Cell
Leading Edge
Essay

An Invitation to the Marriage of Metagenomics and Metabolomics

Peter J. Turnbaugh[1] and Jeffrey I. Gordon[1,*]
[1]Center for Genome Sciences, Washington University, St. Louis, MO 63108, USA
*Correspondence: jgordon@wustl.edu
DOI 10.1016/j.cell.2008.08.025

Metagenomics seeks to characterize the composition of microbial communities, their operations, and their dynamically coevolving relationships with the habitats they occupy without having to culture community members. Uniting metagenomics with analyses of the products of microbial community metabolism (metabolomics) will shed light on how microbial communities function in a variety of environments, including the human body.

Metagenomics is a rapidly growing area of the genome sciences that seeks to define the features of intact microbial communities in their native habitats (Box 1). It allows us to see, with a new and powerful set of lenses, the vast microbial and metabolic diversity that exists in our biosphere. The scale of this diversity is mind-boggling. The estimated total number of microbial cells is 10^{30}: these microbes may harbor as many as 10^{31} phage (Dinsdale et al., 2008). Their genomes encode a wide variety of enzymes that influence major metabolic fluxes within the biosphere, from phototrophy in the sea (DeLong et al., 2006; Frias-Lopez et al., 2008), to phosphate removal in industrial sludge (Martin et al., 2006), to nutrient harvest from the diet in animal guts (Gill et al., 2006; Turnbaugh et al., 2006, 2008; Warnecke et al., 2007). Metagenomics provides us with an unprecedented opportunity to assess the metabolic features of microbial communities without the need to culture their component members. The results promise to impact a wide range of "bio"-related disciplines, including biomedicine, bioenergy, bioremediation, and biodefense (Committee on Metagenomics, 2007).

Metagenomics creates experimental, computational, and conceptual connections between the too-often artificially separated worlds of environmental and medical microbiology. For example, although our development occurs in the apparently microbe-free environment of the uterus, beginning at birth we are colonized by members of all three known domains of life: Bacteria, Archaea, and Eukarya. As adults, the number of microbial cells associated with our adult human bodies likely exceeds the number of our somatic and germ cells by 10-fold. Studies of model organisms suggest that acquisition of this microbiota (Box 1) reflects a combination of legacy effects (a summation of the microbes we are exposed to in our distinctively experienced living environments) and habitat effects (largely ill-defined "filters" that operate to select microbes for residency in our different body sites) (Rawls et al., 2006).

In many ways, we should view ourselves, and all other animals, as "supraorganisms," composed of mixtures of host and microbial cells, genes, and metabolic attributes. Defining the gene content of our microbiota (the "microbiome") represents a first step in exploring the largely mysterious biotransformations that are supported by its constituents. The metabolome—the metabolites generated by one or more organisms under a particular set of physiological/environmental conditions (Box 1)—holds the key to moving beyond descriptions of the composition of our indigenous microbial communities (that is, their component organisms, genes, transcripts, and proteins). The metabolome will enable a better understanding of the dynamic operations of our indigenous microbial communities and their functional contributions to the various body habitats that they occupy. These contributions include metabolic activities that are not encoded in our human genome, for example, the breakdown of otherwise difficult to digest plant glycans, synthesis of vitamins, and the processing of various xenobiotics that we consume (either intentionally or unintentionally).

Metagenomics: Three Reasons to Care

The skeptical reader may suppose that metagenomics is currently just a limited set of evolving DNA-centric experimental and computational techniques, driven in large part by the introduction of massively parallel sequencers that, unlike capillary machines, do not require subcloning of DNA fragments prior to sequencing. Our retort is a "yes but"…there are at least three reasons why data generated by the sequencing of microbial community DNA, including 16S rRNA genes, will be necessary to understand microbial metabolism and therefore community function. First, only a small fraction of microbial diversity in the biosphere can be isolated in pure culture, presumably because of the highly artificial and limited number of conditions used currently for cultivation. This problem has been called the "great plate-count anomaly" and refers to the fact that the number of organisms cultured from a community is typically far less than the number observed using culture-independent methods such as 16S rRNA gene sequencing (e.g., Huber et al., 2007; Roesch et al., 2007). In addition, culture-based approaches do not scale as readily as culture-independent techniques: for example, a single massively parallel pyrosequencer can gen-

