

Analyses of Intestinal Microbiota: Culture versus Sequencing

Andreas Hiergeist, Joachim Gläsner, Udo Reischl, and André Gessner

Andreas Hiergeist, PhD, and Joachim Gläsner, PhD, are senior scientists at the Institute of Clinical Microbiology and Hygiene, University Hospital Regensburg, Germany. Udo Reischl, PhD, is Head of Molecular Diagnostics and an associate professor for Medical Microbiology; and André Gessner, MD, PhD, is Director of the Institute of Clinical Microbiology and Hygiene, University Hospital Regensburg, Germany.

Address correspondence and reprint requests to Prof. André Gessner, Institute of Clinical Microbiology and Hygiene, University Hospital Regensburg, Franz-Josef-Strauss-Allee 11, 93053 Regensburg, Germany or email andre.gessner@ukr.de.

Abstract

Analyzing human as well as animal microbiota composition has gained growing interest because structural components and metabolites of microorganisms fundamentally influence all aspects of host physiology. Originally dominated by culture-dependent methods for exploring these ecosystems, the development of molecular techniques such as high throughput sequencing has dramatically increased our knowledge. Because many studies of the microbiota are based on the bacterial 16S ribosomal RNA (rRNA) gene targets, they can, at least in principle, be compared to determine the role of the microbiome composition for developmental processes, host metabolism, and physiology as well as different diseases. In our review, we will summarize differences and pitfalls in current experimental protocols, including all steps from nucleic acid extraction to bioinformatical analysis which may produce variation that outweighs subtle biological differences. Future developments, such as integration of metabolomic, transcriptomic, and metagenomic data sets and standardization of the procedures, will be discussed.

Key words: culturomics; metagenome; microbiome; next-generation sequencing

Introduction

Historically, the focus of research on microbial interactions with humans was set on single pathogenic organisms. Studies of colonizing, nonpathogenic microbes in the body as a whole were of minor interest because these organisms were thought to be benign, unlikely to have effects on human health like their pathogenic counterparts. The analyses of microbiomes have led to new interest in the communities of nonpathogenic microbes residing in distinct niches of the human body. Describing, ranking, and functional assignment of these organisms to shed light on a specific microbiome and to finally benefit from the knowledge attracts extensive attention.

Following the seminal invention of plating techniques by Robert Koch in 1881, microbiology was entirely culture

dependent throughout a century, requiring an established and proven protocol for growth of an organism to be analyzed as a pre-condition. Members of an unknown microbial community were identified by stains like Gram that used physiological or biochemical properties. This approach limited the range of detectable organisms to those that would proliferate in the setting of actual laboratory culture conditions which necessarily favored easily growing, aerobic organisms such as *Escherichia coli*. However, this bacterial genus accounts for approximately only 0.1% of the microbes inhabiting the average human intestine, whereas the majority of microbial species could never have been cultured, studied, or quantified in a laboratory.

Great progress was made by the advent of DNA-based culture-independent methods in the 1980s. The basic principle of this

methodology is to analyze the DNA extracted directly from a sample derived from the site of interest, in contrast to harvesting the bacterial DNA from in vitro isolated pure cultures. By doing this, researchers received a key tool to investigate several aspects of microbial communities (e.g., taxonomic composition and functional metagenomics) and (theoretically) to deduce potential biological tasks carried out by a community as a whole.

The earliest DNA-based methods probed extracted DNA of a microbial community for genes of interest by using fluorescent in situ hybridization (FISH), in which fluorescently labeled, specific oligonucleotide probes for marker genes are hybridized to the DNA (Amann et al. 1995). Alternatively, specific genes were amplified by polymerase chain reaction (PCR), cloned in *Escherichia coli* and subsequently sequenced (Ward et al. 1990). Although DNA sequencing techniques such as Sanger sequencing have been available since the mid-1970s (Sanger and Coulson 1975; Sanger et al. 1977), this traditional sequencing method was quite expensive and too time consuming for extensive use. Like FISH, sequencing of cloned fragments represents primarily a low-throughput technology and cannot deliver exhaustive insight into microbial diversity.

For more than 30 years, culture-independent microbial profiling has been based on the sequencing of a very important and convenient gene, the 16S ribosomal RNA (rRNA) gene (Olsen et al. 1986). In bacteria, the three rRNA molecules are genetically organized in a ribosome operon and primarily transcribed as a single 30S rRNA precursor that is subsequently cleaved by RNase III into 16S, 23S, and 5S rRNA subunits (Schlesinger et al. 1974). Operon size, sequences, and secondary structures of these three rRNA genes are conserved within a bacterial species (Maidak et al. 1997). The application of this gene for the assessment of bacterial taxa and their relationships was introduced by Carl Woese and colleagues in the late 1970s (Woese and Fox 1977) when it was shown principally that phylogenetic trees could be identified by comparing relatively stable parts of the genome (e.g., the 16S rRNA gene, which is one of several potential marker genes found in all bacteria and archaea). The alternating organization of the 16S rRNA gene featuring highly conserved and hypervariable sequences offers the advantage to employ universal PCR primers matching to constant sections in order to produce amplicons spanning discriminative regions. These regions reveal a sufficient interspecies variability and may be aligned to known sequences in reference databases to track microbial ecology and evolution (Yarza et al. 2014). Molecular rRNA-based microbial ecology dates back to 1990, when, for the first time, clone libraries of 16S rRNA genes from environmental bacteria (natural populations of Sargasso Sea picoplankton) were directly amplified and sequenced by the Sanger method (Giovannoni et al. 1990). This procedure represented a breakthrough that permanently changed the way prokaryotes in the environment were analyzed. Indeed, environmental metagenomic research provided basic tools and preceded application to the human and mouse body (Stein et al. 1996; Vergin et al. 1998).

About one decade later, in 2005, the revolutionary technology of high throughput (synonymously used term: “next-generation”) sequencing was introduced (Metzker 2005), exhibiting substantial advances over the Sanger method in terms of ease and cost of sequencing, as complete bacterial genome sequences could be assayed and dissected in hours or days rather than months or years. To investigate microbial communities efficiently and completely (i.e., to detect all members including the least abundant), deep sampling and high throughput DNA sequencing are the approaches of choice at present, probably in combination with other genome-wide analyses such as transcriptomics,

proteomics, or metabolomics. In this way, sequencing the DNA of an entire sample originating from the environment or an individual organism was definitely becoming economically feasible for numerous scientific institutions.

Shortly after invention of high throughput sequencing, centrally controlled, large-dimensioned research programs were initiated and performed by consortia of scientists mainly from the United States, but also from Europe and other regions throughout the world. Seminal works in this regard have been the *Metagenomics of the Human Intestinal Tract* (MetaHIT) project and the *Human Microbiome Project* (HMP). The MetaHIT was founded in 2008 and aimed to sequence the microbial genomes of fecal samples derived from both diseased (inflammatory bowel disease and obesity) and healthy individuals (Arumugam et al. 2011; Le Chatelier et al. 2013; Li et al. 2014; Qin et al. 2010). The budget of the 4.5-year-project comprised EUR 22 Mio. and was financed at roughly 50% by the European Union. The HMP is a U.S. National Institutes of Health initiative, running between 2008 and 2013 and endowed with U.S. \$115 Mio., with the objective to characterize the diversity of the microbiota sampled at multiple body sites exclusively in healthy humans (Huttenhower et al. 2012; Methé et al. 2012; Peterson et al. 2009; Weinstock 2012). Results are of special interest for future comparison analyses because they catalog the average composition of the gut microbiota from hundreds of apparently healthy individuals, thus serving as a valuable reference.

A series of currently emerging sequencing techniques, summarized by the term “third-generation sequencing,” is based on single-molecule real-time analyses. The PacBio RS II system (Pacific Biosciences) and the MinION nanopore device (Oxford Nanopore Technologies) represent two prominent instruments of this evolving new technology. By applying these sequencing approaches, some of the major problems of the next-generation sequencing (NGS) are resolved, because much longer reads beyond 10 kb are produced in a markedly reduced running time, saving considerable costs. In addition, there is no need for an amplification of samples, thus eliminating potential errors and biases. However, single-molecule reads still contain a high fraction of (stochastically spread) insertions and deletions that are compensated for in part by the high coverage. To date, applications other than microbiome analyses have been the main focus of these new technologies, although it may be expected that technical and analytical advances will make them attractive also for assessing microbial communities in the near future. Combining next- and third-generation sequencing data may be a desirable option in this regard.

In the past few years, the number of culture-independent metagenomic investigations and publications of the human and mouse microbiome has massively expanded, making it one of the most studied and interesting fields of microbiology, and potentially yielding profit to clinical practice. There is a wide range of disease phenotypes linked to the composition of the microbiota: chronic inflammatory diseases, obesity, diabetes, allergies, autism, depression, cardiovascular diseases, some cancer types, and even lung diseases have recently been reported to persist concomitantly with a distinct microbiome constellation (Marsland and Gollwitzer 2014; Sekirov et al. 2010; Shreiner et al. 2015). Although no causative or curative role is known to date for any of the microbial members detected in these approaches, one should estimate the ability to use the taxonomic as well as metagenomic data obtained from gut, lung, mouth, etc. as a diagnostic or prognostic biomarker for certain pathological entities or syndromes in the near future.

