Available online at www.sciencedirect.com
ScienceDirect
Stud. Hist. Phil. Biol. & Biomed. Sci. 38 (2007) 834–846
Studies in History and Philosophy of Biological and Biomedical Sciences
www.elsevier.com/locate/shpsc

Metagenomics and biological ontology

John Dupré, Maureen A. O'Malley

Egenis, ESRC Centre for Genomics in Society, University of Exeter, Byrne House, St Germans Road, Exeter EX4 4PJ, UK

Abstract

Metagenomics is an emerging microbial systems science that is based on the large-scale analysis of the DNA of microbial communities in their natural environments. Studies of metagenomes are revealing the vast scope of biodiversity in a wide range of environments, as well as new functional capacities of individual cells and communities, and the complex evolutionary relationships between them. Our examination of this science focuses on the ontological implications of these studies of metagenomes and metaorganisms, and what they mean for common sense and philosophical understandings of multicellularity, individuality and organism. We show how metagenomics requires us to think in different ways about what human beings are and what their relation to the microbial world is. Metagenomics could also transform the way in which evolutionary processes are understood, with the most basic relationship between cells from both similar and different organisms being far more cooperative and less antagonistic than is widely assumed. In addition to raising fundamental questions about biological ontology, metagenomics generates possibilities for powerful technologies addressed to issues of climate, health and conservation. We conclude with reflections about process-oriented versus entity-oriented analysis in light of current trends towards systems approaches.

Keywords: Ontology; Metagenomics; Metagenome; Metaorganism; Multicellularity; System

1. Introduction

Life is commonly considered to be organized around the pivotal unit of the individual organism, which is traditionally conceived of as an autonomous cell or a group of coordinated cells with the same genome.[1] Hierarchies of other biological entities constitute and are constituted by organisms. Macromolecules are often placed at the bottom of the organism-constituting hierarchy, and they are succeeded by various sub-cellular and cellular levels of organization, including the tissues and organs of multicellular organisms. Above the level of organism rises the organism-constituted hierarchy, in which groups of organisms form ecological communities across space and lineages or species across time. Implicit in this hierarchical approach to the organization of life is an evolutionary timeline that runs from most primitive to most complex. Unicellular organisms are generally regarded as inhabiting the lower end of the complexity spectrum and large mammals are placed at the upper end.

E-mail addresses: j.a.dupre@ex.ac.uk (J. Dupré), m.a.o'malley@ex.ac.uk (M. A. O'Malley)

[1] If viruses (which are included in the usual definitions of microbes) are given organismal status, then an appropriate definition of organism could not insist on cellularity. Although there is a strong case for defining life as fundamentally cellular (most biologists would agree on this), the exclusion of viruses from the category of organisms raises deep conceptual problems due to the increasing recognition of the centrality of viruses to life processes at all scales.

1369-8486/$ - see front matter
doi:10.1016/j.shpsc.2007.09.001

All biologists and philosophers of biology know the difficulties of uniquely dividing different groups of organisms into species or genomes into genes, and both communities of investigators are divided about the inevitably of pluralism or the possibility of defining natural kinds. Our view is that these problems reflect a more fundamental difficulty, that life is in fact a hierarchy of processes (e.g.: metabolic, developmental, ecological, evolutionary) and that any abstraction of an ontology of fixed entities must do some violence to this dynamic reality. Moreover, while the mechanistic models that are constructed on the basis of these abstracted entities have been extraordinarily valuable in enhancing our understanding of life processes, we must remain aware of the idealized nature of such entities, and the limitations of analogies between biological process and mechanism (we shall say a bit more about this last point in the concluding section of the paper). Despite the philosophical challenges just mentioned to species and gene concepts, there are few similar doubts about the notions of unicellularity and multicellularity or, indeed, to the prima facie most unproblematic concept of all, the individual organism. We will raise these questions through a discussion of an emerging scientific perspective, metagenomics, and conclude with some further reflections on biological ontology.

2. Multicellularity

One of the most fundamental divisions in conceptions of the way life is organized is that between unicellular and multicellular lifeforms. The distinction between these two modes of life is usually seen as a major evolutionary transition (Maynard Smith & Szathmáry, 1995). It is standard to understand multicellularity as the differentiation of 'monogenomic' cells (differentiated cells possessing the same genome) and to think it pertains only to plants, animals and fungi (excluding yeasts). The most sophisticated forms of multicellularity are usually attributed to vertebrates because of the complex cell differentiation and coordination necessary throughout the lifespan of these animals. Non-vertebrates and plants are often considered to be less sophisticated multicellular organisms, perhaps because many of them can and do 'revert' to reproducing themselves by non-specialized cells. The complexities of incorporating well known symbiotic forms such as lichens and 'Thiodendron' mats cannot even be addressed within this scheme. This common way of ranking organisms causes problems, however, for an adequate understanding of multicellularity.

Many of the characteristics that are used to define multicellularity do not exclude unicellular life when proper attention is paid to bodies of research that illuminate cellular cooperation, developmental processes, competition and communication strategies amongst unicellular organisms (Shapiro & Dworkin, 1997; Shapiro, 1998; Wimpenny, 2000). Less formal criteria for designating unicellularity, such as visibility, are also problematic, because many supposedly unicellular forms of life live in groups visible to the naked eye, such as filaments, stromatolites (microbial mats that bind sediments and are best known in their fossilized forms) and biofilms, as well as some very large unicellular organisms. The research we refer to above reveals that microbial communities do indeed exhibit a multitude of multicellular characteristics. We have discussed several philosophical aspects of conceiving of microbial communities as multicellular organisms in earlier work (O'Malley & Dupré, 2007). Here, we will focus on a new genomics-based strategy called metagenomics that looks at microbial communities with fewer implicit preconceptions as to what constitutes a single organism.

3. Metagenomics

Metagenomics—also called environmental genomics, community genomics, ecogenomics or microbial population genomics—consists of the genome-based analysis of entire communities of complexly interacting organisms in diverse ecological contexts. The term 'metagenome' was first defined as the collective genome of the total microbiota of a specific environment by Jo Handelsman and colleagues in 1998 (Handelsman et al., 1998) and 'metagenomics' is now applied retrospectively to some earlier studies that preceded the coining of the label (e.g.: Stein et al., 1996). A brief history of where metagenomics fits in both recent molecular biology and microbiology in general will help clarify its scientific and philosophical implications.

3.1. Historical overview

Microbiology has traditionally been a discipline marginal to the mainstream biology of macroorganisms.[2] Its research interests entered the mainstream of biological thinking through genetics and molecular biology in the 1940s, when most molecular and genetic analysis was tied to microorganisms for technical reasons (Brock, 1990). The use of microbes, especially viruses and bacteria, as tools to understand genetic inheritance was based on the conviction and eventual demonstration that microbes possessed not only the same genetic material that multicellular organisms did, but also many similar biological processes, including reproductive ones. As single-gene studies gave way to whole-genome studies in the 1990s, the centrality of microbes and microbial knowledge to molecular biology was further reinforced, especially as the sequencing of prokaryotes and viruses rapidly advanced beyond that of macrobial organisms. Genomic studies of individual iso-

[2] In O'Malley & Dupré (2007) we suggest the word 'macrobe', in contrast to 'microbe', to refer to the diverse group of eukaryotic organisms generally thought of as multicellular.

lated microbes, while adding vast amounts of new information and understanding to comparative evolutionary biology in particular (Ward & Fraser, 2005), has so far provided only limited knowledge about function and about the many millions of uncultured microbial taxa (Nelson, 2003).

Attempts to remedy these shortfalls have resulted in the broadening of molecular microbiology's environmental scope. Ecological studies of microbes have historically been peripheral to both general ecology and mainstream microbiology, the former because of more general tendencies in biology and the latter because of the predominance of the pure culture paradigm (Brock, 1966; Costerton, 2004). Although microbial ecology had late nineteenth century roots in the work of Russian soil microbiologist, S. Winogradsky, and the founder of the famous Delft school of microbiology, M. W. Beijerinck, it took until the late 1960s for microbial ecology to gain real disciplinary recognition. Its unifying theme, whatever the methods used, is that microorganisms have to be understood in their ecological contexts (which include the context of other organisms), rather than as isolated individuals in artificial environments (Brock, 1987). Most microbes live preferentially in complex, often multispecies, communities such as biofilms, and as many as 99% of prokaryote taxa are not currently culturable in 'unnatural' laboratory conditions (Amann et al., 1995).

Taking an environmental approach enabled microbial genomics to extend beyond the sequencing of laboratory cultures of isolated microorganisms to the sequencing of DNA extracted directly from natural environments (Pace et al., 1985; Olsen et al., 1986; Amann et al., 1995). This move out of the laboratory vastly expanded the scope of the data collected as well as understandings of biodiversity and evolutionary relationships (Pace, 1997; Xu, 2006). It took the rest of the 1990s however, to overcome another associated narrowness of approach, in which very limited DNA sequences (often non-protein-coding) were used as markers of species and indicators of biodiversity. A focus on particular genes and an assumption they would be the same in each species—whatever the environment—limited the information gained about the physiological or ecological characteristics of the organisms (Ward, 2006; Tyson & Banfield, 2005; Rodríguez-Valera, 2002).

