Metagenomics, Metatranscriptomics, and Metabolomics Approaches for Microbiome Analysis

Vanessa Aguiar-Pulido[1], Wenrui Huang[1], Victoria Suarez-Ulloa[3], Trevor Cickovski[1,4], Kalai Mathee[2,5,6] and Giri Narasimhan[1,2]

[1]Bioinformatics Research Group (BioRG), School of Computing and Information Sciences, Florida International University, Miami, FL, USA. [2]Biomolecular Sciences Institute, Florida International University, Miami, FL, USA. [3]Chromatin Structure and Evolution Group (Chromevol), Department of Biological Sciences, Florida International University, Miami, FL, USA. [4]Department of Computer Science, Eckerd College, St. Petersburg, FL, USA. [5]Herbert Wertheim College of Medicine, Florida International University, Miami, FL, USA. [6]Global Health Consortium, Florida International University, Miami, FL, USA.

Supplementary Issue: Bioinformatics Methods and Applications for Big Metagenomics Data

ABSTRACT: Microbiomes are ubiquitous and are found in the ocean, the soil, and in/on other living organisms. Changes in the microbiome can impact the *health* of the environmental niche in which they reside. In order to learn more about these communities, different approaches based on data from multiple *omics* have been pursued. Metagenomics produces a taxonomical profile of the sample, metatranscriptomics helps us to obtain a functional profile, and metabolomics completes the picture by determining which byproducts are being released into the environment. Although each approach provides valuable information separately, we show that, when combined, they paint a more comprehensive picture. We conclude with a review of network-based approaches as applied to integrative studies, which we believe holds the key to in-depth understanding of microbiomes.

KEYWORDS: microbiome, metagenomics, metatranscriptomics, metabolomics, networks

SUPPLEMENT: Bioinformatics Methods and Applications for Big Metagenomics Data

CITATION: Aguiar-Pulido et al. Metagenomics, Metatranscriptomics, and Metabolomics Approaches for Microbiome Analysis. *Evolutionary Bioinformatics* 2016:12(S1) 5–16 doi: 10.4137/EBO.S36436.

TYPE: Review

RECEIVED: December 01, 2015. **RESUBMITTED:** January 26, 2016. **ACCEPTED FOR PUBLICATION:** January 31, 2016.

ACADEMIC EDITOR: Jike Cui, Associate Editor

PEER REVIEW: Four peer reviewers contributed to the peer review report. Reviewers' reports totaled 1475 words, excluding any confidential comments to the academic editor.

FUNDING: GN and KM gratefully acknowledge the support from the Alpha-One Foundation. The work of VAP was supported by the College of Engineering and Computing, Florida International University (FIU). VSU was supported by the Department of Biological Sciences, FIU. TC was supported by a Faculty Development Grant from Eckerd College. The authors confirm that the funders had no influence over the study design, content of the article, or selection of this journal.

COMPETING INTERESTS: Authors disclose no potential conflicts of interest.

CORRESPONDENCE: giri.narasimhan@fiu.edu

Introduction

Communities of microbes are found in diverse environmental niches, such as the ocean, soil, and inside host organisms, including all animals, plants, and lower eukaryotes.[1] These communities show characteristics, such as complexity, diversity, interaction, cooperation, dynamism, generosity, danger, and competition.[2] In such communities, microbes may compete for nutrients,[3] share functional genes through horizontal gene transfer,[4] produce toxins that can kill other microbes,[5] produce various metabolites and signaling molecules for sharing and communication,[6] and combine forces to fight common enemies, such as the host immune system.[7] In short, the importance of the microbial community stems from the fact that they are critical to the *health* of the environmental niche in which they reside,[8] and an imbalance in the community could be harmful.[9]

Traditionally, a microbiome has been defined as a microbial community occupying a reasonably well-defined habitat.[10] One of the most common approaches to studying a microbiome is analyzing its constituent microbial genomes through metagenomics. More recently, this definition has evolved to include not only the microbes and their genomes but also the aggregate of environmental and host factors. The inclusion of the host environment as part of the microbiome significantly expands its implications, with the interactions between the host and its associated microbial community now relevant to understanding the dynamics of the microbiome. For evolutionary and functional studies of the microbiome, modifications in the host environment (eg, a diet shift in the host organism or a compositional change in the environmental matrix under study) now become critical and must be taken into consideration. Coevolution processes can then be identified, providing valuable information to understand the relationship of the microbial community with its host. This apparent conceptual shift is accompanied by the recognition that, in order to achieve a more comprehensive study of microbiomes, metagenomics must be combined with other omic approaches. Many relevant omic approaches have been proposed for microbiome

studies. In this article, we discuss metatranscriptomics and metabolomics, which are rapidly becoming critical to microbiome studies.

Metagenomics is the study of the genomes in a microbial community and constitutes the first step to studying the microbiome. As seen in the "Metagenomics" section, metagenomics comes in different flavors. However, its main purpose is to infer the taxonomic profile of a microbial community. Although whole-metagenome sequencing (WMS) provides a partial glimpse into the functional profile of a microbial community, it is better inferred using metatranscriptomics, which involves sequencing the complete (meta)transcriptome of the microbial community. Metatranscriptomics informs us of the genes that are expressed by the community as a whole. With the use of functional annotations of expressed genes, it is possible to infer the functional profile of a community under specific conditions, which are usually dependent on the status of the host. While metagenomics helps address the question "what is the composition of a microbial community under different conditions?", and metatrascriptomics helps answer the question "what genes are collectively expressed under different conditions?", the question considered by metabolomics is "what byproducts are produced under different conditions?". The metabolites released by the microbial community are largely responsible for the health of the environmental niche that they inhabit.

Regardless of whether microbiome studies are biomedical or environmental in their focus, it is clear that the different omic approaches provide invaluable information. However, the best results are obtained by performing *integrative* studies that involve all available omic datasets.[11] While such efforts hold promise, the integration must be done carefully.[12]

As suggested by a variety of different analyses,[13–16] we believe that network-based approaches can lead to a sophisticated in-depth analysis of microbiomes, particularly when applied to integrative studies, and consequently lead to critical insights into the world of microbiomes.

Major microbiome initiatives. *Human microbiome studies.* The National Institute of Health has funded a major initiative that aims to generate resources for a comprehensive characterization of the human microbiome to understand its impact on human health and disease. The first phase, known as the Human Microbiome Project (HMP),[17] focuses on the study of microbial communities that inhabit the human body of healthy individuals,[18,19] with particular emphasis on nasal, oral, skin, gastrointestinal, and urogenital areas.[17,18,20–23] It is known that the amount of microbial cells present in the human body is notably larger than the amount of human cells. These bacterial communities play critical roles, such as assisting in the digestion of food, synthesizing necessary vitamins, and aiding the immune system in defending our body from pathogenic invaders.[24] Human microbiome studies have revealed strong correlations between changes in microbial community profiles and diseases.[22,25–27] These studies have also shown that the structure of the microbial community is significantly different in five areas of the human body (gut, mouth, airways, urogenital, and skin), and that this seems to be independent of gender, age, and ethnicity.[18,19] All the data and protocols associated with this project are available at the HMP Data Analysis and Coordination Center (DACC).[28]

The Integrative HMP (iHMP)[27] is the second phase of this initiative, going a step further by gathering multiple omic data from both the microbiome and the host. This is part of a longitudinal study with a broader objective of understanding host–microbiome interactions using integrative analyses. Another related initiative focused on the human microbiome is the Metagenomics of the Human Intestinal Tract (MetaHIT) project.[29] This project was funded by the European Seventh Framework Programme until 2012. Its goal was to understand the link between the human intestinal microbiota and human health/disease. For this purpose, they focused on two disorders of increasing incidence in Europe: obesity and inflammatory bowel disease. Similarly, the Human Food Project and the American Gut Project[30] focus on the gut microbiome with the aim of determining how to acquire a healthy microbiome through food.

Environmental microbiome studies. The Earth Microbiome Project (EMP) is a remarkable effort started in 2010 to characterize the diversity, distribution, and structure of microbial ecosystems across the planet and has already gathered over 30,000 samples.[31] Their focus is on diverse ecosystems, including not only the ones within the bodies of humans, animals, and plants but also terrestrial, marine, freshwater, sediment, air, and constructed environments, as well as every intersection of these ecosystems.

