Environmental Microbiology (2007) **9**(1), 1–11 doi:10.1111/j.1462-2920.2006.01222.x

Crystal ball – 2007

In this feature, leading researchers in the field of environmental microbiology speculate on the technical and conceptual developments that will drive innovative research and open new vistas over the next few years.

Theory and the microbial world

Tom Curtis, School of Civil Engineering and Geosciences, University of Newcastle upon Tyne, UK

The microbial world is not really a good expression. It is not a world, it's not a planet, or a constellation. It's not even a universe, for there are 1 000 000 000 times more bacteria in the world than stars in the sky. Imagine the microbial 'world' as billion universes each made of thousands or millions of galaxies and you have some idea of the scale of the challenge of microbial ecology.

Remarkably new galaxies arise constantly. Every time a child is born, a seedling germinates, or a leaf falls, microbes find their place, grow and reproduce with clockwork like reliability, actually a lot better than clockwork. No one would seek to understand the universe star by star. However, it sometimes seems that microbiologists are trying to understand the microbial world one cell at a time. Of course every cell and (each star) is fascinating in its own right and there is tremendous satisfaction to be gained from their study. But these simple pleasures will not be enough. Microbiology is going to develop a body of theory. By theory I mean a consilient (Wilson, 1998) and calibrated set of rules to describe and predict the behaviour of the microbial world as a system.

Why theory? Most microbial systems are large and complex and operate at scales that are difficult to observe and, almost certainly, defy intuition alone, which is why we need to seek and quantitatively express those rules. However, our current *modus operandi* is one of description with ever more sophisticated tools. The tools are telling us that the microbial world is very complex. Though we talk of opening the black box we find we are peeling the black onion. For each innovation reveals more wonders. One has to question whether we as a community can really carry on being surprised by this. We really need to sit down and take stock. There is simply no reason to assume that we can gain an adequate understanding by the unguided opportunistic ingenuity we employ today. We can use these tools to test those putative rules and we can use those putative rules to determine what those tools need to be able to do. We also need theory because the microbial world is of profound practical importance. Agronomists, medical practitioners and engineers want to exploit microbes to cure, clean and grow food. We currently do this empirically, but empiricism is subject to the law of diminishing returns and the insights empiricism delivers are often qualitative and situation bound. Practitioners need numbers. It is one thing to infer the action of gravity and determine that you would need a big rocket to escape the Earth's pull. It is another to determine the escape velocity and so rationally design a spacecraft. The link with practice is important. Microbial ecology is going to become 'big science' and need big budgets. If theory can effectively link the basic science with the economy and national it will us to justify those big budgets. If we do not make that link, I suggest that the money might stop. Clearly political and scientific drivers make the development of theory imperative: it has to happen.

How theory? The complex nature of the microbial world may make theoretical descriptions seem impossible. But they are not. For progress will be made using simple ideas, refined iteratively and grounded in truth. Simple truths about a system will describe some part of it. A calibrated model describing that part is a foundation that we can build on and form at least a partial basis for prediction. This may sound like motherhood and apple pie: it ain't. Truth in this context means parameters and verification. These are hard slog and could require all the ingenuity that modern microbial ecology can muster. But it will be worth it. If we do it we will intellectually outstrip much contemporary classical theoretical ecology. This field is hidebound by the difficulty of experimentation and is therefore contaminated by self-congratulatory mathematical castles in the air with invented parameters and little verification. More importantly calibrated theory will open the door to a new age in microbial ecology as we stop merely gawping at the wonder of it all, like pre-renaissance peasants on a star lit night, and start to begin to truly understand.

Reference

Wilson, E.O. (1998) *Consilience: the Unity of Knowledge*. New York: Random House.

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

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

diversity of food sources available to us compared with other species, and our global distribution might be reflected in an exceptionally diverse gut microbiota compared with other species if diet is the primary factor driving diversity in this microbiota and its microbiome. On the other hand, our low levels of (primate lineage) genetic diversity relative to other mammal species (Li and Sadler, 1991; Kaessmann *et al*., 2001) might suggest that our gut microbiota is relatively impoverished if host genetics (perhaps mediated through the immune system) is the primary determinant of microbial diversity.

To answer these questions about human microbial ecology and its variation requires an integration of data about a microbiota's collective genome, transcriptome and metabolome, the physical and chemical attributes of the host's surface habitats occupied by the sampled microbial community, plus information about the genotype, systems physiology, lifestyle, and living environment of the humans being studied. Initial studies that use monozygotic twin pairs and their mothers will help limit some confounding variables.

