LETTER

No evidence that equations cause impeded communication among biologists

In their recent contribution to PNAS, Fawcett and Higginson (1) analyzed article equation density with respect to citation rate. The observed negative correlation is used to conclude that "equation density has a strong negative impact on citation rates," which then "presumably impedes the wider dissemination of theoretical predictions" (1). This conclusion is incorrect as it mistakes correlation for causation.

Although the correlation is found to be significant, no evidence of causality is shown. It is plausible that equations cause low citation rates via impeded communication. However, equations are used specifically to communicate with precision and conciseness. Concepts expressed easily with an equation are often unwieldy when expressed via text. Thus, it is equally plausible that, if equations disappeared en masse from scientific papers, the resultant articles would be unintelligible even to experts, and thus maintain a low citation rate.

For example, the common inequality "$P < 0.05$" expresses the phrase "...conditioned on the null hypothesis, the random variable P has a uniform distribution over the Borel field of the unit interval and is found to have a realized value, conditioned on the data, to be within the fifth lower percentile..." with great brevity and clarity.

Fawcett and Higginson (1) further assume that citation rates reflect communication level, and that said communication reflects understanding. However, consider the following two counterexamples. In bioinformatics, the work of Altschul et al. describing the BLAST tool (2) has 32,000 citations and counting. In statistics, the work of Fisher describing P values (3) has transcended citation. Both these works are models of clear, minimal equation writing, yet the conclusions of each work are routinely misunderstood. BLAST E values are commonly confused with P values, and P values are commonly confused with false-positive error rates. Thus, in neither case does citation rate correspond to communication effectiveness.

In both cases, the mechanism causing impeded communication is from the reader, not the writer. Further, given the importance of these works to bioinformatics and biology in general, we reject the statement that it is too difficult or too long-term to think of strengthening the mathematical expertise of biologists. If anything, as mathematics rapidly becomes a more central tool for biology in general, we believe it is too urgent that biologists gain mathematical sophistication, and do so quickly. For biology to advance, greater mathematical literacy is an unavoidable necessity.

In his bestselling book (4), Nobel Prize-winning physicist Richard Feynman tackles the Herculean task of explaining quantum electrodynamics to the general public. His success in doing so, in his own words, comes at the price of inefficiency. Thus, although the reader will be able to understand the discipline without using mathematics, he will forever be unable to meaningfully work within it.

Andrew D. Fernandes[1]
YouKaryote Genomics, London, ON, Canada N6K 0A1

1. Fawcett TW, Higginson AD (2012) Heavy use of equations impedes communication among biologists. *Proc Natl Acad Sci USA* 109:11735–11739.
2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410.
3. Fisher RA (1925) *Statistical Methods for Research Workers* (Oliver and Boyd, Edinburgh).
4. Feynman RP (1985) *QED: The Strange Theory of Light and Matter* (Princeton Univ Press, Princeton).

Author contributions: A.D.F. wrote the paper.

The author declares no conflict of interest.

[1]E-mail: andrew@youkaryote.com.