DROSOPHILA AND THE COURSE OF RESEARCH

A. H. HERSH,
Biological Laboratory, Western Reserve University

The publication of an exhaustive bibliography[1] on the genetics of Drosophila up to the end of the year 1938 gives an opportunity to trace the course of scientific productivity in certain of its quantitative relations. The pertinent data for those aspects to which this note is limited are given in Table I, where the number of papers published each year and the number of authors publishing them are listed on the left hand side, and on the right hand side is given the number of authors credited with the number of titles indicated. The earliest reference in the bibliography is dated 1818 and for the interval from that date to 1900 there is a total of 12 titles. In 1900 the modern work on Drosophila began.[2] One paper was published in each of the years 1900, 1902 and 1903. Table I begins with 1905, the first year in which more than one paper was published. In a few years thereafter, by 1910–1911, Drosophila was already in the forefront of genetic research, mainly because of the work of Morgan and associates.

When the data for the number of papers published each year since 1910 are plotted semilogrithmically against time it is found that the points fall within a narrow band with straight and approximately parallel edges. In other words the data conform roughly to the simple exponential relation, called sometimes the compound interest law,

$$y = ae^{rt} \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

in which y is the number of papers, t the time, e the base of the system of natural logarithms, a the value of y when $t = 0$ (taken as 1910), and r the relative rate of increase. Furthermore, when y is taken as the number of authors, it is found that the data likewise conform approximately to equation (1). The numerical values of the constants for the two sets of data, determined by the method of averages, follow.

$$\text{papers, } a = 15.5;\ r = 0.106$$

$$\text{authors, } a = 9.5;\ r = 0.112$$

The result shows that the number of papers published per year increased at a compound interest rate of about 10.6 percent per year from 1910 to the end of 1938; a similar figure for the number of authors is 11.2 percent per year. This treatment neglects a certain irregular periodicity in the data. The high points of this periodicity occurred in the years 1911–14, 1929–30, and 1935–37. The first one mentioned inaugurated the work, the next is no doubt a consequence of Muller's discovery in 1927 that x-rays produce a great increase in the rate of mutation, while the last one is in large part the result of a stimulation given to cytogenetic investigation by Painter's discovery of the importance of the salivary gland chromosomes.

---

[1]Muller, H. J. 1939. Bibliography on the genetics of Drosophila, p. 132. Oliver and Boyd, Edinburgh.

[2]Davenport, C. B. 1941. The early history of research with Drosophila. Science, N. S., 93, pp. 305–306.

Since the two sets of data conform in this way to equation (1) it immediately follows that the simple power function

$$y = bx^k \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

is an approximate description of their relation to each other. In this equation y is taken as the number of papers and x as the number of authors for the years indicated; b and k are constants. The value of k with its standard error is 0.95 ± 0.03. Although this value does not differ significantly from one, it may nevertheless be the expression of a real trend, in which case it would be a rough measure of the increasing collaboration in the writing of papers, and as such is better expressed by its reciprocal, 1.05. The numerical value of b is 1.8; it is a rough measure of the extent to which authors tend to publish more than one paper per year.

TABLE I

Years	Titles	Authors	Papers	Authors
1905	3	3	1	418
1906	3	7	2	126
1907	6	5	3	69
1908	6	6	4	44
1909	4	4	5	27
1910	10	8	6	26
1911	22	15	7	13
1912	19	12	8	14
1913	27	16	9	10
1914	30	15	10	8
1915	27	18	11	5
1916	21	15	12	5
1917	40	26	13	6
1918	32	22	14	4
1919	45	19	15	7
1920	45	28	16	1
1921	48	29	17	5
1922	43	40	18	2
1923	64	45	19	3
1924	47	37	20	5
1925	62	50	21	1
1926	65	48	22	2
1927	78	60	23	2
1928	92	70	24	3
1929	134	90	25	1
1930	171	98	26	2
1931	137	108	27	2
1932	160	104	30	2
1933	134	108	35/38/41/49/	...
1934	207	136	52/54/67/	...
1935	315	201	69/89/96/	...
1936	312	200	100/102/131/	...
1937	333	231	1 author each.	...
1938	233	180	...	...

The data on the right hand side of Table I likewise conform to equation (2). When the logarithms of the number of authors are plotted against the logarithms of the number of papers to their credit in the entire bibliography then it is found that the first ten points, which include about 90% of all authors, fall on or quite close to a straight line of negative slope. When the constants b and k are determined by the method of averages for the first ten pairs of values, it is found that the value of k is —1.69 and that of b is 427. Calculation from these values shows that the curve cuts the x-axis at 36. Since this value will undoubtedly increase as the papers continue to accumulate, it seems clear that the entire curve showing the relation will reach higher and higher levels but the slope will undergo an algebraic increase, and approach —1.0 as a limiting value.

Any prediction from the exponential curves is hazardous because of the impossibility of knowing when a point of inflection may be reached. From the data in hand, however, an extrapolation to 1945 indicates that some 633 papers will be published in that year by about 480 authors. One need not be unduly pessimistic to doubt that the present diversion of the energies of the world to war will permit this high level of vigor and vitality to be maintained in such a purely peacetime enterprise as research on Drosophila.

The data show that research, an important sphere of man's social activity, in certain of its quantitative relations conforms to two functions which are among the most frequent types met with in the physical and biological sciences.