Comparisons of the microbiota and microbiome in different groups of humans undoubtedly will be multilayered, making use of data sets of lineage assemblages, gene assemblages and populations of genome types. Integrating these diverse and complex data sets will spark development of advanced computational methods. Our crystal ball shows that the key challenge will be in defining distances: e.g. how far apart are the microbiomes, the transcriptomes, and the metabolomes, of gut microbial communities represented in two stool samples from unrelated, related or the same individual(s)? How can we relate changes in these parameters to changes on different time scales, such as day-to-day diet, health and disease, and the evolutionary ecology of different species? Once these distances are defined, we can use established statistical techniques to place all the data in a single, unified framework that allows us to ask which differences correlate with one another and with human health. A key innovation will be to extend the phylogenetic distance metrics developed for comparing microbial communities using 16S rRNA genes to allow phylogenetic classification of metagenomic samples. The concept of distance is so fundamental a unifying principle that the mantra among students may become 'distance is to microbial ecology as energy is to physics!'. New, accurate methods of assessing lateral gene transfer that combine information about both the composition and phylogenetic history of each gene, will revolutionize our picture of microbial adaptation to the diversity of ecological niches humans provide.

One major question we will soon be able to answer is whether there is a threshold beyond which phylogenetic relatedness is irrelevant to predicting microbial function: just as pigeons and penguins are both birds but occupy very different niches, we may find it is not correct to assume that relatedness is our best guide to bacterial functions in the human microbiome. Understanding whether rare or abundant microbes play the greatest role in determining the function of the human microbiota and its microbiome, in particular by relating abundance and gene expression in these species to human health, will revolutionize our understanding of the $\geq$99% of the genes associated with our body that are carried by microbial genomes.

Once these methods are in hand, the epidemiologist of the future can collect metadata to correlate with variability between his subject's microbiomes. The results of epidemiological studies will be translated into therapies. Our medical insurance cards will contain one chip for our primate genome, and one for our microbiome. As part of the annual physical exam, physicians will take a stool sample to update the microbiome profile. Just as today a rise in blood pressure from one visit to the next signals a risk of developing heart disease, tomorrow changes in the microbiome profile will herald a predisposition to diseases such as obesity. Therapeutic intervention will follow, likely a combination of individualized nutrition, deliberate 're-programming' of the microbiota with addition/removal or stimulation of particular lineages or genetic complements within the microbiome, or use of microbial gene products themselves (or their revealed human gene product targets) as part of our 21st century pharmacopoeia.

As our human population increases in size, and as globalization promotes movement of people around the world, and exposure to one another and to new environments (including those changed by our anthropogenic perturbations), our microbiomes proliferate and mix between populations at rates unprecedented in human evolution. The impact of global change on our human microbial ecology (our 'microevolution') is not currently known but we are surely not detached from it. It behoves us to establish human microbial observatories just as we have established long-term environmental microbial observatories. It is time to breach the institutionalized dichotomy between environmental science and biomedical research, and to study ourselves as an integral and dependent part of our microbe-dominated world.

References

Imhoff, M.L., Bounoua, L., Ricketts, T., Loucks, C., Harriss, R., and Lawrence, W.T. (2004) Global patterns in human consumption of net primary production. *Nature* **429:** 870–873.

Kaessmann, H., Wiebe, V., Weiss, G., and Paabo, S. (2001) Great ape DNA sequences reveal a reduced diversity and an expansion in humans. *Nat Genet* **27:** 155–156.

Li, W.H., and Sadler, L.A. (1991) Low nucleotide diversity in man. *Genetics* **129:** 513–523.

Riding giants

Philip Hugenholtz, DOE Joint Genome Institute, Walnut Creek, CA, USA

We've all seen the graphs showing the exponential rise in sequence data in the public databases. This of course is the result of great strides forward in high-throughput sequencing and, incredibly, new technologies are on the verge of putting us into sequence overdrive. A traditional Sanger sequencer produces per run just ~70 kbp, whereas a 454 pyrosequencer produces 30 Mbp, and Solexa is promising up to 1 Gbp. And other technologies are in the works. So it doesn't take a crystal ball to see that the bottleneck will rapidly become computational. In fact, Darren Platt, head of informatics at JGI, only half jokingly predicts that just storage of the data alone will become limiting and that the storage medium of choice in the future may be the DNA itself, i.e. it will be cheaper to resequence the DNA than store the information electronically. Hopefully things won't get to that extreme and we will step up to the computational grand challenge.

What better target to aim this elephant gun at than microorganisms. After all, they constitute the bulk of the biomass and evolutionary and metabolic diversity on the planet. Ironically, we may run out of characterized microbial isolates to sequence in the not too distant future. Plans are afoot to sequence all ~6000 described species which should take ~200 Gbp; this amounts to 5 years of dedicated work at a production facility like JGI based on current capacity, but likely much less time as the new technologies come on line (200 Solexa runs?).[1] The natural microbial world on the other hand represents a limitless source of sequencing targets, with the added benefit of no cultivation bias. Initial forays into microbial community sequencing (metagenomics) have been very promising, but reinforce our suspicions that we have barely scratched the surface of the microbial world. Indeed, if a recent survey of the deep sea using 16S pyrotags (Sogin *et al*., 2006) is anything to go by, we have barely brushed the surface. And this goes 10-fold for the virosphere.

