University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Faculty Publications in Food Science and Technology

Food Science and Technology Department

2015

Metagenomics for Bacteriology

Erika del Castillo

Jacques Izard

Follow this and additional works at: https://digitalcommons.unl.edu/foodsciefacpub

Part of the Food Science Commons, and the Genomics Commons

Metagenomics for Bacteriology

Erika del Castillo

Tufts University School of Medicine

Jacques Izard

University of Nebraska–Lincoln
jizard@unl.edu

The study of bacteria, or bacteriology, has gone through transformative waves since its inception in the 1600s. It all started by the visualization of bacteria using light microscopy by Antonie van Leeuwenhoek, when he first described "animalcules." Direct cellular observation then evolved into utilizing different wavelengths on novel platforms such as electron, fluorescence, and even near-infrared microscopy. Understanding the link between microbes and disease (pathogenicity) began with the ability to isolate and cultivate organisms through aseptic methodologies starting in the 1700s. These techniques became more prevalent in the following centuries with the work of famous scientists such as Louis Pasteur and Robert Koch, and many others since then. The relationship between bacteria and the host's immune system was first inferred in the 1800s, and to date is continuing to unveil its mysteries. During the last century, researchers initiated the era of molecular genetics. The discovery of the first-generation sequencing technology, the Sanger method, and, later, the polymerase chain reaction technology propelled the molecular genetics field by exponentially expanding the knowledge of relationship

between gene structure and function. The rise of commercially available next-generation sequencing methodologies, in the beginning of this century, is drastically allowing larger amount of information to be acquired, in a manner open to the democratization of the approach.

Healthy Hosts and Microbiomes

Cooperation and association, in their broadest meanings, are ubiquitous and part of the evolutionary processes between bacteria and host. This mutually beneficial association has sustained coevolution through different habitats.

Microbiota–host cooperation starts from the moment development begins in the environment outside of the genetic progenitors, for example, the microbiota changes from postlarvae stage to the adult stage in an oyster, throughout the different stages of metamorphosis for the frog, and from birth to adulthood for mammals.[1–4] Interestingly, it seems that individual-specific strains, when established, are stable in an environment even if their relative abundance changes over time.[5,6]

To redefine the concept of health, the Human Microbiome Project (HMP) consortium recruited subjects without sign of proinflammatory condition or disease.[7,8] The studies from the acquired metagenomic data sets, from multiple body sites, show that diversity of microbes is key to health.[7,8] Other studies have shown that the microbiome influences a wide spectrum of biological events including the immune function and behavior of the host.[9–12]

Our life expectancy has drastically improved in the last 100 years. The impact of these changes on the ancestral mutualistic relationships between humans and microbes has to be part of those progresses but is not well understood. A study on calcified dental plaque has shown that from the Neolithic (remains dated 7550–5450 years before present) to the medieval times, the oral microbiota was more diverse than the present oral microbiota and was relatively stable.[13] A study of 1400-year-old coprolites from northern Mexico shows a more diverse gut microbiota compared with those of modern urbanized populations, however, more similar to rural populations with different modern life-styles.[14] Many questions remain as we are just at the beginning of our understanding on how our own microbiomes are key to our survival.

What about Disease?

Human diseases are not a new burden. At a middle-age monastic site in Germany, adult skeletons were recovered with evidence of mild-to-severe periodontitis (oral microbial infection leading to tooth loss). Using DNA extracted from the teeth of the skeletons, researchers were able to reconstruct the genome of a known pathogen, *Tannerella forsythia*,[15] and identify the molecular signatures of other periodontitis-associated species.

The treatment of disease has been an interest of any society, and microbial modification has always been a component of treatment. While plant-based therapy was probably the way to treat diseases in Neolithic times, refined chemical compounds are now available as pharmaceuticals. Regardless of the source, the microbiome can be targeted by these antimicrobials modifying community structure and metabolic potential.[16,17] Next-generation sequencing is providing a greater depth of understanding of the broader effect during treatment as well as host microbiome recovery post-treatment.[18]

Medical challenges where antimicrobial therapy has been unsuccessful have led to new approaches, such as fecal transplants. Refractory recurrent *Clostridium difficile* infections do not respond to appropriate antibiotic therapy. Fecal transplants offer the possibility of a rapid remodeling of the receiver gut microbiome toward its donor's transplant profile, and at the same time eliminate *C. difficile* challenge.[19,20]

Treatment successes and failures might have to be revisited in the context of the host–microbiome relationship. Therapeutic drugs alter the host–microbiota composition and can colocalize specific bacteria to lymphoid tissue or cells where they can synergistically modulate and influence the efficacy of the therapeutic drugs.[21] Thus, in addition of being the target, the microbiome can also act as a modulator of treatment efficacy by altering the expected effect.[21,22] A thorough understanding of the molecular bases of host–microbiota interactions could lead to the development of new therapeutic strategies for treating human disorders, as well as decreasing the toxicity of some of the present treatments.

While new approaches are being designed, the realms of traditional eastern and western medicine are slowly beginning to intersect with our increased understanding of the microbiome role in health and

disease. Traditional Chinese medicine has been widely used for millennia in the treatment of various diseases in East Asian countries. The analysis of tongue coating, a fundamental practice in Chinese medicine, has been used as a basis to differentiate the microbiota in the case of hot and cold syndromes.[23] The observed differences suggest that tongue-associated microbiomes could be used as a novel holistic biomarker to subtype human host populations.

Food, Biotransformation, and Life

Since food and nutrition are essential to the survival of all living beings on Earth, it comes as no surprise that the first metagenomic studies have focused on the gut microbiota. As the body of publication is significant, we will look at two cross-pollinations among fields.

The comparative genomic analysis of the genome of the giant panda uncovered the presence of the enzymes associated with a carnivorous digestive system while lacking the enzymes to digest cellulose, the principal component of their bamboo diet. The apparent metabolic contradiction was resolved while studying their gut microbiome. The study shows that *Clostridia* bacteria appear to be the microbial symbionts bridging this necessary metabolic gap.[24] Without the presence of *Clostridia* in the gut microbiome, the panda would not be able to survive on a diet of bamboo. The presence of stable and specific cellulose-degrading species in gut microbiome has allowed the giant panda to transition from a carnivore to a herbivore life-style, illustrating a coevolutionary process between the host and its gut microbiome.

