Environmental Microbiology (2004) doi:10.1111/j.1462-2920.2004.00668.x

Microbial population genomics and ecology: the road ahead

Edward F. DeLong*
Massachusetts Institute of Technology, Cambridge, MA, USA.

The current landscape

Microbial biology is experiencing a remarkable period of saltatory evolution. Although many longstanding and fundamental questions endure, emerging new methods, perspectives and theory are accelerating the pace of discovery and synthesis in environmental microbiology and microbial ecology. The widespread application of molecular approaches to ecological questions (Pace, 1997), the pervasive influence of microbial evolutionary perspective (Woese, 1987), the continued development of microbial population biology (Feil and Spratt, 2001) and ecological theory (Horner-Devine *et al*., 2004), and development of innovative cultivation strategies (Connon and Giovannoni, 2002; Rappe *et al*., 2002), all bear witness to the fundamental sea changes now occurring. The recent blossoming of microbial genomics adds significantly to the list. Considering that the first full microbial genomic sequence was determined less than a decade ago (Fleischmann *et al*., 1995), this accelerated pace of discovery is nothing short of astounding. Within this expanding repertoire of scientific activity, a new subdiscipline that exploits genomic technologies to study indigenous microbial assemblages is emerging. Since the beginnings of genomics, it was well appreciated that genomic perspectives could transform microbial biology – especially in the areas of ecology and evolution. It was also recognized early on that genomic techniques might provide new access to the vast and under-appreciated microbial world that had previously escaped scrutiny in laboratory settings (Olsen *et al*., 1994). In a logical extension of Norm Pace's cultivation-independent molecular phylogenetic approach to survey microbial populations (Pace *et al*., 1985), large genome fragment analyses were applied to characterize uncultivated, indigenous microbes (Stein *et al*., 1996). The advantages of genomic approaches for providing access to untapped biodiversity have also long been known and have been practiced for well over a decade now, in biotechnologically oriented bio-prospecting applications (Short, 1997). Today, the application of genomic technologies to characterize naturally occurring microbes is gaining widespread acceptance and broader application. Several terms, including 'environmental genomics', 'microbial population genomics', 'ecogenomics' and 'metagenomics', have been used to describe this emerging research agenda. The foundations of environmental genomics rest on several different technologies and disciplines, which include genomics, ecology, evolution, high throughput DNA sequencing, and bioinformatics. The adaptation of new genomic techniques to address longstanding questions in microbial ecology and evolution is a core activity in environmental genomics. Bio-prospecting for antibiotics, pharmaceuticals, and other bioproducts represents a utilitarian component of these efforts. Several overviews touching on different aspects of microbial environmental genomics have recently appeared, and serve as indicators of expanding activity and interest in this area (DeLong, 2002; Rodriguez-Valera, 2002; Nelson, 2003; Schloss and Handelsman, 2003; DeLong, 2004; Rodriguez-Valera, 2004). The repertoire of tools for 'post-environmental genomics' is also expanding, with microarray, proteomic, and metabolomic applications following fresh on the heels of environmental genomic discoveries. This enthusiasm has to be tempered of course, with the realization that our present capacity to produce genomic information (in both laboratory cultures and field populations), still surpasses our ability to analyse, interpret and use it. Major analytical challenges still lie on the road ahead. Nevertheless, studies of the genomic characterization of uncultivated microbes, gene inventories of diverse microbial assemblages, functional analyses of novel proteins and processes, and quantitative genome surveys, are all now increasing in quantity and quality.

Last summer a meeting entitled 'Metagenomics 2003', organized by Dr Christa Schleper and colleagues, was convened at the Darmstadt Technical University. It was clear from the meeting's presentations that the pace of genomic investigations in environmental microbiology and microbial ecology is accelerating. The microbial habitats now being examined in 'metagenomic' studies are diverse and expanding. Applications reported at last summer's meeting included microbial populations in the human gut, soil microbiota, anaerobic ammonia oxidizing and methane oxidizing consortia, waste water treatment communities, and biofilms on biliary stents and in drinking water.

*For correspondence. E-mail delong@mit.edu; Tel. (+1) 617 253 5271; Fax (+1) 617 258 8850.

Applied research topics, aimed at developing the tools and methods to discover new enzymes, biocatalysts, and antibiotics, were also discussed. While there was much enthusiasm for all the new studies, it was also evident that the field is still quite young, and will require more time, effort, resources, and disciplinary integration to fully mature. The pace, magnitude, and level of sophistication in environmental genomics is however, expanding rapidly. Considerable challenges loom on the horizon due to the unprecedented scale and scope of data that are becoming available, the analytical demands of these immense data sets, and the need to develop new theory, tools, and techniques. This special edition of Environmental Microbiology captures the flavour of some of the work presented at the 'Metagenomics 2003' meeting that was held last summer in Darmstadt.