What will environmental -omics look like in the future? Here are a few predictions: snapshot samples will be replaced by time series and fine-scale spatial sampling, and via this, we will no longer have to try to understand the plot of the film by looking at the corner of one frame, we will see the film in motion. Collection of sequence-associated data, or metadata, will become more standardized and detailed facilitating meaningful correlations between communities and their ecological settings. Viral, bacterial, archaeal and eukaryotic fractions will be routinely sampled and sequenced in parallel so that different trophic levels can be analysed in conjunction. Expressed mRNAs and proteins will be routinely obtained and analysed from the same environmental samples to get a window on community function instead of just metabolic potential. Fractionating individual populations and cells from the community for independent sequencing will become commonplace and greatly facilitate dissection and interpretation of the community data. But moreover, population genomics will mature in its own right and sampling (sequence coverage) of naturally occurring populations will go much, much deeper bringing the evolutionary processes that drive and shape populations into sharp focus. Population geneticists will be drawn to the field in droves. Also, the structure and dynamics of microbial populations will be placed convincingly into their many and varied ecological contexts. The whole process will be much faster, and the data made publicly available much sooner in a fully integrated format.

In order to make sense of these massive data sets, modelling will assume a central role in microbial ecology. As a result, it will transition from a mainly qualitative descriptive discipline to a quantitative predictive one. However, I think that ecosystem predictability will be noisy, more like predicting the stock market than gravitational orbits, and general principles will be hard fought and won. As Tom Curtis is likely to point out in these pages, we would do well to use macro-ecological theory as a guide in this endeavour.

The wild card in the deck is synthetic biology. The idea of treating cells as chasses, genes as parts, and designing your own organism to specification is alien to most ecologists, particularly if the designer organism is assembled from parts that are partitioned as discrete functional units in a community. But I believe both fields stand to benefit enormously from one another. By reinventing life, synthetic biology will fail repeatedly, but in so doing will accelerate our understanding of metabolism and regulation, and improve ecological modelling efforts. And by observing nature in detail at the molecular level, microbial ecology will provide many design constraints to synthetic biologists that they will not have to uncover by trial and error.

[1]The real bottleneck here will be the ability to obtain good quality genomic DNA from these isolates, and to close the draft genomes.

Acknowledgements

I thank Héctor García Martín, Victor Kunin, Gene Tyson and Trina McMahon for feedback on my naval gazing.

Reference

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 future of single-cell environmental microbiology

Marcel M. M. Kuypers and Bo Barker Jørgensen, Max Planck Institute for Marine Microbiology, Bremen, Germany

One of the key objectives in environmental microbiology is to couple identity and function of microorganisms in soil, water, sediment or other ecosystems. The field has come a long way over the past decade, both with respect to 'who is there?' and to 'what are they doing?' The coupling between identification and activity, however, remains the weakest point. Existing approaches need to be further developed and combined. It is a dream to one day do experiments with prokaryotes the way that they are done with higher animals or plants, at the level of the individual organism.

There are indeed already methods in use that enable this to a certain extent. These may be based on pulse-chase experiments during which the microorganisms have been fed a radioactive or an isotopically heavy meal. Those cells that have taken up the substrate may subsequently be identified, e.g. by a combination of microautoradiography and fluorescence *in situ* hybridization (MAR-FISH). MAR-FISH has the advantage that the active substrate uptake can be related to cells. This is not the case with most methods that combine stable isotope tracers and the analysis of DNA, RNA or biomarkers. The fact that MAR uses radioactivity, however, limits its use to those elements that have a radioisotope with a suitable half-life and excludes the study of other elements such as nitrogen.

It was therefore a methodological breakthrough when Orphan and coworkers (Orphan *et al*., 2001) combined FISH with secondary ion mass spectrometry (SIMS). The aim was to show whether those aggregates of archaea and sulfate reducing bacteria, which Boetius and coworkers had discovered 1 year before in sediments where anaerobic oxidation of methane predominated, were indeed living on methane. By analysing FISH-stained aggregates with SIMS it was shown that their highly ^{13}C-depleted cell carbon carried the distinct isotopic signal of methane. It was a limitation of this FISH-SIMS method, however, that the size (~10–15 μm) of the ion beam used to sputter biomass and generate secondary ions to be analysed by mass spectrometry exceeded the average diameter of a microbial cell (~1 μm). This problem has recently been solved, however, by the development of the so-called nanoSIMS, which for the first time makes it possible to determine the chemical, radioisotopic or stable isotopic composition of biomass at the submicron level (Lechene *et al*., 2006; McMahon *et al*., 2006). This technique in combination with pulse-chase experiments with radioactive or stable isotope labelled substrates opens up a vast new area of research possibilities in single-cell environmental microbiology.