erate more than 2 million sequencing reads from 16S rRNA genes each day—a number well beyond the range of our current capacity for culturing microorganisms. Second, multiple representatives of the same "species" (commonly defined as organisms that share ≥97% nucleotide sequence identity between their 16S rRNA genes) can have a surprising amount of variability in their proteomes. For example, comparisons of 32 *Escherichia coli* and *Shigella* genomes (Willenbrock et al., 2007) and an analysis of six strains of *Streptococcus agalactiae* (Tettelin et al., 2005) disclosed a large set of seemingly "dispensable" genes found in subsets of the genomes from each "species." Such observations indicate that a future definition of species will be gleaned from analyses of the functions encoded by the shared (core) set of genes present in a larger "pan-genome" that represents the sum of all genes found in all isolates of a 16S rRNA-defined species-level phylotype (Box 1). The bottom line is that the total number of genome variants in a community is likely to greatly exceed the number of 16S rRNA-based phylotypes. Metagenomic sequencing of total community DNA offers a rapid and efficient way of addressing these issues whether or not the microbes can be cultured. Third, most microbial metabolic processes in nature occur in the context of communities, not pure cultures. There are clearly many advantages to exploring the niche (professions) of organisms in pure cultures, including the opportunity for controlled experimentation. However, metagenomics represents a conceptual framework for analyzing microbial metabolism at a higher level of organization than that of a single organism. Metagenomic studies of microbial communities are revealing how many emergent properties become apparent in the context of a complex system that could not be easily deduced from an analysis limited to its individual components. For example, the global distribution of the dominant Cyanobacteria, *Prochlorococcus*, in large regions of the world's oceans could be inferred through computational analyses of a model marine ecosystem seeded with diverse phytoplankton having randomly assigned physiological characteristics (Follows et al., 2007).

Metagenomics for Predicting Microbial (and Host) Metabolism

Metagenomics has already proven useful in going from DNA sequence data to understanding the metabolism of relatively "simple" microbial communities. An excellent example is the shotgun sequencing of DNA isolated from a low-complexity microbial biofilm community that forms in the acidic water, rich in pyrite (FeS_2), that drains from an abandoned mine in Iron Mountain, California (Tyson et al., 2004). This approach allowed the genomes of two previously uncultured organisms, *Leptospirillum* group II and *Ferroplasma* type II, to be assembled into a relatively small number of contigs. The deep sequence coverage achieved was used to surmise that the *Ferroplasma* type II genome was actually a composite of three ancestral strains that have undergone extensive homologous recombination, producing a population of mosaic genomes. The *Leptospirillum* group II and the less abundant group III genomes encode key enzymes in the Calvin-Benson-Bassham cycle for carbon fixation (type II ribulose 1,5-bisphosphate carboxylase-oxygenase). The *Ferroplasma* type II genome assembly contains multiple ABC-type sugar and amino acid transporters indicative of a heterotrophic lifestyle in which organic matter is used for energy (Tyson et al., 2004). Furthermore, the *Leptospirillum* group III genome assembly has a single nitrogen fixation operon (*nif*), suggesting a role for these low abundance bacteria as the primary nitrogen fixers in the acid mine community. These in silico predictions were subsequently validated experimentally (Tyson et al., 2005).

