

GEOMETRIC MODELS OF CERTAIN GENETIC PROCESSES

I. M. YAGLOM

To H. S. M. Coxeter on his sixtieth birthday

1. Classical genetics studies the production of the descendants of a pair of parents with certain genetic characteristics. In the simplest case the genetic type of an individual is determined by a pair of genes each of which can be of one of two types, G and g . A given individual can have these genes in the combination GG (dominant person), or Gg (genetically not different from the combination gG ; hybrid individual), or, finally, gg (recessive individual); GG - and Gg -individuals do not differ in appearance from one another, but a gg -individual is different from them.

The hypothesis of a purely accidental selection from a given population of parental couples (i.e. the hypothesis of "panmixia" (cf. 2, Chap. 10) leads to a Markov chain, defined by the following matrix of transition probabilities:

$$(1) \quad \begin{matrix} & (DD) & (DH) & (DR) & (HH) & (HR) & (RR) \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \frac{1}{16} & \frac{1}{4} & \frac{1}{8} & \frac{1}{4} & \frac{1}{4} & \frac{1}{16} \\ 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} & & & & & \end{matrix}.$$

The letter D denotes a dominant, H a hybrid, R a recessive individual; the parental couples appear as the conditions of the chain; cf., e.g., (1, §7.4). Hence the transition from a population with fractions d, h, r of dominant, hybrid, and recessive individuals to its descendants according to the scheme defined by the chain (1) leads to the following quadratic transformation in the vector space (d, h, r) , expressing the well-known Hardy-Weinberg law (cf. 2, Chap. 10):

$$(2) \quad \begin{aligned} d' &= d^2 + dh + \frac{1}{4}h^2, \\ h' &= dh + 2dr + \frac{1}{2}h^2 + hr, \\ r' &= \frac{1}{4}h^2 + hr + r^2. \end{aligned}$$

Received May 3, 1967. Translated from the Russian by H. Schwerdtfeger.

The geometrical meaning of the transformation (2) is very simple. The formulae (2) can be rewritten in the form

$$\begin{aligned}
 (2a) \quad d' &= (d + \frac{1}{2}h)^2 = \Gamma^2, \\
 h' &= 2(d + \frac{1}{2}h)(\frac{1}{2}h + r) = 2\Gamma\gamma, \\
 r' &= (\frac{1}{2}h + r)^2 = \gamma^2,
 \end{aligned}$$

where $\Gamma = d + \frac{1}{2}h$ and $\gamma = \frac{1}{2}h + r$ are the fractions of the genes G and g in the population. Thus

$$(3) \quad 4d'r' - h'^2 = 0;$$

and consequently the transformation (2) carries all the points of the equilateral triangle DHR , i.e. the set of the vertices of all the vectors (d, h, r) in the three-dimensional (euclidean) space where $d \geq 0, h \geq 0, r \geq 0$, and $d + h + r = 1$, into the arc DR of a parabola along which this triangle intersects the cone $4dr - h^2 = 0$. On the other hand, since

$$(4) \quad d' - r' = \Gamma^2 - \gamma^2 = (\Gamma - \gamma)(\Gamma + \gamma) = \Gamma - \gamma = d - r,$$

all the points of the triangle are shifted along the straight line $d - r = \text{const.}$, perpendicular to the side DR ; hence the transformation represents the parallel projection of the triangle DHR onto the parabola $h^2 = 4dr$ in the direction of the height of the triangle through the vertex H (cf. Fig. 1).

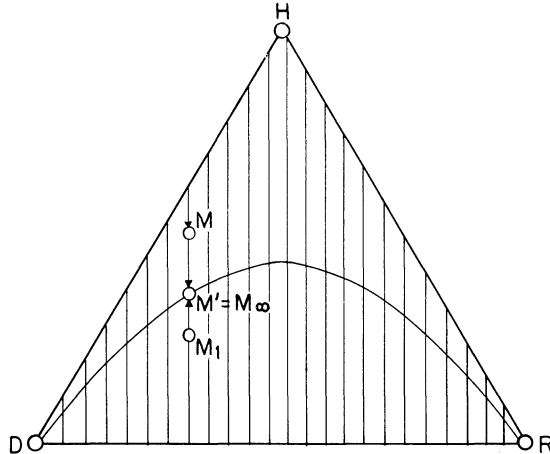


FIGURE 1

From Figure 1 follows, in particular, that under given conditions the evolution is completed in one step and the set of stable populations is characterized by the conditions

$$(5) \quad d : h : r = \Gamma^2 : 2\Gamma\gamma : \gamma^2$$

(this, really, is the Hardy–Weinberg law).

