UNDER THE MICROSCOPE

Metagenomics

R.D. Sleator, C. Shortall and C. Hill

Alimentary Pharmabiotic Centre, University College Cork, Cork, Ireland

Keywords
genome, gut, marine water, metagenome, patho-biotechnology, soil.

Correspondence
Roy D. Sleator, Alimentary Pharmabiotic Centre, University College Cork, Cork, Ireland.
E-mail: r.sleator@ucc.ie

2008/0882: received 23 May 2008, revised 24 June 2008 and accepted 25 June 2008

doi:10.1111/j.1472-765X.2008.02444.x

Abstract

The total number of prokaryotic cells on earth has been estimated to be approximately $4–6 \times 10^{30}$, with the majority of these being uncharacterized. This diversity represents a vast genetic bounty that may be exploited for the discovery of novel genes, entire metabolic pathways and potentially valuable end-products thereof. Metagenomics constitutes the functional and sequence-based analysis of the collective microbial genomes (microbiome) in a particular environment or environmental niche. Herein, we review the most recent sequence-based metagenomic analyses of some of the most microbiologically diverse locations on earth; including soil, marine water and the insect and human gut. Such studies have helped to uncover several previously unknown facts; from the true microbial diversity of extreme environments to the actual extent of symbiosis that exists in the insect and human gut. In this respect, metagenomics has and will continue to play an essential part in the new and evolving area of microbial systems biology.

Introduction

Ninety-nine per cent of all micro-organisms in almost every environment on earth remain, as yet, uncultured (Amann *et al.* 1990). The discipline of metagenomics, defined as the culture-independent genomic analysis of all the micro-organisms in a particular environmental niche (Handelsman *et al.* 1998), evolved as an effort to discover more about the microbial diversity of natural environments such as soil, marine water and the gastrointestinal tracts of vertebrates and invertebrates (López-García and Moreira 2008). Metagenomics is a new and increasingly sophisticated field which in its simplest terms is concerned with the direct isolation of DNA from a defined habitat, followed by cloning (in a surrogate host such as *Escherichia coli*) of the complete genomes of the entire microbial population in that habitat (Langer *et al.* 2006). The resulting DNA library is then analysed for functions and sequences of interest (Fig. 1).

Metagenomics can be divided into sequence-based and function-driven analysis of uncultured micro-organisms (Gabor *et al.* 2007). Functional metagenomics involves screening metagenomic libraries for a particular phenotype, e.g. salt tolerance, antibiotic production or enzyme activity, and then identifying the phylogenetic origin of the cloned DNA (Dinsdale *et al.* 2008). Sequence-based approaches, on the other hand, involve screening clones for the highly conserved 16S rRNA genes for identification purposes and then sequencing the entire clone to identify other genes of interest, or large-scale sequencing of the complete metagenome to search for phylogenetic anchors in the reconstructed genomes (Riesenfeld. *et al.* 2004; Hoff *et al.* 2008). Herein, we review some of the most recent large scale sequence-based metagenomic studies, focusing on the environment (soil and marine water) and the (insect and human) gut.

Environmental metagenomics

The first extensive large-scale environmental sequencing project was carried out by the J. Craig Venter Institute in 2004 in which they sequenced fragments of DNA derived from the entire microbial population of the nutrient-limited Sargasso Sea, an intensively studied region of the Atlantic Ocean close to Bermuda (Venter *et al.* 2004). A shot-gun sequencing approach yielded over 1·6 billion