This importance in energy balance has been underlined in metabolic transfer from bacteria to the host in obesity, in voluntary diet modification, as well as in the forced change of diet due to habitat loss.[25–27] In both humans and mouse models, it has been shown that changes to the gut ratio of Bacteroidetes/Firmicutes modulate the capacity for energy harvest, with a decrease of Bacteroidetes being associated with obesity. This correlation allowed for a better understanding of the physiology of the Australian sea lion metabolism. Their gut has a dominant composition of Firmicutes predisposing this aquatic mammal toward an excess of body fat needed for thermoregulation within their cold oceanic habitat.[25,28,29] Microbiota balance or dysbiosis

depends on the context and physiology of the host. The numbers of bacteria or genes by themselves do not provide a complete story: a larger-scale analysis is required to understand the intricacy of the microbiome relationships sustaining life.[25–27]

Some Practical Usage of the Microbiomes

The utilization of bacteria in food production by many societies/ civilizations/cultures predates modern microbiology. In Asia, before the end of the first millennia AD, a low-temperature lactic acid-based fermentation process was used to preserve food for the winter season. Now kimchi is known worldwide. Metagenomic analysis of the kimchi fermentation process led to a greater understanding of microbial community composition, pH, and respiration-associated function modulation during this month-long process.[30]

During the middle ages, the Europeans developed the process to produce the cheese products that we still enjoy. Today, the Italian Mozzarella, Grana Padano, and Parmigiano Reggiano cheeses, while from different geographic regions, are all produced by microbial communities with similar metabolic functionality, composed of thermophilic, aciduric, and moderately heat-resistant lactic acid bacteria.[31] A few additional examples of the ubiquitous use of microbiota in food are the preparation of cocoa bean in the Americas, the fermentation of millet to make boza drink in the Middle East, and the fermentation of teff to make the sourdough-risen flatbread injera in Africa.

The soil microbiome around the plant rhizosphere is modified by plant roots exudates. In agriculture, metagenomics approaches offer the potential to modify soil microbiome structure using blends of phytochemicals that might support beneficial microbiota with the goal of enhancing crop yields, sustainability, and fend off infections by maximizing a healthy plant–soil interaction.[32]

In aquaculture, metagenomics approaches can help in the design of preventive strategies with the goal of enhancing the health of the fishes by the manipulation of their gut microbiota. Recently, the gut microbiota of commercially valuable warm-water fishes, including the channel catfish and the largemouth bass, has been characterized with the goal of growth optimization and disease control.[33]

Our Societal Choices Influence the Microbiomes

As we move toward a better understanding of the intersection of human behaviors (both individual and societal), the human microbiome, and the environments in which humans live, the overall complexity drastically increases. The choices we make either as an individual or as a society influence our interactions with the diverse microbiomes surrounding us. Furthermore, the impact is not limited to us and can be positive, neutral, or negative to others. For example, the microbial communities in the drinking water distribution systems depend on the source of water, the tubing material (copper, stainless steel, or polyvinyl chloride), and the regimen and selection of disinfection methods on drinking water by private and municipal water services leading to a safe drinking water.[34–36] Although the microbiota present in the drinking water sources might be regionally or locally determined, the need for clean and safe drinking water is universal.

Architectural choices of our homes, schools, and hospitals, by the design of the airflow, the temperature, the relative humidity, and the interactive surfaces in the different sections of the rooms or buildings, influence the surrounding surface and aerosol microbiota.[37,38] Our choice of mode of transportation, whether private or public, also has an influence that might be as equivalent to our exposure to the outdoor conditions from the same environment, showing that safety exist also in numbers.[39,40]

At a larger scale, how different societies use the land and water resources can have long distance and long-term effects in the microbiome of those environments. Hurricanes, for example, are able to aerosolize a large amount of microbial cells to the upper troposphere that can potentially influence the hydrological cycle, clouds, and climate.[41] The microbiome–society interaction is bidirectional and until recently we have been largely blind to this relationship.

Recent developments create greater optimism for a better management of our inner ecology as well as the biosphere. These events include a wider spread of scientific theories, as shown by the large number of individuals taking online scientific courses,[42,43] the increasing strength of citizen science,[44,45] and a greater access to scientific tools through open-source software and scientific literature from open-access publishing.[46]

Diving into a Detailed View of the Scales Involved

Looking at a smaller scale, the coexistence of microorganisms in communities, microbial networking, and community development are at the center of the dynamic aspect of the microbiomes. Bioinformatic approaches are allowing us to redefine our understanding of the relationships between members within the communities, as well as the rules of association, competition, and exclusion.

Metagenomic approaches are finally allowing an in-depth comparative analysis of multiple sites within an individual and across populations. The first large-scale effort of this type was performed in the Sargasso Sea at different oceanic sampling stations.[47] More recently, in the cohort of the HMP, a study of 18 body sites was performed, and later was complemented by an additional selection reaching an overview of 22 human body habitats.[7,48] This biogeography is associated with the presence of relationship networks of diverse structures. Traditional microbiology has shown that these relationships can lead to direct physical interactions associated with the succession of biofilm formation, ultimately leading to an interactome.[49] When analyzing next-generation sequencing data, this network expands to co-occurrence networks, where phylotypes are typically, but not always, present together at a site.[8,23,50] Although we are far from understanding all of these relationships, a metabolic interdependence exists, because of a degradation cascade of nutriments that affect both the microbiome and the host.

Within a microenvironment, horizontal gene transfer seems to be a competitive option to complement the panel of functional capabilities, as shown by the analysis of available genomes.[51] In the specific case of the human gut bacterium *Bacteroides plebius*, the genetic exchange occurred with a marine bacteria. This gene transfer facilitates seaweed digestion in some Japanese individuals carrying *B. plebius* enhanced by this genetic addition.[52] Another available option in multispecies communities is to use mutualistic cooperation to both enhance nutriment intake and protecting themselves from the host.[41,42,53,54]

The complexity of the interactions becomes more apparent as we go deeper into the details of the massive data sets. The initial findings on the gut microbiome, from the MetaHIT project, indicates that microbial genes outnumber human genes by more than 100-fold, predicting over 3 million bacterial genes in the gut alone.[55] Multiple

scales of observation are needed, from the atomic structure modification of proteins during an enzymatic digest to the gradient of molecules within the cell, the chemotactic abilities of cells to improve their nutrient uptake or flee toxics, the surface protein providing direct interaction with other cells and to the assemblage of cells forming biofilms, and the surface to which the biofilm associates. These integrated scales of interactions, mechanistic events, and optimizations are crucial for survival, dormancy, or ability to thrive. It is up to us to understand the rules that have been in place for million of years.

