Proc. Natl. Acad. Sci. USA
Vol. 76, No. 6, pp. 2862–2865, June 1979
Genetics

Continuous variation caused by genes with graduated effects

(polygenic inheritance/multifactorial transmission)

STEVEN MATTHYSSE†, KENNETH LANGE‡, AND DIANE K. WAGENER§

†Mailman Research Center, McLean Hospital, Belmont, Massachusetts 02178; ‡Department of Biomathematics, University of California, Los Angeles, California 90024; and §Western Psychiatric Institute, University of Pittsburgh, Pittsburgh, Pennsylvania 15261

Communicated by Seymour S. Kety, February 26, 1979

ABSTRACT **The classical polygenic theory of inheritance postulates a large number of genes with small, and essentially similar, effects. We propose instead a model with genes of gradually decreasing effects. The resulting phenotypic distribution is not normal; if the gene effects are geometrically decreasing, it can be triangular. The joint distribution of parent and offspring genic value is calculated. The most readily testable difference between the two models is that, in the decreasing-effect model, the variance of the offspring distribution from given parents depends on the parents' genic values. The more the parents deviate from the mean, the smaller the variance of the offspring should be. In the equal-effect model the offspring variance is independent of the parents' genic values.**

In the classical polygenic model of inheritance, the contributions of a large number of loci with small but similar effects are summed to produce a phenotypic distribution that is approximately normal (for exact conditions, see ref. 1). It seems more plausible to assume that the contributions of the loci are unequal, with a few major genes and a larger number of minor genes. Indeed, it may not be an exaggeration to suggest that every locus contributes in some way to any given quantitative trait, although the effects of most loci are vanishingly small. An example of unequal gene contributions to a continuously varying trait is heading time in wheat, for which one locus was detected that accounted for about 80% of the additive variance in the trait, and three other loci that jointly accounted for about 14% of the additive variance. The remainder presumably represents the contributions of genes of even smaller effect (2). The common observation of unequal response of different lines to selection for a quantitative trait may be explained in some cases by segregation of major genes of a polygenic character (3). Because the individual genes contributing to a quantitative trait affect the phenotype in different ways, it is not surprising to find their contributions unequal. For example, in *Drosophila*, wing size is associated predominantly with variance in cell number and to a lesser degree with cell size (4).

We assume an infinite number of loci, each with two possible alleles A_n, B_n; the two alleles actually present at each locus are assumed independent of each other and of the alleles at the other loci. The genetic contributions are additive without dominance, X_n representing the contribution of the paternal allele and Y_n the contribution of the maternal allele at the nth locus. The loci are not assumed to have equal effects; the contributions of each allele at the nth locus are $\delta_n + a_n$ from A_n (probability p_n), $\delta_n - a_n$ from B_n (probability $1 - p_n$) ($a_n \geq 0$). δ_n is so chosen that the means of X_n and Y_n are zero: $\delta_n = (1 - 2p_n)a_n$. Assume the a_n are bounded.

The variance of X_n (or Y_n) is $\sigma_n^2 = 4p_n(1 - p_n)a_n^2$. There are two possibilities: (*i*) $\Sigma_{n=1}^{\infty} \sigma_n^2 = \infty$; (*ii*) $\Sigma_{n=1}^{\infty} \sigma_n^2 < \infty$.

Case i. $\Sigma_{n=1}^{\infty} \sigma_n^2 = \infty$. Let $s_n^2 = \sigma_1^2 + \ldots + \sigma_n^2$. The distribution of the normalized sum $(X_1 + \ldots + X_n)/S_n$ tends to the normal distribution with zero mean and unit variance, as $n \to \infty$ (ref. 5, p. 264).

Case ii. $\Sigma_{n=1}^{\infty} \sigma_n^2 < \infty$. Then $X_1 + \ldots + X_n$ converges almost surely to a random variable X (ref. 6, p. 502). X is the limit of the contributions of the paternal alleles only; the distribution function of genic value $G = X + Y$ is the self-convolution of the distribution function of X. In this case, X (and therefore G also) cannot have a normal distribution, because whenever $X_1 + \ldots + X_n$ converges to a limit X, $X_2 + \ldots + X_n$ converges to a limit X'; $X = X_1 + X'$; and X_1 and X' are independent. If the sum of two independent random variables is normally distributed, each of them also has a normal distribution (ref. 5, p. 525), whereas X_1 is atomic. On the other hand, for certain models of this type, the distributions of X and G can be made arbitrarily close to normal. For example, if $p_n = \frac{1}{2}$ (all n), the normalized distribution $F(x)$ of X satisfies

