

2008) and can be expanded to be of use for nutritional and pharmaceutical interventions as well as to discriminate between health and disease state of the intestinal tract.

Given the abundance and vast coding capacity of the intestinal microbiome, it will be of great interest to follow the developments in the International Human Microbiome Consortium and see whether it can serve as a model for other ecosystems where next-generation sequence technology is applied to expose the coding capacity and function of the microbial world on our planet.

## References

- Ingham, C.J., Sprengels, A., Bomer, J., Molenaar, D., van den Berg, A., van Hylckama Vlieg, J.E., and de Vos, W.M. (2007) The micro-Petri dish, a million-well growth chip for the culture and high-throughput screening of microorganisms. *Proc Natl Acad Sci USA* **104**: 18217–18222.
- Konstantinov, S.R., Smidt, H., de Vos, W.M., Bruijns, S.C., Singh, S.K., Valence, F., *et al.* (2008) S layer protein A of *Lactobacillus acidophilus* NCFM regulates immature dendritic cell and T cell functions. *Proc Natl Acad Sci USA* **105**: 19474–19479.
- Mardis, E.R. (2008) Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* **9**: 387–402.
- Mullard, A. (2008) The inside story. *Nature* **453**: 578–580.
- Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermúdez-Humarán, L.G., and Gratadoux, J.J. (2008) *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci USA* **105**: 16731–16736.
- Zoetendal, E.G., Rajilic-Stojanovic, M., and de Vos, W.M. (2008) High throughput diversity and functionality analysis of the gastrointestinal tract microbiota. *Gut* **57**: 1605–1615.
- problems is the lack of reliable experimental information on community-wide spatial and temporal dynamics of microbial community structure, activities and functions. Because microorganisms are difficult to directly count, obtaining enough data required for modelling is extremely difficult and even impossible with conventional molecular techniques such as PCR-based cloning, *in situ* hybridization and quantitative PCR. Although applications of such conventional molecular techniques over the last two decades have provided new insights into microbial diversity and structure, they have failed to provide community-wide quantitative information in a rapid fashion for modelling and predicting microbial community dynamics. Lack of sufficient reliable data prevents microbial ecologists from addressing quantitatively important ecological and environmental questions such as competition, stability, succession, and adaptation in microbial communities. Such questions are extremely important for the applied manipulation of microbial communities for desired functions related to human health, food, energy production, environmental cleanup, and industrial and agricultural practices.

With the recent development and application of large-scale high-throughput sequencing (see Sogin *et al.*, 2006; Huber *et al.*, 2007; Hamady *et al.* 2008) and associated metagenomics technologies such as GeoChip related functional gene arrays (He *et al.*, 2007; Zhou *et al.*, 2008; Handelsman *et al.*, 2007), community-wide spatial and temporal information on microbial community functional structure and activities can be rapidly obtained. This will be the first time that microbial ecologists will be not frustrated by the paucity of experimental data to address ecological questions. In turn, microbial ecologists may very well be overwhelmed by massive amounts of metagenomics data and will be perplexed as to how to utilize and interpret the data within ecological and environmental contexts. Here, I will argue that we are in the new era of transforming microbial ecology from descriptive studies to a quantitative and predictive science.

Using high-throughput metagenomics technologies and computational modelling, it is possible to tackle some fundamental ecological questions (Zhou *et al.*, 2004), which have been difficult to address before such as: (i) How do the phylogenetic and functional structure of microbial communities change across various spatial (from micrometers, kilometers to tens of thousands kilometers) and temporal scales and what are the forces shaping their diversity and structure? (ii) How can the complex ecological networks in microbial communities be identified and whether are they important to ecosystem functioning? (iii) What is the molecular basis for functional stability and adaptation of microbial communities? (iv) How is the functional stability of a microbial community related to its genetic and metabolic diversity as well as

## Predictive microbial ecology

Jizhong Zhou, Institute for Environmental Genomics and Department of Botany and Microbiology, University of Oklahoma, Norman, OK, USA.

Microbial ecology is a science that studies the relationships between microorganisms and their biotic and abiotic environments. The ultimate goal of microbial ecology is to predict who is where with whom doing what, why and when. To achieve this predictive goal, computational modelling and simulation of microbial community dynamics and behaviour at both the structural and functional levels is essential. However, in contrast to the intensive modelling studies in plant and animal community ecology, the modelling of microbial community behaviour within different environmental contexts has not been initiated. The vast majority of the community studies in microbial ecology are still at a descriptive level rather than at a quantitative and predictive level. One of the major

environmental disturbance? (v) Can the functional stability and future status of a microbial community be predicted based on the metabolic functional conservation and differentiation of individual microbial populations? (vi) Can a microbial community be manipulated to achieve a desired stable function by manipulating the metabolic traits of the community? (vii) How can the information be scaled from molecules to populations, to communities, and to ecosystems for understanding ecosystem behaviours and dynamics? (viii) Can the molecular-level understanding of microbial community structure improve our predictive power of the ecological and evolutionary responses of microbial communities to environmental changes, especially global climate changes?

