

**The following resources related to this article are available online at [www.sciencemag.org](http://www.sciencemag.org) (this information is current as of December 30, 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/313/5789/918a>

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/313/5789/918a#related-content>

This article **cites 3 articles**, 2 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/313/5789/918a#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](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>

# Comment on “Computational Improvements Reveal Great Bacterial Diversity and High Metal Toxicity in Soil”

Igor Volkov,<sup>1,2\*</sup> Jayanth R. Banavar,<sup>1</sup> Amos Maritan<sup>3</sup>

Based on analysis of the reassociation kinetics of bacterial DNA in soil, Gans *et al.* (Reports, 26 August 2005, p. 1387) claimed that millions of microbe species existed in 10 grams of pristine soil and that 99.9% of the diversity was lost as a result of toxic metals. We show that the data do not support these startling conclusions unambiguously.

Gans *et al.* (1) reanalyzed the reassociation kinetics for bacterial DNA from pristine and metal-polluted soils. They claimed that a power law best described the abundance distributions and that more than one million species existed in pristine soil—an increase of two orders of magnitude compared with earlier estimates. Our analysis shows that the data analyzed in (1) constrain neither the species-abundance relationship nor the effective parameters, including the total diversity.

Consider the rate equation

$$\frac{d[C_i]}{dt} = -k\gamma[C_{0i}]^{-1+\gamma}[C_i]^{1+\gamma} \quad (1)$$

where  $[C_i]$  is the concentration of single-stranded DNA of the  $i$ -th species at time  $t$ ,  $[C_{0i}]$  is the concentration at  $t = 0$ , the coefficient  $[(1 + \gamma)/(\gamma)]$  determines the order of the reaction, and  $k$  is a measure of the reaction rate. The empirical retardation factor  $\gamma$  is equal to 1 for a second-order reaction. As a solution, one obtains the basic equation used by Gans *et al.* (1).

$$\frac{[C_i]}{[C_{0i}]} = \frac{1}{(1 + k[C_{0i}]t)^\gamma} \quad (2)$$

Following Gans *et al.*, we reconsidered data fits to equation 2 in (1) obtained for the case of a species-abundance distribution  $P(n)$ . We considered two of the models studied in (1) for the species-abundance relationship, delta and zipf, described by the following equations.

$$P(n) = \delta(n - N/S) \quad (3)$$

where  $S$  is the total number of species and  $N$  is the total population, and

$$P(n) = \frac{1 - z}{n^{2-z}[N_0^{z-1} - (N_0 + \Delta)^{z-1}]} \quad (4)$$

<sup>1</sup>Department of Physics, 104 Davey Lab and <sup>2</sup>Department of Biology, 208 Mueller Lab, Pennsylvania State University, University Park, PA 16802, USA. <sup>3</sup>Dipartimento di Fisica Galileo Galilei, Università di Padova and Istituto Nazionale per la Fisica Della Materia, via Marzolo 8, 35131 Padova, Italy.

\*To whom correspondence should be addressed. E-mail: volkov@psu.edu

where  $N_0 \leq n \leq N_0 + \Delta$  and  $N_0$ ,  $\Delta$ , and  $z$  are all free parameters. One can readily find analytic forms for the Cot equation in the two cases to be

$$\frac{[C]}{[C_0]} = \frac{1}{(1 + \frac{k}{S}[C_0]t)^\gamma} \quad (5)$$

and

$$\frac{[C]}{[C_0]} = \frac{1}{1 - \alpha^z} [{}_2F_1(z, \gamma; 1 + z; -\beta[C_0]t - \alpha^z {}_2F_1(z, \gamma; 1 + z; -\alpha\beta[C_0]t)] \quad (6)$$

respectively, where  ${}_2F_1(\cdot)$  is the hypergeometric function,  $\alpha = 1 + [(\Delta)/(N_0)]$ , and  $\beta = [(kz(1 - \alpha^{z-1})/(S(1 - z)(\alpha^z - 1))]$ .

Figure 1 shows the results of a least-squares fit of Eq. 5 (the delta model) to the data (2) analyzed in (1). For the noncontaminated sample, one obtains  $k/S = 0.033$ ,  $\gamma = 0.092$ , and  $R^2_{adj} = 0.9965$ ; for the low-contaminated sample, one obtains  $k/S = 0.034$ ,  $\gamma = 0.12$ , and  $R^2_{adj} = 0.9925$ ; and for the highly contaminated sample, one obtains  $k/S = 0.035$ ,  $\gamma = 0.16$ , and  $R^2_{adj} = 0.9963$ . Note the high quality of the fits, the notably similar values of  $k/S$ , and the somewhat low values of  $\gamma$ , the retardation

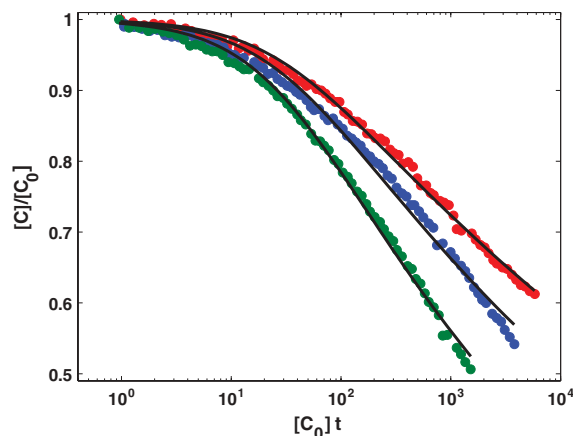


Fig. 1. Cot curves for noncontaminated (red), low-contaminated (blue), and highly contaminated (green) samples, along with the fits to the delta model of species abundance.