J. Craig Venter Institute's (JCVI) Global Oceanic Sampling (GOS) expeditions and the European Tara Oceans initiatives[32–36] have focused on understanding and cataloging the marine microbiome diversity across the planet. JCVI's vessel, *Sorcerer II*, has made multiple oceanic expeditions to collect samples from oceans across the globe. Their multistage processing allows them to exploit size differences to separate different groups of microbes, including large microzooplankton and phytoplankton (3–20 μm), picoplankton and large cyanobacteria (0.8–3 μm), prokaryotes and large viruses (0.1–5 μm), and viroplankton (below 0.1 μm).

Metagenomics

Metagenomics allows us to investigate the composition of a microbial community. *Genomic* studies consider the genetic material of a specific organism, while *metagenomics* (*meta* meaning *beyond*) refers to studies of genetic material of entire communities of organisms. This process usually involves next-generation sequencing (NGS) after the DNA is extracted from the samples. NGS produces a large volume of data in the form of short reads, from which a microbial community profile or other information can be pieced together just like gathering information from the pieces of a puzzle.

Recently, some authors have argued in favor of a terminological distinction between metagenomics (used to describe a broad comprehensive genomic approach to microbiome profiling) and metataxonomics (which uses amplicons from a targeted marker gene in order to make taxonomic inferences).[37] One popular marker gene used in metataxonomic studies is *16S rDNA*.[13,38–42] A large number of databases are available for amplicons targeted in this region[43–45] and to aid in classification of reads and in building taxonomic profiles of a microbiome. With the advancement of technology, studies have shifted toward shotgun approaches,[46] such as WMS. As a result, a number of specialized databases with complete reference genomes have been developed.[47] These databases are then used to construct taxonomic profiles[18,48,49] but are also useful for inferring potential functional profiles for the microbial community based on the collection of genes present in the sample.

Tools and techniques. A variety of tools and analysis pipelines have been developed to analyze metagenomic data.[50] problem solving environments (PSEs[51]) provide user-friendly workbenches to develop flexible scientific analysis pipelines using a menu of available tools. Such workbenches incorporate different ranges of generality. For instance, Galaxy[52] maximizes generality by providing a framework for genomic analysis while allowing the user to supply tools and file formats for various stages in a pipeline. Galaxy can execute jobs remotely, allows for *undoing* or repeating of individual steps, and permits inspection of intermediate results but requires considerable computational and storage resources. QIIME[53] provides a set of integratable scripts for analyzing raw microbial DNA samples including taxonomic classification using marker genes, such as *16S rRNA*, but allows flexible pipelines to be constructed. Mothur[54] was initially designed to target the microbial ecology community but has since been adopted by the human microbiome community as well. It provides an extensible package with functionality accessible through a domain-specific language. Like QIIME, Mothur is also a metataxonomic tool, focusing on marker genes, such as *16S rRNA*. Pathoscope[55] provides a pipeline that can identify bacterial strains present in a series of raw sequences and generate reports of statistics, such as percentages, gene locations, and protein products. Ideally, a PSE should be open source, infinitely extensible, lightweight, and able to accommodate any tool, user, or developer.

As shown in Figure 1, metagenomic analysis pipelines can be divided into three main steps: (1) preprocessing the reads, (2) processing the reads, and (3) downstream analyses.

Preprocessing and processing the reads. The procedures followed in preprocessing and processing of the reads (steps 1 and 2) have become fairly standardized. Hence, we describe them briefly and focus mostly on downstream analysis ("Downstream analyses of metagenomic data" section).

Preprocessing mainly involves removing adapters from reads, filtering reads by quality and length, removing

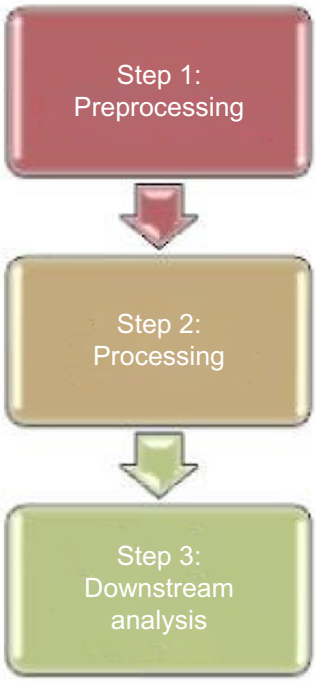

Figure 1. Generic microbiome analysis pipeline.

contaminants, identifying and removing any chimeric sequences that may have been generated during polymerase chain reaction (PCR) amplification, and preparing data for subsequent analysis. A survey of some of the popular tools and techniques currently available for this step can be found in Kim et al.[50]

After preprocessing of the reads, the next step is to classify each read based on the taxa with the highest probability of being the origin of that read. This step often uses a reference database of relevant microbial genomes and produces a microbial profile usually represented as an abundance matrix with microbial taxa as rows, samples as columns, and values representing the abundance of a taxon in the sample.

In the case of metataxonomics, reads are frequently grouped (or clustered) prior to assigning a label. Unlike WMS, which produces a lower coverage and may identify thousands of strains per sample, targeted approaches have reads that come from relatively small regions of the genome, making this extra clustering step valuable in lowering errors in the classification. Groups of reads that result from the clustering process displaying similarity in sequence and/or composition are inferred to have a common origin and referred to as operational taxomonic units (OTUs).

The classification and labeling performed on the reads can be either taxonomy dependent or taxonomy independent. Taxonomy-dependent methods use a database of reference genomes, which has some bias toward data with pathogenic or commercial applications. Methods in this category can be further classified as alignment-based, composition-based, or hybrid. Alignment-based methods usually give the highest

accuracy but are limited by the reference database and by the alignment parameters used and are generally computation and memory intensive. Composition-based methods store only compact models instead of the whole genome, requiring fewer computational resources. These methods use features extracted from the genomes (eg, GC percentage and codon or oligonucleotide usage patterns) to build models but have not yet achieved the accuracy of alignment- based approaches. Hybrid approaches offer a compromise between the two. Taxonomy-independent methods, on the other hand, do not require a priori knowledge. Instead, they segregate reads based on properties, such as distance, k-mers, abundance levels, and frequencies. These methods are typically used if the samples are more likely to have microbes that are not documented in the databases. Chen et al.[56] and Mande et al.[57] reported an extensive review of popular tools and techniques used for processing 16S reads and for processing WMS reads, respectively.

Accurate classification and labeling are challenging because (a) sequencing technologies produce short reads, (b) for economic reasons the datasets often have low coverage of the genomes in the microbiome, (c) some sequencing technologies have a high percentage of sequencing errors, and (d) the reference genome databases used are not comprehensive, often failing to provide an accurate taxonomic context because of lateral gene transfers between microbial taxa.

Downstream analyses of metagenomic data. Once the reads have been assigned labels or classified as best as possible, downstream analyses attempt to extract useful knowledge from the data. Typical questions addressed in this step include "how diverse are the microbial taxa in the sample?", "what is the functional profile of the genes present and/or expressed in the microbial community?", "what microbial taxa are differentially abundant in the samples?", "what phylogenetic groups, functional and metabolic pathways, orthologous groups of genes, and gene ontology terms are particularly enriched or depleted in the samples?", and "what microbial groups tend to co-occur or co-avoid in the samples of interest?". We now review several current tools and techniques for performing downstream analysis.

Richness and diversity are measures that have traditionally been used to characterize a metagenomic sample.[58,59] Richness is a simple count of taxa present in a sample. Diversity refers to a collection of indices and measures (eg, Shannon, Chao, Simpson, and Berger–Parker) that quantify the evenness of the distribution of the abundances of the taxa,[59] often incorporating distance measures or similarity indices (eg, Jaccard, Sorenson, and Bray–Curtis). Richness and diversity offer measures of complexity of the community but disclose little about interactions within the community, which requires more complex downstream analyses.

Visualizing taxonomic profiles is a task that has been addressed by several initiatives. Krona,[60] for example, is a simple and intuitive web-based tool to visualize the taxonomic profile as a pie chart with an embedded hierarchy. In contrast, the Visualization and Analysis of Microbial Population Structure (VAMPS) tool[61] can measure and visualize statistically significant similarities and differences between multiple taxonomic profiles of complex microbial communities.