Culturomics: Methods and Caveats

Since the advent of the novel massively parallel high throughput DNA sequencing approaches and the market launch of 454 *Life Science's* GS20 sequencing machine nearly 10 years ago, microbiome research is undergoing a period of profound changes, mainly focusing now on molecular methods analyzing the composition of the microbiota in a variety of environments. Although molecular approaches, such as 16S rDNA amplicon or whole metagenome shotgun (WMS) sequencing, provide some clear benefits compared with culture-dependent methods by reason of their ability to provide direct and in-depth insights into the composition of the microbiota in a culture-independent manner, they seem to lack in the detection of low-abundant organisms. For example, Hugon and colleagues found that their 16S rDNA sequencing approach underestimated gram negatives when compared with bacterial counts observed from transmission electron microscopy (TEM) and Gram stain (Hugon et al. 2013).

Comparing the detection of bacterial isolates from a systematic culture approach, Lagier and colleagues estimated the threshold of available metagenomic next-generation sequencing methods to be $>10^6$ microorganisms per gram of feces. The lower detection limit was moreover dependent on the sequencing depth (Lagier, Armougom et al. 2012) of the chosen sequencing technique. Besides that depth bias, next-generation molecular approaches, especially those based on the amplification of the highly conserved 16S rRNA gene, inherently fail to detect intra-species variations. Deducing physiological host-microbe relationships from experiments with molecularly defined type strains, which have originally been isolated from completely different habitats, may generate distorted functional relationships. When the results of the HMP were published back in 2012, only few novel but many uncultivated ("most wanted") taxa were found within the operational taxonomic units (OTUs) that were identified from 18 body sites of 200 healthy volunteers (Fodor et al. 2012). These findings clearly demonstrate the need for further advancements of microbial cultivation techniques and targeted isolation approaches.

In general, it is estimated that $<20\%$ of environmental bacteria from all branches of the phylogenetic tree can be grown in defined growth media (Ward et al. 1990). The cultivation of microbes under verified laboratory conditions is still complicated for different possible reasons. Low-abundant and slow-growing oligotrophic microorganisms may be outcompeted by high-abundant and fast-growing species, while others fail to grow on conventional media because of inappropriate conditions regarding pH, redox state, temperature, or availability of essential nutrient molecules. Close interaction such as that observed in interspecies electron transfer between microorganisms in syntrophic relationships facilitating the decomposition of organic matter in anoxic environments (Stams and Plugge 2009) is an example that illustrates inextricable metabolic relationships in natural habitats. Several approaches have been established to cultivate previously unculturable microorganisms. Methods are mainly derived from the field of environmental microbiology and include, among others, the use of mixed culture or cocultivation with helper strains (Davis et al. 2014; Ohno et al. 1999) to facilitate growth. Signaling molecules like cAMP, homoserine lactones, or cell-free supernatants have been added successfully to enable growth of previously unculturable microorganisms (Bruns et al. 2003). Mimicking the native environment by the use of a diffusion chamber was also successful in growing previously uncultivated marine bacterial species (Kaeberlein et al. 2002).

The cultivation of microbes is essential for the understanding of the close physiological relationship between the host and strains of the gut microbiota. Approaches for the culture-based analysis of microorganisms have been carried out already more than 130 years ago. Since then, more than 1,000 different species from all three domains of life—bacteria, archaea, and eukarya—that have been found in the human gastrointestinal tract, are described in the scientific literature (Rajilić-Stojanović and de Vos 2014). Further developments in automatization of cultivation methods together with the simplification and acceleration of species identification by implementing mass spectrometry mark an important step toward a comprehensive high throughput strategy for the cultivation of yet unidentified species of the human gut microbiota (Lagier et al. 2015). Lagier and colleagues (Lagier, Armougom et al. 2012) transferred the knowledge of different approaches for the cultivation of fastidious bacteria into a high throughput approach for the comprehensive large-scale cultivation of gut microbial species. In this approach, referred to as "microbial culturomics," more than 200 different culture conditions were evaluated with variable physicochemical conditions. Dilution and filtration techniques and the targeted lysis of bacteria with the help of bacteriophages were applied for the reduction of biodiversity to enable the isolation of single cells. Approximation of the natural environment was achieved by the use of growth media containing sterile rumen or human fecal extracts or by cocultivation with *Amoeba* spp. for the isolation of intracellular growing microbes (Singh et al. 2013). Analyzing thousands of colonies (e.g., 32,500 in the study of Lagier et al.), the identification strategy is focused on MALDI-TOF mass spectrometry. Unidentifiable species were confirmed by sequencing the 16S rRNA gene.

Using the culturomics approach on the analysis of the fecal microbiota from two lean African and one obese European (Lagier, Armougom et al. 2012), 340 bacterial species were identified including 174 species that were not described in the human gut yet, together with 5 fungi and a new large Senegalvirus. Only 51 OTUs from sequencing the V6 region of the 16S rRNA overlapped with the cultivated species.

The group extended the successful implementation of this methodological principle to further samples analyzing the microbiota of a patient with resistant tuberculosis (Dubourg et al. 2013), anorexia nervosa (Pfleiderer et al. 2013), or the gut microbiota of patients treated with broad-spectrum antibiotics (Dubourg et al. 2014). They repeatedly discovered hitherto uncultivated species, together with not yet identified bacterial species. The overlap between the taxa observed from the molecular 16S rDNA-targeted approach and the species identified with culturomics was again very small.

This discrepancy can possibly be explained by intrinsic strategic differences between molecular and cultivation-based procedures, making a direct comparison difficult. An unbiased, quantitative rendering of the current ecological community status composed of countless interdependent (inter- and intraspecies) metabolic processes (e.g., cell-cell communication, biofilm formation) may be disturbed by preselected culture conditions. On the other side, the large number of cultivated species that were not represented by OTUs from 16S rDNA-based methods may also arise from an inefficient DNA extraction protocol. The application of culturomics in microbiome community analyses surely will have a great impact on the isolation of specific strains and the revelation of physiological links to syndromic diagnoses. The spread and application of this elaborative, time- and space-consuming methodology away from specialized laboratories to a wide-scale implementation will, however, be dependent on

further advancement with regard to automatization and miniaturization. Furthermore, care must be taken to ensure a valid taxonomic description of newly identified strains and the deposition in well-acknowledged culture collections allowing data validation and further phenotypic characterizations.

Another encouraging and interesting approach synergistically combines molecular and culture approaches to enable access to bacterial strains that have been previously identified from extensive metagenomic surveys, for further physiological studies. Ma and colleagues (Ma et al. 2014) developed a genetically targeted method for the cultivation of pure microbial strains, identified from 16S ribosomal RNA or metagenomic next-generation sequencing studies. A “chip wash” method in combination with a target-specific PCR detection on a novel microfluidic device was used to sequentially optimize cultivation conditions, which preferably enables the growth of the desired strain starting from a complex pool of cells. The target organism is then cultivated to microcolonies under the preselected conditions using a second microfluidic device performing 3,200 parallel cultivation experiments in a nanoliter scale. Each microcolony is split into two parts. The first part is used for PCR identification of the target organism, and the second part is used as an inoculum for a scale-up culture. The successful application of this method was proven by the isolation of a previously unidentified member of the Ruminococcaceae family from the HMP’s “most wanted taxa” list. The results of this study impressively illustrate the mutually beneficial effect, reconciling the advantages of molecular and culture-based approaches.

Nucleic Acid-based Analysis—Applications and Pitfalls

The application of NGS technologies and the development of novel protocols raised the field of microbiome analysis to a new level. Within a short time, historically used methods like construction and Sanger sequencing of metagenomic clone libraries, automated ribosomal internal transcribed spacer analysis (ARISA), terminal restriction fragment length polymorphism (T-RFLP), denaturing gradient gel electrophoresis (DGGE), or microarray-based methods were replaced by high throughput next-generation sequencing approaches generating millions to billions of short-fragment data within hours. Essentially, there are two options for the NGS-based investigation of microbial community structures on a genomic level: (1) PCR amplification of phylogenetically conserved marker sequences (e.g., 16S rRNA and 18S rRNA genes, ITS) with subsequent next-generation sequencing of the constructed amplicon library or (2) WMS sequencing of the whole genetic content present in a given complex sample. Both methods include the extraction of genomic DNA, construction of appropriate sequencing libraries, next-generation sequencing, bioinformatical analysis including quality control, and the comparison to reference databases. Both methods are nowadays widely used for the fast and comprehensive culture-independent analysis of microbial diversity and therefrom deduced interpretation of physiological correlations. However, the application of complex methodologies requires an accurate knowledge of the specific methodological pitfalls and their quality-controlled implementation. WMS sequencing enables the analysis of the microbial phylogenetic composition and functional diversity. More particularly, this approach allows conclusions on the altered potential of genetically encoded metabolic features among different conditions like carbohydrate and energy metabolism, biosynthesis of secondary metabolites,

signal transduction, and fermentation pathways, presence of regulatory sequences and other genetic features. On the other hand, marker gene-based approaches only allow the analysis of community structures, although bioinformatical approaches for the prediction of the functional metagenome composition from the combination of marker gene data and a database of reference genomes have been developed (Langille et al. 2013). Metagenomic approaches have some advantages over marker gene surveys in terms of the ability to detect microheterogeneity and genetic intraspecies variations, while bypassing the introduction of additional biases during the PCR amplification steps. On the other hand, comprehensive and quality-checked reference genome databases are not available to the same extent as intensively maintained ribosomal RNA databases (e.g., Greengenes, SILVA, or the Ribosomal Database Project [RDP]), although efforts are made to establish copious repositories such as the catalog of reference genes in the human gut microbiome (Li et al. 2014). Moreover, the presence or abundance of different genes may reveal little about the present state of gene expression in a given sample. In a comparison of metagenomic with metatranscriptomic data, Franzosa and colleagues found that 41% of detected microbial transcripts were not differentially regulated relative to their abundance (Franzosa et al. 2014), and only small parts of the present genes are expressed at a given point in time. The integration of gene expression data will enable researchers to create a more detailed picture of the dynamic response of the microbiota to its environment.

