Hornung et al. Genes & Nutrition (2018) 13:5
https://doi.org/10.1186/s12263-018-0594-6

Genes & Nutrition

REVIEW Open Access

Studying microbial functionality within the gut ecosystem by systems biology

Bastian Hornung[1*], Vitor A. P. Martins dos Santos[1], Hauke Smidt[2] and Peter J. Schaap[1]

Abstract

Humans are not autonomous entities. We are all living in a complex environment, interacting not only with our peers, but as true holobionts; we are also very much in interaction with our coexisting microbial ecosystems living on and especially within us, in the intestine. Intestinal microorganisms, often collectively referred to as intestinal microbiota, contribute significantly to our daily energy uptake by breaking down complex carbohydrates into simple sugars, which are fermented to short-chain fatty acids and subsequently absorbed by human cells. They also have an impact on our immune system, by suppressing or enhancing the growth of malevolent and beneficial microbes. Our lifestyle can have a large influence on this ecosystem. What and how much we consume can tip the ecological balance in the intestine. A "western diet" containing mainly processed food will have a different effect on our health than a balanced diet fortified with pre- and probiotics.

In recent years, new technologies have emerged, which made a more detailed understanding of microbial communities and ecosystems feasible. This includes progress in the sequencing of PCR-amplified phylogenetic marker genes as well as the collective microbial metagenome and metatranscriptome, allowing us to determine with an increasing level of detail, which microbial species are in the microbiota, understand what these microorganisms do and how they respond to changes in lifestyle and diet. These new technologies also include the use of synthetic and in vitro systems, which allow us to study the impact of substrates and addition of specific microbes to microbial communities at a high level of detail, and enable us to gather quantitative data for modelling purposes.

Here, we will review the current state of microbiome research, summarizing the computational methodologies in this area and highlighting possible outcomes for personalized nutrition and medicine.

Keywords: Microbiome, Systems biology, Modelling, NGS, Metagenome, Metatranscriptome, Genome scale metabolic model, Gut, Community interactions, Microbial ecology

Background

The gut is an essential part of the human body. It has so much influence on our well-being that it even has been dubbed a "second brain" by the media [1, 2], and in recent years this "superorgan" inhabited by trillions of microorganisms has triggered a large amount of scientific interest.

The microbial communities residing in the different parts of the gut are among the main contributors to its functioning and therefore also directly influence health. The recent availability of high-throughput methods (metagenomics and other omics) have improved our insights into these ecosystems dramatically. Figure 1 summarizes the current state of meta-omics (all nucleotide sequencing approaches, as well as metaproteomics and meta-metabolomics) research with an intestinal focus (for details regarding the literature search methodology, see Additional file 1). Not surprisingly, the largest body of research has been focused on humans (Fig. 1d), but other (model) organisms including pigs, rodents (mice, rats) and fishes (mainly zebrafish) have also been investigated. Non-model organisms are also under investigation, but for different purposes such as the potential biotechnological applicability of lignin degradation by termite gut microbial species [3].

Over the trajectory of the human gut, the microbiome has a varying degree of complexity [4, 5] (Fig. 1b). In general, microbial density increases from the duodenum until it reaches its maximum in the colon and faeces. At

* Correspondence: Bastian.hornung@gmx.de
[1]Laboratory of Systems and Synthetic Biology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, the Netherlands
Full list of author information is available at the end of the article

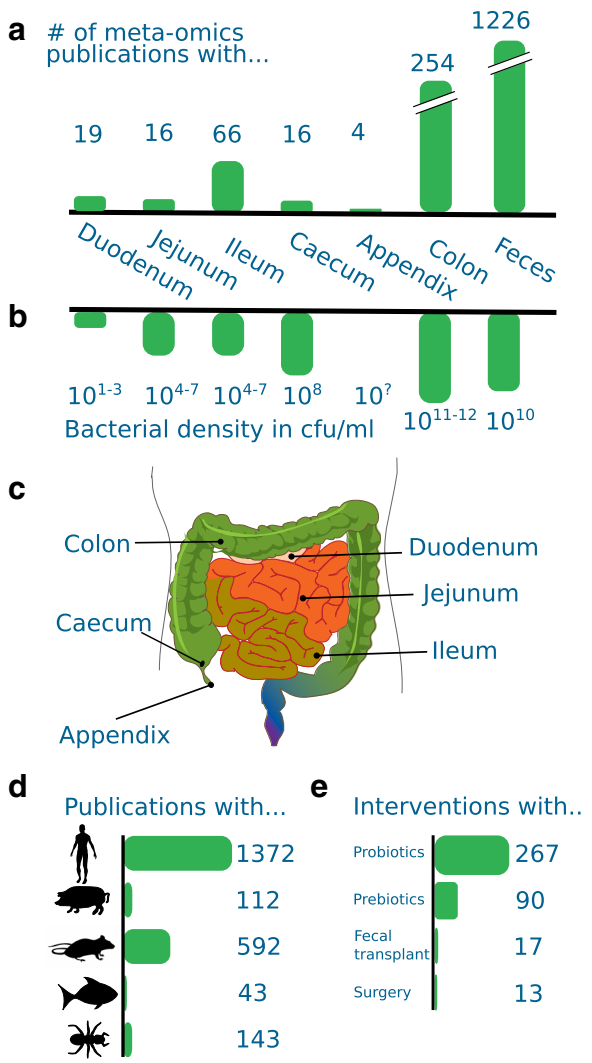

Fig. 1 The gut in the focus of meta-omics science. An overview of **a** main sampling sites and **b** microbial complexity is given, together with **c** an overview over the physiology. **d** The number of the studied hosts and **e** methods to improve gut health are indicated. All data was retrieved via PubMed searches for the corresponding terms. For the exact search terms, please see Additional file 1

the same time, these two parts are also the most studied parts (Fig. 1a). While the high complexity of the community at these specific sites makes them interesting research sites, other parts of the (healthy) human gut remain grossly under-sampled, which is mainly due to inaccessibility. Along the trajectory of the human gut, the focus of microbial metabolic activities changes profoundly, with the small intestine having a higher capacity to degrade simpler carbohydrates [6], whereas in the colon mostly complex carbohydrates are degraded [7].

Most human omics studies are observational, aimed at studying microbial diversity and function as well as host-microbe interactions; however, a number of studies directly aim at improving gut health (and in proxy, individual health, Fig. 1e). These interventional studies can be broadly classified into two categories: pre-clinical and clinical interventions. Pre-clinical interventions focus mostly on improving gut health via changes in nutrition. In this field, the concept of probiotics (administering of beneficial bacteria [8]) is probably the most widely known, also in the eye of the general public, due to a wide array of commercially available products. Most interventional studies have focused on these probiotics, with a smaller part investigating the benefits of prebiotics (substrates enhancing the growth of beneficial bacteria in the gut; for a review, see [7]). Clinical interventions in response to conditions associated with a chronic disruption of intestinal homeostasis such as ulcerative colitis, and IBS with for example faecal transplants and bariatric surgery, have only been reported in a few publications [9, 10].

With all these studies, many important factors have been discovered regarding the ecology of the human microbiome.

The human microbiota: symbiosis, competition and other relationships

Our microbiota is an important part of our personal ecosystem, which is assumed to be composed of more than a trillion microbial cells [11], approximately equalling the amount of human cells in our body [12]. Whereas the microbial ecosystems associated with some niches of the human body like for example the vagina [13] have a low complexity with only a few different inhabitants, most body sites contain hundreds of different microbes [11]. Like in macro-ecology, they perform different roles and thus can have different relationships with each other and with the host. In the microbiota, a broad range of different interactions exist, ranging from mutualistic and commensal to predatory relationships, and competition for the same niche exists. The nature of these relationships has an impact on the habitat itself, and imbalances with respect to the abundance and function of specific members can lead to an imbalance of the whole ecosystem. Many bacteria like for example *Akkermansia muciniphila* [14] have a good symbiotic relationship with their host. They degrade the carbohydrates supplied by the host, and other bacteria benefit from the breakdown products of this degradation process. This leads to the production of host beneficial compounds like short-chain fatty acids (SCFA; mainly acetate, propionate, butyrate) [15], which can be for example used by human colonocytes as energy source [16] or directly be incorporated into the human metabolism as additional carbon sources [17]. In other cases, this symbiosis applies to nutrition-derived carbohydrates that are not (fully) digested by host-derived enzymes in the small intestine such as resistant starch and other complex carbohydrates [7]. These might only be broken down by specific combinations of microorganisms for further catabolization.