Metagenomics is conceived of as an approach through which such limitations can be transcended. Instead of individual genomes ('monogenomes') or single gene markers, metagenomics starts with large amounts of the DNA collected from microbial communities in their natural environments in order to explore biodiversity, functional interactions and evolutionary relationships. To understand the full range of metagenomic approaches and the ambitious agenda those practising it have set for themselves, we will briefly cover the various kinds of insight generated by the field so far. These range from the generation of catalogues of biodiversity to projects in evolutionary and functional metagenomics, and culminate in what we might call 'metaorganismal' metagenomics.

3.2. Biodiversity metagenomics

The primary focus of early metagenomics has been a better appreciation of biodiversity through the analysis of metagenomes or environmental DNA. Currently, metagenomicists take two approaches to the study of metagenomes. The most common practice consists of extracting DNA from environmental samples and cloning it in large-insert libraries.[3] These are screened for clone activity (particular functions expressed in the host cell) or specific gene sequences (Riesenfeld et al., 2004). Genes of interest continue to include ribosomal RNA genes, genes that have been favoured for phylogenetic analysis since the early days of molecular microbiology, and which continue to be used as phylogenetic 'anchors' for further analysis of diversity and function (Tringe & Rubin, 2005). Sampled environments include ocean waters of various depths and temperatures (DeLong et al., 2006; Grzymski et al., 2006; Béjà, Suzuki, et al., 2000), marine sediments (Hallam et al., 2004), agricultural soils (Rondon et al., 2000), the human gut (Manichanh et al., 2006) and mouth (Diaz-Torres et al., 2003), as well as human-made environments such as drinking-water valves (Schmeisser et al., 2003). These approaches deliver a mix of targeted and 'undirected' biological information about all levels of biological organization, from single genes and metabolic pathways to ecosystem activities (Riesenfeld et al., 2004).

The second approach is even more comprehensive and involves the random shotgun sequencing of small insert libraries[4] of all the DNA in an environmental sample. The two 'classic' examples include, first, a study that sequenced all the DNA from an environment with low species density, specifically a biofilm in the highly acidic, metal-rich runoff in a mine that exposed pyrite ore to air and water (Tyson et al., 2004). The genomes of only five taxa were identified in this community's DNA, the relative simplicity of which allowed two individual genomes to be reconstructed from the sequence data. Analysis of genes for metabolic pathways gave insights into the roles of community members in metabolic processes and changed the authors' understanding of how the community functioned (Tringe et al., 2005; Tringe & Rubin, 2005).

The second example involved the DNA from a considerably more complex oceanic community in the low nutrient Sargasso Sea (Venter et al., 2004). This catalogue of subsurface prokaryotic DNA is still the largest genomic dataset for any community and has been described as a

[3] DNA fragments of 100 kb and more can be propagated in bacterial artificial chromosomes (BACs) and up to 40 kb in fosmids (modified plasmids). These cloning vectors were originally used primarily for plant and animal genomics, and BACs were essential to the sequencing of the human genome.

[4] These libraries consist of very small fragments of DNA cloned in plasmids and amenable to high-throughput sequencing.

'megagenome' (Handelsman, 2004). Almost 70,000 divergent genes and 148 previously unknown phylotypes (taxa with divergent sequence in commonly used phylogenetic markers) were amongst this study's findings. Special database (GenBank) provisions had to be made so that the megagenomic data did not overwhelm all the ordinary monogenomic data contained in the databank and skew subsequent comparative analyses (Galperin, 2004). Venter's team continues to gather environmental DNA from oceans around the world (following the route of *The Beagle*) and has also commenced the 'Air Metagenome Project', which will sample and sequence the midtown Manhattan microbial communities in the air (Holden, 2005).

Viral metagenomics, which also focuses on shotgun sequencing of metagenomes, gives insight into the vast and previously untapped diversity of viral communities in, for example, near-shore marine environments (sediments and water column), and human and horse faeces (Breitbart et al., 2002, 2003, 2004; Edwards & Rohwer, 2005). These data, which indicate an even greater amount of biodiversity than that attributed to prokaryote communities, allow further hypotheses to be developed about the role of viral communities in the evolution and ecology of not only microbial communities but the entire natural world (Hamilton, 2006).

An epistemologically paradoxical but methodologically intriguing use of metagenomics was to overcome the technical challenges of sequencing ancient DNA—in the most well known case from a Pleistocene cave bear, *Ursus spelaeus*. By sequencing all the DNA in the sample, which included microbial, fungi, plant and animal contaminants, then comparing the metasequence against modern dog and bear sequences, it became possible to distinguish the small amount (<6%) of cave bear DNA in the sample (Noonan et al., 2005). With similar techniques applied to mammoth DNA (Poinar et al., 2006), the viability of ancient DNA sequencing via 'palaeometagenomics' now seems established and may result in the success of the 'Neanderthal Metagenome Project' (Rubin, in Pennisi, 2005). In these cases, the interest is obviously not the metagenome itself but an individual genome, and 'metagenomics' is clearly conceived of as no more than a bioinformatics technique.

There is much more to metagenomics than shallow catalogues of sequence and biodiversity, however. These inventories also give unique insights into microbial community structure and biogeography. They enable subtle understandings of ecophysiological characteristics of communities, in which adaptations to different environmental gradients result in different metabolic and morphological strategies (e.g.: capacities for movement) that spread vertically and horizontally through community members (DeLong et al., 2006). The genomic heterogeneity in environmental samples shows that the genomes of single isolated organisms can no longer be considered as typical of whole populations or species (Allen & Banfield, 2005). Rich as such understanding is, it is only the beginning of metagenomic analysis because community genome data is also the basis for more complex evolutionary understanding as well as for the investigation of community function and ecological dynamics.

3.3. Evolutionary metagenomics

Metagenomics has amplified insights into and questions about the genetic heterogeneity of populations and the genomic mosaicism of individuals. Understanding this genetic variability requires a deeper understanding of evolutionary processes and the mechanisms of genetic exchange and recombination (Falkowski & de Vargas, 2004). Metagenomics naturally aligns with an area of investigation that is sometimes called 'horizontal genomics' because both are concerned with the plethora of mobile genetic elements available to microbial communities and with the ways in which the metagenomic resources they inherit are shared and utilized (DeLong, 2004).

The metagenomic role of gene cassettes provides an interesting example of how such study is being pursued. Gene cassettes are intergenomically mobile genes that are integrated into genomes in units called integrons. Such cassettes usually carry genes for environmental emergencies, such as antibiotic assaults on prokaryote communities, rather than genes for everyday function. Cassettes, and the integron elements in the host genome that allow the cassettes to be inserted and expressed (or excised), are efficient mechanisms for the movement and expression of genes within and between species, and are implicated heavily in antibiotic resistance (Michael et al., 2004; Holmes et al., 2003; Rowe-Magnus et al., 2002). Gene cassettes were originally studied individually but a metagenomic perspective[5] allows them to be treated as a 'floating' evolutionary resource of high diversity and widespread activity that exists independently of individuals and is likely to have a high impact on bacterial genome evolution (Holmes et al., 2003; Michael et al., 2004).

Although the extent, types and precise effects on the metagenome of mobile resources (cassettes, as well as all the genetic material available for exchange to greater and lesser degrees by conjugation, transformation and transduction) still require much more research, the conceptual implications for evolutionary understanding are already powerful, particularly because such studies back up extensive work done on lateral gene transfer and recombination processes. Metagenomic analysis supports and extends the earlier unexpected findings of comparative microbial genomics, which contradicted the dominant eukaryo-cen-

[5] The metagenomic technique is somewhat different for gene cassettes because in this case PCR assays that target complete ORFs (specifically those flanked by the integration elements that are part of integrons) can be used to select environmental DNA prior to cloning and sequencing (see Stokes et al., 2001).

tric paradigm of vertical inheritance and mutation driven species divisions that give rise to a single tree of life (Allen & Banfield, 2005; Rodríguez-Valera, 2004; DeLong, 2002b; Doolittle, 2005). Rather than focusing on individual organismal lineages, such metagenomic studies enable a shift in scientific and philosophical attention to an overall evolutionary process in which diverse and diversifying metagenomes underlie the differentiation of interactions within evolving and diverging ecosystems. Conceptually, metagenomics implies that the communal gene pool is evolutionarily important and that genetic material can fruitfully be thought of as the community resource for a superorganism or metaorganism, rather than the exclusive property of individual organisms (Sonea & Mathieu, 2001).

