

Unveiling prokaryotic diversity

Sean Nee

School of Biological Sciences, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK

Microscopic life is overwhelmingly important, both quantitatively and qualitatively, but has, until recently, been largely invisible. Molecular technology is now revealing this unseen world to us. A combination of novel theory and empirical research by Tom Curtis and his colleagues provides us with new estimates of prokaryotic diversity and reveals that this can range over four orders of magnitude among habitats.

The natural history of the microscopic world is the most exciting and explosive growth area in the study of biodiversity [1–5]. But one of the many difficulties impeding the birth of a true biodiversity science is that expertise is needed from vastly different research areas. Microbiology provides many examples of this fact: it is estimated, for example, that bacteria in ocean sediments comprise at least 10% of the living carbon on the planet, suggesting that they should be of great interest for biodiversity scientists. However, the startling facts and theories about them are discussed at geology conferences, rather than biological ones (e.g. [6]).

Collaboration is the obvious short-term solution to this problem, and a landmark paper has recently appeared written by a microbiologist, a mathematician and, perhaps oddly, a psychologist [7]. The paper provides estimates of prokaryotic diversity (species number) in the sea, in soil and in sewage. Yes, sewage: we will recognize that biodiversity science is mature when researchers start choosing study sites that are not also holiday destinations.

To estimate prokaryotic diversity, the authors have made some large leaps. First, they have devised a remarkably simple way to estimate the total diversity of a community from a sample of individuals: I believe that this has been missed by previous researchers in the area. This estimation procedure is applicable to both eukaryotes and macroscopic organisms. Second, they completely side-step the problem of what a ‘species’ is. Essentially, it can be whatever you want it to be, in terms of molecular distance, and your estimates of total diversity will adjust themselves accordingly with your definition. Although traditionalists might object to this complete disregard for long-cherished philosophical obsessions concerning the nature of a ‘species’, the description of biodiversity in purely objective, molecular terms is here to stay.

Curtis *et al.* base their method for estimating diversity on the fact that, for large communities of macroscopic organisms, the distribution of the logarithm of species abundances is commonly observed to be bell shaped –

lognormal. The authors make the reasonable assumption that this is also true for prokaryotes. Mathematically, lognormal distributions are usually characterized by their mean and variance, but for the purposes of readily connecting them to estimation problems, it is better to characterize them by using N_{\min} and N_{\max} , the abundances of the rarest and most common species in the community, respectively. If we can estimate N_{tot} , the total number of individuals of all species in a community, and also estimate N_{\max}/N_{tot} , then we can obtain an estimate of the total number of species in the community (Fig. 1). In fact, N_{tot} is readily measured for microbial species and, for macroscopic species, can usually be ‘guesstimated’ quite well. And N_{\max}/N_{tot} can be estimated from a sample of individuals because, apart from sampling variation, it is independent of sample size (except, obviously, for very small samples).

For prokaryotic diversity, the estimate of N_{\max}/N_{tot} currently comes from clone libraries of the gene encoding 16S RNA: the estimate is simply the most abundant sequence:the total number of sequences ratio. Techniques,

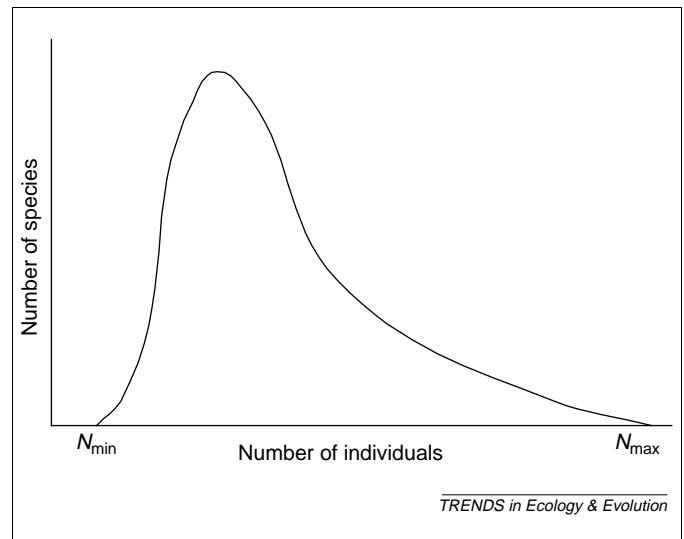


Fig. 1. Estimating diversity. Frequency distribution of species abundances: notice that we are looking at raw abundances, not their logarithms. The total number of individuals of all species, N_{tot} , is given by the integral under the curve. By assuming that the curve is lognormal and that exactly one species has the smallest population size, N_{\min} , and one species has the largest population size, N_{\max} , the shape of the curve has to adjust itself, constrained by its functional form, to accommodate N_{tot} . One can then retrieve the desired information about the total number of species. The formula relating the observed (or estimated) quantities, N_{\max}/N_{tot} and N_{tot} , to total species number is complicated, but readily solved numerically. Curtis and colleagues [7] assume that $N_{\min} = 1$: other assumptions can obviously be made. In any case, for the purposes of rank ordering diversity in different habitats, this assumption does not matter. Unless we delude ourselves as to the precision of the estimates that we will get using this, or any other, technique, intuition suggests that this estimate is robust to relaxations of the assumption of lognormality.