2. Another mathematical model of the process of transfer by heredity of a single characteristic, defined by the genes G and g , is obtained by assuming that the vitality of heterozygotean (hybrid) individuals differs from the vitality of homozygotean (dominant and recessive) individuals. Denoting the conservation coefficient of the hybrids by l , we arrive at Markov chain given by the following matrix of transition probabilities:

$$(6) \quad \begin{bmatrix} (DD) & (DH) & (DR) & (HH) & (HR) & (RR) \\ 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{(l+1)^2} & \frac{2l}{(l+1)^2} & 0 & \frac{l^2}{(l+1)^2} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \frac{1}{4(l+1)^2} & \frac{l}{(l+1)^2} & \frac{1}{2(l+1)^2} & \frac{l^2}{(l+1)^2} & \frac{l}{(l+1)^2} & \frac{1}{4(l+1)^2} \\ 0 & 0 & 0 & \frac{l^2}{(l+1)^2} & \frac{2l}{(l+1)^2} & \frac{1}{(l+1)^2} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

In this case we have the following transformation, which for $l = 1$ will coincide with the transformation (2); it indicates the law of transition from the population (d, h, r) to the hereditary population (d', h', r') :

$$(7) \quad \begin{aligned} d' &= \frac{(d + \frac{1}{2}h)^2}{1 + \frac{1}{2}(l-1)(d + 2h)(2h + r)}, \\ h' &= \frac{2l(d + \frac{1}{2}h)(\frac{1}{2}h + r)}{1 + \frac{1}{2}(l-1)(d + 2h)(2h + r)}, \\ r' &= \frac{(\frac{1}{2}h + r)^2}{1 + \frac{1}{2}(l-1)(d + 2h)(2h + r)}. \end{aligned}$$

For $l \neq 1, \infty$ the transformation (7) has three fixed points: $D(1, 0, 0)$, $R(0, 0, 1)$, and $P_i(1/2(l+1), l/(l+1), 1/2(l+1))$; the last one is the point of intersection of the height $d = r$ of the triangle DHR with the arc of the parabola

$$(8) \quad 4l^2dr = h^2$$

into which the triangle DHR is transformed by the transformation (7) (indeed $4l^2d'r' = h'^2$; cf. Fig. 2). It is obvious that all the points of the line $d = r$ (the height through H of the triangle) are carried into the fixed point P_i by a single application of the transformation (7). As to the other points, observing that for $l > 1$ (the vitality of the hybrids is greater than the norm), we have

$$(9a) \quad |d' - r'| < |d - r|;$$

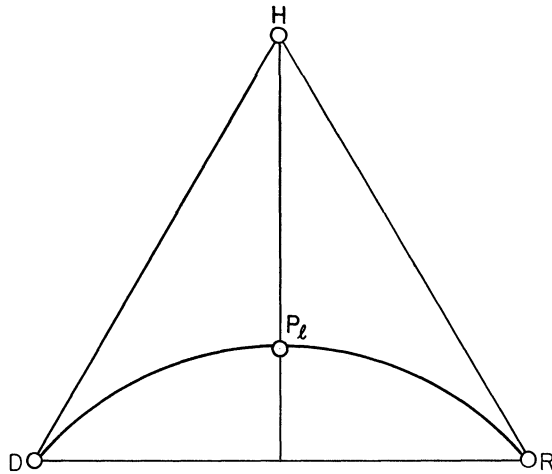


FIGURE 2

thus in the case of multiple application of the transformation (7) all the points of the arc DR of the parabola (8), apart from the end points D and R , are contracted into the point P . Starting from an arbitrary initial composition of the population (not just single dominant or single recessive individuals), in the limit the evolution leads to the population P_i :