factor. Were one to assume that  $k = 5.19$  (as for *Escherichia coli*), one would find that  $S \approx 150$ , which is unrealistically low and might merely reflect the number of abundant species. Interestingly, as shown in (1), the fit is much worse if  $\gamma$  is chosen to be 0.45 and not allowed to be a free parameter.

The presence of a large number of species in the sample leads to the retardation factor  $\gamma$  becoming smaller than that observed for a single species (e.g., *E. coli*). Equation 2 is valid only under the assumption of a special sample in which the species are segregated from each other. However, the actual sample is a mixture of all the species, such that the rate at which two single-stranded DNA molecules of the same species reassociate when in proximity to each other should decrease. This slower reassociation rate may be captured by a decrease of  $\gamma$ , which determines the rate of change of  $[(C')]/(C)] \propto [C]^{1+\gamma}$ . It would be important to carry out careful studies of the influence of sample heterogeneity on the form of the Cot curve and on the values of the empirical parameters. Furthermore, the value of  $\gamma$  could very well depend on the degree of purity of the soil.

Figure 2 shows the fits of the data using Eq. 6 (the zipf model), with  $k = 5.19$  (the reaction rate for *E. coli*) and  $\gamma = 0.45$  (the retardation factor for *E. coli*), as carried out in (1). Interestingly, the quality of the fits does not degrade significantly when  $S$  varies over a wide range.

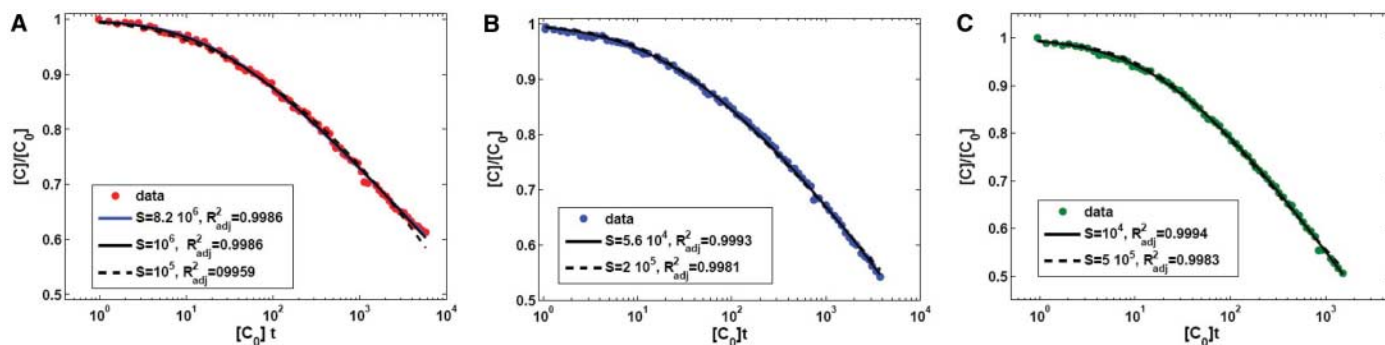
Thus, our reanalysis of the data using the framework developed in (1) suggests that the data, in and of themselves, constrain neither the species-abundance relationship nor the effective parameters, including the total diversity. This problem is exacerbated because there are no error estimates in the experimental data, and one can observe great variation in the model behavior upon changing the empirical retardation factor  $\gamma$ .

Accurately determining microbial diversity and gauging the impact of pollutants on it are extremely vital issues (3) that affect health, agriculture, and geochemical cycles. Unfortunately, the experimental data analyzed in (1) do not allow one to infer either that the diversity of pristine soil is orders of magnitude higher than previously thought or that metal pollutants have a devastating effect on microbial diversity.

## References

1. J. Gans, M. Wolinsky, J. Dunbar, *Science* **309**, 1387 (2005).
2. R. A. Sandaa *et al.*, *FEMS Microbiol. Ecol.* **30**, 237 (1999).
3. T. P. Curtis, W. T. Sloan, *Science* **309**, 1331 (2005).

11 October 2005; accepted 30 March 2006  
10.1126/science.1121225



**Fig. 2.** Cot curves for (A) noncontaminated (red), (B) low-contaminated (blue), and (C) highly contaminated (green) samples, along with the fits to the zipf model of species abundance. In each case, the data are fitted to several versions of the model with varying numbers of species  $S$  as shown in the inset. The blue solid lines represent the best fit in accord with analysis in (1). The

other lines are fits in which the number of species is decreased by factors of 8.2 and 82 for the pristine soil and increased by factors of 3.56 and 50 for the two polluted cases, with little degradation in the quality of fit. The most sensitive region of the Cot curve occurs when  $[C]/[C_0]$  becomes small compared with 1, and the experimental data does not extend to this regime.