Technical Papers

On an Equation Specifying Equilibrium Populations

C. C. Li

Department of Biostatistics, University of Pittsburgh, Pittsburgh, Pennsylvania

Let x, $2y$, and z be the proportions of AA, Aa, and aa, respectively, in a population, where $x+2y+z=1$. It is well known that in a large random-mating population $x=p^2$, $2y=2pq$, and $z=q^2$, where p denotes the frequency of gene A and $q=1-p$ denotes that of a in the population. Hence the equation

$$xz-y^2=0 \qquad (1)$$

specifies all panmictic populations (in equilibrium). Any population may be represented by a point P inside an equilateral triangle XYZ (Fig. 1), letting x, $2y$, and z be the perpendiculars from P to sides YZ, ZX, and XY, respectively; and $x+2y+z=$ altitude of the triangle, which we take as unity. Thus, the parabola (1) is the locus of points representing all panmictic populations (*1*).

Now consider a population with an inbreeding coefficient F which is defined as the (product-moment) correlation coefficient between uniting gametes (*2*). Here, $x=p^2+Fpq=(1-F)p^2+Fp$, $2y=2(1-F)pq$, and $z=q^2+Fpq=(1-F)q^2+Fq$. The parameter F is completely independent of the gene frequencies of a population. With any fixed value of F, the locus of these population points is obviously given by the parabola (*1*):

$$\left.\begin{array}{c} xz-y^2=F(x+y)(y+z) \\ \text{or} \\ xz-y^2=\dfrac{y\cdot F}{1-F}. \end{array}\right\} \qquad (2)$$

The primary purpose of this note is to put (1) and (2) in a new form which can be generalized to the case of multiple alleles, and to reveal the biological meaning of these equations. We note that x is the proportion of gene A existing in homozygous condition, whereas the total frequency of A in the population is $x+y=p$. Hence the proportion of A in homozygous condition among its total frequency is $x/(x+y)=x/p$. The above two equations of parabola are equivalent to

$$\frac{x}{x+y}+\frac{z}{y+z}=\frac{x}{p}+\frac{z}{q}=1 \qquad (1')$$

and

$$\frac{x}{x+y}+\frac{z}{y+z}=\frac{x}{p}+\frac{z}{q}=1+F, \qquad (2')$$

respectively. The new form (1′) shows that, in panmictic populations, the sum of the proportions of genes existing in homozygous conditions among their own respective total frequencies is unity. It will be seen that this is a general theorem applicable to multiple alleles. When there is inbreeding in the population, the sum of such proportions will be in excess of unity by a mupltiple of F as shown in the following. It should also be noted that the second parabola (2′) cuts the perpendiculars PQ from the first parabola to the base, ZX, so that the ratio $PP':P'Q=F:1-F$ (Fig. 1).

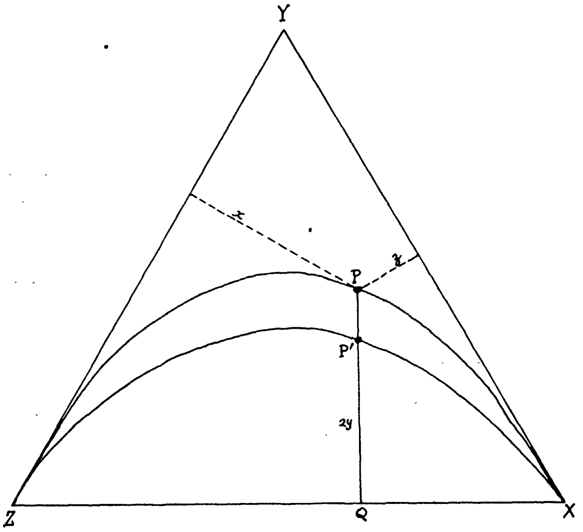


FIG. 1. The equations of the upper and lower parabolas are (1) and (2), respectively. Note that $ZQ:QX=p:q$ along the base and that $PP':P'Q=F:1-F$ along the projection PQ. (Modified from Haldane and Moshinsky, 1939.)