Mega-metagenomics

Early studies in environmental genomics owe their origins to Norm Pace's cultivation-independent survey approach, for studying natural microbial populations (Pace *et al*., 1985; Olsen *et al*., 1986). This approach (proposed before the advent of PCR) consisted of constructing recombinant lambda libraries from mixed microbial biomass, identifying those clones that contained rRNA genes, and sequencing those rRNA genes to deduce the phylogenetic identity of individual population members. A study of bacterioplankton rRNA genes on large insert recombinant lambda phage clones helped to ground-truth this approach, and resulted in the identification of several new 'phylotypes', now known to be ubiquitous and abundant bacterioplankters (Schmidt *et al*., 1991). A logical extension of this approach involved analyses not only of the rRNA gene, but entire, large recombinant DNA inserts as well. The notion was that sequences with 20 kilobases or more of additional genomic sequence, much more could be learned about naturally occurring microbes (Schmidt *et al*., 1991). An early attempt to study planktonic marine *Archaea* proved out the basic principle and application of this approach (Stein *et al*., 1996). More recently, advances in DNA sequencing technology, development of improved cloning vectors like Bacterial Artificial Chromosomes (Kim *et al*., 1992; Shizuya *et al*., 1992), and streamlined cloning techniques, have now rendered the recovery and sequencing of large DNA inserts from naturally occurring microbes a rather routine process. Last summer's meeting in Darmstadt showcased many different examples of large insert DNA cloning and sequencing studies, for characterizing naturally occurring microbes from complex populations. Today, the environmental genomic plot thickens in even more interesting ways. It had been proposed for some time that whole genome shotgun sequencing might also be used to characterize entire microbial assemblages. The main issue, of course, is that the cost and effort for characterization of any particular microbial assemblage is hard to predict. This is because the genome coverage in a microbial population shotgun sequencing effort, will depend critically on several variables, including genome size distributions, species richness, species evenness, and haplotype variability. These parameters are difficult to determine *a priori* in any microbial assemblage. The issue becomes even more complex, when one realizes that a critical parameter element, 'microbial species', is still looking for a concrete definition! Nevertheless, two very recent studies published this year have changed the landscape significantly, and showcase the power and potential of shotgun sequencing approaches to characterize natural microbial populations. The first report focused on an acid mine drainage microbial biofilm, consisting of iron-oxidizing *Bacteria* and *Archaea* (Tyson *et al*., 2004). Via the shotgun sequencing of only 76 Mbp of sequence, Tyson and colleagues (2004) showed it was possible to assemble 'near complete' composite genomes of constituent *Bacteria* (*Leptospirillum* species) and *Archaea* (*Ferroplasma* species). Vagaries introduced by within-population haplotype genetic variability were also very well evidenced in this study. The second spectacular report of a shotgun sequencing effort was that of Venter and colleagues (2004), that described the analyses of shotgun sequencing in the Sargasso Sea. The raw magnitude of sequence data released in this one paper – totalling approximately 10 times the number of peptides currently archived in the SWISSPROT database – dramatically demonstrates the fact that only a minute fraction of microbial genomic diversity has been sampled to date. The authenticity of the contig assemblies reported in the study, and the origins of some of the apparently prevalent microbes in the Sargasso Sea samples, remains to be verified (Falkowski and de Vargas, 2004). Nevertheless, the writing is certainly on the wall with respect to some major new research directions and challenges for contemporary microbial biologists.

Evolutionary modalities

Another topical area that environmental genomics promises to shed significant light on relates to the tempo and mode of microbial genome and species evolution. In particular, the relative influence and importance of vertical inheritance versus lateral gene transfer, on microbial genome evolution and speciation, remains an open question (Doolittle, 2000; Ochman *et al*., 2000; Boucher *et al*., 2003; Daubin *et al*., 2003; Lerat *et al*., 2003). One camp suggests that a frequent occurrence in the literature is the indiscriminate pooling of orthologous and paralogous genes in phylogenetic analyses, that yields erroneous results often used to bolster support for lateral gene

transfer. This ongoing debate has made it abundantly clear that seemingly disparate areas in microbiology, including population biology, genomics and ecology, can all be used to examine different aspects of the same question. These perspectives need to be combined, and focused simultaneously on specific microbial ecosystems, to better define tempo, mode and mechanism in microbial genome evolution. Environmental genomics can now provide unprecedented data on naturally occurring genomic structure and variation, genetic drift, and lateral gene transfer in natural microbial populations. These data should help inform the currently data-poor debate, that revolves about the relative influence and significance of lateral gene transfer events, on microbial genome evolution and species differentiation. The natural microbial world provides some ideal case studies for interpreting process and dynamics in microbial genome evolution. Conversely, genomics provides the tools to study the complexity of form, function and process, that underpins microbial population structure, dynamics and evolution.