$$(10) \quad d: h: r = \frac{1}{2}: l: \frac{1}{2}.$$

If $l < 1$ (i.e. the vitality of the hybrids is smaller than the norm; this includes, in particular, also the case $l = 0$ of the complete lack of vitality of hybrids), then

$$(9b) \quad |d' - r'| > |d - r|;$$

hence in the case of multiple application of the transformation (7) all the points of the "left half" of the triangle DHR (i.e. the points for which $d > r$, which correspond to populations containing an excess number of dominant individuals compared with the number of recessive individuals) are contracted into the point $D(1, 0, 0)$, and all the points of the "right half" of the triangle DHR (i.e. the points for which $d < r$) into the point $R(0, 0, 1)$.

The case $l = 1$ has been analysed above; in the case $l = \infty$ (i.e. complete lack of vitality of the homozygotean individuals) a single application of the transformation (7) reduces the whole triangle DHR (excluding its vertices D and R) into the vertex $H(0, 1, 0)$ of the triangle.

3. A further scheme generalizing the transformation (2) is obtained by replacing the panmixia hypothesis by a hypothesis concerning the mutual attraction between opposite individuals, measured by the attraction coefficient

k . This hypothesis leads to a Markov chain with the following matrix of transition probabilities (cf. 1, p. 176):

$$(11) \quad \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \frac{1}{4(k+3)} & \frac{1}{k+3} & \frac{k(k+1)}{(k+3)(3k+1)} & \frac{1}{k+3} & \frac{2k(k+1)}{(k+3)(3k+1)} & \frac{1}{4(3k+1)} \\ 0 & 0 & 0 & \frac{1}{2(k+1)} & \frac{k}{k+1} & \frac{1}{2(k+1)} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Hence the transition from a population with the fractions d, h, r of dominant, hybrid and recessive individuals to its descendants is described by the following non-linear transformation, which for $k = 1$ includes the transformation (2):

$$(12) \quad \begin{aligned} d' &= \frac{(d + \frac{1}{2}h)^2}{d + h + kr}, \\ h' &= \frac{d + \frac{1}{2}h}{d + h + kr} \left(h + \frac{k(k+1)r}{k(d+h) + r} \right), \\ r' &= \frac{1}{2} \frac{h(\frac{1}{2}h + kr)}{d + h + kr} + \frac{r(\frac{1}{2}kh + r)}{k(d+h)r}. \end{aligned}$$

If $k = 0$ (i.e. the case of the exclusion of marriage between individuals of different appearance), the transformation (11) appears in the form

$$(13) \quad \begin{aligned} d' &= d + \frac{\frac{1}{4}h^2}{d+h}, \\ h' &= h - \frac{\frac{1}{2}h^2}{d+h}, \\ r' &= r + \frac{\frac{1}{4}h^2}{d+h}. \end{aligned}$$

The range of the fixed points of this transformation is the straight line DR (i.e. the line $h = 0$), that is, the subset of the population free from hybrids. Since

$$d' - r' = d - r,$$

we conclude that, in analogy with the case of the transformation (2), all the points of the triangle DHR , as a result of the transformation (13), move along the straight lines $d - r = \text{const.}$, parallel to the height of the triangle through H . Clearly by repetition of this transformation, h tends to zero, i.e. in the limit evolution leads to the disappearance of hybrids (cf. Fig. 3; thus, e.g., the vertex