This can be exemplified by consortia of Bifidobacteria [18], which lead to the liberation of otherwise inaccessible substrates from for example indigestible plant biomass like cellulose components. In both scenarios, the liberated substrates can be further metabolized by other bacteria (e.g. [19]) to host beneficial compounds. Parasitic relationships also exist, like for example between *Actinomyces odontolyticus* and TM7 [20], where the parasitizing TM7 might eventually kill its microbial host. There are also predatory relationships, e.g. bacteria of the genus *Bdellovibrio* prey on other bacteria as source of energy and therefore help to regulate the diversity and balances of bacterial populations [21, 22]. Imbalances in the ecosystem might lead to bacterial overgrowth, which makes the ecosystem in general less resilient to perturbations [23]. Blooms of bacteria, e.g. *Clostridium difficile*, which infects more than half a million individuals per year and leads to 29,000 deaths in the USA alone [24], will have a directly noticeable impact. The produced toxins in such an outbreak will not only affect the microbiota [25] but will also lead to a direct disease state of the host [26]. Therefore, understanding of internal and external factors that affect composition and functioning of this ecosystem, such as for example nutrition intake, antibiotic intake, symbiotic or predatory relationships, are essential for being able to characterize and predict the state and functioning of this ecosystem. All of these challenge the intrinsic emergent community properties such as resilience, stability and its efficiency to provide nutrients for the host.

Metabolic syndrome and the microbiome

The metabolic syndrome is a complex disorder with high associated cost and is mainly characterized by four sub-pathologies: Obesity, elevated blood sugar/insulin resistance/diabetes type II, elevated blood pressure and dyslipidemia [27, 28]. Although genetics [29] and lifestyle [30] play major roles, the microbiome also contributes to all of these main sub-pathologies.

Obesity might provide the most direct link. It has been shown that gut microbiota composition in obese and lean individuals is significantly different [31]. The microbiome is an important factor in carbohydrate degradation and uptake. Microbial metabolism on average contributes to up to 10% of the daily calorie intake [32], and potentially in obese subjects, this contribution could be increased [33]. This is mainly due to the degradation of carbohydrates, which due to the lack of necessary catabolic enzymes, are not directly accessible for the human host. These carbohydrates are converted by the microbiota into SCFA, thereby directly contributing to the energy intake of the host [34]. Since not all microorganisms are capable of such conversions, species diversity and abundance will directly influence the types of carbohydrates that can be converted into SCFA and therefore how much of the non-digestible carbohydrates will be utilized by the host-microbe holobiont. While some bacteria are specialized in carbohydrate breakdown, like for example *Bacteroides thetaiotaomicron* [35], others mainly rely on their peers to scavenge nutrients [36]. A microbial community consisting mainly of carbohydrate degraders will therefore be more beneficial for the host providing valuable nutrients. It is tempting to speculate that in case of obesity this beneficial trait has turned disadvantageous and might contribute to an increased risk towards metabolic syndrome-associated pathologies.

Such differences in microbial composition have also been causally linked to obesity. It has been shown that transplantation of an "obese microbiome" into germ-free animals causes an increase in body fat as compared to control animals inoculated with a "lean microbiome" [33, 37, 38], indicating that the increased capacity to harvest energy is transferred with the microbiome.

The involvement of the gut microbiome in the second most prevalent pathology, elevated blood sugar/insulin resistance leading to diabetes type II, can be explained via an indirect route, starting from inflammation. Even without an obvious disease phenotype, low-grade inflammation might be present [39], caused by yet unidentified bacteria. This inflammation is hypothesized to be one of the causes of the metabolic syndrome [39, 40] and to be an early stage of Inflammatory Bowel Disease, including Ulcerative Colitis and Crohn's Disease [41]. An invasion of bacteria into the intestinal tissue causes the presence of endotoxins (LPS, flagellin) in the blood stream, leading to chronic inflammation in the intestinal tissue. It has been suggested that as a physiological response to inflammation the blood glucose level is increased to serve as additional energy source for the various immune cells [42]. Since the inflammation is chronic, so will be the elevated glucose levels. In the long term this might lead to insulin resistance and type II diabetes [43].

The connection between the composition of the human gut microbiota and the third and fourth pathology, elevated blood pressure and dyslipidemia, is less well characterized [44]. It has been demonstrated with cross-over experiments that gut microbiota from rats with elevated blood pressure will transfer this physiological trait to receiving rats [45]. It has also been shown that inflammatory processes [46] and effects on the nervous system [47] will affect blood pressure, but a full understanding of these relationships is still missing. For dyslipidemia, the relationship is also rather unclear, due to its strong association with obesity [48]. The clearest mode of action until now are effects of the microbiota on bile acid metabolism, which is critical for the absorption of lipids [49], but the observed associations are currently not linked to known mechanisms [50, 51].

Top down: how to investigate the microbiome

In contrast to macro-ecology, in microbial ecology, it is possible to capture nearly the whole biodiversity of a habitat by sequencing its associated total DNA and/or specific phylogenetic marker genes. Different omics techniques can give the researcher information about species diversity and abundance, about their metabolic capabilities and associated symbiosis or pathogenicity factors. Technically, there are different ways of obtaining this information but the ultimate goal of omics approaches is to answer the following set of questions: Who is there, what can they do and what are they actually doing?

While in macro-ecology, specimen can normally be collected and studied in captivity; this is usually not the case for microbial ecosystems. It is assumed that we can only cultivate less than 1% of the bacterial diversity [52]. The rest, the so called "dark matter" cannot be readily captured by cultivation [53], although much progress has been made in recent years with high-throughput culturing, the so called "culturomics" [54]. While bacteria make up most of the diversity of the human microbiota, archaea are also present in humans [55], as well as a high diversity of phages [56]. Fungi and protozoa also exist in this ecosystem, but are less well studied [57]. Why the majority of this biodiversity cannot be cultured is not clear, but different hypotheses exist. One of these hypotheses is that these organisms cannot survive on their own because of community dependencies. They are for instance microorganisms that live in a strict syntrophic relationship and are sharing nutrients and metabolites [58]. Syntrophic relationships might be due to excretion and uptake of common metabolites, but also more intricate cross-feeding networks have been reported to exist [6, 59, 60]. Other types of non-metabolic interactions also exist but are less easily quantifiable. Biofilms, which occur frequently in human-associated microbiomes [61], are often not the product of a single species, but of a community [62]. They are not controlled by direct metabolic dependencies but by other mechanisms like quorum sensing [63].

Omics approaches towards understanding of the who and what of microbial communities

To answer the "who", the "what can they do", the "what are they actually doing" and "how do they respond to a diet or otherwise environmental change", different approaches can be used. To answer the "who", low-cost amplicon sequencing of 16S ribosomal RNA (16S rRNA) encoding genes can be utilized. The 16S rRNA gene is present in all prokaryotes and slowly mutating due to structural and catalytic constraints. Some of the secondary structure elements, called regions V for variable 1 to 9, are less constrained and therefore over time accumulate mutations more rapidly than other more conserved regions. Together, sequence variation within conserved and variable regions can be transformed into an evolutionary distance, allowing interference of the phylogeny of all members within a microbial community. As knowing the community composition in most studies is a prerequisite, next generation sequencing (NGS) of PCR amplicons targeting a selection of these variable regions is the most widely used approach. Despite the fact that no genomes are sequenced, this is often falsely referred to as "metagenomics". This should be avoided and proper terminology should be used [64]. Nevertheless, making use of the currently available information from genomes and metagenomes, species identification in part also allows for predictions of functional capabilities [65, 66], albeit with inherent limitations with respect to their accuracy especially for understudied environments that are less well represented in currently available (meta)genome databases [67]. To more comprehensively answer the question "what can they do", metagenomics can be used. Metagenomics significantly increases both the amount and the complexity of the data. Besides the "who", and the "what can they do", community responses to diets or otherwise environmental changes can be studied by metatranscriptomics to answer the question "what are they doing". Sequencing the full transcriptome of the community provides by proxy insights in which pathways/processes are actually active. The logical progression of technology also leads to metaproteomics, which due to lack of precisely matching reference genomes [68] is still not very widely used and despite interesting results [69, 70] still remains to represent a niche discipline [71]. Meta-metabolomics (also called metabonomics [64], although this term has been used for a different purpose [72]) is currently an even less used technique.

A large body of research applying abovementioned omics approaches is published in well-known journals. Figure 2a provides data up and until 2016. PubMed lists after the initial publications starting in the early 2000s an increasing amount of publications per year, reaching to more than a 1000 per year at the moment (Fig. 2b). The focus of most of these publications is on DNA-based approaches, including 16S rRNA gene sequencing and true metagenomics. This trend is followed distantly by metatranscriptomics, metaproteomics and meta-metabolomics. Since by far the majority of these publications are within the scope of some form of high-throughput nucleotide sequencing (16S rRNA gene, metagenomics, metatranscriptomics), in the following paragraphs, we will focus on these omics approaches.