The nanoSIMS has a beam size of ~50 nm which is sufficiently small to analyse individual cells or even parts of a cell. It is only mildly destructive as it removes only the upper one to three atomic layers (~1 nm) to obtain sufficient sputtered (vaporized) biomass for elemental and isotopic analysis. On the other hand, it is also possible to sputter the entire cell when whole cell isotopic analysis is needed. In a pulse-chase experiment using a stable isotope labelled substrate the individual microbial cells that have assimilated the substrate in an environmental sample can be identified from their isotopic enrichment after 24 h of incubation, assuming a growth rate of just one doubling per week and a typical labelling percentage of the substrate of *c*. 50%. Some of the first such experiments in a $^{15}N_2$-fixing bacterial culture have shown large intercellular and intracellular differences in the degree of ^{15}N labelling. Thus, even the cells in a laboratory culture are not all alike but consist of billions of individuals, each with their different activities and life stages.

The high sensitivity and spatial resolution of the nanoSIMS opens a novel possibility for coupling phylogenetic identity and metabolic function in studies of mixed microbial communities from the environment. We propose that by replacing the fluorescent oligonucleotide probes used for FISH with isotopically labelled probes (stable or radioactive) or halogenated probes, individual hybridized cells can be directly identified by nanoSIMS. The hybridization procedure is essentially the same as that used for FISH and the same oligonucleotide probes can be used. The main difference is that stable isotope or radioactive elements are coupled to the probe or halogenated probes are used instead of a fluorescent dye. By combining this new type of probing with a pulse-chase experiment using an isotope labelled substrate (e.g. with ^{13}C or ^{15}N) the metabolically active cells can at the same time be phylogenetically identified during a single scan in the nanoSIMS. To describe such techniques, we may need to think of new -ISH names in parallel to FISH, like 'RISH' (radioisotope *in situ* hybridization), 'SISH' (stable isotope *in situ* hybridization) or 'HISH' (halogen *in situ* hybridization).

There are, of course, many more possibilities for using this approach in combination with the powerful toolbox of DNA, RNA, biomarker, and protein based techniques. We have used the pulse-chase experiment here only as an example. The nanoSIMS already has amazing specifications compared with more conventional SIMS instruments. Sample preparation is comparable to simple electron microscopy, or cells can even be filtered on a gold-sputtered polycarbonate filter and dried, which makes the sample preparation similar to that of normal FISH hybridization. The sensitivity for detecting ^{14}C following a radiotracer experiment is at least 1000 times that of microautoradiography (Lechene *et al*., 2006). The stan-

dard deviation for stable isotope analyses can be better than ±1‰. So, what's the catch? Well, running this instrument requires extra mass spectrometric and general instrumental expertise, and the price (≥ 2 million euros) is certainly prohibiting for most laboratories. So far little more than a dozen instruments are up and running and they are mostly used for research in material sciences, cosmochemistry, geology and biology. We predict that in the future the nanoSIMS will show up also on the wish-list of many environmental microbiologists.

References

Lechene, C., Hillion, F., McMahon, G., Benson, D., Kleinfeld, A.M., Kampf, J.P., *et al.* (2006) High-resolution quantitative imaging of mammalian and bacterial cells using stable isotope mass spectrometry. *J Biol* **5:** 20. doi:10.1186/jbiol42.

McMahon, G., Glassner, B.J., and Lechene, C.P. (2006) Quantitative imaging of cells with multi-isotope imaging mass spectrometry (MIMS) – nanoautography with stable isotope tracers. *Appl Surf Sci* **252:** 6895–6906.

Orphan, V.J., House, C.H., Hinrichs, K.-U., McKeegan, K.D., and DeLong, E.F. (2001) Methane-consuming Archaea revealed by directly coupled isotopic and phylogenetic analysis. *Science* **293:** 484–487.

Single-cell genomics

Howard Ochman, Departments of Biochemistry and Molecular Biophysics, and Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA

A burgeoning technology that will revolutionize the analysis of microbial communities will be the ability to obtain a complete genome sequence from an individual bacterial cell. At present, the majority of published genome sequences represent bacteria that can be grown in culture. But as most bacteria live in close proximity to or in contact with other organisms and most cannot be propagated in pure culture, there are still severe limitations to the genomic analysis of the most abundant and most diverse group of organisms on the planet. Certain cultivation issues have been minimized by applying procedures (such as filtration, centrifugation, enzymatic treatment or library screening) that return relatively pure samples of DNA in quantities sufficient for genome sequencing. Although contaminating DNA might sometimes offer insights into community or cellular interactions and contents, it certainly does not make the job of genome assembly any easier. The genomic analysis of microbial communities is further complicated by the fact that each of the species can harbour substantial amounts of genetic variation. So even in the event that species-specific DNA can be recovered, it is difficult to ascertain how polymorphisms might assort among the various lineages within the sample.