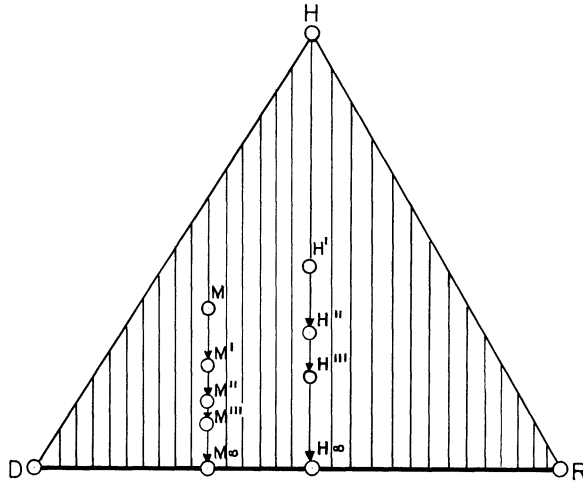


FIGURE 3

$H(0, 1, 0)$ of the triangle after application of the transformation (13) is carried into points of the height $d - r = 0$ for which consequently $h = 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots$.

The transformation

$$(14) \quad \begin{aligned} d' &= 0, \\ h' &= 1 - \frac{\frac{1}{2}h}{d + h}, \\ r' &= \frac{\frac{1}{2}h}{d + h}, \end{aligned}$$

which is the form of the transformation (12) for $k = \infty$ (i.e. the case of the exclusion of marriage between individuals of identical appearance), has a completely different character. This transformation has one fixed point $Q_\infty(0, \frac{1}{2}, \frac{1}{2})$ corresponding to a population consisting half of hybrids and half of recessive individuals. Applying the transformation (14) twice reduces the triangle (from which the segment DH has now to be excluded) into the point Q_∞ (evolution is completed in two generations). If A is some point on the side DH of the triangle DHR , the transformation (14) carries the segment RA into a point A' of the side HR such that $HA' = \frac{1}{2}DA$ (cf. Fig. 4) and thus the triangle DHR into the segment HQ_∞ ; the second application of the transformation (14) carries the segment HQ_∞ into the point Q_∞ .

If $k \neq 1, 0, \infty$, then the transformation (12) has three fixed points:

$$D(1, 0, 0), \quad R(0, 0, 1), \quad Q_k \left(\frac{(\sqrt{k^2 + k} - k)^2}{2k}, \sqrt{k^2 + k} - k, \frac{1}{2} \right)$$

(the last point being the intersection of the middle line $r = \frac{1}{2}$ of the triangle DHR with the parabola $h^2 = 4kdr$; Fig. 5). Multiple application of the transformation (12) leaves the vertices D and R of the triangle fixed, but all the

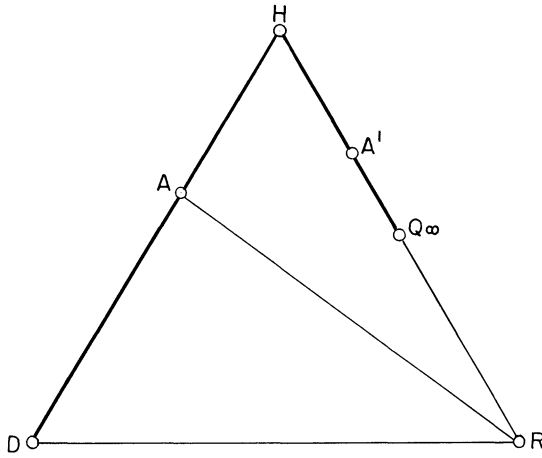


FIGURE 4

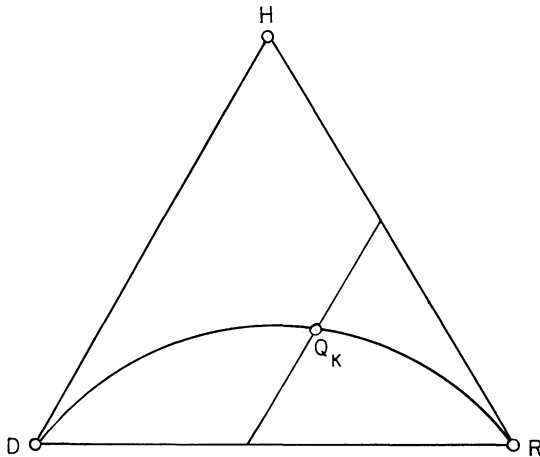


FIGURE 5

other points contract into the point Q_k , corresponding to a population for which

$$(15) \quad d:h:r = (\sqrt{k^2 + k} - k)^2 : 2k(\sqrt{k^2 + k} - k) : k$$

Clearly these considerations can be further complicated by introducing both factors expressed in terms of the coefficients l and k .

