

journal homepage: [www.elsevier.com/locate/csbj](http://www.elsevier.com/locate/csbj)

Mini Review

# Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome

Shirley Bikel <sup>a,c</sup>, Alejandra Valdez-Lara <sup>a,c</sup>, Fernanda Cornejo-Granados <sup>a,c</sup>, Karina Rico <sup>a</sup>, Samuel Canizales-Quinteros <sup>a</sup>, Xavier Soberón <sup>b</sup>, Luis Del Pozo-Yauner <sup>b</sup>, Adrián Ochoa-Leyva <sup>a,c,\*</sup>

<sup>a</sup> Unidad de Genómica de Poblaciones Aplicada la Salud, Facultad de Química, UNAM, Instituto Nacional de Medicina Genómica (INMEGEN), México, D.F. 14610, Mexico

<sup>b</sup> Instituto Nacional de Medicina Genómica (INMEGEN), México, D.F., Mexico

<sup>c</sup> Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad 2001, Cuernavaca C.P. 62210, Mexico

ARTICLE INFO

Article history:

Received 11 March 2015  
 Received in revised form 1 June 2015  
 Accepted 4 June 2015  
 Available online 9 June 2015

Keywords:

Metagenomics  
 Metatranscriptomics  
 Viromics  
 Human microbiome  
 Systems-level  
 Bioinformatics

ABSTRACT

The advances in experimental methods and the development of high performance bioinformatic tools have substantially improved our understanding of microbial communities associated with human niches. Many studies have documented that changes in microbial abundance and composition of the human microbiome is associated with human health and diseased state. The majority of research on human microbiome is typically focused in the analysis of one level of biological information, i.e., metagenomics or metatranscriptomics. In this review, we describe some of the different experimental and bioinformatic strategies applied to analyze the 16S rRNA gene profiling and shotgun sequencing data of the human microbiome. We also discuss how some of the recent insights in the combination of metagenomics, metatranscriptomics and viromics can provide more detailed description on the interactions between microorganisms and viruses in oral and gut microbiomes. Recent studies on viromics have begun to gain importance due to the potential involvement of viruses in microbial dysbiosis. In addition, metatranscriptomic combined with metagenomic analysis have shown that a substantial fraction of microbial transcripts can be differentially regulated relative to their microbial genomic abundances. Thus, understanding the molecular interactions in the microbiome using the combination of metagenomics, metatranscriptomics and viromics is one of the main challenges towards a system level understanding of human microbiome.

© 2015 Bikel et al. Published by Elsevier B.V. on behalf of the Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Contents

1. Introduction	391
2. Sequencing and bioinformatic strategies to study the human microbiome	391
2.1. 16S rRNA gene profile analysis	391
2.2. Metagenomic analysis	393
2.3. Metatranscriptomic analysis	394
2.4. Viromic analysis	395
3. Characteristics of oral and gut microbiomes	395
3.1. Metagenomic and 16S rRNA profiling combined with viromic analyses	396
3.2. Metagenomic combined with metatranscriptomic analyses	397
4. Perspectives and future trends of the human microbiome analyses	398
Acknowledgments	398
References	398

\* Corresponding author at: Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad 2001, Cuernavaca C.P. 62210, Mexico.

E-mail addresses: [adriano8a@gmail.com](mailto:adriano8a@gmail.com), [aochoa@ibt.unam.mx](mailto:aochoa@ibt.unam.mx) (A. Ochoa-Leyva).

## 1. Introduction

The human body is inhabited by a high diversity of bacteria and archaea, as well as fungi, protozoa and viruses. These microbes inhabit in several niches within the human body sections and are collectively known as the human microbiota, whereas their collective genomes form the human metagenome [1]. The advance in high-throughput technologies used to analyze the components of the microbiome have substantially improved our knowledge of the microbial communities associated to human niches [2,3]. The decrease in cost of sequencing using high-throughput technologies has enabled large-scale studies of the human microbiome, revealing high interindividual variability of the microbiota composition and in different body sites within individuals [4–7].

Changes in abundance and composition in the fecal microbiota (dysbiosis) has been observed in patients with several human diseases, ranging from inflammatory bowel disease and obesity to diabetes and neurological disorders [8–17]. The fundamental objectives of human microbiome research is to study the structure and dynamics of microbial communities, the relationships between their members (microorganisms, viral particles and host) and their potential association with health and disease. The study of interactions between the DNAs, RNAs, and viruses that are present in the microbiome, are the main interest of metagenomics, metatranscriptomics, and viromics, respectively. For studying the microbial community of the human microbiome using high throughput sequencing technologies, there are several types of large scale analyses: the 16S profiling analysis which is based in sequencing the hypervariable regions of the 16S rRNA gene and the shotgun analysis which is based in direct sequencing of the total DNA (metagenome) and/or total RNA (metatranscriptome). In addition, the viral component of the microbiome (virome) can be also analyzed by sequencing the total viral particles.

In the last decade, many studies using the sequencing of the 16S rRNA gene to characterize the microbiota composition have been conducted, however, this analysis mainly identifies the abundance and diversity of bacteria and archaea in the sample. Although, there is a computational approach to predict the metagenome functional composition by the 16S rRNA gene sequences [18]. The metagenomic analysis also identifies the abundance and diversity of microbial community, but additionally can identify the gene content and inferred functional potential of proteins encoded in the genomes of the microbial community. The metatranscriptomic analysis allows the identification of expressed transcripts in the microbiome. The transcript numbers can also be used to compare the gene expression profiles between microbial communities. In addition, for comparative study, metatranscriptomic data must be paired with metagenomic data in order to analyze if the transcript abundance is reflecting changes in community composition [19,20]. The use of high throughput sequencing technologies to analyze human metagenomes has also revealed the existence of many bacteriophages in metagenomes [21]. Interestingly, it has been proposed that bacteriophages may have a role in shaping the diversity and composition of the oral and gut bacteria [22,23]. The involvement of phages in microbial dysbiosis may indirectly contribute to the disease. In this regard, a model suggesting that virome may contribute to the intestinal inflammation and bacterial dysbiosis was recently reported in the human gut microbiome [24]. However, studies that involve metagenomic or metatranscriptomic combined with viromic analysis are still necessary to understand the molecular interactions within the human microbiome and their relevance in health and diseased states.

The microbiome has been conceptualized as a dynamic ecological community consisting of multiple taxa each potentially interacting with each other, the host and the environment [25]. Hereafter, we use the term microbiome, to refer to the microbial communities and viruses in conjunction with the environment they inhabit, interacting as a system. In the first section of this review, an overview of the different sequencing, experimental, and bioinformatic procedures that have

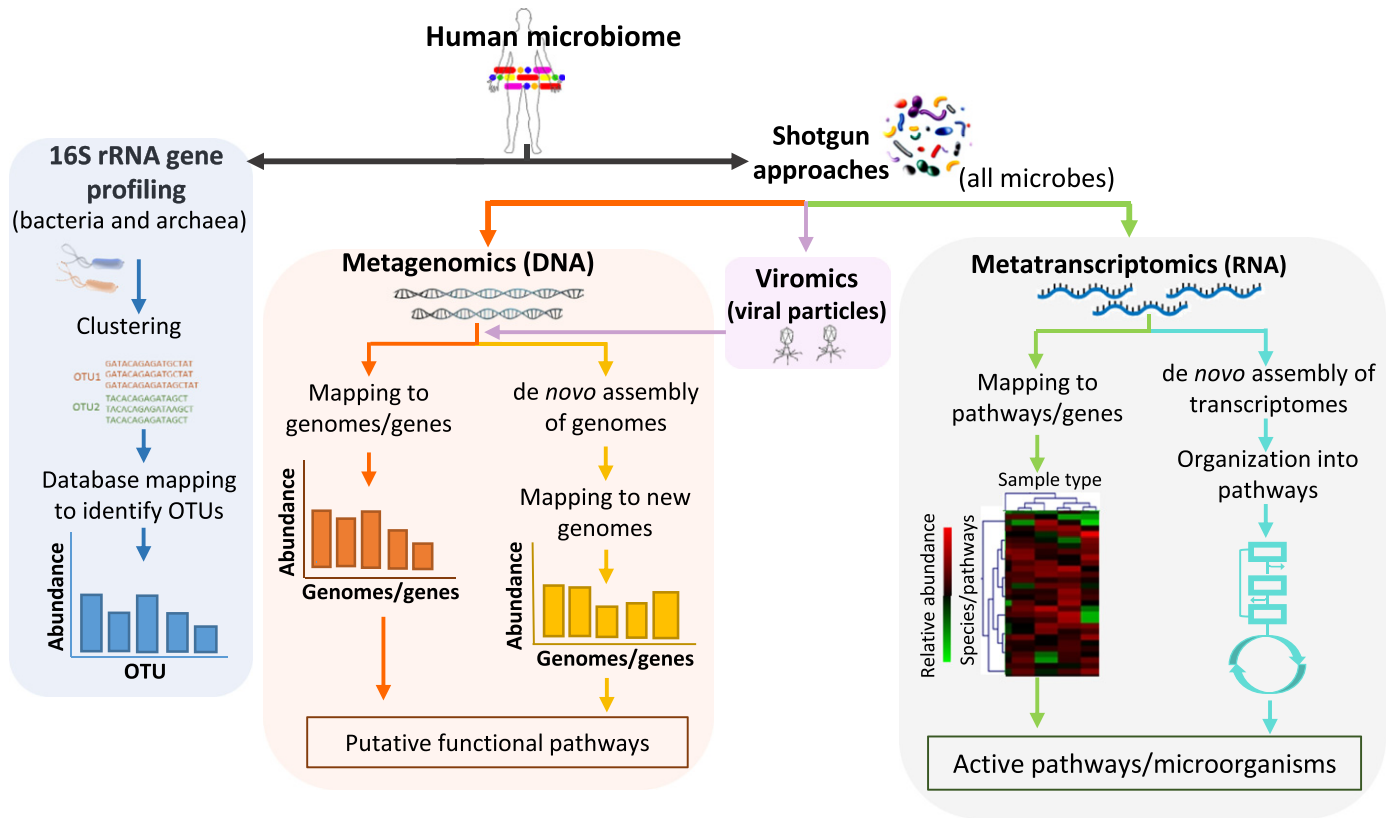
been used to study the human microbiome are discussed. In the second section, the recent advances combining metagenomic, metatranscriptomic and viromic analyses to identify the molecular dialog within the microbiome are discussed. The metabolomic and metaproteomic analyses are not in the focus of this review.