What Would Help to Further the Leap?

Metagenomics heavily relies on reference databases to improve the analysis phase for phylogenetic, metabolic, and functional content including hypothetical small RNAs and proteins. Assessing the biodiversity in greater details also presents the challenge of validation in the laboratory as it is a more controlled environment.

Bacterial Systematics

For over 140 years, the world of bacterial systematics has been evolving because of technological and conceptual advances.[56] As of 2013, the number of validly named taxa rose to about 2000 genera and 10,600 species from 29 phyla (list graciously maintained by Dr Euzéby, available at www.bacterio.net). To this list, additional organisms deposited in culture collections are awaiting naming after isolation and genome sequencing during large-scale efforts such as the HMP[57,58] (list available at www.hmpdacc.org). Beyond traditional methods, whole-genome study allows proper positioning in the phylogenetic hierarchy. However, the move to whole-genomes phylogenetic analysis has been curbed, until recently, by the limited number of whole-genome and high-quality genomic sequence drafts. Additionally, new tools need to be developed to go further and define strain-level phylogeny based on genetic content.[59] This will undoubtedly bring some conflicts with the present classification as it happened when the 16S rRNA gene phylogenetic classification competed with the phenotypic classification.[56] Concurrently, databases such as the Ribosomal Database Project, Greengenes, SILVA, Human Oral Microbiome Database,

and others expand beyond officially named bacteria and maintain our ability to do 16S rRNA gene phylogenetic analyses.[60–63]

Bacterial Quantitation

Refined quantitative analysis to study the relative abundance of different bacteria will have to take into account the copy number of genes including the 16S rRNA gene. As shown in **Table 1**, the number of 16S rRNA genes can vary from 1 to 15 with no specific correlation to genome size, GC%, or membership to a specific genus or phylum. For example, two strains of the Firmicutes *Bacillus subtilis* differ by two copies (8 vs. 10), and their genome size by 4% (Table 1). Within the Proteobacteria, the GC% range from 14% to 75%, while the number of ribosomal operon varies from 1 to 15. Thus, the interpretation of microbial diversity and abundances, (relative abundance distribution, estimate of abundance of different taxa, overall diversity, and similarity measurements) based on the phylogenetically informative 16S rRNA gene quantitation, should consider the variation in both the abundance of organisms and the operon copy numbers per genome. Refined analyses will only be available for a small community where all the partners are defined. A software is available that estimates both 16S rRNA gene copy number and abundance of organisms.[64] Further efforts need to be spent to relate these 16S rRNA gene copy number with genome copy number as discussed in the text below.

Not all bacteria conform to the patterns of genome organization, chromosomal replication initiation, elongation, termination, and genomic segregation during cell division exemplified by *Escherichia coli*, whose genome is distributed in one chromosome and has only one genome copy per cell. To be truly quantitative, we will also need to understand the ploidy of each organism in function of the experimental conditions (**Table 2**). The biological significance of polyploidy will depend on the system studied and might be involved in diverse functions such as DNA recombination among genome copies, replacement of deleterious mutations through homologous recombination of genomes, or to mitigate the accumulation of deleterious mutations over time.[65–69] Additionally, the cells can replicate asynchronously, displaying a heterogeneous DNA content.[70,71] We must contend with the fact that the genome copy number can change in the different phases of growth and that more than one ploidy can be observed in

Table 1. 16S rRNA Gene Copy Numbers in a Subset of Bacterial Genomes

Phylum	Average 16S rRNA Gene Copies in Phylum[a]	Organism Name	16S rRNA Gene Copy Number[b]	Genome Size (bp)[d,e]	GC%[e]
Actinobacteria	3.1 ± 1.7	*Frankia sp.* Cc13	2	5,433,628	70.1
		Frankia sp. EuI1c	3	8,815,781	72.3
		Kineococcus radiotolerans SRS30216	4	4,956,672	74.2
Bacteroidetes	3.5 ± 1.5	Candidatus *Sulcia muelleri* DMIN	1	243,933	22.5
		Tannerella forsythia ATCC 43037	2	3,405,521	47.0
		Porphyromonas gingivalis ATCC 33277	4	2,354,886	48.4
Cyanobacteria	2.3 ± 1.2	*Synechocystis sp.* PCC 6803	2	3,947,019	47.3
Deinococcus-Thermus	2.7 ± 1.0	*Thermus thermophilus* HB-8	2	2,116,056	69.5
		Deinococcus radiodurans R1	3	3,284,156	66.6
Firmicutes	5.8 ± 2.8	*Lactobacillus casei* ATCC 334	5	2,924,325	46.6
		Staphylococcus aureus JH1	6	2,936,936	33.0
		Streptococcus pyogenes M1 GAS (SF370)	6	1,852,441	38.5
		Bacillus subtilis W23	8	4,027,676	43.9
		Bacillus subtilis 168	10	4,215,606	43.5
		Brevibacillus brevis NBRC 100599	15	6,296,436	47.3
Proteobacteria[c]	2.2 ± 1.3 (α)	*Bartonella henselae* Houston-1	2	1,931,047	38.2
		Erythrobacter litoralis HTCC2594	1	3,052,398	63.1
	3.3 ± 1.6 (β)	Candidatus *Zinderia insecticola* CARI	1	208,564	13.5
	2.7 ± 1.4 (δ)	*Anaeromyxobacter dehalogenans* 2CP-C	2	5,013,479	74.9
		Desulfovibrio vulgaris Hildenborough	5	3,773,159	63.2
	3.0 ± 1.1 (ε)	*Helicobacter pylori* 26695	2	1,667,867	38.9
		Campylobacter jejuni 269.97	3	1,845,106	30.4
	5.8 ± 2.8 (γ)	*Buchnera aphidicola* (*Acyrthosiphon pisum*)	1	655,725	26.3
		Francisella tularensis FSC147	3	1,893,886	32.3
		Aggregatibacter actinomycetemcomitans D7S-1	6	2,309,073	44.3
		Haemophilus influenzae 86-028NP	6	1,914,490	38.2
		Escherichia coli K-12 MG1655	7	4,641,652	50.8
		Yersinia pestis 91001	7	4,803,217	47.7
		Klebsiella pneumoniae HS11286	8	5,682,322	57.1
		Vibrio cholerae N16961	8	4,033,464	47.5
		Vibrio vulnificus MO6-24/O	9	5,007,768	47.0
		Aeromonas veronii B565	10	4,551,783	58.7
		Vibrio natriegens ATCC 14048	13	5,131,685	45.0
		Photobacterium profundum SS9	15	6,403,280	42.0
Spirochaetes	2.4 ± 1.0	*Borrelia burgdorferi* N40	1	1,339,539	28.6
		Treponema denticola ATCC 35405	2	2,843,201	37.9
		Treponema pallidum Chicago	2	1,139,281	52.8
Synergistetes	2.5 ± 1.0	*Anaerobaculum mobile* DSM 13181	2	2,160,700	48.0
		Thermanaerovibrio acidaminovorans DSM 6589	3	1,848,474	63.8
Tenericutes	1.6 ± 0.5	*Mycoplasma genitalium* G-37	1	580,076	31.7

