

Alan W. Walker

Abstract

There are a range of methodologies available to study the human microbiota, ranging from traditional approaches such as culturing through to state-of-the-art developments in next generation DNA sequencing technologies. The advent of molecular techniques in particular has opened up tremendous new avenues for research, and has galvanised interest in the study of our microbial inhabitants. Given the dazzling array of available options, however, it is important to understand the inherent advantages and limitations of each technique so that the best approach can be employed to address the particular research objective. In this chapter we cover some of the most widely used current techniques in human microbiota research and highlight the particular strengths and caveats associated with each approach.

Keywords

Microbiota • Techniques • Sequencing • PCR • FISH • Stable isotope • Metabolomics • Proteomics

2.1 Introduction

The Nobel prize winning biologist Sydney Brenner once remarked that “progress in science results from new technologies, new discoveries and new ideas, probably in that order” (Robertson

1980) and this sentiment has undoubtedly been well exemplified in the field of microbiota research. Study of the human microbiota can be traced back to Antonie van Leeuwenhoek’s late Seventeenth Century description of “animalcules” in scrapings from the human mouth (Porter 1976), a discovery that was made possible by van Leeuwenhoek’s ground-breaking work with microscopes. From the pioneering endeavours of Cohn, Pasteur, Koch and others in the Nineteenth Century, through to developments in anaerobic microbiology and molecular biology in the

A.W. Walker (✉)

Microbiology Group, Rowett Institute of Nutrition and Health, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, UK
e-mail: alan.walker@abdn.ac.uk

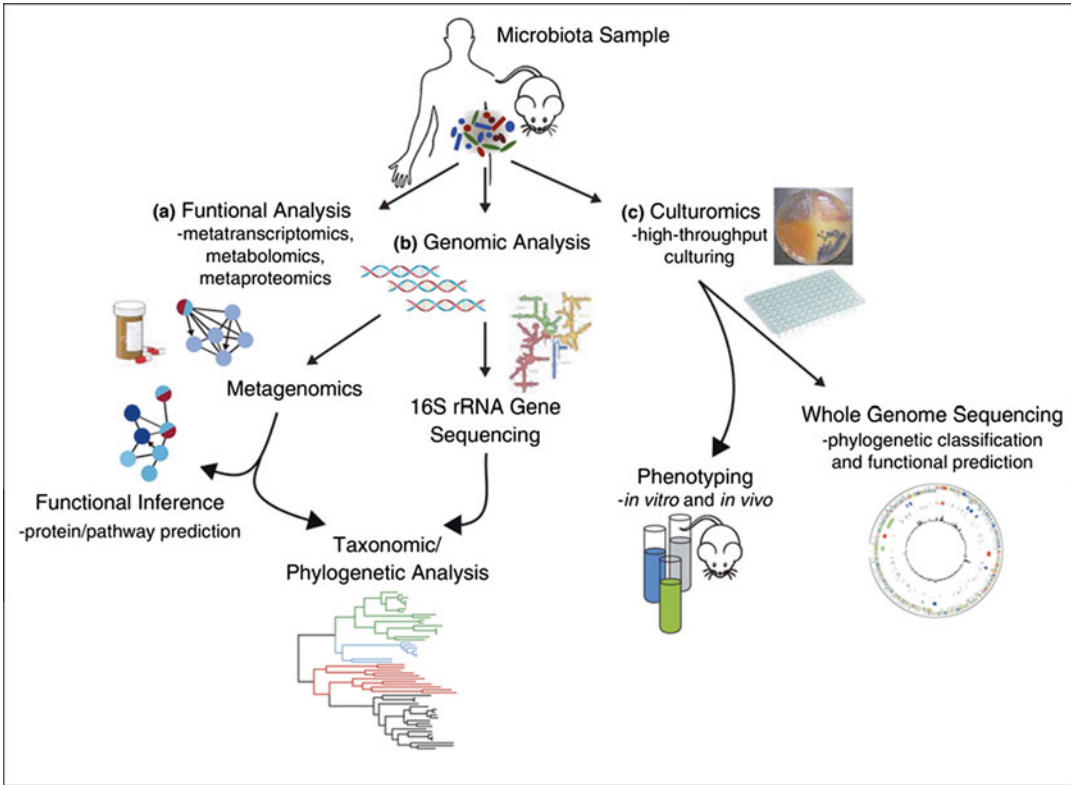


Fig. 2.1 Overview of some of the most common techniques used to study the human microbiota

(a) The functional activities of the microbiota can be studied by monitoring transcription (using RNA-seq/metatranscriptomics), protein production (metaproteomics) or metabolite production (metabolomics). (b) DNA sequence-based techniques are used to determine the composition of the microbiota (e.g. 16S rRNA gene surveys) and the functional encoding capabilities of the microbiome (shotgun metagenomics). (c) Culture remains highly relevant as cultured organisms can be studied in

depth in the laboratory or in animal hosts. Recently, the term “culturomics” has been applied to high-throughput culturing of microbes in multi-welled plates containing highly nutritious growth media. Cultured organisms can also have their genomes sequenced, providing further information about their potential activities *in vivo*. These techniques can be used in combination to generate more comprehensive understandings of the human microbiota. Reprinted in unmodified form from: Pham and Lawley (2014) (Pham and Lawley 2014) under Creative Commons Attribution (CC BY) license

second half of the Twentieth Century, and the Twenty-first Century’s own breakthroughs in genomics and DNA sequencing technologies (McPherson 2014), subsequent developments in the field of microbiota research have been similarly driven by successive waves of technological and methodological advances. As a result, today’s microbiota researcher has the benefit of a staggering array of tools at their disposal (Fig. 2.1). This chapter gives a broad overview of the many techniques that are now available, and attempts to describe the inherent advantages and limitations of each of these techniques.

2.2 Classical Microbiological Methods

2.2.1 Culture

For well over a 100 years microbiologists have used the classical approaches of cultivating microbes in the laboratory, isolating individual colonies and then studying these isolated strains in order to describe their phenotypic characteristics and metabolic capabilities (see Lagier et al. (2015a) for a recent overview of the techniques used). As a result of these extensive efforts, it has

been estimated that over 1000 distinct microbial species have been cultured from the human gastrointestinal tract alone (McPherson 2014), and characterisation of microbes and gene function discovery in the laboratory remains the bedrock upon which many of the more modern molecular techniques that will be described in later sections of this chapter rely upon. A further advantage of having a strain in culture is that it allows potential exploitation for therapeutic purposes should it turn out to have beneficial properties (Walker et al. 2014).

The simplest form of microbial cultivation is to incubate samples or individual strains in batch culture in nutritious or selective growth media. Batch culture studies allow selective enrichment of bacterial groups of interest, comparisons to be made between growth rates and metabolite production on different substrates, and interactions between specific species to be observed and measured (Belenguer et al. 2006). Many microbial inhabitants of humans are obligately anaerobic and therefore exquisitely sensitive to oxygen. As a result, some species can be killed by even very brief exposure to air (Flint et al. 2007), making them much more difficult to grow. To permit laboratory cultivation of these species, culturing must therefore be carried out under strictly anaerobic conditions, for example by using anaerobic cabinets or Hungate roll tubes (Eller et al. 1971). Cultivation of particularly fastidious gut species can also be enhanced by using media containing rumen fluid, filtered stool extracts, or mixtures of short chain fatty acids, which can be utilised by some gut bacteria as growth substrates (Duncan et al. 2002; Lagier et al. 2015b).

A limitation of batch culture is that results can only be obtained over relatively short periods of time before the supply of nutrients in the growth medium is exhausted or toxic by-products accumulate and lead to cessation of microbial growth (Ferenci 1999). A further, and key, disadvantage to using culture is that it is highly labour intensive, and a range of complex growth media are typically required to recover as wide a diversity of organisms from a sample as possible. It is also known that many of the microbial species that inhabit the human body have yet to be grown in

the laboratory (Rajilic-Stojanovic et al. 2007). This problem is particularly acute for bodily sites such as the colon, where the majority of the constituent bacteria are strict anaerobes. As such, culture alone cannot address the sheer complexity of the human microbiota.

Nonetheless, there are many reasons to be optimistic that cultured coverage of the human microbiota can be greatly improved. DNA-sequence based surveys of the gut microbiota, for example, commonly show that many of the most abundant sequences map to cultured species, and that it is the rarer sequences that are less likely to be derived from a cultured isolate (Walker et al. 2014). This suggests that it is insufficient culturing effort rather than an inherent “unculturability” that is the main barrier to successful novel isolations. Furthermore, unlike environments such as soil, which can harbour very slow growing microbes, bacteria living in the human body are often provided with relatively stable environmental conditions, and a generally reliable supply of growth nutrients, and must therefore be capable of multiplying quickly or else face being rapidly outcompeted. Provided the correct conditions can be supplied in artificial growth media it can be assumed therefore that these species will be relatively more amenable to culture. Indeed, novel species continue to be regularly isolated from the human microbiota, and there have been some impressive recent examples of successful high-throughput culturing programmes (Lagier et al. 2015b; Goodman et al. 2011). Such efforts have been dubbed “culturomics”, and have contributed to a reinvigorated interest in the use of culture-based techniques to better characterise the human microbiota. Information gleaned from modern genomics methods can also be used to design improved culture media that support the growth of previously uncultivated species (Bomar et al. 2011).

2.2.2 Continuous Culture

A more sophisticated method to cultivate microbes in the laboratory is the use of continuous culture model systems such as fermentors

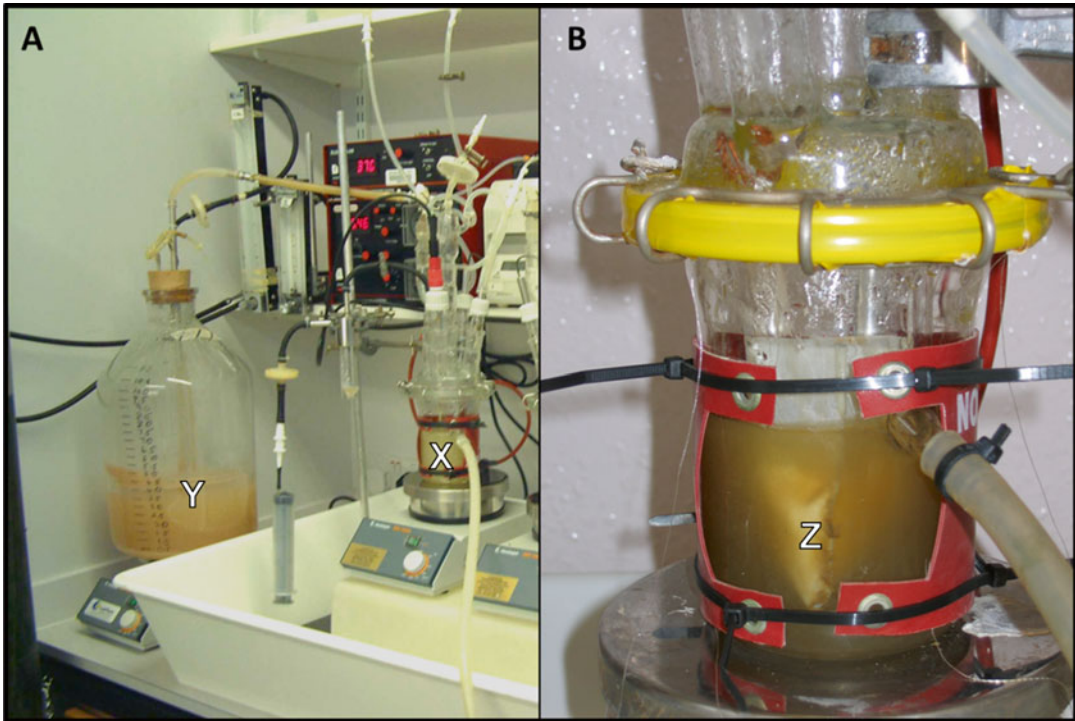


Fig. 2.2 Continuous culture fermentor system

Fermentors are continuous culture model systems, which allow long term cultivation of microbes. (A) An example of a single vessel fermentor system (the culture vessel is labelled with “X”), inoculated with human faeces and fed a constant supply of nutritious growth medium (labelled “Y”). The contents of the culture vessel are gassed with

CO₂ or N₂ to ensure that they remain anaerobic, and can be maintained at defined pH and temperatures, which are constantly monitored. (B) a modified fermentor vessel, incorporating a nylon bag containing insoluble particulate substrates (labelled “Z”), developed to identify fibre-degrading gut bacteria

(Fig. 2.2). In contrast to the batch approach, continuous culture is carried out in an open system, which is continually supplied at one end with fresh growth medium/nutrients, and overflow is allowed to drain from the vessel at the other end, diluting out toxic metabolic by-products and dead cells. Systems such as these reach a “steady state” equilibrium, allowing the researcher to exert an enhanced level of control over prevailing environmental conditions within the culture vessel, and can therefore be run over relatively long time periods (Miller and Wolin 1981). These sort of systems have been commonly used to study colonic microbes, and a number of research groups have made fermentors more advanced by incorporating distinct sequential stages, which aim to mimic the sort of environmental changes microbes are exposed to as they pass along the

length of the gastrointestinal tract (Van den Abbeele et al. 2010). While these model systems are an advance over simple batch culture it should be noted, however, that they still have important limitations. For example, they lack an immune system, and metabolites such as short chain fatty acids (SCFAs) produced by the bacteria are not absorbed, meaning results may not necessarily be directly translatable to the situation *in vivo*.