3.4. Functional metagenomics

Most functional metagenomics involves screening library clones created from environmental DNA for functional activity and specific genes, particularly those for which function has already been established. Many metagenomic studies reconstruct metabolic pathways based on genome sequence by assigning functional roles to different taxa in the community based on those genes (Allen & Banfield, 2005; Tringe et al., 2005). These analyses are not only discovering new genes, but also revealing wholly unanticipated functions and mechanisms such as photobiology in oceanic bacteria. Genes for light-driven energy production (proteorhodopsin genes) were well known in halophilic or salt-loving archaea but never before suspected in oceanic bacteria until discovered (because one gene was serendipitously close to a 16S ribosomal RNA gene being sequenced for phylogenetic purposes) via metagenomic analysis (DeLong, 2005; Béjà, Suzuki et al., 2000). Further analysis involved the expression of these genes in a laboratory *E. coli*, and the correlation of proteorhodopsin with the light available at different ocean depths (Béjà et al. 2001; Béjà, Aravind et al. 2000). Venter's Sargasso Sea inventory also revealed the surprising abundance of proteorhodopsin genes in ocean waters.

This gene-based approach is most comprehensively captured by an innovative comparison of metagenomic data from microbial communities in farm soil and around the skeletons of decomposed whale carcasses in ocean waters (Tringe et al., 2005). Using bioinformatic techniques to predict the protein products of each metagenome, Tringe and colleagues compared their data to the acid mine drainage and Sargasso Sea metagenomes. They identified patterns of gene distribution that are specific to the environmental locations of the communities, thereby adding support to the argument for the importance of understanding microbial genome activity *in situ* (even though current techniques are biased towards the most common taxa). Rather than trying to reconstruct whole individual genomes from the metagenomic data, the study sought to elucidate the 'functional fingerprints'[6] of the DNA of complex communities in a variety of nutrient-rich environments.

Other functional studies of community genomes have uncovered versatile metabolic capacities,[7] such as the ammonia oxidizing activities of archaea in oceans and soils (a function previously known only in bacteria). The metagenomic identification of archaeal gene sequences involved in ammonia oxidation pathways provides a clear framework for deeper biochemical and physiological understanding of the role these mostly uncultured organisms play in the global nitrogen cycle (Hallam et al., 2006; Treusch et al., 2005; Francis et al., 2007). Amino acid analysis based on an Antarctic marine bacterial metagenome has lead to further development of physiological hypotheses about cold adaptation (Grzymski et al., 2006). Another interesting investigation focused on 'reverse methanogenesis' or methane consumption (the reverse of the better studied process of methane production or methanogenesis) in archaeal communities[8] in deep-sea sediments (Hallam et al., 2004). This study may have solved a biogeochemical puzzle of why seabed methane does not escape into the water and is one of the many examples of the potential social and economic relevance of metagenomics to environmental issues such as global warming.

Many of these studies of function however, are either of functions predicted from gene sequence on the basis of homology searches, or are based on low-throughput screening (Wellington et al., 2003; Johnston et al., 2005). Validating gene predictions was the bottleneck for monogenomic analysis, and there are fears this will be the case for metagenomics (Ward, 2006), especially if expression studies can only investigate limited gene activity per assay (Sebat et al., 2003). Several metagenomic research laboratories are now extending their analyses to the transcriptome (comprehensive gene expression in a particular set of conditions as measured by the abundance of mRNA transcripts) and proteome levels (total protein expression). In parallel with the terminology of metagenome, these objects of investigation are referred to as the metatranscriptome and the metaproteome. Community

[6] Tringe and colleagues used partially assembled metagenome fragments, which they called 'enviromental gene tags' or EGTs.

[7] Biotechnological hopes for applications arising from functional metagenomics are high, because the science opens up a previously concealed realm of microbial gene products (apparently inaccessible to standard cultivation techniques). Commercial applications of metagenomics focus on the discovery of 'novel natural products' such as enzymes, antibiotics and other drugs (Cowan et al., 2005; Lorenz & Eck, 2005; Courtois et al., 2003). Soils in particular are perceived to be rich sources of new biomolecules for industry and biomedicine (Daniel, 2004; Voget et al., 2003). In cases such as acid mine drainage, metagenomic data is anticipated to lead to more effective bioremediation techniques (Eyers et al., 2004; Pazos et al., 2003). Even broader insight and applications are sought from metagenomics to address global warming (Committee on Metagenomics, 2007).

[8] These communities were physically simplified before metagenomic analysis (to a greater extent than size filters during sampling would achieve), so are not truly 'natural' samples.

microarray analyses of gene transcription are already extending bioinformatic inferences of function. Using microarray technology for studying metagenome expression and regulation is challenging but possible, according to early efforts (Wu et al., 2001; Dennis et al., 2003; Poretsky et al., 2005; Zhou, 2003).

Metaproteomic analysis is the next step towards achieving better understanding of functional gene expression in a range of environments and specific conditions (Wilmes & Bond, 2006; Powell et al., 2005; Ram et al., 2005; Kan et al., 2005; Schulze et al., 2005). Using mass spectrometry, Ram and colleagues analysed the metaproteome of the same acid mine drainage biofilm as in Tyson et al.'s metagenomic study. Novel gene products as well as expected ones were quantified, and inferences made from metagenomic data were supported. Most crucially, a key iron-oxidizing enzyme involved in acid production was identified, which gave much greater depth to understandings of that particular ecosystem. Metaproteomics is most commonly used for low complexity environments and it is still very hard to apply to more complex ones or to extend the functional understanding gained from these early analyses (Wilmes & Bond, 2006). Nevertheless, existing techniques plus combinations and extensions of them are being used to understand *in situ* function, and users are optimistic about further technical advances (Wilmes & Bond, 2006).

Metametabolomics, the community-based version of individual organism metabolomics (the study of small molecules or metabolites in a particular physiological state of a cell) is only just beginning, and one of the first studies focuses on the metabolic networks of the microbial community in the human gut and the digestive capacities provided by microbes to humans (Gill et al., 2006). Such studies also investigate 'co-metabolomes' or the metabolites that can only be produced by host–microbial interactions (Nicholson et al., 2005). As well as intracellular metabolites, data on extracellular signalling molecules are needed, if communication and coordination processes within communities are to be understood (Buckley, 2004). Integrated analyses of metagenomes, metatranscriptomes, metaproteomes and metametabolomes will be necessary for the ultimate metagenomic analysis, which will involve the investigation of all levels of community activity *in situ*.

3.5. Metaorganismal metagenomics, or microbial systems biology

The next ambitiously anticipated phase of metagenomics is microbial systems science, in which complex biological networks across multiple hierarchical levels are to be analysed by interdisciplinary teams (DeLong, 2002a; Rodríguez-Valera, 2004; Buckley, 2004). Systems biology more generally is now the most rapidly proliferating and heavily funded area of biology. It consists of large groups of molecular and cell biologists, mathematicians and bioinformaticians attempting to integrate huge bodies of 'omic' data with the aim of predicting and intervening in multiple levels of biological activities in wide ranging environmental conditions. Although discussion of how community-based microbial systems biology will be accomplished has so far been limited, it is clear that a 'systems ecological' perspective would meld a molecular approach with a synthesizing perspective on ecosystems (including biogeochemical analysis) and would thereby extend the focus of current systems biology on intracellular interactions (DeLong, 2005; Allen & Banfield, 2005; Doney et al., 2004). More data is not the only requirement for a systems-biological approach to microbial communities. Advances in modelling and *in silico* simulation as well as serious interdisciplinary cooperation will be vital to the field's development (Lovley, 2003). These requirements mirror challenges to systems biology generally (O'Malley & Dupré, 2005).

Despite the limitations of currently available tools,[9] a great deal of metaorganismal analysis is already under way, both of purely microbial systems and of mixed microbial–macrobial systems. The dynamics between plant and nitrogen-fixing microbial communities or between squid and light-generating bacterioplankton are well studied examples of associations in which there are mutual influences on gene expression and developmental processes (Xu & Gordon, 2003). Particularly striking is the growing understanding that symbiotic bacteria are required for the proper development of many vertebrates. It was recently reported, for example, that environmentally acquired digestive tract bacteria in zebrafish regulate the expression of 212 genes (Rawls et al., 2004; Bates et al., 2006). In fact, for the majority of mammalian organism systems that interact with the external world—the integumentary (roughly speaking, the skin), respiratory, excretory, reproductive, immune, endocrine, and circulatory systems—there is strong evidence for the coevolution of microbial consortia in varying levels of functional association (McFall-Ngai, 2002). Germfree (gnotobiotic) rodent studies indicate that removing all or part of a mammalian microbiome leads not only to abnormal physiological development, but also morphological abnormalities and immunological depression (Berg, 1996). The roots of plants lie in the midst of some of the most complex multispecies communities on the planet. Bacteria and filamentous fungi not only form dense cooperative communities in the soil surrounding plant roots (often described as the rhizosphere), but are also found within the roots themselves. The dynamic interactions—both intra- and extra-cellular—between roots, fungi and bacteria provide nutrients, control pathogens and structure function, growth and community composition (Perotto & Bonfante, 1997; Barea et al., 2005). These interactions, many of them obligate,

[9] See Appendix.

break down the boundary between plant and non-plant, and also serve to illustrate that at the heart of every interface between multicellular eukaryotes and the external environment lies a complex multispecies microbial community.