Massively parallel 16S rRNA gene sequencing is, however, less costly and less time consuming than WMS approaches, and pooling of barcoded amplicon libraries allows the analysis and comparison of hundreds of samples at one time. To capture the whole genetic information, WMS-based methods require much more effort, although marker gene-based approaches are also benefiting from higher sampling depths (Smith and Peay 2014). Even though sequencing costs per one megabase dropped from about |U.S. \$1,000 to U.S. \$ 0.1 between 2001 and 2011 (Sboner et al. 2011) and are likely to fall further, the economic factor is still of some importance for the realization of WMS sequencing. Irrespective of the sequencing approach for the analysis of microbial communities, both methods are subject to biases and systematic errors that can significantly affect downstream analyses. Strict observance of uniform sample handling and DNA extraction procedures is a prerequisite for the prevention of “home-grown” intrasample variations. Awareness of the intrinsic vulnerabilities of a sequencing technology and its advantages and disadvantages regarding read length, sequencing depth, and error profiles should be harmonized with the choice of downstream bioinformatical pipelines and reference databases in order to distinguish novel sequences from sequencing errors. However, Pylro and colleagues demonstrated that the same biological conclusion could be reached from results generated by different sequencing methodologies on two widely used next-generation sequencing platforms, provided that stringent downstream bioinformatical practices for clustering OTUs and quality filtering are applied (Pylro et al. 2014).

The overall accuracy of the analysis and the identified taxa in marker gene-based surveys is very much dependent on the choice and the taxa spectrum of the amplification primers and thus on the amplified 16S rDNA variable regions, which has a significant effect on the taxa coverage (Klindworth et al. 2012). Furthermore, selected universal amplification primers are mainly kingdom specific, and analyses are restricted to the examination of eubacterial, archaeal, or fungal sequences. Marker gene-based surveys are therefore often limited to the examination of

eubacterial diversity, while whole metagenome approaches, at least in principal, enable the inventory of genetic data comprising all kingdoms as well as viral sequences.

In addition to a variety of experimental methodologies, a vast range of analytical and bioinformatical tools are available for the processing of sequencing data. Even if bioinformatical pipelines (e.g., QIIME [Caporaso et al. 2010] and *mothur* [Schloss et al. 2009]) are offering valuable open source software packages and algorithms for the quality control, clustering, and graphical presentation of marker gene-based next-generation sequencing data, they often do not provide standardized procedures. With respect to the comparability of microbiome data obtained from different surveys, the methodical realization as well as data processing practices must be taken into account precisely, which is sometimes considerably impeded by inadequate descriptions in some current publications. For example, strong variations are observed after applying different approaches for denoising of sequencing data in order to correct for sequencing or PCR-based errors (Gaspar and Thomas 2013) or algorithms and settings for the clustering of sequencing reads to OTUs (Patin et al. 2013). Changing the stringency of these approaches can lead to over- or underestimation of species richness due to sequencing errors. Also, the general use of library size normalization by rarifying sequencing counts for the detection of differentially abundant species is currently being discussed (McMurdie and Holmes 2014). Beyond that, comparing the taxonomies of the three most commonly used curated 16S rRNA sequence databases, Greengenes, SILVA, and RDP-II, has indicated significant differences in naming and abundance of taxa between these three repositories (Yilmaz et al. 2014).

Currently available next-generation sequencing platforms differ significantly with regard to the obtained read length, sequencing depth, and inherent error profiles. The ongoing technological evolution in many fields of microbiome research and the further integration of additional data from metaproteomics, metabolomics, or metatranscriptomics to study functional microbe–host relationships in depth will require new quality control and analysis strategies. The application of statistical and ecological theories along with the use of diverse bioinformatical and molecular biological methodologies, as well as the medical contextualization, will render microbiome analysis to a highly interdisciplinary area of research. Only when issues that introduce bias can be readily identified and overcome by the implementation of procedures for methodological standardization and quality control, clinical applications of microbiome research will be feasible.

Nontrivial DNA Isolation: Peel it, Cook it or Forget it

Considering the increasing relevance of metagenomic data for interpretation of metabolic, immunological, or neurological disorders, it is essential to fully reflect the present microbial community structure. Great efforts are undertaken to correct for bias in downstream analyses by extensive bioinformatical approaches. Beyond other inherent procedural variations of both 16S rRNA amplicon sequencing and whole metagenome-based methods, the introduction of PCR or sequencing errors (Pinto and Raskin 2012), the initial steps of sample retrieval, and the implementation of a uniform DNA extraction method are often overlooked aspects. These variables are crucial for DNA-based microbial community analyses and the optimal comparability between different datasets and different studies.

As described very recently (Zarrinpar et al. 2014), the gut microbiome is highly dynamic, exhibiting daily cyclical changes that are dependent on the feeding/fasting rhythm. Thus, the exact timing of sampling, in addition to diet, is an important factor influencing the diversity and composition of gut microbiota.

Protocols for the isolation of highly purified DNA from a wide variety of specimens have been developed for many years, and numerous choices exist, including enzymatic lysis with lysozyme, mutanolysin, and Proteinase K; or surfactants like cetrionium bromide (CTAB) or sodium dodecyl sulfate (SDS); strong chaotropic agents; or physical methods like sonication, freezing-thawing or repeated bead-beating. A deliberate combination of protocol steps is critical to keep the balance between a profound cell disruption and low DNA degradation on the one hand and a uniform DNA extraction on the other hand. Applying variable storage times to human fecal and dermal samples, ranging from short-term storage to a maximum of 2 weeks at room temperature prior to deep freezing at -20°C or -80°C , had no significant impact on the relative distribution of bacteria (Lauber et al. 2010), but freezing/thawing may disturb the composition of bacterial communities (Bahl et al. 2012; Mølbak et al. 2006). Proper storage and handling of samples is instead much more critical for metatranscriptome analyses to prevent any degradation of the less stable RNA molecules. Furthermore, the stool water content, typically ranging from about 70% for hard and formed stool to $>85\%$ for liquid stools (Bliss et al. 1999), or sample homogenization seem only to have little effect on the relative abundance of individual bacterial taxa, but strong bias can be introduced into microbiome data with the application of inappropriate DNA extraction protocols (Santiago et al. 2014).

The fecal microbiota consists of hundreds of bacterial species from more than 30 different phyla, most of which belong to Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria, and Verrucomicrobia (Rajilić-Stojanovic et al. 2007). The biggest difference among the microorganisms with regard to rigidity is the gram-positive or gram-negative structure of the cell wall. At least half of the gut-residing bacteria are gram-positive (Gossling and Slack 1974; Lagier, Million et al. 2012), and the commonly hard-to-lyse domain of the Archaea is represented predominantly by methanogenic *Methanobrevibacter* spp. with a variable and overall low prevalence (Hoffmann et al. 2013). Moreover, the microbiota comprises taxa containing aerobic and anaerobic bacteria belonging to the genera of *Bacillus*, *Clostridium*, and others within the Firmicutes, potentially forming endospores, which are probably the most robust cellular structures known. Mechanical disruption of the cells has been shown to be superior to other methods, and repeated bead-beating (RBB) is critical especially for the lysis of gram-positive Eubacteria and Archaea and showed the highest bacterial diversity when compared with methods involving lysis with enzymatic cocktails containing lysozyme and/or mutanolysin (Salonen et al. 2010). The combination of repeated bead-beating with a freeze-thaw protocol for cell lysis also had positive effects on the DNA extraction from challenging gram-positive bacteria or fungi (Sergeant et al. 2012). Comparing the bead-beating-based DNA extraction protocols from fecal specimens used in the two most extensive microbiome studies carried out in the last years, the MetaHIT and the HMP, significant differences in the overall DNA yield and in the Bacteroidetes/Firmicutes ratio were obtained from WMS sequencing. This was obviously caused by variations in the lysis efficiency between both protocols for some of the most frequent genera within the phyla of Bacteroidetes, Firmicutes, and Proteobacteria (Wesolowska-Andersen et al. 2014).

A quality-controlled implementation of a uniform DNA extraction method suitable for all bacterial taxa is challenging. Even after the application of different DNA isolation protocols to defined bacterial communities found in the oral microbiota, species representation varied significantly, arising mainly from the under-representation of gram positives like *Streptococcus mutans* (Abusleme et al. 2014). While the implementation of comprehensive DNA extraction methods is required, too harsh lysis protocols lead to the partial degradation of the released DNA and probably to an underestimation of easily lysed gram-negative bacteria (Hugon et al. 2013) in the assessment of bacterial richness. This effect, however, is less weighty when using sequencing technologies and NGS libraries with short read length (Santiago et al. 2014) but will become more important with the technological development of long-read platforms. For the now evolving third-generation sequencing technologies like nanopore and single-molecule real-time (SMRT) sequencing, reaching read lengths of 10 kilobases and longer, the preparation of high-quality and high-molecular-weight DNA from complex samples will become even more crucial for a comprehensive census of the microbiota.

