Microbiology Insights

Libertas Academica
FREEDOM TO RESEARCH

OPEN ACCESS
Full open access to this and thousands of other papers at http://www.la-press.com.

REVIEW

The Use of Metagenomic Approaches to Analyze Changes in Microbial Communities

Girish Neelakanta and Hameeda Sultana

Center for Molecular Medicine, Department of Biological Sciences, Old Dominion University, Norfolk, VA, USA.
Corresponding author email: hsultana@odu.edu

Abstract: Microbes are the most abundant biological entities found in the biosphere. Identification and measurement of microorganisms (including viruses, bacteria, archaea, fungi, and protists) in the biosphere cannot be readily achieved due to limitations in culturing methods. A non-culture based approach, called "metagenomics", was developed that enabled researchers to comprehensively analyse microbial communities in different ecosystems. In this study, we highlight recent advances in the field of metagenomics for analyzing microbial communities in different ecosystems ranging from oceans to the human microbiome. Developments in several bioinformatics approaches are also discussed in context of microbial metagenomics that include taxonomic systems, sequence databases, and sequence-alignment tools. In summary, we provide a snapshot for the recent advances in metagenomics approach for analyzing changes in the microbial communities in different ecosystems.

Keywords: metagenomics, sequencing, microbial diversity, bioinformatics, microbial changes

Microbiology Insights 2013:6 37–48

doi: 10.4137/MBI.S10819

This article is available from http://www.la-press.com.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article published under the Creative Commons CC-BY-NC 3.0 license.

Introduction

In the natural environment, constant polymicrobial interaction(s) occur between bacteria, viruses, protozoa, protists, archaea and fungi. These microbes do not exist in isolation and are often found in a dynamic "consortia" of different microbial species populations.[1] Understanding microbial population dynamics in a consortium will benefit the genomic information of all coexisting members. Isolating and sequencing the genome of an individual organism from a consortium might not be adequate as the single isolate cannot be a representative of the full genetic and metabolic potential of its associated members. Moreover, achieving culture conditions for isolating a single member from a consortium would be a daunting task. Traditional microbiologists were always dependent on culture-based techniques for the identification of microbes in environmental samples. The challenge of identifying uncultured organisms was totally ignored. However, an explosion of knowledge in the field of microbial physiology and genetics happened during 1960s to mid-1980s wherein some scientists came to believe that cultured microorganisms did not represent the whole microbial world. This was evidenced by the "great plate count anomaly" showing discrepancy in the microbial numbers between dilution plating and microscopy.[2] From then on, several independent studies supported the rise of this uncultured world of microbes.[3]

New non-culture based approaches have recently been developed that can be extensively used for comprehensive analysis of different communities in a microbial consortia.[1,4–6] Metagenomics or genomic studies of microorganisms refer to an non-culture based approach for collectively studying sets of genomes from a mixed population of microbes.[1] The term "Metagenomics" was first coined by Handelsman and his colleagues in their study of natural products from soil microbes.[1] Community genomics, environmental genomics, and population genomics are often used as synonyms for metagenomics. The field of metagenomics was initially started by an idea from Pace in 1985[7] that subsequently lead to several studies, starting from the first cloning of DNA directly from environmental samples in a phage vector,[8] and culminating in the direct random shotgun sequencing of environmental DNA.[9,10] Since the use of the metagenomics approach in these independent studies, several other studies have come to use this approach to study microbial populations in a wide range of samples ranging from the oceans to humans.[8,11–22] Metagenomics has also provided significant information on the "changes" in the microbial community. For example, using a metagenomic approach, studies have elucidated changes in the microbial composition in humans fed on different diets.[22,23] Similarly, metagenomics have provided information in the changes of microbial composition in ticks collected from different geographic regions.[15]

Metagenomic studies can be grouped into four categories based on different screening methods: (a) shotgun analysis using mass genome sequencing; (b) genomic activity-driven studies designed to search for specific microbial functions; (c) genomic sequence studies using phylogenetic or functional gene expression analysis; and (d) next generation sequencing technologies for determining whole gene content in environmental samples.[4,6,18,24–28] These four methods can be sub-classified under unselective (shotgun analysis and next generation sequencing) and targeted (activity-driven and sequence-driven studies) metagenomics.[4,6,15,24,26,28,29] Some studies have used an unselective metagenomic approach extensively because of its cost-effectiveness and simplicity in DNA sequencing.[30]

In this review we summarize recent advances in the field of metagenomics in studying changes in bacterial and viral communities from different ecosystems, provide a snapshot of metagenomic analysis and applications of the metagenomic approach, and discuss the development of some of the approaches to answer the challenges faced in accessing metagenomic data.

Experimental Design for Metagenomic Analysis

A common sequence-based metagenomic approach involves steps that are outlined in Figure 1. Due to high experimental costs incurred in metagenomics projects, there is a definite requirement of proper experimental design with appropriate replication and statistical analyses. For example, the approximate cost to produce metagenomic data from one gram of soil requiring 6000 HiSeq2000 runs would cost $ 267 million.[31] A proper experimental design should ideally start with a question rather than technical or

