Science
AAAS

Response to Comment by Bunge et al. on "Computational Improvements Reveal Great Bacterial Diversity and High Metal Toxicity in Soil"

Jason Gans, *et al.*

Science **313**, 918d (2006);
DOI: 10.1126/science.1126853

The following resources related to this article are available online at www.sciencemag.org (this information is current as of November 9, 2009):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/cgi/content/full/313/5789/918d

A list of selected additional articles on the Science Web sites **related to this article** can be found at:
http://www.sciencemag.org/cgi/content/full/313/5789/918d#related-content

This article **cites 3 articles**, 2 of which can be accessed for free:
http://www.sciencemag.org/cgi/content/full/313/5789/918d#otherarticles

This article appears in the following **subject collections**:
Microbiology
http://www.sciencemag.org/cgi/collection/microbio
Technical Comments
http://www.sciencemag.org/cgi/collection/tech_comment

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:
http://www.sciencemag.org/about/permissions.dtl

Downloaded from www.sciencemag.org on November 9, 2009

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2006 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.

Response to Comment by Bunge *et al.* on "Computational Improvements Reveal Great Bacterial Diversity and High Metal Toxicity in Soil"

Jason Gans, Murray Wolinsky, John Dunbar*

Bunge *et al.* claim that we underestimated the error in our analysis of bacterial diversity in noncontaminated soil. However, they used an unsatisfactory model that exhibited pathological behavior and consequently led to an exceptionally high calculated error. In contrast, the zipf distribution yielded an error estimate only 0.7 times the estimate of the total number of species (*S*), and it is more biologically relevant.

The functional form of DNA reassociation kinetics from Bunge *et al.* [equation 1 in (*1*)] is equivalent to the form presented in Gans *et al.* [equation 2 in (*2*)]. The equivalence of the two forms follows from equations S16 and S19 [SOM for (*2*)] and the change of variables $W = N/\langle N\rangle$ and $\phi(W) = \langle N\rangle P(W\langle N\rangle)$, where N and $\langle N\rangle$ are species abundance and average species abundance, respectively. One can either use the species-abundance distribution, $P(N)dN$, and compute the total number of species (S) by equation S19 (*2*), or use a "reduced distribution," $\phi(W)dW$, like Bunge *et al.* (*1*), and compute S as a free parameter. Using $P(N)dN$, the error in S is calculated using equation S19 (*2*) and standard propagation of errors (*3*). Bunge *et al.* proposed instead that $\phi(W)dW$ with S as a free parameter is more appropriate because the error can be calculated directly using nonlinear regression analysis (*4*). Both approaches are acceptable. To illustrate, we compared the computed error in S for a zipf distribution using each method. For every Cot curve, our propagation of errors calculation yielded a more conservative (i.e., larger) value than the error calculated by nonlinear regression by a factor of 1.02, 1.4, and 1.5 for the noncontaminated, low-metal, and high-metal soils, respectively, which demonstrates that our error estimates were valid.

The higher error observed by Bunge *et al.* arose from their use of an alternative model to fit the Cot curve data. In modeling bacterial species abundance from DNA reassociation data (or any other data type), the ultimate goal is to obtain a biologically relevant model (e.g., zipf) with acceptable fit and parameter errors. For comparison, one can apply more granular models that lack biological realism but are more flexible and can (ideally) provide relatively unbiased estimates of the general shape of the abundance distribution and the total number of species. The "model-free" form we described in (*2*) represents a granular model with little realism, and the three-point-mass model of Bunge *et al.* is an even more extreme example. These unrealistic granular models can potentially be useful guides but are not the desired goal.

Using the three-point-mass model, Bunge *et al.* found that estimation of S had unacceptable error. Their model provides excellent fits to the soil data and shares the same power law envelope as both the zipf and model-free distribution [equation 3 in (*2*)]. Bunge *et al.* did not test the zipf distribution, which differs from the Pareto distribution by an additional parameter that provides a variable upper bound on the range. We compared the three-point-mass and zipf distributions (Fig. 1) and agree that, for the noncontaminated soil, the three-point-mass model yields an unacceptably high standard error (SE) on S and for this reason should be rejected (*5*) as an acceptable model for estimating S. Compared with the three-point-mass model, the zipf distribution (i) has fewer parameters, (ii) provides comparable goodness of fit and, most important, (iii) has significantly smaller SE (0.7, 0.06, and 0.06 times the estimate of S for the noncontaminated, low-metal, and high-metal soils, respectively).