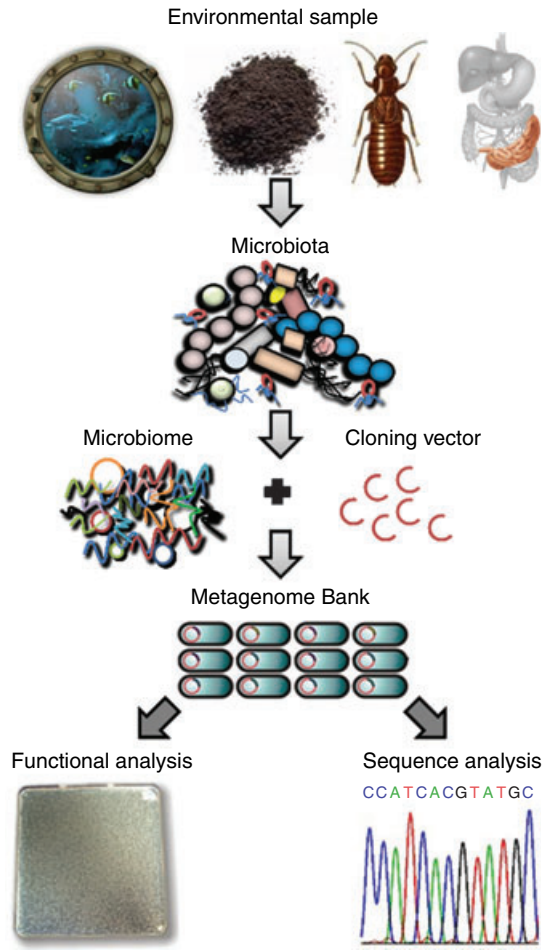

Figure 1 Schematic overview of the metagenomics, divided into functional and sequence-based approaches. While functional screening is focused on the identification of novel processes and proteins produced by heterologous gene expression against a surrogate host, sequence-based analysis provides insights into the metabolic potential and ecological diversity of an environmental community by comparing DNA databases.

base pairs of DNA and led to the discovery of 1·2 million new genes. A total of 794 061 of these genes were assigned to a conserved hypothetical protein group, the functions of which are unknown. A significant number of genes discovered from the phosphorous-limiting Sargasso Sea have been found to contain high similarity to those involved in the uptake of phosphorous such as polyphosphates, phosphonates, pyrophosphates as well as other inorganic phosphorous (Riesenfeld. *et al.* 2004). Such an extensive sequencing project contributed greatly to the sequence richness of 16S rRNA gene databases, indeed special database (GenBank) provisions had to be made so that this 'megagenomic' data would not overwhelm the ordinary 'monogenomic' data already deposited in the databank, skewing subsequent comparative analysis (Galperin 2004).

While the Sargasso Sea was originally chosen for metagenomic analysis because it was assumed to have a relatively simple community, this ultimately proved not to be the case; the community was in fact not simple enough to allow the assembly of microbial genomes as hoped (Schoss and Handelsman 2005). Tyson *et al.* (2004) chose a much simpler community in acid mine drainage (AMD) in the Richmond mine at Iron Mountain, California, one of the most extreme environments on earth. In this environment the microbiota exists as a pink biofilm that forms on the surface of the mine water. The biofilm has a pH of 0·83, a temperature of 43°C and contains high concentrations of iron, zinc, copper and aspartate (Tyson *et al.* 2004). An extended random shot-gun sequencing approach revealed 384 16S rRNA genes and the 5′ and 3′ ends of these genes were sequenced (Baker *et al.* 2003). The AMD microbiota was found to contain three bacterial and three archaeal species (Schoss and Handelsman 2005). The three dominant bacterial genera included *Leptospirillum, Sulfobacillus, Acidomicrobium* and the dominant archaeal species was *Ferraplasma acidomicrobium* (Handelsman 2004). The simplicity of the community structure allowed Tyson *et al.* (2004) to sequence almost all of the microflora with a high degree of coverage. It was noted that the G + C content of the genomes of the dominant taxonomic groups differed to a large extent and thus provided a means to source each of the clones (Bond *et al.* 2000). Metagenomic analysis of the AMD community resulted in the reconstruction of nearly complete genomes of *Leptospirillum* group II and *Ferroplasma* type II. Perhaps not surprisingly, all of the genomes in the AMD were found to be plentiful in genes that function in the removal of elements that would otherwise be toxic to the cell (Handelsman 2004).