## 2. Sequencing and bioinformatic strategies to study the human microbiome

### 2.1. 16S rRNA gene profile analysis

The small ribosome subunit 16S gene (16S rRNA gene) is used as a housekeeping genetic marker to study bacterial phylogeny and taxonomy as it is highly conserved between different species of bacteria and archaea. In addition to highly conserved regions, the 16S rRNA gene contains hypervariable regions that are used to identify between different bacteria. Furthermore, some bacteria have a different copy number of the 16S rRNA gene, often existing as a multigene family, or operons. Hence, the 16S rRNA gene sequencing has become typically used to identify and quantify bacterial taxa present within a microbiome sample. 16S rRNA profiling relies on using PCR 'universal' primers targeted at the conserved regions and designed to amplify a range of different microorganism as wide as possible. The amplified fragments (amplicons) of the gene correspond to selected short-hypervariable regions ranging from V1 to V9, making it faster and cheaper to sequence with high throughput technologies than many other bacteria genes (Figs. 1 and 2). Two of the most significant limitations of 16S rRNA sequencing that should be considered before starting a sequencing project are: (1) the introduction of biases by selection of the 16S rRNA hypervariable regions and (2) the introduction of biases by PCR primer design, which may select for or against particular groups of microorganisms (Fig. 2). To minimize the biases introduced by primer design, the primers include degenerated bases and can be used at lower hybridization temperatures to capture more microbial diversity. Other problems using the PCR is that bacterial contamination of reagents may be affecting the results [26] and that the 16S rRNA gene is also present in different copy numbers in bacterial genomes influencing the apparent relative abundance of a microorganism [26].

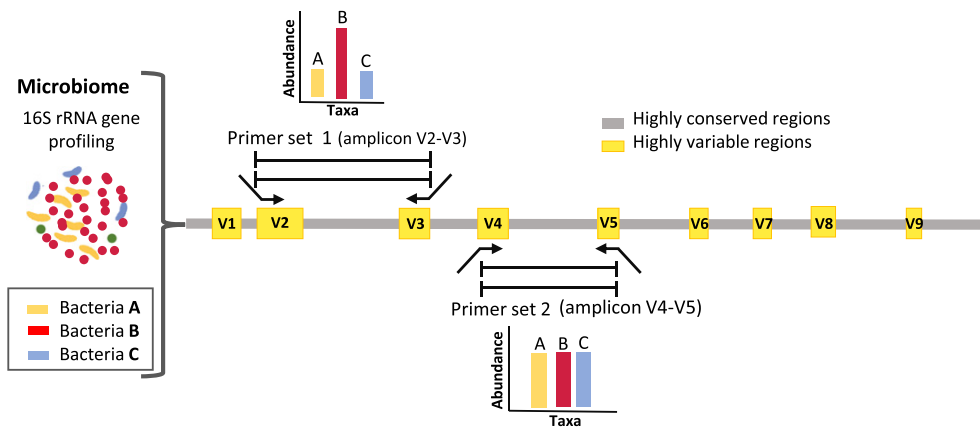
The sequence fragments obtained by high throughput sequencing technologies are typically called sequence reads. Longer read lengths (1000 bp), as the ones obtained by 454/Roche technology, can span multiple hypervariable regions of the 16S rRNA gene, increasing the number of the microorganisms that can be identified at species level. Although, this technology is cost prohibitive and it will be discontinued in 2016. However, short-read length sequences spanning only one hypervariable region has sufficient resolution for the accurate taxonomic assignments [27,28]. The optimal community clustering confirmed with sequence reads of this length is an important advance in amplicon design because sequencing only one hypervariable region is more cost-effective (Table 1). There is a large amount of PCR primers to amplify different hypervariable regions of 16S rRNA gene for sequencing in the short read sequencing platforms. Although, the compatibility of the fragment length should be according with sequencing platform read length capacity. There are many studies that target different regions of the 16S rRNA gene, for example V3–V5 [29], V1–V2 [30], V1–V3 [26], V4–V5 [31], and V8–V9 [32]. There is an active discussion about the hypervariable region that should be sequenced to perform a microbial diversity analysis. For example, the V6 region is not optimal for sequencing analyses that are directed for taxonomic assignment and community clustering, as opposed to sequence reads spanning the V2 and V3 regions [33,2]. The most informative 16S rRNA gene region to amplify may also depend of the analyzed environment, for example, in a study for the diagnosis of pathogenic bacteria Chakravorty et al., showed that in a mix of 110 different bacterial species including common blood borne pathogens, CDC-defined agents and environmental microflora, the V2 and V3 were most suitable for distinguishing all



**Fig. 1.** Different sequencing and bioinformatic strategies for human microbiome analysis. In the 16S rRNA gene profiling the raw sequences obtained are passed through quality filters to minimize the presence of sequencing artifacts. The resulting filtered sequence reads are clustered into operational taxonomic units (OTUs), which represent similar organisms. After that, taxonomic identity is assigned for each OTU based in sequence homology against known 16S rRNA gene databases and the relative abundance of each OTU is calculated for each sample. The resulting OTUs table is also used for quantifying population diversity within and between the samples, as the alpha and beta diversity measurements, respectively. In the shotgun approaches, metagenomic, metatranscriptomic and viromic analyses are performed. In the metagenomic analysis, the DNA sequences obtained can either be mapped to reference genomes/genes or used for *de novo* assembly of genomes. Then the relative abundance of the present genomes/genes and the functional potential of the sequences can be assessed using functional annotated databases. In viromics analysis, first the viral particles (VPs) must be enriched and posteriorly sequenced to obtain the virus genomes. Furthermore, to analyze the active genes and species of the microbiome, the metatranscriptomic analysis is applied and the obtained RNA sequences are mapped to reference pathways and genes. The results are used to identify the active pathways, genes and microorganisms. Thus, the relative abundance of each active pathway/gene/microorganism in the human microbiome is determined. The *de novo* assembly of genomes and transcriptomes can be also performed to identify novel genomes or pathways.

bacterial species to the genus level except for closely related enterobacteriaceae [34]. Additionally, the V2 region was best distinguished among *Mycobacterium* species and V3 among *Haemophilus* species [34]. Another study suggests that sequencing of the V2 and V3 regions

performs well for both community clustering and taxonomic assignments in a wide range of different samples (i.e. the mouse and human gut and in the hypersaline microbial mat from Guerrero Negro). On the other hand, the Vaginal Human Microbiome Project has validated



**Fig. 2.** Importance of primer selection for the amplification of the hypervariable regions of the 16S rRNA gene. The figure illustrates how choosing different sets of primers for the amplification of different hypervariable regions of the 16S rRNA gene has an influence in the resulting abundance of hypothetical bacteria A, B and C. For example, in this figure, the species abundance distribution obtained using primer set 1 shows a more similar distribution to that observed in the microbiome than the abundance obtained from primer set 2. In a similar manner, Kuczynski et al [2], demonstrated that using the universal primer set F515–R806 (which is typically used to amplify a great coverage of bacteria and archaea) in skin samples showed poor results for the identification of *Propionibacterium*, however the use of primer set F27–R338 was better to identify this bacteria [2].

**Table 1**

Typical high throughput sequencing platforms used in 16S rRNA gene profiling and shotgun sequencing approaches.

Sequence read length	Hypervariable regions that can be evaluated	Shotgun utility	Costs	Sequence reads per run	Run time
Platform: Roche 454 GS-FLX + 800 base reads	Up to seven per read; long reads allow a good coverage of 16S rRNA gene, allowing a good taxonomical assignment.	Long reads help with assembly of new genomes.	Cost limits the deep sequencing analysis.	1 million	20 h
Platform: Illumina HiSeq, MiSeq and NextSeq 100–500 base reads	Only one per read (NextSeq) Up to 3 per read (HiSeq and MiSeq); short reads do not seem to limit the taxonomical analysis.	Short reads, but high sequencing output allows a deep sequencing analysis.	Cost-effective deep sampling	25 million, 2 × 300 bp (MiSeq) 130–400 million, 2 × 150 bp (NextSeq) 2.5 billion, 2 × 150 bp (HiSeq)	5–55 h (MiSeq) 12–30 h (NextSeq) 24–84 h (HiSeq)
Platform: Life Technologies ion personal genome machine (PGM) 35–400 base reads	Up to 3 in custom design. Up to seven using the Ion 16S metagenomics kit; in a very fast sequencing time.	Short reads, but mid sequencing output in a very fast sequencing time.	Time-effective with a mid-cost for deep sequencing analysis.	1 × 200 or 1 × 400 and number of reads depends on model of the sequencing chip: 10 million (314) 100 million (316) 1 billion (318)	2–4 h (314) 3–5 h (316) 4–7 h (318)

a protocol that provides species-level classification of V1–V3 16S rRNA sequences from the vaginal microbiome [35]. The primer pair 515F/R806, targeting the V4 region, was highly recommended by several human microbiome studies [20,28,36,37] and a well-established illumina protocol has been reported for this primer set [36,37]. Although no consensus has been established, the V4 region has been suggested as the gold standard for human gut microbiota characterization by MetaHit consortium [6]. Furthermore, the Earth Microbiome Project also has demonstrated that the V4 region can be extensively supported as the standard 16S rRNA region for general community assessment across a range of very different environments [38].

The most widely used bioinformatic pipelines to analyze the amplicons of the 16S rRNA gene, are QIIME [39] Mothur [40], MGRASP [41] and Galaxy [42], which are open source packages. The typical bioinformatic pipeline to analyze amplicon sequences involves three basic steps. First, the amplicon sequences are subjected to data filtering based on several quality filters. The typically used quality filters are read length, base quality (Phred score), ambiguous base calls, homopolymers, low complexity sequences and CG content. In addition, the adapter sequences should be eliminated and errors in barcodes should be corrected. Of those, it is critical to be concerned about barcode errors to avoid assigning sequence reads to the wrong sample. Second, the amplicon sequences are clustered into groups of related sequences based on their sequence similarity at a particular taxonomy level of interest (97% sequence identity is frequently chosen for species). The clusters of similar sequences are referred to as operational taxonomic units (OTUs) or sometimes phylotype, which provide a working name for groups of related bacteria (Fig. 1). OTU counts are summarized in a table of their relative abundances for each sample. Then the OTUs are compared against a reference ribosomal sequence database such as Greengenes [43], RDP [44] or SILVA [45] to assign the taxonomical classification. The third phase of analysis uses the resulting data for quantifying population diversity in the samples [46]. Within a microbial community, several measures including Shannon Index, Chao1 and Simpson's Diversity exist for calculating alpha diversity within sample. These give rise to plots of alpha diversity versus simulating sequencing effort, known as rarefaction curves. Additionally, when comparing multiple populations, beta diversity measures are applied to describe how many taxa are shared between them. In this regard, UniFrac is a beta diversity measure that uses phylogenetic information to compare the taxa shared between multiple samples and when it is coupled with standard multivariate statistical techniques including principal

coordinate analysis (PCoA), identifies factors explaining differences among microbial communities [27].