Another example is the shotgun sequencing of DNA prepared from two relatively low-complexity communities that had been seeded in the laboratory from a man-made industrial sludge. This sludge has ecological importance because of its capacity for microbe-mediated removal of inorganic phosphate from the environment. The results allowed the genome of the dominant but previously uncultured member, *Candidatus* Accumulibacter phosphatis, to be assembled. In silico metabolic reconstructions revealed how the organism

Box 1. Glossary

Metagenomics: An emerging field encompassing culture-independent studies of the structures and functions of microbial communities and their interactions with the habitats they occupy. Metagenomics includes (1) shotgun sequencing of microbial DNA isolated directly from a given environment, (2) high-throughput screening of expression libraries, constructed from cloned community DNA, to identify specific functions such as antibiotic resistance (functional metagenomics), (3) profiling of RNAs and proteins produced by a microbiome (meta-transcriptomics and meta-proteomics), and (4) identification of a community's metabolic network (meta-metabolomics).

Metabolomics: The characterization by mass spectroscopy, NMR, or other analytical methods of metabolites generated by one or more organisms in a given physiological and environmental context.

Microbiota: A microbial community, including Bacteria, Archaea, Eukarya, and viruses, that occupies a given habitat.

Microbiome: In ecology, the term "biome" refers to a habitat and the organisms in it. In this sense, the human "micro"biome would be defined as the collection of microorganisms associated with the human body, and their collective genomes would constitute a metagenome. However, the term microbiome is now commonly used to refer to the collective genomes present in members of a given microbiota.

Phylotype: A phylogenetic group of microbes, currently defined by a threshold percent identity shared among their small subunit (16S) rRNA genes (e.g., ≥97% for a "species" level phylotype).

Pan-genome: The group of genes found in genomes comprising a given phylotype. This includes "core" genes found in all genomes and "dispensable" genes found in a subset of genomes within the phylotype.

Core human microbiome: Whatever is shared in a given body habitat among all or the vast majority of human microbiomes. A core microbiome may include a common set of organisms (genomes), gene or protein families, and/or metabolic capabilities. Microbial genes that are variably represented in different humans may contribute to our distinctive metabolic attributes (metabotypes).

Seed bank: Low-abundance phylotypes within a microbiota that may not significantly alter major metabolic fluxes. These phylotypes can include transient (allochthonous) members of the community.

Gene bank: Low-abundance genes found within a given microbiome or pan-genome. Some of these genes may be the result of lateral gene transfer from allochthonous organisms and may include unique metabolic activities found in members of the seed bank.

may (1) dispose of inorganic phosphate (through import and metabolism to polyphosphate, which is then used to provide energy for importing and storing fatty acids), (2) degrade glycogen (through the Embden-Meyerhof but not the Entner-Doudoroff pathway), and (3) produce NADPH via a novel cytochrome *b/b6* (Martin et al., 2006).

Metagenomic approaches have also been used to investigate the metabolic underpinnings of host-bacterial symbioses in invertebrates. *Olavius algarvensis*, a marine worm that lacks a mouth, gut, or anus, harbors four known Proteobacterial symbionts below its cuticle (Woyke et al., 2006). Shotgun sequencing of microbial community DNA, and in silico reconstructions of the metabolic pathways encoded by the microbiome of this gutless worm, revealed that its symbionts—including a sulfate-reducing δ-Proteobacteria and a sulfur-oxidizing chemoautotroph belonging to the γ-Proteobacteria—have important metabolic capabilities. They are able to provide the host with nutrients (amino acids and vitamins) derived from reduced sulfur compounds, hydrogen, CO_2, or dissolved organic carbon from the environment, and to remove metabolic end-products generated by the worm (ammonium, urea, and short-chain fatty acids).

Likewise, DNA-level metagenomic analysis of the Sharpshooter, a major agricultural pest that feeds on the xylem fluid of plants and acts as a vector for plant pathogens, disclosed that one of its symbionts, the γ-Proteobacteria *Baumannia cicadellinicola*, has the capacity to provide it with vitamins [thiamine (B_1), riboflavin (B_2), niacin (B_3), panthothenic acid (B_5), pyridoxine (B_6), biotin, and folic acid]. Another one of its symbionts, *Sulcia muelleri* (a Bacteroidetes), has the capacity to supply the Sharpshooter with amino acids (threonine, leucine, valine, isoleucine and potentially lysine, arginine, and tryptophan) (Wu et al., 2006).