Such problems will be circumvented by single-cell genomics. These procedures will eliminate the need to cultivate or mechanically purify large samples of cells, to recognize contaminating sequences, or to presume *anything* about the contents of an individual genome. Imagine isolating a single bacterial cell in the morning and obtaining its complete, gap-free and annotated sequence genome just after lunch. Okay, the technology is not quite there, but it is close. Single cells have been isolated by various methods, such as optical tweezers, flow-sorting, and serial dilution. The DNA from individual cells seems sufficient for some form of whole-genome amplification (e.g. Raghunathan *et al*., 2005). Finally, clone-free sequencing methods are producing volumes of DNA sequence information in a matter of hours (e.g. Margulies *et al*., 2005). And in fact, there is already one report of its execution (Zhang *et al*., 2006). Mark my words, by next year you will be contemplating grant proposals, and in 2 years reviewing manuscripts, that invoke this technology.

References

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:** 376–380.

Raghunathan, A., Ferguson, H.R., Jr, Bornarth, C.J., Song, W., Driscoll, M., and Lasken, R.S. (2005) Genomic DNA amplification from a single bacterium. *Appl Environ Microbiol* **71:** 3342–3347.

Zhang, K., Martiny, A.C., Reppas, N.B., Barry, K.W., Malek, J., Chisholm, S.W., and Church, G.M. (2006) Sequencing genomes from single cells by polymerase cloning. *Nat Biotechnol* **24:** 680–686.

Moving to a higher level of abstraction

Ross Overbeek, Integrated Genomics, Inc., Chicago, IL, USA

I vividly remember a point in 1978 when, as a computer science professor, I told a student that the significance of microprocessors was grossly oversold. Since then, I have tried to be somewhat more cautious in my predictions. However, in 1994 I did predict that over 100 genomes would be sequenced by the turn of the century, and thanks largely to the efforts of Craig Venter, that prediction turned out to be reasonably accurate. To increase my odds of success, in this article I will comment only on the immediate future of bacterial genome annotation and include one rather specific prediction for a change of paradigm.

The field of biology is advancing rapidly due to acquisition of data in many forms. The underlying driving force is certainly the continuing drop in costs of sequencing. The annotation and exploration of hundreds, and within

the next 2–3 years thousands, of genomes is becoming a task of central importance. There are a number of related advances that have become clearly predictable. When viewed separately, one has a perception of rapid progress in each of several areas, but when viewed as an integrated whole the advances portend a qualitative shift in how we work with genomes.

The more-or-less independent advances are as follows (these are not so much predictions, but should be thought of as relevant facts):

- It now seems very likely that thousands of genomes from single-celled organisms will become available in the next few years.
- High-throughput, accurate annotation of gene function becomes achievable as the notion of annotation shifts from 'annotate a genome' to 'annotate a single pathway or subsystem consistently through all sequenced genomes'.
- Regulatory sites can now be determined with reasonable speed and accuracy via comparative analysis, and the results can be projected onto sets of genomes.
- High-volume expression data, most commonly in the form of microarray data, are beginning to exist. We will routinely see large collection of expression data covering numerous, diverse conditions for many organisms.
- Construction of accurate metabolic reconstructions automatically from encoded subsystems is becoming possible.
- *In silico* modelling of metabolic reaction networks, based on these new reconstructions, will be used to routinely analyse states of the cell.

These advances are clearly happening, although the speed with which they are occurring may be questioned.

When skilled biologists work with an organism, they constantly view it in terms of abstractions that often remain in the background. My prediction is that in the next few years these advances will support the development of consistent integrations that allow a *state of the cell* to be viewed as a set of *regulons,* a regulon to be viewed as a set of *variants of subsystems*, and a *variant of a subsystem* to be viewed as a set of *gene products implementing a specific outcome*. Thus, my prediction is that we will shift from thinking of a state of the cell or a regulon in terms of specific genes, but rather in terms of a clear conceptual hierarchy. This undoubtedly seems innocuous compared with the 'facts' I listed above, but it is a transition that may or may not happen, and its consequences would be profound. Movement to these higher level constructs is only possible in the presence of consistent, accurate characterization of the underlying reality. I predict that this will occur for a limited set of organisms soon, but for numerous organisms quite rapidly thereafter.

The importance of individuals and scale: moving towards single cell microbiology

Les Dethlefsen and David A. Relman, Departments of Microbiology and Immunology, and Medicine, Stanford University, Stanford, CA, USA

It is commonly said that no two snowflakes are alike. This may in fact be the case with living cells. As we now enter an era of single cell microbiology, the sources of variation between individuals and the nature of individuality (both genetic and otherwise) become more clear. Technological developments will drive discovery in this area during the next few years. And the implications of the resultant findings for our understanding of microbial diversity, interactions of microbial communities with environment, and the function of biological systems will be profound.