## 2.2. Metagenomic analysis

Metagenomics sequencing allows the determination of the functional potential codified in the microbiome. In addition, metagenomic analyses also have been used for the discovery of novel enzymatic functions, microorganisms and genes that may be used for bioremediation [47,48], for understanding the host-pathogens interactions [49] and for novel therapeutic strategies in human diseases [50]. One challenge in metagenomic analysis is addressing the presence of host DNA in samples. For example the Human Microbiome Project (HMP) has reported high levels of human DNA in different microbiome samples, such as mid-vagina, throat and saliva samples [51,52]. The amount of host DNA varies greatly by body site and sample type, for example the samples of sputum or lung tissue in cystic fibrosis usually contain a large amount of human DNA released by neutrophils during the immune response, sometimes representing even more than 99% of the total DNA [53–55]. As a result, only a small percentage of the sequence reads from such samples correspond to microbial genomes and consequently a large percentage of sequences are eliminated. Therefore, obtaining sufficient sequence coverage of the metagenomes can become cost prohibitive. In this regard, improved experimental methods for solving the host DNA problem are needed, which are not only limited to selective cell lysis. It is also important to note that even in the absence of host DNA; the metagenomic sequencing requires a high amount of sequence to get a reasonable coverage of the microbial genomes present in the sample. Contrarily, the 16S rRNA profiling only requires a little amount of sequence to get a reasonable taxonomical census of the microorganisms present in the sample; however, it misses out the determination of gene content. A typical metagenomic experiment involves the isolation of the total DNA from the microbiome followed by its fragmentation to smaller pieces of DNA (the fragment sizes in bp depend of the selected sequencing platform). After that, the 5' and/or 3' ends of DNA library are repaired and adapters (containing sequences to allow hybridization to a flow cell) are ligated. The final steps are library cleanup and amplification, followed by quantification, after which the library is finally ready for sequencing.

The typical bioinformatic pipeline to analyze the sequences obtained by shotgun analysis involves as the first step the data filtering as was previously explained for 16S rRNA profiling analysis. The sequence reads that

passed all quality controls are mapped to known genomes and can be also used for *de novo* assembly of contigs or genomes. However, no genomes are usually recovered for most species using metagenomic approaches. In the mapping strategy, the sequence reads are mapped (located) to a reference genome. There are powerful methods like BLAST and BLAT that are not specialized for the vast amount of data obtained by sequencing platforms. In this regard, different mapper algorithms have been developed for mapping the sequence reads to a reference genome in an efficient and productive manner, some of them are Bowtie [56], SMALT [57], BWA [58] and GEMmapper [59]. The metagenomic sequence reads are usually mapped to the human microbiome reference genome database of the Human Microbiome Project [51] and/or to specific genome databases like the Human Oral Microbiome Database (HOMD) [7] or the gut microbial gene catalog [6]. After mapping, the microbial abundance can be measured as the fraction of sequence reads that mapped to a single species in the database. Additional to the mapping strategies, there are several pipelines that compare metagenomic sequence reads against gene markers using BLAST [60], Usearch [61] or HMMs [62,63] to taxonomically annotate and quantify each metagenomic homologue. (Fig. 1). The sequence identity of the best match can be used to determine the most likely phylogenetic origin of the read. The functional diversity of the microbiome can be estimated by annotating metagenomic sequences with known functions. To this end, the sequence reads that contain protein coding genes are identified and their sequence is homology compared to the coding sequences of protein databases like the Kyoto Encyclopedia of Genes and Genomes (KEGG), protein family annotations (PFAM), gene ontologies (GO) and clusters of orthologous groups (COG). KEGG is a database resource that integrates genomic, chemical and systemic functional information [64] and PFAM is a large collection of protein families, each represented by multiple sequence alignments and hidden Markov models (HMMs) [65]. The COG database consists of clusters of orthologous groups (COG) of proteins found to be orthologous for at least three lineages and are classified into functional groups. The GO provides a controlled vocabulary of terms for describing gene product properties at three different levels: the cellular component, the molecular function and the biological process [66]. Hence, the function of the query sequence is assigned based on its similarity to sequences functionally annotated in all the above mentioned databases. The resulting data is used to describe the number of potential functions and their relative abundance in the metagenome. Furthermore, INFERNAL is a powerful tool that can be used to predict small RNA in the metagenomic data [67]. HUMAnN is an automated pipeline to determine the presence/absence and abundance of microbial pathways and gene families in a community directly from metagenomic sequence [68]. This pipeline, which is an offline platform, converts sequence reads into coverage and abundance tables summarizing the gene families and pathways in a microbial community [68]. Another offline platform used to analyze metagenomic data is the MEtaGenome Analyzer (MEGAN) [69]. Furthermore, there are integrated suites that have been designed to analyze metagenomic data sets online in an automated manner from metagenomic sequence, such as RAST (MG-RAST) [41], IMG/M server [70] and JCVI Metagenomics Reports (METAREP) [71].

In *de novo* assembly strategy, the total sequence reads are used to assemble genomes (Fig. 1). Additionally, the assembly can be also performed only using the sequence reads that were not mapped to known genomes (Fig. 1). There are several bioinformatics tools used for *de novo* assembly like MetaVelvet [72], khmer [73], metamos [74], Meta-IDBA [75], MetaORFA [76] and RayMeta [77]. Metagenomic assemblers generally adapt graph-based reconstruction approaches as the overlap–layout–consensus (OLC) to assemble longer sequences and de Bruijn graph to assemble shorter sequences. However, the assemblers based in the de Bruijn graphs are the most used due to the success of the shorter sequences produced by popular sequencing platforms of Illumina and Life Technologies companies (Table 1). After that, the new assembled genomes can be used for mapping the metagenomics sequence reads to estimate the abundance of these

new genomes in the microbiome. Furthermore, the new genomes are also used for functional annotation when compared against sequences annotated in databases as KEGG or GO (Fig. 1).

### 2.3. Metatranscriptomic analysis

Metagenomics is a powerful tool used to describe the gene content and potential functions encoded in sequenced genomes. However, metagenomics approaches have a very limited role in revealing the microbial activity measured by gene expression. The metatranscriptomic shotgun sequencing (RNAseq) provides the access to the metatranscriptome of the microbiome allowing the whole-genome analysis profiling of the active microbial community under different conditions. In this regard, sequencing of metatranscriptomes has been recently employed to identify RNA-based regulation and expressed biological signatures in human microbiome [78]. However, only few investigations (which are discussed in the second section of this review) have performed a combined analysis using metatranscriptomics with metagenomics. A typical metatranscriptomic experiment involves the isolation of the total RNA from the microbiome followed by RNA enrichment depending on the type of RNA to be sequenced (i.e. mRNA, lincRNA, and microRNA). After that, the RNA is fragmented to smaller pieces (the fragment sizes in bp depend of the selected sequencing platform) followed by cDNA synthesis using reverse transcriptase and random hexamers or oligo(dT) primers. After that, like in the construction of metagenomic libraries, the 5' and/or 3' ends of the cDNA are repaired and adapters are ligated, followed by library cleanup, amplification and quantification and finally the library is sequenced. As converting RNA into cDNA has been shown to introduce biases in quantification of transcripts [79], semi direct RNA sequencing without cDNA synthesis has been developed [80–82]. There are several technical issues affecting the large-scale application of metatranscriptomics: (1) the collection and storage procedures to preserve the RNA of the sample, (2) the limitation to obtain high-quality and sufficient quantity of RNA from human microbiome samples, (3) the mRNA enrichment procedures by removing ribosomal RNA (rRNAs) which represent over 90% of the RNA, (4) the average useful life of mRNA leads to difficulty in the detection of rapid and short-term responses to environmental changes, (5) the transcriptome databases are insufficient, (6) the host RNA contamination which cannot be removed by currently available rRNA purification methods and (7) the poly-A RNA selection kits to capture the mRNA population are not feasible in prokaryotes. However, several efforts have been recently made to tackle these technical issues, for example Franzosa et al. have developed experimental strategies to improve the point 1 [19] and Giannokus et al. [83] have analyzed strategies to improve points 2 and 3. Interestingly, the Ambion's MICROBEnrich Kit uses hybridization capture technology to remove human, mouse, and rat RNA (both mRNA and rRNA) from complex host-bacterial RNA populations, leaving behind enriched microbial total RNA.

The typical bioinformatics pipeline to analyze the data obtained from a metatranscriptomic experiment is similar to the one used in metagenomics and it is also divided in two strategies: (1) mapping sequence reads to reference genomes and genes and (2) *de novo* assembly of new transcriptomes (Fig. 1). In the first strategy, after mapping the RNA sequence reads to different genomes or pathways is possible to identify the taxonomical classification of active microorganism and the functionality of their expressed genes. For example, through the mapping of metatranscriptomic sequences to KEGG database [64], the pathways whose expressed genes are up and down regulated or unchanged in the microbiome during health and disease conditions are obtained [84]. There are several bioinformatic programs used in metagenomics like SOAPdenovo [85], ABySS [86] and Velvet-Oases [87] that have been reported to be successfully applied to the metatranscriptome assembly of microbiomes [86,88–91]. However, Trinity is a program specially developed for *de novo* transcriptome assembly from short-read RNA-seq data and it is a very efficient and

sensitive in recovering full-length transcripts and isoforms and now is one of the most used bioinformatics tools to assembly *de novo* transcriptomes of very different species [92–94].