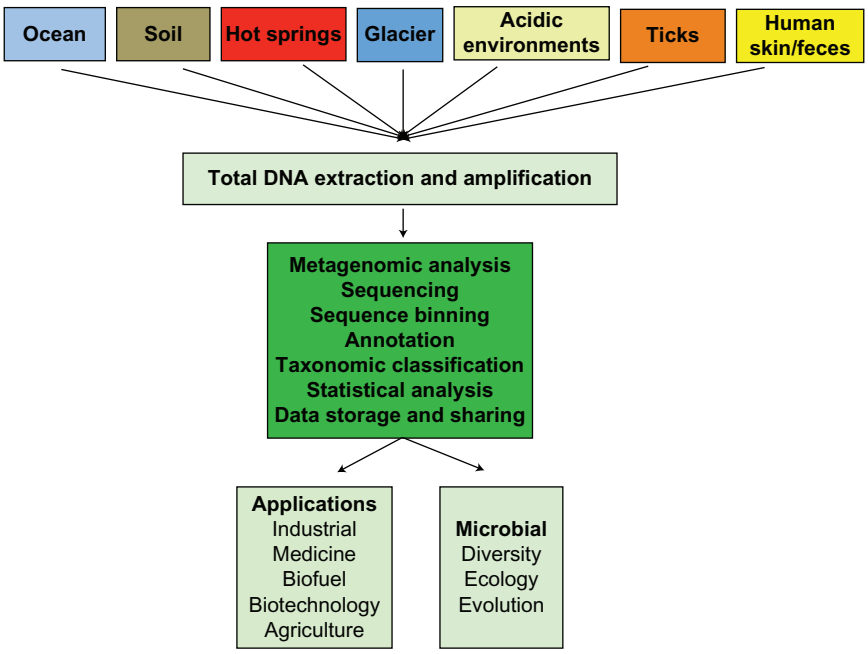

Figure 1. Overview of metagenomic analysis.
Notes: Schematic representation of a typical metagenomic analysis is shown. Samples from various sources such as from Ocean, soil, hot springs, glaciers, acidic environments, ticks and human skin and feces samples are processed for total DNA extraction to amplify microbial sequences. The extracted DNA is then processed for metagenomic analysis that is comprised of the following steps: sequencing; sequence binning; annotation of sequences; taxonomic classification of microbial species; statistical analysis of the metagenomic data; and data storage in central metagenome databases. Some of the potential coding sequences that include but are not limited to enzymes, antibiotics, and proteases are cloned into heterologous expression vectors. The expressed proteins are later used in variety of applications. In addition, the information obtained from typical metagenomic analysis would provide substantial insights in the field of microbial diversity, ecology, and evolution.

operational restriction. As the ultimate aim of metagenomic projects is to link functional and phylogenetic information of microbial communities to the chemical, physical, and other biological parameters that characterize the environment, suitable reference samples for comparison should be considered and emphasized in the experimental design. The biological or technical variations that may arise during the experiment should not be neglected and should instead be considered carefully in planning the experiment. As microbial systems are dynamic, temporal sampling from an environment can have substantial impact on data and interpretation. Proper replicates need to be included in the experimental design and should also consider the level at which replication takes place. In summary, a well-planned experimental design in metagenomic projects would facilitate integration of data sets into new or existing ecological models.[32]

Sample Processing for Metagenomic Analysis

Sample processing is the first step of any metagenomic project. The DNA that will be used for metagenomic analysis should be representative of all cells present in the sample and should be ideal for generation of genomic libraries. High quality DNA extractions that include robust DNA extraction procedures are now readily available.[10,32,33] Some of the common DNA extraction procedures, such as use of fractionation or selective lysis for isolating target DNA associated with a host,[10,32–34] physical separation, and isolation of cells from the samples (eg, soil samples) or Direct lysis of cells in the soil matrix, have been reported.[33] Metagenomic analysis requires high nanogram to microgram amounts of DNA.[33,35] In the case of samples that yield less DNA, amplification methods for the DNA is recommended. Multiple displacement amplification using random hexamers and phage phi29 polymerase has been reported to successfully amplify femtograms of DNA in order to produce micrograms of product.[36,37]

Metagenomic Sequencing

The metagenomics approach was originally focused on bacterial communities, but since been used to explore a wide range of microorganisms.[1,5,20]

Recently, several methods including Shotgun sequencing have been extensively used in metagenomic studies.[38] In Shotgun metagenomics, DNA isolated from an environmental sample is randomly sheared, sequenced in short fragments, and reconstructed into consensus sequences. With this method, detection of several microbes that would otherwise go unnoticed in culturing techniques was successful in environmental samples.[38] With the recent development of next-generation sequencing (NGS) (both 454/Roche and Illumina/Solexa systems), the whole of metagenomic sequencing has shifted from Sanger sequencing technology.[39,40] However, Sanger sequencing is still considered for sequences with large insert sizes and a read length exceeding 700 base pairs.[41] Emulsion polymerase chain reaction is performed to clonally amplify random DNA fragments that are then attached to microscopic beads when NGS is performed using 454/Roche sequencer. The Beads attached to DNA fragments are deposited into picotitre plate followed by individual and parallel pyrosequencing. In the case of the Illumina/Solexa system, DNA fragments are immobilized on a surface and then solid-surface PCR amplification is performed. The amplified DNA fragments are then sequenced using reversible terminators in a sequencing-by-synthesis process.[42]

A typical bacterial metagenomic analysis of environmental bacteria survey requires the use of the whole 16S ribosomal RNA (rRNA) gene.[1,4] However, due to the read length restriction in NGS procedures, most surveys are aimed at characterizing selected hyper-variable regions of the 16S rRNA gene.[15,43,44] The primary and secondary structures of the 16S rRNA gene show nine hyper-variable regions flanked by relatively conserved regions.[15,43,44] This property makes hyper variable regions of 16S rRNA gene an optimal species molecular marker.[45] Recent studies have shown comparable results between sequencing of hyper-variable regions and sequencing of a full-length 16S rRNA gene.[46,47] Based on these studies, it is recommended to design oligonucleotides for the V1-V3 region or V4-V7 region for Archaea and the V1-V3 region or V1-V4 region for bacteria.

Metagenomic Sequence Assembly, Binning, and Annotation

The sequenced DNA fragments are then processed for assembly using one of the two strategies, either reference-based assembly (co-assembly) or de novo assembly. Software packages such as Newbler (Roche), AMOS (http://sourceforge.net/projects/amos/), or MIRA[48] can be employed to perform reference-based assembly. For de novo assembly, tools based on the de Bruijn graphs are created to handle very large amounts of data.[49,50] In addition, two new assembly programs (Meta Velvet and Meta-IDBA)[51] have been developed to deal with the non-clonality of natural populations. The sequenced information is then processed for binning to sort DNA sequences into taxonomic groups that might represent individual or closely related genomes. Several algorithms employing different methods of grouping sequences have been developed, including but not limited to Phylopythia, S-GSOM, PCAHIER, TACAO, IMG/M, MG-RAST, MOTHUR, MEGAN, TANGO, CARMA, SOrt-ITEMS, MetaPhyler, PhymmBL and MetaCluster.[52–63] These algorithms have been developed depending on the type of input data generated from metagenomic sequencing.

