The searchlight and the bucket of microbial ecology

Nicole Dubilier, Max Planck Institute for Marine Microbiology, Bremen, Germany

In a crystal ball article scientists are asked to 'articulate their personal visions on the new conceptual, technical, and theoretical developments that will drive the most exciting progress over the next few years'. While I could easily spin some ideas off the top of my head about future developments in a conversation, particularly after a few beers, I find it quite intimidating to have to put my thoughts in writing: (i) because supposedly crystal balls are widely read (at least according to the editors of *Environmental Microbiology*), meaning I have a large audience for making a fool of myself; (ii) anyone can read this article in a couple of years from now meaning I have a large audience for making a fool of myself in the future, when none of what I have written has actually come true; and (iii) I have not had a few beers (yet). However, I cannot afford to not write this article as I have agreed with one of the editors of *Environmental Microbiology* that I owe him a bottle of Romanée-Conti burgundy if I do not submit this piece by tomorrow. I foolishly made this agreement before I googled the term 'Romanée-Conti' and for those of you that are as clueless about good burgundies as I am, here's just one quote: 'With prices that start at £80 a bottle . . . , and rise with eye-watering increments to £700 a bottle . . . Romanée-Conti . . . is the sort of stuff that precious few can afford'.

One of my favourite pet (bug) causes is diversity research. While botanists and zoologists have had several hundred years to figure out 'Who is out there?', microbiologists wasted a bit of time trying to answer this question using microscopes and agar plates. It was only 30 years ago that Carl Woese brought order to the field of microbial taxonomy by establishing the comparative rRNA sequence approach and microbial diversity research only really took off 20 years ago (Pace, 1997), when automated sequencing methods enabled the analysis of larger data sets.

Twenty years is not a lot of time and we are still very far from answering some very basic questions such as the spatial and temporal distribution patterns of microorganisms, their biogeography and their functional biodiversity. The basis for answering these questions is knowing which organisms are present in our organism, community, or habitat of interest. Yet even in low-diversity ecosystems, the true diversity of microorganisms is often underrated. Just one example for the inherent difficulties we have in estimating microbial diversity: despite extensive 16S rRNA analyses of the extremely low-diversity communities at an acid mine drainage site, a novel lineage of archaea that is ubiquitous at this site remained undiscovered until, fortuitously, random shotgut sequencing recovered a genome fragment with a 16S rRNA gene from one of the archaeal strains. Fittingly named WTF, these archaea have several mismatches in their 16S rRNA sequences to the commonly used broad-specificity primers and were therefore not present in the 16S rRNA libraries (Baker *et al.*, 2006).

If we face these kinds of difficulties in low-diversity ecosystems, what are we missing in high-diversity environments such as oceanic surface waters, marine sediments, or terrestrial soils? Several authors have emphasized how notoriously incomplete most PCR libraries are for describing microbial diversity on the basis of only a few hundred 16S rRNA amplicons at best (Acinas *et al.*, 2004; Kemp and Aller, 2004; Hong *et al.*, 2006). And yet papers are still being published in which at most several hundred 16S rRNA amplicons were analysed and based on this limited data set, often without statistical analyses or quantitative methods such as fluorescence *in situ* hybridization to confirm the sequence data, conclusions about microbial distribution patterns or biogeography are drawn. And as a final flog to a horse I hope I haven't killed yet, we have still not clearly defined what a species is in microbiology and cannot be sure that 16S rRNA gene analyses reveal sufficient information about the functional diversity of microorganisms.

One of the main limiting factors in diversity research has been the time and money involved in sequencing the 16S rRNA gene using the classical Sanger dideoxy chain termination technique. Now there is light at the end of the tunnel or better luciferase. A new sequencing technique called pyrosequencing (a method in which light produced in a cascade of enzymatic reactions ending with luciferase is proportional to the number of incorporated nucleotides) is causing considerable excitement, not only among microbiologists. The major drawback of this technique is that, currently, read lengths are extremely short, at best 200 nucleotides. Despite this limitation, Mitch Sogin has been a key driver in applying this technique to microbial diversity and the results that he and his coauthors have from the analysis of 118 000 (!) 16S rRNA gene amplicons are fascinating (Sogin *et al.*, 2006). Using a hypervariable region in the 16S rRNA gene as a genetic marker, Sogin and colleagues (2006) showed that the microbial diversity of marine communities is at least one to two orders of magnitude higher than previously assumed. More interesting than this sheer increase in numbers, is the kind of organisms Sogin and colleagues (2006) found, namely the thousands of low-abundance populations that account for most of the observed phylogenetic diversity. These organisms were not previously found because most microbial communities are dominated by a relatively small number of high abundance populations. When only a limited number of amplicons are sequenced as in most previous studies, it is the high abundance populations, the

© 2007 The Authors
Journal compilation © 2007 Society for Applied Microbiology and Blackwell Publishing Ltd,

members of the 'terra frequentata' that are most commonly found, while the members of the 'terra incognita' go unrecognized (Curtis and Sloan, 2005). As a good study should, the Sogin and colleagues (2006) paper immediately raises a whole gaggle of further questions: do these 'rare' species represent ancient lineages, so-called living fossils of the microbial world? What are they doing? Do they play a role in major biogeochemical processes? How does the abundance of these species change over time? While I am not sure that I agree with all of the hypotheses that Sogin and colleagues (2006) present to answer these questions, it is exciting to imagine that pyrosequencing may soon enable the kind of comprehensive and exhaustive sequencing analyses needed to shed more light on fundamental questions in microbial ecology.