$$|F(x) - N(x)| \leq \frac{6 \sum_{n=1}^{\infty} a_n^3}{\left(\sum_{n=1}^{\infty} a_n^2\right)^{3/2}}$$

for all x, by the Berry–Esseen theorem [$N(x)$ is the normal distribution with zero mean and unit variance] (ref. 5, p. 544). This quantity can be made vanishingly small by choosing appropriate $\{a_n\}$; e.g., $a_n = n^{-b}$ for b near and greater than $\frac{1}{2}$ or $a_n = a^n$ for a near and less than 1.

Classification of limit distributions

According to a remarkable theorem of Jessen and Wintner (7), a convergent convolution of atomic distributions is either (*i*) atomic, (*ii*) absolutely continuous, or (*iii*) continuous but singular. No mixed cases are possible. In other words, either (*i*) X takes only a discrete set of values; (*ii*) X has a probability density (not necessarily continuous or bounded); (*iii*) no single value has positive probability, but there is a set of Lebesgue measure zero such that X is in this set with probability 1. In order to construct a model that is suitable for genetic applications, it is reasonable to consider only absolutely continuous limits with densities that are continuous or at least bounded.

A necessary and sufficient condition that X have a continuous distribution function is $\prod_{n=1}^{\infty} \max(p_n, 1 - p_n) \to 0$ (ref. 6, p. 537); alternatively, $\Sigma_{n=1}^{\infty} \min(p_n, 1 - p_n) = \infty$. For further study of the classification problem, we will assume $p_n = \frac{1}{2}$ (all n). The limit distribution is then continuous, although not necessarily absolutely continuous. The characteristic function of X is $\phi(t) = \prod_{n=1}^{\infty} \cos(a_n t)$ because $\delta_n = 0$. Wintner (8) discusses such characteristic functions in detail. For example, if $a_n = n^{-b}$ for $b > \frac{1}{2}$, the corresponding distribution has an infinitely differentiable density (ref. 8, p. 147).

If $a_n = a^n$ and $a = (\frac{1}{2})^{1/k}$ ($k = 1, 2, \ldots$), then $\phi(t) = O(|t|^{-k})$ (refs. 8, 9), so the characteristic function of G is $O(|t|^{-2k})$. G therefore has a continuous density with at least $2k - 2$ contin-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

uous derivatives (ref. 8, p. 118). We will treat the case $a = 1/2$ more fully later, because it leads to a computable density.

Surprisingly, even though the distribution corresponding to $a = (1/2)^{1/k}$ is absolutely continuous for $k = 1,2,\ldots, 1/2 \leq a < 1$ does not guarantee the existence of a density for G. An example is $a = 1/2(\sqrt{5} - 1)$, for which $\phi(t)$ does not $\to 0$ as $|t| \to \infty$ (ref. 8, p. 142), so its self-convolution cannot have a density (ref. 5, p. 514).

If $a_n = a^n$, in which $a = 1/k$ $(k = 3,4,\ldots)$, $\phi(t)$ does not $\to 0$ as $|t| \to \infty$ (ref. 8, p. 140), so G is singular. For general a in the range $0 < a < 1/2$, it is known that X is singular (ref. 8, p. 140). This suffices to show that G does not have a bounded or continuous density, because whenever the self-convolution of a symmetric probability distribution F has a bounded density, F itself has a density. Indeed, the characteristic function ϕ of F satisfies $\phi^2 \geq 0$ because F is symmetric (ref. 5, p. 499) and furthermore $\phi^2 \in L$ because $F * F$ has a bounded density (ref. 5, p. 510). Consequently $\phi \in L^2$, and the appropriately normalized inverse Plancherel transform of ϕ is the required density. A continuous density for $F * F$ is also impossible because $F * F$ is concentrated on a finite interval.