#### 2.4. Viromic analysis

Viruses outnumber microbial cells 10:1 in most environments; however, viral DNA only represents 0.1% of the total DNA in a microbial community [6]. Hence, to obtain a deep sequence coverage of the human viruses that are present in the microbiome, the isolation of viral particles (VPs) becomes necessary [95–98]. In solid samples containing viruses, such as human feces, a common approach is to suspend the fecal material in an osmotic neutral buffer followed by serial filtration steps to remove large particles, including undigested or partially digested food fragments and microbial cells [99]. Additionally, other protocols use the ultracentrifugation with a cesium chloride density gradient to separate the viral particles from the microbial cells [100,101]. After that, viral particles are purified and the non-encapsulated free nucleic acids are removed by treatment with DNase and RNase, then the VP-derived nucleic acids are isolated. However, because the amount of DNA extracted from purified VPs is often below the required for sequencing, the amplification of the total viral DNA is typically necessary. To this end, a range of amplification methods have been developed, such as random amplified shotgun library (RASL) [102], linker-amplified shotgun library (LASL) [103] and the Multiple Displacement Amplification (MDA) [104]. The MDA method takes the advantage of the high processivity of the phage-derived  $\phi$ 29 polymerase which synthesize >70,000 nucleotides per association–dissociation cycle and its strong strand displacement capability, which together allow the amplification of complete viral genomes. Although, recent publications have shown that critical biases and contamination are introduced to the sample when the MDA amplification is used [105–109].

The sequencing technologies that prioritize long read lengths over those of short read lengths are preferred, because most viral sequences are novel and they are enriched in regions of low-complexity repeats [110,111]. However, the sequencing technologies of long read lengths as 454/Roche pyrosequencing are about to be discontinued. To resolve this limitation many bioinformatic programs have been developed to analyze viruses from short sequence reads [27]. The initial analysis of the sequences obtained after DNA sequencing of VPs also involves the data quality filtering like the one in the 16S rRNA profiling and metagenomic sections. However, the majority (usually, 60–99%) of sequences in viromes from any environment have no significant similarity to other sequences in databases or have higher homology to prokaryotic or eukaryotic genes [112,103,113–116], therefore, the filtering of bad quality sequences and the decontamination of 16S rRNA, 18S rRNA and human sequences by mapping is still important in viromic analysis. The resulting sequences are compared against individual viral genomes using several mapping algorithms or using programs as BLASTX or USEARCH (as were previously described in the metagenomic section), to analyze the taxonomic composition of viral community.

Although viral sequence databases have considerably expanded due to the start of the viromics era, the number of deposited genomes is far less than the expected number of virotypes [117] and most of the new sequences are poorly annotated [118]. Furthermore, the percentage of sequence reads with similarity to known viral sequences depends on how well the sequences have been filtered and the database that is used, however, it is generally less than 0.01% [119–121]. Although, the number of sequence reads with similarity against databases of viral genomes depends on how well the sequences have been filtered and the database that is used for comparison. There are several databases focused on taxonomical virus classification, such as the Classification of Mobile Genetic Elements (ACLAME) [122] and Phage SEED [123]. However, it is important to note that ACLAME database has not been updated in the last years. After taxonomic and functional assignments a viral community profile characterizing the diversity in the sample is created.

However, given that most of the available viral metagenomic data lacks similarity to sequences in the databases, similarity-independent methods have been developed to better understand viral community structure. One example of that is Phage Communities from Contig Spectrum (PHACS), which is a bioinformatic tool to assess the biodiversity of uncultured viral communities. PHACS was designed to quantify virotypes [112,124] based on the assumption that if a virotype is present in high abundance in a VP sample it is more likely to be assembled into a large contig. Another alternative for identifying shared viruses among different samples is crass [125], an algorithm that allows the simultaneous cross-assembly of all the samples in a data set as opposed to the pairwise assemblies used in MaxiPhi [114], which is based on contig assemblies generated from the pooled VP viromes. The chimeras are a common problem with most assemblers, although most occur between viruses and not so much between virus and bacteria. In this regard, the overlap–layout–consensus (OLC) algorithms have showed efficiency in the viral genomes assembly. One of the most popular of these assemblers is Newbler which has been extensively used in viral and bacterial shotgun metagenomic projects [6,95,126–131]. However, it remains to be determined if Newbler will be discontinued with the 454/Roche in 2016. Additionally, two other new OLC assemblers were developed and tested on viral metagenomics data: Minimo, designed for the assembly of small datasets [132] and used for virome analyzes [95,133]; and VICUNA, an assembler specialized in *de novo* assembly of data from heterogeneous viral populations [134]. De Bruijn graph assemblers, as MetaVelvet [72], are an alternative strategy to the OLC assemblers and have also been used on the assembly of viral metagenomes [135]. Other popular metagenome assemblers for viromes are IBDA [136] and RayMeta [77]. The RNA viruses in the human gut microbiome are mostly influenced by RNA from ingested plants of the food [89], therefore, the sequencing of total RNA viruses is impractical. However, it remains to be seen whether they represent stable members of the gut virome or are transiently present as a result of plant consumption [89]. Finally, in the Table 1 are summarized the most common high throughput sequencing technologies used in 16S rRNA profiling and shotgun approaches.

### 3. Characteristics of oral and gut microbiomes

The oral and gut microbiomes represent the two best-studied human microbiomes to date. The human gastrointestinal tract involves an extremely complex and dynamic microbial community, that includes archaea, bacteria, viruses and eukaryote [3]. However, most of the microorganisms that inhabit the gastrointestinal tract are bacteria and 70% of them inhabit the colon [137]. The gut microbial community plays an important role in protecting the host against pathogenic microbes [138,139], modulating immunity [140, 141] and regulating metabolic processes [142,143]. Also, the human gut has been considered as a neglected endocrine organ [144]. In the gut microbiome, bacteria can interact with each other while they must also compete with each other and phages are expected to have a significant role in driving the biodiversity of this complex ecosystem. Even if the role of human gut microbiome has been well reviewed at a metagenomic level [145–147], its integration with viromic and metatranscriptomic analyses has been little studied yet.

The human oral cavity is the second most important niche of the human body, due to the enormous amount of microorganisms that inhabit it like viruses, fungi, protozoa, archaea and bacteria. The most representative microorganisms at the oral cavity are the bacteria and viruses and their different abundances and composition have been associated with several dental diseases [148]. The bacterial diversity in the oral microbiome is approximately of 1000 species where the most representative phyla are Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria, Spirochaetes and Fusobacteria.

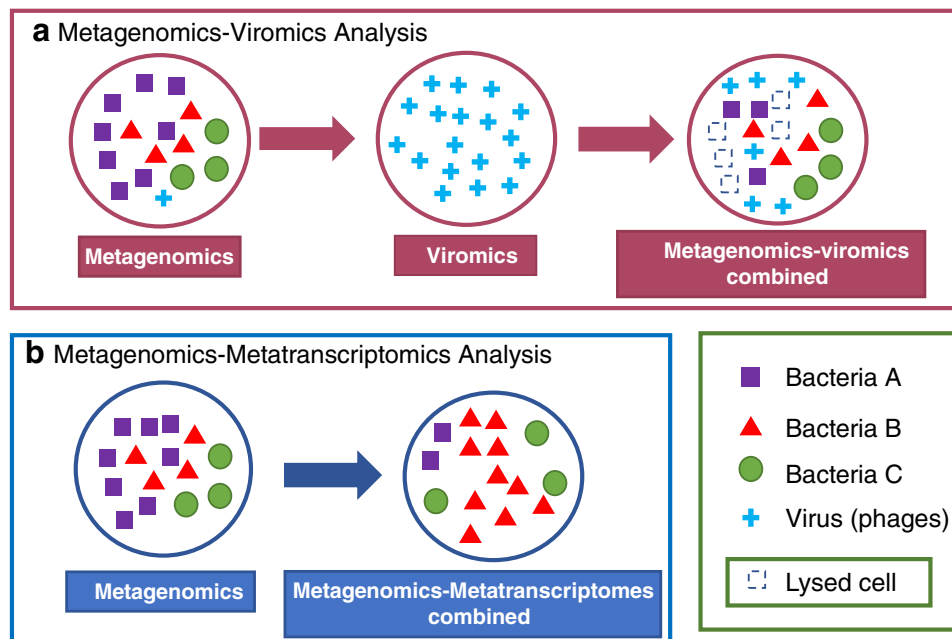
### 3.1. Metagenomic and 16S rRNA profiling combined with viromic analyses

Over the past decade, an increasing number of studies have indicated that changes in bacterial abundance and/or composition are associated with the presence of several human diseases [1,16,137]. Bacteria are the most enriched microbes inhabiting human body sites and their viruses (bacteriophages) are significantly more prevalent than eukaryotic viruses [103,149]. The bacteriophages (phages) are a natural antibacterial able to regulate bacterial populations due to the induction of bacterial lysis or by providing functional advantages to their host [150] (Fig. 4a). According to this, the application of phages to the treatment of chronic bacterial infections has proved their potential role in human health by taking advantage of their capacity to destroy pathogens [150,151]. However, there are very few examples studying the interaction between phages and their microbial hosts in human microbiome combining viromics and metagenomics. In this regard, Norman et al. used 16S rRNA gene profiling and virome sequencing to suggest that decreased diversity of enteric bacterial community observed in patients with Crohn's disease and ulcerative colitis was associated with an abnormal virome composition [24]. In particular, the bacteria family *Bacteroidaceae* showed a reduction in their relative abundance that was correlated with an increase in several *Caudovirales* bacteriophages in Crohn's disease (CD) [24]. In contrast, the families *Enterobacteriaceae* and *Pasteurellaceae* were increased in their relative abundance in patients with CD [24]. Although other studies based only on 16S rRNA profiling or metagenomic analyses [152–154] also showed a low bacterial diversity in inflammatory bowel disease, the Norman et al. study demonstrated that bacteriophages may have a role in this disease through interactions with the bacterial community of the microbiome [24]. This study shows how powerful it is to combine the viromics with 16S rRNA profiling for a better comprehension of the human microbiome.