Integrating additional information in metagenomic analyses is extremely valuable in order to provide improved perspectives of the microbial profiles. Based on this premise, a number of approaches have sought the use of phylogenetic information to enhance the labeling and classification of reads, as is the case with Amphora2,[62] which performs phylogenetic inference using phylum-specific marker databases. This type of inference can be done algorithmically as well, through edge principal component analysis (PCA) and squash clustering.[63] Phymm[64,65] is a software package that classifies sequence fragments into phylogenetic groups using interpolated Markov models. Finally, PPlacer[66] performs phylogenetic placement using a fixed reference tree and maximum-likelihood inference with distance calculations to indicate uncertainty and can be executed in parallel.

A more significant improvement is possible with the help of functional annotations of the genes to which the reads are mapped.[67,68] Although many analytical metagenomic approaches focus on the composition or structure of the samples, functional profiling is also essential, as it provides insight into the underlying biological processes. Other useful resources for annotation include gene ontology (GO),[69,70] Kyoto Encyclopedia of Genes and Genomes (KEGG),[71,72] and Clusters of Orthologous Groups (COG).[73,74] As a part of the HMP initiative to analyze WMS data, a methodology called HUMAnN[75] was developed for inferring the functional and metabolic potential of a microbial community.

Alternatively, other existing tools, such as IMG/M,[76] CAMERA,[77] METAREP,[78] MEGAN,[79] and CoMet,[80] can also be used to obtain functional profiles of microbiomes. IMG/M, METAREP, and CoMet provide a web-based user interface, while CAMERA aims to offer a state-of-the-art computational structure for high-performance network access and grid computing as a part of a distributed architecture. In contrast, MEGAN is a standalone computer program. METAREP and CoMet annotate the data with GO and KEGG, whereas MEGAN uses the NCBI taxonomy to summarize and order the results obtained after performing BLAST. METAREP also offers the option to annotate the data with taxonomic information, and IMG/M uses BLAST to infer phylogenetic information from the sample. However, IMG/M is more oriented toward protein-related information by annotating the results with resources, such as COG, Pfam, TIGRFAMs, ENZYME, and KEGG. IMG/M was developed by the Joint Genome Institute and contains data from the HMP and the Genome Encyclopedia of Bacterial and Archaea Genomes. CAMERA has been designed for environmental and ecological purposes with the aim of providing new ways of visualizing and interacting with data and was

applied to data from GOS. METAREP, on the other hand, was developed at JCVI. It performs statistical tests and mutidimensional scaling (MDS) and can also produce graphical summaries, heatmaps and hierarchical clustering plots. MEGAN uses the lowest common ancestor algorithm to label the reads and has been applied to datasets, such as the Saragaso Sea dataset, and data from mammoth bone. Finally, CoMet combines open reading frame finding and assignment of protein sequences to Pfam domain families with comparative statistical analysis, providing the user with comprehensive tabular data files and visualizations in the form of hierarchical clustering and MDS. It was applied to 454 data.

Obtaining the functional profile is typically not possible with targeted approaches, since it provides no direct evidence of the functional capabilities of the microbial community. However, the tool Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) shows how to infer a functional profile of a microbial community directly from taxonomic profiles of marker genes, such as the *16S rDNA*, and a database of reference genomes.[81] Their results provide useful insights on uncultivated microbial communities, prior to which only marker gene surveys were available.

Discussion. In summary, metataxonomics helps us to compute the taxonomic profile of a microbial community, while metagenomics helps us to compute the functional profile by focusing on the gene content and using the available functional annotations of the corresponding proteins. While metagenomics is powerful, solely using it to study a microbiome is limited in value. Many experts have confirmed that the percentage of documented bacteria is very low compared to the estimate of bacterial species on our planet.[82] This may be due partially to the impossibility of culturing complex environments or replicating in the laboratory the real conditions in which the microbiome exists. Either way, the reference databases used to classify and label bacteria are limited to what has been cataloged. Current methods typically either discard reads from undocumented microbes or label them based on the *closest* documented microbe from the database. Thus, inevitably, results will be based on a biased percentage of bacteria present in the samples, representing the first shortcoming of these methods. Another limitation is that metagenomics cannot reveal dynamic properties, such as the spatiotemporal activity of the community and the impact of the environment on these activities. The only information that can be obtained at a functional level is the potential of the microbiome to display functional properties associated with the presence of genes with no information about their expression levels or lack thereof. The need to monitor gene expression patterns brings us to the topic of our next section, metatranscriptomics.

Metatranscriptomics

By focusing on what genes are expressed by the entire microbial community, metatranscriptomics sheds light on the *active* functional profile of a microbial community.[83] The metatranscriptome provides a snapshot of the gene expression in a given sample at a given moment and under specific conditions by capturing the total mRNA. Pioneering studies aiming to identify expressed genes in environmental samples date back to 2005[84,85] and represent the dawn of metatranscriptomics. However, these were limited to a relatively narrow group of genes. As for metagenomics, it is now possible to perform whole metatranscriptomics shotgun sequencing. This (meta)genome-wide expression provides the expression and functional profile of a microbiome.[48,86,87]

When processing reads, a typical metatranscriptomics analysis pipeline will either (1) map reads to a reference genome or (2) perform de novo assembly of the reads into transcript contigs and supercontigs. The first strategy, in a manner similar to the alignment-based methods in WMS, maps reads to reference databases, thus gathering information to infer the relative expression of individual genes. The second strategy infers the same but with assembled sequences. The first strategy is limited by the information in the database of reference genomes. The second strategy is limited by the ability of software programs to assemble contigs and supercontigs correctly from short reads data.

Tools and techniques. The application of metatranscriptomics to the study of the microbiome is far less common relative to other omics reviewed in this article. Most analysis pipelines described in the literature were built ad hoc. The majority of these methods follow the aforementioned first strategy based on read mapping.[88–92] In this case, metatranscriptomic reads are generally mapped to specialized databases (usually downloaded from the NCBI) using alignment tools, such as Bowtie2, BWA, and BLAST. The results are then annotated using resources, such as GO, KEGG, COG, and Swiss-Prot. Finally, different types of downstream analysis are carried out depending on the goal of the study (eg, PCA-based phylogenetic analysis or enrichment analysis). The latest metatranscriptomics techniques include stable isotope probing (SIP), which has been used to retrieve specific targeted transcriptomes of aerobic microbes in lake sediment.[93] This not only helps to target specific organisms but also contributes significantly to metabolomics studies.

The second strategy requires assembling metatranscriptomic reads into longer fragments called contigs. For this purpose, numerous software packages are available. Celaj et al.[94] compared de novo sequence assemblers to reference-based *mapping* tools. The compared tools included Trinity,[95] MetaVelvet,[96] Oases,[97] AbySS, Trans-Abyss, and SOAPdenovo,[98–100] as well as tools such as Scripture and Cufflinks.[101,102] It was found that compared to other tools Trinity not only outperformed all of them but also appeared to be best tuned for sensitivity across the broadest range of expression levels. This was particularly noticeable in reconstructing transcripts within the highest expression quintiles, in which other de novo strategies failed to perform well.[95] Li and Dewey[103] developed RNA-Seq by Expectation Maximization (RSEM),

a quantitative pipeline for transcriptomic analysis, currently provided as stand-alone software or a plug-in within Trinity. RSEM takes as input a reference transcriptome or assembly (most likely obtained through Trinity) along with RNA-Seq reads generated from the sample and calculates normalized transcript abundance (ie, the number of RNA-Seq reads corresponding to each reference transcriptome or assembly).[104,105] Although both Trinity and RSEM were designed for transcriptomic datasets (ie, obtained from a single organism), it may be possible to apply them to metatranscriptomic data (ie, obtained from a whole microbial community). MEGAN annotates results with GO to perform enrichment analysis.[106]

Discussion. Although current metatranscriptomic techniques are promising, there are still several obstacles that limit their large-scale application. First, much of the harvested RNA comes from ribosomal RNA, and its dominating abundance can dramatically reduce the coverage of mRNA, which is the main focus of transcriptomic studies. Some efforts have been made to effectively remove rRNA.[107] Second, mRNA is notoriously unstable, compromising the integrity of the sample before sequencing. Third, differentiating between host and microbial RNA can be challenging, although commercial enrichment kits are available. This may also be done *in silico* if a reference genome is available for the host, as in the work of Perez-Losada et al.[108] who consider the impact of host–pathogen interactions on the human airway microbiome. Finally, transcriptome reference databases are limited in their coverage.