Even if we insist on retaining a fundamentally monogenomic conception of the human organism, there are multiple ways in which studies of macrobe–microbe interactions can improve our understanding of human biology. A human body is a symbiotic system composed of 90% prokaryotes and 10% human cells (Savage, 1977), a ratio that is made more dramatic when viruses (including the bacteriophages hosted by every prokaryote) are considered. Humans and their microbial partners (the microbiome) form a highly coordinated system that appears to be the product of coevolution. Microorganismal studies of human–microbial interactions conceive microbial communities to be at least as important for human health as they are for disease. Indeed, it may turn out that diseases caused by microbial pathogens are best seen not so much as an invasion by a hostile organism, but rather as a kind of holistic dysfunction of the microbiome.

Analyses of the diversity of the human gut metagenome (e.g.: Gill et al., 2006; Zoetendel et al., 2006) allow richer understandings of how this many celled 'organ' functions and maintains human life and health. Transcriptional studies further illuminate how this community modulates host gene expression and communicates with host cells (Hooper et al., 2001). Environmental perturbations of human microbiomes, such as may be caused by dietary changes or by antibiotic use, are linked to heart disease, cancers, obesity, asthma and diabetes (Ordovas & Mooser, 2006; Bäckhed et al., 2004, 2005; Ley et al., 2005, 2006). Human drug response is highly dependent on molecular interactions with microbially generated molecules (Nicholson et al., 2005). The human immune system recognizes a huge variety of prokaryotes organised in different communities as 'self' rather than non-self (the objects of immunological defence). Together, commensal prokaryote communities and the human host form an integrated immune system that provides benefits to all the organisms involved (Kitano & Oda, 2006a, b).

Such 'self-extending symbioses' argue Kitano & Oda (2006a), are the evolutionary norm because they are highly adaptive and robust against environmental perturbation. All of these insights, whether delivered by metagenomics or more general molecular microbiology, encourage us to see any animal as a composite of all three domains (bacteria, archaea[10] and eukaryotes) and our own primary genome as a metagenome of microbial and human DNA (Gordon et al., 2005). It is even suggested that humans and other animals could be regarded as 'advanced fermenters', the main role of which is to house, nourish and assist the reproduction of an enormous array of microbes (Nicholson et al., 2005). The original human genome sequencing projects were, from this perspective, about only a tiny and unrepresentative complement of our genes: a limitation that may ultimately be remedied by a human metagenome project (Relman & Falkow, 2001).

If we think about symbioses (the fundamental condition of life) in more general terms than those symbioses limited to the outer surfaces of animal bodies, then metaorganismal metagenomics has very obvious contributions to make to understandings of ecosystem diversity, function and dynamics. The applications of microbially grounded systems ecology are extensive. Bioremediation, one of the great hopes of genetic engineering, applied genetics and microbiology, is unlikely to be successful without taking a systems-based approach (Cases & de Lorenzo, 2005). The problem of previous approaches and the reason for their failures, argue some commentators, is that the genetic engineering of isolated strains took too limited a perspective on microbial interactions. This perspective can only be advanced upon by a multilevel 'eco-engineering' approach to the metabolic activities of indigenous microbial communities in their natural environments. There are already promising signs from systems-biologic approaches to microbial bioremediation of environmental pollutants (Pazos et al., 2003).

It is becoming increasingly clear that a range of fundamental questions about life on this planet will find their answers only with advances in system based understandings of microbial communities in global environments (Hunter-Cevera et al., 2005). Will warming oceans disturb the world's primary oxygen producers, the marine cyanobacteria *Prochlorococcus* and endanger oxygen dependent lifeforms (everything apart from prokaryotes[11])? Will the thawing of the polar icecaps lead to intensified global warming as dormant methanogenic prokaryotes become active and release more methane (a contributor to global warming)? Will global changes in human habitat and diet modify the microbiome in human bodies and have significant health consequences? Will antibiotics still be effective in twenty years or will we see the return of high fatality rates from infections such as tuberculosis and pneumonia with the worldwide circulation of antibiotic resistant genes in the microbial metacommunity? All of these are questions that metagenomics is beginning to address, and it offers real hope that they will be answered (Committee on Metagenomics, 2007).

Dynamic system perspectives, which take networks of biological activity in ecological settings as their objects of

[10] Much less is known about archaeal symbionts than bacterial ones, but numerous recent studies are investigating their presence and role in all organisms, including humans. Methanogens (methane generating archaea) are the most frequently found symbionts, in host environments such as the human gut and mouth (Lange et al., 2005). There are still no known pathogenic archaea, but the limitations of detection tools may be the most appropriate interpretation of this non-finding (Eckberg et al., 2003).

[11] It appears unlikely that there are any truly obligate anaerobic eukaryotes (Bryant, 1991; Lloyd, 2004).

study, have some of the most interesting implications for the philosophy of biology and its traditional considerations about the units of fundamental ontological importance. In system-level understandings of microbial communities, the metaorganism is conceived of as deriving causal powers from the interactions of the individual components from which it is constituted. At the same time, however, those components are themselves understood to be controlled and coordinated in various ways by the causal capacities of the metaorganism. Although metagenomics might be far away from the full implementation of such a causally dialectical research programme, every effort to think in systems-oriented ways and find techniques and approaches to answer the questions raised by this perspective inevitably points to the necessity of integrating multiple levels of separately realized biological insight. As mentioned earlier, this is one of the main challenges being addressed by systems biology generally.

4. Concepts of metagenome and metaorganism

For many of the field's practitioners, metagenomics is a technique that provides access to otherwise inaccessible microbial communities. We believe that a more complex line of thought informs metagenomics and has probably informed it since its inception (e.g.: Rodríguez-Valera, 2004). This perspective takes very seriously the proposal that metagenomes are communal resources and that the entity to which the resource is available is a coordinated, developing, multifunctional, multicellular organism composed of large numbers of cells of different varieties and capabilities, able to work in ways in which the collectivity regulates the functions of individuals[12]. Individual organisms, from this viewpoint, are an abstraction from a much more fundamental entity. A lot of evidence for this perspective has been generated by a variety of methods over the last two decades (e.g.: Shapiro, 1997; Kolenbrander, 2000; Kaiser, 2001). Metagenomic analysis is able to take these avenues of research into account and begin to treat the metagenome as the concept that most effectively captures the ways in which microbes survive and flourish in such a remarkable range of diverse environments.

Ultimately, metagenomics is about 'the community of all communities' on this planet, and, uncomfortable as the notion may be for many philosophers, for some practitioners metagenomics is perceived to be moving us inexorably in the direction of a Gaia-like concept of the world (Committee on Metagenomics, 2007). Taking the interconnectedness of ecosystems into account does not mean taking on board all of Lovelock's philosophical notions or empirical predictions (in fact, going back to Baas Becking, who applied the word in 1931, is probably a better idea[13]), but it does mean that adequate understandings of climate change, global human health and international economic prosperity will all depend on metagenomic knowledge.

4.1. Competition versus cooperation

It is plausible to conclude that the fundamental activity of cells, beyond self-organization and maintenance, is to form cooperative associations in a plurality of forms. This may suggest that the organization of life is determined not by competition but by the ability to cooperate (albeit competitively). To put another way, although there is little doubt that competition and selection have been essential to the evolutionary process, it may be that the main respect in which organisms (or cells) have competed has been with respect to their ability to cooperate in complex single- or multi-species communities. If this is the case, then we might expect—contrary to the orthodox evolutionary view that altruism is exceptional and requires special explanation[14]—that the norm among organisms is a disposition to act for the benefit of other organisms or cells. Darwin was right about the general picture of the organic environment being fundamental to the determination of fitness, but his account of the relationships between organisms in those environments needs to be supplemented.

4.2. Process versus entities

A final crucial point about metaorganisms is that they are paradigmatically dynamic entities and therefore very clear illustrations of the ultimate necessity of a process-oriented approach to biological investigation. None of the entities that constitute organisms, or which organisms constitute, are static. Genomes, cells, and ecosystems are in constant interactive flux: subtly different in every iteration, but similar enough to constitute a distinctive process. The greatest significance of this point is perhaps that its appreciation will prevent us from taking too literally mechanistic models of biological processes. A good machine starts with all its parts precisely constructed to interact together in the way that will generate its intended functions. The technical manual for a car specifies exactly the ideal state of every single component. Though the parts of a machine are not unchanging, of course, their changes constitute a relentless, unidirectional trend towards failure. As friction, corrosion, and so on gradually transform these components from their ideal forms, the function of the car deteriorates. For a while these failing components can be replaced with replicas, close to the ideal types specified in the manual, but eventually too many parts will have deviated too far from

[12] There are, of course, some microbiologists who explicitly reject this perspective and argue for the retention of a single-organism focus, even if metagenomics has proved its use as a tool (e.g.: Buckley, 2004).