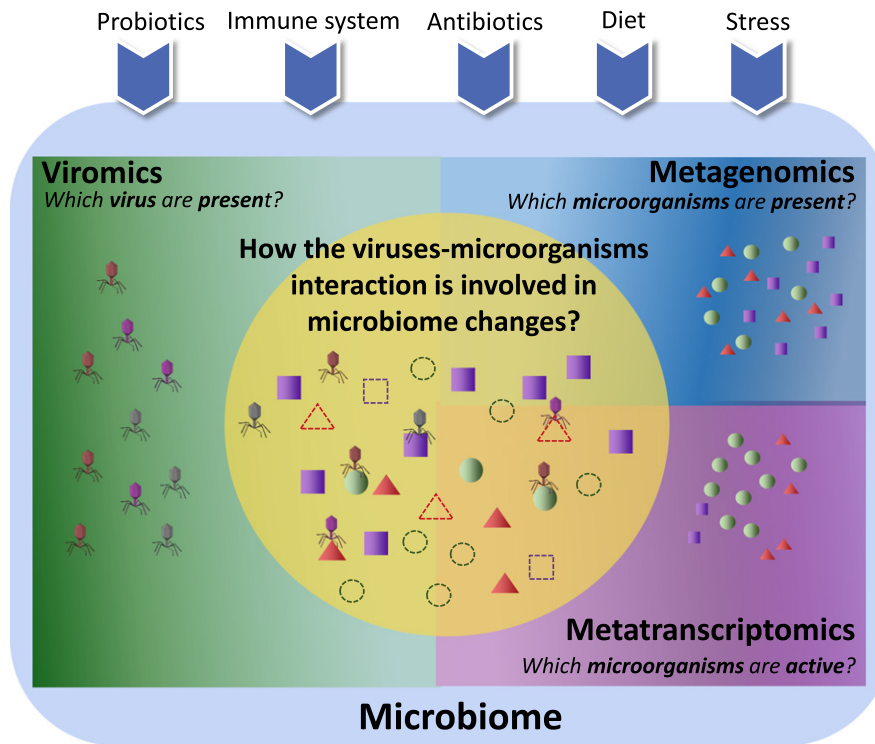
Other studies in the gut and oral microbiomes through the integration of metagenomics and viromics have shown that virus diversity

correlates with their microbial host counterparts [21,95,100,127,155–158]. For example, the analysis of all publicly available fecal metagenomes showed that the new bacteriophage crAssphage was the predominant phage in all samples [21]. Dutilh et al. predicted that the crAssphage host belongs to Bacteroidetes phylum and they found that this virus comprises up to 90% and 22% of all sequence reads in viral particle derived metagenomes and total community metagenomes, respectively [21]. Interestingly, the 90% of the human gut microbiome comprises some combination of Bacteroidetes and Firmicutes, hence Dutilh et al. suggested, that a high amount of crAssphage can be due to the high proportion of their host in the human gut microbiome [21]. Another study that also showed a strong correlation between the bacterial diversity (measured by 16S rRNA sequencing) and viral diversity (measured by VPs sequencing) in gut microbiome is the one published by Minot et al. [100]. The study showed that correlation on diversity is maintained over time (1–8 days), suggesting that the observed stability in bacterial diversity is also reflected in their host viruses through time [100]. Interestingly, the functional potential reflected by the 16S rRNA gene profiling was different to the one observed by VPs sequencing of the same sample. However, whether the changes in phage abundance are a result of changes in abundance of their hosts, or whether additional mechanisms (as lysogenic induction) are involved will require further work combining metagenomics with viromics data (Fig. 3a).

Interestingly, it was demonstrated that phages can accelerate the genomic evolution of its bacterial host in the microbiome [159,160] and these genomic changes can lead to functional adaptations of host's bacterial community [156]. In this regard, it has been demonstrated that gut and oral viromes are dominated by temperate phages integrated into the genome of their hosts, therefore this integration may alter host-bacterial phenotype by lysogenic conversion [95,100,161,162]. An example of that functional interaction was recently demonstrated in an animal model, where a prophage was liberated from its host cell after that cell was exposed to a fecal community [95]. The authors



**Fig. 3.** Molecular interactions explored using metagenomics–viromics and metagenomics–metatranscriptomics analyses. The interactions between microorganisms in the human microbiome can be better studied combining omics analysis. (a) In this panel is illustrated how phages can interact and affect the microbial diversity by infecting their host bacteria and thus promoting homeostasis or dysbiosis [179,24]. This type of interaction can be explored using viromics combined with metagenomics. (b) The species abundance of the three hypothetical bacteria can be different depending on the used analysis (metagenomics or metagenomics combined with metatranscriptomics) [171]. The data integration and normalization when metagenomics is combined with metatranscriptomics is important because the metatranscriptomics data can correspond to different species abundance and/or to differentially expressed transcriptomes.



**Fig. 4.** Towards a systems level understanding of human microbiome. The use of only one analysis to study the human microbiome (viromics, metagenomics or metatranscriptomics) provides a partial view of the complete ecological system. In a combined approach, the metagenomic analysis can give us a view of the microorganism's abundance and functions available in the microbiome, while the metatranscriptomic analyses combined with metagenomics can show us which of these microorganisms and functions are actually active. Finally, the integration of viromics analysis with the other omics data can provide information about the role that viruses play within the microbiome. The combined analyses can offer a better understanding of the role that external factors like diet, immune system and probiotics are playing in shaping the human microbiome abundance and composition. Thus, an integrated systems analysis (orange circle) seems necessary to have a better understanding of molecular mechanisms and their interactions in human microbiome.

determined that after two weeks of colonization of gnotobiotic mice with two bacteria containing prophages those were differentially activated in cecal and fecal samples [95]. Moreover, a differential prophage induction was also observed in human fecal samples [100] by comparing metagenomic with viromic sequences [100]. The later study also suggests that viruses are a reservoir for functions that can be used by their host under specific biological conditions, like stress or diet (Fig. 3a). There are multiple ways in which a prophage offers evolutionary advantages to their hosts through genetic diversity [163]. In this regard, a recent study demonstrated that activation of the composite phage (0 V1/7) in *E. faecalis* is controlled by the nutrient availability in the mouse intestine [156]. This study suggests that prophages can impact the dynamics of bacterial abundance in the mammalian gut after the exposition to nutrient availability [156]. Furthermore, the phage-encoded proteins can increase virulence either indirectly by aiding bacterial adaptation to the niche and/or directly through expression of toxins and other virulence factors [164–166]. For example, in oral microbiome the virulence of *Corynebacterium diphtheria* was associated with a small region of a putative prophage [167].

The articles discussed above demonstrated that microbial abundance and composition in the gut and oral microbiomes are also influenced by phage members and that virus activation can be triggered by the human diet and habits. This suggests that virus–bacteria interaction could be present in complex diseases, like obesity and diabetes. However, the genetic regulation that bacteria offer against the viruses is also important. For example, the CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats) which are spacer sequences that interfere with viral replication have been detected in human microbiome [168,169]. This suggests that CRISPRs mechanism contribute to the interaction between bacteria and viruses in the human microbiome. Hence, the interactions between phages and their hosts

are of great interest for understanding their impact in shaping the abundance and composition of the human microbiome [158]. However, further studies are required to understand if the interaction between metagenomes and viromes play a significant role in the progression or impediment of human diseased states associated with changes in microbiome abundance and composition.

The most common microorganism used as a probiotic is Bifidobacteria. However, in the majority of Bifidobacterial genomes the existence of prophage sequences has been confirmed, suggesting that also these intestinal commensals are targeted by phage predation [170]. Interestingly, in probiotics trials there are always subsets of individuals who do not respond to probiotics interventions, leading us to think about the role that the interaction of phages–bacteria and/or the inter-individual differences in bacterial composition are playing in shaping the microbial populations under probiotics treatments. However, more detailed investigation on the interactions between phage, bacteria and probiotics are necessary [170]. For example, determining the virome combined with metagenome before and after probiotic treatment may be an effective method to study the dynamics of gut microbial community.

### 3.2. Metagenomic combined with metatranscriptomic analyses

The metatranscriptomics is becoming increasingly practical as a tool to analyze the regulation and dynamics of transcriptionally active microbial community. The application of metatranscriptomic combined with metagenomic analysis has showed that gut microbiome contains distinctive sets of active microorganisms between individuals [19,171] (Fig. 4b). Interestingly, the induction of microbial genes as a response to host targeted exposure of xenobiotics has been observed using metagenomic combined with metatranscriptomic analysis of the gut microbiome [172]. The later study also shows that the level of metabolic activity can define the gut microbiome members [172]. However more



studies are necessary to understand how metabolic activity can influence bacterial fitness and how external factors like diet are implicated in their expression (Figs. 3a and 4). The potential of combining metatranscriptomic with metagenomic analysis is clearly observed in the Franzosa study [19]. This study showed that a substantial fraction (59%) of microbial transcripts was differentially regulated relative to their genomic abundances. The study also demonstrated that several gene families that are less abundant at the metagenomic level can be very active at the metatranscriptomic level and vice versa [19], suggesting that performing metagenomic analyses alone, could overestimate or underestimate the functional relevance of the encoded genes in metagenomes [19]. The authors also suggest that functional diversity at transcriptional level shows a pattern of subject-specific metagenome regulation; this is measured by the function of the top 10 gene families of each analyzed individual [19]. Interestingly, the metatranscriptomics profiles were significantly more individualized than DNA-level functional profiles, suggesting a subject-specific whole-community regulation. Additionally, the Franzosa study also demonstrates that the gut microbiome seems to be more stable between samples at a functional level independently of the microorganisms that are producing them. Another recent study performed by Gosalbes et al. [173], sequenced the total RNA of 10 fecal samples from healthy volunteers and the relative abundance at family level of the 16S rRNA sequences indicated that phylogenetic composition is not uniformly distributed among individuals. Contrarily, the functional analysis of putative mRNAs against COG database showed and increased homogeneity distribution of the functional COG categories between the samples [173]. Interestingly, Gosalbes et al. suggest a health related functional profile showing some differences with those indicated by the potential functions of predicted genes in DNA-based surveys. In another work on oral microbiome, the relative abundance of bacterial genera with metagenomic data was also different to that obtained with metatranscriptomic data [171].

The studies discussed in this section suggest that metatranscriptomic combined with metagenomic analyses allow a deep understanding on microbial interactions within human microbiome. Furthermore, these combined omics analyses provide useful insights about the microorganisms that have relevant functions and at the same time it allows knowing the active genes and pathways that can be related to a diseased phenotype (Fig. 4). Furthermore, the interactions between prophages and bacteria within the human microbiome can also be explored through the identification of the expressed genes from the prophages under particular ambient conditions (Fig. 4). Applying such approach to study the expression of phage genes in the gut microbiome can indicate the role of encoded prophage genes in microbial physiology and determine the dynamics that exist between active phages and their microbial host.