WMS approaches provide information on the taxonomic profile of a microbial community as well as its potential functional profile; in contrast, whole metatranscriptome sequencing describes the active functional profile. This would help in studying the dynamics of functional profiles with varying conditions. We now discuss metabolomics, which studies the consequences of the shifts in the collective gene expression of the microbial community that modifies the very medium where the microbial community must feed, grow, reproduce, and cooperate or compete to survive.

Metabolomics

Metabolomics is the comprehensive analysis by which all metabolites of a sample (small molecules released by the organism into the immediate environment) are identified and quantified.[109] The metabolome is considered the most direct indicator of the health of an environment or of the alterations in homeostases (ie, dysbiosis).[110] Variation in the production of signature metabolites are related to changes in activity of metabolic routes, and therefore, metabolomics represents an applicable approach to pathway analysis.[111] Additionally, the application of metabolomics for drug discovery and pharmacogenomics represents a promising avenue for personalized medicine.[112]

The metabolomic profile associated with the microbiome may show a strong dependence on environmental factors (eg, diet, exposure to xenobiotics, and environmental stressors), providing valuable information not just about the characteristics of the microbiome but also about the interactions of the microbial community with the host environment.[113–115] Thus, metabolomics aims to improve our understanding of the role of the microbiome in the transformation of nutrients and pollutants as well as other abiotic factors that may affect the homeostasis of the host environment. Microbial communities exert a strong influence on critical biogeochemical cycles, and the study of their metabolome can help to develop predictive biomarkers for environmental stressors.[116] The microbiome is regarded as a biological reactor that, based on its genetic pool, can transform resources and hazardous elements into products that are either beneficial or detrimental to the health of its environment. A good example is bioremediation and its application to reduce the consequences of pollution.[117]

Most interestingly, the metabolome can illustrate signaling processes involved during communication between bacteria, such as quorum sensing, which relates gene expression responses to changes in cell population density.[118–123] A deeper understanding of the communication mechanisms within microbial communities could possibly revolutionize the current strategies in areas such as infections disease control, and optimize agricultural exploitation in environmental conservation. Thus, metabolomics complements the information provided by the other omics (mentioned earlier) by describing not just biological systems themselves, but how they interact internally and externally.

Generating metabolomics data differs significantly from generating metagenomics and metatranscriptomics data, which rely heavily on sequencing. Identifying and quantifying metabolites is typically carried out using a combination of chromatography techniques (ie, liquid chromatography, LC, and gas chromatography, GC) and detection methods, such as mass spectrometry (MS) and nuclear magnetic resonance (NMR). For a more detailed review of these technologies and their many variants, we refer the reader to a recent review by Aldridge and Rhee.[124] These technologies produce spectra consisting of patterns of peaks that allow both the identification and quantification of metabolites. These patterns (either predicted or experimentally obtained) are stored in spectral databases, allowing automated analysis and generation of metabolomic profiles. With these technological resources, metabolomics fulfills the requirements of a high-throughput analytical method, and thus data analysis represents a critical step in knowledge generation. As a result, we have seen a rise in software development, large data repositories, and initiatives for standardization. This in turn paves the road for data integration.

Tools and techniques. The analysis pipeline for spectral metabolomic data involves three steps: (1) preprocessing, (2) statistical analysis, and (3) machine learning techniques for pattern recognition.[125] In the first step, denoising and peak-picking improve the quality of the data to be processed.

Once the peak pattern has been established, a comparison against spectral databases identifies the metabolites in the sample and the area below the peaks their respective quantities. To automate this process, spectral databases are maintained and curated by specialized international consortia that emphasize standardization. These include the following: the Human Metabolome Database, a cross-referenced database about the small metabolites found in the human body[126–128]; the BioMagResBank, which works as a central repository for experimental NMR data including both small metabolites and macromolecules[129]; the Madison-Qingdao Metabolomics Consortium Database,[130] which includes both NMR and MS data thoroughly annotated collected from other databases and literature; MassBank,[131] which merges spectral data from different collision-induced dissociation conditions to improve the precision in the identification of compounds; the Golm Metabolome Database,[132] which stores spectral data with retention indexes, useful for automated identification of compounds analyzed with GC–MS; and the METLIN Metabolite Database,[133] which contains curated spectral information of biological metabolites without information of the environmental context from which the samples where obtained. Each of them differs slightly in functionality but pursues similar goals, serving as repositories of spectral data and offering links to their biological interpretation.

Discussion. By cataloging all metabolites present in a sample, metabolomics offers a powerful way to relate the metabolites to the cellular processes of which they are the byproducts. The combination of metabolomic and pathways information can lead to new hypotheses. One important challenge of this approach is difficulty in determining whether a metabolite was generated by the host or by the microbiome. In addition, if conclusions are to be made about which genes, enzymes, or pathways are associated with a specific metabolite, the results obtained from a metabolomic study must be combined with other omic data. This highlights the need for new approaches that deal with integrated omics, as discussed in the "Integrating multiomic data" section.

Integrating Multiomic Data

Standard analyses of individual omic datasets focus on the community structure and functional roles of individual taxa or groups of taxa. The remaining challenge lies in elucidating the large, dynamic, and complex network of interactions between its constituent entities. With the increasing availability of heterogeneous multiomic datasets,[11] the need for integrative analyses has become even more urgent. A reasonable approach (Fig. 2) is to perform separate analysis, adding an extra integrative step within *downstream analysis*.

Integrating multiple omic datasets is a problem that researchers are just beginning to tackle.[12] Bringing together different studies will allow researchers to build and test mathematical models of microbial activity and interaction, enabling a better understanding of the interplay between the environment and the microbial community.[134,135] For example, the combination of metagenomics and metatranscriptomics may reveal overexpression or underexpression of particular functions and, in some cases, the activities of specific organisms.[90,136–138] The addition of metabolomics could provide insight into the outcome of those changes in gene expression, which may lead to differential expression of specific metabolites that impact the health of the host environment.[139–144] Understanding the whole ecosystem opens new avenues and exciting approaches for generating new knowledge. By combining multiple (potentially noisy and heterogeneous) data types, we can build support for specific hypotheses; if independent lines of evidence arrive at the same conclusion, then our confidence in that conclusion will grow.

Tools and techniques. Current studies indicate that integrating metagenomics and metatranscriptomics has the potential of attributing functional changes in gene expression to specific members of the microbial community. Franzosa et al.[145] showed a relationship between genomic abundances and differential regulations of microbial transcripts, discovering up- and downregulated pathways within the human gut microbiome. Shi et al.[146] applied this integrative approach relating the functional and taxonomic profiles of marine environmental samples. Current studies also indicate that integrating the results of metagenomics with metabolomics can provide insight into how members of a microbial community interact with each other and with their environment.[147] For example, Lu et al.[148] observed a simultaneous effect on both microbiome composition and metabolite production upon introducing arsenic into the mouse gut environment. Zhang et al.[149] performed a similar study with the introduction of disinfection byproducts from drinking water. These studies illustrate that the different omics are interdependent and that an integrated approach can lead to more useful discoveries.

Several current studies suggest that integrating all three omic data – metagenomics, metatranscriptomics, and metabolomics – would provide a complete picture from genes to phenotype.[150,151] With the wealth of datasets available but not currently integrated, Abram[152] argues for a system-based approach to multiomics, which would allow predictive modeling. In particular, he points out that studying interrelationships between entities (which he refers to as SIP-omics) would provide some guidance to establishing linkages between various datasets.