Soil and marine water

Soil

As demonstrated by the metagenomic study of the Sargasso Sea, increased complexity of the environment under study increases the complexity of metagenomic analysis (Venter *et al.* 2004). The problems of assembly following shot-gun cloning from complex microbial environments are compounded when even more diverse ecosystems are targeted by this approach (Kowalchuk *et al.* 2007). This was clearly demonstrated by Tringle *et al.* (2005) in their analysis of a soil metagenomic library. Soil borne microorganisms are one of the earth's greatest source of biodiversity (Curtis *et al.* 2002), with estimates ranging between 3000 and 11 000 microbial genomes per gram of

soil (Schmeisser *et al.* 2007). Moreover, nearly 140 mega bases of sequence taken from Minnesota farmland soil contained <1% of sequences with any overlaps and formed no contigs (assemblies of overlapping individual clones), demonstrating that complete sequencing of highly diverse environments is virtually impossible with current technologies (Kowalchuk *et al.* 2007). However, because of the huge diversity of soil and its history as a source of commercially important molecules in agriculture, chemical, industrial and pharmaceutical industries, it remains the most common target for studies of functional metagenomics (MacNeil *et al.* 2001; Courtois *et al.* 2003; Daniel 2005).

Marine water

The marine environment has recently been pursued as a target of metagenomic studies (Li and Qin 2005; Pedrós-Alió 2006). It is the largest contiguous ecosystem on earth, occupying 71% of the earth's surface with an average depth of 4 km (Karl 2007). It is no surprise then that the earth's oceans represent one of the most significant yet least understood microbial-driven natural environments on the planet (Martin-Cuadrado *et al.* 2007). Two large scale metagenomic analyses of deep-sea communities have been carried out to date: the Pacific gyre water column at the A Long-Term Oligotrophic Habitat Assessment station (ALOHA; 22°45′N, 158°00′W – located 100 km north of Oahu, Hawaii), and more recently a single depth of 3000 m was sampled in the Ionian Sea located south-east of Sicily in the deep Mediterranean (Martin-Cuadrado *et al.* 2008). A fosmid library was constructed from the 3000-m deep Mediterranean plankton and analysed by phylogenetic analysis of 16S rRNA genes and fosmid end sequencing. Sequence analysis revealed a high similarity with genomes from *Rhizobiales* within the *Alphaproteobacteria*, *Cenarchaem symbiosum*, *Planctomycetes*, *Acidobacteria*, *Chloroflexi* and *Gammaproteobacteria* (Martin-Cuadrado *et al.* 2007). With the exclusion of the crenarchaeota, deep sea microbes assume a heterotrophic lifestyle, a presumption further verified by the detection of metabolic genes that function in the catabolism, transport and degradation of organic matter. A strongly similar community structure was found in the subtropical ALOHA water column of the Pacific. A considerable proportion of genes encoding dehydrogenases, including *cox* genes, involved in aerobic carbon monoxide oxidation have also been detected in the warm Mediterranean. Carbon monoxide dehydrogenase genes are more frequent in deep layers of the ALOHA water column, indicating that they play a significant role as an additional energy source in deep sea communities. Comparative studies of the Pacific gyre water column data and the deep Mediterranean Iondian Sea revealed that the bathypelagic (4000 m-deep) Mediterranean microbes correspond to the mesopelagic (500–700 m-deep) Pacific microbes, suggesting that (with disregard to light) temperature serves as a principal stratifying determinant of community structure (Martin-Cuadrado *et al.* 2007). In further support of this the Global Ocean Sampling metagenomic study carried out by Rusch *et al.* (2007) on surface waters found that the effect of temperature on microbial communities as a function of depth paralleled that of the deep Mediterranean and the ALOHA aphotic zone (Martin-Cuadrado *et al.* 2007).

The gut metagenome

The insect gut

In 2000 the complete genome of the fruit-fly *Drosophila melanogaster* was sequenced, pioneering an era of insect genomics (Adams *et al.* 2000; Myers *et al.* 2000). Further genomic research on insect maps; including the completion of over 30 additional whole genome sequences and more than 2 million ESTs, has effectively classified insects as the most diverse group of organisms on earth (Behura 2006).