#### 4. Perspectives and future trends of the human microbiome analyses

The study of human microbiome through the combination of metagenomic, metatranscriptomic and viromic analyses allows a deep understanding of molecular interactions within microorganisms and their role in human health and disease (Fig. 4). Of these combined analyses the identification of potential genes, pathways and viruses that can be associated with health and disease is also possible. These driver genes and pathways could be explored as a potential pharmacological target to treat diseases that are associated with a microbiome dysbiosis. Although bacteria have been directly associated with human diseases, the role of the virome in the microbial community should be explored. However, the lack of a conserved region in virus genomes (like the 16S rRNA gene of bacteria), make the study of viruses more difficult to analyze in large cohorts. Hence, the development of novel experimental and bioinformatic strategies for a better comprehension of the role that viruses play in the microbial dysbiosis is necessary.

Probiotics can profoundly alter the human microbial community [170–175], opening the possibility to use them for the treatment of diseases associated with microbial dysbiosis. However, the reestablishment of microbial diversity has not been observed in all individuals subjected to probiotic therapy trials [170], [176,177]. Therefore, the combination of metagenomics, metatranscriptomics and viromics is an important approach to investigate the role that the interaction between viruses and bacteria play in probiotic therapies. The systems level study of human microbiome opens the opportunity to identify novel molecular targets for the treatment of microbial dysbiosis associated to human diseases.

#### Acknowledgments

We acknowledge the support provided by CONACyT grants CB-2013-223279 and SALUD-2014-C01-234188. The authors would like to acknowledge the reviewers for their helpful suggestions.

#### References

- [1] Hooper LV, Gordon JL. Commensal host–bacterial relationships in the gut. *Science* 2001;292(5519):1115–8.
- [2] Kuczynski J, Lauber CL, Walters WA, Parfrey LW, Clemente JC, Gevers D, et al. Experimental and analytical tools for studying the human microbiome. *Nat Rev Genet* 2012;13(1):47–58.
- [3] Weinstock GM. Genomic approaches to studying the human microbiota. *Nature* 2012;489(7415):250–6.
- [4] Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, et al. The NIH human microbiome project. *Genome Res* 2009;19(12):2317–23.
- [5] Blaser MJ. Harnessing the power of the human microbiome. *Proc Natl Acad Sci U S A* 2010;107(14):6125–6.
- [6] Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 2010;464(7285):59–65.
- [7] Chen T, Yu WH, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE. The human oral microbiome database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database* 2010;2010(1):baq013.
- [8] Mülle JG, Sharp WG, Cubells JF. The gut microbiome: a new frontier in autism research. *Curr Psychiatry Rep Feb* 2013;15(2):337.
- [9] Bhattacharjee S, Lukiw WJ. Alzheimer's disease and the microbiome. *Front Cell Neurosci* 2013;7:153.
- [10] D'Argenio V, Salvatore F. The role of the gut microbiome in the healthy adult status. *Clin Chim Acta* 2015 [in press].
- [11] He J, Li Y, Cao Y, Xue J, Zhou X, et al. The oral microbiome diversity and its relation to human diseases. *Folia Microbiol (Praha)* 2014;60(1):69–80.
- [12] Smith MI, Yatsunenkov T, Manary MJ, Trehan I, Mkakosya R, Cheng J, et al. Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. *Science* 2013;339(6119):548–54.
- [13] Wong VW, Tse CH, Lam TT, Wong GL, Chim AM, Chu WC, et al. Molecular characterization of the fecal microbiota in patients with nonalcoholic steatohepatitis—a longitudinal study. *PLoS One* 2013;8(4):e62885.
- [14] Cammarota G, Ianiro G, Cianci R, Bibbo S, Gasbarrini A, Curro D. The involvement of gut microbiota in inflammatory bowel disease pathogenesis: potential for therapy. *Pharmacol Ther* 2015;149:191–212.
- [15] Sanchez M, Panahi S, Tremblay A. Childhood obesity: a role for gut microbiota? *Int J Environ Res Public Health* 2014;12(1):162–75.
- [16] Dicksved J, Halfvarson J, Rosenquist M, Järnerot G, Tysk C, Apajalahti J, et al. Molecular analysis of the gut microbiota of identical twins with Crohn's disease. *ISME J* 2008;2(7):716–27.
- [17] Tai N, Wong FS, Wen L. The role of gut microbiota in the development of type 1, type 2 diabetes mellitus and obesity. *Rev Endocr Metab Disord* 2015;16(1):55–65.
- [18] Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* 2013;31(9):814–21.
- [19] Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM, et al. Relating the metatranscriptome and metagenome of the human gut. *Proc Natl Acad Sci U S A* 2014;111(22):E2329–38.
- [20] David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 2014;505(7484):559–63.
- [21] Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GG, Boling L, et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat Commun* 2014;5:4498.
- [22] Dalmaso M, Hill C, Ross RP. Exploiting gut bacteriophages for human health. *Trends Microbiol* 2014;22(7):399–405.
- [23] Mills S, Shanahan F, Stanton C, Hill C, Coffey A, Ross RP. Movers and shakers: influence of bacteriophages in shaping the mammalian gut microbiota. *Gut Microbes* 2013;4(1):4–16.

- [24] Norman JM, Handley SA, Baldrige MT, Droit L, Liu CY, Keller BC, et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* 2015;160(3):447–60.
- [25] Foxman B, Goldberg D, Murdock C, Xi C, Gilsdorf JR, et al. Conceptualizing human microbiota: from multicelled organ to ecological community. *Interdiscip Perspect Infect Dis* 2008;6:13979.
- [26] Grahn N, Olofsson M, Ellnebo-Svedlund K, Monstein HJ, Jonasson J. Identification of mixed bacterial DNA contamination in broad-range PCR amplification of 16S rDNA V1 and V3 variable regions by pyrosequencing of cloned amplicons. *FEMS Microbiol Lett* 2003;219(1):87–91.
- [27] Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R. Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res* 2007;35(18):e120.
- [28] David LA, Materna AC, Friedman J, Campos-Baptista MI, Blackburn MC, Perrotta A, et al. Host lifestyle affects human microbiota on daily timescales. *Genome Biol* 2014;15(7):R89.
- [29] Ahn J, Yang L, Paster BJ, Ganly I, Morris L, Pei Z, et al. Oral microbiome profiles: 16S rRNA pyrosequencing and microarray assay comparison. *PLoS One* 2011;6(7):e22788.
- [30] Salipante SJ, Sengupta DJ, Rosenthal C, Costa G, Spangler J, Sims EH, et al. Rapid 16S rRNA next-generation sequencing of polymicrobial clinical samples for diagnosis of complex bacterial infections. *PLoS One* 2013;8(5):e65226.
- [31] Schmalenberger A, Schwieger F, Tebbe CC. Effect of primers hybridizing to different evolutionarily conserved regions of the small-subunit rRNA gene in PCR-based microbial community analyses and genetic profiling. *Appl Environ Microbiol* 2001;67(8):3557–63.
- [32] Bricheux G, Morin L, Le Moal G, Coffe G, Balestrino D, Charbonnel N, et al. Pyrosequencing assessment of prokaryotic and eukaryotic diversity in biofilm communities from a French river. *Microbiol Open* 2013;2(3):402–14.
- [33] Liu Z, DeSantis TZ, Andersen GL, Knight R. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Res* 2008;36(18):e120.
- [34] Chakravorty S, Helb D, Burday M, Connell N, Alland D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods* 2007;69(2):330–9.
- [35] Fettweis JM, Serrano MG, Sheth NU, Mayer CM, Glascock AL, Brooks JP, et al. Species-level classification of the vaginal microbiome. *BMC Genomics* 2012;13(Suppl. 8):S17.
- [36] Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A* 2010;108(Suppl. 1):4516–22.
- [37] Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 2012;6(8):1621–4.
- [38] Gilbert JA, Jansson JK, Knight R. The Earth Microbiome Project: successes and aspirations. *BMC Biol* 2014;12:69.
- [39] Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;7(5):335–6.
- [40] Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, 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.
- [41] Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, et al. The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinforma* 2008;9:386.
- [42] Kosakovsky Pond S, Wadhawan S, Chiaromonte F, Ananda G, Chung WY, Taylor J, et al. Windshield splatter analysis with the Galaxy metagenomic pipeline. *Genome Res* 2009;19(11):2144–53.
- [43] DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006;72(7):5069–72.
- [44] Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 2013;42:D633–42.
- [45] Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;41:D590–6.
- [46] Morgan XC, Huttenhower C. Chapter 12: human microbiome analysis. *PLoS Comput Biol* 2012;8(12):e1002808.
- [47] Russell JR, Huang J, Anand P, Kucera K, Sandoval AG, Dantzer KW, et al. Biodegradation of polyester polyurethane by endophytic fungi. *Appl Environ Microbiol* 2011;77(17):6076–84.
- [48] Lovley DR. Cleaning up with genomics: applying molecular biology to bioremediation. *Nat Rev Microbiol* 2003;1(1):35–44.
- [49] Vazquez-Castellanos JF, Garcia-Lopez R, Perez-Brocail V, Pignatelli M, Moya A. Comparison of different assembly and annotation tools on analysis of simulated viral metagenomic communities in the gut. *BMC Genomics* 2014;15:37.
- [50] Suez J, Korem T, Zeevi D, Zilberman-Schapira G, Thaiss CA, Maza O, et al. Artificial sweeteners induce glucose intolerance by altering the gut microbiota. *Nature* 2014;514(7521):181–6.
- [51] Huttenhower C, Gevers D. Structure, function and diversity of the healthy human microbiome. *Nature* 2012;486(7402):207–14.
- [52] Schmieder R, Edwards R. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One* 2011;6(3):e17288.
- [53] Breitenstein S, Tummler B, Romling U. Pulsed field gel electrophoresis of bacterial DNA isolated directly from patients' sputa. *Nucleic Acids Res* 1995;23(4):722–3.
- [54] Shak S, Capon DJ, Hellmiss R, Marsters SA, Baker CL. Recombinant human DNase I reduces the viscosity of cystic fibrosis sputum. *Proc Natl Acad Sci U S A* 1990;87(23):9188–92.
- [55] Lethem MI, James SL, Marriott C, Burke JF. The origin of DNA associated with mucus glycoproteins in cystic fibrosis sputum. *Eur Respir J* 1990;3(1):19–23.
- [56] Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009;10(3):R25.
- [57] Ning Z, Cox AJ, Mullikin JC. SSAHA: a fast search method for large DNA databases. *Genome Res* 2001;11(10):1725–9.
- [58] Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009;25(14):1754–60.
- [59] Marco-Sola S, Sammeth M, Guigo R, Ribeca P. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat Methods* 2012;9(12):1185–8.
- [60] Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade specific marker genes. *Nat Methods* 2012;9(8):811–4.
- [61] Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;26(19):2460–1.
- [62] Wu M, Eisen JA. A simple, fast, and accurate method of phylogenomic inference. *Genome Biol* 2008;9(10):R151.
- [63] Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 2011;39:W29–37.
- [64] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;28(1):27–30.
- [65] Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. *Nucleic Acids Res* 2013;42:D222–30.
- [66] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25(1):25–9.
- [67] Nawrocki EP, Eddy SR. Computational identification of functional RNA homologs in metagenomic data. *RNA Biol* 2013;10(7):1170–9.
- [68] Abubucker S, Segata N, Goll J, Schubert AM, Izard J, Cantarel BL, et al. Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput Biol* 2012;8(6):e1002358.
- [69] Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. *Genome Res* 2007;17(3):377–86.
- [70] Markowitz VM, Ivanova NN, Szeto E, Palaniappan K, Chu K, Dalevi D, et al. IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res* 2008;36:D534–8.
- [71] Goll J, Rusch DB, Tanenbaum DM, Thiagarajan M, Li K, Methe BA, et al. METAREP: JCVI metagenomics reports—an open source tool for high-performance comparative metagenomics. *Bioinformatics* 2010;26(20):2631–2.
- [72] 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.
- [73] Pell J, Hintze A, Canino-Koning R, Howe A, Tiedje JM, Brown CT. Scaling metagenome sequence assembly with probabilistic de Bruijn graphs. *Proc Natl Acad Sci U S A* 2012;109(33):13272–7.
- [74] Treangen TJ, Koren S, Sommer DD, Liu B, Astrovskaya I, Ondov B, et al. MetAMOS: a modular and open source metagenomic assembly and analysis pipeline. *Genome Biol* 2011;14(1):R2.
- [75] Peng Y, Leung HC, Yiu SM, Chin FY. Meta-IDBA: a de novo assembler for metagenomic data. *Bioinformatics* 2011;27(13):i94–i101.
- [76] Ye Y, Tang H. An ORFome assembly approach to metagenomics sequences analysis. *J Bioinform Comput Biol* 2009;7(3):455–71.
- [77] Boisvert S, Raymond F, Godzaridis E, Laviolette F, Corbeil J. Ray Meta: scalable de novo metagenome assembly and profiling. *Genome Biol* 2012;13(12):R122.
- [78] He S, Kunin V, Haynes M, Martin HG, Ivanova N, Rohwer F, et al. Metatranscriptomic array analysis of 'Candidatus *Accumulibacter phosphatis*'—enriched enhanced biological phosphorus removal sludge. *Environ Microbiol* 2010;12(5):1205–17.
- [79] Liu D, Graber JH. Quantitative comparison of EST libraries requires compensation for systematic biases in cDNA generation. *BMC Bioinforma* 2006;7:77.
- [80] Ozsolak F, Milos PM. Single-molecule direct RNA sequencing without cDNA synthesis. *Wiley Interdiscip Rev RNA* 2011;2(4):565–70.
- [81] Ozsolak F, Platt AR, Jones DR, Reifengerber JG, Sass LE, McInerney P, et al. Direct RNA sequencing. *Nature* 2009;461(7265):814–8.
- [82] Hickman SE, Kingery ND, Ohsumi TK, Borowsky ML, Wang LC, Means TK, et al. The microglial sensome revealed by direct RNA sequencing. *Nat Neurosci* 2013;16(12):1896–905.
- [83] Giannoukos G, Ciulla DM, Huang K, Haas BJ, Izard J, Levin JZ, et al. Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biol* 2012;13(3):R23.
- [84] Jorth P, Ciulla DM, Huang K, Haas BJ, Izard J, Levin JZ. Metatranscriptomics of the human oral microbiome during health and disease. *MBio* 2014;5(2):e01012–4.
- [85] Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, et al. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 2009;25(15):1966–7.
- [86] Birol I, Jackman SD, Nielsen CB, Qian JQ, Varhol R, Stazyk G, et al. De novo transcriptome assembly with ABySS. *Bioinformatics* 2009;25(21):2872–7.
- [87] 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.