Particularly in the context of the great diversity and complexity of samples of human origin, like fecal specimens, the presence of complex polysaccharides, bile salts, lipids, urate, or other PCR inhibitors may require additional extraction steps like chromatographic purification, chloroform extraction, treatment with activated carbon, sample dilution, addition of BSA, or the selection of resistant polymerases to cope with the presence of inhibitory substances (Schrader et al. 2012). In addition to the uniform lysis of microbial cells, the deployment of highly pure DNA extracts suitable for the subsequent PCR amplification of marker genes or ligation of oligonucleotide adaptors for the generation of next-generation sequencing libraries is essential.

16S rRNA-based Amplicon Sequencing: Length Matters

Since the invention of massively parallel NGS technologies, the DNA sequencing market is rapidly changing. Even if new third-generation real-time sequencing technologies still have to prove their operational capability for the analysis of microbiota, it can be anticipated that new developments will, among other improvements, imply changes in respect to throughput, sequencing depth, and read length. Especially read length and sequencing depth as well as the platform-specific error profiles will significantly determine the outcomes of marker gene-based NGS surveys for the assessment of microbial community structures. Even if a great sequencing depth together with very long reads of highest quality is desirable, current sequencing technologies are only capable of striking the balance between these intended objectives. According to the manufacturer's current specifications, sequencing with the GS FLX+ instrument (Roche/454) allows users to achieve 1 million sequencing reads with a maximum length of 1,000 bp in 23 hours, while the popular MiSeq benchtop sequencer (Illumina) produces up to 25 million 2×300 bp paired-end reads. The PacBio RS II platform, as an example of a third-generation real-time sequencing platform, produces up to 50,000 reads with a maximal length of 40 kilobases (modal length > 14 kb) per SMRT cell in 30–240 min (Liu et al. 2012). Error rates vary from 0.1% to 2% for pyrosequencing, semiconductor sequencing (e.g., IonTorrent), and Illumina sequencing platforms to more than 10% for (single-pass) SMRT sequencing reads. Read quality positionally decreases in Roche/

454 pyrosequencing and Illumina's sequencing by synthesis approach, and errors are accumulating especially at the 3' end of the reads, whereas PacBio RS II errors are stochastically distributed (Fichot and Norman 2013). The accumulation of errors can easily generate false positive variants overestimating species richness in 16S-based microbiome data. Read length and error characteristics have to be considered together when analyzing next-generation sequencing data. To improve sensitivity and accuracy in next-generation sequencing experiments for the analysis of microbial community structures, one straightforward approach is to increase sequencing depth (Smith et al. 2014). The implementation of a quality control pipeline, including accurate sequencing error correction procedures, is important, and only high-quality reads should be considered for the subsequent analysis of microbial diversity.

DNA sequencing of the universally distributed 16S rRNA gene has been used for a long time as a gold standard to determine the phylogenetic relationships of prokaryotes. It is currently the only taxonomic marker, for which curated databases containing comprehensive taxonomic information exist. Discriminating sites of the 16S rRNA gene are located in nine variable regions (V1 – V9), which are important for accurate richness estimations of microbial diversity, but other regions are also contributing significantly to the discriminative power (Vinje et al. 2014). The design of oligonucleotides for the amplification of 16S rRNA gene and the encompassed variable regions should be optimized in respect to their overall coverage and their phyletic spectrum (Klindworth et al. 2012).

The 16S rRNA gene sequence of *Escherichia coli* is 1542 bp long. In order to overcome the shortage of currently available sequencing technologies regarding read length, there have already been various successful attempts to reach a higher resolution by integrating short reads obtained from multiple variable regions of the 16S rRNA gene (Amir et al. 2013) or by the assembly of the whole 16S rDNA sequence from short reads (Miller et al. 2013). However, the alignment of short reads, especially when assembling genes containing highly conserved regions like in the 16S rRNA gene, can be challenging and prone to errors. Comparing microbial diversity and taxonomic assignment by trimming 16S rRNA (V1 – V4) sequencing data of rice root microbiomes from conventional and pyrosequencing to obtain reads of varying length, Okubo and colleagues found significant differences on the genus level caused by the overestimation of *Bradyrhizobium spp.*, while no deviations were observed up to the family level (Okubo et al. 2012). Reaching similar conclusions, Yarza and colleagues predicted from OTU clustering of partial 16S ribosomal RNA sequences that near full-length fragments longer than 1,300 nucleotides are required for a comprehensive and reliable estimation of taxa richness, especially for the classification of high taxonomic ranks (Yarza et al. 2014). Currently, the GS FLX+ sequencer (Roche/454) using Titanium XL+ chemistry is capable of reaching read lengths up to 1,000 bp. With the development of new sequencing platforms, a full-length analysis of the 16S rRNA gene will be possible. One limiting factor is that currently only 23% of the 16S rRNA sequences published are longer than 900 bp (Yarza et al. 2014). The evolution of sequencing technology and the feeding of databases with full-length reads are mutually dependent, and the extent of high-quality full-length 16S rRNA data will rapidly increase. Mosher and colleagues analyzed the capability of the PacBio RS II sequencer to obtain full-length reads of 16S rRNA amplicons generated from metagenomic environmental samples. While high error rates (17%–18%) dramatically overestimated species richness in their initial study back in 2013 (Mosher et al. 2013), further improvements of the sequencing chemistry

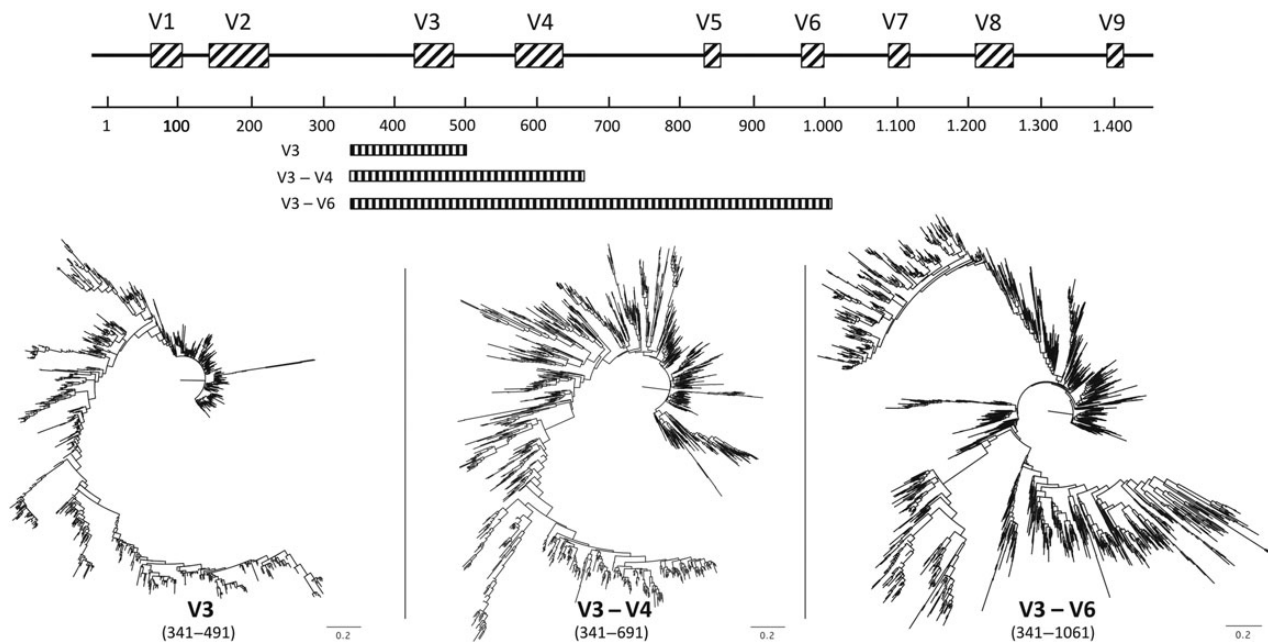


Figure 1 Effect of sequencing read length on the phylogenetic representation of the mouse gut microbiota. PCR amplicons were generated from mouse fecal DNA using the universal eubacterial 16S rDNA primers 341F and 1061R and sequenced on the Roche/454 GS Junior device. 131,634 raw full-length reads (790 nt, 16S rRNA gene positions 341–1061, spanning V3–V6) were truncated to 150 nt (positions 341–491, spanning V3) and to 350 nt (positions 341–691, spanning V3–V4), respectively. OTUs were de novo clustered to 97% using UCLUST v1.2.22q algorithm, and approximately maximum-likelihood phylogenetic trees were generated with FastTree v2.1.3 after PyNAST 1.2.2.2 alignment of a representative sequence set to the *Greengenes* v13.08 reference database. Trees in the lower part of the figure were represented graphically with FigTree 1.4.2. In the upper part of the figure, the rRNA gene sequence with variable regions V1 to V9 (hatched boxes) is depicted.

improved the outcomes significantly, allowing the accurate identification of microorganisms to the species level in environmental samples (Mosher et al. 2014). This demonstrates the fast-paced development and the capacity of new innovative sequencing technologies after careful error monitoring and troubleshooting to accommodate the methodological requirements.

In conclusion, long read lengths are clearly superior to short fragments, not only with regard to 16S ribosomal RNA-based surveys (Figure 1), but also in WMS sequencing experiments performed to analyze the phylogenetic composition and functional diversity within microbial communities. Short reads (150–400 bp) miss a significant amount of BLAST homologs, and genetic functional classes within WMS libraries are better detected by increased read length (750 bp) than by greater sampling depth (Wommack et al. 2008).