Generally, metagenomic sequences are annotated in two steps: (a) Feature prediction is performed by identifying characteristics of interest within genes; and (b) functional annotation is performed by assigning putative gene functions and taxonomic neighbors. Several tools such as MG-RAST, IMG/M, FragGeneScan, MetaGeneMark, Metagene, and Orphelia have been developed for classifying sequence stretches as either coding or non-coding.[52–58,64–70] BLAST-based searches are also used for potentially identifying any missing information from these programs. Some of the other tools that are employed for predicting non-protein coding genes are tRNAs, Signal peptides. and CRISPRs.[71–73] Other primary online sources for obtaining annotated nucleotides sequence information include the International Nucleotide Sequence Database Collaboration (INSDC), the DNA Data Bank of Japan, the European Nucleotide Archive, GenBank, and the Sequence Read Archive (SRA). By mid-September 2010, the SRA had accumulated more than 500 billion reads consisting of 60 trillion base pairs available for download.[74] SRA contained 80% of the sequencing data from the Illumina GA platform, as well as 15% and 5% from the SOLiD TM and Roche/454 platforms, resepectively.[74] Functional annotation of the metagenomic data is a major challenge, as only a small percentage of metagenomic

sequences are annotated.[54,65,75] The sequences that cannot be annotated, either because they might simply reflect erroneous coding sequences, because they might be real genes but encode for unknown biochemical functions, or because they may not have homology to known genes, are all grouped as ORFans.[76] Additional reference databases such as KEGG, eggNOG, COG/KOG, PFAM, and TIGRFAM are all available online tools that can be used to study functional properties of ORFans.[76]

Statistical Analysis and Data Sources

A typical metagenomic project contains an enormous amount of data that needs careful evaluation using proper statistical methods. Primer-E-Package is a popular tool that can perform a range of multivariate statistical analysis.[77] This package includes generation of multidimensional scaling plots, analysis of similarities (ANOSIM), identification of the species, and identification of gene functions (SIMPER). There is also a web-based tool called Metastats that has been used in recent studies.[78] The ShotgunFunctionalizeR package also provides several statistical programs to evaluate functional differences between samples.[29] Due to the increasing number of metagenomic studies, it is important to deposit large sets of metagenomic data into databases. Deposition of metagenomic data in centralized services would not only facilitate comparative analysis of different metagenomic data but also facilitate a new level of organization and collaboration among researchers. Services like IMG/M, CAMERA, and MG-RAST are three prominent metagenomic databases that are available for large-scale metagenomic analysis.[54,57,75]

Metagenomics to Study Microbial Diversity in Environment

In the last decade, several studies have used the metagenomic approach and provided comprehensive data on microbial communities in different ecosystems. It is estimated that, depending on the sample and methods used, the number of bacteria in soil may vary from 467 species to 500,000 species.[19,79–81] Curtis and colleagues have speculated that bacterial content may range up to 4×10^6/ton of soil and the numbers of bacteria are unlikely to exceed 2×10^6 in the sea.[19] These comparisons clearly suggest that microbial content is several orders of magnitude less in the sea in comparison to soil environments. The members of the archaeal phylum *Crenarchaeota* are shown to be predominant microorganisms found in the depth of ocean with estimated numbers 1.3×10^{28} in global oceans,[82] and total bacterial cells are estimated to be 3.1×10^{28}.[82] It has also been noticed that in certain regions and at certain times, 50% of the bacterial community in surface waters consists of members belonging to SAR11 clade.[83] Metagenomic ocean surveys have also led to several surprising discoveries. For example, using anchored chromosome walking, a 130 Kb BAC clone was isolated from uncultivated SAR86 bacterium (a bacterium belonging to alpha proteobacteria that is abundantly found in ocean surface waters).[12] Sequencing of the 130 Kb fragment resulted in the identification of a new class of genes of the rhodopsin family for the first time in bacteria.[12] Further studies on this class of genes in *Escherichia coli* proved its function as a light-driven proton pump.[12] In summary, this study discovered a new type of light-driven energy generation in oceanic bacteria which subsequently led to the identification of several photoproteins in the Sargasso Sea.[84]

Soil is one of the most challenging environmental sources to analyze microbial diversity. Several parameters of soil, such as particle size, permeability, porosity, water content, mineral composition, and plant cover, can influence microbial composition.[35,85,86] In addition, other factors such as collection and storage of soil sample, DNA extraction methods, host-vector systems used for DNA cloning, and representative soil sampling, can also influence the results of microbial content.[35,85,86] With the advent of various technical developments, several landmark studies have been performed using the metagenomics approach.[10,12,20,29,83,87–89] By direct cloning into plasmid, cosmid, or BAC vectors, novel genes from soil microbes that encode enzymes and antibiotics have been discovered.[90] These genes share little homology with known genes, thus illustrating the enormous potential of soil metagenomics in isolating novel classes of genes. Some of the genes that were isolated from soil microorganisms include lipases, proteases, oxidoreductases, amylases, antibiotics, antibiotic resistance enzymes, and membrane proteins.[21,87,91–93]

Using a metagenomics approach several studies have provided a wealth of information on microbial diversity in extreme environmental conditions.

Studies from Barns and colleagues have provided information on microbial diversity in hot spring environments.[94,95] Archaea similar to *Crenarchaeota* phylotype are found to be the abundant species in Yellowstone National Park hot springs.[94,95] Analysis in the same hot spring revealed more bacterial numbers distributed in twelve new division-level lineages.[94,95] Furthermore, Blank and colleagues showed differences in microbial content in the samples collected from different Yellowstone National park hot springs at close proximity with similar temperatures and comparable pH values.[13] Sequencing of polar ice caps has revealed the presence of algal population and several heterotrophic bacteria in ice matric at low temperatures and low levels of light.[14,96] Similar findings were noted in the analysis of microbial composition in the cryoconite hole of a glacier.[16] Recent studies have also found dominance of archaea *Salinibacter ruber* in hyper-saline environments.[97] Non-thermal environments with extreme acidic conditions have also been shown to contain archaea of *Ferroplasma* and *Thermoplasma* groups. In addition, several bacterial species that include *Acidiphilium*, *Acidithiobacillus*, *Leptospirillum* and *Sulfobacillus* have also been found to be abundant in extreme acidic environments.[11]

Although these studies indicate the important role of microorganisms in biogeochemical cycles, many details remain unclear; until we fully understand the nature of microbial diversity in different environments, this will remain as an important area of investigation.