[13] See Baas Becking (1931) as well as L. Margulis's work on this topic (e.g., Margulis, 1998).

[14] Altruism and its interpretation have been much deliberated in the biological literature. See Lyon (this issue) for a discussion of altruism and cooperation in relation to microbes.

the ideal, and the car will be abandoned, crushed, and recycled. No such unidirectional tendency towards failure characterizes biological processes: although perhaps all organisms die in the end, many exhibit high levels of stability for centuries and millennia. This is not, however, a stability based on parts so durable that they take centuries to deteriorate, but rather the dynamic stability of processes that constantly recreate or maintain their essential constituents. Both the promise and the challenge of systems-biologic approaches is that they offer the hope of providing techniques for the modelling of such self-maintaining dynamic systems.

A further philosophical point about dynamically self-sustaining systems as opposed to mechanistic[15] ones, is that in the latter causation can be seen to run not merely upwards from part to whole, but also downwards from whole to part[16]. The behaviour of individual cells, for instance, whether in multicellular eukaryotes or complex microbial systems, is in fundamental respects determined by the features of the system of which it is part. We suggest that if useful object-like abstractions can successfully be produced from the flux of dynamic biological processes, then this reflects the fact that these 'objects' are temporarily stable nexuses in the flow of upward and downward causal interaction. So, for instance, a gene is a part of the genome that is a target for external (that is, cellular) manipulation of genome behaviour and, at the same time, carries resources through which the genome can influence processes in the cell more broadly. Since the analysis of biological systems into entities is not determinate, for some purposes of inquiry a monogenomic organism is the most appropriate, whereas for others the appropriate focus is on polygenomic systems. Answering the question, 'What is an organism?', requires seeing that there is a great variety of ways in which cells—sometimes genomically homogeneous, sometimes not—combine to form integrated biological wholes. The concept of a multicellular organism is, therefore, a far more complex and diverse one than the simple, straightforward category we outlined in the introduction.

5. Conclusion

A specific and central philosophical aim of this paper is to encourage scepticism about a concept that has often been treated as self-evident and unproblematic in theoretical biology: the monogenomic concept of an organism[17]. Its 'obviousness' is sustained by the distinction between unicellularity and multicellularity, where it is assumed that each organism has exactly one genome. If that genome appears in sharply differentiated but functionally interrelated cells, we think of the organism as multicellular; if cells with that genome seem more or less functionally homogeneous and more or less contingently functionally interrelated to other cells, we call it unicellular.

However, if we think of the organism as being simply whichever cooperative systems of cells are most usefully recognized for exploring biological function, then the assumption of 'one genome, one organism' starts to look like a poorly grounded dogma. As highly organized microbial communities such as biofilms illustrate, genomically diverse groups of cells can form organism-like communities (Stoodley et al., 2002; Costerton et al., 1995; O'Toole et al., 2000). Moreover, in understanding the functions of even such paradigmatically multicellular organisms as ourselves, many perspectives (including those of nutrition, development and immune response) suggest that the relevant system is actually a much broader one, which includes many genomically different kinds of cells. Indeed, as we have noted, the majority of cells in the systems we think of as humans are actually microbial rather than what we have conventionally thought of as human cells. No doubt that which we are describing as a dogma is seen by many biologists to be firmly grounded in a conception of evolution associated with genetically isolated lineages. We have also argued however (and in more detail elsewhere, O'Malley & Dupré, 2007), that this view of evolution is itself based on a perspective that is dangerously macrobocentric. The prevalence of horizontal genetic transfer makes it seem increasingly implausible to see the evolution of microbes as fitting within this isolated lineages model, and indeed suggests that evolving units may turn out quite typically to be multispecies communities. Proper attention to microbes may force us to rethink some fundamental ideas about the evolutionary process.

Philosophers of biology have for many years raised concerns about central ontological categories: most notably the species and more recently the gene. We are urging that the organism should join this group of problematic biological categories. We also propose here a diagnosis of these ontological problems. Analysing biological processes into things is necessarily to make an abstraction. Life is not composed in a machine-like way out of unchanging individual constituents. Genomes, cells, organisms, lineages are all assemblages of constantly changing entities in constant flux. Unlike the case of a machine, the stability of life processes is not maintained by the constant interactions of unchanging parts, but by dynamic, self-sustaining and self-repairing processes. There is no doubt that mechanistic investigations of life processes have provided profound insights, and this very fact is enough to show that quasi-mechanistic elements are fundamental constituents of life

[15] Or at any rate machine-like ones. The concept of a mechanism developed recently by some philosophers (Machamer et al., 2000; Craver 2005) is in many ways well suited to the representation of dynamic biological systems.

[16] Craver and Bechtel (forthcoming) provide an interesting but rather different perspective on top-down causation in complex systems from that assumed here.

[17] The concept is not wholly monolithic because of known exceptions, such as genomic mosaicism.

processes. But there are limits to how far conventional mechanistic investigations can take us in understanding the dynamic stability of processes at this hierarchy of different levels. Such understanding will require models that incorporate both the capacities provided by mechanistic or quasi-mechanistic constituents, and the constraints and causal influences provided by properties of the wider systems of which these constituents are parts. It remains to be seen how far we have the abilities to construct such models, but this is what enthusiasts for the rapidly growing project of dynamic metaorganismal metagenomics and systems biology should be aiming to offer.

Acknowledgement

We thank the participants at the Philosophical and Social Dimensions of Microbiology Workshop (University of Exeter, July 2007) for discussion and ideas, and our two referees for helpful comments on the paper. We gratefully acknowledge research funding from the UK Arts and Humanities Research Council, and workshop funding from the Wellcome Trust. John Dupré did part of the work for this paper while holding the Spinoza chair in Philosophy at the University of Amsterdam and would like to thank the Department of Philosophy for hospitality and support. The research in this paper is part of the programme of the Economic and Social Research Council's Centre for Genomics in Society (Egenis).

Appendix. Problems and limitations of metagenomic analysis

At present, metagenomics is in an intensive discovery phase not unlike that of early single-organism genomics. It is more concerned with constructing inventories of environmental sequence and gene products than ready to give extensive insight into ecosystem function and physiology (Schloss & Handelsman, 2003; Chen & Pachter, 2005). Even if inventories were the only aim, the full metagenome sequence of the most complex and diverse communities (especially in soils) is still beyond the reach of current technologies because of the size and complexity of soil-based communal genomes, which require formidably high numbers of clones and sequence coverage to accurately represent their genetic composition (Riesenfeld et al., 2004). Shallower sequencing coverage means lower representation of less common taxa and functions (Foerstner et al., 2006). Various technical biases in regard to sampling, lysing cells for DNA extraction, cloning and expression systems are all being worked around and may eventually be overcome (Handelsman et al., 2002; Streit & Schmitz, 2004; Liles et al., 2003; Béjà, 2004).

There are additional sources of bias imposed by technological limitations. The extension of tools devised originally for individual genome and proteome analysis is not unproblematic. Reconstructing individual genomes from shotgunned metagenomes (which would accord the latter more reliability) is very difficult, especially in complex communities with a lot of genomic microheterogeneity (DeLong, 2005). New assembly algorithms and more extensive sequencing coverage are amongst the potential solutions (Edwards & Rohwer, 2005; Allen & Banfield, 2005; Chen & Pachter, 2005), as well as combinations of small-insert and large-insert library construction (DeLong, 2005). Comparisons of metagenomic libraries with monogenomic data are still crude, because of the low diversity of monogenomic data and (possibly) inadequate search tools. While comparative metagenomics, or the comparison of the genomic diversity and activity of different microbial communities has already begun (Tringe et al., 2005), it is currently shallow. Likewise, assigning functional roles to large numbers of environmental proteins is still impossible, which restricts understanding of metabolic activity in communities (Allen & Banfield, 2005). Metagenomic expression methods developed so far analyse only bacterial gene expression because they are restricted to *E. coli* based techniques (de Lorenzo, 2005). Many metagenomic studies are more 'proof of principle' than wholly successful extensions of existing tools, but the potential rewards for overcoming such challenges are considered to provide more than enough motivation for investing time and effort in their development (Chen & Pachter, 2005; Nicholson et al., 2004).