a. From Vetrovsky and Baldrian.[73]
b. From the following sources: ribosomal RNA database (rrnDB).[74]
c. Values are provided for each subdivisions. (α) *Alphaproteobacteria*, (β) *Betaproteobacteria*, (δ) *Deltaproteobacteria*, (ε) *Epsilonproteobacteria, and* (γ) *Gammaproteobacteria.*
d. "bp" stands for base pairs.
e. From National Center for Biotechnology Information (NCBI) Genome Information (NCBI) Genome Information by Organism (www.ncbi.nlm.nih.gov/genome) and Kyoto Encyclopedia of Genes and Genomes (KEGG) Complete Genomes.[75]

Table 2. Genome Copy Numbers Per Cell of a Subset of Bacterial and Archaeal Species

Phylum	Organism Name	Genome Copy Number (Average or Range)	Ploidy	Generation Time, Growth Phase Environment-Free Living/Facultative-Obligate Symbiont	References
Bacteroidetes	*Blattabacterium sp.*	323–353 10–18	Polyploid Polyploid	Obligate endosymbiont of cockroach *Blattella oreintalis* Obligate endosymbiont of cockroach *Periplaneta americana*	Lopez-Sanchez et al.[76]
	Candidatus Sulcia muelleri DMIN	140–880	Polyploid	Obligate endosymbiont of green sharpshooter *Draeculecephala minerva*	Woyke et al.[77]
	Aphanizomenon ovalisporum	84–122 1–4	Polyploid Oligoploid	Akinetes (dormant spore-like cells) Vegetative cells	Sukenik et al.[78]
	Synechococcus PCC 7942	4	Oligoploid	Exponential and stationary phaes (generation time 1440 min)	Griese et al.[67]
	Synechocystis PCC 6803 Motile wild-type	218 58 58	Polyploid	Exponential phase Linear phase (1200 min) Stationary phase	Griese et al.[67]
Deinococcus-Thermus	*Deinococcus radiodurans*	10 4–8	Oligoploid	Exponential phase Stationary phase	Hansen,[79] Minton[80]
	Thermus thermophiles HB8	4–5	Oligoploid	Exponential and stationary phase (slow growth conditions)	Ohtani et al.[81]
Firmicutes	*Epulopiscium sp.* Type B	20,000–400,000 49,000–120,000	Polyploid	Symbiont of the unicornfish *Naso tonganus* symbiont	Mendell et al.,[82] Angert[83]
	Lactobacillus lactis subsp. lactis IL1403	2	Diploid	Doubling time 223 min (slow growing culture)	Michelsen et al.,[84]
Proteobacteria	*Azotobacter vinelandii*	>40 >80 >100	Polyploid	Late exponential phase Early stationary phase Late stationary phase	Nagpal et al.,[85] Maldonado et al.[86]
	Buchnera sp.	120 (50–200)	Polyploid	Obligate endosymbiont of the aphid *Acyrthosiphon pisum;* genome copy number varies with host developmental stage	Komaki and Ishikawa[69,87]
	Caulobacter crescentus	2.1	Monoploid	Doubling time 93 min	Pecoraro et al.[68]
	Desulfovibrio vulgaris	4	Oligoploid	Doubling time 2400 min	Postgate et al.[88]
	Escherichia coli	2.5/1.2[a] 6.8/1.7[a]	Monoploid Merooligoploid	Doubling time 103 min Doubling time 25 min	Pecoraro et al.[68]
	Neisseria gonorrhoeae	3	Oligoploid	Exponential phase (generation time 60 min)	Tobiason and Seifert[89]
	Pseudomonas putida	20/14[a]	Polyploid	Doubling time 46 min	Pecoraro et al.[68]
	Wolinella succinogenes	0.9	Monoploid	Doubling time 96 min	Pecoraro et al.[68]
Spirochaetes	*Borrelia hermsii*	5 14 (12–17)	Oligoploid Polyploid	Late exponential phase (maintained in laboratory) Isolated from mice	Kitten and Barbour[90]
Euryarchaeota	*Methanococcus maripaludis*	55 30	Polyploid	Exponential phase Stationary phase	Hildenbrand et al.[91]
	Methanothermobacter thermoautotrophicus	2 1–2	Diploid	Exponential phase Stationary phase	Majernik et al.[92]

a. Based on gene copy number near origin/gene copy number near the termini.

a population.[70,72] An understanding of the role of polyploidy and replication will provide insights into the extent the structure and content of the genome influences the phenotypic features of cells with multiple genomes, as well as influence the data from each "omics" platforms. In some remarkable cases there is a complementation of the physiology of both hosts and their polyploid symbionts, and these functional interactions remain to be elucidated.