Viral Metagenomics

The development of metagenomic approaches has revolutionized evaluation of viral particles in environmental samples. The results from more than 24 independent studies have already been published.[98] These studies highlight that 50% of viral sequences are "unknown". Of the remaining 50% "known" sequences, many had low amino acid similarities to known viral proteins and thus represent an uncategorized group.[20] These findings suggest a more complex diversity of viral genomes in comparison to bacteria in environmental samples. This is consistent with the findings that 30% of the open reading frames in sequenced viral genomes are ORFans, compared to 9% ORFans from bacteria.[99] Despite these challenges, viral metagenomics have developed methods to catalogue viruses in environmental samples based on identifiable sequences. Full genome sequences of novel viruses that were identified from different environments have already been reported and assembled.[100] Based on genomic structure and taxonomic metagenomic analysis, some of the studies have linked viruses with their potential hosts.[101,102]

Over the past decade, several studies using metagenomics have provided a substantial amount of information in the identification of new viruses from human samples.[103–106] Most of the infectious diseases caused by viruses were documented before the identification of their causative agent. For example, Egyptian literature from approximately 3700 BC provided information on poliomyelitis. However, the causative agent for this disease was identified as poliomyelitis virus in 1909 AD.[107] Similar descriptions of clinical conditions likely caused by Smallpox were found in ancient literature from India 1500 BC long before the isolation of the Variola virus.[108,109] With the steady rise in the development of viral metagenomics, several novel viruses have been isolated within a short amount of time that are associated with disease outcomes in humans.[98,104–106] Novel viruses including Borna virus, Arena virus, Paralysis virus, LUJO virus, Astrovirus as etiology of mink shaking syndrome, Simian hemorrhagic fever virus, and Klassevirus have been identified by metagenomic approaches as causes of diseases in humans and other mammals.[110–116] In addition, a recent study has provided important information in the identification of several viruses in a public-health setting.[104] These studies highlight future perspectives on the use of metagenomic approaches for generating enormous amounts of data in the identification of unknown and potentially infectious agents to humans, all in a short amount of time. Recent metagenomic analysis also addressed changes in the viral communities in Cystic Fibrosis and compared them to those of non-cystic fibrosis individuals.[105] In addition, interest in tapping the vast novelty of viral genetic information, especially phages, has brought great attention to the use of metagenomics in this field.[101] Overall, metagenomics has provided substantial insights to virus–host interactions and viral diversity in different environments.

Tick Metagenomics

Ticks are medically important arthropod vectors that transmit pathogens causing various human diseases.[117]

The advancement of metagenomic approaches has facilitated research in studying microbial communities associated with medically important arthropod vectors.[15,118] Using 454/Roche and Illumina-based metagenomic sequencing, Carpi et al have evaluated pathogen load and microbiome in *Ixodes ricinus* ticks. MEGAN comparison of the bacterial taxonomic profiles determined that a total of 108 genera belonging to all bacterial phyla were present in *I. ricinus* ticks.[15] Their study determined that, in addition to mutualistic bacteria such as *Wolbachia* and *Rickettsiella*, pathogenic bacteria such as *Borrelia*, *Rickettsia* and *Candidatus Neoehrlichia* were also present in ticks. The bacterial content varied in ticks collected from different geographic regions and at different life stages, which might be due to the changes in environmental factors and host-selection behaviors of ticks.[15] Metagenomic-based studies such as this one would not only facilitate epidemiological surveillance of several zoonotic pathogens, but would also lead to the development of better strategies to control vector-borne human diseases.

Industrial Metagenomics

With the advent of the metagenomic approach to discover novel genes that encode various enzymes, antibiotics, photoproteins, and membrane proteins from environmental uncultured bacteria, several industries have shown interest in exploiting these resources for the development of commercially available compounds. Metagenomics has provided access to novel enzymes and biocatalysts that were not initially achievable by conventional cultivable bacteria.[21,87,119–121] In fact, global sales for enzymes were estimated to be $ 2.3 million in 2003, a figure that includes sales of enzymes in detergents, food applications, agriculture/feed, textile processing, pulp/paper, leather, and production of fine and bulk chemicals.[122] In light of increasing energy costs, environmental pollution, public health hazards, and recent global economic crises, the discovery of novel enzymes from metagenomic approaches can be viewed both as an opportunity and as a necessity. For example, Diversa, the largest biotech company focusing on the commercialization of metagenome technologies, has constructed and screened for various nitrilase gene sequences isolated from diverse environmental libraries.[123] This nitrilase enzyme library was marketed to several fine-chemical and pharmaceutical industries.[123] In summary, metagenomics has played a significant role in the identification of several bioactive molecules that have attracted interest from both academia and industrial companies.[1,87,119–121]

Metagenomic Application to Study Human Gut and Skin Microbiome

Over the past decade, metagenomics have provided great insights to the human microbiome. Waddington used a metaphor and regarded microbiota as an essential "organ" of the human body capable of performing metabolic functions that human cells might not be able to perform.[124,125] Several factors such as specific microbial species colonizing the gut, niches they occupy, time, space, factors unique to the environment of each human being such as different dietary needs, and interactions with host cells can all influence taxonomic composition of the human microbiome. Metagenomics have uncovered nearly 1000 human-associated microorganisms' draft genome sequences, along with 3.3 million unique microbial genes derived from the intestinal tract of over 100 European adults.[17,126,127] Analysis of intestinal microbial content of humans across various continents revealed that microbes were clustered in 3 groups that are termed as enterotypes.[17,22,126,127] Metagenomics of the human gut microbiome also revealed interesting functions carried by microorganisms within the gut, ranging from its role in newly discovered signaling mechanisms, vitamin production synthesis, glycan production, amino-acid, and xenobiotic metabolism. Several studies have also reported that microbial composition of the human gut is greatly affected by genetic background, age, diet, and health status of the host.[17,22,126–128] Differences in microbial content were seen in all age groups of human beings. Babies (breast fed and formula fed), healthy and malnourished infants, youngsters, the elderly, humans that were either lean or obese, and humans with inflammatory bowel diseases (IBD) showed differences in microbial composition.[23,88,129–131] A metagenomic study from De Filippo et al showed that European children who consumed a diet high in animal protein, sugar, starch, and fat, and low in fiber showed differences in gut microbial content in comparison to children fed on vegetarian diet consisting of carbohydrates, fiber, and non-animal protein.[127] Interestingly,

the microbiome of European children was enriched with *Firmicutes* and *Proteobacteria*, whereas the African microbiome was enriched with *Actinobacteria* and *Bacteroidetes*. Members of *Xylanibacter* and *Prevotella* were only present in Children from Europe. These results clearly suggest that host dietary habits influence gut microbial content. Metagenomics have also revealed an interesting link with microbial content in the gut to host metabolism and disease development.[132] The causes of intestinal diseases such as IBD, Crohn's disease (CD), and ulcerative colitis (UC) have all been linked with both human gene- and microbiome-associated factors.[132] Several metagenomic and microarray studies have also revealed differences in microbial composition in CD patients in comparison to healthy individuals.[89,131,133,134]

Recent studies have used metagenomic approaches in looking at the microbial diversity of the human skin.[135–138] Skin serves as a good host of microbes that include both commensal and pathogenic bacteria. Determination of microbial diversity in skin revealed several interesting findings.[135–138] Bacteria belonging to *Proteobacteria*, such as *Pseudomonas* species and *Janthinobacterium* species, were found to be abundant in both human and mice skin biopsies.[136–138] The presence of other bacteria belonging to *Alphaproteobacteria*, *Gammaproteobacteria*, *Betaproteobacteria*, *Actinobacteria* species, such as *Kocuria* species, *Pripionibacteria* species, *Firmicutes*, and *Bacteroidetes* were all evident in human skin biopsies.[136–138] There is substantial evidence that viruses also represent a significant part of the skin microbiome.[135] The presence of beta and gamma-human papillomaviruses, polyomaviruses, and circoviruses on normal-appearing skin has been reported.[135] With the increasing interest in understanding the role of the microbiome on human development, aging and disease, the field of metagenomics is rapidly advancing with new techniques and has become a most effective tool in this area of investigation.