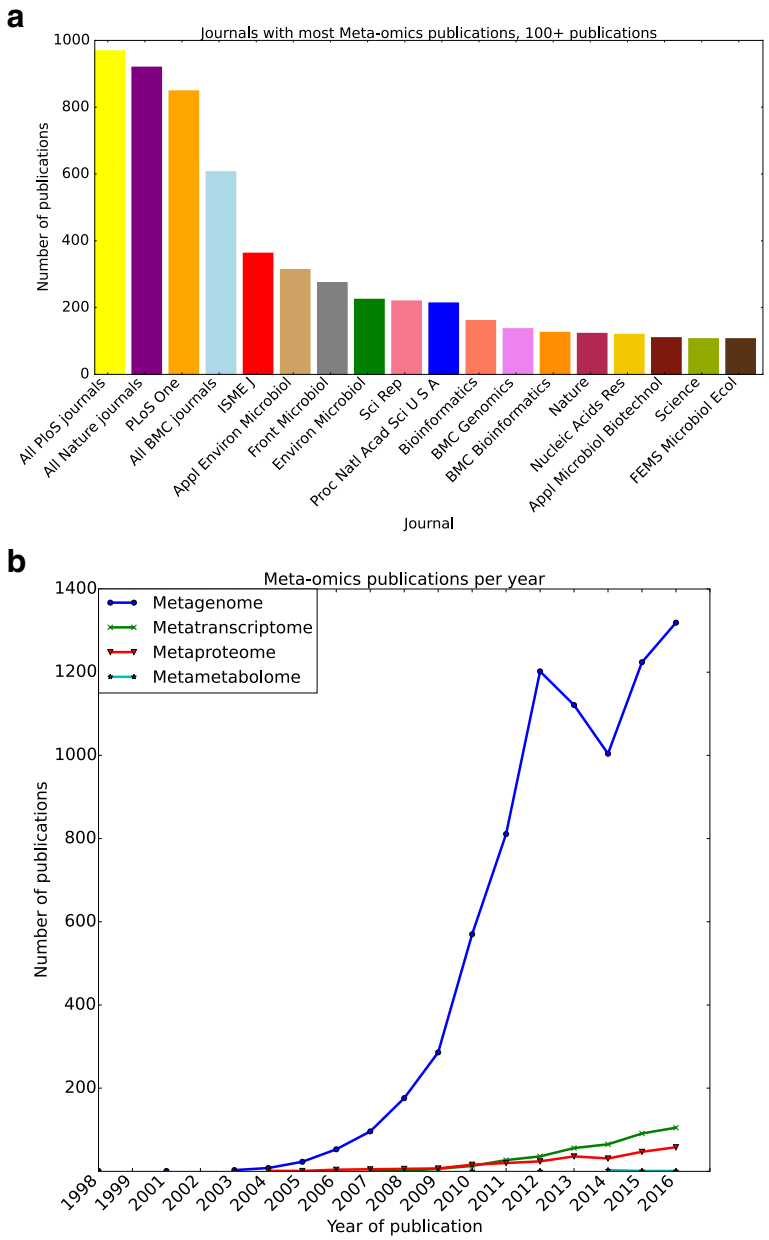

Fig. 2 a Journals with the most gut-related meta-omics publications. **b** Overview of gut-related omics publications per year. 16S rRNA gene sequencing and metagenomics are combined, since these cannot be easily distinguished via title/abstract searches due to the erroneous labelling of amplicon sequencing approaches as metagenomics by many researchers. All data was retrieved via PubMed searches for the corresponding terms. For the exact search terms, please see Additional file 1

Differences within the omics technologies

The methods used for amplicon sequencing, metagenomics and metatranscriptomics are summarized under the term NGS technologies (also called second generation technologies; for a review see [73]), including highly automated technologies represented by Illumina sequencing machines like HiSeq or MiSeq, the Roche 454, Ion Torrent and SOLiD technologies. These technologies are a follow-up of Sanger sequencing, which still has the highest level of accuracy but has a rather low throughput due to limited parallelisation possibilities. NGS technologies allow millions of fragments to be sequenced in a single run. The DNA is randomly sheared, and all resulting fragments are sequenced with fluorescent nucleotides, which emit at incorporation in the new formed DNA strand certain light wavelengths. These can automatically be recorded by current systems and allow high-throughput sequencing information by generating millions

of short reads. One lane on a typical Illumina HiSeq machine can generate up to 360 million reads, currently with lengths up to 350 bases. The limitation in this approach is mainly the used DNA polymerase for the extension of the newly formed DNA fragments, which tends to lose precision with increasing read length, making longer reads more error prone. Especially in metagenomics obtaining longer read lengths is important. Besides providing more information per single read, which is in general desirable in many cases, specifically for metagenomics it will (i) lead to a higher chance of uniquely assigning reads to a single microbial taxon leading to a better resolution in strain and species separation, (ii) make it easier to capture gene functionality and (iii) allow for a higher confidence during the assembly of the data, especially in those cases when the community harbours phylogenetically close species.

The new sequencing technologies (third generation sequencing) from Pacific Biosciences (PacBio) and Oxford Nanopore are ameliorating this problem. Both technologies can produce very long reads, up to 60,000 bases (PacBio) and more (Nanopore). PacBio circumvents the loss of precision of the polymerase by repeatedly sequencing the same DNA fragment [74]. Oxford Nanopore channels single-stranded DNA through a pore which carries an electric current, and measures the change in current as the DNA passes by, with each of the bases causing a different change. This technology does not lose precision with increased length, but generating longer fragments and stably channelling them is the limitation [75]. Current drawbacks of both technologies as compared to the second generation technologies are a higher error rate, requirement of a significantly larger amount of template DNA and higher sequencing costs. PacBio [76–81] and Oxford Nanopore [82] have already been used in microbiota sequencing and their use will most likely increase when the technologies further mature.

Extraction of information from 16S rRNA amplicon sequencing data

The 16S rRNA molecule shows a high degree of structural and sequence conservation in all prokaryotic organisms. Being part of the ribosome, it is a crucial part of the translation machinery. Because the specific secondary structure and function constraints evolutionary drift, it is, albeit with some limitations [83], possible to work with "universal" or species-independent primers and therefore amplicon sequence analysis remains the standard approach to investigate microbial diversity. If two or multiple complete rRNA gene sequences have more than 97% identity, they belong to the same species. The 97% identity threshold is due to historical reasons because this value was found to be in agreement with DNA-DNA hybridization results, but otherwise no coherent species definition exists [84, 85]. In order to make clear that the actual species/genotype is often not known and might actually differ, 97% identity clusters of rRNA sequences are also referred to as "operational taxonomic units" (OTU).

The 16S rRNA gene is approximately 1500 nucleotides in size and for the highest confidence the complete sequence is required. Due to the read length limitations of second generation technologies researchers have therefore investigated, which sequence range of the rRNA showed the highest degree of variability and will therefore result in the best resolution [84, 86]. Using second generation sequencing techniques, these regions (variable regions V1-V9) are therefore preferentially sequenced (for a review see [87]). Here, region-primer combinations need to be carefully matched as these choices can have a high impact on the results [88].

In eukaryotes, like for example fungi, the situation is more complicated. Sequencing 18S rRNA genes does not provide the required resolution, and often internal transcribed spacers (ITS) are sequenced instead [89].

After the amplicon sequencing data has been generated, the next step is to derive corresponding information regarding community composition. In general, since sequencing of single phylogenetic marker genes (fragments) requires less throughput than whole genomes, also the costs per sample are considerably lower, providing the necessary statistical power for a more detailed analysis [90].

Using second generation sequencing techniques, there are multiple considerations involved, e.g. how similar the sequences are expected to be in the variable regions of choice, which reference database to use (SILVA [91], RDP [92] or Greengenes [93]), the significance of base-calling error rates intrinsic to high-throughput sequences data [94] and how erroneous sequences can be detected. Due to these challenges, sophisticated pipelines for taxonomic assignment have been developed, like for example Qiime [95], Mothur [96], Phyloseq [97], MICCA [98] and NG-Tax [99], the latter of which has been developed in our laboratories and provides computationally efficient and accurate taxonomic assignments and quantification of OTUs per sample with improved robustness against choice of region and other technical biases associated with 16S rRNA gene amplicon sequencing studies.

A range of different methods coming from macroecology is used to investigate a habitat's diversity. The species richness or mean species diversity of a sample is often referred to as alpha-diversity and the amount of variation in species composition among the samples (beta-diversity) can also be investigated. A range of different alpha-diversity measures is being used, including those that account for species richness (defined as the absolute count of individual populations per habitat), phylogenetically weighted richness (Faith's Phylogenetic Diversity [100]), and species diversity, including Shannon index [101] and Simpson index [102] (for a

review, see [103]). Diversity indices also try to incorporate the evenness of the species distribution [104] because different conclusions need to be drawn if an ecosystem is dominated by a single species with a plethora of other rare species, or if the distribution is rather even. Another important aspect is under-sampling. To estimate if the true richness of species has been captured, different methods like rarefaction analysis, Chao1 [105] or ACE [106] estimators can be used (for a review, see [107]).

Analyses of beta-diversity make use of a number of different measures of pairwise community similarity, including for example Jaccard index [108], Bray Curtis dissimilarity [109] and UniFrac distance [110], the latter of which is phylogenetically weighted.

In most cases, a first look at the data is done with unconstrained multivariate statistical approaches such as Principle Component and Principle Coordinate Analysis (PC(o)A). These two methods try to fit highly dimensional data (e.g. a high amount of samples and different species in them) into a plot with two (or three) dimensions, trying to display as much of the variation in the data as possible. Factors that are potentially related to the observed variation, including for example environmental conditions, time points or the objective of the research, can be projected a posteriori, and their significance can be tested post hoc.

