sigma_{\delta_q}^2\phi(q)} = \frac{2\Delta q}{\sigma_{\delta_q}^2}$$

$$(47)\quad \log[\chi(q)-\chi(1)] = \log(C/2) + 2\int(\Delta q/\sigma_{\delta_q}^2)dq$$

$$(48)\quad [\chi(q)-\chi(1)] = (C/2)e^{2\int(\Delta q/\sigma_{\delta_q}^2)dq}$$

Equating the two expressions for $[\chi(q)-\chi(1)]$ of (45) and (48), we get the desired formula in which the constant C is such that $\int_0^1\phi(q)dq = 1$.

$$(49)\quad \phi(q) = (C/\sigma_{\delta_q}^2)e^{2\int(\Delta q/\sigma_{\delta_q}^2)dq}$$

This is the general formula for the distribution of a gene frequency when a steady state has been reached.