Several metagenomic studies have been carried out on the hindgut and midgut of the wood-feeding 'higher' termite and the gypsy moth (*Lymantrisexia dispar*) respectively. Termites are widely known as economically important wood-degrading organisms (Sugimoto *et al.* 2000) with essential environmental roles in the turnover of carbon as well as serving as prospective sources of biochemical catalysts which can be used in the conversion of wood to biofuels (Warnecke *et al.* 2007). Significant data have recently emerged which suggests that the symbiotic bacteria resident in the hindgut of the termite play a functional role in the hydrolysis of both cellulose and xylan (Tokuda and Watanabe 2007). In order to gain a greater appreciation of the diversity of biological mechanisms governing lignocellulose degradation, metagenomic analysis of the microbiota of the hindgut paunch of a wood-feeding 'higher' *Nasutitermes* (which do not contain hindgut flagellate protozoa) species was carried out in search of a large, complex set of bacterial genes commonly employed for cellulose and xylan hydrolysis (Warnecke *et al.* 2007). The complete microbial community DNA of the paunch hindgut compartment was extracted, cloned and sequenced. Fifteen fosmids were chosen for further analysis after the initial end sequencing screen. A total of 1750 bacterial rRNA genes were amplified by PCR identifying a vast array of bacteria categorized into 12 phyla and 216 phylotypes. Following PCR, the genus *Treponeme* and the phylum *Fibrobacters* comprised the

most frequently recovered phylotypes with *Treponemes* being the most dominant making up 68% of marker genes and *Fibrobacters* comprising 13% of recovered fragments. A conservative analysis based on global alignment techniques identified more than 100 gene modules analogous to catalytic domains of glycoside hydrolases. This study also illustrated other potentially important functions of the microbial population of the paunch hindgut of the 'higher' wood-feeding termite, such as hydrogen metabolism, carbon dioxide-reductive acetogenesis and nitrogen fixation (Warnecke *et al.* 2007).

The human gut

Microbial communities occupy all surfaces of the human body with a combined microbial cell population approximately 10 times that of human cells (Kurokawa *et al.* 2007). The distal colon has been identified as the most densely populated natural bacterial ecosystem, encompassing more bacterial cells than all of our microbial communities combined (Marchesi and Shanahan 2007; Frank and Pace 2008). The total number of genes encoded by their collective genomes (the gut microbiome) surpasses that of the human genome by at least one order of magnitude (Bäckhed *et al.* 2005). It comes as no surprise then that the human gastrointestinal microbiota is essential; bestowing metabolic functions that are otherwise absent in the host, such as improved strategies of energy harvest from ingested foods, synthesis of essential vitamins and the degradation of complex plant polysaccharides (Gill *et al.* 2006; Kurokawa *et al.* 2007). Indeed it is not uncommon for imbalances to occur in this intestinal microbial community structure with the potential for causing diseases such as Crohn's disease inflammatory bowel disease, allergy, obesity and cancer (Manichanh *et al.* 2006; Kurokawa *et al.* 2007).

The first comprehensive mining effort of the human gut metagenome was undertaken by members of the Relman lab and The Institute for Genomic Research (TIGR) in 2005 in an effort to uncover the diversity within the gastrointestinal microflora, or to at least expand our understanding of it (Eckburg *et al.* 2005). 13 335 16S rRNA sequences were produced from mucosal biopsy samples collected from the proximal to the distal colons of three healthy adults and one faecal sample from each person. The results yielded the largest database of 16S rRNA sequences (11 831 bacterial sequences, 1524 archael sequences) from a single study of any environment with the identification of 395 bacterial phylotypes and one archael phylotype (Eckburg *et al.* 2005). The PCR amplified 16S rDNA products were cloned and sequenced bi-directionally, revealing that the vast majority of the 395 bacterial phylotypes discovered are members of the *Bacteriodetes* (48%) and the *Firmicutes* (51%); with the remaining phylotypes being represented by *Proteobacteria*, *Verrucomicrobia*, *Fusbacteria*, *Cyanobacteia*, *Spirochetes* and VadinBE9V.