Several of these statistical tools are standardly embedded in sequence analysis pipelines like Mothur [96], Qiime [95] or Phyloseq [97] and allow to capture measures of alpha- and beta-diversity. Choices can be made between default analysis routines and more customized procedures where users can adjust specific settings.

With these methods, it has been found that for example the alpha-diversity in the microbiota of obese subjects is significantly reduced in contrast to the alpha-diversity in lean subjects [111]. Other successful studies in this field have already revealed that gut microbiota is transmitted vertically and that obese mice have a considerably less diverse microbiota than their lean counterparts [112]. Furthermore, it has been shown that the gut microbiota changes during human development starting at birth and is different depending on geographic location [113], during long-term dietary interventions [114] or when consuming specific diets even during a single day [115].

Extraction of functional information from metagenome data

In principle, full genomic information can be captured with metagenomics. Seminal projects in this area like MetaHit [11] and the human microbiome project [116] made great efforts to sequence the metagenomes of diverse cohorts with many subjects to investigate the full functional capacity of the different microbiomes. The amount of data required makes deeper sequencing necessary, which complicates the workflow to extract information from metagenomics data (Fig. 3).

High-throughput sequencing data is noisy, and quality control is a critical first step (review see [117]). One crucial step for which settings have not yet been universally agreed upon is the quality trimming [118], and no consensus advice can be given.

For simple read mapping there are a number of strategies that can be applied. BLAST [119] or Diamond [120] can be used to match reads directly to KEGG, to quantify the functions based on the number of matching reads (e.g. applied in [38]). A higher resolution is obtained when reads are mapped to a set of reference genomes [111, 121], which also allows for a taxonomic classification of observed functions [122]. If the phylogenetic distance between the reference set and the sample is small this has the advantage of speeding up the analysis. Furthermore, associated functional annotations can be directly utilized, making a separate annotation step unnecessary. A major drawback for this type of workflow is that only known species can be analysed, whereas new strains with novel functions, horizontal gene transfer and other evolutionary events will not be captured, and micro-diversity will be lost.

An alternative approach therefore is to assemble reads into larger contigs and extract genomes directly from metagenome data [123] (Fig. 3). Today obtaining a high quality single genome can still be a challenge [124], and with a community genome assembly approach these challenges can multiply. Examples are chimeric assemblies between genomes due to presence of multiple strains of the same species (although miss-assemblies should not occur very often [125]), and a low coverage of low abundant species. At this point, it is also important to consider the mapping rate after the assembly. While we expect for a single organism that after the genome assembly most of the reads will map to the assembly, this can deviate for metagenomics. This is mainly due to the species richness and species evenness of the community under investigation. A complex species-rich sample of high evenness (i.e. similar abundance of many community members) will require more data to assemble the top-ranking species than a sample where a few high-ranking species have much higher abundances. Therefore species richness and species evenness need to be taken into account to evaluate if the mapping rate is appropriate for further analysis.

Some of these challenges have been tackled with specific metagenome assemblers like MetaVelvet [126], which take different properties of the sequencing data into account like for example the different abundances of the potentially present species. Currently, a community-derived assembly will also not lead to closed genomes. The next

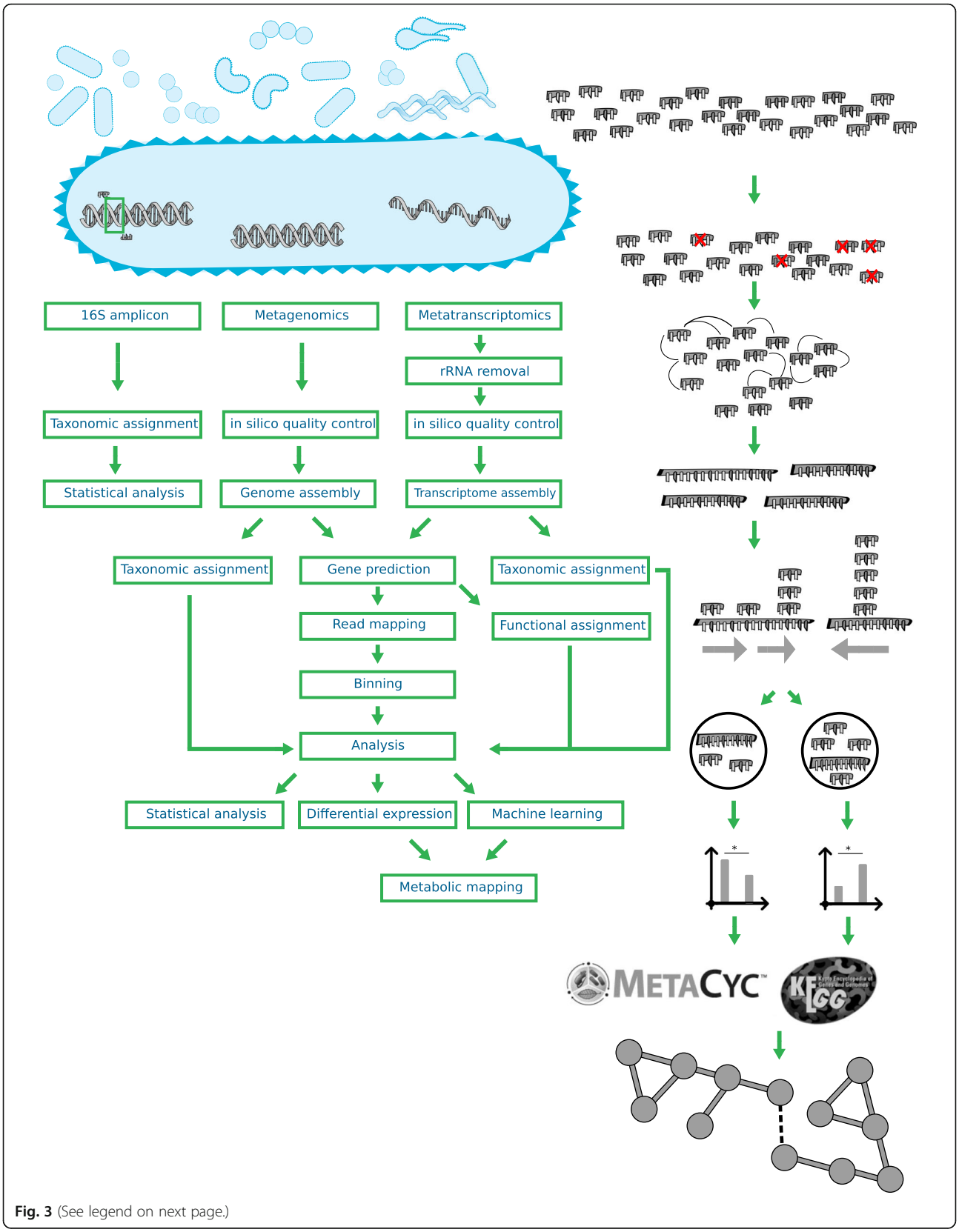

Fig. 3 (See legend on next page.)

(See figure on previous page.)
Fig. 3 Overview of the different steps in the meta-omics analysis workflow. The different workflows are depicted, from left to right for 16s amplicon data, metagenomics data and metatranscriptomic data. The main steps for 16s amplicon data is the definition of OTUs together with taxonomic assignment, followed by statistical analysis. For metagenome data, first steps involve quality control steps, followed by a metagenome assembly. The workflow splits afterwards into two directions, one being the taxonomic assignment, the other one the definition of metagenomic bins and the functional annotation. Genes can be predicted from the genome assembly, which can be functionally profiled. With the coverage information of the genes, it is also possible to define genome bins. After this step is done, the same statistics as for 16s amplicon data can be performed, as well as differential expression/abundance analysis together with pattern detection through machine learning, and finally analysis of the metabolism. The workflow for metatranscriptomic data is in general the same, except that rRNA, which does not provide any information in this setting, needs to be removed before most of the steps, and that no binning is possible with transcriptome data

challenge is therefore to determine which of the assembled contigs/scaffolds belong to a single species. This process has been termed binning, and several tools such as MaxBin [127] or MetaCluster [128] have been developed to determine the amount of bins required and to assign contigs to bins. To do so, these tools take different types of information into account, such as k-mers frequency in the data or contig read coverage. The quality control of this step is critical, since this process is also error prone, especially when phylogenetically close organisms of similar abundance occur in a community.

The most widely used method to test for correctness of binning is based on single copy marker genes, like in for example CheckM [129]. Based on the presence of these necessary genes, both the coverage of a genome in a bin as well as the amount of contamination from other genomes can be determined. A problem with this approach is that it is limited to contigs/scaffolds containing these core functions.