A recent study of obligate γ-Proteobacterial symbionts that live in the posterior midgut of the stinkbug demonstrated that pest behavior (fitness on crop legumes) is dependent on this microbe. When symbiont strains, with 99.9% identical 16S rRNA gene sequences, are experimentally exchanged between a stinkbug pest species (*Megacopta punctatissima*) and a non-pest species (*Megacopta cribraria*), the pest phenotype is reversed (Hosokawa et al., 2007). Sequencing of these uncultured symbiont genomes will likely reveal candidate pest-promoting microbial genetic determinants.

Metagenomics has also identified distinguishing metabolic features of very complex microbial communities, including those associated with an agricultural soil, sunken whale carcasses (whale falls), and the ocean (Tringe et al., 2005). Comparisons of the soil and Sargasso Sea microbiomes showed an enrichment (significantly higher relative abundance) of genes involved in degradation of plant material in the soil (e.g., predicted genes encoding cellobiose phosphorylase) and enrichment for light-driven proton pumps (bacteriorhodopsins) in the Sargasso Sea (Tringe et al., 2005). Furthermore, the ocean is composed of stratified microbial assemblages: studies of microbiome gene content at multiple depths (10 m–4000 m) identified differences in the relative abundance of specific categories of genes. The more productive photic zone community has a higher relative abundance of photosynthesis genes than the deep water community, a lower representation of genes indicative of lateral gene transfer (transposases), and a higher relative abundance of cyanophage (DeLong et al., 2006).

Metagenomic studies of complex communities are also linking microbiomes to mechanisms underlying biomass conversion in a variety of biological bioreactors, including the mouse and human intestine (Turnbaugh et al., 2006, 2008; Gill et al., 2006; Kurokawa et al., 2007). Surveys of 16S rRNA sequences present in the distal gut microbiota of lean and obese mice and humans revealed an association between adiposity and the relative proportions of the two dominant bacterial divisions in these mammalian ecosystems—the Bacteroidetes and the Firmicutes (Ley et al., 2006). Sequencing DNA isolated from the distal intestines of genetically obese (*ob/ob*) mice and their lean (*ob*/+, +/+) littermates disclosed a significant enrichment for genes in the *ob/ob* gut microbiome that are involved in harvesting energy from polysaccharides. These in silico metabolic predictions were supported by biochemical analyses, and by microbiome transplantation experiments. The latter involved transfer of unfractionated distal gut microbial communities from *ob/ob* and lean (+/+) donors to wild-type germ-free mouse recipients. The obese mouse gut community produced a significantly greater increase in adiposity in recipients over a 2 week period (Turnbaugh et al., 2006).

Follow-up studies of a mouse model of diet-induced obesity caused by consumption of a high-fat, high-sugar "Western" diet revealed that obesity was associated with marked but reversible structural and functional changes in the gut microbiome. In particular, there was a bloom in a single uncultured clade of bacteria within the Mollicutes class of the Firmicutes division and an increased ability of this community to promote adiposity in germ-free recipients compared to recipients of a microbiota harvested from lean mice consuming a standard chow diet. Metagenomic sequencing of the gut microbiome, biochemical assays, plus sequencing, and in silico metabolic reconstructions of a related human gut-associated Mollicute (*Eubacterium dolichum*) revealed features that may provide a competitive advantage for members of the bloom in the Western diet nutrient milieu, including genes involved in import and metabolism of simple sugars (Turnbaugh et al., 2008).