4. The scheme described above becomes more complicated in the case of transfer of sex-linked hereditary characteristics: colour blindness (daltonism) or haemophilia in human beings, rose-coloured eyes in the drosophilla (fly), etc.; in order to fix ideas we shall speak of daltonism. The peculiarity of the

process of hereditary transmission of daltonism is connected with the fact that the corresponding genes are localized in the *X*-chromosome defining the sex (the male individual has only one *X*-chromosome, and the female has two *X*-chromosomes). Thus with respect to the characteristic of a male individual there may be only two types corresponding to the presence of a dominant gene *G* (in our case: securing normal sight) in an individual, or of a recessive gene *g*; a female individual may be of all three types: dominant, hybrid, or recessive. In the inheritance of this characteristic a male individual receives his unique gene from the mother; a female individual receives one gene from the father and one from the mother. Correspondingly six types of parental couples are possible. If the letters *D*, *H*, *R* have the same meaning as before and the first letter refers to the father and the second to the mother, the matrix of the transition probabilities of the Markov chain with the six conditions has the form:

$$(16) \quad \begin{matrix} & (DD) & (DH) & (DR) & (RD) & (RH) & (RR) \\ \left[\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right]. \end{matrix}$$

We shall denote the fractions of dominant, hybrid, and recessive individuals in the female part of the population, as before, by the letters *d*, *h*, and *r* respectively, and the fractions of dominant and recessive individuals in the male part of the population by δ and ρ respectively. Thus the genetic composition of a population is characterized by five "coordinates," *d*, *h*, *r*, δ , ρ ; the set of the end points of all vectors in the five-dimensional (euclidean) space (*d*, *h*, *r*, δ , ρ) corresponding to all possible populations then fills a triangular prism: the direct product of the triangle $d \geq 0, h \geq 0, r \geq 0, d + h + r = 1, \delta = \rho = 0$ and the segment $d = h = r = 0, \delta \geq 0, \rho \geq 0, \delta + \rho = 1$. The transition from a point (*d*, *h*, *r*, δ , ρ) which characterizes a given population, to a point (*d'*, *h'*, *r'*, δ' , ρ') which corresponds to a set of descendants of this population, is given by the formulae

$$(17) \quad \begin{cases} d' = d\delta + \frac{1}{2}h\delta, & h' = \frac{1}{2}h\delta + r\delta + d\rho + \frac{1}{2}h\rho, & r' = \frac{1}{2}h\rho + r\rho, \\ \delta' = d + \frac{1}{2}h, & \rho' = \frac{1}{2}h + r. \end{cases}$$

Of the coordinates *d*, *h*, *r*, δ , ρ , only three are independent; as independent coordinates characterizing the points of our prism $DHRD_1H_1R_1$ (the vertices of the prism have the coordinates $D(1, 0, 0, 1, 0)$, $D_1(1, 0, 0, 0, 1)$, etc.) we take the magnitudes $\Gamma = d + \frac{1}{2}h$ (the fraction of the genes *G* in the female part of the population), *h* (the fraction of the hybrids in the female part of the population), and δ (the fraction of the genes *G* or of the dominant individuals

in the male part of the population). In the coordinates (Γ, h, δ) the transformation (17) appears in the form

$$(17a) \quad \Gamma' = \frac{1}{2}(\Gamma + \delta), \quad h' = \Gamma + \delta - 2\Gamma\delta, \quad \delta' = \Gamma.$$

Since $h' = 2\Gamma' - 4\Gamma'\delta + 2\delta'^2$, the transformation (17a) carries the whole triangular prism into the hyperbolic paraboloid

$$(18) \quad h = 2\Gamma - 4\Gamma\delta + 2\delta^2$$

(more accurately, into the part of the surface of this hyperbolic paraboloid which is included inside the prism); the edges DH and $H_1 R_1$ of the prism are straight-line generators of the hyperbolic paraboloid (18) of one family, and the diagonals DH_1 and HR_1 of faces of the prism are straight-line generators from a second family (cf. Fig. 6). The fixed points of the transformation (17a) can be found from the relations