Interrelationships also form the basis of the reverse ecology algorithm,[153] which attempts to connect microbial communities with properties of their environment under the assumption that adaptation to the environment is most fundamental to their structure and topology. The set of metabolites that are acquired by an organism from external sources is called the *seed set* and represents the metabolic *interface* with the environment. Borenstein et al.[154] showed how to compute the seed set for individual organisms and how it can be used

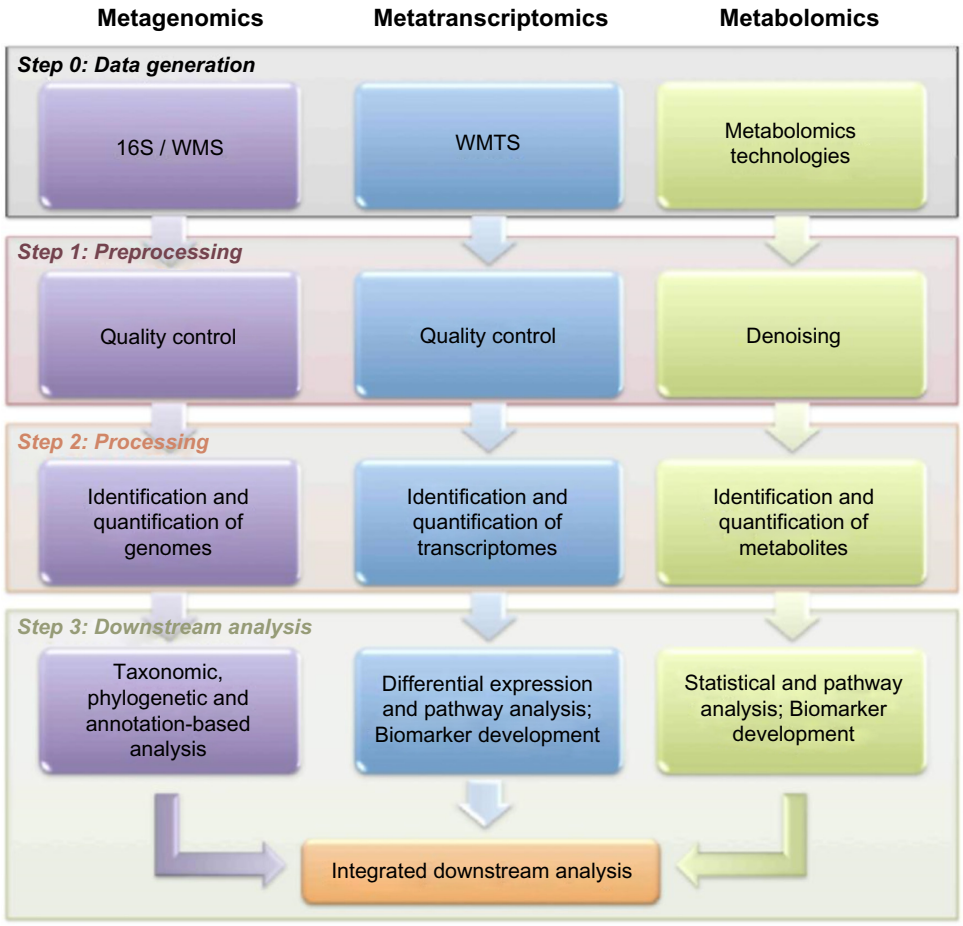

Figure 2. Generic multiomic analysis pipeline.

to characterize the effective biochemical habitat. Ebenhöh et al.[155] offered predictive models of an organism's ability to flourish in specific environments.

Conclusion and Future Directions

In this article, we have discussed how three different omic approaches – metagenomics, metatranscriptomics, and metabolomics – provide useful information toward understanding microbiomes. We also discussed how the value of an integrative approach is greater than the sum of its parts.

Biological networks have long been used to model interactions between biological entities, with applications to areas, such as gene regulation, metabolic and signaling pathways, protein–protein networks, and food webs in ecology.[156–159] With its proven application to analyzing interrelationships and their critical role in multiomics, we believe biological network analysis will be critical to future multiomic approaches to studying the microbiome. In addition, network analyses offer the possibility of exploring both local (eg, relationship with neighbors) as well as global properties (eg, connectivity) of a community. Dutkowski et al.[160] studied the assignment of ontologies using networks and developed tools, such as Cytoscape,[161] to perform these analyses.

Metagenomic studies have shown that interactions within a microbiome can be naturally modeled using a network representation,[14,42,162] with properties closely related to *social networks*.[15,24] Macroscale community structures have been observed in these types of networks, indicating *clubs* (ie, groups of co-occurring bacteria) as well as *rival clubs* (ie, groups of bacteria that tend to not co-occur).[15,42]

In order to integrate data from various omic sources, microbiomes can also be modeled as heterogeneous networks (Fig. 3), which provides a visual description of what such a network in the context of the microbiome would look like. A heterogeneous network would allow researchers to generate new interesting hypotheses that involve entities from the different omics described in this article (represented in the figure by nodes with different shapes and colors). For instance, we could potentially have a *club* that includes genes, microbes, and metabolites. Heterogeneous networks have been used in other applications, such as associations between genetic interactions and protein–protein interactions in order to infer cellular function.[163] Another study couples these same types of networks to infer gene dependencies and new processes, such as DNA damage repair, and also different types of co-expression networks.[164]

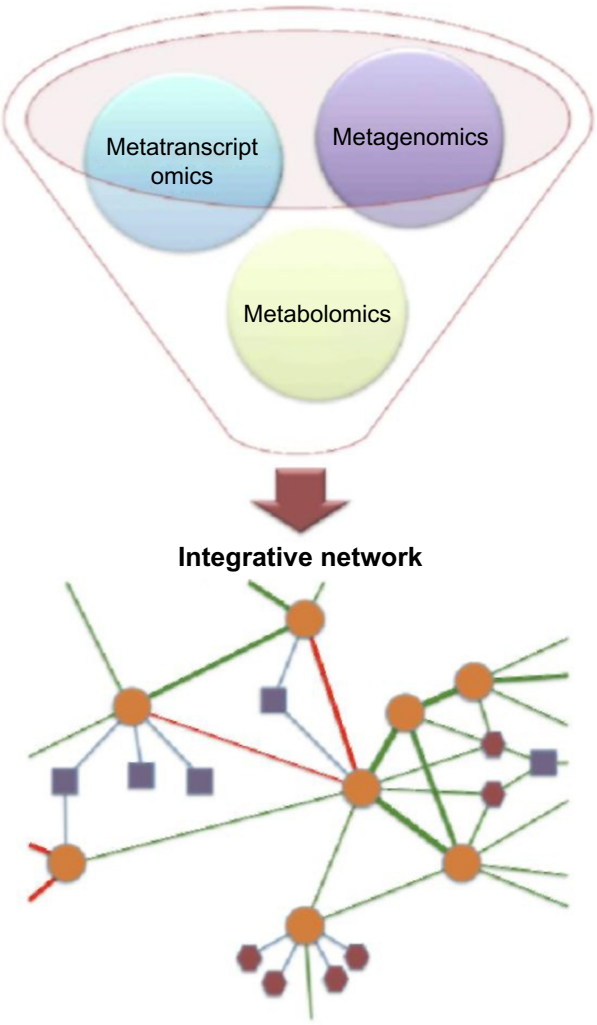

Figure 3. Integrated networks for multiomic data.

Many types of omic networks were also integrated to study gene regulation in the bacterium *Mycobacterium tuberculosis*.[165] Other omic areas not included in this study include metaproteomics, metalipidomics, and metaglycomics. We believe that analyzing heterogeneous networks built across multiple omic datasets is critical to linking the different levels of complexity inherent to biological systems, thus establishing a more comprehensive understanding of the nature and dynamics of microbiomes.