With multiple alleles, $A_1, A_2, \ldots, A_k$ of frequencies, $q_1, q_2, \ldots, q_k$ where $\Sigma q=1$, the zygotic proportions of a panmictic population are given by

$$[\sum_i q_iA_i]^2=\sum_i q_i^2A_iA_i+2\Sigma q_iq_jA_iA_j \ (i<j),$$

and those of a population with inbreeding coefficient F are (*3*)

$$(1-F)\ [\Sigma q_iA_i]^2+F\sum_i q_iA_iA_i$$

$$=\sum_i[(1-F)q_i^2+Fq_i]A_iA_i+2(1-F)\Sigma q_iq_jA_iA_j(i<j).$$

If we use x_{ii} to denote the proportion of homozygotes A_iA_i in the population, then the equations specifying the above two populations are, respectively,

$$\sum_i \frac{x_{\ i}}{q_i}=1 \qquad (3)$$

$$\sum_i \frac{x_{ii}}{q_i}=1+F(k-1). \qquad (4)$$

Their biological meaning is obvious. Note that q_i is equal to x_{ii} plus half the proportion of all heterozygotes carrying A_i. Letting $2x_{ij}$ be the proportion of A_iA_j, then $q_i=\sum_j x_{ij}$ (all j including $j=i$).

With k alleles, there are $K=k(k+1)/2$ genotypes. Such a population may be represented, using homogeneous coordinates as before, by a point P inside a regular simplex of $(K-1)$-dimensional space, letting the perpendiculars from P to the K-bounding flats represent the zygotic proportions x_{ii}, $2x_{ij}$, etc., whose sum is unity. Eq. (3) or (4) defines the surface within the simplex on which all equilibrium population points lie.

Eq. (4) also furnishes us a ready method of estimating F from random samples. Let a_{ii} and $2a_{ij}$ be the observed numbers of A_iA_i and A_iA_j, $n_i = \sum_j a_{ij}$ (all j) and $N = \sum_i n_i$, the total number of individuals in the sample. Our sample estimate of F is then

$$f = \frac{1}{k-1}\left\{ \sum_i \frac{a_{ii}}{n_i} - 1 \right\}.$$

When $k = 2$, this estimate reduces to

$$f = \frac{a_{11}}{a_{11} + a_{12}} + \frac{a_{22}}{a_{12} + a_{22}} - 1,$$

which is identical with

$$f = \frac{a_{11}a_{22} - a_{12}^2}{(a_{11} + a_{12})(a_{12} + a_{22})},$$

as given by Haldane (*4*) and Li (*5*). Other methods of estimating F will be discussed elsewhere.

References

1. HALDANE, J. B. S., and MOSHINSKY, P. *Ann. Eugen.*, **9**, 321 (1939).
2. WRIGHT, S. *Am. Naturalist*, **56**, 330 (1922).
3. ———. In G. L. Jepsen, E. Mayr, and G. G. Simpson (Eds.), *Genetics, Paleontology and Evolution.* Princeton, N. J.: Princeton Univ. Press (1949).
4. HALDANE, J. B. S. *J. Genet.*, **36**, 213 (1938).
5. LI, C. C. *Chinese J. Agr.*, **1**, 43 (1948).

Manuscript received August 29, 1952.

The Reaction of Diazonium Salts with Alcohols

Nathan Kornblum[1] and Alec E. Kelley

Departments of Chemistry, Purdue University, Lafayette, Indiana,* and *University College, London, England

The most remarkable thing about this "well-known" reaction is that it continues to be completely misunderstood in a great many quarters. Griess (*1*) in 1864 reported that benzene diazonium sulfate or nitrate reacts with ethanol to give benzene and acetaldehyde:

$$\text{Aryl—N}_2^+ X^- + \text{CH}_3\text{CH}_2\text{OH} \rightarrow \text{Aryl—H} + \text{CH}_3\text{CHO} + \text{N}_2 + \text{H}X.$$

This came to be regarded as a general reaction despite the fact that numerous examples of an alternative process, replacement of the $—N_2^+$ by $—OCH_2CH_3$, were soon found:

$$\text{Aryl—N}_2^+ X^- + \text{CH}_3\text{CH}_2\text{OH} \rightarrow \text{Aryl—OCH}_2\text{CH}_3 + \text{N}_2 + \text{H}X.$$

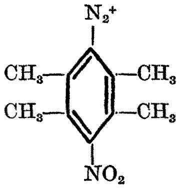
In 1887, it was shown by Remsen and Orndorff (*2*) that Griess had been in error even in the case of benzene diazonium sulfate and nitrate. These react with ethanol to give phenetole contaminated with a little benzene.