An initial analysis of the fecal microbiomes of two unrelated humans (Gill et al., 2006) showed that compared to all previously sequenced microbial genomes and the human genome, the human gut microbiome is enriched for genes involved in a number of metabolic pathways. These genes include those for (1) 2-methyl-D-erythritol 4-phosphate pathway-mediated biosynthesis of essential vitamins (e.g., thiamine and pyridoxine) and isoprenoids (derivatives of isopentenyl pyrophosphate), (2) the metabolism of amino acids, xenobiotics (including carcinogens), and dietary polysaccharides, and (3) methanogenesis. Methanogenesis is a key metabolic pathway, present in members of the Euryarchaeota division of Archaea, that removes the H_2 end-product of bacterial fermentation of dietary polysaccharides,

thereby preventing inhibition of this process. Inhibition occurs because the flux of sugars through the glycolytic pathway leads to accumulation of NADH. Removal of hydrogen by methanogenic Archaea, by phylogenetically diverse acetogens that use the Wood-Ljungdhal pathway for synthesis of acetyl-CoA from CO_2, and/or by Proteobacteria that reduce sulfate to sulfide (a genotoxic compound) allows NADH to be oxidized to NAD^+ (Drake et al., 2006).

These metagenomic analyses were subsequently extended to the fecal microbiomes of 13 healthy humans, including 4 suckling infants, 2 children, and 7 adults (Kurokawa et al., 2007). Comparative metagenomics revealed that the infant fecal microbiome is highly variable, consistent with a recent 16S rRNA gene sequence-based study of microbial community assembly in the gut of infants during the first year after birth (Palmer et al., 2007). In silico functional analyses of the adult human fecal microbiome versus previously sequenced microbial genomes confirmed that it is enriched for genes involved in carbohydrate and amino acid metabolism, including a number of polysaccharide-degrading glycoside hydrolases. Additionally, the human gut microbiome contains many conjugative transposon sequences, emphasizing the potential role of lateral gene transfer in shaping gut genomes (Kurokawa et al., 2007).

One notion that can be abstracted from these studies of mammalian gut microbiomes is that the nutrient/energetic value of food is a relative, not an absolute term that can be influenced by metabolic activities encoded in the microbiome. In turn, our food choices may become imprinted into our supraorganismal metabolome in part by affecting the structure and activity of the gut microbiota (Rezzi et al., 2007; also see the Forum by Holmes et al., page 714 of this issue). This metabolic imprinting could be initiated at an early stage of life. It follows that metagenomic studies should allow us to obtain a deeper understanding of the nutritional needs of humans. The results could yield microbiome-based biomarkers for identifying those at risk for obesity or malnutrition, dietary recommendations based on the nutrient-processing capacities of the gut microbiota of various human populations, and a new understanding about why certain prevention strategies and therapeutic interventions fail in some but succeed in others (Prentice et al., 2008).

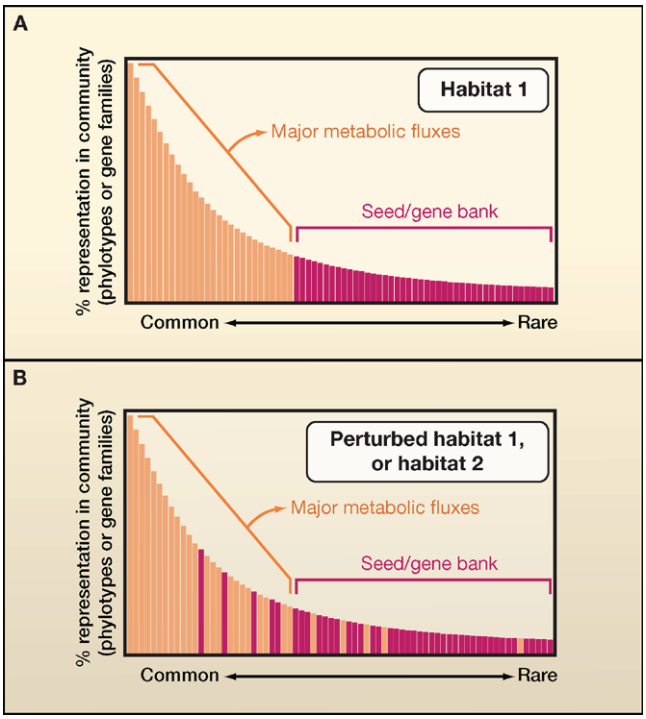

Figure 1. Common and Rare Components of the Microbiome

A theoretical framework for the global dispersal of phylotypes (defined by percent identity of 16S rRNA genes) and gene families (defined through homology, Hidden Markov Models of protein domains, and/or enzymatic activity).