Next the taxonomic origin of the various bins can be determined (Fig. 3). All programs and workflows which can perform this are reference based, but work with different mechanisms. One approach is to use BLAST [119] to compare all the metagenomic contigs against a database, like the NCBI NT database, or specialized databases like for example the human microbiome project [11]. The accuracy of the taxonomic assignments is proportional to the similarity score of the alignments. One of the first programs to deal with this problem is MEGAN [130], which also gives the user a graphical interface for direct analysis. The biggest drawbacks of this method are that (i) it can be computationally prohibitive to use a large database and (ii) that closely related species cannot be differentiated from each other. A computationally more efficient alignment free method for the taxonomy determination is to compare the k-mer profiles of the metagenomics contigs with k-mer profiles obtained from a reference database. This has been implemented in tools like Kraken [131] or PhyloPythia [132] (for a review of programs, see [133]),

To understand the underlying causes of a community change and potential effect, functional profiling needs to be performed (Fig. 3). This part of the analysis is for a metagenome mainly different to a single genome in regards to the quantity, but the basic processes are the same. First gene prediction needs to be performed with gene callers like example prodigal [134], which have special settings for this kind of data. A low-level profiling can be obtained with a COG analysis [135]. The COG ontology consists of limited number of broad categories, which allow the detection of extensive changes. When more data is available a higher resolution can be obtained. These can be for example (i) EC number prediction, which can be obtained via PRIAM [136] and can be linked to metabolic pathways using databases like KEGG or Metacyc [137], (ii) lists of carbohydrate active enzymes [138] can be obtained via dbCAN [139] and (iii) full domain profiles including GO terms [140] via for example InterproScan [141] or via second generation annotation tools [142]. With these so called full functional profiles, it is possible to reconstruct the metabolism of the bin [143–145], and bin-specific auxotrophies or special metabolic capabilities can be investigated. If someone wants to draw statistical conclusions for the difference in the metabolism by for example investigating for overrepresented functions (e.g. GO enrichment [146]), it should not be forgotten that, even for genomic information, replication is necessary [147]. If it is not possible to obtain all this data, due to lacking computational resources, also web services like IMG/M [148] or EBI metagenomics [149] can be used, which normally also have a user friendly interface, but only offer a limited depth of analysis.

Extraction of functional information from metatranscriptome data

The transcriptome approach will allow the investigator to focus on functions that are actually expressed in a given sample. A highly abundant species may show a low expression of functions of interest and vice versa (e.g. [150]). In fact, since DNA is also highly stable, the metagenomics approach might also take non-viable cell populations into account, which could falsify the conclusions, but also separate measures, like removal of non-viable cells, can be taken to prevent this [151]. Thus, the metatranscriptome provides a more accurate account of actual functionality.

Most relevant steps, including QC, are the same as for single organism transcriptomics (for a review, see [117]; workflow, see Fig. 3). Not mentioned in [117], but necessary for metatranscriptome data is the in silico removal of spurious rRNA reads [152] as in vitro removal of rRNA prior to sequencing will most likely not remove all of it.

Like in metagenomics mRNA reads can either be mapped or de novo assembled. Mapping can be done if a set of reference genomes is available. If binning has been performed before, then the transcriptome should not be mapped to the different bins separately. If bins were separated before mapping, then the assignment of reads would be skewed if phylogenetically related bins are present (incorrect multiple assignment of reads). If no reference metagenome is available, it can be attempted to map the RNAseq data to related datasets. In this case again, the absolute mapping rate of the data needs to be cautiously taken into account, because an unsuitable reference (due to large phylogenetic distance or missing species) will exhibit low mapping rate and will prevent a full analysis of the data. Alternatively, a de novo transcriptome assembly can be performed. Specific metatranscriptome assemblers have been developed to deal with the complexity of such data (for a review, see [153]). Subsequent mapping of the same mRNA reads onto the de novo assembly allows for differential expression analysis, which can be performed with known tools like for example edgeR [154] or DESeq2 [155].

In many regards, metatranscriptome analysis can function as a substitute for a metagenomics analysis while adding an additional layer of information. For instance, metatranscriptome analysis has already revealed that activity of carbohydrate degrading enzymes can be underestimated if only genomic information is considered, or how the activity of the gut microbiome responds to different diets [156, 157]. In principle, similar conclusions could also be obtained from a combined metagenomics/metaproteomics approach [158] albeit at lower resolution.

A pure transcriptome assembly has the drawback that binning is not possible, since many of the binning approaches rely on the fact that in a metagenome all contigs from one species will exhibit similar coverage, which is not the case for a transcriptome. It will also not be possible to assemble very long contigs, because many intergenic regions will not be transcribed. Important changes at the ecosystem level can be assessed by analysing the expression levels of the microbiota in the community provided that species abundances are also taken into account; a 50% increase in abundance might appear as a 50% higher gene expression, but in this case does not reflect a transcriptional response on a per-microbe basis, but rather a compositional response at community level.

From information to understanding

As exemplified above many computational tools and pipelines exist that are able to extract biological information from high-throughput data. Understanding the unique chemical and functional capabilities of the human microbiome and deciphering the biological roles of individual species is much more difficult. Linking microbial activities with gene expression and enzyme functionalities is just the first step. In early years of genomic research, "hairball" graphs had their appearance in many publications, showing connectivity within the available pile of data, rather than focusing on the biologically informative parts. With the increasing number of samples being analysed for example from patients, from replicates, from different conditions, different types of sequencing data combined with different types of computationally derived data such as EC number and domain predictions, which methods can be used to gain useful information?

The most obvious approach, especially with pure abundance data, is looking for correlations (also possible via regression [159]). It can be assumed that correlating species/OTUs have a symbiotic relationship with each other and/or with a third OTU, whereas anti-correlation can (but does not have to) indicate antagonistic behaviour. There are, however, several pitfalls. For example, OTUs, which are present only in very few samples, will be highly correlated due to the common absence in multiple samples. While this general conclusion can be true, it needs to be considered that absence in sequencing data does not have to mean absence of the organism. It can also indicate abundance below the detection threshold, or simply a failure in detecting the organism with the current pipelines.

The same methods described above for the analysis of 16S rRNA gene amplicon sequence data can also be utilized for metagenomics data. Multivariate visualization tools such as PCA can be used to see if specific sample groups, e.g. defined by specific interventions or states of health, cluster together, or if other factors are more prevalent in explaining the observed variation in the data. Nevertheless, for the in-depth analysis, more sophisticated methods should be used such as for example pattern recognition, which enables the researcher to find useful information in big data. This field is broadly classified into two approaches, i.e. supervised and unsupervised learning. In supervised learning, the researcher tries to classify unknown samples into categories for which already known samples exist. If, for example, samples from lean and obese subjects have been obtained, an algorithm can be trained to determine if samples of

unknown origin were obtained from a lean or obese person. While supervised learning has been already used in microbiome research with great success, e.g. [160, 161] (for reviews of the methodologies, see [162, 163]), and is currently researched for the application in many different fields and termed "life changing" for the general public (e.g. deep learning [164]), this approach is often hampered by the fact that samples from different studies are not comparable due to different methodological approaches with respect to for example DNA extraction or sequencing method and depth.

Unsupervised learning, also called clustering, does not rely on prior information. Clustering algorithms, including hierarchical clustering, k-means and dbscan, try to find unknown patterns in the given data, e.g. different patterns of gene expression over multiple conditions. This approach has also been used for example to determine the enterotypes [165] but also suffers from a wide array of challenges. The choice of clustering algorithm is not trivial and depends on the structure of the data, which can often not be determined in an easy way [166]. Furthermore these algorithms often rely on user-defined parameters such as the amount of clusters to find. Determining the best parameter set is its own research field, given that more than 30 different algorithms for this purpose exist [167], and not all are applicable to all clustering algorithms [166]. If at the end, wrong parameters are chosen; it might lead to erroneous conclusions, like for example if not the optimal amount of clusters (in this case, enterotypes [168]) is selected. Otherwise, a cluster might be split into multiple, or multiple distinct clusters might be treated as one.

Having said that many of these algorithms have been implemented in different programs like ELKI [169] or WEKA [170] and can also be utilized by inexperienced users, although the final evaluation still often requires expert knowledge.

If useful patterns have been obtained after the machine learning, the last level is the biological understanding and interpretation. Simple approaches include just mapping extracted functional information such as EC numbers and KO numbers to pathway databases like KEGG [171]. More sophisticated solutions try to automatically extract the useful information from these networks, e.g. MetaModules [172]. If also other non-metabolic functions should be investigated, then a broader type of classification can be used. The most common analysis is the GO enrichment analysis, which aims to identify overrepresented functions in the dataset [146].

It also needs to be considered that the microbiome data does not have to stand on its own. If clinical or nutritional data is available, these can be used as well. Correlating such metadata with microbiome data has shown that factors like age or stool consistency are highly related to microbiome composition [173], as well as the hosts genetics [174]. Furthermore, it is also possible to revert this and use microbiome data together with clinical data to predict a persons' glycemic response to food intake [175].

Since this type of data can be highly connected, visualization of this connectivity might be necessary for a better understanding. While some visualization forms are standard, for example depicting the distribution of species/OTUs per sample in a bar chart, and metabolic networks as networks, sometimes more sophisticated methods are necessary. For analysis purposes, the Krona library [176] can be a useful visualization tool to explore quantitative hierarchical relationships between taxonomical groups. In many cases, there are no standard recipes for the analysis workflow, and custom solutions have to be developed. For these cases it is necessary to consider what type of data should be shown, and with which method they are obtained. Several visualization methods are available [177, 178], but standard packages for many of these are not necessarily developed yet or easily accessible.