2.2.3 Animal Models

Microbes of interest can also be cultivated and maintained in animal models. Until relatively recently, for example, the only way to grow segmented filamentous bacteria, which have been shown to have important pro-inflammatory

effects in mice, was in animal models (Klaasen et al. 1991). One disadvantage of using animal models is that, while the microbiota composition at the phylum level generally appears to be similar between humans and other animals, at the species and strain level there is considerable divergence, likely to be due to underlying differences in host anatomy/physiology, and dietary regimes (Nguyen et al. 2015). However, recent work has shown that it may be possible to mitigate this issue somewhat as a significant proportion of human-associated bacterial species appear to be able to successfully colonise the intestines of animal models following faecal microbiota transfer (Ellekilde et al. 2014). Germ-free, or gnotobiotic, mice are another appealing option as these mice can be specifically inoculated with microbial strains of interest (Goodman et al. 2011; Seedorf et al. 2014). This permits a more reductionist approach to study host-microbe and microbe-microbe interactions, separated from the potentially perplexing background complexity of the wider microbiota. A further particular advantage of using mouse models is that extensive genotyping analyses have been carried out, and there are a range of knockout mouse lines available to allow the study of interactions between specific host genetic components and the microbiota (Kostic et al. 2013).

There are, however, a number of important limitations to using animal models, particularly rodent models. For example, co-housing, and the practice of coprophagy, generally leads to rapid transfer of microbiota between cage mates, and this can confound results by being a stronger determinant of intestinal microbiota composition than either host genotype or experimental variables (Lees et al. 2014; Ericsson et al. 2015). Furthermore, recent work has indicated that rodents who are handled by male experimenters are likely to be more stressed than those handled by females (Sorge et al. 2014), and it is possible that stress may impact microbiota structure (Cryan and Dinan 2012). Finally, emerging evidence suggests that host diet may have a greater impact on microbiota structure and composition in rodents than in humans (explaining around 60 % of variance vs 10 % respectively), raising

concerns as to whether or not rodent models are most appropriate for studies investigating links between the microbiota and, for example, diet-dependent diseases such as obesity (Salonen et al. 2014). A recent review by Nguyen et al. (2015) extensively documents the inherent advantages and disadvantages of using mouse models, and discusses the translatability of findings in mice to humans.

2.3 Sequence-Based Approaches

While culture remains an important tool, human microbiota research has been completely revolutionised over the last decade by molecular methods, and in particular by the falling costs and vastly increased throughput of DNA sequencing technologies (Fig. 2.3). This rapidly moving, and highly innovative, field continues to produce exciting and novel technologies, with the latest generation of sequencing machines capable of generating data at a depth of billions of individual sequence reads (Illumina HiSeq), or at comparatively long read lengths (PacBio), or even via miniaturised devices that can be plugged into the USB port of a laptop (Oxford Nanopore's MinION) (Reuter et al. 2015).

The key advantage to sequence-based approaches is that, by circumventing the requirement to grow microorganisms in the laboratory, they generally give much more comprehensive overviews of the species present in a sample. They are also typically far less labour intensive than classical microbiological techniques, and as a result it is now possible to carry out experiments at a scale that would have been unthinkable just a decade ago. Indeed, recent global research initiatives such as the Human Microbiome Project (HMP) and MetaHIT, for example, have taken advantage of these new sequencing technologies to produce staggering amounts of freely available data (Human Microbiome Project Consortium 2012a; Li et al. 2014). There are a number of ways in which the power of DNA sequencing can be used to study the human microbiota, which are detailed in the following text. In addition, Table 2.1 summarises

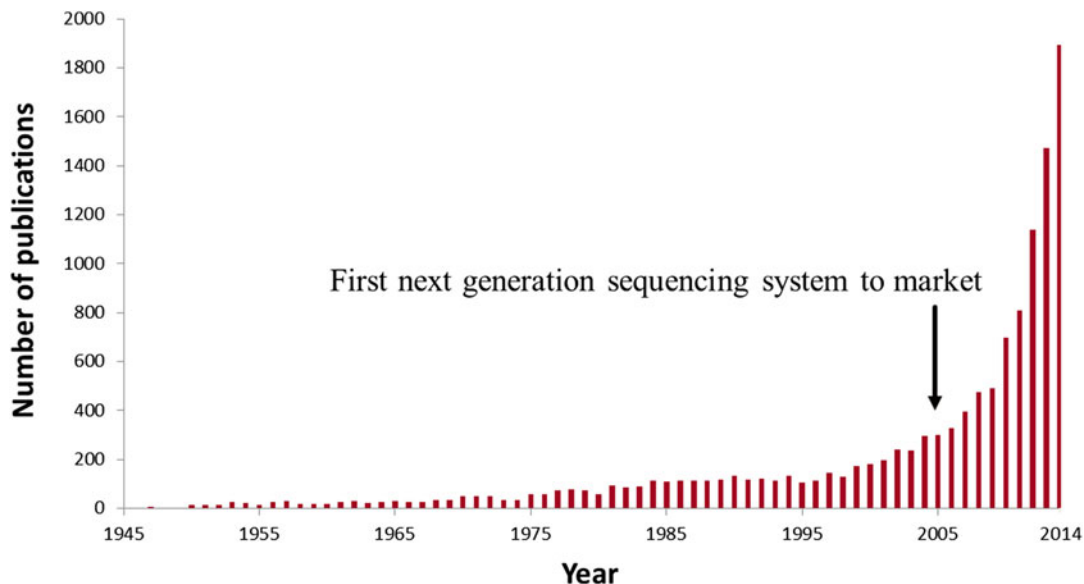


Fig. 2.3 DNA sequencing approaches have revolutionised microbiota research

Chart showing the meteoric rise in publications mentioning the gut microbiota since the advent and market release of next generation sequencing platforms such as 454 pyrosequencing and Illumina. Data collected by searching

Pubmed (search date Dec. 1st, 2014) for the terms “gut flora” OR “gut microflora” OR “gut microbiota” OR “gut microbiome” OR “intestinal flora” OR “intestinal microflora” OR “intestinal microbiota” OR “intestinal microbiome” OR “colonic flora” OR “colonic microflora” OR “colonic microbiota” OR “colonic microbiome”

the most common uses, and outlines the inherent advantages and limitations of each approach.

2.3.1 Marker Gene Surveys

One common sequence-based approach is to carry out surveys of universal marker genes, which provide a broad census of the microbial species present within a sample. While these sort of surveys have been carried out since the 1980s recent developments in next generation sequencing technologies mean it is possible to survey microbial communities at previously unimaginable depth and scales (Tringe and Hugenholtz 2008; Caporaso et al. 2011). The most widely used universal marker genes are the small subunit ribosomal RNA (SSU rRNA) genes (16S rRNA gene for bacteria and archaea, 18S rRNA gene for eukaryotes). Within these genes there are regions of DNA sequence that are highly conserved, but there are also other regions that are more variable, and which are unique to certain

microbial groups or genera (Woese and Fox 1977). Following DNA extraction from the human tissue sample, the SSU rRNA genes are typically PCR-amplified using primers targeted towards highly conserved regions of the gene. The aim here is to generate a mixed pool of PCR amplicons that are derived from as many of the bacterial species present in the original sample as possible, which are then sequenced *en masse*. Typically, the resulting data is then clustered by sequence similarity into Operational Taxonomic Units (OTUs), with the assumption being that these OTUs will be a reasonable approximation of the underlying species content of a given sample. It should be noted though that, due to the wide variation in 16S rRNA gene operon copy numbers between individual strains, results are not truly quantitative (Vetrovsky and Baldrian 2013). Furthermore, the chosen OTU sequence similarity threshold is both artificial and subjective and will not be able to accurately capture diversity correctly across the full range of genera present in a sample. Nonetheless, when the full

Table 2.1 Comparison between different sequence-based approaches used to study the human microbiota

Method	Advantages	Limitations
Species profiling via marker gene surveys (e.g. 16S rRNA gene)	Provides overview of species present in a sample	Relatively insensitive – can be impossible to derive species-level classifications for some genera
	Much cheaper than other sequencing methods	Only provides information on community composition, does not provide direct functional capability data
	Analysis requires less computational power	Single marker genes such as 16S rRNA genes typically only describe the bacterial/archaeal fraction of microbial communities. Does not describe viruses, fungi etc. that may also be present.
	Larger sample sets increase statistical power	Results can be heavily impacted by sampling, storage, PCR and DNA extraction biases
	Broad functional capabilities can often be inferred from 16S rRNA gene sequences by comparing to closely related isolates with fully sequenced genomes	16S rRNA gene is usually multi-copy, and the number of copies is variable between species, meaning results are not truly quantitative Usually does not discriminate between active and inactive/dead cells
Whole genome sequencing	Provides information on the complete coding potential of an organism	Usually requires that the organism be cultivated prior to sequencing the genome
	Draft bacterial genomes can now be generated very quickly and cheaply	Modern, short read, sequencing technologies will typically generate draft, not complete, genomes
	Data generated can be used for epidemiological purposes, e.g. for strain typing	Many constituent genes will be of unknown function
Metagenomics	Allows simultaneous profiling of both the functional capabilities and species composition of microbial communities	Can require very deep sequencing to achieve reasonable genome coverage, making it comparatively expensive
	Can simultaneously obtain genomic data of bacterial, archaeal, eukaryotic and viral origin	Often limited to small numbers of samples, which reduces statistical power
	Complete genomes of constituent species, including uncultured organisms, can be assembled	Data analysis may require large computational resources.
	No PCR bias	Assembling genomes can be challenging
		Gaps in reference databases mean that large proportions of the genomic data are of unknown function Biases introduced during sampling, storage and DNA extraction can impact results Usually does not discriminate between active and inactive/dead cells

(continued)

Table 2.1 (continued)

Method	Advantages	Limitations
Single-cell genomics	Provides genomic data from uncultured species	Isolating single cells typically requires access to expensive equipment (e.g. flow cytometry, micromanipulators)
	Allows placement of genomic data in a phylogenetic context	Genome amplification step introduces biases, making complete genome assembly challenging
	Data generated from uncultured species can improve reference databases for metagenomic analyses.	Sensitivity of the genome amplification step means that contamination is a constant concern and must be mitigated against
		Biases introduced during sampling, storage and DNA extraction can impact results
		Cannot discriminate between active and inactive/dead cells
Metatranscriptomics	Gives data on the functional activity of microbial communities	Short half-life of mRNA is an important limitation; sample selection and preservation are key concerns
	Focusses on the active members of the microbiota, results not as impacted by dead/inactive cells as other sequencing methods	Can be technically challenging, often need to deplete the far more abundant rRNA before sequencing mRNA
	Can often attribute source organisms to transcripts	Gaps in reference databases mean that large proportions of the genomic data are of unknown function
		Biases introduced during sampling, storage and RNA extraction can impact results

1500 bp sequence of the 16S rRNA gene is available, clustering into OTUs with 98.7–99% sequence similarity appears to best fit species-level designations derived from culture work (Stackebrandt and Ebers 2011). However, as next generation sequencing technologies typically generate comparatively short read lengths, which are focussed on hyper-variable regions of the gene, slightly less stringent clustering is required and it is now most common to cluster OTUs with 97% sequence similarity (Schloss and Westcott 2011).