- [88] Shi CY, Yang H, Wei CL, Yu O, Zhang ZZ, Jiang CJ, et al. Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. *BMC Genomics* 2011;12:131.
- [89] Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, et al. De novo assembly and analysis of RNA-seq data. *Nat Methods* 2010;7(11):909–12.
- [90] Garg R, Patel RK, Tyagi AK, Jain M. De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Res* 2010;18(1):53–63.
- [91] Ness RW, Siol M, Barrett SC. De novo sequence assembly and characterization of the floral transcriptome in cross- and self-fertilizing plants. *BMC Genomics* 2011;12:298 [936].
- [92] Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011;29(7):644–52.
- [93] Ghaffari N, Sanchez-Flores A, Doan R, Garcia-Orozco KD, Chen PL, Ochoa-Leyva A, et al. Novel transcriptome assembly and improved annotation of the whiteleg shrimp (*Litopenaeus vannamei*), a dominant crustacean in global seafood mariculture. *Sci Rep* 2014;4:7081.
- [94] Luria N, Sela N, Yaari M, Feygenberg O, Kobiler I, Lers A, et al. De-novo assembly of mango fruit peel transcriptome reveals mechanisms of mango response to hot water treatment. *BMC Genomics* 2014;15:957.
- [95] Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* 2010;466(7304):334–8.
- [96] Caporaso JG, Knight R, Kelley ST. Host-associated and free-living phage communities differ profoundly in phylogenetic composition. *PLoS One* 2011;6(2):e16900.
- [97] Handley SA, Thackray LB, Zhao G, Presti R, Miller AD, Droit L, et al. Pathogenic simian immunodeficiency virus infection is associated with expansion of the enteric virome. *Cell* 2012;151(2):253–66.
- [98] Hurwitz BL, Deng L, Poulos BT, Sullivan MB. Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environ Microbiol* 2013;15(5):1428–40.
- [99] Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer. Laboratory procedures to generate viral metagenomes. *Nat Protoc* 2009;4(4):470–83.
- [100] Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, et al. The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res* 2011;21(10):1616–25.
- [101] Minot S, Grunberg S, Wu GD, Lewis JD, Bushman FD. Hypervariable loci in the human gut virome. *Proc Natl Acad Sci U S A* 2011;109(10):3962–6.
- [102] Rohwer F, Seguritan V, Choi DH, Segall AM, Azam F. Production of shotgun libraries using random amplification. *Biotechniques* 2001;31(1):108 [–12, 114–6, 118].
- [103] Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P, et al. Metagenomic analyses of an uncultured viral community from human feces. *J Bacteriol* 2003;185(20):6220–3.
- [104] Hutchison III CA, Smith HO, Pfannkoch C, Venter JC. Cell-free cloning using phi29 DNA polymerase. *Proc Natl Acad Sci U S A* 2005;102(48):17332–6.
- [105] Lasken RS, Stockwell TB. Mechanism of chimera formation during the multiple displacement amplification reaction. *BMC Biotechnol* 2007;7:19.
- [106] Abulencia CB, Wyborski DL, Garcia JA, Podar M, Chen W, Chang SH, et al. Environmental whole-genome amplification to access microbial populations in contaminated sediments. *Appl Environ Microbiol* 2006;72(5):3291–301.
- [107] Yilmaz S, Allgaier M, Hugenholz P. Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nat Methods* 2010;7(12):943–4.
- [108] Zhang K, Martiny AC, Reppas NB, Barry KW, Malek J, Chisholm SW, et al. Sequencing genomes from single cells by polymerase cloning. *Nat Biotechnol* 2006;24(6):680–6.
- [109] Kim KH, Chang HW, Nam YD, Roh SW, Kim MS, Sung Y, et al. Amplification of uncultured single-stranded DNA viruses from rice paddy soil. *Appl Environ Microbiol* 2008;74(19):5975–85.
- [110] Wommack KE, Bhavsar J, Ravel J. Metagenomics: read length matters. *Appl Environ Microbiol* 2008;74(5):1453–63.
- [111] Bibby K, Viau E, Peccia J. Viral metagenome analysis to guide human pathogen monitoring in environmental samples. *Lett Appl Microbiol* 2011;52(4):386–92.
- [112] Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, et al. Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci U S A* 2002;99(22):14250–5.
- [113] Breitbart M, Felts B, Kelley S, Mahaffy JM, Nulton J, Salamon P, et al. Diversity and population structure of a near-shore marine sediment viral community. *Proc Biol Sci* 2004;271(1539):565–74.
- [114] Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, et al. The marine viromes of four oceanic regions. *PLoS Biol* 2006;4(11):e368.
- [115] Schoenfeld T, Patterson M, Richardson PM, Wommack KE, Young M, Mead D. Assembly of viral metagenomes from Yellowstone hot springs. *Appl Environ Microbiol* 2008;74(13):4164–74.
- [116] 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.
- [117] Rohwer F, Thurber RV. Viruses manipulate the marine environment. *Nature* 2009;459(7244):207–12.
- [118] Brister JR, Ako-Adjei D, Bao Y, Blinkova O. NCBI viral genomes resource. *Nucleic Acids Res* 2014;43:D571–7.
- [119] Yozwiak NL, Skewes-Cox P, Stenglein MD, Balmaseda A, Harris E, DeRisi JL. Virus identification in unknown tropical febrile illness cases using deep sequencing. *PLoS Negl Trop Dis* 2012;6(2):e1485.
- [120] Yang J, Yang F, Ren L, Xiong Z, Wu Z, Dong J, et al. Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. *J Clin Microbiol* 2011;49(10):3463–9.
- [121] Xu B, Liu L, Huang X, Ma H, Zhang Y, Du Y, et al. Metagenomic analysis of fever, thrombocytopenia and leukopenia syndrome (FTLS) in Henan Province, China: discovery of a new bunyavirus. *PLoS Pathog* 2011;7(11):e1002369.
- [122] Leplae R, Lima-Mendez G, Toussaint A. ACLAME: a Classification of Mobile genetic Elements, update 2010. *Nucleic Acids Res* 2010;38:D57–61.
- [123] Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* 2005;33(17):5691–702.
- [124] Angly F, Rodriguez-Brito B, Bangor D, McNairne P, Breitbart M, Salamon P, et al. PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. *BMC Bioinforma* 2005;6:41.
- [125] Dutilh BE, Schmieder R, Nulton J, Felts B, Salamon P, Edwards RA, et al. Reference-independent comparative metagenomics using cross assembly: crAss. *Bioinformatics* 2012;28(24):3225–31.
- [126] Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. *Nature* 2011;473(7346):174–80.
- [127] Perez-Brocal V, Garcia-Lopez R, Vazquez-Castellanos JF, Nos P, Beltran B, Latorre A, et al. Study of the viral and microbial communities associated with Crohn's disease: a metagenomic approach. *Clin Transl Gastroenterol* 2013;4:e36.
- [128] Williamson SJ, Allen LZ, Lorenzi HA, Fadrosch DW, Bami D, Thiagarajan M, et al. Metagenomic exploration of viruses throughout the Indian Ocean. *PLoS One* 2012;7(10):e42047.
- [129] Yu JM, Li JS, Ao YY, Duan ZJ. Detection of novel viruses in porcine fecal samples from China. *Virol J* 2013;10:39.
- [130] Methé B. A framework for human microbiome research. *Nature* 2012;486(7402):215–21.
- [131] Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, et al. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 2004;304(5667):66–74.
- [132] Treangen TJ, Sommer DD, Angly FE, Koren S, Pop M. Next generation sequence assembly with AMOS. *Curr Protoc Bioinformatics* 2011:11–8.
- [133] Allen HK, Bunge J, Foster JA, Bayles DO, Stanton TB. Estimation of viral richness from shotgun metagenomes using a frequency count approach. *Microbiome* 2013;1(1):5.
- [134] Yang X, Charlebois P, Gnerre S, Coole MG, Lennon NJ, Levin JZ, et al. De novo assembly of highly diverse viral populations. *BMC Genomics* 2012;13:475.
- [135] Hurwitz BL, Sullivan MB. The Pacific Ocean virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS One* 2013;8(2):e57355.
- [136] Peng Y, Leung HC, Yiu SM, Chin FY. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 2012;28(11):1420–8.
- [137] Ley RE, Turnbaugh PJ, Klein S, Gordon JL. Microbial ecology: human gut microbes associated with obesity. *Nature* 2006;444(7122):1022–3.
- [138] Endt K, Stecher B, Chaffron S, Slack E, Tchitchek N, Benecke A, et al. The microbiota mediates pathogen clearance from the gut lumen after non-typhoidal Salmonella diarrhea. *PLoS Pathog* 2010;6(9):e1001097.
- [139] Fukuda S, Toh H, Taylor TD, Ohno H, Hattori M. Acetate-producing bifidobacteria protect the host from enteropathogenic infection via carbohydrate transporters. *Gut Microbes* 2012;3(5):449–54.
- [140] Maynard CL, Elson CO, Hatton RD, Weaver CT. Reciprocal interactions of the intestinal microbiota and immune system. *Nature* 2012;489(7415):231–41.
- [141] Viaud S, Saccheri F, Mignot G, Yamazaki T, Dailere R, Hannani D, et al. The intestinal microbiota modulates the anticancer immune effects of cyclophosphamide. *Science* 2013;342(6161):971–6.
- [142] Tremaroli V, Backhed F. Functional interactions between the gut microbiota and host metabolism. *Nature* 2012;489(7415):242–9.
- [143] Cani PD. Metabolism in 2013: the gut microbiota manages host metabolism. *Nat Rev Endocrinol* 2014;10(2):74–6.
- [144] Clarke G, Stilling RM, Kennedy PJ, Stanton C, Cryan JF, Dinan TG, et al. Minireview: gut microbiota: the neglected endocrine organ. *Mol Endocrinol* 2014;28(8):1221–38.
- [145] Sommer F, Backhed F. The gut microbiota—masters of host development and physiology. *Nat Rev Microbiol* 2013;11(4):227–38.
- [146] Clemente JC, Ursell LK, Parfrey LW, Knight R. The impact of the gut microbiota on human health: an integrative view. *Cell* 2012;148(6):1258–70.
- [147] Owyang C, Wu GD. The gut microbiome in health and disease. *Gastroenterology* 2014;146(6):1433–6.
- [148] Abeles SR, Robles-Sikisaka R, Ly M, Lum AG, Salzman J, Boehm TK, et al. Human oral viruses are personal, persistent and gender-consistent. *ISME J* 2014;8(9):1753–67.
- [149] Pride DT, Salzman J, Haynes M, Rohwer F, Davis-Long C, White III RA, et al. Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. *ISME J* 2012;6(5):915–26.
- [150] Wittebole X, De Rook S, Opal SM. A historical overview of bacteriophage therapy as an alternative to antibiotics for the treatment of bacterial pathogens. *Virulence* 2013;5(1):226–35.
- [151] Maura D, Galtier M, Le Bouguenec C, Debarbieux L. Virulent bacteriophages can target O104:H4 enterohemorrhagic *Escherichia coli* in the mouse intestine. *Antimicrob Agents Chemother* 2012;56(12):6235–42.
- [152] Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux K, Pelletier E, Frangeul L, et al. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 2006;55(2):205–11.
- [153] Scher JU, Ubeda C, Artacho A, Attur M, Isaac S, Reddy SM, et al. Decreased bacterial diversity characterizes the altered gut microbiota in patients with psoriatic arthritis, resembling dysbiosis in inflammatory bowel disease. *Arthritis Rheum* 2015;67(1):128–39.