In the past, research in microbiology has often been described as being 'method-limited' with progress dependent on the tools and methodology needed to study small microorganisms. More recently, however, the tables have turned and new methods in sequencing, genomics, proteomics, bioinformatics and imaging are producing vast amounts of information at an ever increasing speed. While some scientists fear that these techniques will lead to a dominance of 'data-driven discoveries', my crystal ball shows a golden era taking shape. What lies ahead of us is a truly challenging period in which we can hunt through the buckets of information we are amassing using our hypothesis-driven searchlights and focusing them on the key questions we have in microbial ecology. These are, and perhaps always will be: 'Who is there?' and 'What are they doing?'

References

Acinas, S.G., Klepac-Ceraj, V., Hunt, D.E., Pharino, C., Ceraj, I., Distel, D.L., and Polz, M.F. (2004) Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* **430:** 551–554.

Baker, B.J., Tyson, G.W., Webb, R., Hugenholtz, P., Allen, E.E., and Banfield, J.F. (2006) Community sequencing of novel ultra-small Euryarchaea from an extremely acid subsurface mine, Iron Mountain. Joint Genome Institute User Meeting, Walnut Creek, California.

Curtis, T.P., and Sloan, W.T. (2005) Exploring microbial diversity – a vast below. *Science* **309:** 1331–1333.

Hong, S.-H., Bunge, J., Jeon, S.-O., and Epstein, S.S. (2006) Predicting microbial species richness. *Proc Natl Acad Sci USA* **103:** 117–122.

Kemp, P.F., and Aller, J.Y. (2004) Estimating prokaryotic diversity: when are 16S rDNA libraries large enough? *Limnol Oceanogr Methods* **2:** 114–125.

Pace, N.R. (1997) A molecular view of microbial biodiversity and the biosphere. *Science* **276:** 734–740.

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., *et al.* (2006) Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci USA* **103:** 12115–12120.

The human microbiome: eliminating the biomedical/environmental dichotomy in microbial ecology

Ruth E. Ley, Center for Genome Sciences, Washington University, St Louis, MO, USA
Rob Knight, Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO, USA
Jeffrey I. Gordon, Center for Genome Sciences, Washington University, St Louis, MO, USA

When a new human being emerges from its mother, a new island pops up in microbial space. Although a human lifespan is a blink in evolutionary time, the human island chain has existed for several million years, and our ancestors stretch back over the millennia in a continuous archipelago. Microbes thrive on us: we provide wonderfully rich and varied habitats, from our UV-exposed, oxic and desiccating skin to our dark, wet, anoxic and energy-rich gut that serves as a home to the vast majority of our 100 trillion microbial (bacterial and archaeal) partners. A sobering or inspiring fact: we contain 10 times more microbial than human cells and an estimated 100 times more microbial genes. How our association with microbes has evolved, the forces that shape it, what about it might be uniquely 'human', how changes in our biosphere are affecting it, and how it impacts our health, all are challenging questions for the future because they require a level of engineering and computational sophistication that is still emerging. Our crystal ball sees the epidemiologist of the future describing how changes in kilometre-scale macro-ecosystems affect micrometer-scale microbial ecosystems associated with populations of meter-scale human beings, on time scales of an infection, a human lifespan, or the rise and fall of a society.

Step forward into the world of metagenomics, and we start to see ourselves as supra-organisms whose genome evolved with associated microbial genomes (the microbiome). Although the primate-lineage component of the human genome is decoded, sequencing of the microbiome is just beginning. Our first glance of the microbiota, from recent extensive surveys of its organismal lineages (based on 16S rRNA), and initial DNA-based metagenomic analyses of its microbiome, raises a long list of basic questions: is there a core microbiome shared between humans and passed along as a family heirloom? How does the microbiome of humans differ from those of other animals? How does it change with ageing, travel, marriage, sickness? Is its composition an unappreciated determinant of our well-being, and/or a contributing factor to diseases such as obesity?

We humans have an extraordinary impact on the environment: although we comprise about 0.5% of the total heterotroph biomass on earth, we consume 14–26% of terrestrial net primary production (70% in some regions of south-central Asia) (Imhoff *et al*., 2004). The remarkable

© 2007 The Authors