Bottom up: mechanistic insights into the microbiome

The next step after collecting data and investigating the communities is building models and testing hypotheses. While with single species this is very well doable, microbial communities pose more challenges to the researcher. For a single culturable species, it will be possible to collect the necessary data. It is possible to reconstruct the full metabolism (according to current knowledge), manually curate it, and measure a vast array of metabolites. In contrast, all these factors pose challenges in a community like the intestinal microbiota.

The sum is more than its parts

A community is more than an accumulation of multiple single organisms. The different microbes interact within a dynamic environment; they will behave differently, depending on who is in the surrounding, and what they are doing. Even for a single species, species abundance can lead to emergent properties for example via quorum sensing, which can alter the behaviour of individual cells and the entire population dramatically [179]. In biofilms, the formation itself is an emergent property, which would not be possible to observe if only single cells are considered. It also leads to the change in behaviour of the different cells, as some will get advantages in this environment (protection), whereas the cells on the surface are less protected, but also have more access to nutrients. Other forms of symbiotic relationships can also lead to emergent properties where for example some species in the community provide the means to overcome amino acid

auxotrophies or vitamin deficiencies of others or of the host [180–182]. Another unrelated example from the oceanic microbiome is the detoxification the environment [183]. This case is commensalistic, since a big part of the microbial community benefits from the ability of one member to detoxify oxygen radicals, giving the other members a benefit, which lead in this case to genome streamlining by loss of genes related to oxidative stress. The authors even expanded their observation into the "Black Queen Hypothesis", stating that this streamlining together with a dependency on helper organisms with leaky beneficial functions might be an universal concept. This is only possible to observe at the community level, and the investigation of a single species would not lead to such conclusions.

Numerous additional examples exist, also in the gut environment (for a more complete review, see [184]).

How to predict the sum from its parts

How should the behaviour of such a community be predicted? The apparent approach is to model the metabolism of the whole community as a single entity or "supra-organism", neglecting species boundaries [185]. While this can give an idea about the metabolic capabilities, it is an oversimplification and will miss critical steps like metabolite exchanges and interdependencies between organisms. The extension of this approach would be to model single organisms, and connect these models to one community model.

Producing a good model of a single organism is the first step in this process. There exist high-throughput methods, like ModelSEED [186], Pathway Tools [187] or KBase [188], which can automatically construct a genome scale metabolic model (GSMM) from the given genomic information. Although such reconstructions can be of high quality, it is still likely that the model will contain errors or gaps, which need to be solved by manual curation [189].

If different models for the relevant organisms can be obtained, the next challenge is combining them. If the models are based on different databases/coming from different sources, then this could result in incongruences in the final model. While this should in general be avoided, it is sometimes necessary, because high quality models of different organisms exist (e.g. *Homo sapiens* [190], *Escherichia coli* [191]), and it is not feasible to integrate this work into the high-throughput frameworks. For such cases, an integration of different model sources needs to be performed. The challenge is to match all the metabolites that need to be shared between all relevant models. Due to different problems, like the lack of unique identifiers, matching these names is not a trivial task, can be very error prone and requires the application of specialized tools (e.g. [192]).

Different hypotheses can be tested after a multi-organism model has been finally generated (e.g. [193, 194]). One of the first approaches should be to investigate ecological compatibility. This can be done for example via reverse ecology [195], by matching the metabolites in the different organisms to each other to see possible interconnections and metabolic dependencies. More advanced challenges are to actually simulate this metabolism. Finding the target, the objective function of a model, will depend on the underlying biology. Maximization of biomass is often used in single-organism models [196] (among others) and has also been used in multi-organism models (e.g. [193, 197]). This is not applicable in all cases because for example competition or parasitic relationships can exist in an ecosystem and often the objective is not to maximize the biomass of the competitors in the surrounding. Therefore, more sophisticated methods like D-OptCom [198] have been developed, which break the community optimization problem into multiple single problems. These consist of smaller optimization problems for each community member, and the main problem is to optimize the community. Others have extended this to even include spatial structures [199]. This allows the simulation of each bacterium's growth independently, giving a more realistic result than simulating community growth.

Metabolic models are not the only models which can be employed, metabolism is also not the only type of process which can be simulated, and the bacterial level is not the only scale which can be considered. Different kinds of kinetic models of the metabolism have been developed, some especially for the gut [200, 201], and also for related ecosystems [202], but this field is still in its infancy. The mentioned models also simulate metabolism, predicting the flow of carbohydrates into acids or extracellular polysaccharides, including different non-metabolic parameters like peristaltic movement of the gut. Also non-metabolic models exist, with the focus on antibiotic resistance in the gut [203] or the succession of organisms in the gut [204]. As it can be seen, the field is still far away from a comprehensive virtual gut model. In fact, already the whole cell model [205] is extremely complex, and contains for example different scales which might be lacking full integration into the model. With all the different factors to consider, integrating more data into the models with proper feedback systems, until up to the ecosystem level, will probably be a research objective for many years to come [206].

How to change the sum, and its parts

Modelling cannot be only done in silico. With synthetic biology, artificial model systems of the gut environment have been created [207]. These models vary in their complexity and capabilities to simulate the environment. It is important to differentiate which part of the gut is

modelled, if there need to be multiple compartments, and if for example each of them needs to be pH controlled. These systems were shown to simulate parts of the gut appropriately [208], and [209] showed the contributions of intestinal movement to the development of inflammation in the gut.

But since these systems do not (yet) perfectly model the gut, final proof has often to be provided from animal models. Gnotobiotic animals [210] offer the possibility for controlled interventions. In contrast to the in vitro systems, the in vivo system will be able to incorporate all the necessary factors to evaluate gut functioning. Inoculation of the sterile animals with a defined microbiota ("synthetic ecology") allows studying the niches of specific bacteria [211, 212], the development of the microbiota over time [204], during development [213] and the interactions between different bacteria [58, 60, 214]. Gnotobiotic animal models have also been used, as mentioned earlier, to show that the microbiota does not only change with obesity, but that it also contributes to it [33, 37, 38, 215].

At the end, it still needs to be taken into account that animal models do not represent humans, and ways to influence our gut microbiota in a rational way are only partially understood. One of these rational methods is the gastric bypass. It is one of the last resorts for morbidly obese patients to lose weight, will have a significant effect on a subject carbohydrate consumption and will alter the gut microbiota in different ways [216–219] (mainly an increase in Gammaproteobacteria), due to different changing factors like for example the distribution of bile acids. This is the most drastic method for a targeted microbiota change besides antibiotics and faecal transplantation. The latter has been used to treat severe diseases like *Clostridium difficile* infection (e.g. [220, 221]) or Ulcerative Colitis [222]. Faecal transplantation replaces a patient's gut microbiome with that of healthy donors, however, mechanisms underlying success or failure of the treatment have not yet been fully understood in all cases. The main factors do not only include the gut microbiota itself or the host genetics [174], but potentially also other factors like excreted metabolites [223, 224]. Due to the difficulties of understanding the mechanisms, it has not yet been possible to rationally design a medicine from this therapy, which would simplify the production and legal issues [225, 226], but progress is likely to be made within the coming years [184, 227].

Microbiome changes do not only have clinical impact. Pre-clinical applications are also possible. Nutritional methods can be rationally employed, without having dramatic impact on the everyday life and include mainly pre- and probiotics. The substances and microorganisms consumed are not new, and have been already consumed for millennia, e.g. as fermented milk products. But also their mode of action is not fully understood, and in some cases their usefulness is even debated [228]. Probiotics like *Lactobacillus* and *Bifidobacterium* (e.g. [229, 230]) might act in different ways. Tested hypotheses are that they might change the gut environment to make it inhospitable for pathogens [231, 232], produce antimicrobial compounds like SCFAs [233–235], alter the composition by releasing compounds from otherwise indigestible substrates (e.g. prebiotics) [229, 236] or reverse/prevent dietary effects [237, 238]. But even in such controlled setups it is too simple to attribute changes to single organisms, since the breakdown of prebiotics (leading to "postbiotics", which might be the actual bioactive compound) can involve multiple organisms (see for example the summary about quercetin in [239]).