Ecological revelation

Just as perspective on microbial evolution will benefit from environmental genomic endeavours, so will our understanding of the inner workings of microbial ecosystems. Already, the first few metagenomic glimpses have provided impressive perspective on extant, naturally occurring microbial diversity. It will require time, effort, and new approaches to place this genetic diversity into proper biological context. A first step includes straightforward studies of the ebb and flow of this diversity across space and time. Microbial ecologists will also begin to think more deeply about the meaning of the genetic and functional inventories that metagenomic studies will reveal. Community gene content has important implications with respect to population biology, dynamics, ecophysiology and community interactions. As metabolic maps of 'whole community metabolism' arise from functional annotation of environmental genomic datasets, then pattern and process in ecosystem function might be examined in much greater detail and in unprecedented ways. Repeated correlation of specific genes with one another may for instance point to linkages between specific metabolic pathways, or unanticipated symbiotic interactions. The new metagenomic data will undoubtedly lead to the identification of unsuspected biological functions in particular habitats. Indeed, examples of such discoveries already exist (Béjà *et al*., 2000). Predictions of previously unknown or unsuspected details of biogeochemical cycles are also a likely outcome of this research agenda. Comparative eco-physiology will most likely occur on a much grander scale than it does now. The new subjects of study will be not individual species, but rather whole assemblages and ecosystems. It is probable that such comparative ecosystem genomics will yield new perspectives on function and pattern within natural microbial ecosystems. Microbial ecology and genomics can now reciprocally inform one another, to provide new conceptual understanding and advanced theoretical models.

The road ahead

Many challenges lie on the road ahead, some more easily tackled than others, because they only require simple methodological or economic solutions. More daunting problems will arise that relate to bridging the sometimes wide cultural and conceptual barriers separating disparate disciplines. Some of these disciplinary and conceptual gaps really do need to be bridged, before we can traverse to the places that newfound datasets have potential to lead us to. The informatics challenges and opportunities are profound. Questions that might be addressed in upcoming years include: how do we manage the new types of genomic databases that will include complete genomes, as well as incomplete genome fragments that originate from environmental samples? How do we facilitate interoperability between genomic databases, and corresponding environmental and ecological datasets? What sorts of new computational tools are required to compare and interrelate metagenomic datasets? What other sorts of data (e.g. new microbial isolate genome sequences, and large insert clone sequences) will be required to ground-truth, organize and interpret environmental shotgun sequence datasets? What new approaches will be necessary to test the many hypotheses that will originate from all the new metagenomic data? These are not simple questions, and the field will certainly require time to evolve and adapt to address them. The American Academy of Microbiology has recently held a number of colloquia and is organizing more, to delve into these and similar issues (Tiedje and Stahl, 2002; Buckely, 2004).

New and holistic ways of thinking about microbial assemblages, their interaction with one another and their environment, and their (genomic) similarities and differences may be required in the coming years. Integrative systems biology approaches, that will include the blending of environmental, bioinformatic, proteomic, metabolomic and ecological datasets, techniques, and analyses, are likely to play a key role in the future. The study of individual organisms will be joined by analyses of complex biological networks across multiple hierarchical levels. It is not entirely apparent how all these activities will take shape and evolve, as more traditional scientific disciplines adapt to and incorporate them.

What is clear, however, is that integration of a wide range of perspectives, including systems engineering, biogeochemistry, genomics, biochemistry, physiology and ecology, will all be required components of this complex cocktail.

Environmental genomics promises to produce unique and informative new tools perspectives for interpreting the microbial world around us. The new gestalt has potential to shape the way we think about and conduct our own activities as well. A deeper appreciation of form and function in sustainable microbial ecosystems may serve as a template and guide, for designing more sustainable human interaction with functioning ecosystems.

References

Béjà, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P. *et al.* (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289:** 1902–1906.

Boucher, Y., Douady, C.J., Papke, R.T., Walsh, D.A., Boudreau, M.E., Nesbo, C.L. *et al.* (2003) Lateral gene transfer and the origins of prokaryotic groups. *Annu Rev Genet* **37:** 283–328.