Addressing the above questions in a quantitative and predictive way requires rigorous experimental designing and systematic intensive sampling of the microbial systems studied. Selection of experimental systems with appropriate complexity and replications could be very important. I believe that two general strategies can be employed. One is to focus on surveying complex natural microbial systems by using high-throughput metagenomics technologies to systematically compare the commonality and differences of microbial community diversity patterns, metabolic capacities, and functional activities across various spatial and temporal scales. While such survey-based approach provides rich information on microbial community diversity patterns and dynamics, it could be difficult to establish detailed definitive mechanistic linkages between microbial diversity and ecosystem functioning because the microbial systems in natural settings are generally very complex. Another complementary strategy is to establish well-controlled laboratory systems such as bioreactors with simplified communities to systematically examine the responses of microbial communities to environmental changes and the impacts of their responses on ecosystem functioning. Such laboratory systems are important to establish cause-and-effect relationships, because they have great advantages in terms of system controls, monitoring, data collection, replications and modelling. Determination of cause-and-effect relationships is much easier with simpler, engineered, laboratory-based bioreactor systems than with complex natural communities, as input and output parameters can be controlled, along with environmental conditions. Although the community in a controlled system is not a natural community, such systems would offer the best opportunity to acquire mechanistic understanding of the fundamental principles of interactions among various microorganisms and the molecular level ecological and evolutionary responses of microbial communities to environmental changes. Therefore, well-controlled laboratory engineered systems will be critical to predictive microbial ecology studies.

Predictive microbial ecology requires not only high-throughput experimental tools but also high performance computational capabilities. System-level understanding of the dynamic behaviour of microbial community structure, functions and their relationships to ecosystem functioning faces several grand computational challenges. First, microbial diversity is extremely high. The number of genes in a genome or populations in a community far exceeds the number of sample measurements due to high cost of measurements. It is difficult to apply classical mathematical tools such as differential equations to simulate high-throughput metagenomics data because no sole solution can be obtained for the constructed models. New mathematical theories and approaches are needed to deal with such dimensionality problems. Second, metagenomic data from analyses of transcriptomes, proteomes and metabolomes, as well as physiological and geochemical data, are heterogeneous. Synthesizing various types of large-scale data together to make biological sense is also difficult. Rapid high performance parallel computational tools are needed for data processing, computation and visualization. In addition, because the dynamic behaviours of biological systems at various levels (cells, individuals, populations, communities, and ecosystems) are measured on different temporal and spatial scales, linking cellular-level genomic information to ecosystem-level functional information for predicting ecosystem dynamics is even more challenging. Novel mathematical framework and computational tools are needed for achieving systems-level understanding and prediction of microbial community dynamics, behaviour and functional stability.

With the rapid continuing advances of metagenomics-based high-throughput experimental technologies and associated high performance computational tools, microbiologists should be able to perform more quantitative modelling studies of microbial systems as macroecologists have done since last half century. There is no doubt that the era of quantitative predictive microbial ecology is coming.

## References

- Hamady, M., Walker, J.J., Harris, J.K., Gold, N.J., and Knight R. (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat Methods* **5**: 235–237.
- Handelsman, J., Tiedje, J.M., Alvarez-Cohen, L., Ashburner, M., Cann, I.K.O., DeLong, E.F., *et al.* (2007) *Committee on Metagenomics: Challenges and Functional Applications*. Washington, DC, USA: National Academy of Sciences.
- He, Z., Gentry, T.J., Schadt, C.W., Wu, J., Liebich, S.C., Chong, Z., *et al.* (2007) GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J* **1**: 67–77.

- Huber, J.A., Mark Welch, D., Morrison, H.G., Huse, S.M., Neal, P.R., Butterfield D.A., and Sogin M.L. (2007) Microbial population structures in the deep marine biosphere. *Science* **318**: 97–100.
- 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.
- Zhou, J.-Z., Kang, S., Schadt, C.W. Garten, C.T., Jr (2008) Spatial scaling of functional gene diversity across various microbial taxa. *Proc Nat Acad Sci USA* **105**: 7768–7773.
- Zhou, J.-Z., Thompson, D.K., Xu, Y. and Tiedje, J.M. (2004) *Microbial Functional Genomics*. Hoboken, NJ, USA: John Wiley & Sons.