Future Directions

Over the past decade metagenomics has undoubtedly benefited the scientific world in rapidly analyzing changes in microbial communities in different environments. Despite exhaustive research efforts, both in financial and intellectual terms, the underlying mechanisms of the relationship between the microbial communities to the environment or to human gut metabolism, aging, and disease remains unclear. Therefore, improvements in metagenomic techniques that involve functional microbiomic approaches need to be developed. In addition, development of novel metagenomic approaches that consider several geochemical parameters is highly warranted to evaluate the complexity of microbial population in extreme environments. Metagenomics has provided identification of several new microbial genes from different environmental samples.[11,12,14,15,17,19,91,93,135] Heterologous gene expression is an important and challenging approach that is required to identify the function of new genes identified by metagenomic studies.[139–142] Studies have successfully used a heterologous gene expression system to identify several antibiotic resistance genes.[140,142–145] *E. coli* or other domesticated bacteria are commonly used to express genes identified from metagenomic approaches.[1,140,142,144,145] However, the important limitation in this approach is that many genes, and indeed perhaps most genes, are not expressed in these bacteria.[1,140,142,144,145] Therefore, improvement in heterologous gene expression systems and production of functional recombinant proteins would speed up the discovery of important biomolecules from different environments. Due to the sheer volume of metagenomic data that is continuously being generated, development of novel methods for analysis, data storage, and sharing is warranted. This would not only facilitate best use of metagenomic data for researchers, but would also lead researchers to the answers to some of the fundamental questions about microbes—namely "who they are" and "what they are doing"—in this complex world of microorganisms.

Author Contributions

Conceived and designed the experiments: GN, HS. Analyzed the data: GN, HS. Wrote the first draft of the manuscript: GN, HS. Contributed to the writing of the manuscript: GN, HS. Agree with manuscript results and conclusions: GN, HS. Jointly developed the structure and arguments for the paper: GN, HS. Made critical revisions and approved final version: HS. All authors reviewed and approved of the final manuscript.

Funding
This work was supported by independent start-up funds from Old Dominion University to GN and HS.

Competing Interests
Author(s) disclose no potential conflicts of interest.

Disclosures and Ethics
As a requirement of publication the authors have provided signed confirmation of their compliance with ethical and legal obligations including but not limited to compliance with ICMJE authorship and competing interests guidelines, that the article is neither under consideration for publication nor published elsewhere, of their compliance with legal and ethical guidelines concerning human and animal research participants (if applicable), and that permission has been obtained for reproduction of any copyrighted material. This article was subject to blind, independent, expert peer review. The reviewers reported no competing interests. Provenance: the authors were invited to submit this paper.

References
1. Handelsman J. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev*. 2004;68:669–85.
2. Staley JT, Konopka A. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol*. 1985;39:321–46.
3. Whang K, Hattori T. Oligotrophic bacteria from rendzina forest soil. *Antonie Van Leeuwenhoek*. 1988;54:19–36.
4. Riesenfeld CS, Schloss PD, Handelsman J. Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet*. 2005;38:525–52.
5. Schleper C, Jurgens G, Jonuscheit M. Genomic studies of uncultivated archaea. *Nat Rev Microbiol*. 2005;3:479–88.
6. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol*. 2008;26:1135–45.
7. Pace NR, Stahl DA, Lane DJ, Olsen GJ. Analyzing natural microbial populations by rRNA sequences. *ASM News*. 1985;51:4–12.
8. Schmidt TM, DeLong EF, Pace NR. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J Bacteriol*. 1991;173:4371–8.
9. Tyson GW, Chapman J, Hugenholtz P, et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*. 2004;428:37–43.
10. Venter JC, Remington K, Heidelberg JF, et al. Environmental genome shotgun sequencing of the Sargasso Sea. *Science*. 2004;304:66–74.
11. Baker BJ, Banfield JF. Microbial communities in acid mine drainage. *FEMS Microbiol Ecol*. 2003;44:139–52.
12. Beja O, Aravind L, Koonin EV, et al. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science*. 2000;289:1902–6.
13. Blank CE, Cady SL, Pace NR. Microbial composition of near-boiling silica-depositing thermal springs throughout Yellowstone National Park. *Appl Environ Microbiol*. 2002;68:5123–35.
14. Bowman JP, McCammon SA, Brown MV, Nichols DS, McMeekin TA. Diversity and association of psychrophilic bacteria in Antarctic sea ice. *Appl Environ Microbiol*. 1997;63:3068–78.
15. Carpi G, Cagnacci F, Wittekindt NE, et al. Metagenomic profile of the bacterial communities associated with Ixodes ricinus ticks. *PLoS One*. 2011;6:e25604.
16. Christner BC, Kvitko BH 2nd, Reeve JN. Molecular identification of bacteria and Eukarya inhabiting an Antarctic cryoconite hole. *Extremophiles*. 2003;7:177–83.
17. Claesson MJ, Cusack S, O'Sullivan O, et al. Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc Natl Acad Sci U S A*. 2011;108 Suppl 1:4586–91.
18. Cox-Foster DL, Conlan S, Holmes EC, et al. A metagenomic survey of microbes in honey bee colony collapse disorder. *Science*. 2007;318:283–7.
19. Curtis TP, Sloan WT, Scannell JW. Estimating prokaryotic diversity and its limits. *Proc Natl Acad Sci U S A*. 2002;99:10494–9.
20. Edwards RA, Rohwer F. Viral metagenomics. *Nat Rev Microbiol*. 2005;3:504–10.
21. Knietsch A, Waschkowitz T, Bowien S, Henne A, Daniel R. Metagenomes of complex microbial consortia derived from different soils as sources for novel genes conferring formation of carbonyls from short-chain polyols on Escherichia coli. *J Mol Microbiol Biotechnol*. 2003;5:46–56.
22. Maccaferri S, Biagi E, Brigidi P. Metagenomics: key to human gut microbiota. *Dig Dis*. 2001;29:525–30.
23. Monira S, Nakamura S, Gotoh K, et al. Gut microbiota of healthy and malnourished children in bangladesh. *Front Microbiol*. 2011;2:228.
24. Harismendy O, Ng PC, Strausberg RL, et al. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol*. 2009;10:R32.
25. Turnbaugh PJ, Ley RE, Mahowald MA, et al. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*. 2006;444:1027–31.
26. Palenik B, Ren Q, Tai V, Paulsen IT. Coastal Synechococcus metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environ Microbiol*. 2009;11:349–59.
27. Biddle JF, Fitz-Gibbon S, Schuster SC, Brenchley JE, House CH. Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. *Proc Natl Acad Sci U S A*. 2008;105:10583–8.
28. Edwards RA, Rodriguez-Brito B, Wegley L, Haynes M, Breitbart M, et al. Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics*. 2006;7:57.
29. Turnbaugh PJ, Gordon JI. The core gut microbiome, energy balance and obesity. *J Physiol*. 2009;587:4153–8.
30. Chen K, Pachter L. Bioinformatics for whole-genome shotgun sequencing of microbial communities. *PLoS Comput Biol*. 2005;1:106–12.
31. Desai N, Antonopoulos D, Gilbert JA, Glass EM, Meyer F. From genomics to metagenomics. *Curr Opin Biotechnol*. 2012;23:72–6.
32. Burke C, Kjelleberg S, Thomas T. Selective extraction of bacterial DNA from the surfaces of macroalgae. *Appl Environ Microbiol*. 2009;75:252–6.
33. Delmont TO, Robe P, Clark I, Simonet P, Vogel TM. Metagenomic comparison of direct and indirect soil DNA extraction approaches. *J Microbiol Methods*. 2011;86:397–400.
34. Thomas T, Rusch D, DeMaere MZ, et al. Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *ISME J*. 2010;4:1557–67. Sessitsch A, Weilharter A, Gerzabek MH, Kirchmann H, Kandeler E. Microbial population structures in soil particle size fractions of a long-term fertilizer field experiment. *Appl Environ Microbiol*. 2001;67:4215–24.
35. Ishoey T, Woyke T, Stepanauskas R, Novotny M, Lasken RS. Genomic sequencing of single microbial cells from environmental samples. *Curr Opin Microbiol*. 2008;11:198–204.
36. Lasken RS. Genomic DNA amplification by the multiple displacement amplification (MDA) method. *Biochem Soc Trans*. 2009;37:450–3.
37. Eisen JA. Environmental shotgun sequencing: its potential and challenges for studying the hidden world of microbes. *PLoS Biol*. 2007;5:e82.
38. Metzker ML. Sequencing technologies—the next generation. *Nat Rev Genet*. 2010;11:31–46.
39. Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends Genet*. 2008;24:133–41.