Distribution of genic value when $p_n = 1/2$ (all n)

When $p_n = 1/2$ (all n), G has characteristic function $\theta(t) = [\prod_{n=1}^{\infty} \cos(a_n t)]^2$. Because the a_n are bounded, for $|t|$ sufficiently small $\log \theta(t)$ can be expanded as $2\sum_{k=1}^{\infty} \log\cos(a_k t)$, and each term has the power series expansion

$$\log\cos(a_k t) = \sum_{n=1}^{\infty} (-1)^n \frac{2^{2n-1}(2^{2n}-1)B_{2n}}{n(2n)!} a_k^{2n} t^{2n},$$

in which B_n is the nth Bernoulli number. This expression is derived by integrating term by term the expansion for the tan function (ref. 9, p. 20). The double series can be rearranged because all terms are negative, and when compared with the expansion of characteristic functions in terms of cumulants (ref. 10, p. 115),

$$\log\theta(t) = \sum_{h=1}^{\infty} \frac{\kappa_h}{h!}(it)^h,$$

yields for the cumulants:

$$\kappa_{2n} = \frac{2^{2n}(2^{2n}-1)B_{2n}}{n} \sum_{k=1}^{\infty} a_k^{2n} \quad (n \geq 1),$$

the odd cumulants vanishing. From the cumulants, the moments may be derived (ref. 11, p. 69).

If, in addition, the a_n satisfy $a_n = 2^{-(n+1)}$, X is uniformly distributed in $[-1/2, 1/2]$ (ref. 6, p. 557). The self-convolution of a uniform distribution is triangular (ref. 5, p. 27), so the density function of G is $1 - |x|$ for $|x| \leq 1$ and 0 for $|x| > 1$ (Fig. 1).

Joint distribution of parents and offspring

Assume $p_n = 1/2$ (all n) and $a_n = -2^{(n+1)}$, other assumptions as above. The joint distribution of the genic values of two parents (G, G') and one child (G''), assuming random mating, can be computed explicitly by using characteristic functions. We will show that the conditional distribution of offspring genic value, given the parents' genic values, is trapezoidal. Most important, the genic variance of the offspring is not independent of the parents' genic values, and this constitutes a major difference from the classical "polygenic" model.

Let the contribution of the father's paternal allele at the nth locus be X_n $(= \pm 2^{-(n+1)})$ and that of his maternal allele be Y_n; the same quantities for the mother will be called X'_n and Y'_n. The contributions of the child's paternal and maternal allele will be

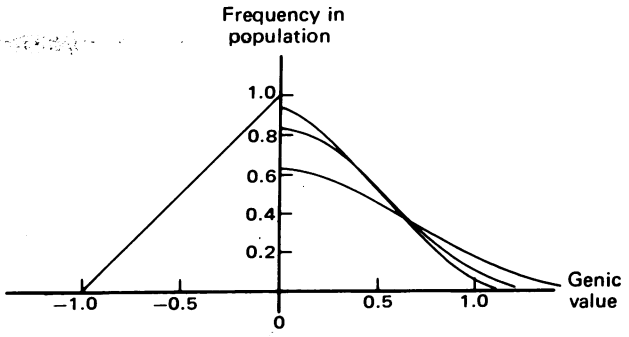

FIG. 1. Frequency distribution of genic value in the population according to the geometrically-decreasing-effect model. The curves on the right include independent normally distributed environmental components with standard deviations 0.1, 0.25, and 0.5 (from top to bottom at frequency intercepts).

called X''_n and Y''_n. The characteristic function of the random variable (G_n, G'_n, G''_n) is $\theta_n(t_1,t_2,t_3) =$

$$E\{\exp[it_1(X_n + Y_n) + it_2(X'_n + Y'_n) + it_3(X''_n + Y''_n)]\}.$$

For fixed $X_n, Y_n, X'_n, Y'_n, X''_n$ takes on the values X_n and Y_n with equal probability and Y''_n independently takes on the values X'_n and Y'_n with equal probability. Consequently

$$\begin{aligned} &E\{\exp[it_3(X''_n + Y''_n)] | X_n, Y_n, X'_n, Y'_n\} \\ &\qquad \times \exp[it_1(X_n + Y_n) + it_2(X'_n + Y'_n)] \\ &= 1/4 \exp[i(t_1 + t_3)X_n + it_1 Y_n + i(t_2 + t_3)X'_n + it_2 Y'_n] \\ &+ 1/4 \exp[i(t_1 + t_3)Y_n + it_1 X_n + i(t_2 + t_3)X'_n + it_2 Y'_n] \\ &+ 1/4 \exp[i(t_1 + t_3)X_n + it_1 Y_n + i(t_2 + t_3)Y'_n + it_2 X'_n] \\ &+ 1/4 \exp[i(t_1 + t_3)Y_n + it_1 X_n + i(t_2 + t_3)Y'_n + it_2 X'_n]. \end{aligned}$$