In an effort to identify genomic features common to all human gut microbiomes Kurokawa *et al.* (2007) recently carried out a comparative metagenomic analysis of faecal samples from 13 healthy individuals of various ages, including unweaned infants. This study revealed that the gut microbiota from adults and children are diverse and exhibit high functional uniformity regardless of age. A total of 273 gene families were identified to be rich in adult-type microbiomes and 136 families in infant-type microbiomes, with a small overlap. Furthermore, a conjugative transposon family was found to be explosively amplified in the human microbiome, a finding which strongly suggests that the intestine is a 'hotspot' for horizontal gene transfer between microbes. This observation raises serious biosafety considerations concerning the use of so-called designer probiotics (Sleator and Hill 2008).

Perhaps the most significant recent advance in metagenomic research to date is the establishment of the Human Gut Microbiome Initiative; an interdisciplinary world wide effort to gain a greater appreciation for and understanding of the microbial components of the human genetic and metabolic landscape and how they contribute to normal physiology and predisposition to disease (Turnbaugh *et al.* 2007). The aim of this initiative, which has been described as a logical conceptual and experimental extension of the human genome project, is to produce deep draft whole metagenome sequences for reference genomes [100 species indicative of the bacterial divisions or superkingdoms within the distal gut (Gordon *et al.* 2005)] followed by a shallower 16S rRNA gene and community dataset from a moderate number of samples. Many outcomes are predicted for the Human Microbiome Project (HMP), including the identification of new biomarkers for health and medicine, new enzymes capable of degrading xenobiotics in biotechnology and ultimately a more complete understanding of the nutritional requirements of humans (Turnbaugh *et al.* 2007).

Future prospects

While the advantages of metagenomic research are obvious, there are notable limitations to this new science: (i) low resolution (due to lack of sequence coverage in complex environments), (ii) inability to classify short metagenomic fragments and (iii) a lack of functional verification continues to hamper progress in the field (Warnecke and Hugenholtz 2007).

However, combining basic metagenomics approaches with existing and new complementary technologies may

help to address these deficiencies. Dividing microbial communities into simpler subsets on the basis of cell properties such as size or nucleic acid content, using fluorescence-activated cell sorting and microfluidics (Brehm-Stecher and Johnson 2004; Weibel *et al.* 2007) for example, may compensate for the problems of low resolution, thus helping to link identity with function. In addition to cell sorting, emerging high throughput sequencing-by-synthesis technologies are paving the way to cloning independent, low-cost random 'shotgun' approaches to metagenomic sequence analysis. Particular progress has been made in the areas of randomly proliferating limited-source DNA, isolating specific target sequences from highly complex template mixtures, and cDNA cloning from extracted mRNA to directly clone actively expressed genes from microbial consortia (Gabor *et al.* 2007).

Ultimately, integrated analysis of metagenomes, metatranscriptomes, metaproteomes and metametabolomes will be needed to sustain the logical next phase of metagenomics – microbial systems biology – the interdisciplinary analysis of complex biological networks across multiple hierarchical levels *in situ*. Such systems based understanding of microbial communities in a global context may ultimately frame a number of important questions such as climate change, human nutrition and health and the emergence and management of antibiotic resistance (Dupré and O'Malley 2007).

Acknowledgements

Dr Roy Sleator is a Health Research Board (HRB) and Alimentary Pharmabiotic Centre (APC) principal investigator. The authors wish to acknowledge the continued financial assistance of both the HRB and APC.

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W. *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195.

Amann, R.I., Binder, B.L., Chisholm, S.W., Devereux, R. and Stahl, D.A. (1990) Combination of 16S rRNA targeted oligonucleotide probes with Flow-cemetry for analysing mixed microbial populations. *Appl Environ Microbiol* **56**, 1919–1925.

Bäckhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A. and Gordon, J.I. (2005) Host-bacterial mutualism in the human intestine. *Science* **307**, 1915–1920.

Baker, B.J., Hugenholtz, P., Dawson, S.C. and Banfield, J.F. (2003) Extremely acidophilic protists from acid mine drainage host Rickettsiales-lineage endosymbionts that have intervening sequences in their 16S rRNA genes. *Appl Environ Microbiol* **69**, 5512–5518.

Behura, S.K. (2006) Molecular marker systems in insects: current trends and future avenues. *Mol Ecol* **15**, 3087–3113.

Bond, P.L., Smriga, S.P. and Banfield, J.F. (2000) Phylogeny of microorganisms populating thick, subaerial, lithotrophic biofilm at an extreme acid mine drainage site. *Appl Environ Microbiol* **66**, 3842–3849.

Brehm-Stecher, B.F. and Johnson, E.A. (2004) Single-cell microbiology: tools, technologies, and applications. *Microbiol Mol Biol Rev* **68**, 538–559.

Courtois, S., Cappellano, C.M., Ball, M., Francou, F.X. and Normand, P. (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl Environ Microbiol* **69**, 49–55.

Curtis, T.P., Sloan, N.T. and Scannell, J.N. (2002) Estimating prokaryote diversity and its limits. *Proc Natl Acad Sci USA* **99**, 10494–10499.

Daniel, R. (2005) The metagenomics of soil. *Nat Rev Microbiol* **3**, 470–478.

Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., Furlan, M., Desnues, C. *et al.* (2008) Functional metagenomic profiling of nine biomes. *Nature* **452**, 629–632.

Dupré, J. and O'Malley, M.A. (2007) Metagenomics and biological ontology. *Stud Hist Philos Biol Biomed Sci* **38**, 834–846.

Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E. *et al.* (2005) Diversity of the human intestinal microbial flora. *Science* **308**, 1635–1638.

Frank, D.N. and Pace, N.R. (2008) Gastrointestinal microbiology enters the metagenomics era. *Curr Opin Gastroenterol* **24**, 4–10.

Gabor, E., Liebeton, K., Niehaus, F., Eck, J. and Lorenz, P. (2007) Updating the metagenomics toolbox. *Biotechnol J* **2**, 201–206.

Galperin, M.Y. (2004) Metagenomics: from acid mine to shining sea. *Environ Microbiol* **6**, 543–545.

Gill, S.R., Pop, M., DeBoy, R., Eckburg, P.B., Taurnbaugh, P.J. *et al.* (2006) Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1255–1359.

Gordon, J.I., Ley, R.E., Wilson, R., Mardis, E., Xu, J., Fraiser, C.M. and Relman, D.A. (2005) *Extending Our View of Self: The Human Gut Microbiome Initiative (HGMI)*. National Human Genome Research Institute. http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/HGMISeq.pdf.

Handelsman, J. (2004) Metagenomics: applications of genomics to uncultured micro-organisms. *Microbiol Mol Biol Rev* **68**, 669–685.

Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J. and Goodman, R.M. (1998) Molecular biological access to the

chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* **5**, R245–R249.

Hoff, K.J., Tech, M., Lingner, T., Daniel, R., Morgenstern, B. and Meinicke, P. (2008) Gene prediction in metagenomic fragments: a large scale machine learning approach. *BMC Bioinformatics* **9**, 217.

Karl, D.M. (2007) Microbial oceanography: paradigms, processes and promise. *Nat Rev Microbiol* **5**, 759–769.

Kowalchuk, G.A., Speksnijder, A.G.C.L., Zhang, K., Goodman, R.M. and Veen, J.A. (2007) Finding the needles in the metagenome haystack. *Microb Ecol* **53**, 475–485.

Kurokawa, K., Itoh, T., Kuwahara, T., Oshima, K., Hoh, H. *et al.* (2007) Comparative metagenomics revealed commonly enriches gene sets in human gut microbiomes. *DNA Res* **14**, 169–181.