40. Goltsman DS, Denef VJ, Singer SW, et al. Community genomic and proteomic analyses of chemoautotrophic iron—oxidizing "Leptospirillum rubarum" (Group II) and "Leptospirillum ferrodiazotrophum" (Group III) bacteria in acid mine drainage biofilms. *Appl Environ Microbiol*. 2009;75:4599–615.
41. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008;456:53–9.
42. Al Masalma M, Lonjon M, Richet H, et al. Metagenomic analysis of brain abscesses identifies specific bacterial associations. *Clin Infect Dis*. 2012;54: 202–10.
43. Lazarevic V, Whiteson K, Gaia N, et al. Analysis of the salivary microbiome using culture-independent techniques. *J Clin Bioinforma*. 2012;2:4.
44. Shah N, Tang H, Doak T, et al. Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. *Pac Symp Biocomput*. 2011:165–76.
45. Jeraldo P, Chia N, Goldenfeld N. On the suitability of short reads of 16S rRNA for phylogeny-based analyses in environmental surveys. *Environ Microbiol*. 2011;13:3000–9.
46. Kim M, Morrison M, Yu Z. Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *J Microbiol Methods*. 2011;84:81–7.
47. Chevreux B, Wetter T, Suhai S. Genome sequence assembly using trace signals and additional sequence information computer science and biology. *Proc German Conf Bioinform*. 1999;99:45–56.
48. Miller JR, Koren S, Sutton G. Assembly algorithms for next-generation sequencing data. *Genomics*. 2010;95(6):315–27.
49. Pevzner PA, Tang H, Waterman MS. An Eulerian path approach to DNA fragment assembly. *Proc Natl Acad Sci U S A*. 2001;98(17):9748–53.
50. Peng Y, Leung HC, Yiu SM, Chin FY. Meta-IDBA: a de Novo assembler for metagenomic data. *Bioinformatics*. 2011;27(13):i94–101.
51. Chan CK, Hsu AL, Halgamuge SK, Tang SL. Binning sequences using very sparse labels within a metagenome. *BMC Bioinformatics*. 2008;9:215.
52. Diaz NN, Krause L, Goesmann A, Niehaus K, Nattkemper TW. TACOA: taxonomic classification of environmental genomic fragments using a kernelized nearest neighbor approach. *BMC Bioinformatics*. 2009;10:56.
53. Glass EM, Wilkening J, Wilke A, Antonopoulos D, Meyer F. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb Protoc*. 2010;2010(1):pdb.prot5368.
54. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. *Genome Res*. 2007;17(3):377–86.
55. Krause L, Diaz NN, Goesmann A, et al. Phylogenetic classification of short environmental DNA fragments. *Nucleic Acids Res*. 2008;36(7):2230–9.
56. Markowitz VM, Chen IM, Chu K, et al. IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Res*. 2012;40(D1):D123–9.
57. McHardy AC, Martin HG, Tsirigos A, Hugenholtz P, Rigoutsos I. Accurate phylogenetic classification of variable-length DNA fragments. *Nat Methods*. 2007;4(1):63–72.
58. Zheng H, Wu H. Short prokaryotic DNA fragment binning using a hierarchical classifier based on linear discriminant analysis and principal component analysis. *J Bioinform Comput Biol*. 2010;8(6):995–1011.
59. Monzoorul Haque M, Ghosh TS, Komanduri D, Mande SS. SOrt-ITEMS: Sequence orthology based approach for improved taxonomic estimation of metagenomic sequences. *Bioinformatics*. 2009;25(14):1722–30.
60. Liu B, Gibbons T, Ghodsi M, Treangen T, Pop M. Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences. *BMC Genomics*. 2011;12 Suppl 2:S4.
61. Brady A, Salzberg SL. Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. *Nat Methods*. 2009;6(9): 673–6.
62. Leung HC, Yiu SM, Yang B, et al. A robust and accurate binning algorithm for metagenomic sequences with arbitrary species abundance ratio. *Bioinformatics*. 2011;27(11):1489–95.
63. Aziz RK, Bartels D, Best AA, et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*. 2008;9:75.
64. Markowitz VM, Mavromatis K, Ivanova NN, et al. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics*. 2009;25(17):2271–8.
65. Lukashin AV, Borodovsky M. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res*. 1998;26(4):1107–15.
66. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res*. 1999; 27(23):4636–41.
67. Noguchi H, Taniguchi T, Itoh T. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res*. 2008;15(6):387–96.
68. Hoff KJ, Lingner T, Meinicke P, Tech M. Orphelia: predicting genes in metagenomic sequencing reads. *Nucleic Acids Res*. 2009;37(Web sServer issue): W101–5.
69. Yok NG, Rosen GL. Combining gene prediction methods to improve metagenomic gene annotation. *BMC Bioinformatics*. 2011;12:20.
70. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1997;25(5):955–64.
71. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol*. 2004;340(4):783–95.
72. Bland C, Ramsey TL, Sabree F, et al. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics*. 2007;8:209.
73. Leinonen R, Sugawara H, Shumway M. The sequence read archive. *Nucleic Acids Res*. 2011;39(Database issue):D19–21.
74. Sun S, Chen J, Li W, et al. Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource. *Nucleic Acids Res*. 2011;39(Database issue):D546–51.
75. Yooseph S, Sutton G, Rusch DB, et al. The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol*. 2007;5(3):e16.
76. Clarke KR. Non-parametric multivariate analyses of changes in community structure. *Australian J Ecolology*. 1993;18:117–43.
77. White JR, Nagarajan N, Pop M. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput Biol*. 2009;5(4):e1000352.
78. Hughes JB, Hellmann JJ, Ricketts TH, Bohannan BJ. Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl Environ Microbiol*. 2001;67(10):4399–406.
79. Torsvik V, Daae FL, Sandaa RA, Ovreas L. Novel techniques for analysing microbial diversity in natural and perturbed environments. *J Biotechnol*. 1998;64(1):53–62.
80. Dykhuizen DE. Santa Rosalia revisited: why are there so many species of bacteria? *Antonie Van Leeuwenhoek*. 1998;73(1):25–33.
81. Karner MB, DeLong EF, Karl DM. Archaeal dominance in the mesopelagic zzone of the Pacific Ocean. *Nature*. 2001;409(6819):507–10.
82. Morris RM, Rappe MS, Connon SA, et al. SAR11 clade dominates ocean surface bacterioplankton communities. *Nature*. 2002;420(6917):806–10.
83. de la Torre JR, Christianson LM, Beja O, et al. Proteorhodopsin genes are distributed among divergent marine bacterial taxa. *Proc Natl Acad Sci U S A*. 2003;100(22):12830–5.
84. Girvan MS, Bullimore J, Pretty JN, Osborn AM, Ball AS. Soil type is the primary determinant of the composition of the total and active bacterial communities in arable soils. *Appl Environ Microbiol*. 2003;69(3):1800–9.
85. Kowalchuk GA, Buma DS, de Boer W, Klinkhamer PG, van Veen JA. Effects of above-ground plant species composition and diversity on the diversity of soil-borne microorganisms. *Antonie Van Leeuwenhoek*. 2002;81(1–4):509–20.
86. Henne A, Schmitz RA, Bomeke M, Gottschalk G, Daniel R. Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on Escherichia coli. *Appl Environ Microbiol*. 2000;66(7):3113–6.
87. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: human gut microbes associated with obesity. *Nature*. 2006;444(7122):1022–3.
88. Qin J, Li R, Raes J, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010;464(7285):59–65.

