Mini Review

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

Shirley Bikel\textsuperscript{a,c}, Alejandra Valdez-Lara\textsuperscript{a,c}, Fernanda Cornejo-Granados\textsuperscript{a,c}, Karina Rico\textsuperscript{a}, Samuel Canizales-Quinteros\textsuperscript{a}, Xavier Soberón\textsuperscript{b}, Luis Del Pozo-Yaunerb, Adrián Ochoa-Leyva\textsuperscript{a,c,⁎}

\textsuperscript{a} Unidad de Genómica de Poblaciones Aplicada la Salud, Facultad de Química, UNAM, Instituto Nacional de Medicina Genómica (INMEGEN), México, D.F. 14610, Mexico
\textsuperscript{b} Instituto Nacional de Medicina Genómica (INMEGEN), México, D.F., Mexico
\textsuperscript{c} Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de Mexico, Avenida Universidad 2001, Cuernavaca C.P. 62210, Mexico

\textbf{A B S T R A C T}

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/).

\textbf{Contents}

1. Introduction .............................................................. 391
2. Sequencing and bioinformatic strategies to study the human microbiome .......................................................... 391
   2.1. 16S rRNA gene profiling 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
References ........................................................................ 398

⁎ Corresponding author at: Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de Mexico, Avenida Universidad 2001, Cuernavaca C.P. 62210, Mexico.
E-mail addresses: adriano8a@gmail.com, aochoa@ibt.unam.mx (A. Ochoa-Leyva).

http://dx.doi.org/10.1016/j.csbj.2015.06.001
2001-0370/© 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/).
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 [23]. 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 microbiomes, 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 relative 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 [33,32]. 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 at 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
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...
and Galaxy, which are open source packages. The typical biohuman microbiome studies \[20,28,36,37\] and a well-established sequences from the vaginal microbiome\[35\]. The primer pair 515F/ also has demonstrated that the V4 region can be extensively supported the gold standard for human gut microbiota characterization by no consensus has been established, the V4 region has been suggested as amplicons of the 16S rRNA gene, are QIIME\[39\] Mothur \[40\], MGRAST steps. First, the amplicon sequences are subjected to data formatic pipeline to analyze amplicon sequences involves three basic
generation, after which the library is finally ready.

| Platform: Roche 454 GS-FLX + | Shotgun utility | Costs | Sequence reads per run | Run time |
|-----------------------------|-----------------|-------|------------------------|----------|
| 800 base reads              | Long reads help with assembly of new genomes. | Cost limits the deep sequencing analysis. | 1 million | 20 h |
| Platform: Illumina HiSeq, MiSeq and NextSeq | Only one per read (NextSeq); short reads do not seem to limit the taxonomical analysis. | Cost-effective deep sampling | 25 million, 2 × 300 bp (MiSeq) | 5–55 h (MiSeq) |
| 100–500 base reads          | Short reads, but high sequencing output allows a deep sequencing analysis. | (NextSeq) | 130–400 million, 2 × 150 bp | 12–30 h (NextSeq) |
| Platform: Life Technologies ion personal genome machine (PCM) | Time-effective with a mid-cost for deep sequencing analysis. | 2.5 billion, 2 × 150 bp (HiSeq) | 2–4 h (314) |
| 35–400 base reads           | Short reads, but mid sequencing output in a very fast sequencing time. | 1 × 200 or 1 × 400 and number of reads depends on model of the sequencing chip: | (316) |
| 16S metagenomics kit;       | in a very fast sequencing time. | 100 million (314) | 3–5 h |
| 35–400 base reads           | Up to 7 in custom design. | 1 billion (318) | (316) |
| 16S metagenomics kit;       | Up to 7 using the Ion | | |
| in a very fast sequencing time. | 130–400 million, 2 × 150 bp | | |
| 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 MEGA Genome 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 assembled into 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 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 MICROenRich 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 c29 polymerase which synthesizes >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 are novel and they are enriched in regions of low-complexity repeats [118–119] 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 c29 polymerase which synthesizes >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 are novel and they are enriched in regions of low-complexity repeats [118–119] 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 c29 polymerase which synthesizes >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 are novel and they are enriched in regions of low-complexity repeats [118–119] 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 c29 polymerase which synthesizes >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 are novel and they are enriched in regions of low-complexity repeats [118–119] 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 c29 polymerase which synthesizes >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 are novel and they are enriched in regions of low-complexity repeats [118–119] 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 c29 polymerase which synthesizes >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 are novel and they are enriched in regions of low-complexity repeats [118–119] 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 c29 polymerase which synthesizes >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 are novel and they are enriched in regions of low-complexity repeats [118–119] 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 c29 polymerase which synthesizes >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 are novel and they are enriched in regions of low-complexity repeats [118–119] 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 c29 polymerase which synthesizes >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 are novel and they are enriched in regions of low-complexity repeats [118–119] 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 c29 polymerase which synthesizes >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].
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 disbiosis [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.
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 contributes 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 Francoza 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 Francoza 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 rDNA gene of bacteria), make the study of viruses more difficult to analyze in large cohorts. Hence, the development of novel experimental and bioinformatics 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.