Regardless of sequence similarity used, OTUs can be mapped against comprehensive reference databases such as SILVA, RDP, EzTaxon and Greengenes in order to assign taxonomic classifications to them (Quast et al. 2013; Cole et al. 2014; Chun et al. 2007; DeSantis et al. 2006). This provides information about which taxa were present in the original sample, and allows the researcher to monitor differences in microbiota composition between samples and between study cohorts. A range of software options are now

available, such as mothur, QIIME, VAMPS and GUSTA ME (Schloss et al. 2009; Caporaso et al. 2010; Huse et al. 2014; Buttigieg and Ramette 2015) which allow the researcher to carry out all of the stages involved in processing marker gene survey data, from quality control steps to statistical comparisons and visualisation of results.

While broad marker-gene surveys using universal markers such as the 16S rRNA gene are the most commonly applied variation of this technique it is also possible to carry out focussed surveys of functional genes that have more limited dissemination throughout the microbiota (Walker et al. 2014). The principle here is similar; degenerate PCR primers are targeted towards conserved regions of these functional genes, creating a mixed pool of amplicons, which are then sequenced. This approach has been used, for example, to identify novel groups of butyrate/propionate producing bacteria from the human colon, and cellulolytic bacteria from the rumen (Louis et al. 2010; Reichardt et al. 2014; Brulc et al. 2011). A disadvantage of this targeted

approach is that the PCR primers may not efficiently amplify all of the functional genes of interest in a given sample. Untargeted approaches such as metagenomics (see “*Metagenomics*” section below) may circumvent this issue, but at the cost of having to generate far greater amounts of data, which is considerably more expensive to produce and more difficult to analyse (Prakash and Taylor 2012).

2.3.2 Whole Genome Sequencing

The first bacterial genome to be completely sequenced was that of *Haemophilus influenzae*, in 1995 (Fleischmann et al. 1995). Then, sequencing was carried out using the traditional Sanger method (Sanger et al. 1992) and it took many years, and hundreds of thousands of dollars, to complete a whole bacterial genome. Advances in DNA sequencing technology since then mean that draft bacterial genomes can now be generated in a matter of hours, and at a cost that is thousands of times cheaper (Loman et al. 2012; Koser et al. 2012). Given the extremely high-throughput nature of next generation sequencing platforms such as Illumina it is now common to simultaneously sequence many microbial genomes on a single sequencing run. This is done by multiplexing samples via the addition of a unique sequence “tags” and then bioinformatically separating reads from each of the combined samples post-sequencing (Lennon et al. 2010). “Shotgun” sequencing, whereby DNA is randomly fragmented prior to sequencing and then the resulting overlapping sequence data is pieced together bioinformatically into contiguous stretches (contigs), is the standard method (Fleischmann et al. 1995). Genomes are typically pieced together by either mapping data on to an existing reference genome (if one is available) or by assembling the data *de novo*. There is a wide range of software available for the genome assembly step, with the optimal choice of assembler depending on the sequencing platform used (Loman et al. 2012).

There are now a large, and constantly increasing, number of genomes available from human-

associated microbes. The Human Microbiome Project alone, for example, aims to have generated over 3000 draft genomes once the first phase is complete (Human Microbiome Project Consortium 2012b). Genome sequence data provides critical information on the putative functional capabilities of a given species, although it should be acknowledged that there are often a large number of unannotated genes due to paucity of close, well-characterised matches in reference databases. Indeed, even with *E. coli* K-12, which has been extensively studied and used as a model organism over many decades, around a quarter of the constituent genes remain unannotated (Conway et al. 2014). Nonetheless, as reference databases expand, and techniques for high-throughput, genome-wide, functional probing such as transposon insertion sequencing are developed (van Opijnen and Camilli 2013), this situation will improve. A further, and flourishing, use for whole genome sequencing is in the field of epidemiology, and there are now numerous examples of using whole genome sequence data to trace both global and local dissemination of microbes within human populations (Parkhill and Wren 2011; Eppinger et al. 2014), and to monitor evolutionary changes in genomic content (He et al. 2010; Schuenemann et al. 2013).

2.3.3 Metagenomics

An important limitation of whole genome sequencing is that it typically requires the organism to be grown in culture first, so that enough DNA can be extracted for subsequent sequencing. 16S rRNA gene-based surveys have revealed, however, that the majority of human microbiota species have yet to be cultivated in the laboratory (Eckburg et al. 2005). As a result complementary methods such as metagenomics, which can provide genomic insights into this uncultured majority, are attractive options, and have gained increasing favour in recent years. With metagenomics, the researcher directly shotgun sequences DNA extracted from an environmental sample. They then either attempt to bioinformatically piece together the resulting sequence data, which

will be comprised of fragments of DNA derived from the range of different species that were present in the original sample, into contiguous stretches of sequence data derived from each individual constituent species, or use the unassembled sequence data directly as a means to assess the functional capabilities of the microbial community as a whole entity (Handelsman 2004).

Metagenomic sequencing in this manner was first applied to samples from the human gut in 2006 (Gill et al. 2006), and has since been used numerous times to study the human microbiota. This technique can be hugely powerful, and it is possible to generate in depth profiles of the functional potential of a given microbial community, including uncultured constituents. It is important to note, however, that the high complexity of many human-associated microbial habitats, such as the colon, means that very deep sequencing is often required in order to generate sufficient sequence data from a representative cross-section of the microbes that are present. Luckily, the development of next-generation sequencing platforms such as Illumina mean that this is now possible, and large-scale metagenomics studies incorporating many individual samples are now being carried out (Hu et al. 2013). Metagenomics is also the only technique that allows effective, in depth, monitoring of the viral communities (or “viromes”) that are present in the human body as there are no marker genes equivalent to SSU rRNA that are universally detected in all viruses and so can be used for sequence-surveys (Minot et al. 2011). Further key advantages of metagenomics over other sequence-based techniques are outlined in Table 2.1.

There are, however, some important limitations to the use of metagenomics. For example, this sort of study is far more expensive than marker gene surveying, and comes with a requirement for appropriate computational infrastructure and expertise in order to be able to process the data effectively. Unfortunately, these factors mean that sample sizes tend to be quite small, and large-scale metagenomics studies are currently out of reach for many laboratories. This situation

will likely improve though as sequencing costs fall and the use of cloud computing facilities becomes more wide-spread (Angiuoli et al. 2011). As with other DNA-based approaches, the sample storage, preparation, and processing methodologies used will also have significant impacts on the quality of the final metagenomics data (see section “*Common pitfalls of sequence based approaches*” below).

The task of assembling genomes from such a complex collection of microbes, where there will also be great divergences in genome coverage depth based on the relative abundance of each species in the original sample, is also daunting, particularly when trying to assemble genomes from closely related strains and species, or highly fragmented genomes where there is only limited coverage (Nielsen et al. 2014). Although these issues have still not been completely surmounted, there have been great improvements in this area in recent years, and various bioinformatics tools have been developed to aid the genome assembly and species assignment processes (Peng et al. 2011; Namiki et al. 2012; Bankevich et al. 2012; Alneberg et al. 2014).

A further concern is that the current reference databases that are routinely used to classify the DNA sequences are not comprehensive enough. As a result, a large fraction of metagenomics data often goes uncharacterised as there are simply no close matches in the reference database to base a classification on (Thomas et al. 2012). This also means that results tend to be heavily weighted towards well characterised housekeeping genes, which are comparatively well covered in reference databases (Walker et al. 2014). This situation will improve, however, as novel gene functions and pathways are continually elucidated, and reference databases incorporate genomes from a more phylogenetically diverse array of isolates (Walker 2014).

2.3.4 Single-Cell Genomics

Single cell genomics (SCG) is an emerging and complementary technique to metagenomics, and

is a more targeted approach to generating genomes from uncultured microbes. With this technique, individual microbial cells are isolated from environmental samples, and their genomic DNA subsequently amplified by a whole genome amplification technique (typically multiple displacement amplification) (Walker and Parkhill 2008). This exquisitely powerful amplification step generates sufficient DNA from just a single cell that subsequent shotgun sequencing becomes feasible (Blainey 2013). Moreover, combining SCG with a form of targeted cell selection, such as fluorescent *in situ* hybridisation (Amann and Fuchs 2008), stable isotope probing or Raman microspectroscopy, allows the researcher to potentially recover specific cells that are derived from a particular phylogenetic background, or that carry out a function of interest. As such, SCG complements metagenomics by allowing recovery of genomic information from species that may be rare in the microbial community, and allows the researcher to understand which organisms are capable of carrying out a particular function, even if the genes that are responsible for carrying out this function are unknown or missing from reference databases (Walker et al. 2014).

There are some important limitations to this technique however (see Table 2.1), which have so far hindered wide-scale implementation. Of particular relevance are the issues of contamination (with such a small starting DNA input, any amount of contaminating DNA can easily overwhelm the sequence data that is derived from the cell of interest), and of biases introduced during the amplification step, which can confound genome assembly software, and typically mean that only partial genome coverage can be achieved (Raghunathan et al. 2005). Nonetheless, SCG has been used to characterise novel human-associated bacteria from rare and understudied phyla such as TM7 and *Chloroflexi* (Marcy et al. 2007; Campbell et al. 2014) and holds great promise for wider future application. Results generated can also greatly aid metagenomics-based analyses by broadening reference databases and providing reference genomes to aid with the assembly steps (Rinke et al. 2013).

2.3.5 Metatranscriptomics

A further emerging sequence-based technique with applicability to the human microbiota is metatranscriptomics (also termed RNA-seq). Transcriptomics is the study of the RNA transcripts produced by a given species, whereas metatranscriptomics is the study of combined transcripts from an entire microbial community. Thus, in contrast to metagenomics, metatranscriptomics allows insights into the functional *activity* of the microbiota at a given time and under prevailing environmental conditions, not just the functional *potential*. Typically, this technique involves isolating RNA from environmental samples and using this to create reverse transcribed cDNA libraries, which can then be shotgun sequenced using modern high throughput sequencing platforms such as Illumina (Reck et al. 2015). Shotgun sequenced data is then typically assembled by either mapping back to reference genomes, or by carrying out *de novo* assembly. Recent RNA-seq developments now allow strand-specific identification of transcripts, permitting enhanced detection of both messenger and non-coding RNAs, and providing new insights into the roles that the latter may play in cellular function (Croucher and Thomson 2010).

Metatranscriptomics is considerably more technically challenging than metagenomics as it requires additional processing steps such as creating cDNA and depleting host and bacterial rRNAs, which typically make up the vast majority of RNA present in a sample (Giannoukos et al. 2012). Furthermore, transcriptomics is also commonly used in combination with reference genomes, as mapping transcripts back to a reference allows the researcher to understand how a given species responds to changes in environmental conditions. Metatranscriptomic analyses of human microbiota samples are therefore rendered more complex by the fact that there are often no reference genomes available for many members of the microbial community. As such, the raw data may require complex *de novo* assembly prior to analyses, a process which has been

improved in recent years with the advent of novel software programmes (Tjaden 2015).

A key limitation of metatranscriptomics is that, due to the very short half-life of mRNA molecules (typically measured in minutes (Reck et al. 2015)), it may not always be entirely representative of microbial activities *in situ*. For example, microbial transcriptional activities measured in faecal samples may not be reflective of gene expression occurring in areas such as the proximal colon. A further limitation is that, as with metagenomics, many of the transcribed genes will be of unknown function due to extensive gaps in reference databases.

Given these inherent complexities and limitations, metatranscriptomics has yet to be applied to human microbiota samples to the same extent as metagenomics, although uptake of this technique is increasing (Jorth et al. 2014; Leimena et al. 2013; Maurice et al. 2013; Macklaim et al. 2013). Moreover, direct comparisons between metagenomic and metatranscriptomic datasets demonstrate the worth of this approach, as highly significant differences between the two datasets are detected, reflective of the fact that microbes are constantly altering their gene expression profiles in response to prevailing environmental conditions (Franzosa et al. 2014).

2.4 Common Pitfalls of Sequence Based Approaches

While sequence-based approaches have undoubtedly revolutionised the field of microbiota research there are a number of key caveats, particularly in the areas of sample handling and processing, that should be considered when applying them. Analyses of mock bacterial communities prepared for the Human Microbiome Project, for example, showed that samples clustered together based upon which of four sequencing centres generated the data, illustrating the impact that sample processing steps can have on final sequencing results (Schloss et al. 2011). Furthermore, it is clear from comparisons

between techniques that sequence-based approaches commonly “miss” a significant fraction of species present in a sample due to their inherent biases (Shade et al. 2012; Lagier et al. 2012). Awareness of these inherent limitations and biases is therefore important to ensure that erroneous conclusions are not drawn from sequence data (Degnan and Ochman 2012).