Defining What Is a Strain

Bacteria, both in the laboratory and in nature, are studied at the population level. Bacterial populations are not composed of millions of identical individuals. During cell duplication, the genomes of individual cells are subjected to mutations, producing a genetically heterogeneous population within a species. Large-scale metagenomic studies reveal that microbial communities are predominantly organized in sequence-discrete populations, and the genomes of the organisms within those populations share higher than 94% average nucleotide identity (ANI). These sequence-discrete populations are important units within natural microbial communities. Bacteria that belong to a particular population, but of different environment, significantly show less genetic identity to other co-occurring populations, typically less than 80–85% ANI. This genetic metric offers higher resolution than the widely used 16S rRNA gene sequencing analysis.[93,94] Defining strain might be contextual at first, until we have a more complete view of cell evolution. To facilitate the process, culture-independent "*omics*" techniques (transcriptomics, proteomics, and metabolomics) might further refine the taxonomical assignment and provide ecologically relevant properties of natural microbial populations. Quantification of yet-to-be-cultivated bacteria can be improved with the characterization of ecologically appropriate genes and pathways in sequence-discrete populations, which uniquely define the population genomic signatures.

Expanding Gene Catalogs

Identifying the genetic content of a microbiome is the first layer provided by the new generations of sequencing machines. From a metagenome or a metatranscriptome, the avalanche of information needs to be transformed in order to go beyond a simple comparison of gene counts. Genomic sequences from reference genomes are used

in multiple aspects of the analysis, including gene definition, gene function, taxonomy, and so on. The first genome sequenced was isolated from the bacteria *Haemophilus influenza* in 1995.[95] Since then, the number of genome sequences has been growing rapidly and can be found in international depositories comprising DNA Data Bank of Japan (DDBJ), the European Nucleotide Archive (ENA), and the genetic sequence database of the National Center for Biotechnology Information (NCBI) of the United States (GenBank), as well as more specialized repositories. However, the number of reference genomes needs to increase to keep pace with advancements in metagenomics. Beyond cultivability, gene catalogs and single cell genomes will increase the pool of information to infer additional layers of analysis.[96–101]

Making Reference Strains Available

Presently, the number of cultivable strains deposited in reference strain depositories that are not yet sequenced is decreasing because of international efforts. The next frontier is in obtaining strains that were previously thought to be uncultivable. Some of the strains previously classified as "yet-to-be-cultivable" are now deposited at the American Type Culture Collection (ATCC) and sequenced by the means of the HMP,[99] awaiting further functional studies.

For a successful understanding of the microbiome and its interaction with the environment, novel large-scale investigations into the biology of single organisms and ecological models that integrate phylogenetic and functional relationships among organisms are required. Bacterial isolates available now or in the future will enable both biochemical-based study of their dynamic genomes and culture-based studies of their functional role in microbial communities. This will aid in improving assembly and annotation of metagenomes, and in quantification of microbial communities in their residing habitats.

Metabolic Potential

Bacteria exist in a wide range of environments and have extremely diverse physiological capabilities. Microbiome functionality can be derived either from gene-based knowledge or the intersection of other omics including metagenomics. Metabolism is key for the living cell. Databases such as KEGG, MetaCyc, Carbohydrate-Active enZYmes Database (CAZy), and Braunschweig Enzyme Database (BRENDA) considerably enhance our ability to create inferences leading to a greater

understanding of single species or a complex community.[102–105] However, metabolism is not the only cell function, of which many aspects still remain unknown. For example, there is a large number of conserved proteins in international depositories for which a function needs to be identified to improve our understanding of the proteome.[106–108]

Learning About Archaea

Most previous work has focused on bacteria, as information about archaea is still nascent. Limited information is emerging about human– archaeal associations and the role of these organisms in human physiology. Much remains to be known about archaeal phylogenetic diversity, abundance, and biochemistry *in situ*. Current molecular approaches can reveal the genomic dynamics of methanogenic archaea associated with humans. These include *Methanobrevibacter smithii,* a methane producer predominant in human colon and also present in the vagina, *Methanobrevibacter oralis,* which has been associated with subgingival diseases and is capable to thrive at low pH in the stomach, and various other methanogens including *Methanosphaera stadtmanae*, *Methanobrevibacter millerae*, and *Methanobrevibacter arboriphilus*.[91,109–112] In the upcoming years, we need to expand our understanding of the role of archaea in the human microbiomes, as their transcripts are overabundant compared with their cell relative abundance.[113]

In closing, novel approaches are essential to properly integrate metagenomics, proteomics, lipidomics, and metabolomics in a comprehensive and integrative conceptual framework. Proper annotation of data sets is the first step in this direction by using minimum information standards when depositing the data sets and the annotations, and standardizing the names of body sites as well as of other descriptive components.[114–116] This opportunity is offered to all of us.

References

1. Ringel-Kulka T, Cheng J, Ringel Y, Salojarvi J, Carroll I, Palva A, et al. Intestinal microbiota in healthy U.S. young children and adults – a high throughput microarray analysis. PloS One 2013;8(5):e64315.

2. Jost T, Lacroix C, Braegger CP, Chassard C. New insights in gut microbiota establishment in healthy breast fed neonates. PloS One 2012;7(8):e44595.

3. Trabal Fernandez N, Mazon-Suastegui JM, Vazquez-Juarez R, Ascencio-Valle F, Romero J. Changes in the composition and diversity of the bacterial microbiota associated with oysters (*Crassostrea corteziensis, Crassostrea gigas* and *Crassostrea sikamea*) during commercial production. FEMS Microbiol Ecol 2014;88(1):69–83.

4. Kohl KD, Cary TL, Karasov WH, Dearing MD. Restructuring of the amphibian gut microbiota through metamorphosis. Environ Microbiol Rep 2013;5(6):899–903.

5. Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, et al. The long-term stability of the human gut microbiota. Science 2013;341(6141):1237439.

6. Schloissnig S, Arumugam M, Sunagawa S, Mitreva M, Tap J, Zhu A, et al. Genomic variation landscape of the human gut microbiome. Nature 2013;493(7430):45–50.

7. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature 2012;486(7402):207–14.

8. Segata N, Haake SK, Mannon P, Lemon KP, Waldron L, Gevers D, et al. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. Genome Biol 2012;13(6):R42.

9. Holdeman LV, Good IJ, Moore WE. Human fecal flora: variation in bacterial composition within individuals and a possible effect of emotional stress. Appl Environ Microbiol 1976;31(3):359–75.