System-oriented analysis of the complex functions of microbial communities in their environments is still rudimentary (DeLong, 2005; Ram et al., 2005; Handelsman, 2004; Torsvik & Øvreås, 2002). Advances in the integration of 'omic' data modeling techniques for single organism, lab based, monogenomic systems biology (Joyce & Palsson, 2006; Palsson, 2000) are highly likely to benefit metaorganismal metagenomic analyses. The development of a systems-based molecular approach to microbial communities will also require tighter integration of microbial ecology, population genomics, phylogeny and biogeochemistry (DeLong, 2005; Rodríguez-Valera, 2002). More hypothesis-driven approaches (combining, for example, functional and sequencing aspects of metagenomics in the same investigation) and supplementary data from traditional microbiological techniques (e.g.: culturing, microscopy and other visualization techniques) are seen as desirable (Schloss & Handelsman 2005; Ward, 2006). If metagenomic-based systems biology progresses even a little, it will revolutionize the way in which life is investigated.

References

Allen, E. E., & Banfield, J. F. (2005). Community genomics in microbial ecology and evolution. *Nature Reviews Microbiology*, *3*, 489–498.

Amann, R. I., Ludwig, W., & Schleifer, K.-H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbial Reviews*, *59*, 143–169.

Baas Becking, L. G. M. (1931). *Gaia of leven en aarde*. The Hague: Nijhoff.

Bäckhed, F., Ding, H., Wang, T., Hooper, L. V., Koh, G. Y., Nagy, A., Semenkovich, C. F., & Gordon, J. I. (2004). The gut microbiota as an environmental factor that regulates fat storage. *Proceedings of the National Academy of Sciences*, *101*, 15718–15723.

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

Barea, J.-M., Pozo, M. J., Azcón & Azcón-Aguilar, C. (2005). Microbial cooperation in the rhizosphere. *Journal of Experimental Botany*, *56*, 1761–1778.

Bates, J. M., Mittge, E., Kuhlman, J., Baden, K. N., Cheesman, S. E., & Guillemin, K. (2006). Distinct signals for the microbiota promote different aspects of zebrafish gut differentiation. *Developmental Biology*, *297*, 374–386.

Béjà, O. (2004). To BAC or not to BAC: Marine ecogenomics. *Current Opinion in Biotechnology*, *15*, 187–190.

Béjà, O., Aravind, L., Koonin, E. V., Suzuki, M. T., Hadd, A., Nguyen, L. P., Jovanovich, S. B., Gates, C. M., Feldman, R. A., Spudich, J. L., Spudich, E. N., & DeLong, E. F. (2000). Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science*, *289*, 1902–1906.

Béjà, O., Spudich, E. N., Spudich, J. L., Leclerc, M., & DeLong, E. F. (2001). Proteorhodopsin phototrophy in the ocean. *Nature*, *411*, 786–789.

Béjà, O., Suzuki, M. T., Koonin, E. V., Aravind, L., Hadd, A., Nguyen, L. P., Villacorta, R., Amjadi, M., Garrigues, C., Jovanovich, S. B., Feldman, R. A., & DeLong, E. F. (2000). Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environmental Microbiology*, *2*, 516–529.

Berg, R. D. (1996). The indigenous gastrointestinal microflora. *Trends in Microbiology*, *4*, 430–435.

Breitbart, M., Felts, B., Kelley, S., Mahaffy, J. M., Nulton, J., Salamon, P., & Rohwer, F. (2004). Diversity and population structure of a near-shore marine-sediment viral community. *Proceedings of the Royal Society London*, *271B*, 565–574.

Breitbart, M., Hewson, I., Felts, B., Mahaffy, J. M., Nulton, J., Salamon, P., & Rohwer, F. (2003). Metagenomic analyses of an uncultured viral community from human feces. *Journal of Bacteriology*, *185*, 6220–6223.

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., Azam, F., & Rohwer, F. (2002). Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences*, *99*, 14250–14255.

Brock, T. D. (1966). *Principles of microbial ecology*. Englewood Cliffs, NJ: Prentice-Hall.

Brock, T. D. (1987). The study of microorganisms in situ: Progress and problems. *Symposium of the Society for General Microbiology*, *41*, 1–17.

Brock, T. D. (1990). *The emergence of bacterial genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Bryant, C. (Ed.). (1991). *Metazoan life without oxygen*. London: Chapman & Hall.

Buckley, M. R. (2004). *Systems biology: Beyond microbial genomics*. Washington, DC: American Academy of Microbiology.

Cases, I., & de Lorenzo, V. (2005). Genetically modified organisms for the environment: Stories of success and failure and what we have learned from them. *International Microbiology*, *8*, 213–222.

Chen, K., & Pachter, L. (2005). Bioinformatics of whole-genome shotgun sequencing of microbial communities. *PLoS Computational Biology*, *1*(2), 106–112. doi:10.1371/journal.pcbi.0010024, e24.

Committee on Metagenomics: Challenges and Functional Applications, National Research Council. (2007). *The new science of metagenomics: Revealing the secrets of our microbial planet*. Washington, DC: National Academies Press.

Costerton, B. (2004). Microbial ecology comes of age and joins the general ecology community. *Proceedings of the National Academy of Sciences*, *101*, 16983–16984.

Costerton, J. W., Lewandowski, Z., Caldwell, D. E., Korber, D. R., & Lappin-Scott, H. M. (1995). Microbial biofilms. *Annual Review of Microbiology*, *49*, 711–745.

Courtois, S., Cappellano, C. M., Ball, M., Francou, F.-X., Normand, P., Helynck, G., Martinez, A., Kolvek, S. J., Hopke, J., Osburne, M. S., August, P. R., Nalin, R., Guérineau, M., Jeannin, P., Simonet, P., & Pernodet, J.-L. (2003). Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Applied and Environmental Microbiology*, *69*, 49–55.

Cowan, D., Meyer, Q., Stafford, W., Muyanga, S., Cameron, R., & Wittwer, P. (2005). Metagenomic gene discovery: Past, present and future. *Trends in Biotechnology*, *23*, 321–329.

Craver, C. F. (2005). Beyond Reductionism: Mechanisms, multifield integration, and the unity of science. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *36C*, 373–396.

Craver, C. F., & Bechtel, W. (forthcoming). Top-down causation without top-down causes. *Biology and Philosophy*. (Available at doi:10.1007/s10539-006-9028-8)

Daniel, R. (2004). The soil metagenome—a rich resource for the discovery of novel natural products. *Current Opinion in Biotechnology*, *15*, 199–204.

DeLong, E. F. (2002a). Towards microbial systems science: Integrating microbial perspective from genomes to biomes. *Environmental Microbiology*, *4*, 9–10.

DeLong, E. F. (2002b). Microbial population genomics and ecology. *Current Opinion in Microbiology*, *5*, 520–524.

DeLong, E. F. (2004). Microbial population genomics and ecology: The road ahead. *Environmental Microbiology*, *6*, 875–878.

DeLong, E. F. (2005). Microbial community genomics in the ocean. *Nature Reviews Microbiology*, *3*, 459–469.

DeLong, E. F., Preston, C. M., Mincer, T., Rich, V., Hallam, S. J., Frigaard, N.-J., Martinez, A., Sullivan, M. B., Edwards, R., Brito, B. R., Chisholm, S. W., & Karl, D. M. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science*, *311*, 496–503.

de Lorenzo, V. (2005). Problems with metagenomic screening. *Nature Biotechnology*, *23*, 1045.

Dennis, P., Edwards, E. A., Liss, S. N., & Fulthorpe, R. (2003). Monitoring gene expression in mixed microbial communities by using DNA microarrays. *Applied and Environmental Microbiology*, *69*, 769–778.

Diaz-Torres, M. L., McNab, R., Spratt, D. A., Villedieu, A., Hunt, N., Wilson, M., & Mullany, P. (2003). Novel tetracycline resistance determinant from the oral metagenome. *Antimicrobial Agents and Chemotherapy*, *47*, 1430–1432.

Doney, S. C., Abbott, M. R., Cullen, J. J., Karl, D. M., & Rothstein, L. (2004). From genes to ecosystems: The ocean's new frontier. *Frontiers in Ecology and Environment*, *2*, 457–466.

Doolittle, W. F. (2005). If the tree of life fell, would we recognize the sound? In J. Sapp (Ed.), *Microbial phylogeny and evolution: Concepts and controversies* (pp. 119–133). Oxford: Oxford University Press.

Eckberg, P. B., Lepp, P. W., & Relman, D. A. (2003). Archaea and their potential role in human disease. *Infection and Immunity*, *71*, 591–596.

Edwards, R. A., & Rohwer, F. (2005). Viral metagenomics. *Nature Reviews Microbiology*, *3*, 504–510.

Eyers, L., George, I., Schuler, L., Stenuit, B., Agathos, S. N., & El Fantroussi, S. (2004). Environmental genomics: Exploring the unmined richness of microbes to degrade xenobiotics. *Applied Microbiology and Biotechnology*, *66*, 123–130.