Because X_n, Y_n, X'_n, and Y'_n are independent with identical distributions, the expected value of each term is the same, and

$$\theta_n(t_1,t_2,t_3) = \phi_n(t_1 + t_3)\phi_n(t_1)\phi_n(t_2 + t_3)\phi_n(t_2),$$

in which ϕ_n is the characteristic function of X_n. Because of the independence of the alleles at distinct loci, the characteristic function of the random variable (G,G',G'') is the product of the θ_n; $\theta(t_1,t_2,t_3) = \phi(t_1 + t_3)\phi(t_1)\phi(t_2 + t_3)\phi(t_2)$, in which ϕ is the characteristic function of X. Setting $t_2 = 0$ yields the characteristic function $\theta(t_1,t_3) = \phi(t_1 + t_3)\phi(t_1)\phi(t_3)$ of (G,G''). Let u_δ denote the uniform density in $[-1/2 - \delta, 1/2 - \delta]$. $u_\delta(x) = u(x + \delta)$. Let $g(s;x_1,x_2) = ((uu_{x_1}) * (uu_{x_2}))(s)$, in which uu_x denotes the product of u and u_x and $*$ denotes convolution. Consider the Fourier transform of $g(x_3 - x_1 - x_2; x_1, x_2)$

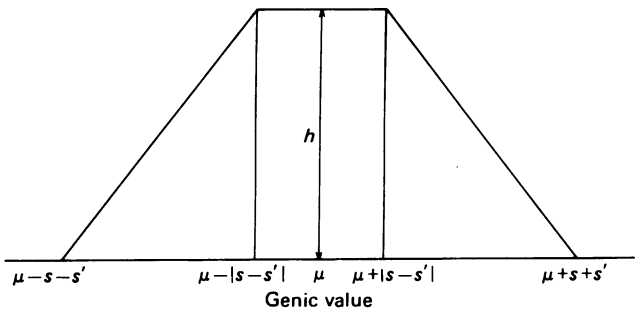

FIG. 2. Conditional distribution of offspring genic value for given parents' genic values. μ, s, and s' are defined in the text. $h = 1/[2\max(s,s')]$.

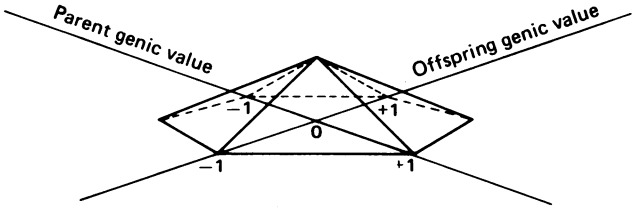

FIG. 3. Joint distribution of the genic values of parent and offspring.

$$\int\int\int((uu_{x_1}) * (uu_{x_2}))(x_3 - x_1 - x_2) \times \exp[i(t_1x_1 + t_2x_2 + t_3x_3)]\, dx_1dx_2dx_3$$
$$= \int\int\int((uu_{x_1}) * (uu_{x_2}))(s)\exp[i(t_1x_1 + t_2x_2 + t_3(x_1 + x_2))] \times \exp(it_3s)\, dsdx_1dx_2$$
$$= \int \widehat{uu_{x_1}}(t_3)\exp[i(t_1 + t_3)x_1]dx_1 \times \int \widehat{uu_{x_2}}(t_3)\exp[i(t_2 + t_3)x_2]\, dx_2$$

($\wedge$ denoting Fourier transform) because $\widehat{(uu_{x_1}) * (uu_{x_2})} = \widehat{uu_{x_1}}\widehat{uu_{x_2}}$. The first of these integrals is

$$\int\int u(s)u(s + x_1)\exp(it_1x_1)\exp[it_3(s + x_1)]\, ds\, dx_1 = \int\int u(s)u(w)\exp[i(t_1 + t_3)w]\exp(-it_1s)\, dw\, ds \quad (w = s + x_1)$$
$$= \phi(-t_1)\phi(t_1 + t_3),$$