10. Tillisch K, Labus J, Kilpatrick L, Jiang Z, Stains J, Ebrat B, et al. Consumption of fermented milk product with probiotic modulates brain activity. Gastroenterology 2013;144(7):1394–401.

11. Round JL, Mazmanian SK. The gut microbiota shapes intestinal immune responses during health and disease. Nat Rev Immunol 2009;9(5):313–23.

12. Diaz Heijtz R, Wang S, Anuar F, Qian Y, Bjorkholm B, Samuelsson A, et al. Normal gut microbiota modulates brain development and behavior. Proc Natl Acad Sci USA 2011;108(7): 3047–52.

13. Adler CJ, Dobney K, Weyrich LS, Kaidonis J, Walker AW, Haak W, et al. Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. Nat Genet 2013;45(4):450–5.

14. Tito RY, Knights D, Metcalf J, Obregon-Tito AJ, Cleeland L, Najar F, et al. Insights from characterizing extinct human gut microbiomes. PloS One 2012;7(12):e51146.

15. Warinner C, Rodrigues JF, Vyas R, Trachsel C, Shved N, Grossmann J, et al. Pathogens and host immunity in the ancient human oral cavity. Nat Genet 2014;46(4):336–344.

16. Perez-Cobas AE, Artacho A, Knecht H, Ferrus ML, Friedrichs A, Ott SJ, et al. Differential effects of antibiotic therapy on the structure and function of human gut microbiota. PloS One 2013;8(11):e80201.

17. Dethlefsen L, Huse S, Sogin ML, Relman DA. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. PLoS Biol 2008;6(11):e280.

18. Hill DA, Hoffmann C, Abt MC, Du Y, Kobuley D, Kirn TJ, et al. Metagenomic analyses reveal antibiotic-induced temporal and spatial changes in intestinal microbiota with associated alterations in immune cell homeostasis. Mucosal Immunol 2010;3(2):148–58.

19. Brace C, Gloor GB, Ropeleski M, Allen-Vercoe E, Petrof EO. Microbial composition analysis of *Clostridium difficile* infections in an ulcerative colitis patient treated with multiple fecal microbiota transplantations. J Crohn's Colitis 2014;8(9):1133–1137.

20. Song Y, Garg S, Girotra M, Maddox C, von Rosenvinge EC, Dutta A, et al. Microbiota dynamics in patients treated with fecal microbiota transplantation for recurrent *Clostridium difficile* infection. PloS One 2013;8(11):e81330.

21. Viaud S, Saccheri F, Mignot G, Yamazaki T, Daillere R, Hannani D, et al. The intestinal microbiota modulates the anticancer immune effects of cyclophosphamide. Science 2013; 342(6161):971–6.

22. Goleva E, Jackson LP, Harris JK, Robertson CE, Sutherland ER, Hall CF, et al. The effects of airway microbiome on corticosteroid responsiveness in asthma. Am J Respir Crit Care Med 2013;188(10):1193–201.

23. Jiang B, Liang X, Chen Y, Ma T, Liu L, Li J, et al. Integrating next-generation sequencing and traditional tongue diagnosis to determine tongue coating microbiome. Sci Rep 2012;2:936.

24. Zhu L, Wu Q, Dai J, Zhang S, Wei F. Evidence of cellulose metabolism by the giant panda gut microbiome. Proc Natl Acad Sci USA 2011;108(43):17714–9.

25. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. Nature 2009;457(7228):480–4.

26. Amato KR, Yeoman CJ, Kent A, Righini N, Carbonero F, Estrada A, et al. Habitat degradation impacts black howler monkey (*Alouatta pigra*) gastrointestinal microbiomes. ISME J 2013;7(7):1344–53.

27. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly and reproducibly alters the human gut microbiome. Nature 2014;505(7484):559–63.

28. Lavery TJ, Roudnew B, Seymour J, Mitchell JG, Jeffries T. High nutrient transport and cycling potential revealed in the microbial metagenome of Australian sea lion (*Neophoca cinerea*) faeces. PloS One 2012;7(5):e36478.

29. Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, et al. Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science 2013;341(6150): 1241214.

30. Jung JY, Lee SH, Kim JM, Park MS, Bae JW, Hahn Y, et al. Metagenomic analysis of kimchi, a traditional Korean fermented food. Appl Environ Microbiol 2011;77(7):2264–74.

31. De Filippis F, La Storia A, Stellato G, Gatti M, Ercolini D. A selected core microbiome drives the early stages of three popular italian cheese manufactures. PloS One 2014;9(2):e89680.

32. Badri DV, Chaparro JM, Zhang R, Shen Q, Vivanco JM. Application of natural blends of phytochemicals derived from the root exudates of Arabidopsis to the soil reveal that phenolic-related compounds predominantly modulate the soil microbiome. J Biol Chem 2013;288(7):4502–12.

33. Larsen AM, Mohammed HH, Arias CR. Characterization of the gut microbiota of three commercially valuable warmwater fish species. J Appl Microbiol 2014;116(6):1396–1404.

34. Hwang C, Ling F, Andersen GL, LeChevallier MW, Liu WT. Microbial community dynamics of an urban drinking water distribution system subjected to phases of chloramination and chlorination treatments. App Environ Microbiol 2012;78(22):7856–65.

35. Hong PY, Hwang C, Ling F, Andersen GL, LeChevallier MW, Liu WT. Pyrosequencing analysis of bacterial biofilm communities in water meters of a drinking water distribution system. App Environ Microbiol 2010;76(16):5631–5.

36. Holinger EP, Ross KA, Robertson CE, Stevens MJ, Harris JK, Pace NR. Molecular analysis of point-of-use municipal drinking water microbiology. Water Res 2014;49:225–35.

37. Kembel SW, Jones E, Kline J, Northcutt D, Stenson J, Womack AM, et al. Architectural design influences the diversity and structure of the built environment microbiome. ISME J 2012;6(8):1469–79.

38. Meadow JF, Altrichter AE, Kembel SW, Moriyama M, O'Connor TK, Womack AM, et al. Bacterial communities on classroom surfaces vary with human contact. Microbiome 2014;2(1):7.

39. Stephenson RE, Gutierrez D, Peters C, Nichols M, Boles BR. Elucidation of bacteria found in car interiors and strategies to reduce the presence of potential pathogens. Biofouling 2014;30(3):337–46.