Buckely, M. (2004) *The Global Genome Question: Microbes as the Key to Understanding Evolution and Ecology*. DeLong, E.F., and Relman, D. (eds). Washington, DC, USA: American Academy of Microbiology.

Connon, S.A., and Giovannoni, S.J. (2002) High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl Environ Microbiol* **68:** 3878–3885.

Daubin, V., Moran, N.A., and Ochman, H. (2003) Phylogenetics and the cohesion of bacterial genomes. *Science* **301:** 829–832.

DeLong, E.F. (2002) Microbial population genomics and ecology. *Curr Opin Microbiol* **5:** 520–524.

DeLong, E.F. (2004) Microbial population genomics and ecology: a new frontier. In *Microbial Genomics*. Fraser, C.M., Nelson, K.E., and Read, T.D. (eds). Totowa, NJ, USA: Human Press, pp. 419–442.

Doolittle, W.F. (2000) Uprooting the tree of life. *Sci Am* **282:** 90–95.

Falkowski, P.G., and de Vargas, C. (2004) Genomics and evolution. Shotgun sequencing in the sea: a blast from the past? *Science* **304:** 58–60.

Feil, E.J., and Spratt, B.G. (2001) Recombination and the population structures of bacterial pathogens. *Annu Rev Microbiol* **55:** 561–590.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R. *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269:** 496–512.

Horner-Devine, M.C., Carney, K.M., and Bohannan, B.J. (2004) An ecological perspective on bacterial biodiversity. *Proc R Soc Lond B Biol Sci* **271:** 113–122.

Kim, U.-J., Shizuya, H., Dejong, P., Birren, B., and Simon, M. (1992) Stable propagation of cosmid sized human DNA inserts in an F-factor based vector. *Nucleic Acids Res* **20:** 1083–1185.

Lerat, E., Daubin, V., and Moran, N.A. (2003) From gene trees to organismal phylogeny in prokaryotes: the case of the gamma-Proteobacteria. *PLoS Biol* **1:** E19.

Nelson, K.E. (2003) The future of microbial genomics. *Environ Microbiol* **5:** 1223–1225.

Ochman, H., Lawrence, J.G., and Groisman, E.A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405:** 299–304.

Olsen, G.J., Lane, D.J., Giovannoni, S.J., Pace, N.R., and Stahl, D.A. (1986) Microbial ecology and evolution: a ribosomal RNA approach. *Annu Rev Microbiol* **40:** 337–365.

Olsen, G.J., Woese, C.R., and Overbeek, R. (1994) The winds of (evolutionary) change: breathing new life into microbiology. *J Bacteriol* **176:** 1–6.

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

Pace, N.R., Stahl, D.A., Olsen, G.J., and Lane, D.J. (1985) Analyzing natural microbial populations by rRNA sequences. *American Society for Microbiology News* **51:** 4–12.

Rappe, M.S., Connon, S.A., Vergin, K.L., and Giovannoni, S.J. (2002) Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418:** 630–633.

Rodriguez-Valera, F. (2002) Approaches to prokaryotic biodiversity: a population genetics perspective. *Environ Microbiol* **4:** 628–633.

Rodriguez-Valera, F. (2004) Environmental genomics, the big picture? *FEMS Microbiol Lett* **231:** 153–158.

Schloss, P.D., and Handelsman, J. (2003) Biotechnological prospects from metagenomics. *Curr Opin Biotechnol* **14:** 303–310.

Schmidt, T.M., DeLong, E.F., and Pace, N.R. (1991) Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J Bacteriol* **173:** 4371–4378.

Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Natl Acad Sci USA* **89:** 8794–8797.

Short, J.M. (1997) Recombinant approaches for accessing biodiversity. *Nat Biotechnol* **15:** 1322–1323.

Stein, J.L., Marsh, T.L., Wu, K.Y., Shizuya, H., and DeLong, E.F. (1996) Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J Bacteriol* **178:** 591–599.

Tiedje, J., and Stahl, D.A. (2002) *Microbial Ecology and Genomics: a Crossroads of Opportunity*. Tiedje, J., and Stahl, D.A. (eds). Washington, DC, USA: American Academy of Microbiology.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M. *et al.* (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428:** 37–43.

Venter, J.C., Remington, K., Heidelberg, J. F., Halpern, A.L., Rusch, D., Eisen, J.A., *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66–74.

Woese, C.R. (1987) Bacterial evolution. *Microbiol Rev* **51:** 221–271.