PCR Contaminations

In innumerable cases, the application of PCR has demonstrated its power to generate billions of molecules from very small amounts (e.g., even a single copy) of template nucleic acids. However, the high analytical sensitivity of nucleic acid amplification techniques includes one important deficiency. The high level of vulnerability to contamination events raises a major problem, which has taught us to implement comprehensive procedures and advanced spatial concepts to accurately prevent carry-over events of template DNA or PCR products, especially in diagnostic laboratories (Scherczinger et al. 1999). Salter and colleagues impressively demonstrated that the introduction of contaminating microbial DNA in nucleic acid-based microbiome analyses is a considerable burden for both 16S rRNA gene sequencing and WMS surveys (Salter et al. 2014). The isolation of metagenomic DNA from assumed “ultrapure water” using different extraction

kits in various laboratories with subsequent 16S rRNA gene amplicon sequencing resulted in the observance of contaminating water- and soil-dwelling bacteria of the genera *Burkholderia*, *Mesorhizobium*, *Hydrotaea*, and *Bradyrhizobium*, which are frequently associated with nitrogen fixation (nitrogen blanketing is widely used in water storage tanks). *Bradyrhizobium* was found to be a common contaminant in microbiome datasets. When WMS sequencing of DNA extracted from *Salmonella bongori* culture dilutions was performed, the contaminating genera were especially predominant in diluted samples, implying that contamination is more critical when analyzing samples of low bacterial biomass like blood or lung tissues. Possible consequences are the distortion of the microbial composition and loss of low-abundant bacteria due to competition between template DNA and broad range amplification primers.

Widely used decontamination procedures like exonuclease treatment with DNase I or ultraviolet irradiation of PCR reagents often lower the overall sensitivity of PCR and, in addition, the success is variable and undefined (Mennerat and Sheldon 2014). The implementation and sequencing of negative controls beyond the exclusion of contaminating reads from microbiome datasets using bioinformatical techniques are possible approaches to get rid of polluting DNA sequences. Potential sources of contamination are used reagent kits for DNA isolation (Erlwein et al. 2011; Evans et al. 2003), PCR reagents (Tilburg et al. 2010), water, PCR primers (Goto et al. 2005), or aerosols (Witt et al. 2009). PCR amplicons from broad-range 16S rDNA amplification experiments can be considered in particular as a source of bias, and a prevention of cross-contamination is often only accomplished by implementing efficient but costly strategies (Champlot et al. 2010). Another conceivable source of bias is the cross-contamination of reads due to a crosstalk of barcode sequences (unique sample-specific DNA-based identifiers). This barcode bias can either be

introduced during the process of primer synthesis and purification (Quail et al. 2014) or be due to incorrect bioinformatical correction of sequencing errors in the barcode sequence.

The impact of microbial DNA originating from ingested food on the contamination of fecal microbiome datasets is also not fully elucidated so far. In vivo studies of DNA persistence in the gastrointestinal tract of germfree and humanized rats showed no degradation of DNA in the lower part, although partial degradation of plant DNA occurred in the upper part of the intestine. The food matrix protected DNA from low pH and degradation in the stomach. Chloroplast-derived DNA was found along the whole gastrointestinal tract, and directly fed plasmid DNA was not degraded and was even biologically active, as chemically competent *E. coli* cells were still transformable with fecal DNA extracts derived from the respective animals (Wilcks et al. 2004). Autoclaving effectively kills bacterial cells, but the integrity of bacterial DNA is not affected by the sterilization process (Yap et al. 2013). Sterilized animal chow is probably still contaminated with microbial DNA, which could affect downstream nucleic acid-based analyses of the microbiota in experimental animal models. Further investigation will be necessary to detect so far undiscovered sources of contamination, and the reasonable implementation of process controls will support progress in the prevention of contamination events.

Dealing with PCR Artifacts: LEA-PCR

Amplicon sequencing of the bacterial 16S rRNA gene from fecal microbiota has revealed an individual collection of species with great variation in species numbers. While culture-based techniques indicated ~100 species, several times this number of species are suggested by the results of 16S rRNA amplicon sequencing, even after *in silico* attempts to remove chimeric molecules formed during PCR or errors in the sequencing process. These artifacts lead to uncalculable numbers of false positives and complicate tracking of individual bacterial taxa across time. To circumvent these problems, Faith and colleagues (Faith et al. 2013) developed a novel method for 16S rRNA amplicon sequencing to assay the bacterial composition of the gut microbiota at higher depth and precision. They found that sequencing a sample beyond 10,000 reads did not substantially increase the lower detection limit possible at high precision. Increasing sequence quantity rather than sequence quality is the strategy in WMS, where redundant sequencing of genomes at 10- to 50-fold coverage results in far lower error rates than single reads. To redundantly sequence DNA fragments, it is necessary to create a limited DNA pool that is smaller than the amount of sequencing reads available and to individually label the molecules in the pool. To adapt these techniques to 16S rRNA amplicon sequencing, Faith and colleagues developed a method called low-error amplicon sequencing (LEA-Seq). In this method, a bottleneck is created by a linear PCR extension of the template DNA using a barcoded and diluted oligonucleotide primer solution where each oligonucleotide is labeled with a distinct random barcode positioned at the 5' end of the universal 16S rRNA primer sequence. In the next step, the linear PCR pool is exponentially PCR amplified, using primers that specifically amplify the linear PCR molecules. During the exponential PCR, an index primer is added to the amplicons with a third primer to allow pooling of multiple samples in the subsequent sequencing run. The pool of products from the exponential PCR is finally sequenced at sufficient depth to redundantly (~20-fold coverage) sequence the initial linear amplicons. In this way, the multiple reads for each barcode allow the generation of an error-corrected consensus

sequence for the initial template molecule. In the newly developed LEA protocol, the linear PCR primers were diluted to concentrations that generate ~150,000 amplicon reads at the above-mentioned coverage per amplicon. Using this promising new strategy, Faith and colleagues could demonstrate that the majority of the bacterial strains in an individual's microbiota persist for years and thus potentially shape different aspects of the host physiology over long periods of time.

Identifying the “Unculturables”

Bioinformatics for Genome Assembly

Current computational analysis strategies for analyzing metagenomic data rely on comparisons to reference genomes, but the diversity of the individual microbiota extends far beyond what is covered by reference databases. *De novo* assembly of complex metagenomic data into complete separate genome information of particular bacterial strains or viruses was not possible until recently. In 2014, the MetaHIT consortium published a method based on binning co-abundant genes across a series of metagenomic samples of the same type (Nielsen et al. 2014). This allows for the first-time comprehensive discovery of new microbial organisms and viruses, enabling the assembly of microbial genomes without the need for reference genome sequences. The method was applied on data from nearly 400 human gut microbiome samples and identified more than 7,000 co-abundance gene groups, which were used to assemble 238 high-quality microbial genomes, including 181 new genomes from previously unsequenced species. This method thus appears to be suitable for comprehensive profiling of the diversity within complex metagenomic microbiome samples.

Prediction of Culture Conditions

One interesting approach to growing uncultured bacteria uses transcript information from bacteria to determine the particular aspect of their habitat that is important to their growth rather than empirically testing numerous media additives and growth conditions. Bomar and colleagues used high throughput sequencing of RNA transcripts (RNA-seq) to determine that a previously uncultured *Rikenella*-like bacterium in the leech gut was utilizing mucin as an energy and carbon source (Bomar et al. 2011). With this information, the authors succeeded in culturing this isolate on medium containing mucin. This led to the suggestion that the RNA sequence information might be more useful for this purpose than the genomic DNA sequence, as RNA-seq indicates which genes are actually being expressed in the growing bacterium.

In vivo Culturomics: SFB in Germfree Mice

As mentioned above, the microbes responsible for many aspects of host physiology remain unclear so far. However, the use of germfree animals or gnotobiotics associated with defined and limited numbers of particular microbes, including not-yet-cultivable bacteria, is expected to resolve many of these questions.

A breakthrough example for an “in vivo culturomics” approach concerning so-far-uncultivable bacteria was published in 1991 by Klaasen and colleagues for segmented filamentous bacteria (SFB) (Klaasen et al. 1991). SFB are spore-forming, gram-positive bacteria that were originally identified in the ilea of mice and rats. The first publication about these unique SFB appeared in 1974 (Davis and Savage 1974). Since then, SFB have been

identified in the ilia of several other species including rabbits, guinea pigs, cattle, cats, horses, turkeys, pigs, and, most recently, humans. Interestingly, SFB appear to be strictly host specific, and metagenomic comparisons have identified these bacteria as close relatives of *Clostridium* spp. (for review see [Ericsson et al. 2014](#)).

Klaasen and colleagues described the successful mono-association of germfree mice via intraileal inoculation of ethanol-treated ileal contents of donor mice and presented evidence that cage mates of the recipient mice were also mono-associated with SFB. The availability of these animals (i.e., in vivo monocultures of SFB) allowed for the first-time molecular and taxonomic characterization of these bacteria. In addition, a number of studies meanwhile demonstrated that SFB influence the development of the gut immune system (e.g., T helper cell development and activation as well as IgA production) and that these bacteria play important roles in different autoimmune diseases ([Ericsson et al. 2014](#)).