40. Robertson CE, Baumgartner LK, Harris JK, Peterson KL, Stevens MJ, Frank DN, et al. Culture-independent analysis of aerosol microbiology in a metropolitan subway system. App Environ Microbiol 2013;79(11):3485–93.

41. DeLeon-Rodriguez N, Lathem TL, Rodriguez RL, Barazesh JM, Anderson BE, Beyersdorf AJ, et al. Microbiome of the upper troposphere: species composition and prevalence, effects of tropical storms, and atmospheric implications. Proc Natl Acad Sci USA 2013;110(7):2575–80.

42. Costello V. Students Aged 9 to 65+ Study PLOS Research in Marine Megafauna MOOC. Web blog post Diverse perspectives on science and medicine PLOS Blogs 2014, 4 April 2014.

43. Ho AD, Reich J, Nesterko SO, Seaton DT, Mullaney T, Waldo J, et al. HarvardX and MITx: The First Year of Open Online Courses, Fall 2012-Summer 2013 Social Science Research Network SSRN Electronic Journal pp: 33 2014.

44. Crall AW, Jordan R, Holfelder K, Newman GJ, Graham J, Waller DM. The impacts of an invasive species citizen science training program on participant attitudes, behavior, and science literacy. Public Understand Sci 2013;22(6):745–64.

45. Janssens AC, Kraft P. Research conducted using data obtained through online communities: ethical implications of methodological limitations. PLoS Med 2012;9(10):e1001328.

46. Davis PM, Walters WH. The impact of free access to the scientific literature: a review of recent research. J Med Lib Assoc 2011;99(3):208–17.

47. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, et al. Environmental genome shotgun sequencing of the Sargasso Sea. Science 2004;304(5667):66–74.

48. Zhou Y, Gao H, Mihindukulasuriya KA, Rosa PS, Wylie KM, Vishnivetskaya T, et al. Biogeography of the ecosystems of the healthy human body. Genome Biol 2013;14(1):R1.

49. Kolenbrander PE, Palmer RJ, Periasamy S, Jakubovics NS. Oral multispecies biofilm development and the key role of cell-cell distance. Nat Rev Micro 2010;8(7):471–80.

50. Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, et al. Microbial Co-occurrence Relationships in the Human Microbiome. PLoS Computat Biol 2012;8(7):e1002606.

51. Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ. Ecology drives a global network of gene exchange connecting the human microbiome. Nature 2011;480(7376):241–4.

52. Hehemann JH, Correc G, Barbeyron T, Helbert W, Czjzek M, Michel G. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. Nature 2010;464(7290):908–12.

53. Ramsey MM, Whiteley M. Polymicrobial interactions stimulate resistance to host innate immunity through metabolite perception. Proc Natl Acad Sci USA 2009;106(5):1578–83.

54. Sieber JR, McInerney MJ, Gunsalus RP. Genomic insights into syntrophy: the paradigm for anaerobic metabolic cooperation. Annu Rev Microbiol 2012;66:429–52.

55. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 2010;464(7285):59–65.

56. Stackebrandt E. Forces shaping bacterial systematics. Microbe 2007;2(6):283–8.

57. Nelson KE, Weinstock GM, Highlander SK, Worley KC, Creasy HH, Wortman JR, et al. A catalog of reference genomes from the human microbiome. Science 2010;328(5981):994–9.

58. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu WH, et al. The human oral microbiome. J Bacteriol 2010;192(19):5002–17.

59. Huang K, Brady A, Mahurkar A, White O, Gevers D, Huttenhower C, et al. MetaRef: a pan-genomic database for comparative and community microbial genomics. Nucleic Acids Res 2014;42(1):D617–24.

60. Chen T, Yu WH, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE. The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. Database: J Biol Databases Curat 2010;2010. baq013.

61. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res 2014;42(Database issue): D633–642.

62. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J 2012;6(3):610–8.

63. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 2013;41(Database issue):D590–6.

64. Kembel SW, Wu M, Eisen JA, Green JL. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. PLoS Computat Biol 2012;8(10):e1002743.

65. Slade D, Lindner AB, Paul G, Radman M. Recombination and replication in DNA repair of heavily irradiated *Deinococcus radiodurans*. Cell 2009;136(6):1044–55.

66. Zahradka K, Slade D, Bailone A, Sommer S, Averbeck D, Petranovic M, et al. Reassembly of shattered chromosomes in *Deinococcus radiodurans*. Nature 2006;443(7111):569–73.

67. Griese M, Lange C, Soppa J. Ploidy in cyanobacteria. FEMS Microbiol Lett 2011;323(2): 124–31.

68. Pecoraro V, Zerulla K, Lange C, Soppa J. Quantification of ploidy in proteobacteria revealed the existence of monoploid, (mero-)oligoploid and polyploid species. PloS One 2011;6(1):e16392.

69. Komaki K, Ishikawa H. Genomic copy number of intracellular bacterial symbionts of aphids varies in response to developmental stage and morph of their host. Insect Biochem Mol Biol 2000;30(3):253–8.

70. Muller S, Babel W. Analysis of bacterial DNA patterns – an approach for controlling biotechnological processes. J Microbiol Methods 2003;55(3):851–8.

71. Muller S, Nebe-von-Caron G. Functional single-cell analyses: flow cytometry and cell sorting of microbial populations and communities. FEMS Microbiol Rev 2010;34(4):554–87.

72. Caro A, Gros O, Got P, De Wit R, Troussellier M. Characterization of the population of the sulfur-oxidizing symbiont of *Codakia orbicularis* (Bivalvia, Lucinidae) by single-cell analyses. Appl Environ Microbiol 2007;73(7):2101–9.

73. Vetrovsky T, Baldrian P. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. PloS One 2013;8(2):e57923.

74. Lee ZM, Bussema C 3rd, Schmidt TM. rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. Nucleic Acids Res 2009;37(Database issue):D489–93.

75. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking genomes to life and the environment. Nucleic Acids Res 2008;36(Database issue):D480–4.

76. Lopez-Sanchez MJ, Neef A, Patino-Navarrete R, Navarro L, Jimenez R, Latorre A, et al. Blattabacteria, the endosymbionts of cockroaches, have small genome sizes and high genome copy numbers. Environ Microbiol 2008;10(12):3417–22.