such as fluorescence microscopy, can be used to count the number of prokaryotes in a sample, providing us with N_{tot} . On the basis of these numbers, the authors make the following diversity estimates: oceans, 160 'species' ml^{-1} ; soil, 6400–38 000 g^{-1} ; sewage works, 70 ml^{-1} . This last number is a nice example of the paradox of enrichment, the phenomenon whereby fertilization (in the agricultural sense) reduces diversity. The low diversity estimated for the oceans and the high diversity for soils are consistent with results from studies using very different methodologies [8]. Understanding this difference is as important as understanding the latitudinal diversity gradients observed for macroscopic organisms. Also, the low diversity associated with high nutrient levels has also been observed in prokaryotic communities in sediments beneath fish farms [8].

My previous remark about the statistics of $N_{\text{max}}/N_{\text{tot}}$ in small samples requires elaboration. Suppose that, in the community, $N_{\text{max}}/N_{\text{tot}}$ is actually quite small. Then it is possible that our sample will not actually contain the most abundant species in the community. If we know the identity of this species, then we can restrict our analysis to samples that contain it, although this would require a modification of the estimation formula to accommodate this extra level of sampling. But if we have no means of restricting our analysis to such samples, then there appears to be no way to estimate diversity by this method, although this could simply be a failure of imagination on my part.

Finally, this new paper raises very starkly an extremely interesting question: just what, precisely, is the community being sampled? What does it mean to say that there are 70 species ml^{-1} in sewage works? It could mean anything from 'there are 70 species in sewage works, period' or 'there are billions of species in sewage works'.

Turning to macroscopic organisms, if we used this technique to estimate, say, tree diversity then we would need an understanding of β diversity (how diversity changes through space) to go beyond statements about numbers of trees ha^{-1} to numbers of trees in the world. In fact, it is entirely possible that the oceans of the world constitute a single prokaryotic community, just as aquatic microbial eukaryotes appear to be ubiquitous as a consequence of the abundance and dispersal abilities of the organisms concerned [9]. But it is not obvious that there is no β diversity in soils [8], and this needs further investigation.

References

- 1 Floyd, R. *et al.* (2002) Molecular barcodes for soil nematode identification. *Mol. Ecol. Evol.* 11, 839–850
- 2 Lopez-Garcia, P. *et al.* (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* 409, 603–607
- 3 Moon-van der Staay, S.V. *et al.* (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* 409, 607–610
- 4 Zettler, L.A.A. *et al.* (2002) Eukaryotic diversity in Spain's River of Fire – this ancient and hostile ecosystem hosts a surprising variety of microbial organisms. *Nature* 417, 137–137
- 5 Moreira, D. and Lopez-Garcia, P. (2002) The molecular ecology of microbial eukaryotes unveils a hidden world. *Trends Microbiol.* 10, 31–38
- 6 Kerr, R.A. (2001) Life – potential, slow, or long dead? *Science* 294, 1820–1821
- 7 Curtis, T.P. *et al.* (2002) Estimating prokaryotic diversity and its limits. *Proc. Natl Acad. Sci. U. S. A.* 99, 10494–10499
- 8 Torsvik, V. *et al.* (2002) Prokaryotic diversity – magnitude, dynamics and controlling factors. *Science* 296, 1064–1066
- 9 Finlay, B.J. (2002) Global dispersal of free-living microbial eukaryote species. *Science* 296, 1061–1063

0169-5347/02/\$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved.
PII: S0169-5347(02)00043-5

Letters

Shortcuts in systematics? A commentary on DNA-based taxonomy

Ole Seberg¹, Chris J. Humphries², Sandy Knapp², Dennis Wm. Stevenson³,
Gitte Petersen¹, Nikolaj Scharff⁴ and Nils Møller Andersen⁴

¹Department of Evolutionary Botany, Botanical Institute, University of Copenhagen, Gothersgade 140, DK-1123 Copenhagen K, Denmark

²The Natural History Museum, Cromwell Road, London SW7 5BD, UK

³The New York Botanical Garden, Bronx, New York, NY 10458, USA

⁴Entomology Department, Zoological Museum, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark

The primary aims of taxonomy are to name, circumscribe, describe and classify species. The first goal is convention but the remainder are science. The International Codes of Nomenclature are legislative instruments and nomenclature is simply a mechanism to ensure that a species

name is legitimately attached to a type specimen, regardless of scientific status. The type of a species does not serve, as Tautz *et al.* ([1,2], but see [3]) assert, as 'the central reference for comparisons'. The crucial link between names and scientific investigation is species circumscription followed by description. The Codes require Linnaean binomials: a genus name and a species epithet.

Corresponding author: Ole Seberg (oles@bot.ku.dk).