References

[1] Hooper LV, Gordon JL. Commensal host–bacterial relationships in the gut. Science 2001;292(5523):115–8.
[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 microbiome gene catalogue established by metagenomic sequencing. Nature 2010;464(7285):69–65.
[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 taxonomy and genetic information. Database 2010;2010(1):baq013.
[8] Mulu JC, Sharp WC, Cubells JF. The gut microbiome: a new frontier in autism research. Curr Psychiatry Rep Feb 2013;15(2):337.
[9] Bhattacherjee S, Lukow 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 (Prague) 2014;60(1):59–80.
[12] Smith MJ, Yatsunenko T, Manary MJ, Trehan I, Mikoszyna R, Cheng J, et al. Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. Science 2013;339(6119):548–54.
[13] Wong YW, 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, Ianario G, Cinelli R, Bibbo S, Gasbarrini A, Curro D. The involvement of gut microbiota in inflammatory bowel disease pathogenesis: potential for therapy. Pharmacol Ther 2015;150:1–12.
[15] Sanchez M, Panahi S, Tremblay A. Childhood obesity: a role for gut microbiota? Int J Environ Res Public Health 2014;11(12):162–75.
[16] Dickovsk J, Halfvarson J, Rosenquist M, Jarnerot 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] Francoza 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 JB, 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, Bolog L, et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat Commun 2014;5:4808.
[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, Ross RP. Exploiting gut bacteriophages for human health. Trends Microbiol 2014;22(7):399–405.
[27] Liu Z, Lozupone C, Hanady 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 ML, Blackburn MC, Perrotta A, et al. Host lifestyle affects human microbiota on daily timescales. Genome Biol 2014;15(7):R88.

[29] Ahn J, Yang L, Paster BJ, Ganly I, Morris L, Pei Z, et al. Oral microbiome profiling of 16S rRNA pyrosequencing and microarray assay comparison. PLoS One 2011;6(7):e22288.

[30] Salipante SJ, Sengupta DJ, Rosenblatt CA, Costa G, Sanger J, Sims EH, et al. Rapid 16S rRNA next-generation sequencing of polymericial 2x2 samples for diagnosis of complex bacterial infections. PLoS One 2013;8(5):e65226.

[31] Schmalenberger A, Schiewer F, Tbecce EB. 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] Brichete G, Morin L, Le Moal G, Cotte C, Balestrino D, Charbonnel N, et al. Pyrosequencing assessment of prokaryotic and eukaryotic diversity in biofilm communities from a French river. Microbiologica 2013;3(3):402–14.

[33] Liu Z, DeSantis TZ, Andersen GL, Knight R. Accurate taxon assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. Nucleic Acids Res 2008;36(18):e129.

[34] Chaikravy 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. Spectral bacterial RNA gene segments for the diagnosis of pathogenic bacteria. J Microbiol Methods 2013;94(2):222–3.