Sample preservation is a critical, and often under looked, first step. Emerging evidence suggests that prior freezing of faecal samples can lead to systematic distortions in molecular profiling results. Specifically, it appears that *Bacteroides*-derived DNA may be gradually depleted if samples have been previously held in long term frozen storage (Maukonen et al. 2012; Bahl et al. 2012). Furthermore, evidence suggests that bacterial community profiles obtained from sputum samples may be perturbed by being kept for greater than 12 h at room temperature prior to being placed in long-term frozen storage, and also by repeated freeze-thaw cycles prior to DNA extraction and sequencing (Cuthbertson et al. 2014, 2015).

DNA extraction is another key step, and it is known that choice of extraction kit/method can have major impacts on the final sequencing results obtained (Ferrand et al. 2014; Kennedy et al. 2014). If the chosen DNA extraction method is not robust enough to break open the cell walls of certain microbes then DNA from these species will not be recovered and so will not be observed in the final sequencing libraries. For this reason kits with only chemical-based extraction are not recommended, as they typically generate results with an over-abundance of the more easily extracted Gram negative organisms present in a sample compared to the more recalcitrant Gram positive organisms, which have a stronger cell wall that is less likely to be broken down by chemical lysis only (Walker et al. 2015). DNA extraction kits with a mechanical lysis, or bead-beating, step, which is far more effective at breaking open Gram positive cell walls, are therefore typically recommended (de Boer et al. 2010). However, it should be noted that some bead-beating kits are more

effective than others (see Fig. 2.4a) (Kennedy et al. 2014).

For sequence-based approaches requiring prior amplification of specific genes, such as 16S rRNA genes, PCR primer design is a further critically important consideration. It is known that certain groups, for example the Actinobacteria, are systematically under-represented in studies using the commonly used 27f primer (Frank et al. 2008). An example of this is the *Bifidobacterium* genus, typically the dominant member of the gut microbiota in breast fed infants, which has three mismatches to 27f (Fig. 2.4b), therefore this primer should not be used with infant faecal samples as results will not reflect the true microbiota content (Walker et al. 2015). Incorporating degenerate bases into primer design is one way to effectively widen the range of target organisms (Fig. 2.4b). Sim et al (2012), for example, were able to show that improved primers resulted in far better recovery of bifidobacterial sequences from infant faecal samples (Sim et al. 2012).

Primer choice is also important if there are specific groups of bacteria that a researcher is interested in. Next generation sequencing platforms currently generate relatively short reads, meaning that it is typical to target sub-sections of the 16S rRNA gene. Unfortunately, no specific variable region, or combination of variable regions, is able to fully capture the diversity that can be described with full-length 16S rRNA gene sequences. It is therefore prudent to ensure that the species of interest can be differentiated using the variable regions targeted prior to initiating a study (Fig. 2.4c).

A further complicating factor with amplification-based approaches such as marker gene surveys and single-cell genomics is that chimeric molecules can be created during the amplification step (Edgar et al. 2011). Indeed, it is estimated that a significant proportion of DNA sequences submitted to 16S rRNA gene databases, for example, may in fact be chimeric in nature (Ashelford et al. 2005). Chimeric molecules inflate microbial diversity estimates (Schloss et al. 2011), and in the case of single-

cell genomics can confound genome assembly software (Lasken and Stockwell 2007). Errors generated during the sequencing process itself can also vastly inflate diversity measures if steps are not taken to account for their impact (Huse et al. 2010). Repeated PCR cycling may also lead to an over-representation of some groups and the under-representation of others. For this reason it has been recommended that the number of PCR cycles should be kept as low as is feasible (Bonnet et al. 2002).

A further potential pitfall is the presence of contamination. Sequence-based approaches are exquisitely sensitive, which means they are an attractive means with which to investigate areas of the body traditionally thought of as “sterile”, or that have very low abundance of colonising microbes that are difficult to grow. Unfortunately, contaminating DNA or cells can be introduced to the sample of interest at many processing stages, including from reagents in common laboratory DNA extraction and PCR kits (Tanner et al. 1998) (Fig. 2.4d). Recent work by Salter et al has indicated that, when sequencing is applied to low biomass samples (i.e. sample containing less than 10^4 cells), background contamination effectively “swamps” the targeted DNA from the sample and becomes the dominant feature of sequencing results (Salter et al. 2014). Therefore, any researcher working with low biomass samples should ideally make use of copious “negative” sequencing controls. This involves running “blank” DNA extractions and PCR reactions with no sample or template added, and then sequencing these alongside the samples of interest. Any contaminating species detected in the negative controls can then be removed from the sequencing results from the actual samples.

The choice of DNA sequencing platform is a further important consideration. A recent comparative analysis between the Illumina MiSeq and Ion Torrent platforms, for example, indicated that a peculiarity of the Ion Torrent sequencing process can lead to premature truncation of sequence reads derived from certain microbial groups. The effect of this would be to

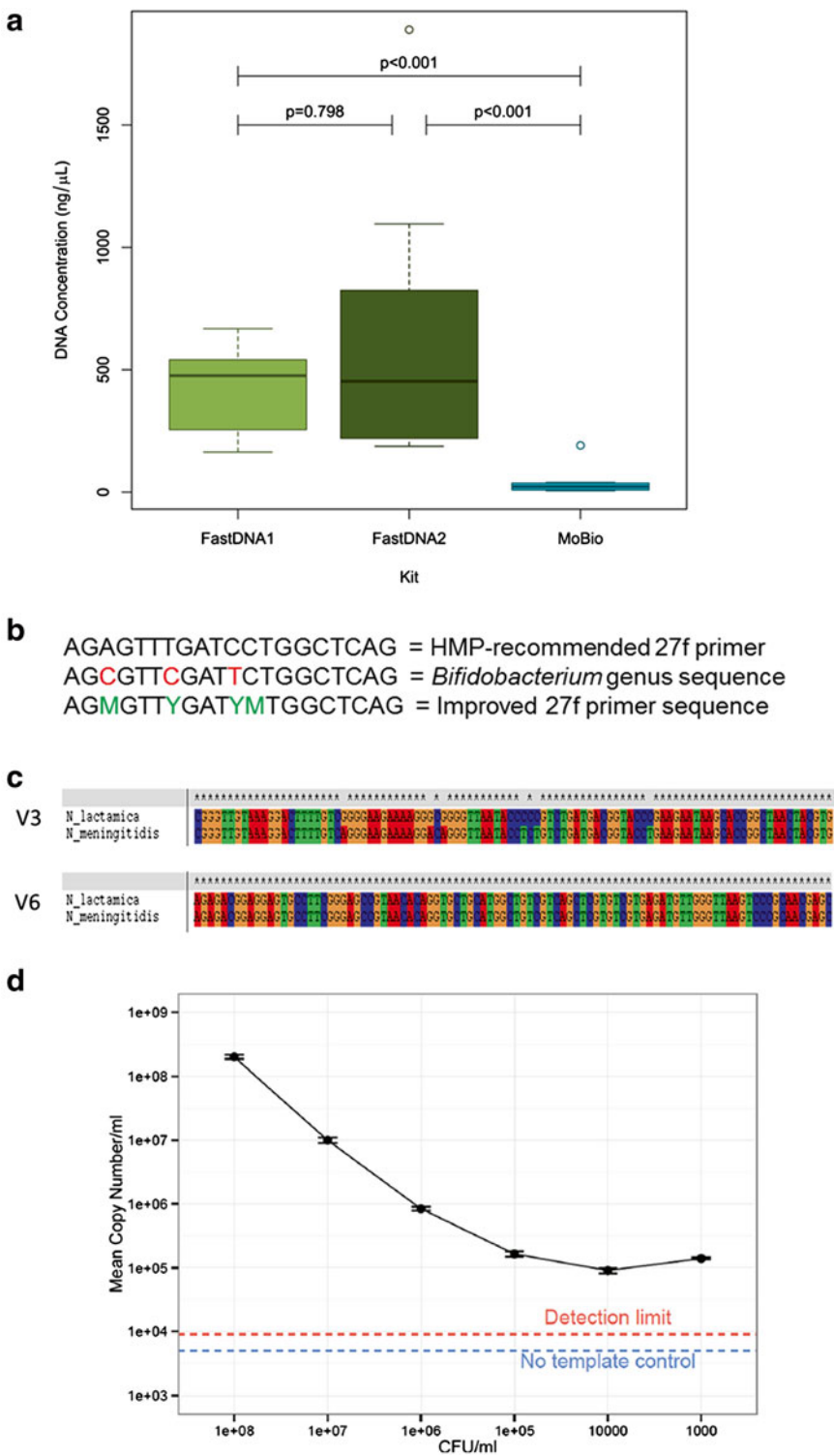


Fig. 2.4 Importance of optimising sample processing protocols
(a) DNA extraction methodology can impact recovery of DNA from microbiota samples. In this example it can be

seen that the yield from a MoBio kit-based protocol is much lower than that from two variations of the FastDNA kit-based protocols. This panel is reprinted in unmodified form from: Kennedy NA et al. The impact of different

bias the results against these groups, and to therefore give misleading estimates of their presence and/or abundance in the original samples. Furthermore, error rates appear to be higher on the Ion Torrent platform, which would artificially inflate measures of diversity (Salipante et al. 2014). The common practise of multiplexing many samples together on a single DNA sequencing run can also introduce bias to the PCR step (Berry et al. 2011) and lead to problems with misidentification of barcoded samples (Esling et al. 2015).

Finally, as DNA can persist in the environment after the death of the host organism, sequencing results (aside from perhaps metatranscriptomics, due to the short half-life of RNA compared to DNA) are unable to distinguish between live and dead/inactive microbes. Results may not therefore accurately represent the active microbiota at the site of interest. However, pre-treatment of samples with agents such as propidium monazide, which can bind to free DNA, and DNA contained within dead or damaged cells, make it possible to make sequencing results more representative of the living or active populations within the microbiota (Rogers et al. 2013).

The combined influence of all of these potentially confounding factors should be particularly borne in mind when conducting meta-analyses incorporating data generated across many different studies where different methodologies have

been used since they have the potential to have a greater influence on results obtained than any underlying experimental variable (Wesolowska-Andersen et al. 2014).

2.5 Other Community Profiling Approaches

2.5.1 Community Fingerprinting Techniques

Due to their falling costs and increased output sequence-based approaches have become the most widely adopted microbial community profiling techniques in recent years. Nonetheless, there are other molecular techniques, such as temperature/denaturing gradient gel electrophoresis (T/DGGE) (Muyzer et al. 1993), terminal restriction fragment length polymorphism (T-RFLP) (Marsh 1999) and automated ribosomal intergenic spacer analysis (ARISA) (Popa et al. 2009), that allow rapid profiling of human-associated microbial communities. These approaches are termed community fingerprinting techniques since they usually give representative overviews of the species present in a sample, without providing direct detailed information about the actual species present. Thus, although these approaches are relatively quick and cheap, the resolution and sensitivity is often much lower than that obtained with direct DNA sequencing (Kovacs et al. 2010; Kisand and Wikner 2003).



Fig. 2.4 (continued) DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA genesequencing. (Kennedy et al. 2014) under Creative Commons Attribution (CC BY) license
(b) Primer sequence can impact the recovery of species in 16S rRNA gene surveys. In this example it can be seen that the commonly used 27f primer has three mismatches with the important intestinal genus *Bifidobacterium*. As a result this genus is often under-represented in DNA sequence libraries. The bottom configuration shows the same primer with four degenerate bases, which widens the specificity of the primer and improves coverage of groups such as the bifidobacteria (Walker et al. 2015)
(c) Choice of 16S rRNA gene variable region can impact species-specificity of sequence results. In this example the

V3 region allows differentiation of two *Neisseria* species (*N. meningitidis* and *N. lactamica*) but the sequences from both species are identical over the V6 region, meaning differentiation would not be possible. Therefore, if the researcher was particularly interested in distinguishing these two species, primers targeting the V6 region could not be used

(d) Contamination in laboratory reagents. The panel shows the qPCR quantification of a serial dilution of a pure culture of *Salmonella bongori*. The bacterial quantification should reduce in a linear manner as the number of target cells reduces. Instead, the quantification plateaus after three dilutions, indicating the presence of background foreign contamination in the DNA extraction. This panel is reprinted in unmodified form from: Salter et al. (2014) under Creative Commons Attribution (CC BY) license

Although these techniques are gradually falling out of favour, recent work suggests that, while they are not as sensitive as modern next-generation sequencing, they can still generate broadly robust results (van Dorst et al. 2014). It should also be noted that, as they are all DNA extraction and PCR-dependent, and typically make use of marker genes such as 16S rRNA genes, they share many of the limitations and biases of the sequence-based approaches outlined previously in the “*Common pitfalls of sequence based approaches*” section, and in Table 2.1.