- [154] Ott SJ, Musfeldt M, Wenderoth DF, Hampe J, Brant O, Folsch UR, et al. Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. *Gut* 2004;53(5):685–93.
- [155] Sharon I, Morowitz MJ, Thomas BC, Costello EK, Relman DA, Banfield JF. Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Res* 2013;23(1):111–20.
- [156] Duerkop BA, Clements CV, Rollins D, Rodrigues JL, Hooper LV. A composite bacteriophage alters colonization by an intestinal commensal bacterium. *Proc Natl Acad Sci U S A* 2012;109(43):17621–6.
- [157] Robles-Sikisaka R, Ly M, Boehm T, Naidu M, Salzman J, Pride DT. Association between living environment and human oral viral ecology. *ISME J* 2013;7(9):1710–24.
- [158] Pride DT, Salzman J, Relman DA. Comparisons of clustered regularly interspaced short palindromic repeats and viromes in human saliva reveal bacterial adaptations to salivary viruses. *Environ Microbiol* 2012;14(9):2564–76.
- [159] Pal C, Macia MD, Oliver A, Schachar I, Buckling A. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* 2007;450(7172):1079–81.
- [160] Brockhurst MA, Morgan AD, Fenton A, Buckling A. Experimental coevolution with bacteria and phage. The *Pseudomonas fluorescens*-Phi2 model system. *Infect Genet Evol* 2007;7(4):547–52 [1090].
- [161] Hitch G, Pratten J, Taylor PW. Isolation of bacteriophages from the oral cavity. *Lett Appl Microbiol* 2004;39(2):215–9.
- [162] Rohwer F. Global phage diversity. *Cell* 2003;113(2):141.
- [163] Brussow H, Canchaya C, Hardt WD. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* 2004;68(3):560–602 [table of contents].
- [164] Canchaya C, Proux C, Fournous G, Bruttin A, Brussow H. Prophage genomics. *Microbiol Mol Biol Rev* 2003;67(2):238–76 [table of contents].
- [165] Canchaya C, Fournous G, Brussow H. The impact of prophages on bacterial chromosomes. *Mol Microbiol* 2004;53(1):9–18.
- [166] Filee J, Forterre P, Laurent J. The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies. *Res Microbiol* 2003;154(4):237–43 [1102].
- [167] Al-Jarbou AN. Genomic library screening for viruses from the human dental plaque revealed pathogen-specific lytic phage sequences. *Curr Microbiol* 2012;64(1):1–6.
- [168] Stern A, Mick E, Tirosh I, Sagy O, Sorek R. CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. *Genome Res* 2012;22(10):1985–94.
- [169] Abeles SR, Pride DT. Molecular bases and role of viruses in the human microbiome. *J Mol Biol* 2014;426(23):3892–906.
- [170] Ventura M, Sozzi T, Turrone F, Matteuzzi D, van Sinderen D. The impact of bacteriophages on probiotic bacteria and gut microbiota diversity. *Genes Nutr* 2011;6(3):205–7.
- [171] Benitez-Paez A, Belda-Ferre P, Simon-Soro A, Mira A. Microbiota diversity and gene expression dynamics in human oral biofilms. *BMC Genomics* 2014;15:311.
- [172] Ursell LK, Knight R. Xenobiotics and the human gut microbiome: metatranscriptomics reveal the active players. *Cell Metab* 2013;17(3):317–8.
- [173] Gosalbes MJ, Durban A, Pignatelli M, Abellan JJ, Jimenez-Hernandez N, Perez-Cobas AE, et al. Metatranscriptomic approach to analyze the functional human gut microbiota. *PLoS One* 2011;6(3):e17447.
- [174] Buffie CG, Jarchum I, Equinda M, Lipuma L, Gouberne A, Viale A, et al. Profound alterations of intestinal microbiota following a single dose of clindamycin results in sustained susceptibility to *Clostridium difficile*-induced colitis. *Infect Immun* 2012;80(1):62–73.
- [175] Tormo-Badia N, Hakansson A, Vasudevan K, Molin G, Ahrne S, Cilio CM. Antibiotic treatment of pregnant non-obese diabetic mice leads to altered gut microbiota and intestinal immunological changes in the off- spring. *Scand J Immunol* 2014;80(4):250–60.
- [176] Chen J, Novick RP. Phage-mediated intergeneric transfer of toxin genes. *Science* 2009;323(5910):139–41.
- [177] Ubeda C, Maiques E, Knecht E, Lasa I, Novick RP, Penades JR. Antibiotic-induced SOS response promotes horizontal dissemination of pathogenicity island-encoded virulence factors in staphylococci. *Mol Microbiol* 2005;56(3):836–44.
- [179] De Paepe M, Leclerc M, Tinsley CR, Petit MA. Bacteriophages: an underestimated role in human and animal health? *Front Cell Infect Microbiol* 2014;4:39.