Author Contributions

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

REFERENCES

1. Ley RE, Peterson DA, Gordon JI. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*. 2006;124(4):837–48.
2. Costello EK, Lauber CL, Hamady M, et al. Bacterial community variation in human body habitats across space and time. *Science*. 2009;326:1694–7.
3. Hibbing ME, Fuqua C, Parsek MR, Peterson SB. Bacterial competition: surviving and thriving in the microbial jungle. *Nat Rev Microbiol*. 2010;8:15–25.
4. Liu L, Chen X, Skogerb G, et al. The human microbiome: a hot spot of microbial horizontal gene transfer. *Genomics*. 2012;100(5):265–70.
5. Proft T. *Microbial Toxins: Current Research and Future Trends*. Caister Academic Press, Norfolk, UK; 2009.
6. Sharon G, Garg N, Debelius J, Knight R, Dorrestein PC, Mazmanian SK. Specialized metabolites from the microbiome in health and disease. *Cell Metab*. 2014;20(5):719–30.
7. Kau AL, Ahern PP, Griffin NW, Goodman AL, Gordon JI. Human nutrition. *Nature*. 2011;474(7351):327–36.
8. Foxman B, Martin ET. Use of the microbiome in the practice of epidemiology: a primer on -omic technologies. *Am J Epidemiol*. 2015;182(1):1–8.
9. Betts K. A study in balance: how microbiomes are changing the shape of environmental health. *Environ Health Perspect*. 2011;119(8):340–6.
10. Whipps JM, Lewis K, Cooke RC. *Mycoparasitism and Plant Disease Control*. Manchester University Press, Manchester, UK; 1988:161–87.
11. Segata N, Boernigen D, Tickle TL, Morgan XC, Garrett WS, Huttenhower C. Computational metaomics for microbial community studies. *Mol Syst Biol*. 2013;9(1):666–80.
12. Franzosa EA, Hsu T, Sirota-Madi A, et al. Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nat Rev Microbiol*. 2015;13(6):360–72.
13. Barberan A, Bates ST, Casamayor EO, Fierer N. Using network analysis to explore cooccurrence patterns in soil microbial communities. *ISME J*. 2011;6:343–51.
14. Faust K, Raes J. Microbial interactions: from networks to models. *Nat Rev Microbiol*. 2012;10:538–50.
15. Fernandez M, Riveros JD, Campos M, Mathee K, Narasimhan G. Microbial "Social Networks". *BMC Genomics*. 2015;16(suppl 11):S6.
16. Fernandez M, Aguiar-Pulido V, Huang W, et al. Microbiome analysis: state-of-the-art and future trends. In: Mandoiu I, Zelikovsky A, eds. *Computational Methods for Next Generation Sequencing Data Analysis*. Wiley, Hoboken, NJ;2015:333–51.
17. Peterson J, Garges S, Giovanni M, et al. The NIH Human Microbiome Project. *Genome Res*. 2009;19:2317–23.
18. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;486:207–14.
19. Human Microbiome Project Consortium. A framework for human microbiome research. *Nature*. 2012;486(7402):215–21.
20. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. *Nature*. 2007;449:804–10.
21. Turnbaugh PJ, Gordon JI. The core gut microbiome, energy balance and obesity. *J Physiol (Lond)*. 2009;587:4153–8.
22. Marrazzo JM, Martin DH, Watts DH, et al. Bacterial vaginosis: identifying research gaps proceedings of a workshop sponsored by DHHS/NIH/NIAID. *Sex Transm Dis*. 2010;37:732–44.
23. Qin J, Li R, Raes J, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010;464:59–65.
24. Ackerman J. The ultimate social network. *Sci Am*. 2012;306(6):36–43.
25. Brown K, DeCoffe D, Molcan E, Gibson DL. Diet-induced dysbiosis of the intestinal micro-biota and the effects on immunity and disease. *Nutrients*. 2012;4(8):1095.
26. Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. *Nat Rev Genet*. 2012;13:260–70.
27. Integrative HMP (iHMP) Research Network Consortium. The integrative human microbiome project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. *Cell Host Microbe*. 2014;16(3):276–89.
28. Human Microbiome Project Consortium. *HMSCP – Shotgun Community Profiling*. Available at: http://hmpdacc.org/HMSCP/. Last accessed: Jan. 2016.
29. Ehrlich SD, MetaHIT Consortium. Metagenomics of the intestinal microbiota: potential applications. *Gastroenterol Clin Biol*. 2010;34:S23–8.
30. Goedert JJ, Hua X, Yu G, Shi J. Diversity and composition of the adult fecal microbiome associated with history of cesarean birth or appendectomy: analysis of the American gut project. *EBioMedicine*. 2014;1(2):167–72.
31. Gilbert JA, Jansson JK, Knight R. The earth microbiome project: successes and aspirations. *BMC Biol*. 2014;12(1):69.
32. Venter JC, Remington K, Heidelberg JF, et al. Environmental genome shotgun sequencing of the Sargasso sea. *Science*. 2004;304(5667):66–74.
33. Nealson KH, Venter JC. Metagenomics and the global ocean survey: what's in it for us, and why should we care? *ISME J*. 2007;1(3):185.
34. Lima-Mendez G, Faust K, Henry N, et al. Determinants of community structure in the global plankton interactome. *Science*. 2015;348(6237):1262073.