Conclusions

Up until now, the vast majority of microbiome studies were limited to descriptions of the taxa and (relative) numbers of microbial organisms that exist in or on a specific part of the mammalian body or in an environmental system. More rigorous standardization of analysis procedures, including timing of sampling, pre-analytic handling, and DNA preparation, as well as PCR and bioinformatics, are required to enhance comparability of studies performed in different laboratories. Furthermore, technical advances should be channeled and emphasis should be shifted toward the biological function in order to get insight into regulatory relationships within a microbiome population as well as mutualistically between host and microbiota. Undoubtedly, this will be a very profitable task.

Acknowledgments

We thank Claudia Deinzer, Christine Irtenkauf, and Nadja Reul for excellent technical assistance and Holger Melzl, Rainer Spang, and Frank Stämmler for constant discussion. This work was supported by the German Research Society (DFG, SPP 1656, GE 671/14-1).

References

- Abusleme L, Hong BY, Dupuy AK, Strausbaugh LD, Diaz PI. 2014. Influence of DNA extraction on oral microbial profiles obtained via 16S rRNA gene sequencing. *J Oral Microbiol* 23:6. doi:10.3402/jom.v6.23990.
- Amann RI, Ludwig W, Schleifer KH. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59:143–169.
- Amir A, Zeisel A, Zuk O, Elgart M, Stern S, Shamir O, Turnbaugh PJ, Soen Y, Shental N. 2013. High-resolution microbial community reconstruction by integrating short reads from multiple 16S rRNA regions. *Nucleic Acids Res* 41:e20541.
- Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Doré J; MetaHIT Consortium, Antolín M, Artiguenave F, Blottiere HM, Almeida M, Brechot C, Cara C, Chervaux C, Cultrone A, Delorme C, Denariáz G, Dervyn R, Foerstner KU, Friss C, van de Guchte M, Guedon E, Haimet F, Huber W, van Hylckama-Vlieg J, Jamet A, Juste C, Kaci G, Knol J, Lakhdari O, Layec S, Le Roux K, Maguin E, Mérieux A, Melo Minardi R, M'rini C, Muller J, Oozeer R, Parkhill J, Renault P, Rescigno M, Sanchez N, Sunagawa S, Torrejon A, Turner K, Vandemeulebrouck G, Varela E, Winogradsky Y, Zeller G, Weissenbach J, Ehrlich SD, Bork P. 2011. Enterotypes of the human gut microbiome. *Nature* 473:174–180.
- Bahl MI, Bergström 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* 329:193–197.
- Bliss DZ, Savik K, Jung H, Jensen L, LeMoine M, Lowry A. 1999. Comparison of subjective classification of stool consistency and stool water content. *J Wound Ostomy Cont Nurs* 26:137–141.
- Bomar L, Maltz M, Colston S, Graf J. 2011. Directed culturing of microorganisms using metatranscriptomics. *mBio* 2:e00012–11. doi:10.1128/mBio.00012-11.
- Bruns A, Nübel U, Cypionka H, Overmann J. 2003. Effect of signal compounds and incubation conditions on the culturability of freshwater bacterioplankton. *Appl Environ Microbiol* 69:1980–1989.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JL, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336.
- Champlot S, Berthelot C, Pruvost M, Bennett EA, Grange T, Geigl E-M. 2010. An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PLoS One* 5:e13042.
- Davis IJ, Bull C, Horsfall A, Morley I, Harris S. 2014. The Unculturables: targeted isolation of bacterial species associated with canine periodontal health or disease from dental plaque. *BMC Microbiol* 14:196.
- Davis CP, Savage DC. 1974. Habitat, succession, attachment, and morphology of segmented, filamentous microbes indigenous to the murine gastrointestinal tract. *Infect Immun* 10:948–956.
- Dubourg G, Lagier JC, Armougom F, Robert C, Hamad I, Brouqui P, Raoult D. 2013. The gut microbiota of a patient with resistant tuberculosis is more comprehensively studied by culturomics than by metagenomics. *Eur J Clin Microbiol Infect Dis* 32:637–645.
- Dubourg G, Lagier JC, Robert C, Armougom F, Hugon P, Metidji S, Dione N, Dangui NPM, Pfliegerer A, Abrahao J, Musso D, Papazian L, Brouqui P, Bibi F, Yasir M, Vialettes B, Raoult D. 2014. Culturomics and pyrosequencing evidence of the reduction in gut microbiota diversity in patients with broad-spectrum antibiotics. *Int J Antimicrob Agents* 44:117–124.
- Ericsson AC, Hagan CE, Davis DJ, Franklin CL. 2014. Segmented filamentous bacteria: Commensal microbes with potential effects on research. *Comparative Med* 64:90–98.
- Erlwein O, Robinson MJ, Dustan S, Weber J, Kaye S, McClure MO. 2011. DNA extraction columns contaminated with murine sequences. *PLoS One* 6:e23484.
- Evans GE, Murdoch DR, Anderson TP, Potter HC, George PM, Chambers ST. 2003. Contamination of Qiagen DNA extraction kits with *Legionella* DNA. *J Clin Microbiol* 41:3452–3453.

- Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, Clemente JC, Knight R, Heath AC, Leibel RL, Rosenbaum M, Gordon JI. 2013. The long-term stability of the human gut microbiota. *Science* 341:6141–6145.
- Fichot EB, Norman RS. 2013. Microbial phylogenetic profiling with the Pacific Biosciences sequencing platform. *Microbiome* 1:10.
- Fodor AA, DeSantis TZ, Wylie KM, Badger JH, Ye Y, Hepburn T, Hu P, Sodergren E, Liolios K, Huot-Creasy H, Birren BW, Earl AM. 2012. The “most wanted” taxa from the human microbiome for whole genome sequencing. *PLoS One* 7:e41294.
- Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM, Giannoukos G, Boylan MR, Ciulla D, Gevers D, Izard J, Garrett WS, Chan AT, Huttenhower C. 2014. Relating the meta-transcriptome and metagenome of the human gut. *Proc Natl Acad Sci USA* 111:E2329–E2338.
- Gaspar JM, Thomas WK. 2013. Assessing the consequences of denoising marker-based metagenomic data. *PLoS One* 8:e60458.
- Giovannoni SJ, Britschgi TB, Moyer CL, Field KG. 1990. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* 345:60–63.
- Gossling J, Slack JM. 1974. Predominant gram-positive bacteria in human feces: numbers, variety, and persistence. *Infect Immun* 9:719–729.
- Goto M, Ando S, Hachisuka Y, Yoneyama T. 2005. Contamination of diverse nifH and nifH-like DNA into commercial PCR primers. *FEMS Microbiol Lett* 246:33–38.
- Hoffmann C, Dollive S, Grunberg S, Chen J, Li H, Wu GD, Lewis JD, Bushman FD. 2013. Archaea and fungi of the human gut microbiome: Correlations with diet and bacterial residents. *PLoS One* 8:e66019.
- Hugon P, Lagier J-C, Robert C, Lepolard C, Papazian L, Musso D, Vialettes B, Raoult D. 2013. Molecular studies neglect apparently gram-negative populations in the human gut microbiota. *J Clin Microbiol* 51:3286–3293.
- Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, Creasy HH, Earl AM, FitzGerald MG, Fulton RS, Giglio MG, Hallsworth-Pepin K, Lobos EA, Madupu R, Magrini V, Martin JC, Mitreva M, Muzny DM, Sodergren EJ, Versalovic J, Wollam AM, Worley KC, Wortman JR, Young SK, Zeng Q, Aagaard KM, Abolude OO, Allen-Vercoe E, Alm EJ, Alvarado L, Andersen GL, Anderson S, Appelbaum E, Arachchi HM, Armitage G, Arze CA, Ayvaz T, Baker CC, Begg L, Belachew T, Bhonagiri V, Bihan M, Blaser MJ, Bloom T, Bonazzi V, Brooks J, Buck GA, Buhay CJ, Busam DA, Campbell JL, Canon SR, Cantarel BL, Chain PS, Chen IM, Chen L, Chhibba S, Chu K, Ciulla DM, Clemente JC, Clifton SW, Conlan S, Crabtree J, Cutting MA, Davidovics NJ, Davis CC, DeSantis TZ, Deal C, Delehaunty KD, Dewhirst FE, Deych E, Ding Y, Dooling DJ, Dugan SP, Dunne WM, Durkin A, Edgar RC, Erlich RL, Farmer CN, Farrell RM, Faust K, Feldgarden M, Felix VM, Fisher S, Fodor AA, Forney LJ, Foster L, Di Francesco V, Friedman J, Friedrich DC, Fronick CC, Fulton LL, Gao H, Garcia N, Giannoukos G, Giblin C, Giovanni MY, Goldberg JM, Goll J, Gonzalez A, Griggs A, Gujja S, Haake SK, Haas BJ, Hamilton HA, Harris EL, Hepburn TA, Herter B, Hoffmann DE, Holder ME, Howarth C, Huang KH, Huse SM, Izard J, Jansson JK, Jiang H, Jordan C, Joshi V, Katancik JA, Keitel WA, Kelley ST, Kells C, King NB, Knights D, Kong HH, Koren O, Koren S, Kota KC, Kovar CL, Kyrpidis NC, La Rosa PS, Lee SL, Lemon KP, Lennon N, Lewis CM, Lewis L, Ley RE, Li K, Liolios K, Liu B, Liu Y, Lo CC, Lozupone CA, Lunsford R, Madden T, Mahurkar AA, Mannon PJ, Mardis ER, Markowitz VM, Mavromatis K, McCorrison JM, McDonald D, McEwen J, McGuire AL, McInnes P, Mehta T, Mihindukulasuriya KA, Miller JR, Minx PJ, Newsham I, Nusbaum C, O’Laughlin M, Orvis J, Pagani I, Palaniappan K, Patel SM, Pearson M, Peterson J, Podar M, Pohl C, Pollard KS, Pop M, Priest ME, Proctor LM, Qin X, Raes J, Ravel J, Reid JG, Rho M, Rhodes R, Riehle KP, Rivera MC, Rodríguez-Mueller B, Rogers YH, Ross MC, Russ C, Sanka RK, Sankar P, Sathirapongsasuti J, Schloss JA, Schloss PD, Schmidt TM, Scholz M, Schriml L, Schubert AM, Segata N, Segre JA, Shannon WD, Sharp RR, Sharpton TJ, Shenoy N, Sheth NU, Simone GA, Singh I, Smillie CS, Sobel JD, Sommer DD, Spicer P, Sutton GG, Sykes SM, Tabbaa DG, Thiagarajan M, Tomlinson CM, Torralba M, Treangen TJ, Truty RM, Vishnivetskaya TA, Walker J, Wang L, Wang Z, Ward DV, Warren W, Watson MA, Wellington C, Wetterstrand KA, White JR, Wilczek-Boney K, Wu Y, Wylie KM, Wylie T, Yandava C, Ye L, Ye Y, Yooseph S, Youmans BP, Zhang L, Zhou Y, Zhu Y, Zoloth L, Zucker JD, Birren BW, Gibbs RA, Highlander SK, Methé BA, Nelson KE, Petrosino JF, Weinstock GM, Wilson RK, White O. The Human Microbiome Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486:207–214.
- Kaeberlein T, Lewis K, Epstein SS. 2002. Isolating “uncultivable” microorganisms in pure culture in a simulated natural environment. *Science* 296:1127–1129.
- 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* 156:148–151.
- Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glockner FO. 2012. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41:e1.
- Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, Bittar F, Fournous G, Gimenez G, Maraninchi M, Trape JF, Koonin EV, La Scola B, Raoult D. 2012. Microbial culturomics: Paradigm shift in the human gut microbiome study. *Clin Microbiol Infect* 18:1185–1193.
- Lagier JC, Hugon P, Khelafia S, Fournier PE, La Scola B, Raoult D. 2015. The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. *Clin Microbiol Rev* 28:237–264.
- Lagier JC, Million M, Hugon P, Armougom F, Raoult D. 2012. Human gut microbiota: Repertoire and variations. *Front Cell Infect Microbiol* 2:136.
- Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkpile DE, Vega Thurber RL, Knight R, Beiko RG, Huttenhower C. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* 31:814–821.
- Lauber CL, Zhou N, Gordon JI, Knight R, Fierer N. 2010. Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. *FEMS Microbiol Lett* 307:80–86.
- Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M, Batto JM, Kennedy S, Leonard P, Li J, Burgdorf K, Garup N, Jørgensen T, Brandslund I, Nielsen HB, Juncker AS, Bertalan M, Levenez F, Pons N, Rasmussen S, Sunagawa S, Tap J, Tims S, Zoetendal EG, Brunak S, Clément K, Doré J, Kleerebezem M, Kristiansen K, Renault P, Sicheritz-Ponten T, de Vos WM, Zucker JD, Raes J, Hansen T; MetaHIT consortium, Bork P, Wang J, Ehrlich SD, Pedersen O. 2013. Richness of human gut microbiome correlates with metabolic markers. *Nature* 500:541–546.

- Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, Arumugam M, Kultima JR, Prifti E, Nielsen T, Juncker AS, Manichanh C, Chen B, Zhang W, Levenez F, Wang J, Xu X, Xiao L, Liang S, Zhang D, Zhang Z, Chen W, Zhao H, Al-Aama JY, Edris S, Yang H, Wang J, Hansen T, Nielsen HB, Brunak S, Kristiansen K, Guarner F, Pedersen O, Doré J, Ehrlich SD, Bork P, Wang J. 2014. An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol* 32:834–841.
- Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. 2012. Comparison of next-generation sequencing systems. *J Biomed Biotechnol* 2012:251364.
- Ma L, Kim J, Hatzenpichler R, Karymov Ma, Hubert N, Hanan IM, Chang EB, Ismagilov RF. 2014. Gene-targeted microfluidic cultivation validated by isolation of a gut bacterium listed in Human Microbiome Project's Most Wanted taxa. *Proc Natl Acad Sci USA* 111:9768–9773.
- Maidak BL, Olsen GJ, Larsen N, Overbeek R, McCaughey MJ, Woese CR. 1997. The RDP (Ribosomal Database Project). *Nucleic Acids Res* 25:109–111.
- Marsland BJ, Gollwitzer ES. 2014. Host-microorganism interactions in lung diseases. *Nat Rev Immunol* 14:827–835.
- McMurdie PJ, Holmes S. 2014. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput Biol* 10:e1003531.
- Mennerat A, Sheldon BC. 2014. How to deal with PCR contamination in molecular microbial ecology. *Microb Ecol* 68:834–841.
- Méthé BA, Nelson KE, Pop M, Creasy HH, Giglio MG, Huttenhower C, Gevers D, Petrosino JF, Abubucker S, Badger JH, Chinwalla AT, Earl AM, FitzGerald MG, Fulton RS, Hallsworth-Pepin K, Lobos EA, Madupu R, Magrini V, Martin JC, Mitreva M, Muzny DM, Sodergren EJ, Versalovic J, Wollam AM, Worley KC, Wortman JR, Young SK, Zeng Q, Aagaard KM, Abolude OO, Allen-Vercoe E, Alm EJ, Alvarado L, Andersen GL, Anderson S, Appelbaum E, Arachchi HM, Armitage G, Arze CA, Ayvaz T, Baker CC, Begg L, Belachew T, Bhonagiri V, Bihan M, Blaser MJ, Bloom T, Bonazzi VR, Brooks P, Buck GA, Buhay CJ, Busam DA, Campbell JL, Canon SR, Cantarel BL, Chain PS, Chen IM, Chen L, Chhibba S, Chu K, Ciulla DM, Clemente JC, Clifton SW, Conlan S, Crabtree J, Cutting MA, Davidovics NJ, Davis CC, DeSantis TZ, Deal C, Delehaunty KD, Dewhirst FE, Deych E, Ding Y, Dooling DJ, Dugan SP, Dunne W Jr, Durkin A, Edgar RC, Erlich RL, Farmer CN, Farrell RM, Faust K, Feldgarden M, Felix VM, Fisher S, Fodor AA, Forney L, Foster L, Di Francesco V, Friedman J, Friedman DC, Fronick CC, Fulton LL, Gao H, Garcia N, Giannoukos G, Giblin C, Giovanni MY, Goldberg JM, Goll J, Gonzalez A, Griggs A, Gujja S, Haas BJ, Hamilton HA, Harris EL, Hepburn TA, Herter B, Hoffmann DE, Holder ME, Howarth C, Huang KH, Huse SM, Izard J, Jansson JK, Jiang H, Jordan C, Joshi V, Katancik JA, Keitel WA, Kelley ST, Kells C, Kinder-Haake S, King NB, Knight R, Knights D, Kong HH, Koren O, Koren S, Kota KC, Kovar CL, Kyrpides NC, La Rosa PS, Lee SL, Lemon KP, Lennon N, Lewis CM, Lewis L, Ley RE, Li K, Liolios K, Liu B, Liu Y, Lo CC, Lozupone CA, Lunsford R, Madden T, Mahurkar AA, Mannon PJ, Mardis ER, Markowitz VM, Mavrommatis K, McCorrison JM, McDonald D, McEwen J, McGuire AL, McInnes P, Mehta T, Mihindukulasuriya KA, Miller JR, Minx PJ, Newsham I, Nusbaum C, O'Laughlin M, Orvis J, Pagani I, Palaniappan K, Patel SM, Pearson M, Peterson J, Podar M, Pohl C, Pollard KS, Priest ME, Proctor LM, Qin X, Raes J, Ravel J, Reid JG, Rho M, Rhodes R, Riehle KP, Rivera MC, Rodriguez-Mueller B, Rogers YH, Ross MC, Russ C, Sanka RK, Sankar P, Sathirapongsasuti J, Schloss JA, Schloss PD, Schmidt TM, Scholz M, Schriml L, Schubert AM, Segata N, Segre JA, Shannon WD, Sharp RR, Sharpton TJ, Shenoy N, Sheth NU, Simone GA, Singh I, Smillie CS, Sobel JD, Sommer DD, Spicer P, Sutton GG, Sykes SM, Tabbaa DG, Thiagarajan M, Tomlinson CM, Torralba M, Treangen TJ, Truty RM, Vishnivetskaya TA, Walker J, Wang L, Wang Z, Ward DV, Warren W, Watson MA, Wellington C, Wetterstrand KA, White JR, Wilczek-Boney K, Wu YQ, Wylie KM, Wylie T, Yandava C, Ye L, Ye Y, Yooseph S, Youmans BP, Zhang L, Zhou Y, Zhu Y, Zoloth L, Zucker JD, Birren BW, Gibbs RA, Highlander SK, Weinstock GM, Wilson RK, White O. 2012. A framework for human microbiome research. *Nature* 486:215–221.
- Metzker ML. 2005. Emerging technologies in DNA sequencing. *Genome Res* 15:1767–1776.
- Miller CS, Handley KM, Wrighton KC, Frischkorn KR, Thomas BC, Banfield JF. 2013. Short-read assembly of full-length 16S amplicons reveals bacterial diversity in subsurface sediments. *PLoS One* 8:e56018.
- Mølbak L, Sommer HM, Johnsen K, Boye M, Johansen M, Møller K, Leser TD. 2006. Freezing at –800 degrees C distorts the DNA composition of bacterial communities in intestinal samples. *Curr Issues Intest Microbiol* 7:29–34.
- Mosher JJ, Bernberg EL, Shevchenko O, Kan J, Kaplan LA. 2013. Efficacy of a 3rd generation high-throughput sequencing platform for analyses of 16S rRNA genes from environmental samples. *J Microbiol Methods* 95:175–181.
- Mosher JJ, Bowman B, Bernberg EL, Shevchenko O, Kan J, Korlach J, Kaplan LA. 2014. Improved performance of the PacBio SMRT technology for 16S rDNA sequencing. *J Microbiol Methods* 104:59–60.
- Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, Plichta DR, Gautier L, Pedersen AG, Le Chatelier E, Pelletier E, Bonde I, Nielsen T, Manichanh C, Arumugam M, Batto JM, Quintanilha Dos Santos MB, Blom N, Borruel N, Burgdorf KS, Boumezbour F, Casellas F, Doré J, Dworzynski P, Guarner F, Hansen T, Hildebrand F, Kaas RS, Kennedy S, Kristiansen K, Kultima JR, Léonard P, Levenez F, Lund O, Moumen B, Le Paslier D, Pons N, Pedersen O, Prifti E, Qin J, Raes J, Sørensen S, Tap J, Tims S, Ussery DW, Yamada T, Sicheritz-Ponten T, Bork P, Wang J, Brunak S, Ehrlich SD, MetaHIT Consortium. 2014. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat Biotechnol* 32:822–828.
- Ohno M, Okano I, Watsuji T, Kakinuma T, Ueda K, Beppu T. 1999. Establishing the independent culture of a strictly symbiotic bacterium *Symbiobacterium thermophilum* from its supporting *Bacillus* strain. *Biosci Biotechnol Biochem* 63:1083–1090.
- Okubo T, Ikeda S, Yamashita A, Terasawa K, Minamisawa K. 2012. Pyrosequence read length of 16S rRNA gene affects phylogenetic assignment of plant-associated bacteria. *Microbes Environ* 27:204–208.
- Olsen GJ, Lane DJ, Giovannoni SJ, Pace NR, Stahl DA. 1986. Microbial ecology and evolution: a ribosomal RNA approach. *Annu Rev Microbiol* 40:337–365.
- Patin NV, Kunin V, Lidström U, Ashby MN. 2013. Effects of OTU clustering and PCR artifacts on microbial diversity estimates. *Microb Ecol* 65:709–719.
- Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C, Baker CC, Di Francesco V, Howcroft TK, Karp RW, Lunsford RD, Wellington CR, Belachew T, Wright M, Giblin C, David H,