2.5.2 Microarrays

A microarray is a grid-like collection of microscopic spots of DNA that are anchored to a solid surface. These can be used to probe for the presence of complementary stretches of DNA extracted from a sample of interest by hybridising against the array. Microarrays can therefore be designed to be used in a number of different ways, for example to monitor changes in gene expression, or to mine for the presence of particular functional or marker genes (Paliy and Agans 2012; Tu et al. 2014). Phylogenetic microarrays (sometimes also referred to as phylochips) are a profiling method used in human microbiota research. This technique typically involves creating custom arrays seeded with short oligonucleotides (usually targeting the SSU rRNA genes) that are selected so that they collectively encompass the taxonomic range of organisms expected to be present within a given environmental sample type (Loy et al. 2010). DNA is extracted from the sample of interest, the SSU rRNA genes PCR amplified and labelled with a fluorescent marker and then hybridised against the microarray. When particular DNA spots on the array retain a positive fluorescent signal post-hybridisation, this indicates that the targeted taxonomic group is present in the original sample. By measuring the relative strength of the signal obtained for each positive spot post-hybridisation it may also be possible to semi-quantitatively assess the abundance of different taxa in a sample (Rajilic-Stojanovic et al. 2009).

A potential advantage that the microarray approach has over other profiling techniques is that it typically allows the researcher to simultaneously detect the presence of even quite low abundance organisms, which may not be detected reliably with even a sequence-based approach unless very deep sequencing is carried out. One major limitation though is that, unlike random sequencing approaches, detection is of course limited to the organisms that are targeted by the range of probes that are included on the initial array. Fortunately, there are now comprehensive custom arrays for a range of human-associated habitats such as the gut (Rajilic-Stojanovic et al. 2009; Ladirat et al. 2013; Tottey et al. 2013), vaginal tract (Gautam et al. 2015) and oral cavity (Crielaard et al. 2011), and the range of oligonucleotide probes that are included in these can be expanded as novel species are detected using sequence-based approaches (Rajilic-Stojanovic et al. 2009). It can also be difficult to design arrays where the hybridisation conditions are standardised for all of the probes included. As such it is prudent to control for potential false positives/negatives by including more than one probe for each taxonomic group targeted (Roh et al. 2010). Microarrays also share the same methodological limitations associated with the DNA extraction and PCR steps as other DNA-based techniques (see “*Common pitfalls of sequence based approaches*” section, and Table 2.1).

2.6 Quantitative Approaches

There are two widely used molecular methods, namely quantitative PCR and fluorescent *in situ* hybridisation, that allow the enumeration or quantification of dominant groups of microbes within the microbiota. For both of these techniques 16S rRNA gene sequences are typically the underlying basis, with different variable regions targeted with oligonucleotide probes and primers that are specific for particular phylogenetic groups.

2.6.1 Quantitative PCR

Quantitative PCR (qPCR), sometimes also referred to as real-time PCR, is a technique based on measuring fluorescence released during PCR amplification (Malinen et al. 2003). The amount of fluorescent signal generated, and the rate at which it accumulates, as the number of PCR cycles increases allows the researcher to quantify the amount of targeted DNA present in a given extraction. This approach is often used to quantify total bacterial cell numbers in a sample, but it can also be used to concurrently quantify the population levels of a number of different bacterial groups by using a range of targeted primer sets (Ramirez-Farias et al. 2009). This is a highly sensitive method and cell densities as low as 10^1 to 10^3 cells per sample may be accurately detected (Ott et al. 2004). One limitation of qPCR, however, is that it only allows monitoring of groups that have been specifically targeted by the chosen PCR primers. As a result, untargeted groups will not be observed in the results, and extensive monitoring of microbial communities typically requires the use of multiple different primer sets. Recent efforts have therefore been made to make this approach more high-throughput (Hermann-Bank et al. 2013). Primers must also be extensively tested first, to rule out non-specific binding to non-target DNA. As with all other DNA-based approaches, qPCR is also highly dependent on the choice of DNA extraction methodology.

2.6.2 Fluorescent *in situ* Hybridisation (FISH)

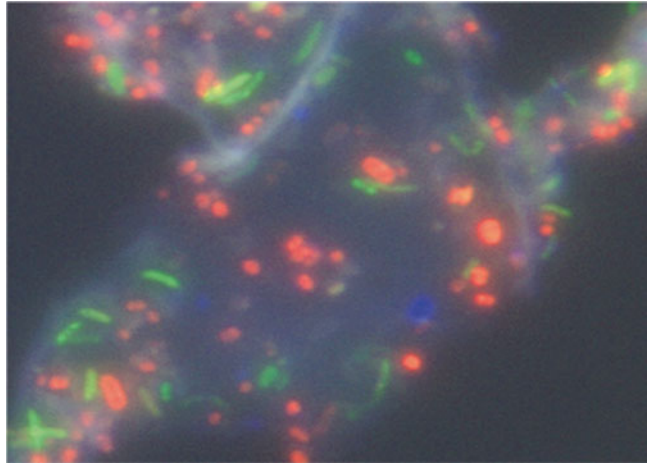
FISH is another widely used quantitative technique, with the added advantage that it does not require a DNA extraction step so is free from some of the biases associated with DNA-based methodologies. With FISH, bacterial cells are first fixed using chemicals such as paraformaldehyde and then permeabilised to allow access of fluorescently-labelled oligonucleotide probes. These oligonucleotides are typically between 15 and 30 bases in length and are com-

monly designed to target regions of rRNA that are specific for chosen phylogenetic groups of bacteria (Amann and Fuchs 2008). Probes may be targeted towards a broad range of bacteria by selecting a highly conserved section of the 16S rRNA gene or towards a narrower range by targeting more specific stretches of the gene (Amann and Ludwig 2000). After entering the fixed cell, the probes hybridise to any sequence of rRNA that is complementary to that of their own. As ribosomes are highly abundant, and distributed throughout the bacterial cell, the targeted cell fluoresces, which allows direct visualisation and enumeration by epifluorescent microscopy (Harmsen et al. 2002). FISH therefore, as well as being a quantitative approach, has the singular advantage that it allows observation of cells of interest *in situ*. For example, it is possible to determine the composition of specific consortia of microbes present on mucosal surfaces, or on the surfaces of particles (Fig. 2.5). A further strength of this approach is that it can be used to link phylogeny to function by employing it in conjunction with techniques such as microautoradiography (MAR-FISH) (Nielsen et al. 2010), Raman microspectroscopy (Raman-FISH) (Wagner 2009) or Secondary Ion Mass Spectrometry (FISH-SIMS) (Musat et al. 2012).

However, there are some important limitations to the use of FISH. It is a far less sensitive quantitative technique than qPCR because a critical mass of bacterial cells (typically around 10^6 cells/ml of sample) is required per microscopic field of view for accurate visual enumeration. As a result, FISH is most often used to monitor bacterial populations at broader taxonomic levels as individual species only rarely reach the required density for accurate monitoring (Harmsen et al. 2002). As with qPCR, it should also be noted that a further limitation is that FISH only allows monitoring of the microbial groups specifically targeted with oligonucleotide probes, and results can be confounded by false positive/negative results. It is therefore imperative that all newly designed oligonucleotides be tested for specificity prior to use with samples.

Fig. 2.5 Fluorescent *in situ* hybridisation

A key advantage of FISH is that it allows direct visualisation of bacteria in environmental samples. In this example we can see groups of bacteria colonising an insoluble fibre particle recovered from a human faecal sample. Cells coloured *green* belong to the *Lachnospiraceae* family, those labelled *red* belong to the *Ruminococcus* genus and those in *blue* are labelled with the universal DAPI stain and do not belong to either of these bacterial groups. Thus it can be seen that the majority of cells attaching to this fibre are derived from the *Lachnospiraceae* and *Ruminococcaceae*



2.7 Functional Analyses

Community profiling techniques can only provide an overview of the microbial composition in a given sample or, in the case of shotgun metagenomics, can only provide an overview of the encoding potential of a microbial ecosystem. Indeed, while we now have a much clearer picture of the kind of microbes that inhabit the various niches associated with the human body we know comparatively far less about the roles that each individual species plays. Fortunately, there are now a number of complementary techniques, beyond the culture-based and metatranscriptomics methods described previously in this chapter, that can be used to assess the functionality of the microbiota.

2.7.1 Functional Metagenomics

In contrast to whole shotgun metagenomics, where the aim is to generate deep sequencing-based profiles of the entire functional capability of the microbiota, with functional metagenomics the aim is instead to identify specific functional genes by cloning and expressing them in a surrogate bacterial species (Handelsman et al. 1998). Typically this involves large-scale cloning of random environmental DNA fragments into a host species such as *E. coli* and then screening for

activity by growing the transformed host species on agar plates containing a substrate of interest. Where functional activity is observed, the cloned gene can then be sequenced to provide supporting genomic data. This approach has been used, for example, to identify complex-carbohydrate degrading enzymes derived from the human gut (Tasse et al. 2010). Functional metagenomics is therefore a potentially hugely powerful approach, with the key advantage that it allows the researcher to simultaneously identify novel genes encoding specific functions from a broad range of bacterial species, including those that may not be amenable to culture in the laboratory (Uchiyama and Miyazaki 2009). A further advantage is that the functional annotation of previously unknown genes enhances reference databases, which can then be used to improve classification success rates and accuracy for sequence-based shotgun metagenomics studies.

There are, however, a number of important limitations, which has so far limited the use of functional metagenomics in comparison to the sequence-based shotgun metagenomics approach. For example, it is typically highly laborious and inefficient; millions of random DNA fragments may need to be cloned in order to identify activities of interest. Furthermore, there are important technological barriers that impinge upon the effectiveness of the approach. Many of the cloned fragments will be poorly

expressed by foreign hosts such as *E. coli*, meaning that alternative hosts/approaches may need to be considered (Liebl et al. 2014). In addition, the DNA extraction step is crucially important as the researcher must reach a balance between using a protocol that is stringent enough to extract DNA from as wide a range of species in the original sample as possible, but is not so stringent that it shears the resulting DNA to the extent that many of the cloned genes and gene clusters are disrupted (Kakirde et al. 2010). A further limitation is that, while this approach may identify products formed from individual genes or relatively simple contiguous gene clusters, it is unlikely to be able to identify gene products that result from complex metabolic pathways (Walker et al. 2014).

2.7.2 Metaproteomics

Metaproteomics is the study of the complement of proteins produced by mixed microbial communities (Wilmes and Bond 2009). As such, it provides functional information by allowing the researcher to monitor changes in protein expression by the entire microbiota in response to changes in prevailing environmental conditions. With this technique, proteins must first be extracted from the environmental sample of interest, and then separated prior to characterisation with mass spectrometry and subsequent bioinformatics-based comparisons with reference databases (Hettich et al. 2012). Until recently proteins (or peptides) were most commonly separated by using gel electrophoresis approaches (Magdeldin et al. 2014), but they are now increasingly separated by using liquid chromatography instead. Recent technological advances in the field mean that it is now possible to carry out very high throughput liquid chromatography-mass spectrometry based analyses, where many thousands of different proteins/peptides can be separated and characterised (Hettich et al. 2013).

Metaproteomics offers some key advantages over the metatranscriptomics approach described

previously in that, by measuring proteins rather than mRNA, it provides a broader, more representative picture of the functional activity of the microbiota as it also accounts for the impact of processes such as post-translational modifications (Cain et al. 2014). Proteins are also typically more stable than mRNA molecules, meaning that results obtained may not be so dependent on the speed with which the samples are processed. A particular advantage over DNA-based metagenomics is that metaproteomics is faster, and cheaper (Verberkmoes et al. 2009). The relatively untargeted nature of metaproteomics also means that it may be possible to identify marker proteins that are indicative of a healthy or diseased human host status.