89. Henne A, Daniel R, Schmitz RA, Gottschalk G. Construction of environmental DNA libraries in Escherichia coli and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Appl Environ Microbiol*. 1999;65(9):3901–7.
90. Lee SW, Won K, Lim HK, et al. Screening for novel lipolytic enzymes from uncultured soil microorganisms. *Appl Microbiol Biotechnol*. 2004;65(6):720–6.
91. Santosa DA. Rapid extraction and purification of environmental DNA for molecular cloning applications and molecular diversity studies. *Mol Biotechnol*. 2001;17(1):59–64.
92. Yun J, Kang S, Park S, et al. Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. *Appl Environ Microbiol*. 2004;70(12):7229–35.
93. Barns SM, Fundyga RE, Jeffries MW, Pace NR. Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc Natl Acad Sci U S A*. 1994;91(5):1609–13.
94. Barns SM, Delwiche CF, Palmer JD, Pace NR. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc Natl Acad Sci U S A*. 1996;93(17):9188–93.
95. Brown MV, Bowman JP. A molecular phylogenetic survey of sea-ice microbial communities (SIMCO). *FEMS Microbiol Ecol*. 2001;35(3):267–75.
96. Benlloch S, Lopez-Lopez A, Casamayor EO, et al. Prokaryotic genetic diversity throughout the salinity gradient of a coastal solar saltern. *Environ Microbiol*. 2022;4(6):349–60.
97. Rosario K, Breitbart M. Exploring the viral world through metagenomics. *Curr Opin Virol*. 2011;1(4):289–97.
98. Yin Y, Fischer D. Identification and investigation of ORFans in the viral world. *BMC Genomics*. 2008;9:24.
99. Meyer F, Paarmann D, D'Souza M, et al. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*. 2008;9:386.
100. Skennerton CT, Angly FE, Breitbart M, et al. Phage encoded H-NS: a potential achilles heel in the bacterial defence system. *PLoS One*. 2011;6(5):e20095.
101. Culley AI, Lang AS, Suttle CA. The complete genomes of three viruses assembled from shotgun libraries of marine RNA virus communities. *Virol J*. 2007;4:69.
102. Breitbart M, Hewson I, Felts B, et al. Metagenomic analyses of an uncultured viral community from human feces. *J Bacteriol*. 2003;185(20):6220–3.
103. Svraka S, Rosario K, Duizer E, et al. Metagenomic sequencing for virus identification in a public-health setting. *J Gen Virol*. 2010;91(Pt 11): 2846–56.
104. Willner D, Furlan M, Haynes M, et al. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS One*. 2009;4(10):e7370.
105. Finkbeiner SR, Allred AF, Tarr PI, et al. Metagenomic analysis of human diarrhea: viral detection and discovery. *PLoS Pathog*. 2008;4(2):e1000011.
106. Paul JR. *A History of Poliomyelitis*. Yale Studies in the History of Science and Medicine Yale University Press; 1971.
107. Fenner F, Henderson DA, Arita I, Jezek Z, Ladnyl ID. In: Fenner F, editor. *Smallpox and its Eradication*. WHO; 1988.
108. Hopkins DR. *The Greatest Killer-Smallpox in History*. Chicago: Chicago University of Chicago Press; 1983.
109. Blomstrom AL, Widen F, Hammer AS, Belak S, Berg M. Detection of a novel astrovirus in brain tissue of mink suffering from shaking mink syndrome by use of viral metagenomics. *J Clin Microbiol*. 2010;48(12):4392–6.
110. Briese T, Paweska JT, McMullan LK, et al. Genetic detection and characterization of Lujo virus, a new hemorrhagic fever—associated arenavirus from southern Africa. *PLoS Pathog*. 2009;5(5):e1000455.
111. Djikeng A, Halpin R, Kuzmickas R, et al. Viral genome sequencing by random priming methods. *BMC Genomics*. 2008;9:5.
112. Greninger AL, Runckel C, Chiu CY, et al. The complete genome of klassevirus—a novel picornavirus in pediatric stool. *Virol J*. 2009;6:82.
113. Honkavuori KS, Shivaprasad HL, Williams BL, et al. Novel borna virus in psittacine birds with proventricular dilatation disease. *Emerg Infect Dis*. 2008;14(12):1883–6.
114. Lauck M, Hyeroba D, Tumukunde A, et al. Novel, divergent simian hemorrhagic fever viruses in a wild Ugandan red colobus monkey discovered using direct pyrosequencing. *PLoS One*. 2011;6:e19056.
115. Palacios G, Druce J, Du L, et al. A new arenavirus in a cluster of fatal transplant-associated diseases. *N Engl J Med*. 2008;358(10):991–8.