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

Foerstner, K. U., von Mering, C., & Bork, P. (2006). Comparative analyses of environmental sequences: Potential and challenges. *Philosophical Transactions of the Royal Society*, *361B*, 519–523.

Francis, C. A., Beman, J. M., & Kuypers, M. M. M. (2007). New processes and players in the nitrogen cycle: The microbial ecology of anaerobic and archaeal ammonia oxidation. *International Society for Microbial Ecology Journal*, *1*, 19–27.

Galperin, M. Y. (2004). Metagenomics: From acid mine to shining sea. *Environmental Microbiology*, *6*, 543–545.

Gill, S. R., Pop, M., DeBoy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., Gordon, J. I., Relman, D. A., Fraser-Liggett, C. M., & Nelson, K. E. (2006). Metagenomic analysis of the human distal gut microbiome. *Science*, *312*, 1355–1359.

Gordon, J. I., Ley, R. E., Wilson, R., Mardis, E., Xu, J., Fraser, C. M., & Relman, D. A. (2005). *Extending our view of self: The Human Gut Microbiome Initiative (HGMI)*. http://www.genome.gov/Pages/

Research/Sequencing/SeqProposals/HGMIseq.pdf. (Accessed 24 June 2007)

Grzymski, J. J., Carter, B. J., DeLong, E. F., Feldman, R. A., Ghadiri, A., & Murray, A. E. (2006). Comparative genomics of DNA fragments from six Antarctic marine planktonic bacteria. *Applied and Environmental Microbiology*, *72*, 1532–1541.

Hallam, S. J., Mincer, T. J., Schleper, C., Preston, C. M., Roberts, K., Richardson, P. M., & DeLong, E. F. (2006). Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. *Publis Library of Science: Biology*, *4*(4), e95.

Hallam, S. J., Putnam, N., Preston, C. M., Detter, J. C., Rokhsar, D., Richardson, P. M., & DeLong, E. F. (2004). Reverse methanogenesis: Testing the hypothesis with environmental genomics. *Science*, *305*, 1457–1462.

Hamilton, G. (2006). The gene weavers. *Nature*, *441*, 683–685.

Handelsman, J. (2004). Metagenomics: Application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews*, *68*, 669–685.

Handelsman, J., Liles, M., Mann, D., Riesenfeld, C., & Goodman, R. M. (2002). Cloning the metagenome: Culture-independent access to the diversity and functions of the uncultivated microbial world. *Methods in Microbiology*, *33*, 241–255.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chemistry and Biology*, *5*, R245–R249.

Holden, C. (2005). Life in the air. *Science*, *307*, 1558.

Holmes, A. J., Gillings, M. R., Nield, B. S., Mabbutt, B. C., Nevalainen, K. M. H., & Stokes, H. W. (2003). The gene cassette metagenome is a basic resource for bacterial genome evolution. *Environmental Microbiology*, *5*, 383–394.

Hooper, L. V., Wong, M. H., Thelin, A., Hansson, L., Falk, P. G., & Gordon, J. I. (2001). Molecular analysis of commensal host–microbial relationships in the intestine. *Science*, *291*, 881–884.

Hunter-Cevera, J., Karl, D., & Buckley, M. (2005). *Marine microbial diversity: The key to earth's habitability*. Washington, DC: American Academy of Microbiology.

Johnston, A. W. B., Li, Y., & Ogilivie, L. (2005). Metagenomic marine nitrogen fixation—feast or famine? *Trends in Microbiology*, *13*, 416–420.

Joyce, A. R., & Palsson, B. Ø. (2006). The model organism as a system: Integrating 'omics' data sets. *Nature Reviews Molecular Cell Biology*, *7*, 198–210.

Kaiser, D. (2001). Building a multicellular organism. *Annual Review of Genetics*, *35*, 103–123.

Kan, J., Hanson, T. E., Ginter, J. M., Wang, K., & Chen, F. (2005). Metaproteomic research of Chesapeake Bay microbial communities. *Saline Systems*, *1*(7). doi:10.1186/1746-1448-1-7.

Kitano, H., & Oda, K. (2006a). Robustness trade-offs and host–microbial symbiosis in the immune system. *Molecular Systems Biology*, *2*. doi:10.1038/msb4100039.

Kitano, H., & Oda, K. (2006b). Self-extending symbiosis: A mechanism for increasing robustness through evolution. *Biological Theory*, *1*, 61–66.

Kolenbrander, P. E. (2000). Oral microbial communities: Biofilms, interactions, and genetic systems. *Annual Review of Microbiology*, *54*, 413–437.

Lange, M., Westermann, P., & Ahring, B. K. (2005). Archaea in protozoa and metazoa. *Applied Microbiology and Biotechnology*, *66*, 465–474.

Ley, R. E., Bäckhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., & Gordon, J. I. (2005). Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences*, *102*, 11070–11075.

Ley, R. E., Peterson, D. A., & Gordon, J. I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*, *124*, 837–848.

Liles, M. R., Manske, B. F., Bintrim, S. B., Handelsman, J., & Goodman, R. M. (2003). A census of rRNA genes and linked genomic sequences within a soil metagenomic library. *Applied and Environmental Microbiology*, *69*, 2684–2691.

Lloyd, D. (2004). 'Anaerobic protists': Some misconceptions and confusions. *Microbiology*, *150*, 1115–1116.

Lorenz, P., & Eck, J. (2005). Metagenomics and industrial application. *Nature Reviews Microbiology*, *3*, 510–516.

Lovley, D. R. (2003). Cleaning up with genomics: Applying molecular biology to bioremediation. *Nature Reviews Microbiology*, *1*, 35–44.

Machamer, P. K., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, *67*, 1–25.

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

Margulis, L. (1998). *Symbiotic planet: A new look at evolution*. New York: Basic Books.

Maynard Smith, J., & Szathmáry, E. (1995). *The major transitions in evolution*. Oxford: Oxford University Press.

McFall-Ngai, M. J. (2002). Unseen forces: The influence of bacteria on animal development. *Developmental Biology*, *242*, 1–14.

Michael, C. A., Gillings, M. R., Holmes, A. J., Hughes, L., Andrew, N. R., Holley, M. P., & Stokes, H. W. (2004). Mobile gene cassettes: A fundamental resource for evolution. *The American Naturalist*, *164*, 1–12.

Nelson, K. E. (2003). The future of microbial genomics. *Environmental Microbiology*, *5*, 1223–1225.

Nicholson, J. K., Holmes, E., Lindon, J. C., & Wilson, I. D. (2004). The challenges of modelling mammalian biocomplexity. *Nature Biotechnology*, *22*, 1268–1274.

Nicholson, J. K., Holmes, E., & Wilson, I. D. (2005). Gut microorganisms, mammalian metabolism and personalized health care. *Nature Reviews Microbiology*, *3*, 431–438.

Noonan, J. P., Hofreiter, M., Smith, D., Priest, J. R., Rohland, N., Rabeder, G., Krause, J., Detter, J. C., Pääbo, S., & Rubin, E. M. (2005). Genomic sequencing of Pleistocene cave bears. *Science*, *309*, 597–599.

Olsen, G. J., Lane, D. J., Giovannoni, S. J., & Pace, N. R. (1986). Microbial ecology and evolution: A ribosomal RNA approach. *Annual Review of Microbiology*, *40*, 337–365.

O'Malley, M. A., & Dupré, J. (2005). Fundamental issues in systems biology. *BioEssays*, *12*, 1270–1276.

O'Malley, M. A., & Dupré, J. (2007). Size doesn't matter: Towards an inclusive philosophy of biology. *Biology and Philosophy*, *22*, 155–191.

Ordovas, J. M., & Mooser, V. (2006). Metagenomics: The role of the microbiome in cardiovascular diseases. *Current Opinion in Lipidology*, *17*, 157–161.

O'Toole, G., Kaplan, H. B., & Kolter, R. (2000). Biofilm formation as microbial development. *Annual Review of Microbiology*, *54*, 49–79.

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., & Lane, D. J. (1985). Analyzing natural microbial populations by rRNA sequences. *ASM News*, *51*, 4–12.

Palsson, B. (2000). The challenges of in silico biology. *Nature Biotechnology*, *18*, 1147–1150.

Pazos, F., Valencia, A., & De Lorenzo, V. (2003). The organization of the microbial biodegradation network from a systems biology perspective. *EMBO Reports*, *4*, 994–999.

Pennisi, E. (2005). Reading ancient DNA the community way. *Science*, *308*, 1401.

Perotto, S., & Bonfante, P. (1997). Bacterial associations with mycorrhizal fungi: Close and distant friends in the rhizosphere. *Trends in Microbiology*, *5*, 496–501.

Poinar, H. N., Schwarz, C., Qi, J., Shapiro, B., MacPhee, R. D. E., Buiges, B., Tikhonov, A., Huson, D. H., Tomsho, L. P., Auch, A., Rampp, M., Miller, W., & Schuster, S. C. (2006). Metagenomics to paleoge-

nomics: Large-scale sequencing of mammoth DNA. *Science*, *311*, 392–394.