35. Karsenti E, Acinas SG, Bork P, et al. A holistic approach to marine eco-systems biology. *PLoS Biol*. 2011;9(10):e1001177.
36. Sunagawa S, Coelho LP, Chaffron S, et al. Structure and function of the global ocean microbiome. *Science*. 2015;348(6237):1261359.
37. Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. *Microbiome*. 2015;3:31.
38. Chaffron S, Rehrauer H, Pernthaler J, von Mering C. A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Res*. 2010;20:947–59.
39. Gonzalez A, Knight R. Advancing analytical algorithms and pipelines for billions of microbial sequences. *Curr Opin Biotechnol*. 2012;23:64–71.
40. Freilich S, Kreimer A, Meilijson I, Gophna U, Sharan R, Ruppin E. The large-scale organization of the bacterial network of ecological co-occurrence interactions. *Nucleic Acids Res*. 2010;38(12):3857–68.
41. Kuczynski J, Liu Z, Lozupone C, McDonald D, Fierer N, and Knight R. Microbial community resemblance methods differ in their ability to detect biologically relevant patterns. *Nat Methods*. 2010;7:813–9.
42. Faust K, Sathirapongsasuti JF, Izard J, et al. Microbial co-occurrence relationships in the human microbiome. *PLoS Comput Biol*. 2012;8(7):e1002606.
43. Cole JR, Wang Q, Fish JA, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res*. 2014;42:D633–42.
44. Pruesse E, Quast C, Knittel K, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res*. 2007;35(21):7188–96.
45. DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol*. 2006;72:5069–72.
46. Sharpton TJ. An introduction to the analysis of shotgun metagenomic data. *Front Plant Sci*. 2014;5:209.
47. Nelson KE, Weinstock GM, Highlander SK, et al. A catalog of reference genomes from the human microbiome. *Science*. 2010;328:994–9.
48. Frias-Lopez J, Shi Y, Tyson GW, et al. Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci (PNAS)*. 2008;105(10):3805–10.
49. Chain PS, Grafham DV, Fulton RS, et al. Genome project standards in a new era of sequencing. *Science*. 2009;326:236–7.
50. Kim M, Lee K-H, Yoon S-W, Kim B-S, Chun J, and Yi H. Analytical tools and databases for metagenomics in the next-generation sequencing era. *Genomics Inform*. 2013;11(3):102–13.
51. Gallopoulos E, Houstis E, Rice J. Computer as thinker/doer: problem-solving environments for computational science. *Comput Sci Eng IEEE*. 1994;1:11–23.
52. Goecks J, Nekrutenko A, Taylor J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol*. 2010;11:R86.
53. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7:335–6.
54. Schloss PD, Westcott SL, Ryabin T, et al. Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol*. 2009;75(23):7537–41.
55. Hong C, Manimaran S, Shen Y, et al. PathoScope 2.0: a complete computational framework for strain identification in environmental or clinical sequencing samples. *Microbiome*. 2014;2(1):33.
56. Chen W, Zhang CK, Cheng Y, Zhang S, Zhao H. A comparison of methods for clustering 16s rRNA sequences into OTUs. *PLoS One*. 2013;8(8):e70837.
57. Mande SS, Mohammed MH, Ghosh TS. Classification of metagenomic sequences: methods and challenges. *Brief Bioinform*. 2012;13(6):669–81.
58. Colwell R. *Estimates, Version 7.5: Statistical Estimation of Species Richness and Shared Species from Samples (Software and Users Guide)*; 2005. Available at: http://viceroy.eeb.uconn.edu/estimates
59. Colwell RK. Biodiversity: concepts, patterns, and measurement. In: Levin SA, ed. *The Princeton Guide to Ecology*. Princeton University Press, Princeton, NJ;2009:257–63.
60. Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a web browser. *BMC Bioinformatics*. 2011;12(1):385.
61. Huse SM, Welch DBM, Voorhis A, et al. VAMPS: a website for visualization and analysis of microbial population structures. *BMC Bioinformatics*. 2014;15(1):41.
62. Wu M, Scott AJ. Phylogenomic analysis of bacterial and archaeal sequences with amphora2. *Bioinformatics*. 2012;28(7):1033–4.
63. Matsen F IV, Evans SN. Edge principal components and squash clustering: using the special structure of phylogenetic placement data for sample comparison. *PLoS One*. 2013;8:3.
64. Brady A, Salzberg SL. Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. *Nat Methods*. 2009;6:673–6.
65. Brady A, Salzberg S. Phymmbl expanded: confidence scores, custom databases, parallelization and more. *Nat Methods*. 2011;8(5):367–7.
66. Matsen FA, Kodner RB, Armbrust EV. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*. 2010;11(1):538.
67. 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(1):386.
68. Stark M, Berger SA, Stamatakis A, von Mering C. MItreemap-accurate maximum likelihood placement of environmental DNA sequences into taxonomic and functional reference phylogenies. *BMC Genomics*. 2010;11(1):461.
69. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. *Nat Genet*. 2000;25(1):25–9.
70. Gene Ontology Consortium. Gene ontology consortium: going forward. *Nucleic Acids Res*. 2015;43(D1):D1049–56.
71. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28:27–30.
72. Kotera M, Moriya Y, Tokimatsu T, Goto S. Kegg and genomenet, new developments, metagenomic analysis. In: Nelson KE, ed. *Encyclopedia of Metagenomics*. Springer, New York;2015:329–39.
73. Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. *Science*. 1997;278:631–7.
74. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The cog database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res*. 2000;28(1):33–6.
75. Abubucker S, Segata N, Goll J, et al. Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput Biol*. 2012;8(6):e1002358.
76. Markowitz VM, Chen I-MM, Palaniappan K, et al. IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res*. 2012;40:D115–22.
77. Seshadri R, Kravitz SA, Smarr L, et al. Camera: a community resource for metagenomics. *PLoS Biol*. 2007;5(3):e75.
78. Goll J, Rusch DB, Tanenbaum DM, et al. METAREP: JCVI metagenomics reports – an open source tool for high-performance comparative metagenomics. *Bioinformatics*. 2010;26(20):2631–2.
79. Huson DH, Mitra S, Ruscheweyh H-J, Schuster SC. Integrative analysis of environmental sequences using MEGAN4. *Genome Res*. 2011;21(9):1552–60.
80. Lingner T, Aßhauer KP, Schreiber F, Meinicke P. Comet – a web server for comparative functional profiling of metagenomes. *Nucleic Acids Res*. 2011;39(Web Server issue):W518–23.
81. Langille MG, Zaneveld J, Caporaso JG, et al. Predictive functional profiling of microbial communities using 16 s rrna marker gene sequences. *Nat Biotechnol*. 2013;31(9):814–21.
82. Eisen J. Environmental shotgun sequencing: its potential and challenges for studying the hidden world of microbes. *PLoS Biol*. 2007;5(3):e82.
83. Moran MA. Metatranscriptomics: eavesdropping on complex microbial communities. *Microbiome*. 2009;4(7):329–34.
84. Poretsky RS, Bano N, Buchan A, et al. Analysis of microbial gene transcripts in environmental samples. *Appl Environ Microbiol*. 2005;71(7):4121–6.
85. Botero LM, D'imperio S, Burr M, McDermott TR, Young M, Hassett DJ. Poly (a) polymerase modification and reverse transcriptase PCR amplification of environmental RNA. *Appl Environ Microbiol*. 2005;71(3):1267–75.
86. Carvalhais LC, Dennis PG, Tyson GW, Schenk PM. Application of metatranscriptomics to soil environments. *J Microbiol Methods*. 2012;91(2):246–51.
87. Gilbert JA, Field D, Huang Y, et al. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PLoS One*. 2008;3(8):e3042.
88. Leimena MM, Ramiro-Garcia J, Davids M, et al. A comprehensive metatranscriptome analysis pipeline and its validation using human small intestine microbiota datasets. *BMC Genomics*. 2013;14(1):530.
89. Yost S, Duran-Pinedo AE, Teles R, Krishnan K, Frias-Lopez J. Functional signatures of oral dysbiosis during periodontitis progression revealed by microbial metatranscriptome analysis. *Genome Med*. 2015;7(1):27.
90. Duran-Pinedo AE, Chen T, Teles R, et al. Community-wide transcriptome of the oral microbiome in subjects with and without periodontitis. *ISME J*. 2014;8(8):1659–72.
91. Jorth P, Turner KH, Gumus P, Nizam N, Buduneli N, Whiteley M. Metatranscriptomics of the human oral microbiome during health and disease. *MBio*. 2014;5(2):e1012–4.
92. Xiong X, Frank DN, Robertson CE, et al. Generation and analysis of a mouse intestinal meta-transcriptome through illumina based RNA-sequencing. *PLoS One*. 2012;7(4):e36009.
93. Dumont MG, Pommerenke B, Casper P. Using stable isotope probing to obtain a targeted metatranscriptome of aerobic methanotrophs in lake sediment. *Environ Microbiol Rep*. 2013;5(5):757–64.
94. Celaj A, Markle J, Danska J, Parkinson J. Comparison of assembly algorithms for improving rate of metatranscriptomic functional annotation. *Microbiome*. 2014;2(1):39.
95. Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol*. 2011;29(7):644–52.