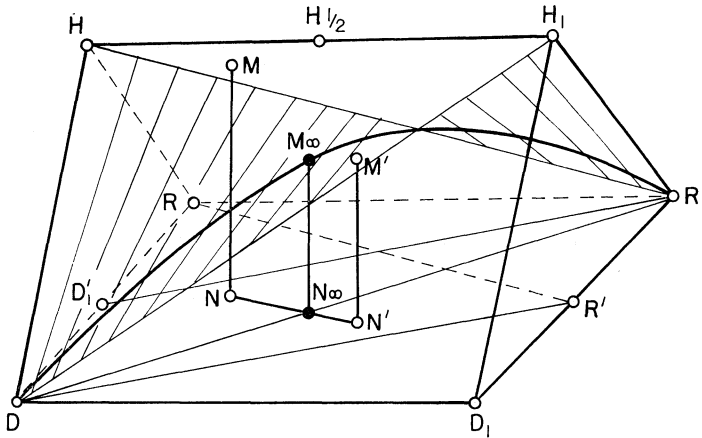


FIGURE 6

$$(19) \quad \Gamma = \frac{1}{2}(\Gamma + \delta), \quad h = \Gamma + \delta - 2\Gamma\delta, \quad \delta = \Gamma;$$

they generate the arc DR_1 of a parabola which appears as the intersection of the hyperbolic paraboloid (18) and the “diagonal” plane $\Gamma = \delta$ of the prism $DHRD_1 H_1 R_1$ (the plane $DH_{\frac{1}{2}} R_1$, where $H_{\frac{1}{2}}(\frac{1}{2}, 1, \frac{1}{2})$ is the mid-point of the edge HH_1 of the prism).

Since the transformation (17a) is completely defined by the linear transformation

$$(20) \quad \Gamma' = \frac{1}{2}(\Gamma + \delta), \quad \delta' = \Gamma$$

of the square $DRR_1 D_1$, it is natural to study first the transformation (20). It carries the square $DRR_1 D_1$ into the parallelogram $DR'R_1 D'_1$, where R' and D'_1 are the mid-points of the edges $R_1 D_1$ and DR respectively of the prism. An arbitrary point N of the square $DRR_1 D_1$ is carried onto a point N' such that

$NN' || RR'$ and the distance $N_\infty N'$ from N' to the diagonal DR_1 of the square DRR_1D_1 equals half of the distance NN_∞ from N to DR_1 . Thus an arbitrary point M of the prism $DHRD_1H_1R_1$, if projected into the plane $h = 0$ at the point N , is carried by the transformation (17a) into a point M' of the hyperbolic paraboloid (18) which is projected into the point N' . Multiple application of the transformation (17a) carries the point M into the point M_∞ of the parabola (19) and this point is transformed into the point N_∞ , the intersection of the segment NN' with the diagonal DR_1 .

The genetic condition $\Gamma = \delta$ corresponds to the equal distribution

$$\Gamma: \gamma = (d + \frac{1}{2}h): (\frac{1}{2}h + r)$$

and $\delta: \rho$ of the genes G and g in the female and in the male part of a population; stable population (18) corresponds to the conditions

$$(21) \quad \Gamma: \gamma = \delta: \rho \quad \text{and} \quad d: h: r = \Gamma^2: 2\Gamma\gamma: \gamma^2$$

which express a purely accidental combination of genes G and g in pairs in the female part of the population.

The author wishes to express his gratitude to I. A. Rombe for her collaboration.

REFERENCES

1. J. G. Kemeny and J. L. Snell, *Finite Markov chains* (Princeton, Toronto, New York, London, 1960).
2. C. Stern, *Principles of human genetics* (San Francisco and London, 1960).

*Moscow State Pedagogical Institute,
Moscow, U.S.S.R.*