Poretsky, R. S., Bano, N., Buchan, A., LeCleir, G., Kleikemper, J., Pickering, M., Pate, W. M., Moran, M. A., & Hollibaugh, J. T. (2005). Analysis of microbial gene transcripts in environmental samples. *Applied and Environmental Microbiology*, *71*, 4121–4126.

Powell, M. J., Sutton, J. N., Del Castillo, C. E., & Timperman, A. T. (2005). Marine proteomics: Generation of sequence tags for dissolved proteins in seawater using tandem mass spectrometry. *Marine Chemistry*, *95*, 183–198.

Ram, R. J., VerBerkmoes, N. C., Thelen, M. P., Tyson, G. W., Baker, B. J., II, Blake, R. C., II, Shah, M., Hettich, R. L., & Banfield, J. F. (2005). Community proteomics of a natural microbial biofilm. *Science*, *308*, 1915–1920.

Rawls, J. F., Samuel, B. S., & Gordon, J. I. (2004). Gnotobiotic zebrafish reveal evolutionarily conserved responses to the gut microbiota. *Proceedings of the National Academy of Sciences*, *101*, 4596–4601.

Relman, D. A., & Falkow, S. (2001). The meaning and impact of the human genome sequence for microbiology. *Trends in Microbiology*, *9*, 206–208.

Riesenfeld, C. S., Schloss, P. D., & Handelsman, J. (2004). Metagenomics: Genomic analysis of microbial communities. *Annual Reviews of Genetics*, *38*, 525–552.

Rodríguez-Valera, F. (2002). Approaches to prokaryotic diversity: A population genetics approach. *Environmental Microbiology*, *4*, 628–633.

Rodríguez-Valera, F. (2004). Environmental genomics: The big picture. *FEMS Microbiology Letters*, *231*, 153–158.

Rondon, M. R., August, P. R., Bettermann, A. D., Brady, S. F., Grossman, T. H., Liles, M. R., Loiacono, K. A., Lynch, B. A., MacNeil, I. A., Minor, C., Tiong, C. L., Gilman, M., Osburne, M. S., Clardy, J., Handelsman, J., & Goodman, R. M. (2000). Cloning the soil metagenome: A strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied and Environmental Microbiology*, *66*, 2541–2547.

Rowe-Magnus, D. A., Guerot, A.-M., & Mazel, D. (2002). Bacterial resistance evolution by recruitment of super-integron gene cassettes. *Molecular Microbiology*, *43*, 1657–1669.

Savage, D. C. (1977). Microbial ecology of the gastrointestinal tract. *Annual Review of Microbiology*, *31*, 107–133.

Schloss, P. D., & Handelsman, J. (2003). Biotechnological prospects from metagenomics. *Current Opinion in Biotechnology*, *14*, 303–310.

Schloss, P. D., & Handelsman, J. (2005). Metagenomics for studying unculturable organisms: Cutting the Gordian knot. *Genome Biology*, *6*, 229. (Available at doi:10.1186/gb-2005-6-8-229)

Schmeisser, C., Stöckigt, C., Raasch, C., Wingender, J., Timmis, K. N., Wenderoth, O. F., Flemming, H.-C., Ziesegang, H., Schmitz, R. A., Jaeger, K. E., & Streit, W. R. (2003). Metagenome survery of biofilms in drinking-water networks. *Applied and Environmental Microbiology*, *69*, 7298–7309.

Schulze, W. X., Gleixner, G., Kaiser, K., Guggenberger, G., Mann, M., & Schulze, E.-D. (2005). A proteomic fingerprint of dissolved organic carbon and of soil particles. *Oecologia*, *142*, 335–343.

Sebat, J. L., Colwell, F. S., & Crawford, R. L. (2003). Metagenomic profiling: Microarray analysis of an environmental genomic library. *Applied and Environmental Microbiology*, *69*, 4927–4934.

Shapiro, J. A. (1997). Multicellularity: The rule, not the exception. In J. A. Shapiro, & M. Dworkin (Eds.), *Bacteria as multicellular organisms* (pp. 14–49). New York: Oxford University Press.

Shapiro, J. A. (1998). Thinking about bacterial populations as multicellular organisms. *Annual Review of Microbiology*, *52*, 81–104.

Shapiro, J. A., & Dworkin, M. (Eds.). (1997). *Bacteria as multicellular organisms*. New York: Oxford University Press.

Sonea, S., & Mathieu, L. G. (2001). Evolution of the genomic systems of prokaryotes and its momentous consequences. *International Microbiology*, *4*, 67–71.

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

Stokes, H. W., Holmes, A. J., Nield, B. S., Holley, M. P., Nevalainen, K. M. H., Mabbutt, B. C., & Gillings, M. R. (2001). Gene cassette PCR: Sequence-independent recovery of entire genes from environmental DNA. *Applied and Environmental Microbiology*, *67*, 5240–5246.

Stoodley, P., Sauer, K., Davies, D. G., & Costerton, J. W. (2002). Biofilms as complex differentiated communities. *Annual Review of Microbiology*, *56*, 187–209.

Streit, W. R., & Schmitz, R. A. (2004). Metagenomics—the key to uncultured microbes. *Current Opinion in Microbiology*, *7*, 492–498.

Torsvik, V., & Øvreås, L. (2002). Microbial diversity and function in soil: From genes to ecosystems. *Current Opinion in Molecular Biology*, *5*, 240–245.

Treusch, A. H., Leininger, S., Kletzin, A., Schuster, S. C., Klenk, H.-P., & Schleper, C. (2005). Novel genes for nitrate reduction and Amo-related proteins indicate a role of uncultivated mesophilic crenarchaeota in nitorgen cycling. *Environmental Microbiology*, *7*, 1985–1995.

Tringe, S. G., Mering, C. v., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., Podar, M., Short, J. M., Mathur, E. J., Detter, J. C., Bork, P., Hugenholz, P., & Rubin, E. M. (2005). Comparative metagenomics of microbial communities. *Science*, *308*, 554–557.

Tringe, S. G., & Rubin, E. M. (2005). Metagenomics: DNA sequencing of environmental samples. *Nature Reviews Genetics*, *6*, 805–814.

Tyson, G. W., & Banfield, J. F. (2005). Cultivating the uncultivated: a community genomics perspective. *Trends in Microbiology*, *13*, 411–415.

Tyson, G. W., Chapman, J., Hugenholz, P., Allen, E. E., Ram, R. J., Richardson, P. M., Solovyev, V. V., Rubin, E. M., Rokhsar, D. S., & Banfield, J. F. (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., Wu, D., Paulsen, I., Nelson, K. E., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, H.-Y., & Smith, H. O. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, *304*, 66–74.

Voget, S., Leggewie, C., Uesbeck, A., Raasch, C., Jaeger, K.-E., & Streit, W. R. (2003). Prospecting for novel biocatalysts in a soil metagenome. *Applied and Environmental Microbiology*, *69*, 6235–6242.

Ward, N. (2006). New directions and interactions in metagenomics. *FEMS Microbial Ecology*, *55*, 331–338.

Ward, N., & Fraser, C. M. (2005). How genomics has affected the concept of microbiology. *Current Opinion in Microbiology*, *8*, 564–571.

Wellington, E. M. H., Berry, A., & Krsek, M. (2003). Resolving functional diversity in relation to microbial community structure in soil: Exploiting genomics and stable isotype probing. *Current Opinion in Microbiology*, *6*, 295–301.

Wilmes, P., & Bond, P. L. (2006). Metaproteomics: Studying functional gene expression in microbial ecosystems. *Trends in Microbiology*, *14*, 92–97.

Wimpenny, J. (2000). An overview of biofilms as functional communities. In D. G. Allison, P. Gilbert, H. M. Lappin-Scott, & M. Wilson (Eds.), *Community structure and cooperation in biofilms* (pp. 1–24). Cambridge: Cambridge University Press.

Wu, L., Thompson, D. K., Li, G., Hurt, R. A., Tiedje, J. M., & Zhou, J. (2001). Development and evaluation of functional gene arrays for detection of selected genes in the environment. *Applied and Environmental Microbiology*, *67*, 5780–5790.

Xu, J. (2006). Molecular ecology in the age of genomics and metagenomics: Concepts, tools, and recent advances. *Molecular Ecology*, *15*, 1713–1731.

Xu, J., & Gordon, J. I. (2003). Honor thy symbionts. *Proceedings of the National Academy of Sciences*, *100*, 10452–10459.

Zhou, J. (2003). Microarrays for bacterial detection and microbial community analysis. *Current Opinion in Microbiology*, *6*, 288–294.

Zoetendel, E. G., Vaughan, E. E., & de Vos, W. M. (2006). A microbial world within us. *Molecular Microbiology*, *59*, 1639–1650.