However, although the technology involved in metaproteomics is rapidly improving, there are a range of important limitations, and this technique is currently far less commonly applied in comparison to DNA-based approaches. Although resolution is improving, metaproteomics can only currently characterise thousands out of the millions of proteins/peptides that might be present in a complex microbiota sample at one time (Kolmeder and de Vos 2014). As such, only proteins produced by the most dominant members of the microbiota can be expected to be captured with reasonable coverage (Verberkmoes et al. 2009). It can also be difficult to differentiate similar proteins or ascribe them to particular phylogenetic groups (Lichtman et al. 2015), and, as with metagenomics studies, a large proportion of the data recovered will have no close matches to available reference databases (Verberkmoes et al. 2009). The methodology chosen during the protein extraction step will also have significant impacts on the representativeness of the protein complement recovered, and it is important to extract proteins with reasonable efficiency from both Gram positive and Gram negative constituents (Tanca et al. 2014). Human-derived proteins will also be present, and can be a highly significant component in samples such as biopsies, meaning it is sometimes necessary to carry out selective steps to enrich for microbial proteins (Kolmeder

and de Vos 2014). There are also issues surrounding reproducibility between samples, particularly when using gel electrophoresis to separate proteins (Magdeldin et al. 2014).

2.7.3 Metabolomics

Metabolomics is the study of the metabolites/small molecules present within a given sample at the time of sampling. As with the metaproteomics approach outlined above, metabolomics therefore offers distinct advantages over other functional approaches such as metatranscriptomics as it allows the direct monitoring of the end products of bacterial metabolism (Ursell et al. 2014). With metabolomics, metabolites are typically isolated from bodily samples such as urine, faeces and blood and measured using technologies such nuclear magnetic resonance (NMR) microscopy or mass spectrometry (Nicholson and Lindon 2008). The end result of these approaches are a series of characteristic spectra or peaks derived from the range of metabolites that are present within the original sample (Savorani et al. 2013). Depending on the approach used, metabolomic screens can either be carried out in a targeted way for particular groups of metabolites (for example, short chain fatty acids), or on a more global basis (Griffiths et al. 2010). In the latter case, the main challenge is to assign particular spectra from the complex mixture of peaks to specific compounds, and then to attempt to correlate presence/absence of these compounds with markers of host health (Lenz and Wilson 2007). By simultaneously capturing both host and microbial-derived metabolites, metabolomics has particular appeal as an approach to characterise host-microbe interactions (Wikoff et al. 2009).

A key limitation of this technique is that it can be difficult to accurately determine which microbial species are producing particular metabolites. While attempts are often made to correlate metabolite production with microbial composition data generated in tandem by sequence survey or metagenomic approaches, these can be con-

founded by the presence of DNA derived from dead or inactive species in the sequence-based results, and by the fact that there can be considerable metabolic flux within complex ecosystems, such that metabolites associated with taxa that are dominant in sequence surveys may not actually be produced by them (Abram 2015). Furthermore, many metabolites, for example short chain fatty acids, are rapidly absorbed by the host, meaning that production levels cannot be accurately defined or ascribed to particular species (Kolmeder and de Vos 2014). An additional important disadvantage is that reference databases are generally lacking, even more so than those for DNA and proteins, meaning that only a small fraction of metabolomics data can currently be assigned to known metabolites (Baker 2011). Finally, as with metaproteomics, resolution limits (even with the most modern instruments) mean that it is only possible to accurately monitor a small subset of the wide range of metabolites that may be present in a complex sample such as faeces (Goedert et al. 2014).

It can be seen, therefore, that all four key modern “omics” technologies (metagenomics for DNA, metatranscriptomics for RNA, metaproteomics for proteins, and metabolomics for metabolites) have distinct strengths and limitations. As a result, there is increasing interest in integrating the output from each of these approaches in order to enhance their overall power and provide a more comprehensive, systems biology-based, overview of the human microbiota. Effective integration of these complex datasets remains to some extent an unfulfilled ambition, but one that is being rapidly guided by improvements in computing infrastructure, bioinformatics, mathematical modelling and statistical approaches (Abram 2015).

2.7.4 Stable Isotope Probing

One final functional approach with strong applicability to the study of the human microbiota is stable isotope probing (SIP). With this technique, mixed microbial communities are incubated with

labelled substrates containing heavy stable isotopes such as ^{13}C , ^{15}N , and ^{18}O . Species that are able to grow on the labelled substrate incorporate the isotope markers into cellular biomass, which can then be studied by looking at components such as DNA (DNA-SIP), RNA (RNA-SIP), proteins (protein-SIP) or phospholipid-derived fatty acids (PFLA-SIP). Approaches like density gradient ultracentrifugation (Dunford and Neufeld 2010) or advanced single-cell resolution techniques such as Raman microspectroscopy and Secondary Ion Mass Spectrometry (SIMS) are used to distinguish the active microbes from species that did not incorporate the marker (Eichorst et al. 2015). Regardless of the actual cellular components targeted SIP is therefore an attractive basis for uncovering which microbes within complex microbial communities carry out particular functions (Uhlik et al. 2013).

SIP is an emerging means with which to unravel the complex activities of the human microbiota. Early studies used this technique in tandem with community profiling approaches like T-RFLP and FISH to characterise the microbes that were able to actively utilise labelled substrates such as resistant starch and oligofructose (Kovatcheva-Datchary et al. 2009; Reichardt et al. 2011). When used in combination with more modern “omics” techniques SIP has the potential to be particularly powerful. For example, fractionated DNA or RNA containing the stable isotopes can then be sequenced using marker gene surveys, metagenomics or metatranscriptomics in order to identify the species that were active during incubation with the labelled substrate (Chen and Murrell 2010). Similarly, advances in micro-manipulation technologies such as optical tweezers mean that whole cells that have been shown by techniques like Raman microspectroscopy to have incorporated the stable isotopes can then be isolated from the sample and either cultured or, if that is not possible, put forward for genome sequencing via single cell genomics (Berry et al. 2015).

While SIP approaches can be hugely powerful there are important caveats, which have limited widespread application of these techniques

thus far. SIP is far more technically challenging than approaches such as SSU rRNA gene surveys or metagenomics, and the modern single-cell resolution techniques such as Raman microspectroscopy and SIMS can be prohibitively expensive (Wagner 2009). Similarly, use of SIP is limited by the supply and cost of labelled substrates (Uhlik et al. 2013). Recent innovations though, such as the use of the cheap and readily available heavy water (D_2O) as a general marker of cellular growth, allows SIP to be carried out without specific labelled carbon or nitrogen sources (Berry et al. 2015). A further limitation is that SIP requires microbes to be grown in the presence of the labelled tracer so that it can be incorporated into active cells. Often this means growing mixed communities under artificial laboratory conditions, meaning that results may not entirely reflect the activity of the microbiota *in vivo* (Uhlik et al. 2013). Nonetheless, impressive new innovations have, for example, allowed researchers to identify microbes growing *in vivo* that forage host-derived proteins for growth (Berry et al. 2013). Finally, there is considerable metabolic flux within complex microbial communities, with cross feeding between species a common feature. This means that stable isotopes such as ^{13}C may “flow” from the primary degrader of a labelled substrate to many other species that are present within the community, potentially impeding the ability to detect the initial utilising species (Dumont and Murrell 2005).

2.8 Conclusions

There are now many different ways in which the human microbiota can be studied, and each methodology has inherent advantages and limitations. Ultimately, the best technical approach for a given situation will clearly depend on the question that the researcher wishes to address. Although each technique has largely been considered in isolation in this review it should be emphasised here that, where possible, the synergistic use of multiple methodological approaches

offers perhaps the greatest power with which to uncover novel insights.

Looking towards the future, it is clear that further improvements in sequence-based technologies, molecular methods, model systems and bioinformatics will continue to open up novel avenues for research. The synergistic adoption of such approaches will greatly enhance our ability to take a systems biology-based view of the human microbiota, and how it interacts with the host. Traditional techniques such as culture will also retain an important role as we seek to translate omics-based observations into interventions such as probiotics and pharmabiotics aimed at improving host health (Reardon 2014). We have come a long way since Antonie van Leeuwenhoek's first glimpses of the human microbiota, and are now quickly entering an era where our increased understanding of our microbial inhabitants is being put to practical therapeutic use (Shanahan 2015). Further technological advances can only accelerate this process.

Acknowledgements The author, and The Rowett Institute of Nutrition and Health, University of Aberdeen, receives core funding support from the Scottish Government Rural and Environmental Science and Analysis Service (RESAS).

References

- Abram F (2015) Systems-based approaches to unravel multi-species microbial community functioning. *Comput Struct Biotechnol J* 13:24–32
- Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ et al (2014) Binning metagenomic contigs by coverage and composition. *Nat Methods* 11(11):1144–1146
- Amann R, Fuchs BM (2008) Single-cell identification in microbial communities by improved fluorescence *in situ* hybridization techniques. *Nat Rev Microbiol* 6(5):339–348
- Amann R, Ludwig W (2000) Ribosomal RNA-targeted nucleic acid probes for studies in microbial ecology. *FEMS Microbiol Rev* 24(5):555–565
- Angiuoli SV, White JR, Matalka M, White O, Fricke WF (2011) Resources and costs for microbial sequence analysis evaluated using virtual machines and cloud computing. *PLoS One* 6(10), e26624
- Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2005) At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl Environ Microbiol* 71(12):7724–7736
- Bahl MI, Bergstrom A, Licht TR (2012) Freezing fecal samples prior to DNA extraction affects the Firmicutes to Bacteroidetes ratio determined by downstream quantitative PCR analysis. *FEMS Microbiol Lett* 329(2):193–197
- Baker M (2011) Metabolomics: from small molecules to big ideas. *Nat Methods* 8:117–121
- Bankovich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS et al (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19(5):455–477
- Belenguer A, Duncan SH, Calder AG, Holtrop G, Louis P, Lobley GE et al (2006) Two routes of metabolic cross-feeding between *Bifidobacterium adolescentis* and butyrate-producing anaerobes from the human gut. *Appl Environ Microbiol* 72(5):3593–3599
- Berry D, Ben Mahfoudh K, Wagner M, Loy A (2011) Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Appl Environ Microbiol* 77(21):7846–7849
- Berry D, Stecher B, Schintlmeister A, Reichert J, Brugiroux S, Wild B et al (2013) Host-compound foraging by intestinal microbiota revealed by single-cell stable isotope probing. *Proc Natl Acad Sci U S A* 110(12):4720–4725
- Berry D, Mader E, Lee TK, Woebken D, Wang Y, Zhu D et al (2015) Tracking heavy water (D₂O) incorporation for identifying and sorting active microbial cells. *Proc Natl Acad Sci U S A* 112(2):E194–E203
- Blainey PC (2013) The future is now: single-cell genomics of bacteria and archaea. *FEMS Microbiol Rev* 37(3):407–427
- Bomar L, Maltz M, Colston S, Graf J (2011) Directed culturing of microorganisms using metatranscriptomics. *mBio* 2(2):e00012–11
- Bonnet R, Suau A, Dore J, Gibson GR, Collins MD (2002) Differences in rDNA libraries of faecal bacteria derived from 10- and 25-cycle PCRs. *Int J Syst Evol Microbiol* 52(Pt 3):757–763
- Burck JM, Yeoman CJ, Wilson MK, Berg Miller ME, Jeraldo P, Jindou S et al (2011) Cellulosomics, a gene-centric approach to investigating the intraspecific diversity and adaptation of *Ruminococcus flavefaciens* within the rumen. *PLoS One* 6(10), e25329
- Buttigieg PL, Ramette A (2015) A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. *FEMS Microbiol Ecol* 90:543–550
- Cain JA, Solis N, Cordwell SJ (2014) Beyond gene expression: the impact of protein post-translational modifications in bacteria. *J Proteome* 97:265–286
- Campbell AG, Schwientek P, Vishnivetskaya T, Woyke T, Levy S, Beall CJ et al (2014) Diversity and genomic insights into the uncultured *Chloroflexi* from the human microbiota. *Environ Microbiol* 16(9):2635–2643

- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK et al (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7(5):335–336
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ et al (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A* 108(Suppl 1):4516–4522
- Chen Y, Murrell JC (2010) When metagenomics meets stable-isotope probing: progress and perspectives. *Trends Microbiol* 18(4):157–163
- Chun J, Lee JH, Jung Y, Kim M, Kim S, Kim BK et al (2007) EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. *Int J Syst Evol Microbiol* 57(Pt 10):2259–2261
- Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y et al (2014) Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 42(Database issue):D633–D642
- Conway T, Creevy JP, Maddox SM, Grissom JE, Conkle TL, Shadid TM et al (2014) Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. *mBio* 5(4):e01442-14
- Crielaard W, Zaura E, Schuller AA, Huse SM, Montijn RC, Keijser BJ (2011) Exploring the oral microbiota of children at various developmental stages of their dentition in the relation to their oral health. *BMC Med Genet* 4:22
- Croucher NJ, Thomson NR (2010) Studying bacterial transcriptomes using RNA-seq. *Curr Opin Microbiol* 13(5):619–624
- Cryan JF, Dinan TG (2012) Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nat Rev Neurosci* 13(10):701–712
- Cuthbertson L, Rogers GB, Walker AW, Oliver A, Hafiz T, Hoffman LR et al (2014) Time between collection and storage significantly influences bacterial sequence composition in sputum samples from cystic fibrosis respiratory infections. *J Clin Microbiol* 52(8):3011–3016
- Cuthbertson L, Rogers GB, Walker AW, Oliver A, Hoffman LR, Carroll MP et al (2015) Implications of multiple freeze-thawing on respiratory samples for culture-independent analyses. *J Cyst Fibros* 14(4):464–467
- de Boer R, Peters R, Gierveld S, Schuurman T, Kooistra-Smid M, Savelkoul P (2010) Improved detection of microbial DNA after bead-beating before DNA isolation. *J Microbiol Methods* 80(2):209–211
- Degnan PH, Ochman H (2012) Illumina-based analysis of microbial community diversity. *ISME J* 6(1):183–194
- DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K et al (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72(7):5069–5072
- Dumont MG, Murrell JC (2005) Stable isotope probing – linking microbial identity to function. *Nat Rev Microbiol* 3(6):499–504
- Duncan SH, Hold GL, Harmsen HJ, Stewart CS, Flint HJ (2002) Growth requirements and fermentation products of *Fusobacterium prausnitzii*, and a proposal to reclassify it as *Faecalibacterium prausnitzii* gen. nov., comb. nov. *Int J Syst Evol Microbiol* 52(Pt 6):2141–2146
- Dunford EA, Neufeld JD (2010) DNA stable-isotope probing (DNA-SIP). *J Vis Exp JoVE* 42
- Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M et al (2005) Diversity of the human intestinal microbial flora. *Science* 308(5728):1635–1638
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27(16):2194–2200
- Eichorst SA, Strasser F, Woyke T, Schintlmeister A, Wagner M, Wobken D (2015) Advancements in the application of NanoSIMS and Raman microspectroscopy to investigate the activity of microbial cells in soils. *FEMS Microbiol Ecol* 91(10):fiv106
- Ellekilde M, Selfjord E, Larsen CS, Jaksevic M, Rune I, Tranberg B et al (2014) Transfer of gut microbiota from lean and obese mice to antibiotic-treated mice. *Sci rep* 4:5922
- Eller C, Crabill MR, Bryant MP (1971) Anaerobic roll tube media for nonselective enumeration and isolation of bacteria in human feces. *Appl Microbiol* 22(4):522–529
- Eppinger M, Pearson T, Koenig SS, Pearson O, Hicks N, Agrawal S et al (2014) Genomic epidemiology of the Haitian cholera outbreak: a single introduction followed by rapid, extensive, and continued spread characterized the onset of the epidemic. *mBio* 5(6):e01721
- Ericsson AC, Davis JW, Spollen W, Bivens N, Givan S, Hagan CE et al (2015) Effects of vendor and genetic background on the composition of the fecal microbiota of inbred mice. *PLoS One* 10(2), e0116704
- Esling P, Lejzerowicz F, Pawlowski J (2015) Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Res* 43(5):2513–2524
- Ferenci T (1999) ‘Growth of bacterial cultures’ 50 years on: towards an uncertainty principle instead of constants in bacterial growth kinetics. *Res Microbiol* 150(7):431–438
- Ferrand J, Patron K, Legrand-Frossi C, Fripiat JP, Merlin C, Alauzet C et al (2014) Comparison of seven methods for extraction of bacterial DNA from fecal and cecal samples of mice. *J Microbiol Methods* 105:180–185
- Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR et al (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269(5223):496–512
- Flint HJ, Duncan SH, Scott KP, Louis P (2007) Interactions and competition within the microbial community of

- the human colon: links between diet and health. *Environ Microbiol* 9(5):1101–1111
- Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ (2008) Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl Environ Microbiol* 74(8):2461–2470
- Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM et al (2014) Relating the metatranscriptome and metagenome of the human gut. *Proc Natl Acad Sci U S A* 111(22):E2329–E2338
- Gautam R, Borgdorff H, Jaspers V, Francis SC, Verhelst R, Mwaura M et al (2015) Correlates of the molecular vaginal microbiota composition of African women. *BMC Infect Dis* 15:86
- Giannoukos G, Ciulla DM, Huang K, Haas BJ, Izard J, Levin JZ et al (2012) Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biol* 13(3):R23
- Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS et al (2006) Metagenomic analysis of the human distal gut microbiome. *Science* 312(5778):1355–1359
- Goedert JJ, Sampson JN, Moore SC, Xiao Q, Xiong X, Hayes RB et al (2014) Fecal metabolomics: assay performance and association with colorectal cancer. *Carcinogenesis* 35(9):2089–2096
- Goodman AL, Kallstrom G, Faith JJ, Reyes A, Moore A, Dantas G et al (2011) Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc Natl Acad Sci U S A* 108(15):6252–6257
- Griffiths WJ, Koal T, Wang Y, Kohl M, Enot DP, Deigner HP (2010) Targeted metabolomics for biomarker discovery. *Angew Chem* 49(32):5426–5445
- Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 68(4):669–685
- Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5(10):R245–R249
- Harmsen HJ, Raangs GC, He T, Degener JE, Welling GW (2002) Extensive set of 16S rRNA-based probes for detection of bacteria in human feces. *Appl Environ Microbiol* 68(6):2982–2990
- He M, Sebaihia M, Lawley TD, Stabler RA, Dawson LF, Martin MJ et al (2010) Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proc Natl Acad Sci U S A* 107(16):7527–7532
- Hermann-Bank ML, Skovgaard K, Stockmarr A, Larsen N, Molbak L (2013) The Gut Microbiotassay: a high-throughput qPCR approach combinable with next generation sequencing to study gut microbial diversity. *BMC Genomics* 14:788
- Hettich RL, Sharma R, Chourey K, Giannone RJ (2012) Microbial metaproteomics: identifying the repertoire of proteins that microorganisms use to compete and cooperate in complex environmental communities. *Curr Opin Microbiol* 15(3):373–380
- Hettich RL, Pan C, Chourey K, Giannone RJ (2013) Metaproteomics: harnessing the power of high performance mass spectrometry to identify the suite of proteins that control metabolic activities in microbial communities. *Anal Chem* 85(9):4203–4214
- Hu Y, Yang X, Qin J, Lu N, Cheng G, Wu N et al (2013) Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nat Commun* 4:2151
- Human Microbiome Project C (2012a) Structure, function and diversity of the healthy human microbiome. *Nature* 486(7402):207–214
- Human Microbiome Project Consortium (2012b) A framework for human microbiome research. *Nature* 486(7402):215–221
- Huse SM, Welch DM, Morrison HG, Sogin ML (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ Microbiol* 12(7):1889–1898
- Huse SM, Mark Welch DB, Voorhis A, Shipunova A, Morrison HG, Eren AM et al (2014) VAMPS: a website for visualization and analysis of microbial population structures. *BMC Bioinforma* 15:41
- Jorth P, Turner KH, Gumus P, Nizam N, Buduneli N, Whiteley M (2014) Metatranscriptomics of the human oral microbiome during health and disease. *M Bio* 5(2):e01012–e01014
- Kakirde KS, Parsley LC, Liles MR (2010) Size does matter: application-driven approaches for soil Metagenomics. *Soil Biol Biochem* 42(11):1911–1923
- Kennedy NA, Walker AW, Berry SH, Duncan SH, Farquarson FM, Louis P et al (2014) The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. *PLoS One* 9(2), e88982
- Kisand V, Wikner J (2003) Limited resolution of 16S rDNA DGGE caused by melting properties and closely related DNA sequences. *J Microbiol Methods* 54(2):183–191
- Klaasen HL, Koopman JP, Van den Brink ME, Van Wezel HP, Beynen AC (1991) Mono-association of mice with non-cultivable, intestinal, segmented, filamentous bacteria. *Arch Microbiol* 156(2):148–151
- Kolmeder CA, de Vos WM (2014) Metaproteomics of our microbiome – developing insight in function and activity in man and model systems. *J Proteome* 97:3–16
- Koser CU, Ellington MJ, Cartwright EJ, Gillespie SH, Brown NM, Farrington M et al (2012) Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog* 8(8), e1002824
- Kostic AD, Howitt MR, Garrett WS (2013) Exploring host-microbiota interactions in animal models and humans. *Genes Dev* 27(7):701–718

- Kovacs A, Yacoby K, Gophna U (2010) A systematic assessment of automated ribosomal intergenic spacer analysis (ARISA) as a tool for estimating bacterial richness. *Res Microbiol* 161(3):192–197
- Kovatcheva-Datchary P, Egert M, Maathuis A, Rajilic-Stojanovic M, de Graaf AA, Smidt H et al (2009) Linking phylogenetic identities of bacteria to starch fermentation in an in vitro model of the large intestine by RNA-based stable isotope probing. *Environ Microbiol* 11(4):914–926
- Ladirat SE, Schols HA, Nauta A, Schoterman MH, Keijser BJ, Montijn RC et al (2013) High-throughput analysis of the impact of antibiotics on the human intestinal microbiota composition. *J Microbiol Methods* 92(3):387–397
- Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C et al (2012) Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin Microbiol Infect* 18(12):1185–1193
- Lagier JC, Edouard S, Pagnier I, Mediannikov O, Drancourt M, Raoult D (2015a) Current and past strategies for bacterial culture in clinical microbiology. *Clin Microbiol Rev* 28(1):208–236
- Lagier JC, Hugon P, Khelaifa S, Fournier PE, La Scola B, Raoult D (2015b) The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. *Clin Microbiol Rev* 28(1):237–264
- Lasken RS, Stockwell TB (2007) Mechanism of chimera formation during the multiple displacement amplification reaction. *BMC Biotechnol* 7:19
- Lees H, Swann J, Poucher SM, Nicholson JK, Holmes E, Wilson ID et al (2014) Age and microenvironment outweigh genetic influence on the Zucker rat microbiome. *PLoS One* 9(9), e100916
- Leimena MM, Ramiro-Garcia J, Davids M, van den Bogert B, Smidt H, Smid EJ et al (2013) A comprehensive metatranscriptome analysis pipeline and its validation using human small intestine microbiota datasets. *BMC Genomics* 14:530
- Lennon NJ, Lintner RE, Anderson S, Alvarez P, Barry A, Brockman W et al (2010) A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454. *Genome Biol* 11(2):R15
- Lenz EM, Wilson ID (2007) Analytical strategies in metabolomics. *J Proteome Res* 6(2):443–458
- Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S et al (2014) An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol* 32(8):834–841
- Lichtman JS, Sonnenburg JL, Elias JE (2015) Monitoring host responses to the gut microbiota. *ISME J* 9(9):1908–1915
- Liebl W, Angelov A, Juergensen J, Chow J, Loeschke A, Drepper T et al (2014) Alternative hosts for functional (meta) genome analysis. *Appl Microbiol Biotechnol* 98(19):8099–8109
- Loman NJ, Constantinidou C, Chan JZ, Halachev M, Sergeant M, Penn CW et al (2012) High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol* 10(9):599–606
- Louis P, Young P, Holtrop G, Flint HJ (2010) Diversity of human colonic butyrate-producing bacteria revealed by analysis of the butyryl-CoA:acetate CoA-transferase gene. *Environ Microbiol* 12(2):304–314
- Loy A, Pester M, Steger D (2010) Phylogenetic microarrays for cultivation-independent identification and metabolic characterization of microorganisms in complex samples. *Methods Mol Biol* 688:187–206
- Macklaim JM, Fernandes AD, Di Bella JM, Hammond JA, Reid G, Gloor GB (2013) Comparative meta-RNA-seq of the vaginal microbiota and differential expression by *Lactobacillus iners* in health and dysbiosis. *Microbiome* 1(1):12
- Magdeldin S, Enany S, Yoshida Y, Xu B, Zhang Y, Zureena Z et al (2014) Basics and recent advances of two dimensional- polyacrylamide gel electrophoresis. *Clin proteomics* 11(1):16
- Malinen E, Kassinen A, Rinttila T, Palva A (2003) Comparison of real-time PCR with SYBR Green I or 5'-nuclease assays and dot-blot hybridization with rDNA-targeted oligonucleotide probes in quantification of selected faecal bacteria. *Microbiology* 149(Pt 1):269–277
- Marcy Y, Ouverney C, Bik EM, Losekann T, Ivanova N, Martin HG et al (2007) Dissecting biological “dark matter” with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proc Natl Acad Sci U S A* 104(29):11889–11894
- Marsh TL (1999) Terminal restriction fragment length polymorphism (T-RFLP): an emerging method for characterizing diversity among homologous populations of amplification products. *Curr Opin Microbiol* 2(3):323–327
- Maukonen J, Simoes C, Saarela M (2012) The currently used commercial DNA-extraction methods give different results of clostridial and actinobacterial populations derived from human fecal samples. *FEMS Microbiol Ecol* 79(3):697–708
- Maurice CF, Haiser HJ, Turnbaugh PJ (2013) Xenobiotics shape the physiology and gene expression of the active human gut microbiome. *Cell* 152(1–2):39–50
- McPherson JD (2014) A defining decade in DNA sequencing. *Nat Methods* 11(10):1003–1005
- Miller TL, Wolin MJ (1981) Fermentation by the human large intestine microbial community in an in vitro semicontinuous culture system. *Appl Environ Microbiol* 42(3):400–407
- Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD et al (2011) The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res* 21(10):1616–1625
- Musat N, Foster R, Vagner T, Adam B, Kuypers MM (2012) Detecting metabolic activities in single cells, with emphasis on nanoSIMS. *FEMS Microbiol Rev* 36(2):486–511

