

Embracing the mantra of modellers and synthesizing omics, experiments and models

Gregory J. Dick, Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, MI, USA.

When I peer into the future of environmental microbiology, I see the emergence of the power of prediction. Driven by the synthesis of modelling with environmental omics and experimental approaches, this will mark the maturation of microbiome science into a rigorously quantitative field in which we use mathematical equations to simulate microbial processes across scales from molecular to cellular and from community to global. In this new era, we have moved beyond defining ‘who is doing what’ (genes, pathways, organisms) and towards understanding how organisms and metabolisms interact with each other to determine outcomes of microbial processes that govern critical biogeochemical, engineered and human host systems. We continue to fumble in the darkness of the vast unknown taxonomic and genetic diversity of the microbial world, and to marvel at the uncovering of new players and pathways that overturn long-held tenets. However, rather than being haunted by unknowns, we are guided by robust quantitative approaches that constrain the boundaries of our ignorance and provide blueprints for breaking down key roadblocks.

For years we have posed and said that our molecular insights will lead to a better understanding of microbial ... you name it. But in how many cases have we actually used the data streams from omics to *quantify* functional aspects of microbial communities? Yes, there have been stunning insights that change the way we view the role of microbes in mediating various processes. But the promise of synthesizing this data, and especially in translating findings into tangible real-world products is largely unmet. The growing piles of omics data represent collections of observations, perhaps advancing specific research areas and opening up new views of our ignorance of the microbial world, but too often they stop short of fully predicting *outcomes* of microbial communities at higher levels. This is a shame because there is no shortage of critical societal issues for which simulations of microbial communities and predictions of their emergent properties are needed to inform policy and applied technologies (Blaser *et al.*, 2016). Modelling offers a path to achieving such

synthetic and predictive capabilities, and the explicit inclusion of microbes into models can alter predictions relative to conventional models (Treseder *et al.*, 2012).

What types of models are we talking about here? Mechanistic models are the most powerful, and challenging. They integrate biological knowledge about the function and properties of particular genes, enzymes and pathways (kinetics, inhibitors, limiting substrates) as well as fundamental processes such as rates and controls on transcription and translation, degradation of mRNAs and proteins and growth efficiency. Such models also simulate environmental conditions that are directly tied to microbial processes: the concentrations of metabolites, nutrients and growth substrates and the rates at which they are produced, consumed, transformed and transported. These models can test and advance our understanding of microbial systems by synthesizing physiological knowledge with microbial community data derived from multi-omics methods with geochemical concentrations and process rates from biogeochemical methods (Reed *et al.*, 2014; 2015; Louca *et al.*, 2016). An important philosophy of such modelling is that one need not have *a priori* knowledge about all the key biological and environmental parameters. Rather, models can be used to study, constrain and discover microbially mediated processes. Though uncomfortable at first, microbiologists will grow to embrace the mantra of modellers: ‘all models are wrong, but some are useful’.

The stage is already set for a marriage between omics approaches and modelling to produce outcome-oriented offspring. Looking back from the future it will be easy to recognize that the signposts were all around us; pockets of this revolution were already underway in the 2nd decade of the 21st century. First, the rapidly accumulating omics data provide an incredibly rich resource for numerical models. Gene abundances and distributions in metagenomic datasets serve as markers of microbially mediated processes – whether we’re talking about cryptic biogeochemical cycles (Canfield *et al.*, 2010), transformations of toxins and pollutants (Podar *et al.*, 2015) or markers of human diseases (Pasolli *et al.*, 2016). Second, a long tradition of modelling is already in place in biogeochemistry, environmental engineering, ecology, physiology and allied fields. Such

models are valuable in their own right, yet given the potential for cryptic and complex biogeochemical cycles, often they cannot resolve exactly what pathways are operative – thus they lack mechanistic understanding and predictive power. This is a critical gap that multi-omics methods can now fill. Third, modelling has already started infiltrating the field of microbial ecology. Equations are used to describe fluid dynamics and microbial processes and interactions, and to connect microscale dynamics with larger scale processes (Smruga *et al.*, 2016). Metabolic models of microbial communities started with simple communities (e.g. Freilich *et al.*, 2011) and have begun to incorporate omics data (Larsen *et al.*, 2011; Reed *et al.*, 2014; 2015).

What will it take to break down barriers to widespread integration of omics data and modelling in microbial ecology? A cultural shift is required, and indeed it's already starting to happen. The Gordon and Betty Moore Foundation Marine Microbiology Initiative was an early driving force of modelling approaches for microbial ecology (Fuhrman *et al.*, 2013). Here at the University of Michigan a Burroughs-Wellcome Fund training grant emphasizes the combination of modelling, laboratory and population approaches for graduate students in microbiome sciences across campus. Perhaps equally important to training students, it uses a co-advising strategy to bring together faculty with expertise in these different approaches. In addition to bringing together the right people and disciplines, full leverage of omics data in models will also require more coordination and standardization of effective data collection, storage and dissemination, a feat made possible only by the sustained long-term federal funding of cyberinfrastructure specialized for microbial communities.