(A) The graph depicts a rank-abundance plot of the phylotypes or gene families present in a given habitat (e.g., ocean, soil, or gut); phylotypes or gene families are plotted in decreasing abundance across the *x* axis. The orange portion of the curve can be defined as the portion of the community undergoing active growth and the abundant gene families that are primarily responsible for the major metabolic fluxes within a given habitat. The red portion of the curve comprises the seed/gene bank. The seed/gene bank represents the low-abundance members that may or may not significantly alter features of community metabolism depending upon the environmental context (the seed bank) and the rare genes that may be found in a small subset of genomes within a given phylotype (dispensable genes), or in the genomes of rare (including transient or allochthonous) members of the community (the gene bank).

(B) This distribution can be altered in a different habitat, or in the same habitat after perturbation by disease or a change in available nutrients.

Biodiversity and the Global Dispersal of Microbiomes

An issue facing the marriage of metagenomics and metabolomics relates to the global dispersal of microbes. Some have proposed that microbes are not geographically isolated, given a generally large population size of each microbial phylotype in the biosphere and presumed minimal barriers to global dispersal. This would mean that in many environments there are members of all or the vast majority of phylotypes, and the habitat (continuously) selects to determine relative abundance. This conceptual framework can be depicted in a rank abundance plot of diversity (Figure 1). In this plot, the highly abundant phylotypes are viewed as responsible for most active growth and carbon and energy flux in a given ecosystem, with the rare phylotypes forming a "seed bank" (a term defined here as a set of organisms present at low abundance that may or may not significantly alter major metabolic fluxes) (Pedros-Alio, 2006; Box 1).

If this formulation were extended to microbiomes, it would suggest that if the vast majority of phylotypes are present in a given environment, then most microbial genes might also be present in that environment. This does not appear to be the case in low-complexity microbial communities (e.g., the acid mine drainage) or in certain host-symbiont relationships (e.g., the Sharpshooter, the gutless worm, and the light organ of the squid *Euprymna scolopes*, which is colonized by a single dominant bacterium, *Vibrio fischeri*). However, it may be true for environments with potentially less restricted barriers to colonization, such as the ocean, soil, or even the human alimentary tract (Hooper et al., 2008). Furthermore, given the extent of lateral gene transfer in environments such as the

distal human intestine, where microbial densities approach 10^{12} organisms per milliliter of gut luminal contents (Xu et al., 2007), plus the results of initial studies of bacterial pan-genomes, it seems likely that a large number of rare genes may be found in members of a wide variety of phylotypes.

This point becomes important when considering how microbial communities adjust their metabolic activities to changing environmental conditions, and how to define a "core microbiome." The core human microbiome has been viewed as whatever is shared in a given body habitat among the vast majority of human microbiomes (e.g., Turnbaugh et al., 2007). Relative abundance needs to be factored into this definition of a core microbiome: the "gene bank" will likely contain a large number of genes that are present at incredibly low relative abundance in all human-associated communities and are thus extremely difficult to detect (Figure 1A). A more readily defined facet of the core microbiome could be the set of abundant genes that influence major metabolic cycles. The more elusive gene bank would include a vast assortment of rare genes that may not markedly affect carbon, nitrogen, or other nutrient and energy fluxes within a microbial community, and between a microbiota and its host. Nonetheless such a gene bank could contribute other important features to community and host metabotypes, and to the resiliency of a microbial community when its structure is perturbed (Figure 1B). However, many of these rare microbial genes may not encode functions that confer a detectable selective advantage to the host under various circumstances.