and the second, likewise, is $\phi(-t_2)\phi(t_2 + t_3)$. Because $\phi(-t) = \phi(t)$, the Fourier transform of $g(x_3 - x_1 - x_2; x_1, x_2)$ is $\theta(t_1,t_2,t_3)$. Consequently $g(x_3 - x_1 - x_2; x_1,x_2)$ is the density of the random variable (G,G',G''). It is easy to verify that $g(x_3 - x_1 - x_2; x_1, x_2)$ is the Lebesgue measure of the intersection $[x_3 - x_1 - x_2 - \frac{1}{2}, x_3 - x_1 - x_2 + \frac{1}{2}] \cap [x_3 - x_2 - \frac{1}{2}, x_3 - x_2 + \frac{1}{2}] \cap [-\frac{1}{2},\frac{1}{2}] \cap [-x_2 - \frac{1}{2}, -x_2 + \frac{1}{2}]$; i.e., $g(x_3 - x_1 - x_2; x_1, x_2) = \max[0,1 + \min(x_3 - x_1,x_3,x_2,0) - \max(x_3 - x_1,x_3,x_2,0)]$. It follows that the conditional distribution of offspring genic value given the parents genic values is trapezoidal (Fig. 2), in which $\mu = \frac{1}{2}(G + G')$ is the mean of the offspring distribution;[¶] $s = \frac{1}{2}(1 - |G|)$ and $s' = \frac{1}{2}(1 - |G'|)$. Similarly, the density of the random variable (G,G'') is $\max[0,1 + \min(x_3 - x_1,x_3,0) - \max(x_3 - x_1,x_3,0)]$ (Fig. 3).[‖]

Nonhomoscedasticity of offspring variances

The variance of the offspring distribution from given parents is $\frac{1}{3}[s^2 + (s')^2]$. In the classical polygenic model, the variance of offspring genic value does not depend on the parents, because of the assumption of multivariate normality. The nonhomoscedasticity of the new model is its most distinctive and readily testable feature. The more the parents deviate from the mean, the smaller the offspring variance should be. In this respect the model behaves as if there were a small number of loci (12).

The correlation between within-sibship variance and an estimator of familial deviation from the mean has been proposed, as a method of detecting major locus effects, by Fain (13) and by Smith *et al.* (14). In order to estimate familial deviation from the mean, Smith *et al.* use the phenotypic values of the parents; Fain uses the within-sibship mean itself. The latter measure has the advantage that data on only one generation is needed; it turns out, however, to be less powerful for testing the decreasing-effects model (see below).

It is possible to calculate exactly the theoretical correlation,

[¶] The offspring mean can be shown to be the average of the parents' genic values whenever $\Sigma_{n=1}^{\infty}\sigma_n^2 < \infty$.

[‖] The joint parent (G)-gamete (g) distribution is uniform within the rhomboid $\{G,g \mid |g| \leq \frac{1}{2}, |G - g| \leq \frac{1}{2}\}$.

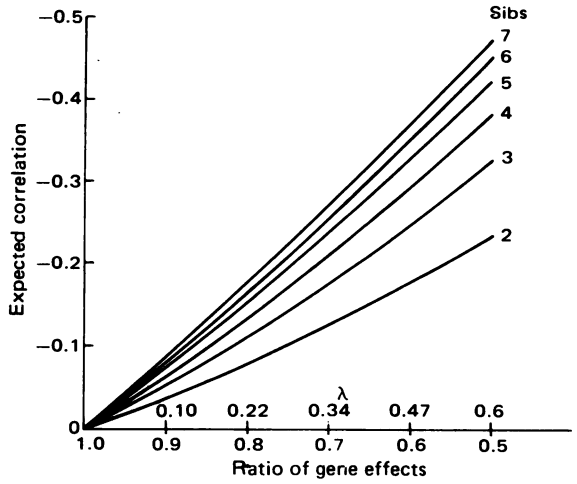

FIG. 4. Expectation value of the correlation between within-sibship variance and mean-square parental deviation, as a function of sibship size and rate of decrease of gene effects.

r_t, between the within-sibship variance for an n-sib family,

$$V = \frac{1}{n-1}\sum_{j=1}^{n}(G_j'' - \overline{G''})^2,$$

in which G_j'' are the genic values of the sibs and $\overline{G''}$ their mean, and either the mean-square parental deviation

$$D^2 = \tfrac{1}{2}(G^2 + (G')^2)$$

or the square of the within-sibship mean ($\overline{G''^2}$) itself. A gene frequency of $\frac{1}{2}$ is assumed at each locus, but the gene effects a_n are allowed to be arbitrary. Both correlations depend on only two parameters: the sibship size n and a measure of the rate of decrease of gene effects at successive loci

$$\lambda = \frac{\sum_{n=1}^{\infty} a_n^4}{\left(\sum_{n=1}^{\infty} a_n^2\right)^2}.$$