96. Namiki T, Hachiya T, Tanaka H, Sakakibara Y. Metavelvet: an extension of velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res*. 2012;40(20):e155.
97. Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*. 2012;28(8):1086–92.
98. Birol I, Jackman SD, Nielsen CB, et al. De novo transcriptome assembly with abyss. *Bioinformatics*. 2009;25(21):2872–7.
99. Li R, Zhu H, Ruan J, et al. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res*. 2010;20:265–72.
100. Robertson G, Schein J, Chiu R, et al. De novo assembly and analysis of RNA-seq data. *Nat Methods*. 2010;7(11):909–12.
101. Guttman M, Garber M, Levin JZ, et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincrnas. *Nat Biotechnol*. 2010;28(5):503–10.
102. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28(5):511–5.
103. Li B, Dewey CN. Rsem: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12(1):323.
104. Haas BJ, Papanicolaou A, Yassour M, et al. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat Protoc*. 2013;8(8):1494–512.
105. De Bona F, Ossowski S, Schneeberger K, Rätsch G. Optimal spliced alignments of short sequence reads. *BMC Bioinformatics*. 2008;9(suppl 10):i174–80.
106. Cao HX, Schmutzer T, Scholz U, Pecinka A, Schubert I, Vu GTH. Metatranscriptome analysis reveals host-microbiome interactions in traps of carnivorous genlisea species. *Front Microbiol*. 2015;6:526.
107. Peano C, Pietrelli A, Consolandi C, et al. An efficient rRNA removal method for RNA sequencing in GC-rich bacteria. *Microb Inform Exp*. 2013;3(1):1.
108. Perez-Losada M, Castro-Nallar E, Bendall ML, Freishtat RJ, Crandall KA. Dual transcriptomic profiling of host and microbiota during health and disease in pediatric asthma. *PLoS One*. 2015;10:e0131819.
109. Fiehn O. Metabolomics – the link between genotypes and phenotypes. *Plant Mol Biol*. 2002;48(1–2):155–71.
110. Bernini P, Bertini I, Luchinat C, et al. Individual human phenotypes in metabolic space and time. *J Proteome Res*. 2009;8(9):4264–71.
111. Krumsiek J, Mittelstrass K, Do KT, et al. Gender-specific pathway differences in the human serum metabolome. *Metabolomics*. 2015;11(6):1815–33.
112. Mastrangelo A, Armitage EG, García A, Barbas C. Metabolomics as a tool for drug discovery and personalised medicine. A review. *Curr Top Med Chem*. 2014;14(23):2627–36.
113. Xu J, Mahowald MA, Ley RE, et al. Evolution of symbiotic bacteria in the distal human intestine. *PLoS Biol*. 2007;5(7):e156.
114. Manor O, Levy R, Borenstein E. Mapping the inner workings of the microbiome: genomic-and metagenomic-based study of metabolism and metabolic interactions in the human microbiome. *Cell Metab*. 2014;20(5):742–52.
115. Wu GD, Compher C, Chen EZ, et al. Comparative metabolomics in vegans and omnivores reveal constraints on diet-dependent gut microbiota metabolite production. *Gut*. 2014;65(1):63–72.
116. Lankadurai BP, Nagato EG, Simpson MJ. Environmental metabolomics: an emerging approach to study organism responses to environmental stressors. *Environ Rev*. 2013;21(3):180–205.
117. Kimes NE, Callaghan AV, Aktas DF, et al. Metagenomic analysis and metabolite profiling of deep-sea sediments from the gulf of Mexico following the deepwater horizon oil spill. *Front Microbiol*. 2013;4:50.
118. Bassler BL, Greenberg EP, Stevens AM. Cross-species induction of luminescence in the quorum-sensing bacterium *Vibrio harveyi*. *J Bacteriol*. 1997;179(12):1943–5.
119. Miller MB, Bassler BL. Quorum sensing in bacteria. *Ann Rev Microbiol*. 2001;55(1):165–99.
120. Bassler BL. Small talk: cell-to-cell communication in bacteria. *Cell*. 2002;109(4):421–4.
121. Henke JM, Bassler BL. Three parallel quorum-sensing systems regulate gene expression in *Vibrio harveyi*. *J Bacteriol*. 2004;186(20):6902–14.
122. Waters CM, Bassler BL. Quorum sensing: cell-to-cell communication in bacteria. *Annu Rev Cell Dev Biol*. 2005;21:319–46.
123. Camilli A, Bassler BL. Bacterial small-molecule signaling pathways. *Science*. 2006;311(5764):1113–6.
124. Aldridge BB, Rhee KY. Microbial metabolomics: innovation, application, insight. *Curr Opin Microbiol*. 2014;19:90–6.
125. Smolinska A, Blanchet L, Buydens LM, Wijmenga SS. NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review. *Anal Chim Acta*. 2012;750:82–97.
126. Wishart DS, Tzur D, Knox C, et al. HMDB: the human metabolome database. *Nucleic Acids Res*. 2007;35(suppl 1):D521–6.
127. Wishart DS, Knox C, Guo AC, et al. HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res*. 2009;37(suppl 1):D603–10.
128. Wishart DS, Jewison T, Guo AC, et al. HMDB 3.0-the human metabolome database in 2013. *Nucleic Acids Res*. 2012;41(Database issue):D801–7.
129. Ulrich EL, Akutsu H, Doreleijers JF, et al. Biomagresbank. *Nucleic Acids Res*. 2008;36(suppl 1):D402–8.
130. Cui Q, Lewis IA, Hegeman AD, et al. Metabolite identification via the Madison metabolomics consortium database. *Nat Biotechnol*. 2008;26(2):162–4.
131. Horai H, Arita M, Kanaya S, et al. Massbank: a public repository for sharing mass spectral data for life sciences. *J Mass Spectrom*. 2010;45(7):703–14.
132. Kopka J, Schauer N, Krueger S, et al. Gmdcsb.db: the Golm metabolome database. *Bioinformatics*. 2005;21(8):1635–8.
133. Smith CA, O'Maille G, Want EJ, et al. Metlin: a metabolite mass spectral database. *Ther Drug Monit*. 2005;27(6):747–51.
134. Reigstad CS, Kashyap PC. Beyond phylotyping: understanding the impact of gut microbiota on host biology. *Neurogastroenterol Motil*. 2013;25(5):358–72.
135. Aw W, Fukuda S. Toward the comprehensive understanding of the gut ecosystem via metabolomics-based integrated omics approach. *Semin Immunopathol*. 2015;37(1):5–16.
136. Mason OU, Hazen TC, Borglin S, et al. Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to deepwater horizon oil spill. *ISME J*. 2012;6(9):1715–27.
137. McNulty NP, Yatsunenko T, Hsiao A, et al. The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. *Sci Transl Med*. 2011;3(106):106ra106.
138. Maurice CF, Haiser HJ, Turnbaugh PJ. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. *Cell*. 2013;152(1):39–50.
139. Verberkmoes NC, Russell AL, Shah M, et al. Shotgun metaproteomics of the human distal gut microbiota. *ISME J*. 2009;3(2):179–89.
140. Weir TL, Manter DK, Sheflin AM, Barnett BA, Heuberger AL, Ryan EP. Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults. *PLoS One*. 2013;8(8):e70803.
141. Wang Z, Klipfell E, Bennett BJ, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*. 2011;472(7341):57–63.
142. Koeth RA, Wang Z, Levison BS, et al. Intestinal microbiota metabolism of l-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat Med*. 2013;19(5):576–85.
143. Kaddurah-Daouk R, Baillie RA, Zhu H, et al. Enteric microbiome metabolites correlate with response to simvastatin treatment. *PLoS One*. 2011;6(10):e25482.
144. Haiser HJ, Gootenberg DB, Chatman K, Sirasani G, Balskus EP, Turnbaugh PJ. Predicting and manipulating cardiac drug inactivation by the human gut bacterium *Eggerthella lenta*. *Science*. 2013;341(6143):295–8.
145. Franzosa EA, Morgan XC, Segata N, et al. Relating the metatranscriptome and metagenome of the human gut. *Proc Natl Acad Sci*. 2014;111(22):E2329–38.
146. Shi Y, Tyson GW, Eppley JM, DeLong EF. Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *ISME J*. 2011;5(6):999–1013.
147. Turnbaugh PJ, Gordon JI. An invitation to the marriage of metagenomics and metabolomics. *Cell*. 2008;134(5):708–13.
148. Lu K, Abo RP, Schlieper KA, et al. Arsenic exposure perturbs the gut microbiome and its metabolic profile in mice: an integrated metagenomics and metabolomics analysis. *Environ Health Perspect*. 2014;122(3):284–91.
149. Zhang Y, Zhao F, Deng Y, Zhao Y, Ren H. Metagenomic and metabolomic analysis of the toxic effects of trichloroacetamide-induced gut microbiome and urine metabolome perturbations in mice. *J Proteome Res*. 2015;14(4):1752–61.
150. Narayanasamy S, Muller EE, Sheik AR, Wilmes P. Integrated omics for the identification of key functionalities in biological wastewater treatment microbial communities. *Microb Biotechnol*. 2015;8(3):363–8.
151. Muller EE, Glaab E, May P, Vlassis N, Wilmes P. Condensing the omics fog of microbial communities. *Trends Microbiol*. 2013;21(7):325–33.
152. Abram F. Systems-based approaches to unravel multi-species microbial community functioning. *Comput Struct Biotechnol J*. 2015;13:24–32.
153. Levy R, Borenstein E. Reverse ecology: from systems to environments and back. In: Soyer OS, ed. *Evolutionary Systems Biology*. Springer, New York;2012:329–45.
154. Borenstein E, Kupiec M, Feldman MW, Ruppin E. Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proc Natl Acad Sci*. 2008;105(38):14482–7.
155. Ebenhöh O, Handorf T, Heinrich R. Structural analysis of expanding metabolic networks. *Genome Inform*. 2004;15(1):35–45.
156. Bachmaier C, Brandes U, Schreiber F. Chapter 20: Biological networks. In: Tamassia R, ed. *Handbook of Graph Drawing and Visualization*. CRC Press, Boca Raton, FL; 2013:621–51.
157. Wuchty S, Ravasz E, Barabási A-L. The architecture of biological networks. In: Deisboek TS, Kresh JY, eds. *Complex Systems Science in Biomedicine*. Springer, New York; 2006:165–81.
158. Barabási A-L, Oltvai ZN, Wuchty S. Characteristics of biological networks. In: Ben-Naim E, Frauenfelder H, Tonoczkai Z, eds. *Complex Networks*. Springer-Verlag, Berlin; 2004:443–57.

159. Pawson T, Nash P. Protein-protein interactions define specificity in signal transduction. *Genes Dev*. 2000;9:1027–47.
160. Dutkowski J, Kramer M, Surma MA, et al. A gene ontology inferred from molecular networks. *Nat Biotechnol*. 2013;31:38–45.
161. Demchak B, Hull T, Reich M, et al. Cytoscape: the network visualization tool for genomespace workflows. *F1000Res*. 2014;2014(3):151–63.
162. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. *PLoS Comput Biol*. 2012;8(9):e1002687.
163. Srivas R, Hannum G, Ruscheinski J, et al. Assembling global maps of cellular function through integrative analysis of physical and genetic networks. *Nat Protoc*. 2011;6(9):1308–23.
164. Amar D, Shamir R. Constructing module maps for integrated analysis of heterogeneous biological networks. *Nucleic Acids Res*. 2014;42(7):4208–19.
165. van Dam JC, Schaap PJ, dos Santos VAM, Suárez-Diez M. Integration of heterogeneous molecular networks to unravel gene-regulation in mycobacterium tuberculosis. *BMC Syst Biol*. 2014;8(1):111.