[36] Shak S, Capone DJ, Hellmiss R, Marsters SA, Baker CL. Recombinant human DNase I transforming. Nucleic Acids Res 2000;28:R125.

[37] Ahn J, Yang L, Paster BJ, Ganly I, Morris L, Pei Z, et al. Oral microbiome profiling of 16S rRNA pyrosequencing and microarray assay comparison. PLoS One 2011;6(7):e22288.

[38] Ng Z, Cox AJ, Mullikin JC. SSS: a fast search method for large DNA databases. Genome Res 2001;11(10):1725–9.

[39] Lihur Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 2009;25(14):1754–60.

[40] Marco-Sola S, Sammeth M, Guigo R, Ribeca P. The CEM mapper: fast, accurate and versatile alignment by filtration. Nat Methods 2012;9(12):1185–8.

[41] 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.

[42] Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinfo- matics 2010;26(19):286–94.

[43] Wu M, Eisen JA. A simple, fast, and accurate method of phylogenetic inference. Genome Res 2005;15(8):1332–9.

[44] Abubaker S, Segata N, Goll J, Schubert AM, Ijaz Z, Cantarel BL, et al. Metabolic reconstruction for metagenomic data and its application to the human microbiome. BMC Genomics 2013;14:245.

[45] Huson DH, Auch AF, Qi J, Schuster SC. MEGAAN analysis of metagenomic data. Genome Res 2007;17(3):377–86.

[46] Goll J, Rusch DB, Tanenbaum DM, Thiagarajan M, Li K, Metz BA, et al. METAP: de novo metagenomic assembly of short reads. Nucleic Acids Res 2008;36:13272–8.

[47] Shi CY, Yang H, Wei CL, Yu O, Zhang ZZ, Jiang CJ, et al. Deep sequencing of the C. elegans transcriptome reveals candidate genes for major metabolic pathways. BMC Genomics 2011;12:D633.

[48] Jorth P, Ciulla DM, Huang K, Haas BJ, Izard J, Levin JZ. Metatranscriptomics of the mouse cecum sensome revealed by direct RNA sequencing. Nat Neurosci 2013;16(12):1589–97.

[49] Szurek EM, Feller DG, Liu S, Gray MA, Cogburn EE, et al. The placental transcriptome is altered during placental insufficiency. PLoS One 2013;8(8):e72788.

[50] Breitstein S, TumiJler B, Romling U. Pulsed field gel electrophoresis of bacterial DNA is isolated directly from patients’ sputa. Nucleic Acids Res 1995;23(18):3722–3.

[51] ShaK S, Capon DJ, Hellmiss R, Marsters SA, Baker CL. Reconstituent human DNA I reduces the viscosity of cystic fibrosis sputum. Proc Natl Acad Sci U S A 1990;87(20):10880–4.

[52] Lethem MI, James SL, Marriott C, Burke JF. The origin of DNA associated with mucous glycoproteins in cystic fibrosis sputum. Eur Respir J 1990;3(1):19–23.

[53] 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.
Yilmaz S, Allgaier M, Hugenholtz P. Multiple displacement amplification of microbial communities. Environ Microbiol 2011;13(5):1428–40.

Thurber RV, Haynes R, Breitbart M, Waglewy R. Laboratory procedures to culture single-stranded DNA viruses from rice paddy soil. Appl Environ Microbiol 2011;77(7):2325–52.

Chaffari N, Sanchez-Flores A, Doan R, Garcia-Dorazio KD, Chen PL, Ochoa-Levya A, et al. Novel transcriptome assembly and improved annotation of the whitelipped shrimp (Liptopeneus vannamei), a dominant crustacean in global seafood mariculture. Mar Biotechnol 2014;16(6):1019–31.

Luria R, Sela N, Yaari M, Feygenson O, Kobiler L, Lers A, et al. De novo assembly of mango peel transcriptome reveals mechanisms of mango response to hot water treatment. BMC Genomics 2014;15:547.

Rohwer F, Thurber RV. Viruses manipulate the marine environment. Nature 2009;471(7346):174–80.