Bunge *et al.* also noted two experimental concerns. First, the soil bacterial DNA used by Sandaa *et al.* (*6*) may have contained eukaryotic DNA. We cannot rule out this possibility. However, the likelihood of obtaining eukaryotic contaminants in a bacterial pellet from soil depends on numerous factors (e.g., sample composition, collection depth, and researcher expertise). The authors who performed the soil bacterial DNA reassociation studies previously conducted 4′,6′-diamidino-2-phenylindole staining and microscopy, like Bunge *et al.*, to check the efficacy of the bacterial extraction method with various soil samples and did not observe contaminating eukaryotic structures (*7*). Second, Bunge *et al.* claim that measuring DNA reassociation by optical absorbance "can greatly underestimate the reassociation of repetitive sequences" and is highly inaccurate compared with measuring by hydroxyapatite binding—a claim unsupported by the authors' citations. For example, Graham *et al.* (*8*) stated, "it has long been clear that the amount of hyperchromic shift is a measure of the degree of base pairing....in

Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87501, USA.

*To whom correspondence should be addressed. E-mail: jdunbar@lanl.gov

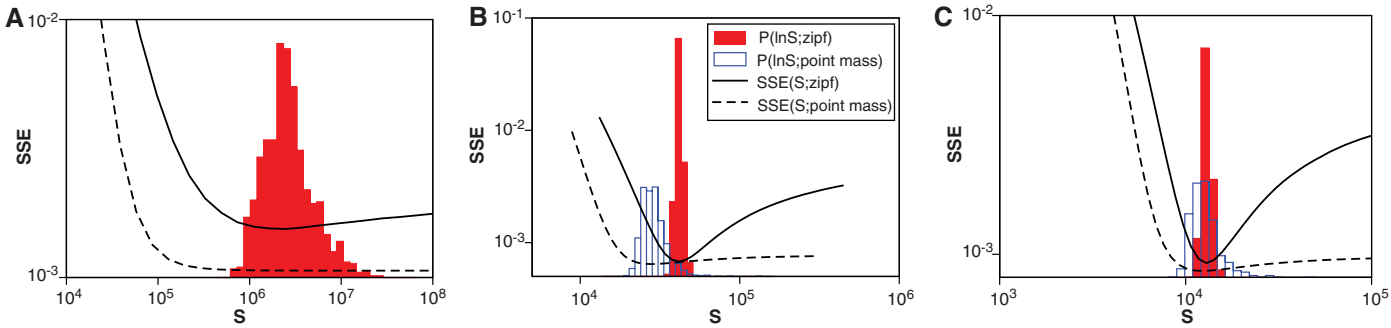

Fig. 1. Computed sum of squared errors (SSE) as a function of S superimposed on the normalized distributions of ln$S$ for both the zipf and three-point-mass distributions fit to (**A**) the noncontaminated, (**B**) the low-metal, and (**C**) the high-metal soil Cot curves. For each soil, $P(\ln S)$ is approximated as the histogram of ln$S$ computed by 10^3 Monte Carlo (MC) trials. No histogram is computed for the three-point-mass distribution fit to the noncontaminated soil, because the insensitivity of SSE to changes in S prevents convergence of the MC calculation.

Downloaded from www.sciencemag.org on November 9, 2009

fact the hyperchromicity is very nearly proportional to the fraction of nucleotides paired." Furthermore, we incorporated the possibility for heteroduplex formation in our error analysis [SOM in (*2*)]. Although we believe that the concerns of Bunge *et al*. are greatly overstated, inclusion of rigorous controls to reduce ambiguities would certainly improve future reassociation experiments.

References

1. J. Bunge, S. S. Epstein, D. G. Peterson, *Science* **313**, 918 (2006); www.sciencemag.org/cgi/content/full/313/5789/918c.
2. J. Gans, M. Wolinsky, J. Dunbar, *Science* **309**, 1387 (2005).
3. J. R. Taylor, *An Introduction to Error Analysis* (Oxford Univ. Press, Oxford, 1982).
4. W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, *Numerical Recipes in C* (Cambridge Univ. Press, Cambridge, 1991).
5. S. H. Hong, J. Bunge, S. O. Jeon, S. S. Epstein, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 117 (2006).
6. R. A. Sandaa *et al.*, *FEMS Microbiol. Ecol.* **30**, 237 (1999).
7. V. Torsvik, F. L. Daae, personal communication.
8. D. E. Graham, B. R. Neufeld, E. H. Davidson, R. J. Britten, *Cell* **1**, 127 (1974).

3 April 2006; accepted 18 July 2006
10.1126/science.1126853

Downloaded from www.sciencemag.org on November 9, 2009