Conclusions

The currently available body of research has shown that it is important to take the ecosystem as a whole into account to understand its health implications. Recently, this trend is increasingly being picked up. After the first human genomes were sequenced, it was believed that it would change how medicine works. It was thought that every aspect of a human would be understood and that all treatments would be personalized [240, 241]. Although personal genome sequencing is still on the rise [242], this prediction has not turned out to be fully true [243], although it should be noted that there have also been significant successes (see for example table 1 in [244]). While we for sure do not yet fully understand the human genome [245], we need to be aware now that it is not the only factor. The personal well-being is not only influenced by our genetic traits. Our complete ecosystem, the whole holobiont, needs to be taken into account. It is already clear that we cannot understand obesity if we do not understand our microbiome, and if we do not understand its connections to the host. With discoveries like the enterotypes [165] (caution for the results [168], as they have been discussed widely, with the notion that gradients are more likely than separate clusters), the next step after the personal genome might even be the personalized metagenome (and the first companies are even trying to market it). If people have different microbiomes, they might need to be treated differently to combat for example obesity. With enough data, and the understanding of its meaning, it might also be possible to prevent this lifestyle epidemic, in combination with personalized nutrition, as it is even already becoming potentially feasible [175]. We might also be able to go further, and even prevent diseases. The preventive measures are normally not part of the regular mainstream medicine, but ideas exist how incorporate preventive measures, pioneered as "4P medicine" (predictive, preventive, personalized, participatory) [246, 247]. If we know a person's microbiome, we will

be able to predict if they are for example more prone to obesity or other risk factors (which is for some disease states already possible [160, 161]). If we understand the functionality, we will be able to take countermeasures with dietary interventions like pre- and probiotics. Since all these ecosystems are different, this approach will need to be personalized. Not only to take the personal genome and the personal microbiome into account, but also the compatibility with lifestyle, because even the best treatment might not suffice if a subject consumes by default a high fat "western diet" without any exercise. And this is all not possible, if the population does not participate. This approach will rely on everyone's personal data, which needs to be acquired. And it will only work, if the results are communicated clearly.

All of these points are future challenges. We do not yet fully understand the microbiome. With diet we are taking counter measures, but not always in rational ways. Medicine is already personalized, but not all treatments have the necessary data to be personalized. And while communication can already work (e.g. the whole "quantified self" movement is relying on achievements being communicated back), it is not always the case, and wrong communication, resulting in wrong expectations, will even discourage the users (e.g. [248]). The researchers in the microbiome field need to be aware that this hype can also happen to the microbiome [249, 250].

Current microbiome research aims to overcome some of these challenges. Obesity research is likely to contribute in the close future to a better understanding of the underlying mechanisms, and the 4P medicine might partially become achievable in not too distant future, leading to better health and combating epidemics like obesity.

Additional file

Additional file 1: Supplementary Materials and Methods. Description on how data for Figs. 1 and 2 were obtained. (DOCX 16 kb)

Acknowledgements
The authors want to thank Ruben van Heck, Maria Suarez-Diez (Wageningen University and Research, Laboratory of Systems and Synthetic Biology), Joan Edwards and Gerben Hermes (Wageningen University and Research, Laboratory of Microbiology) for helpful discussions.

Funding
B. Hornung is supported by Wageningen University and the Wageningen Institute for Environment and Climate Research (WIMEK) through the IP/OP program Systems Biology (project KB-17-003.02-023).

Availability of data and materials
Not applicable

Authors' contributions
BH and PJS drafted the structure of the manuscript. BH and PJS wrote the manuscript with input from the other authors. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Laboratory of Systems and Synthetic Biology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, the Netherlands. [2]Laboratory of Microbiology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, the Netherlands.

Received: 22 August 2017 Accepted: 13 February 2018
Published online: 06 March 2018

References
1. Hadhazy A. Think twice: how the gut's "second brain" influences mood and well-being. In: Scientific American. New York: Nature America, Inc.; 2010.
2. Brown H. The other brain also deals with many woes. In: The New York times. New York: The New York Times Company; 2005.
3. Berasategui A, Shukla S, Salem H, Kaltenpoth M. Potential applications of insect symbionts in biotechnology. Appl Microbiol Biotechnol. 2016;100(4):1567–77.
4. O'Hara AM, Shanahan F. The gut flora as a forgotten organ. EMBO Rep. 2006;7(7):688–93.
5. Marteau P, Pochart P, Dore J, Bera-Maillet C, Bernalier A, et al. Comparative study of bacterial groups within the human cecal and fecal microbiota. Appl Environ Microbiol. 2001;67(10):4939–42.
6. Zoetendal EG, Raes J, van den Bogert B, Arumugam M, Booijink CC, et al. The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. ISME J. 2012;6(7):1415–26.
7. Flint HJ, Scott KP, Duncan SH, Louis P, Forano E. Microbial degradation of complex carbohydrates in the gut. Gut Microbes. 2012;3(4):289–306.
8. Gerritsen J, Smidt H, Rijkers GT, de Vos WM. Intestinal microbiota in human health and disease: the impact of probiotics. Genes Nutr. 2011;6(3):209–40.
9. Fofanova TY, Petrosino JF, Kellermayer R. Microbiome-epigenome interactions and the environmental origins of inflammatory bowel diseases. J Pediatr Gastroenterol Nutr. 2016;62(2):208–19.
10. Lopez J, Grinspan A. Fecal microbiota transplantation for inflammatory bowel disease. Gastroenterol Heaptology. 2016;12(6):374–9.
11. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010;464(7285):59–65.
12. Sender R, Fuchs S, Milo R. Revised estimates for the number of human and bacteria cells in the body. PLoS Biol. 2016;14(8):e1002533.
13. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, et al. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A. 2011;108(Suppl 1):4680–7.
14. van Passel MW, Kant R, Zoetendal EG, Plugge CM, Derrien M, et al. The genome of Akkermansia muciniphila, a dedicated intestinal mucin degrader, and its use in exploring intestinal metagenomes. PLoS One. 2011;6(3):e16876.
15. den Besten G, van Eunen K, Groen AK, Venema K, Reijngoud DJ, et al. The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. J Lipid Res. 2013;54(9):2325–40.
16. WEW R. Role of anaerobic bacteria in the metabolic welfare of the colonic mucosa in man. Gut Microbes. 1980;21:793–8.
17. den Besten G, Lange K, Havinga R, van Dijk TH, Gerding A, et al. Gut-derived short-chain fatty acids are vividly assimilated into host carbohydrates and lipids. Am J Physiol Gastrointest Liver Physiol. 2013;305(12):G900–10.
18. Turroni F, Milani C, Duranti S, Mancabelli L, Mangifesta M, et al. Deciphering bifidobacterial-mediated metabolic interactions and their impact on gut microbiota by a multi-omics approach. ISME J. 2016;10(7):1656–68.