The Human Microbiome Project and Metabolomics

Human microbiome projects are being launched that will survey microbiomes in a number of body habitats of people of varying ages and genotypes, living in different ecosystems, in societies with varying lifestyles (Turnbaugh et al., 2007). As these efforts seek to discover connections between our supraorganismal metabolism and our microbial ecology, they will have to address a number of the questions alluded to above. For example, what is the underlying population structure of the microbial communities that occupy our various body habitats at various time points over a life span? What is the provenance of the organismal lineages that comprise these communities: how much vertical transmission is there from our mothers and/or other family members? How much functional redundancy exists between the myriad bacterial strains that comprise our microbiota, and between communities whose diversity has only been described in terms of 16S rRNA-based phylogenies? How can we improve predictions about the metabolic capabilities and activities of a microbial community from metagenomic data? The latter question embraces issues of how to generate more reliable predictions of gene function, how to create better algorithms for identifying protein families and metabolic pathways (Godzik et al., 2007; Yooseph et al., 2007), and how to integrate metabolite data from NMR and mass spectrometry with DNA, mRNA, and protein data sets.

The task of answering these questions is daunting and the outcomes uncertain. We believe that addressing many of these questions will require concomitant metagenomic studies of genetically malleable model host organisms with defined genotypes, raised under germ-free conditions, that are then colonized at varying times during their life with defined consortia of sequenced microbial phylotypes (a "type community" much as there are "type strains"). These gnotobiotic models should allow general principles to be gleaned about the coevolution and interoperability of microbial and host metabolomes, as many of the confounding variables that apply to humans (diet, genotype, environmental exposures) can be constrained. Much will be learned from careful energy and nutrient balance studies of these gnotobiotic animals, from characterization of host and microbial community genomes, transcriptomes, proteomes, and metabolomes, and from identifying microbiome-based biomarkers of metabolic activities. This knowledge can then be tested in humans with the goal of achieving more comprehensive phenotyping of our physiology, more accurate definitions of our health, more insightful stratification of our risks for various metabolic disorders, and new ways for intentionally manipulating our microbiomes to optimize their functions in the context of our individual biological milieus.

ACKNOWLEDGMENTS

We thank C. Semenkovich, J. Sonnenburg, P. Sudarsanam, A. Goodman, M. Mahowald, and F. Rey for helpful comments.

REFERENCES

Committee on Metagenomics (2007). (Washington, D.C.: The National Academies Press).

DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.U., Martinez, A., Sullivan, M.B., Edwards, R., Brito, B.R., et al. (2006). Science *311*, 496–503.

Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., Furlan, M., Desnues, C., Haynes, M., Li, L., and McDaniel, L. (2008). Nature *452*, 629–632.

Drake, H.L., Kusel, K., and Matthies, C. (2006). Acetogenic prokaryotes. In The Prokaryotes, 3rd Edition, Volume 2 (New York: Springer), pp. 354–420.

Follows, M.J., Dutkiewicz, S., Grant, S., and Chisholm, S.W. (2007). Science *315*, 1843–1846.

Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chisholm, S.W., and Delong, E.F. (2008). Proc. Natl. Acad. Sci. USA *105*, 3805–3810.

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., and Nelson, K.E. (2006). Science *312*, 1355–1359.

Godzik, A., Jambon, M., and Friedberg, I. (2007). Cell. Mol. Life Sci. *64*, 2505–2511.

Hooper, S.D., Raes, J., Foerstner, K.U., Harrington, E.D., Dalevi, D., and Bork, P. (2008). PLoS ONE *3*, e2607. 10.1371/journal.pone.0002607.

Hosokawa, T., Kikuchi, Y., Shimada, M., and Fukatsu, T. (2007). Proc. Biol. Sci. *274*, 1979–1984.

Huber, J.A., Welch, D.B., Morrison, H.G., Huse, S.M., Neal, P.R., Butterfield, D.A., and Sogin, M.L. (2007). Science *318*, 97–100.