In 1901, Hantzsch and Jochem (*3*) made a very careful study of the action of absolute ethyl alcohol on benzene diazonium chloride and benzene diazonium sulfate. Confirming Remsen and Orndorff, they found that phenetole is the main product, along with a very small quantity of benzene; e.g.:

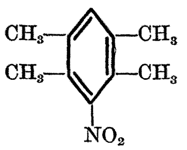
$$\text{C}_6\text{H}_5\text{—N}_2^+\,\text{Cl}^- + \text{CH}_3\text{CH}_2\text{OH} \rightarrow \underset{(61\%)}{\text{C}_6\text{H}_5\text{OCH}_2\text{CH}_3} + \underset{(5\%)}{\text{C}_6\text{H}_6}$$

When they treated benzene diazonium salts with methyl alcohol the methyl ether (anisole) was formed in 70% yield; there was no evidence of benzene formation. Thus, with *either* ethanol or methanol, benzene diazonium salts give ethers as the major product.

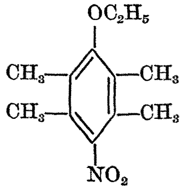
Despite the numerous demonstrations of the unreliability of ethanol as a reagent for replacing the diazonium group by hydrogen, its use persisted. Cain (*4*), in 1909, reporting on the reaction of I with ethanol, was so confident that the product was nitrodurene, II, that he did not have it analyzed. Not until 1952 was it recognized that Cain's "nitrodurene" was, in fact, the ether, III (*5, 6*).



I



II



III

(Special interest attaches to this case because nitrodurene was employed in an early test of the concept of steric inhibition of resonance; the sample used had been prepared according to Cain and was actually the ether, III.)

In recent years the view that treatment of diazonium salts with ethyl alcohol results in replacement by hydrogen has been given renewed importance by totally incorrect statements concerning the course of the reaction of benzene diazonium chloride with ethyl alcohol. Incredible as it may seem, the paper by Hantzsch and Jochem (*3*) is cited as the authority for these statements. Thus, a review (1937) reads:

> The reactions of benzene diazonium chloride with methyl and ethyl alcohols are to an appreciable extent qualitatively different, as shown by Hantzsch and Jochem:
>
> $$\text{C}_6\text{H}_5\text{N}_2\text{Cl} + \text{CH}_3\text{OH} \rightarrow \text{C}_6\text{H}_5\text{OCH}_3 + \text{N}_2 + \text{HCl}$$
> $$\text{C}_6\text{H}_5\text{N}_2\text{Cl} + \text{CH}_3\text{CH}_2\text{OH} \rightarrow \text{C}_6\text{H}_6 + \text{CH}_3\text{CHO} + \text{N}_2 + \text{HCl.}$$[2]

Numerous statements made in subsequent years are even less guarded and are exemplified by the following excerpt from a well-known monograph (1949):

> It had previously been shown by Hantzsch and Jochem that the reactions with methyl and ethyl alcohols take different paths.
>
> $$\text{C}_6\text{H}_5\text{N}_2\text{Cl} + \text{CH}_3\text{OH} \rightarrow \text{C}_6\text{H}_5\text{OCH}_3 + \text{N}_2 + \text{HCl}$$
> $$\text{C}_6\text{H}_5\text{N}_2\text{Cl} + \text{CH}_3\text{CH}_2\text{OH} \rightarrow \text{C}_6\text{H}_6 + \text{CH}_3\text{CHO} + \text{N}_2 + \text{HCl.}$$[2]

---

[1] Fellow of the John Simon Guggenheim Memorial Foundation and the United States Educational Commission in the United Kingdom, 1952–53.

[2] It would be invidious to give the references to these two quotations since they are but examples of a generalized situation.