- Muyzer G, de Waal EC, Uitterlinden AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* 59(3):695–700
- Namiki T, Hachiya T, Tanaka H, Sakakibara Y (2012) MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res* 40(20), e155
- Nguyen TL, Vieira-Silva S, Liston A, Raes J (2015) How informative is the mouse for human gut microbiota research? *Dis model mech* 8(1):1–16
- Nicholson JK, Linton JC (2008) Systems biology: metabonomics. *Nature* 455(7216):1054–1056
- Nielsen JL, Nielsen PH (2010) Combined microautoradiography and fluorescence *in situ* hybridization (MAR-FISH) for the identification of metabolic active microorganisms. In: Timmis KN (ed) *Handbook of hydrocarbon and lipid microbiology*. Springer, Berlin, pp 4093–4102
- Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S et al (2014) Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat Biotechnol* 32(8):822–828
- Ott SJ, Musfeldt M, Ullmann U, Hampe J, Schreiber S (2004) Quantification of intestinal bacterial populations by real-time PCR with a universal primer set and minor groove binder probes: a global approach to the enteric flora. *J Clin Microbiol* 42(6):2566–2572
- Paliy O, Agans R (2012) Application of phylogenetic microarrays to interrogation of human microbiota. *FEMS Microbiol Ecol* 79(1):2–11
- Parkhill J, Wren BW (2011) Bacterial epidemiology and biology—lessons from genome sequencing. *Genome Biol* 12(10):230
- Peng Y, Leung HC, Yiu SM, Chin FY (2011) Meta-IDBA: a de Novo assembler for metagenomic data. *Bioinformatics* 27(13):i94–i101
- Pham TA, Lawley TD (2014) Emerging insights on intestinal dysbiosis during bacterial infections. *Curr Opin Microbiol* 17:67–74
- Popa R, Popa R, Mashall MJ, Nguyen H, Tebo BM, Brauer S (2009) Limitations and benefits of ARISA intra-genomic diversity fingerprinting. *J Microbiol Methods* 78(2):111–118
- Porter JR (1976) Antony van Leeuwenhoek: tercentenary of his discovery of bacteria. *Bacteriol Rev* 40(2):260–269
- Prakash T, Taylor TD (2012) Functional assignment of metagenomic data: challenges and applications. *Brief Bioinform* 13(6):711–727
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P et al (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41(Database issue):D590–D596
- Raghunathan A, Ferguson HR Jr, Bornarth CJ, Song W, Driscoll M, Lasken RS (2005) Genomic DNA amplification from a single bacterium. *Appl Environ Microbiol* 71(6):3342–3347
- Rajilic-Stojanovic M, Smidt H, de Vos WM (2007) Diversity of the human gastrointestinal tract microbiota revisited. *Environ Microbiol* 9(9):2125–2136
- Rajilic-Stojanovic M, Heilig HG, Molenaar D, Kajander K, Surakka A, Smidt H et al (2009) Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. *Environ Microbiol* 11(7):1736–1751
- Ramirez-Farias C, Slezak K, Fuller Z, Duncan A, Holtrop G, Louis P (2009) Effect of inulin on the human gut microbiota: stimulation of *Bifidobacterium adolescentis* and *Faecalibacterium prausnitzii*. *Br J Nutr* 101(4):541–550
- Reardon S (2014) Microbiome therapy gains market traction. *Nature* 509(7500):269–270
- Reck M, Tomasch J, Deng Z, Jarek M, Husemann P, Wagner-Dobler I et al (2015) Stool metatranscriptomics: a technical guideline for mRNA stabilisation and isolation. *BMC Genomics* 16:494
- Reichardt N, Barclay AR, Weaver LT, Morrison DJ (2011) Use of stable isotopes to measure the metabolic activity of the human intestinal microbiota. *Appl Environ Microbiol* 77(22):8009–8014
- Reichardt N, Duncan SH, Young P, Belenguer A, McWilliam Leitch C, Scott KP et al (2014) Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *ISME J* 8(6):1323–1335
- Reuter JA, Spacek DV, Snyder MP (2015) High-throughput sequencing technologies. *Mol Cell* 58(4):586–597
- Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF et al (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499(7459):431–437
- Robertson M (1980) Biology in the 1980s, plus or minus a decade. *Nature* 285(5764):358–359
- Rogers GB, Cuthbertson L, Hoffman LR, Wing PA, Pope C, Hooftman DA et al (2013) Reducing bias in bacterial community analysis of lower respiratory infections. *ISME J* 7(4):697–706
- Roh SW, Abell GC, Kim KH, Nam YD, Bae JW (2010) Comparing microarrays and next-generation sequencing technologies for microbial ecology research. *Trends Biotechnol* 28(6):291–299
- Salipante SJ, Kawashima T, Rosenthal C, Hoogstraal DR, Cummings LA, Sengupta DJ et al (2014) Performance comparison of Illumina and ion torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. *Appl Environ Microbiol* 80(24):7583–7591
- Salonen A, Lahti L, Salojärvi J, Holtrop G, Korpela K, Duncan SH et al (2014) Impact of diet and individual variation on intestinal microbiota composition and fermentation products in obese men. *ISME J* 8(11):2218–2230

- Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF et al (2014) Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 12:87
- Sanger F, Nicklen S, Coulson AR (1992) DNA sequencing with chain-terminating inhibitors. 1977 [classical article]. *Biotechnology* 24:104–108
- Savorani F, Rasmussen MA, Mikkelsen MS, Engelsen SB (2013) A primer to nutritional metabolomics by NMR spectroscopy and chemometrics. *Food Res Int* 54(1):1131–1145
- Schloss PD, Westcott SL (2011) Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. *Appl Environ Microbiol* 77(10):3219–3226
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB et al (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75(23):7537–7541
- Schloss PD, Gevers D, Westcott SL (2011) Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One* 6(12), e27310
- Schuenemann VJ, Singh P, Mendum TA, Krause-Kyora B, Jager G, Bos KI et al (2013) Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science* 341(6142):179–183
- Seedorf H, Griffin NW, Ridaura VK, Reyes A, Cheng J, Rey FE et al (2014) Bacteria from diverse habitats colonize and compete in the mouse gut. *Cell* 159(2):253–266
- Shade A, Hogan CS, Klimowicz AK, Linske M, McManus PS, Handelsman J (2012) Culturing captures members of the soil rare biosphere. *Environ Microbiol* 14(9):2247–2252
- Shanahan F (2015) Separating the microbiome from the hyperbolome. *Genome med* 7(1):17
- Sim K, Cox MJ, Wopereis H, Martin R, Knol J, Li MS et al (2012) Improved detection of bifidobacteria with optimised 16S rRNA-gene based pyrosequencing. *PLoS One* 7(3), e32543
- Sorge RE, Martin LJ, Isbester KA, Sotocinal SG, Rosen S, Tuttle AH et al (2014) Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nat Methods* 11(6):629–632
- Stackebrandt E, Ebers J (2011) Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* 33:152–155
- Tanca A, Palomba A, Pisanu S, Deligios M, Fraumene C, Manghina V et al (2014) A straightforward and efficient analytical pipeline for metaproteome characterization. *Microbiome* 2(1):49
- Tanner MA, Goebel BM, Dojka MA, Pace NR (1998) Specific ribosomal DNA sequences from diverse environmental settings correlate with experimental contaminants. *Appl Environ Microbiol* 64(8):3110–3113
- Tasse L, Bercovici J, Pizzut-Serin S, Robe P, Tap J, Klopp C et al (2010) Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome Res* 20(11):1605–1612
- Thomas T, Gilbert J, Meyer F (2012) Metagenomics – a guide from sampling to data analysis. *Microb inf exp* 2(1):3
- Tjaden B (2015) De novo assembly of bacterial transcriptomes from RNA-seq data. *Genome Biol* 16:1
- Tottey W, Denonfoux J, Jaziri F, Parisot N, Missaoui M, Hill D et al (2013) The human gut chip “HuGChip”, an explorative phylogenetic microarray for determining gut microbiome diversity at family level. *PLoS One* 8(5), e62544
- Tringe SG, Hugenholtz P (2008) A renaissance for the pioneering 16S rRNA gene. *Curr Opin Microbiol* 11(5):442–446
- Tu Q, He Z, Li Y, Chen Y, Deng Y, Lin L et al (2014) Development of HuMiChip for functional profiling of human microbiomes. *PLoS One* 9(3), e90546
- Uchiyama T, Miyazaki K (2009) Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr Opin Biotechnol* 20(6):616–622
- Uhlik O, Leewis MC, Strejcek M, Musilova L, Mackova M, Leigh MB et al (2013) Stable isotope probing in the metagenomics era: a bridge towards improved bioremediation. *Biotechnol Adv* 31(2):154–165
- Ursell LK, Haider HJ, Van Treuren W, Garg N, Reddivari L, Vanamala J et al (2014) The intestinal metabolome: an intersection between microbiota and host. *Gastroenterology* 146(6):1470–1476
- Van den Abbeele P, Grootaert C, Marzorati M, Possemiers S, Verstraete W, Gerard P et al (2010) Microbial community development in a dynamic gut model is reproducible, colon region specific, and selective for Bacteroidetes and Clostridium cluster IX. *Appl Environ Microbiol* 76(15):5237–5246
- van Dorst J, Bissett A, Palmer AS, Brown M, Snape I, Stark JS et al (2014) Community fingerprinting in a sequencing world. *FEMS Microbiol Ecol* 89(2):316–330
- van Opijnen T, Camilli A (2013) Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat Rev Microbiol* 11(7):435–442
- Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J et al (2009) Shotgun metaproteomics of the human distal gut microbiota. *ISME J* 3(2):179–189
- Vetrovsky T, Baldrian P (2013) The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One* 8(2), e57923
- Wagner M (2009) Single-cell ecophysiology of microbes as revealed by Raman microspectroscopy or secondary ion mass spectrometry imaging. *Annu Rev Microbiol* 63:411–429
- Walker A (2014) Adding genomic ‘foliage’ to the tree of life. *Nat Rev Microbiol* 12(2):78
- Walker A, Parkhill J (2008) Single-cell genomics. *Nat Rev Microbiol* 6(3):176–177

- Walker AW, Duncan SH, Louis P, Flint HJ (2014) Phylogeny, culturing, and metagenomics of the human gut microbiota. *Trends Microbiol* 22(5):267–274
- Walker AW, Martin JC, Scott P, Parkhill J, Flint HJ, Scott KP (2015) 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice. *Microbiome* 3:26
- Wesolowska-Andersen A, Bahl MI, Carvalho V, Kristiansen K, Sicheritz-Ponten T, Gupta R et al (2014) Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. *Microbiome* 2:19
- Wikoff WR, Anfora AT, Liu J, Schultz PG, Lesley SA, Peters EC et al (2009) Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc Natl Acad Sci U S A* 106(10):3698–3703
- Wilmes P, Bond PL (2009) Microbial community proteomics: elucidating the catalysts and metabolic mechanisms that drive the Earth's biogeochemical cycles. *Curr Opin Microbiol* 12(3):310–317
- Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* 74(11):5088–5090

Microbiota of the Human Body

Implications in Health and Disease

Schwiertz, A. (Ed.)

2016, VII, 157 p. 25 illus., 21 illus. in color., Hardcover

ISBN: 978-3-319-31246-0