- Mills M, Salomon R, Mullins C, Akolkar B, Begg L, Davis C, Grandison L, Humble M, Khalsa J, Little AR, Peavy H, Pontzer C, Portnoy M, Sayre MH, Starke-Reed P, Zakhari S, Read J, Watson B, Guyer M. 2009. The NIH human microbiome project. *Genome Res* 19:2317–2323.
- Pfleiderer A, Lagier J-C, Armougom F, Robert C, Vialettes B, Raoult D. 2013. Culturomics identified 11 new bacterial species from a single anorexia nervosa stool sample. *Eur J Clin Microbiol Infect Dis* 32:1471–1481.
- Pinto AJ, Raskin L. 2012. PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PLoS One* 7:e43093.
- Pyro VS, Roesch LFW, Morais DK, Clark IM, Hirsch PR, Tótolá MR. 2014. Data analysis for 16S microbial profiling from different benchtop sequencing platforms. *J Microbiol Methods* 107:30–37.
- Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Doré J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J; MetaHIT Consortium, Bork P, Ehrlich SD, Wang J. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464:59–65.
- Quail MA, Smith M, Jackson D, Leonard S, Skelly T, Swerdlow HP, Gu Y, Ellis P. 2014. SASI-Seq: sample assurance Spike-Ins, and highly differentiating 384 barcoding for Illumina sequencing. *BMC Genomics* 15:110.
- Rajilić-Stojanović M, de Vos WM. 2014. The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiol Rev* 38:996–1047.
- Rajilić-Stojanović M, Smidt H, De Vos WM. 2007. Diversity of the human gastrointestinal tract microbiota revisited. *Environ Microbiol* 9:2125–2136.
- Salonen A, Nikkilä J, Jalanka-Tuovinen J, Immonen O, Rajilić-Stojanović M, Kekkonen RA, Palva A, de Vos WM. 2010. Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: Effective recovery of bacterial and archaeal DNA using mechanical cell lysis. *J Microbiol Methods* 81:127–134.
- Salter S, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman N, Walker AW. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 12:87.
- Sanger F, Coulson AR. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* 94:441–448.
- Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467.
- Santiago A, Panda S, Mengels G, Martinez X, Azpiroz F, Dore J, Guarner F, Manichanh C. 2014. Processing faecal samples: a step forward for standards in microbial community analysis. *BMC Microbiol* 14:112.
- Sboner A, Mu XJ, Greenbaum D, Auerbach RK, Gerstein MB. 2011. The real cost of sequencing: higher than you think!. *Genome Biol* 12:125.
- Scherzinger CA, Ladd C, Bourke MT, Adamowicz MS, Johannes PM, Scherzinger R, Beesley T, Lee HC. 1999. A systematic analysis of PCR contamination. *J Forensic Sci* 44:1042–1045.
- Schlessinger D, Ono M, Nikolaev N, Silengo L. 1974. Accumulation of 30S preribosomal ribonucleic acid in an *Escherichia coli* mutant treated with chloramphenicol. *Biochemistry* 13:4268–4271.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537–7541.
- Schrader C, Schielke A, Ellerbroek L, Johne R. 2012. PCR inhibitors - occurrence, properties and removal. *J Appl Microbiol* 113:1014–1026.
- Sekirov I, Russell SL, Antunes L, Caetano M, Finlay BB. 2010. Gut microbiota in health and disease. *Physiol Rev* 90:859–904.
- Sergeant MJ, Constantinidou C, Cogan T, Penn CW, Pallen MJ. 2012. High-throughput sequencing of 16S rRNA gene amplicons: Effects of extraction procedure, primer length and annealing temperature. *PLoS One* 7:e38094.
- Shreiner AB, Kao JY, Young VB. 2015. The gut microbiome in health and in disease. *Curr Opin Gastroenterol* 31:69–75.
- Singh S, Eldin C, Kowalczywska M, Raoult D. 2013. Axenic culture of fastidious and intracellular bacteria. *Trends Microbiol* 21:92–99.
- Smith DP, Peay KG. 2014. Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. *PLoS One* 9:e90234.
- Stams AJM, Plugge CM. 2009. Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nat Rev Microbiol* 7:568–577.
- Stein JL, Marsh TL, Wu KY, Shizuya H, DeLong EF. 1996. Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J Bacteriol* 178:591–599.
- Tilburg JJHC, Nabuurs-Franssen MH, van Hannen EJ, Horrevorts AM, Melchers WJG, Klaassen CHW. 2010. Contamination of commercial PCR master mix with DNA from *Coxiella burnetii*. *J Clin Microbiol* 48:4634–4635.
- Vergin KL, Urbach E, Stein JL, DeLong EF, Lanoil BD, Giovannoni SJ. 1998. Screening of a fosmid library of marine environmental genomic DNA fragments reveals four clones related to members of the order Planctomycetales. *Appl Environ Microbiol* 64:3075–3078.
- Vinje H, Almøy T, Liland KH, Snipen L. 2014. A systematic search for discriminating sites in the 16S ribosomal RNA gene. *Microb Inform Exp* 4:2.
- Ward DM, Weller R, Bateson MM. 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* 345:63–65.
- Weinstock GM. 2012. Genomic approaches to studying the human microbiota. *Nature* 489:250–256.
- Wesolowska-Andersen A, Bahl MI, Carvalho V, Kristiansen K, Sicheritz-Pontén T, Gupta R, Licht TR. 2014. Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. *Microbiome* 2:19.
- Wilcks A, van Hoek AH, Joosten RG, Jacobsen BB, Aarts HJ. 2004. Persistence of DNA studied in different ex vivo and in vivo rat models simulating the human gut situation. *Food Chem Toxicol* 42:493–502.
- Witt N, Rodger G, Vandesompele J, Benes V, Zumla A, Rook GA, Huggett JF. 2009. An assessment of air as a source of DNA

- contamination encountered when performing PCR. *J Biomol Tech* 20:236–240.
- Woese CR, Fox GE. 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc Natl Acad Sci USA* 74:5088–5090.
- Wommack KE, Bhavsar J, Ravel J. 2008. Metagenomics: read length matters. *Appl Environ Microbiol* 74:1453–1463.
- Yap JM, Goldsmith CE, Moore JE. 2013. Integrity of bacterial genomic DNA after autoclaving: possible implications for horizontal gene transfer and clinical waste management. *J Hosp Infect* 83:247–249.
- Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer KH, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* 12:635–645.
- Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO. 2014. The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. *Nucleic Acids Res* 42:643–648.
- Zarrinpar A, Chaix A, Yooseph S, Panda S. 2014. Diet and feeding pattern affect the diurnal dynamics of the gut microbiome. *Cell Metab* 20:1006–1017.