Perez-Brocal V, Garcia-Lopez R, Vazquez-Castellanos JF, Nos P, Beltran B, Latore A, et al. Study of the viral and bacterial community associated with Crohn’s disease: a metagenomic approach. Clin Transl Gastroenterol 2013;4:e36.

Williamson SJ, Allen LZ, Lorenzi HA, Fadrows DW, Brami D, Thigairajan M, et al. Metagenomic exploration of viruses throughout the Indian Ocean. PLoS One 2012;7(10):e42047.

Yu JM, Li JS, Ao YY, Duan ZJ. Detection of novel viruses in porcine fecal samples from China. Virol J 2013;10:39.

Methlin C. A framework for human microbiome research. Nature 2012;486(7402):215–21.

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(5676):66–74.

Treongen TJ, Sommer DD, Angley FE, Koren S, Prog M. Next generation sequence assembly with AMOS. Curr Protoc Bioinformatics 2011:111–8.

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.

Yang X, Chateflos P, Gareau S, Coole MG, Lennon NJ, Levin J, et al. De novo assembly of highly diverse viral populations. BMC Genomics 2012;13:475.

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.

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.

Lei YY, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: human gut microbiome associated with obesity. Nature 2006;444(7122):1022–3.

Endt K, Stecher B, Chaffron S, Slack E, Titchhonc N, Benecke A, et al. The microbiome necrobiome pathogen clearance in mice reveals the gut lumen after non-typhoidal Salmonella diarrhoea. PLoS Pathog 2010;6(6):e1000197.

Fukuda S, Toh H, Taylor TD, Ohno H, Hattori M. Acetate-producing bifidobacteria protect the host from enteropathogenic infection via carbohydrate transporters. PLoS Pathog 2012;5(9):e1003149.

Maneyrol CL, Elson CO, Hatton RD, Weaver CT. Reciprocal interactions of the intestinal microbiota and immune system. Nature 2012;489(7415):231–41.

Vaidh S, Saccheri F, Migot G, Yamazaki T, Dahlere R, Hannani D, et al. The intestinal microbiota is associated with anticancer immune effects of cyclophosphamide. Science 2013;342(6161):971–6.

Tremaroli V, Backhed F. Functional interactions between the gut microbiota and host metabolism. Nature 2012;489(7415):242–9.

Can PD. Metabolism in 2013: the gut microbiota manages host metabolism. Nat Rev Endocrinol 2014;10(2):74–6.

Clarke G, Stillling RM, Kennedy PJ, Stanton C, Cryan JF, Dinan TG, et al. Microbiome–gut–brain axis: the neglected endocrine organ. Mol Endocrinol 2012;26(8):1221–38.

Somer M, Backhed F. The gut microbiota—masters of host development and physiology. Nat Rev Microbiol 2011;13(4):227–38.

Clemente JC, Ursell LK, Parfrey LW, Knight R. The impact of the gut microbiota on health. Annu Rev Med 2010;61:257–67.

Owyang C, Wu GD. The gut microbiome in health and disease. Gastroenterology 2014;146(6):1433–6.

Abel S, Roberstad SR, Bjørgedal LS, Lie MB, Landboe T, Simonsen A, et al. The intestinal microbiota is associated with the anticancer effects of cyclophosphamide. Science 2013;342(6161):971–6.

Tremaroli V, Backhed F. Functional interactions between the gut microbiota and host metabolism. Nature 2012;489(7415):242–9.

Cañ P. Metabolism in 2013: the gut microbiota manages host metabolism. Nat Rev Endocrinol 2014;10(2):74–6.

Clarke G, Stillling RM, Kennedy PJ, Stanton C, Cryan JF, Dinan TG, et al. Microbiome–gut–brain axis: the neglected endocrine organ. Mol Endocrinol 2012;26(8):1221–38.

Somer M, Backhed F. The gut microbiota—masters of host development and physiology. Nat Rev Microbiol 2011;13(4):227–38.