In particular, if the effects at successive loci decrease geometrically ($a_n = a^n$), $\lambda = (1 - a^2)/(1 + a^2)$. The results of these calculations are:

$$E[r_t(V,D^2)] = \frac{-\lambda}{\left[\frac{2}{n-1} + \left(\frac{1}{2} - \frac{n-2}{n(n-1)}\right)\lambda\right]^{1/2}[4 - 2\lambda]^{1/2}}$$

$$E[r_t(V,\overline{G''^2})] = \frac{-\left[\frac{n^3 - 2n^2 + 5n - 4}{2n^2(n-1)}\right]\lambda}{\left[\frac{2}{n-1} + \left(\frac{1}{2} - \frac{n-2}{n(n-1)}\right)\lambda\right]^{1/2} \times \left[\frac{2(n+1)^2}{n^2} - \frac{n^3 + 6n^2 - 3n + 4}{2n^3}\lambda\right]^{1/2}}$$

(Figs. 4 and 5). From a practical point of view, large sibships are needed to test these predictions.

A model of continuous variation based on genes with graduated effects may be useful in the analysis of common diseases without evident Mendelian patterns of inheritance. Gottesman and Shields have proposed that such a model may account for the transmission of schizophrenia: "Some high value genes may

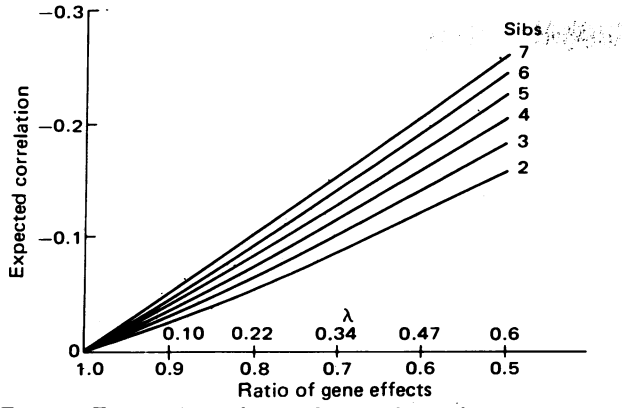

FIG. 5. Expectation value of the correlation between within-sibship variance and the square of the within-sibship mean.

be rather specific for some partitionable aspects of the disorder (e.g., catatonia, paranoid features, or some aspect of blood chemistry, brain protein, or neurophysiology); such possibilities would then encourage both formal segregation analyses and searches for linkage relationships to known genetic markers, but for *facets* of the syndrome." (ref. 15, p. 520).

S.M. was supported by Research Scientist Development Award, U.S. Public Health Service Grant K2-MH-00108, and U.S. Public Health Service Grants MH-16674 and MH-25515. K.L. was supported by National Institutes of Health Special Resource Grant RR-3 and The National Foundation–March of Dimes, Grant 6-55.

1. Lange, K. L. (1978) *J. Math. Biol.* **6**, 59–66.
2. Wehrhahn, C. & Allard, R. W. (1965) *Genetics* **51**, 109–119.
3. Robertson, A. (1958) *Proc. Int. Congr. Genet., 10th* **1**, 199–205.
4. Robertson, F. W. (1959) *Genetics* **44**, 869–896.
5. Feller, W. (1971) *An Introduction to Probability Theory and Its Applications* (Wiley, New York), 2nd Ed., Vol. 2.
6. Kawata, T. (1972) *Fourier Analysis in Probability Theory* (Academic, New York).
7. Jessen, B. & Wintner, A. (1935) *Trans. Am. Math. Soc.* **38**, 48–88.
8. Wintner, A. (1947) *The Fourier Transforms of Probability Distributions* (Edwards, Ann Arbor, MI).
9. Rademacher, H. (1973) *Topics in Analytic Number Theory* (Springer, New York).
10. Wilks, S. S. (1962) *Mathematical Statistics* (Wiley, New York).
11. Kendall, M. G. & Stewart, A. (1963) *The Advanced Theory of Statistics* (Griffin, London), 2nd Ed., Vol. 1.
12. Cavalli-Sforza, L. L. & Kidd, K. K. (1972) *Neurosci. Res. Prog. Bull.* **10**, 416–419.
13. Fain, P. R. (1978) *Ann. Hum. Genet.* **42**, 109–120.
14. Smith, C. A. B., Loesch, D. Z. & Bener, A. (1978) *Ann. Hum. Genet.* **42**, 121–128.
15. Gottesman, I. I. & Shields, J. (1971) *Hum. Hered.* **21**, 517–522.