The origins of single cell microbiology have been associated with Leeuwenhoek, who first saw individual bacteria in 1683. Following his basic observations of cell shape and motility under the microscope, numerous reagents and devices have been developed that reveal structural, chemical, metabolic and phylogenetic details of single microbes. However, these techniques are generally applied to samples containing many millions of cells or more, only a tiny fraction of which are ever examined, and eve then with relatively poor resolution. Other venerable techniques of microbiology, such as obtaining pure cultures via streaking or dilution, are rooted in the concept of isolating single microbial cells. For many decades, though, any investigation of an isolated cell has required many rounds of replication in culture to bring the mass of microbes above the detection limit of available tools. One is left with homogenized, population-wide measurements.

Recent progress in our ability to handle tiny volumes of liquid, along with advances in detection and measurement technology, herald the prospect of microbial experiments at the scale of the microbes themselves. For example, tools and procedures have been developed that will soon allow us to obtain the complete genome sequence of a single cell belonging to an uncultivated microbial species, directly from the environment (Ottesen *et al*., 2006; Zhang *et al*., 2006). Functional chemostats have been demonstrated with working volumes measured in nanolitres or even picolitres, containing at most hundreds or thousands of cells. An interconnected landscape has been etched on a silicon wafer at a scale such that the migration of a single bacterium can exert a significant influence on the population that develops in an individual habitat patch (Keymer *et al*., 2006).

As with other technological advances dating back to Leeuwenhoek's exquisite single-lens microscopes, we predict that microfluidics, nanofabrication and highly sensitive analytical techniques will enable the discovery and investigation of phenomena that deepen our understand-

ing of microbiology, and indeed of all life. For example, tracking the replication of individual *Escherichia coli* cells and their descendants across multiple generations has recently revealed senescence in organisms that reproduce by symmetric binary fission (Stewart *et al*., 2005); this last refuge of biological immortality has proven to be an illusion. On the other hand, the biochemistry of ageing via oxidative damage and life history tradeoffs between longevity and reproduction may be shared between us and the humblest of bacteria. Such a discovery would have been impossible without the ability to monitor individual cells over time.

Perhaps the most straightforward application of single-cell techniques may be the characterization of seemingly homogenous bacterial populations, including those comprised of clonal descendants from single cells. Given the large number of cells in traditional microbiological experiments and our current estimates of genome-wide mutation rates, most microbial measurements have almost certainly involved heterogeneous populations. In the past, we assumed (hoped) that novel mutants remained rare and did not disturb population-wide measurements too much. The rapidity with which a Growth Advantage in Stationary Phase mutant can take over a stationary phase culture, and reports of high frequency genetic rearrangements, suggests that, at least in some cases, these assumptions may not have been well justified. As we begin to work with many fewer total number of cells in an experiment, the expected time before mutation or rearrangement in this population will increase proportionately.

Among other important potential applications, these techniques will facilitate experiments that examine the importance of stochastic fluctuations ('noise'), and cell-to-cell variation in features such as gene expression capacity, pathway capacity, and the partitioning of cellular components to daughter cells during replication (for example in yeast cells, see Colman-Lerner *et al*., 2005). Both modelling and empirical data indicate that the behaviour of individual cells can deviate considerably from the average behaviour of a large number of cells. Variation in gene expression can lead to dramatic differences in the fate of genetically identical bacterial cells, such as the 'suicide bomber' phenotype displayed by a small proportion of cells belonging to a colicin-producing strain, or the rare 'persister' cells that 'voluntarily' shut down their growth activities but concomitantly gain phenotypic resistance to antibiotics and other stresses. These phenomena were discovered because natural selection exploited the wide variance in the output of certain gene regulatory circuits to confer an obvious fitness benefit. Many less obvious microbial traits may also turn out to depend on the variance in expression or pathway capacity among cells, rather than on the mean.

The ability to isolate and cultivate small numbers of cells will greatly enhance studies of evolutionary adaptation. Given a sufficient number of cells, certain mutations arise predictably in the laboratory in response to a particular selective regime. One example is the set of 'wrinkly spreader' mutations that arise in a single operon in *Pseudomonas fluorescens*, allowing this organism to colonize the air–medium interface in static broth cultures. Does such predictability in this organism indicate the availability of just one adaptive pathway, or does it indicate the disproportionately large immediate benefit conferred by such mutations – so that they sweep to fixation and out-compete other mutations with a smaller benefit that would otherwise have set the organism on a different adaptive trajectory? Experiments using small numbers of cells might reveal details of the adaptive landscape that cannot otherwise be discerned. Such an approach might help answer whether multiple distinct pathways eventually converge towards the same optimal phenotype, or whether certain early mutations predetermine subsequent paths towards different fitness optima.