The integration of omics approaches with modelling will be a big step forward for our community, but it will not be sufficient to achieve predictive power for microbial communities due to the preponderance of missing physiological and biochemical data on uncultured microbial groups. Hence, another key advance will be linking experiments to omics and modelling efforts. By this I mean experiments with communities – including naturally complex communities, simpler subsets obtained by enrichment or physical separation or synthetic communities – in which process rates are tracked alongside omics under controlled conditions, and the data integrated into models of the types described above. Of course traditional experimental approaches will continue to be central to the generation of new physiological, biochemical and genetic knowledge, and exciting breakthroughs in cultivation methods will enable indispensable work on a growing diversity of microbes in pure cultures. Yet the challenge of uncultured organisms, and the importance of interacting consortia in determining emergent properties of microbial communities, will lead to a growing

emphasis on communities as the experimental unit, with synthetic communities serving as 'standards' (Zomorodi and Segre, 2016). Simulations of microbial processes and interactions will be calibrated with results from these experiments as well as field observations, allowing connection of genome-scale models to ecosystem-scale models (Louca and Doebeli, 2015).

As generation and analysis of omics data gets cheaper and easier, high-resolution spatiotemporal datasets will further improve the parameterization and power of models. High-resolution data, combined with the new modelling culture, will yield whole new pathways of scientific inquiry. In this new paradigm, modelling will be used not only for predictions and data synthesis, but also to determine key unknown physiological parameters (kinetic properties, rates of transcription and translation and of mRNA and protein degradation) (Larsen *et al.*, 2011; Louca *et al.*, 2016). Hence, my crystal ball shows an exciting new era in which the dimensions of space and time are incorporated into models that project moving pictures of microbial processes.

Acknowledgements

I am grateful to the excellent group of students and post-docs that I have been fortunate to work with at the University of Michigan. In particular, Dr. Daniel Reed introduced me the world of modelling, with motivation and financial support from the Gordon and Betty Moore Foundation Marine Microbiology Initiative.

References

- Blaser, M.J., Cardon, Z.G., Cho, M.K., Dangl, J.L., Donohue, T.J., Green, J.L., *et al.* (2016) Toward a predictive understanding of earth's microbiomes to address 21st century challenges. *MBio* **7**: e00714–16.
- Canfield, D.E., Stewart, F.J., Thamdrup, B., Brabandere, L.D., Dalsgaard, T., DeLong, E.F., *et al.* (2010) A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean coast. *Science* **330**: 1375–1378.
- Freilich, S., Zarecki, R., Eilam, O., Segal, E.S., Henry, C.S., Kupiec, M., *et al.* (2011) Competitive and cooperative metabolic interactions in bacterial communities. *Nat Commun* **2**: 589.
- Fuhrman, J., Follows, M.J., and Forde, S. (2013) Applying “-omics” data in marine microbial oceanography. *Eos* **94**: 241.
- Larsen, P.E., Collart, F.R., Field, D., Meyer, F., Keegan, K.P., Henry, C.S., *et al.* (2011) Predicted relative metabolic turnover (PRMT): determining metabolic turnover from a coastal marine metagenomic dataset. *Microb Inform Exp* **1**: 4.
- Louca, S., and Doebeli, M. (2015) Calibration and analysis of genome-based models for microbial ecology. *Elife* **4**: e08208.
- Louca, S., Hawley, A.K., Katsev, S., Torres-Beltran, M., Bhatia, M.P., Kheirandish, S., *et al.* (2016) Integrating

- biogeochemistry with multiomic sequence information in a model oxygen minimum zone. *Proc Natl Acad Sci USA* E5925–E5933 (in press).
- Pasolli, E., Truong, D.T., Malik, F., Waldron, L., and Segata, N. (2016) Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. *PLoS Comput Biol* **12**: e1004977.
- Podar, M., Gilmour, C.C., Brandt, C.C., Soren, A., Brown, S.D., Crable, B.R., *et al.* (2015) Global prevalence and distribution of genes and microorganisms involved in mercury methylation. *Sci Adv* **1**: e1500675.
- Reed, D.C., Algar, C.K., Huber, J.A., and Dick, G.J. (2014) Gene-centric approach to integrating environmental genomics and biogeochemical models. *Proc Natl Acad Sci USA* **111**: 1879–1884.
- Reed, D.C., Breier, J.A., Jiang, H., Anantharaman, K., Klausmeier, C.A., Toner, B.M., *et al.* (2015) Predicting the response of the deep-ocean microbiome to geochemical perturbations by hydrothermal vents. *ISME J* **9**: 1857–1869.
- Smriga, S., Fernandez, V.I., Mitchell, J.G., and Stocker, R. (2016) Chemotaxis toward phytoplankton drives organic matter partitioning among marine bacteria. *Proc Natl Acad Sci USA* **113**: 1576–1581.
- Treseder, K.K., Balsler, T.C., Bradford, M.A., Brodie, E.L., Dubinsky, E.A., Eviner, V.T., *et al.* (2012) Integrating microbial ecology into ecosystem models: challenges and priorities. *Biogeochemistry* **109**: 7–18.
- Zomorodi, A.R., and Segre, D. (2016) Synthetic ecology of microbes: mathematical models and applications. *J Mol Biol* **428**: 837–861.