Langer, M., Gabor, E.M., Liebeton, K., Meurer, G., Niehaus, F., Schulze, R., Eck, J. and Lorenz, P. (2006) Metagenomics: an inexhaustible access to nature's diversity. *Biotechnol J* **1**, 815–821.

Li, X. and Qin, L. (2005) Metagenomics-based drug discovery and marine microbial diversity. *Trends Biotechnol* **23**, 539–543.

López-García, P. and Moreira, D. (2008) Tracking microbial biodiversity through molecular and genomic ecology. *Res Microbiol* **159**, 67–73.

MacNeil, I.A., Tiong, C.L., Minor, C., August, P.R., Grossman, T.H. *et al.* (2001) Expression and isolation of antimicrobial small molecules from soil DNA libraries. *J Mol Microbiol Biotechnol* **3**, 301–308.

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., Nalin, R. and Jarrin, C. (2006) Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* **55**, 205–211.

Marchesi, J. and Shanahan, F. (2007) The normal intestinal microbiota. *Curr Opin Infect Dis* **20**, 508–513.

Martin-Cuadrado, A.B., Lopez-Garcia, P., Alba, J.C., Moreina, D., Strittmatter, A. *et al.* (2007) Metagenomics of the deep Mediterranean, a warm Bathypelagic habitat. *PLoS ONE* **2**, e914.

Martin-Cuadrado, A.B., Rodriguez-Valera, F., Moreira, D., Alba, J.C., Ivars-Martínez, E., Henn, M.R., Talla, E. and López-García, P. (2008) Hindsight in the relative abundance, metabolic potential and genome dynamics of uncultivated marine archaea from comparative metagenomic analyses of bathypelagic plankton of different oceanic regions. *ISME J* doi:10.1038/ismej.2008.40.

Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., Flanigan, M.J., Kravitz, S.A., Mobarry, C.M. *et al.* (2000) A whole-genome assembly of Drosophila. *Science* **287**, 2196–2204.

Pedrós-Alió, C. (2006) Genomics and marine microbial ecology. *Int Microbiol* **9**, 191–197.

Riesenfeld., C.S., Schoss, P.D. and Handelsman, J. (2004) Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet* **38**, 525–552.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S. *et al.* (2007) Northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5**, e77.

Schmeisser, C., Steele, H. and Streit, W.R. (2007) Metagenomics, biotechnology with non-culturable microbes. *Appl Microbiol Biotechnol* **75**, 955–962.

Schoss, P.P. and Handelsman, J. (2005) Metagenomics for studying unculturable microorganisms: cutting the gordian knot. *Genome Biol* **6**, 229.

Sleator, R.D. and Hill, C. (2008) New frontiers in probiotic research. *Lett Appl Microbiol* **46**, 143–147.

Sugimoto, A., Bigneil, D.E. and MacDonald, J.A. (2000) Global impact of termites on the carbon cycle and atmospheric trace gases. In *Termites: Evolution, Sociality, Symbiosis, Ecology*. ed. Abe, T., Bignell, D.E. and Higashi, M. pp. 409–435. Dordrecht: Kluwer Academic.

Tokuda, G. and Watanabe, H. (2007) Hidden cellulases in termites: revision of an old hypothesis. *Biol Lett* **3**, 336–339.

Tringle, S.G., Von Mering, C., Kobuyashi, A., Salamov, A.A., Chen, K. *et al.* (2005) Comparative metagenomics of microbial communities. *Science* **308**, 554–557.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R. and Gordon, J.I. (2007) The human microbiome project. *Nature* **449**, 804–810.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J. *et al.* (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **438**, 37–43.

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

Warnecke, F. and Hugenholtz, P. (2007) Building on basic metagenomics with complementary technologies. *Genome Biol* **8**, 231.

Warnecke, F., Luginbuhl, P., Ivanova, N., Ghassemian, M., Richardson, T.H. *et al.* (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**, 560–565.

Weibel, D.B., Diluzio, W.R. and Whitesides, G.M. (2007) Microfabrication meets microbiology. *Nat Rev Microbiol* **5**, 209–218.