Experiments using microdevices with single cells teach us microbial ecology at scales relevant in the natural world. The colonization of a copepod faecal pellet by marine microbes or the early stages of gut colonization in a newborn may involve only a few microbes. Does it matter which species or strains are present? Or does the physical and chemical environment exert such an influence during later time points that the initial events of colonization have no particular bearing on the community that ultimately develops? The evolution of antagonistic or cooperative relationships often depends on the extent to which organisms and their descendants continue to share the same environment. The answers to these questions no doubt differ depending on the situation. However, the patchiness of environmental conditions and the heterogeneity of microbial distributions that have been found in many environments when investigated at the scale of the microbes themselves suggest that our ability to conduct experiments at these scales will help reveal the forces that shape the evolution and ecology of microbes in the natural world.

References

Colman-Lerner, A., Gordon, A., Serra, E., Chin, T., Resnekov, O., Endy, D., *et al.* (2005) Regulated cell-to-cell variation in a cell-fate decision system. *Nature* **437:** 699–706.

Keymer, J.E., Galajda, P., Muldoon, C., Park, S., and Austin, R.H. (2006) Bacterial metapopulations in nanofabricated landscapes. *Proc Natl Acad Sci USA* **103:** 17290–17295.

Ottesen, E.A., Hong, J.W., Quake, S.R., and Leadbetter, J.R. (2006) Microfluidic digital PCR enables multigene analysis of individual environmental bacteria. *Science* **314:** 1464–1467.

Stewart, E.J., Madden, R., Paul, G., and Taddei, F. (2005) Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol* **3:** e45.
Zhang, K., Martiny, A.C., Reppas, N.B., Barry, K.W., Malek, J., Chisholm, S.W., and Church, G.M. (2006) Sequencing genomes from single cells by polymerase cloning. *Nat Biotechnol* **24:** 680–686.

Real-time microbial ecology

Forest Rohwer, Department of Biology, San Diego State University, CA, USA

The future will be real-time measurements of microbial dynamics, community composition and metabolic processes.

Determining the composition and turnover of microbial species will be some of the first real-time measurements to be made. There are already systems that monitor plankton based on image analyses (FlowCAM; http://www.bigelow.org/flowcam/flo_r2.html). Satellites and *in situ* instruments (CytoByou; http://www.cytobuoy.com/) monitor phytoplankton based on autofluorescence. And nucleic acid hybridization-based and ELISA platforms have been deployed (e.g. the MBARI Environmental Sample Processor; http://www.mbari.org/esp/). Automated, high-density microarray analyses are just around the corner. The approaches can determine the types and relative abundances of microbes within an environment, but for the most part, the above techniques search for microbes that we already know are there. In the future real-time, discovery-based sequencing will come to dominate the field. In the immediate future, look for the deployment of 454-like pyrosequencers (http://www.454.com/) or similar technologies. Expect automated nucleic acid isolation combined with sequencing platforms to be deployed within 5 years. Proteomic and metabolomic based analyses will soon follow, or in some cases even precede the nucleic acid methods. Even massive sequencing of RNA populations will become routine and replace the current array technologies.

The nucleic acid, protein and metabolite profiling will be the easy part. These technologies essentially already exist and just need to be modified for field deployment. To do real-time monitoring, we will also need to measure microbial growth rates. Current methods rely on adding radiolabelled subunits of biological polymers (e.g. nucleotides and amino acids) as tracers and then incubating for some time. At the end of the incubation, the polymers are biochemically isolated from the free subunits and the amount of incorporated radiation is measured. In turn, conversion factors are used to determine how much 'production' has occurred. These methods are not readily applicable to real-time approaches, because of the hazardous waste and complicated manipulations. Also, the necessary conversion factors, which are derived from laboratory cultures, are suspect in their application to environmental samples. These methods do not measure respiration, so the impact of microbes on their surrounding environments cannot be determined. Finally, these protocols are very hard to apply to soils and sediments. In the immediate future, selective probes including microprobes will provide some real-time data on microbial activities. There are already platforms that can remotely deploy microprobes to measure rates in soils and sediments. Oxygen microprobes and respiration chambers should also be adaptable to real-time systems. In the more distant future, real-time measurements of energy will be made using calorimeters and mass spectrometry plus natural isotopes will be used to measure rates of important metabolic pathways.

The main limitation to the real-time measurement of microbial ecology will not be the techniques, but rather the computing power and analyses. Currently, it takes days to months to analyse the massive metagenome data sets that can be generated by a 454 pyrosequencer. Just wait, soon metagenomic data and other measurements will be pouring in from the field and laboratories around the world. Microbiologists will need to develop new visualization tools that will allow them to sort through incredibly massive data sets, new algorithms will be needed, and we are going to need a whole lot more computing power.

