



Noncoding DNA, Zipf's Law, and Language

Andrzej K. Konopka; Colin Martindale

Science, New Series, Vol. 268, No. 5212. (May 12, 1995), p. 789.

Stable URL:

<http://links.jstor.org/sici?sici=0036-8075%2819950512%293%3A268%3A5212%3C789%3ANDZLAL%3E2.0.CO%3B2-O>

Science is currently published by American Association for the Advancement of Science.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/aaas.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

Noncoding DNA, Zipf's Law, and Language

Faye Flam (Research News, 25 Nov. 1994, p. 1320) reports that Eugene Stanley and his colleagues (1) have found that Zipf's law (2) applies somewhat better to noncoding than to protein-coding DNA sequences. The article implies that the statistical difference between protein-coding and noncoding DNA sequences is a surprising new discovery and that noncoding DNA resembles some sort of language.

The fact that nucleotide sequences of protein-coding regions have a different statistical structure than those of various kinds of noncoding regions (such as introns or intergenic spacers) has been well known since at least 1981 (3). In fact, many routine methods for discriminating between coding and noncoding DNA regions are based on such differences (4). It is therefore difficult to appreciate the alleged novelty of the findings of Stanley and his colleagues.

Zipf's distribution is not specific to language. Zipf himself said that it is far more general. Diverse examples of log-rank distributions that fit Zipf's law include relative sizes of cities (2, p. 416), income (2, p. 484; 5), number of species per genus (2, p. 231), and number of papers per scientist in a

given field of research (2, p. 514; 6). There is no reason to conclude that a general population is a language even if a sample drawn from this population is characterized by Zipf's distribution.

The oligonucleotide frequency distribution in noncoding DNA does not appear to fit Zipf's law any better than does the distribution in coding regions. As may be seen clearly in the figure accompanying Flam's article, both log-rank distributions are similar and both display a nonlinear, rather than a linear, trend. In both cases, only a portion of the range can be approximated by a linear function when the data are plotted on log-log coordinates. A reasonable conclusion is that both coding and noncoding regions fit Zipf's law rather poorly, if at all.

Andrzej K. Konopka

*Biolingua Research, 1415 Key Parkway,
Frederick, MD 21702, USA*

Colin Martindale

*Department of Psychology,
University of Maine,
Orono, ME 04469, USA*

References

1. R. N. Mantegna *et al.*, *Phys. Rev. Lett.* **73**, 3169 (1994).
2. G. K. Zipf, *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Boston, 1949).
3. M. J. Shulman *et al.*, *J. Theor. Biol.* **88**, 409 (1981); J. W. Fickett, *Nucleic Acids Res.* **10**, 5303 (1982); J.-M. Claverie and L. Bougueleret, *ibid.* **14**, 179 (1986); P. Salamon and A. K. Konopka, *Comput. Chem.* **16**, 117 (1992).
4. E. C. Uberbacher and R. J. Mural, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 11261 (1991); M. Borodovsky and J. McIninch, *Comput. Chem.* **17**, 123 (1993); E. E. Snyder and G. D. Stormo, *Nucleic Acids Res.* **21**, 607 (1993); S. Karlin and L. R. Cardon, *Annu. Rev. Microbiol.* **48**, 619 (1994); A. K. Konopka, *Biocomputing: Informatics and Genome Projects*, D. Smith, Ed. (Academic Press, San Diego, CA, 1994), pp. 119–174.
5. V. Pareto, *The Mind and Society* (Harcourt Brace, New York, 1935).
6. A. J. Lotka, *J. Washington Acad. Sci.* **16**, 317 (1926).

Corrections and Clarifications

In the report "Continent-ocean chemical heterogeneity in the mantle based on seismic tomography" by Alessandro M. Forte *et al.* (21 Apr., p. 386), note 14 (p. 388) should have included the following sentence at the end. "We note, however, that this classical measure of significance does not take into account the red spectrum of the observed nonhydrostatic geoid, whose harmonic coefficients cannot be properly regarded as a random distribution; therefore, the statistical significance of the measured correlation coefficient is possibly less than 99%."

Call for a free sample.

Standard Immunodetection Method

1 2 3 4 5 6 7 8 9 10

NEW Rapid Immunodetection Method

1 2 3 4 5 6 7 8 9 10

Standard and NEW Rapid Immunodetection Methods
Using BCIP as the Substrate

Eliminate the blocking step in Western blots.

The standard immunodetection method for blotted proteins can be very time-consuming. That's because conventional membranes must be blocked to prevent non-specific antibody binding. Extensive washes are also required to reduce the background for a better signal-to-noise ratio.

Cut your detection time up to 2 hours with Immobilon-P™ Transfer Membranes from Millipore. Unique membrane properties eliminate the blocking step and dramatically reduce the number and length of washes required – without compromising specificity or sensitivity.

Call or fax to request a free sample of Immobilon-P Transfer Membranes and a copy of the new rapid protocol. U.S. and Canada call Technical Services: 1-800-MILLIPORE; Japan: (03) 3474-9111. In Europe, fax: +33.88.38.91.95.

MILLIPORE

Millipore Lab Catalogue on **Internet:** access **URL** menu and type: **http://www.millipore.com**