116. de la Fuente J, Estrada-Pena A, Venzal JM. Overview: Ticks as vectors of pathogens that cause disease in humans and animals. *Front Biosci*. 2008;13:6938–46.
117. van Overbeek L, Gassner F, van der Plas CL, et al. Diversity of Ixodes ricinus tick-associated bacterial communities from different forests. *FEMS Microbiol Ecol*. 2008;66:72–84.
118. Entcheva P, Liebl W, Johann A, Hartsch T, Streit WR. Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Appl Environ Microbiol*. 2001;67:89–99.
119. Gabor EM, de Vries EJ, Janssen DB. Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. *Environ Microbiol*. 2004;6:948–58.
120. Gupta R, Beg QK, Lorenz P. Bacterial alkaline proteases: molecular approaches and industrial applications. *Appl Microbiol Biotechnol*. 2002;59:15–32.
121. Analysts. GI. *Industrial Enzymes—A Global Multi-Client Market Research Project.* GIA, San José, California, USA; 2004.
122. DeSantis GZZ, Greenberg WA, Wong K, et al. An enzyme library approach to biocatalysis: development of nitrilases for enantioselective production of carboxylic acid derivatives. *J Am Chem Soc*. 2002;124:9024–5.
123. Waddington CH. *The Strategy of the Genes. A Discussion of Some Aspects of Theoretical Biology.* London: George Allen and Unwin; 1957.
124. Hall BK, editor. *Evolutionary Developmental Biology*. London: Chapman & Hall; 1992.
125. Biagi E, Nylund L, Candela M, et al. Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. *PLoS One*. 2010;5:e10667.
126. De Filippo C, Cavalieri D, Di Paola M, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci U S A*. 2010;107:14691–6.
127. Dutton RJ, Turnbaugh PJ. Taking a metagenomic view of human nutrition. *Curr Opin Clin Nutr Metab Care*. 2012;15:448–54.
128. Schwartz S, Friedberg I, Ivanov IV, et al. A metagenomic study of diet-dependent interaction between gut microbiota and host in infants reveals differences in immune response. *Genome Biol*. 2012;13:r32.
129. Zhang H, DiBaise JK, Zuccolo A, et al. Human gut microbiota in obesity and after gastric bypass. *Proc Natl Acad Sci U S A*. 2009;106:2365–70.
130. Willing BP, Dicksved J, Halfvarson J, et al. A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology*. 2010;139:1844–54.
131. Pflughoeft KJ, Versalovic J. Human microbiome in health and disease. *Annu Rev Pathol*. 2012;7:99–122.
132. Manichanh C, Rigottier-Gois L, Bonnaud E, et al. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut*. 2006;55:205–11.
133. Kang S, Denman SE, Morrison M, et al. Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. *Inflamm Bowel Dis*. 2010;16:2034–42.
134. Foulongne V, Sauvage V, Hebert C, et al. Human skin microbiota: high diversity of DNA viruses identified on the human skin by high throughput sequencing. *PLoS One*. 2012;7:e38499.
135. Grice EA, Kong HH, Conlan S, et al. Topographical and temporal diversity of the human skin microbiome. *Science*. 2009;324:1190–2.
136. Grice EA, Kong HH, Renaud G, et al. A diversity profile of the human skin microbiota. *Genome Res*. 2008;18:1043–50.
137. Kong HH, Oh J, Deming C, et al. Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. *Genome Res*. 2012;22:850–9.
138. Courtois S, Cappellano CM, Ball M, et al. Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl Environ Microbiol*. 2003;69:49–55.

139. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol*. 1998;5:R245–9.
140. MacNeil IA, Tiong CL, Minor C, et al. Expression and isolation of antimicrobial small molecules from soil DNA libraries. *J Mol Microbiol Biotechnol*. 2001;3:301–8.
141. Rondon MR, August PR, Bettermann AD, et al. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol*. 2000;66:2541–7.
142. Gillespie DE, Brady SF, Bettermann AD, et al. Isolation of antibiotics turbomycin a and B from a metagenomic library of soil microbial DNA. *Appl Environ Microbiol*. 2002;68:4301–6.
143. Rondon MR, Goodman RM, Handelsman J. The Earth's bounty: assessing and accessing soil microbial diversity. *Trends Biotechnol*. 1999;17:403–9.
144. Rondon MR, Raffel SJ, Goodman RM, Handelsman J. Toward functional genomics in bacteria: analysis of gene expression in Escherichia coli from a bacterial artificial chromosome library of Bacillus cereus. *Proc Natl Acad Sci U S A*. 1999;96:6451–5.