Data storm

Marc Strous, Department of Microbiology, Radboud University, Nijmegen, the Netherlands

Okasaki, Japan.

Back in the hotel room, I unpack the crystal from its wrappings. It seems to me that something must be terribly wrong when scientists start peering in crystal balls. But that only adds to the excitement. I trace the clouds moving faster and faster inside and wait greedily until the future of the field will be unveiled.

Our discipline unfolds. Microbiologists performing most-probable-number-counts. Biochemists unravelling the workings of central metabolism and many catabolic pathways. The ecological-minded such as Winogradsky, already very much aware of the limitations of agar plates. The geologically minded, tying microbes to the history of the Earth and the functioning of current ecosystems. The blossoming of molecular genetics. Applications in crop protection, wastewater treatment. Confident professors building successful research teams around a central microbe or biogeochemical process.

Next comes the great plate count anomaly. The coming of age of molecular ecology. A 'natural' classification of microbes, no longer based on morphology or

lifestyle, but on molecular clocks such as 16S rRNA. Technological innovation seems unstoppable. Improved sensing techniques yield detailed knowledge on nutrient profiles and community changes in space and time. Whole genomes are being unravelled; even complete communities are targeted by the Sanger sequencer, now followed by the pyrosequencer. Analytical chemists practice community proteomics, speak about 'single-cell-proteomics'. System scientists start telling stories about predicting whole ecosystems without the need for new experiments.

> In the future there will be so much going on that no one will be able to keep track of it (Byrne, 1985).

The data storm forces environmental microbiologists to not only know about cultivation, geology and physiology of a *single* microorganism (or clade of microorganisms). To make sense we will be familiar with the biochemistry of *all* sequenced microorganisms, will be proficient in database mining, microsensing, heterologous expression, use of biomarkers, systems sciences, etc.

To make this happen we have to collaborate. The data storm is a vast treasure full of opportunities for new ecological and geochemical discovery. However, for correct interpretation and to prevent false leads there is an urgent need to involve biochemists of many kinds. They will share their expertise in a public interactive online forum based on a uniform definition of a comprehensive set of clusters of orthologous groups (an 'annotation wikipedia') and a moral of shared authorship.

Exciting discoveries will come from the hunt for the remaining phyla and clades exposed by the storm but presently without cultivated representatives. We see them in clone libraries, metagenomes and even under the microscope: what are they doing there? The remaining unknowns may not be so strikingly abundant as was the case for the marine crenarchaeotes, but also at 5% of the population chemolithoautotrophs or secondary-metabolite producers do make a big difference to an ecosystem. Further, there is no doubt that knowledge on their biochemistry, cell biology and membrane lipids will shed new light on key problems like the interpretation of paleomolecular data and the evolution of bacteria.

Most importantly, the storm opens the door to the reconciliation of the concepts *species* and *niche*. Currently we are confronted with closely related microorganisms which occupy very different ecological niches – for example the genus *Burkholderia* contains animal parasites, plant symbionts, polychlorinated hydrocarbon degraders, etc. On the other hand a single ecological niche can be occupied by completely unrelated microorganisms – for example alpha-, gamma-, deltaproteobacteria and Nitrospira all include aerobic nitrite oxidizers. Conventionally substrate range and affinity, the general environmental conditions (such as pH) and in some cases the dynamics of an ecosystem are considered a reasonable definition of a niche.

The data storm will lead future microbiologists to other properties of microorganisms and conceptually new ecological niches. Other properties may include the nature of the cell envelope, membrane composition, starvation responses and the structure of central metabolism. New concepts of niches may include the capability to grow very slowly or fast, the cell concentration in a given habitat, resistance to phages and predators, and the disposition to genetic change. One day we might know what it means to be a *Burkholderia* and understand that nitrite oxidation is a trivial property of otherwise completely different microorganisms. At that moment it is possible to draw a map of the 'holey adaptive landscape' (Gavrilets, 2003) that defines and explains the evolution of the prokaryotes.

I guess the crystal shows that the unity of emerging concepts of species and niche is essential for a convincing narrative of the way evolving microbes shaped and shape the Earth. It will help us to cope with climate change, a problem that is so urgent and complex that it remains to be seen if the data storm will facilitate timely answers.

It is essential to transmit our findings to the public, to inspire young people and attract more students to the field. This should be possible. The revolution in molecular biology in the previous century also attracted many talented young people to science. To translate the abstract data storm into a powerful narrative that can be communicated to teachers and the media are all-important in a changing world.

References

Byrne, D. (1985) 'In the future', from the album 'Music for the knee plays'. Munich, Germany: ECM Records.

Gavrilets, S. (2003) Models of speciation: what have we learned in 40 years? *Evolution* **57:** 2197–2215.