77. Woyke T, Tighe D, Mavromatis K, Clum A, Copeland A, Schackwitz W, et al. One bacterial cell, one complete genome. PloS One 2010;5(4):e10314.

78. Sukenik A, Kaplan-Levy RN, Welch JM, Post AF. Massive multiplication of genome and ribosomes in dormant cells (akinetes) of Aphanizomenon ovalisporum (Cyanobacteria). ISME J 2012;6(3):670–9.

79. Hansen MT. Multiplicity of genome equivalents in the radiation-resistant bacterium *Micrococcus radiodurans*. J Bacteriol 1978;134(1):71–5.

80. Minton KW. DNA repair in the extremely radioresistant bacterium *Deinococcus radiodurans*. Mol Microbiol 1994;13(1):9–15.

81. Ohtani N, Tomita M, Itaya M. An extreme thermophile, *Thermus thermophilus*, is a polyploid bacterium. J Bacteriol 2010;192(20):5499–505.

82. Mendell JE, Clements KD, Choat JH, Angert ER. Extreme polyploidy in a large bacterium. Proc Natl Acad Sci USA 2008;105(18):6730–4.

83. Angert ER. DNA replication and genomic architecture of very large bacteria. Annu Rev Microbiol 2012;66:197–212.

84. Michelsen O, Hansen FG, Albrechtsen B, Jensen PR. The MG1363 and IL1403 laboratory strains of *Lactococcus lactis* and several dairy strains are diploid. J Bacteriol 2010;192(4): 1058–65.

85. Nagpal P, Jafri S, Reddy MA, Das HK. Multiple chromosomes of *Azotobacter vinelandii*. J Bacteriol 1989;171(6):3133–8.

86. Maldonado R, Jimenez J, Casadesus J. Changes of ploidy during the *Azotobacter vinelandii* growth cycle. J Bacteriol 1994;176(13):3911–9.

87. Komaki K, Ishikawa H. Intracellular bacterial symbionts of aphids possess many genomic copies per bacterium. J Mol Evol 1999;48(6):717–22.

88. Postgate JR, Kent HM, Robson RL, Chesshyre JA. The genomes of *Desulfovibrio gigas* and *D. vulgaris*. J Gen Microbiol 1984;130(7):1597–601.

89. Tobiason DM, Seifert HS. The obligate human pathogen, *Neisseria gonorrhoeae*, is polyploid. PLoS Biol 2006;4(6):e185.

90. Kitten T, Barbour AG. The relapsing fever agent *Borrelia hermsii* has multiple copies of its chromosome and linear plasmids. Genetics 1992;132(2):311–24.

91. Hildenbrand C, Stock T, Lange C, Rother M, Soppa J. Genome copy numbers and gene conversion in methanogenic archaea. J Bacteriol 2011;193(3):734–43.

92. Majernik AI, Lundgren M, McDermott P, Bernander R, Chong JP. DNA content and nucleoid distribution in *Methanothermobacter thermautotrophicus*. J Bacteriol 2005;187(5):1856–8.

93. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 2007;57(Pt 1):81–91.

94. Caro-Quintero A, Konstantinidis KT. Bacterial species may exist, metagenomics reveal. Environ Microbiol 2012;14(2):347–55.

95. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 1995;269(5223):496–512.

96. Kyrpides NC, Hugenholtz P, Eisen JA, Woyke T, Goker M, Parker CT, et al. Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. PLoS Biol 2014;12(8):e1001920.

97. Beall CJ, Campbell AG, Dayeh DM, Griffen AL, Podar M, Leys EJ. Single cell genomics of uncultured, health-associated *Tannerella* BU063 (Oral Taxon 286) and comparison to the closely related pathogen *Tannerella forsythia*. PLoS One 2014;9(2):e89398.

98. Mason OU, Hazen TC, Borglin S, Chain PS, Dubinsky EA, Fortney JL, et al. Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to deepwater horizon oil spill. ISME J 2012;6(9):1715–27.

99. Human Microbiome Project Consortium. A framework for human microbiome research. Nature 2012;486(7402):215–21.

100. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 2010;464(7285):59–65.

101. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, et al. Insights into the phylogeny and coding potential of microbial dark matter. Nature 2013;499(7459):431–7.

102. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 2014;42(1):D490–5.

103. Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA, Gilham F, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res 2010;38(Database issue):D473–9.

104. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res 2010;38 (Database issue):D355–60.

105. Schomburg I, Chang A, Placzek S, Sohngen C, Rother M, Lang M, et al. BRENDA in 2013 integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. Nucleic Acids Res 2013;41(Database issue):D764–72.

106. Wilke A, Harrison T, Wilkening J, Field D, Glass EM, Kyrpides N, et al. The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. BMC Bioinform 2012;13:141.

107. UniProt C. Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res 2014;42(Database issue):D191–8.

108. Goodacre NF, Gerloff DL, Uetz P. Protein domains of unknown function are essential in bacteria. mBio 2013;5(1):e00744–e813.

109. Samuel BS, Hansen EE, Manchester JK, Coutinho PM, Henrissat B, Fulton R, et al. Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. Proc Natl Acad Sci USA 2007;104(25):10643–8.

110. Dridi B, Raoult D, Drancourt M. Archaea as emerging organisms in complex human microbiomes. Anaerobe 2011;17(2):56–63.

111. Aminov RI. Role of archaea in human disease. Front Cell Infect Microbiol 2013;3:42.

112. Khelaifia S, Garibal M, Robert C, Raoult D, Drancourt M. Draft genome sequence of a human-associated isolate of *Methanobrevibacter arboriphilicus*, the Lowest-G+C-content archaeon. Genome Announce 2014;2(1):e01181-13.

113. Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM, et al. Relating the metatranscriptome and metagenome of the human gut. Proc Natl Acad Sci USA 2014;111(22): E2329–2338.

114. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, et al. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat Biotechnol 2011;29(5):415–20.

115. Field D, Amaral-Zettler L, Cochrane G, Cole JR, Dawyndt P, Garrity GM, et al. The Genomic Standards Consortium. PLoS Biol 2011;9(6):e1001088.

116. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol 2008;26(5):541–7.