Kurokawa, K., Itoh, T., Kuwahara, T., Oshima, K., Toh, H., Toyoda, A., Takami, H., Morita, H., Sharma, V.K., Srivastava, T.P., et al. (2007). DNA Res. *14*, 169–181.

Ley, R.E., Turnbaugh, P.J., Klein, S., and Gordon, J.I. (2006). Nature *444*, 1022–1023.

Martin, H.G., Ivanova, N., Kunin, V., Warnecke, F., Barry, K.W., McHardy, A.C., Yeates, C., He, S., Salamov, A.A., Szeto, E., et al. (2006). Nat. Biotechnol. *24*, 1263–1269.

Palmer, C., Bik, E.M., DiGiulio, D.B., Relman, D.A., and Brown, P.O. (2007). PLoS Biol. *5*, e177 10.1371/journal.pbio.0050177.

Pedros-Alio, C. (2006). Trends Microbiol. *14*, 257–263.

Prentice, A.M., Gershwin, M.E., Schaible, U.E., Keusch, G.T., Victora, C.G., and Gordon, J.I. (2008). J. Clin. Invest. *118*, 1322–1329.

Rawls, J.F., Mahowald, M.A., Ley, R.E., and Gordon, J.I. (2006). Cell *127*, 423–433.

Rezzi, S., Ramadan, Z., Martin, F.P., Fay, L.B., Bladeren, P.V., Lindon, J.C., Nicholson, J.K., and Kochhar, S. (2007). J. Proteome Res. *6*, 4469–4477.

Roesch, L.F.W., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K.M., Kent, A.D., Daroub, S.H., Camargo, F.A.O., Farmerie, W.G., and Triplett, E.W. (2007). The ISME Journal *1*, 283–290.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., et al. (2005). Proc. Natl. Acad. Sci. USA *102*, 13950–13955.

Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., Podar, M., Short, J.M., Mathur, E.J., Detter, J.C., et al. (2005). Science *308*, 554–557.

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and Gordon, J.I. (2006). Nature *444*, 1027–1031.

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

Turnbaugh, P.J., Bäckhed, F., Fulton, L., and Gordon, J.I. (2008). Cell Host Microbe *3*, 213–223.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., and Banfield, J.F. (2004). Nature *428*, 37–43.

Tyson, G.W., Lo, I., Baker, B.J., Allen, E.E., Hugenholtz, P., and Banfield, J.F. (2005). Appl. Environ. Microbiol. *71*, 6319–6324.

Warnecke, F., Luginbuhl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., Cayouette, M., McHardy, A.C., Djordjevic, G., Aboushadi, N., et al. (2007). Nature *450*, 560–565.

Willenbrock, H., Hallin, P.F., Wassenaar, T.M., and Ussery, D.W. (2007). Genome Biol. *8*, R267.

Woyke, T., Teeling, H., Ivanova, N.N., Huntemann, M., Richter, M., Gloeckner, F.O., Boffelli, D., Anderson, I.J., Barry, K.W., Shapiro, H.J., et al. (2006). Nature *443*, 950–955.

Wu, D., Daugherty, S.C., Van Aken, S.E., Pai, G.H., Watkins, K.L., Khouri, H., Tallon, L.J., Zaborsky, J.M., Dunbar, H.E., Tran, P.L., et al. (2006). PLoS Biol. *4*, e188 10.1371/journal.pbio.0040188.

Xu, J., Mahowald, M.A., Ley, R.E., Lozupone, C.A., Hamady, M., Martens, E.C., Henrissat, B., Coutinho, P.M., Minx, P., Latreille, P., et al. (2007). PLoS Biol. *5*, e156 10.1371/journal.pbio.0050156.

Yooseph, S., Sutton, G., Rusch, D.B., Halpern, A.L., Williamson, S.J., Remington, K., Eisen, J.A., Heidelberg, K.B., Manning, G., Li, W., et al. (2007). PLoS Biol. *5*, e16 10.1371/journal.pbio.0050016.