Clemente JC, Ursell LK, Parfrey LW, Knight R. The impact of the gut microbiota on health. Annu Rev Med 2010;61:257–67.

Owyang C, Wu GD. The gut microbiome in health and disease. Gastroenterology 2014;146(6):1433–6.

Abel S, Roberstad SR, Bjørgedal LS, Lie MB, Landboe T, Simonsen A, et al. The intestinal microbiota is associated with the anticancer effects of cyclophosphamide. Science 2013;342(6161):971–6.

Tremaroli V, Backhed F. Functional interactions between the gut microbiota and host metabolism. Nature 2012;489(7415):242–9.

Cañ PD. Metabolism in 2013: the gut microbiota manages host metabolism. Nat Rev Endocrinol 2014;10(2):74–6.

Clarke G, Stillling RM, Kennedy PJ, Stanton C, Cryan JF, Dinan TG, et al. Microbiome–gut–brain axis: the neglected endocrine organ. Mol Endocrinol 2012;26(8):1221–38.

Somer M, Backhed F. The gut microbiota—masters of host development and physiology. Nat Rev Microbiol 2011;13(4):227–38.

Clemente JC, Ursell LK, Parfrey LW, Knight R. The impact of the gut microbiota on health. Annu Rev Med 2010;61:257–67.

Owyang C, Wu GD. The gut microbiome in health and disease. Gastroenterology 2014;146(6):1433–6.

Abel S, Roberstad SR, Bjørgedal LS, Lie MB, Landboe T, Simonsen A, et al. The intestinal microbiota is associated with the anticancer effects of cyclophosphamide. Science 2013;342(6161):971–6.

Tremaroli V, Backhed F. Functional interactions between the gut microbiota and host metabolism. Nature 2012;489(7415):242–9.

Cañ PD. Metabolism in 2013: the gut microbiota manages host metabolism. Nat Rev Endocrinol 2014;10(2):74–6.

Clarke G, Stillling RM, Kennedy PJ, Stanton C, Cryan JF, Dinan TG, et al. Microbiome–gut–brain axis: the neglected endocrine organ. Mol Endocrinol 2012;26(8):1221–38.
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.

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.

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].

Hitch G, Pratten J, Taylor PW. Isolation of bacteriophages from the oral cavity. Lett Appl Microbiol 2004;39(2):215–9.

Rohwer F. Global phage diversity. Cell 2003;113(2):141.

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].

Canchaya C, Fournous G, Brussow H. The impact of prophages on bacterial chromosomes. Mol Microbiol 2004;53(1):9–18.

Filee J, Forterre P, Laurent J. The role played by viruses in the evolution of their hosts: a view based on informational protein phyllogenies. Res Microbiol 2003;154(4):237–43 [1102].

Ali-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.

Stern A, Mick E, Tirouh 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.

Abeles SR, Pride DT. Molecular bases and role of viruses in the human microbiome. J Mol Biol 2014;426(23):3892–906.

Ventura M, Sozzi T, Turroni F, Matteuzzi D, van Sinderen D. The impact of bacteriophages on probiotic bacteria and gut microbiota diversity. Genes Nutr 2011;6(3):205–7.

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.

Ursell LK, Knight R. Xenobiotics and the human gut microbiome: metatranscriptomics reveal the active players. Cell Metab 2013;17(3):317–8.

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.

Buffie CG, Jarchum I, Equinda M, Lipuma I, Gobourne 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.

Tormo-Badia N, Hakansson A, Vasudevan K, Molin G, Ahne S, Cilio CM. Antibiotic treatment of pregnant non-obese diabetic mice leads to altered gut microbiota and intestinal immunological changes in the offspring. Scand J Immunol 2014;80(4):250–60.

Chen J, Novick RP. Phage-mediated intergeneric transfer of toxin genes. Science 2009;323(5910):139–41.

Ubeda C, Malques 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.

De Paepe M, Leduc M, Tinsley CR, Petit MA. Bacteriophages: an underestimated role in human and animal health? Front Cell Infect Microbiol 2014;4:39.