19. Riviere A, Gagnon M, Weckx S, Roy D, De Vuyst L. Mutual cross-feeding interactions between Bifidobacterium longum subsp. longum NCC2705 and Eubacterium rectale ATCC 33656 explain the bifidogenic and butyrogenic effects of arabinoxylan oligosaccharides. Appl Environ Microbiol. 2015;81(22): 7767–81.
20. He X, McLean JS, Edlund A, Yooseph S, Hall AP, et al. Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. Proc Natl Acad Sci U S A. 2015;112(1):244–9.
21. Dwidar M, Monnappa AK, Mitchell RJ. The dual probiotic and antibiotic nature of Bdellovibrio bacteriovorus. BMB Rep. 2012;45(2):71–8.
22. Atterbury RJ, Hobley L, Till R, Lambert C, Capeness MJ, et al. Effects of orally administered Bdellovibrio bacteriovorus on the well-being and Salmonella colonization of young chicks. Appl Environ Microbiol. 2011;77(16):5794–803.
23. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. Nature. 2012;489(7415): 220–30.
24. Lessa FC, Mu Y, Bamberg WM, Beldavs ZG, Dumyati GK, et al. Burden of Clostridium difficile infection in the United States. N Engl J Med. 2015; 372(9):825–34.
25. Stein RR, Bucci V, Toussaint NC, Buffie CG, Ratsch G, et al. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. PLoS Comput Biol. 2013;9(12):e1003388.
26. Voth DE, Ballard JD. Clostridium difficile toxins: mechanism of action and role in disease. Clin Microbiol Rev. 2005;18(2):247–63.
27. Grundy SM. A constellation of complicatons: the metabolic syndrome. Clin Cornerstone. 2005;7(2/3):36–45.
28. O'Neill S, O'Driscoll L. Metabolic syndrome: a closer look at the growing epidemic and its associated pathologies. Obes Rev. 2015;16(1):1–12.
29. Xia Q, Grant SF. The genetics of human obesity. Ann N Y Acad Sci. 2013; 1281:178–190
30. Swinburn BA, Caterson I, Seidell JC, James WP. Diet, nutrition and the prevention of excess weight gain and obesity. Public Health Nutr. 2007; 7(1a):123–46.
31. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, et al. Richness of human gut microbiome correlates with metabolic markers. Nature. 2013; 500(7464):541–6.
32. McNeil NI. The contribution of the large intestine to energy supplies in man. Am J Clin Nutr. 1984;39:338–42.
33. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, et al. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006;444(7122):1027–31.
34. Bergman EN. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. Physiol Rev. 1990;70(2):567–90.
35. Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, et al. A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. Science. 2003;299(5615):2074–6
36. Lammerts van Bueren A, Saraf A, Martens EC, Dijkhuizen L. Differential metabolism of exopolysaccharides from probiotic lactobacilli by the human gut symbiont Bacteroides thetaiotaomicron. Appl Environ Microbiol. 2015; 81(12):3973–83
37. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, et al. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Genet Diet. 2009;1(6):6ra14
38. Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, et al. Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science. 2013; 341(6150):1241214-1 - 1241214-10
39. Minihane AM, Vinoy S, Russell WR, Baka A, Roche HM, et al. Low-grade inflammation, diet composition and health: current research evidence and its translation. Br J Nutr. 2015;114(7):999–1012
40. Chassaing B, Gewirtz AT. Has provoking microbiota aggression driven the obesity epidemic? BioEssays. 2016;38(2):122–8
41. Chassaing B, Gewirtz AT. Gut microbiota, low-grade inflammation, and metabolic syndrome. Toxicol Pathol. 2014;42(1):49–53
42. MacIver NJ, Jacobs SR, Wieman HL, Wofford JA, Coloff JL, et al. Glucose metabolism in lymphocytes is a regulated process with significant effects on immune cell function and survivial. J Leukoc Biol. 2008;84(4):949–57
43. Grundy SM, Brewer HB Jr, Cleeman JI, Smith SC Jr, Lenfant C, et al. Definition of metabolic syndrome: report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. Circulation. 2004;109(3):e13–8
44. Al Khodor S, Reichert B, Shatat IF. The microbiome and blood pressure: can microbes regulate our blood pressure? Front Pediatr. 2017;5(138)
45. Durgan DJ, Ganesh BP, Cope JL, Ajami NJ, Phillips SC, et al. Role of the gut microbiome in obstructive sleep apnea-induced hypertension. Hypertension. 2016;67(2):469–74
46. Schiffrin EL. Immune mechanisms in hypertension and vascular injury. Clin Sci (Lond). 2014;126(4):267–74
47. Pluznick JL, Protzko RJ, Gevorgyan H, Peterlin Z, Sipos A, et al. Olfactory receptor responding to gut microbiota-derived signals plays a role in renin secretion and blood pressure regulation. Proc Natl Acad Sci U S A. 2013; 110(11):4410–5
48. Ghazalpour A, Cespedes I, Bennett BJ, Allayee H. Expanding role of gut microbiota in lipid metabolism. Curr Opin Lipidol. 2016;27(2):141–7
49. Joyce SA, MacSharry J, Casey PG, Kinsella M, Murphy EF, et al. Regulation of host weight gain and lipid metabolism by bacterial bile acid modification in the gut. Proc Natl Acad Sci U S A. 2014;111(20):7421–6
50. Fu J, Bonder MJ, Cenit MC, Tigchelaar EF, Maatman A, et al. The gut microbiome contributes to a substantial proportion of the variation in blood lipids. Circ Res. 2015;117(9):817–24
51. Raza GS, Putaala H, Hibberd AA, Alhoniemi E, Tiihonen K, et al. Polydextrose changes the gut microbiome and attenuates fasting triglyceride and cholesterol levels in western diet fed mice. Sci Rep. 2017;7(1)
52. Staley JT, Konopka A. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annu Rev Microbiol. 1985;39:321–46
53. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, et al. Insights into the phylogeny and coding potential of microbial dark matter. Nature. 2013;499(7459):431–7
54. Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect. 2012;18(12):1185–93
55. Moissl-Eichinger C, Huber H. Archaeal symbionts and parasites. Curr Opin Microbiol. 2011;14(3):364–70
56. Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature. 2010; 466(7304):334–338
57. Parfrey LW, Walters WA, Knight R. Microbial eukaryotes in the human microbiome: ecology, evolution, and future directions. Front Microbiol. 2011;2
58. Samuel BS, Gordon JI. A humanized gnotobiotic mouse model of host-archaeal-bacterial mutualism. Proc Natl Acad Sci U S A. 2006;103(26):10011–6
59. Turroni F, Ozcan E, Milani C, Mancabelli L, Viappiani A, et al. Glycan cross-feeding activities between bifidobacteria under in vitro conditions. Front Microbiol. 2015;6
60. Mahowald MA, Rey FE, Seedorf H, Turnbaugh PJ, Fulton RS, et al. Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. Proc Natl Acad Sci U S A. 2009;106(14):5859–64
61. de Vos WM. Microbial biofilms and the human intestinal microbiome. NPJ Biofilms Microbiomes. 2015;1
62. Elias S, Banin E. Multi-species biofilms: living with friendly neighbors. FEMS Microbiol Rev. 2012;36(5):990–1004
63. Parsek MR, Greenberg EP. Sociomicrobiology: the connections between quorum sensing and biofilms. Trends Microbiol. 2005;13(1):27–33
64. Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. Microbiome. 2015;3
65. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol. 2013;31(9):814–21
66. Asshauer KP, Wemheuer B, Daniel R, Meinicke P. Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. Bioinformatics. 2015; 31(17):2882–2884
67. Iwai S, Weinmaier T, Schmidt BL, Albertson DG, Poloso NJ, et al. Piphillin: improved prediction of metagenomic content by direct inference from human microbiomes. PLoS One. 2016;11(11)
68. Rooijers K, Kolmeder C, Juste C, Dore J, de Been M, et al. In iterative workflow for mining the human intestinal metaproteome. BMC Genomics. 2011;12(6)
69. Walker A, Pfitzner B, Neschen S, Kahle M, Harir M, et al. Distinct signatures of host-microbial meta-metabolome and gut microbiome in two C57BL/6 strains under high-fat diet. ISME J. 2014;8(12):2380–96
70. El Aidy S, Derrien M, Merrifield CA, Levenez F, Dore J, et al. Gut bacteria-host metabolic interplay during conventionalisation of the mouse germfree colon. ISME J. 2013;7(4):743–55

71. Wilmes P, Heintz-Buschart A, Bond PL. A decade of metaproteomics: where we stand and what the future holds. Proteomics. 2015;15(20):3409–17
72. Nicholson JK, Lindon JC. Metabonomics. Nature. 2008;455:1054–6
73. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17(6):333–51
74. Eid J, Fehr A, Gray J, Luong K, Lyle J, et al. Real-time DNA sequencing from single polymerase molecules. Science. 2009;323:133–8
75. Deamer D, Akeson M, Branton D. Three decades of nanopore sequencing. Nat Biotechnol. 2016;34(5):518–24
76. Frank JA, Pan Y, Tooming-Klunderud A, Eijsink VG, McHardy AC, et al. Improved metagenome assemblies and taxonomic binning using long-read circular consensus sequence data. Sci Rep. 2016;6
77. Marshall CW, Ross DE, Fichot EB, Norman RS, May HD. Electrosynthesis of commodity chemicals by an autotrophic microbial community. Appl Environ Microbiol. 2012;78(23):8412–20
78. Singer E, Bushnell B, Coleman-Derr D, Bowman B, Bowers RM, et al. High-resolution phylogenetic microbial community profiling. ISME J. 2016;10(8):2020–32
79. Schloss PD, Jenior ML, Koumpouras CC, Westcott SL, Highlander SK. Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. PeerJ. 2016;4
80. Wagner J, Coupland P, Browne HP, Lawley TD, Francis SC, et al. Evaluation of PacBio sequencing for full-length bacterial 16S rRNA gene classification. BMC Microbiol. 2016;16(1)
81. Mosher JJ, Bowman B, Bernberg EL, Shevchenko O, Kan J, et al. Improved performance of the PacBio SMRT technology for 16S rDNA sequencing. J Microbiol Methods. 2014;104:59–60
82. Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, et al. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med. 2015;7
83. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 2013;41(1)
84. Stackebrandt E. Goebel BM: taxonomic note: a place for DNA-DNA reassociation and 16s rRNA sequence analysis in the prsent species definition in bacteriology. Int J Syst Evol Microbiol. 1994;44:846–849
85. Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, et al. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int J Syst Evol Microbiol. 1987;37(4):463–4
86. van de Peer Y, Chapelle S. De Wachter R: a quantitative map of nucleotide substitution rates in bacterial rRNA. Nucleic Acids Res. 1996;24(17):3381–91
87. Hermes GD, Zoetendal EG, Smidt H. Molecular ecological tools to decipher the role of our microbial mass in obesity. Benef Microbes. 2015;6(1):61–81
88. Claesson MJ, Wang Q, O'Sullivan O, Greene-Diniz R, Cole JR, et al. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. Nucleic Acids Res. 2010;38(22)
89. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. Proc Natl Acad Sci U S A. 2012;109(16):6241–6
90. McMurdie PJ. Holmes S: waste not, want not: why rarefying microbiome data is inadmissable. PLoS Comput Biol. 2014;10(4)
91. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41 Database issue:D590–596
92. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, et al. Ribosomal database project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 2014;42 Database issue:D633–42
93. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72(7):5069–72
94. Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, et al. Sequence-specific error profile of Illumina sequencers. Nucleic Acids Res. 2011;39(13)
95. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7:335–6
96. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75(23):7537–41
97. McMurdie PJ, Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013;8(4)
98. Albanese D, Fontana P, De Filippo C, Cavalieri D, Donati C. MICCA: a complete and accurate software for taxonomic profiling of metagenomic data. Sci Rep. 2015;5
99. Ramiro-Garcia J, Hermes GDA, Giatsis C, Sipkema D, Zoetendal EG, et al. NG-Tax, a highly accurate and validated pipeline for analysis of 16S rRNA amplicons from complex biomes. F1000Res. 2016;5
100. Faith D. Conservation evaluation and phylogenetic diversity. Biol Conserv. 1992;61:1–10
101. Shannon CE. A mathematical theory of communication. Bell Syst Tech J. 1948;27:523–656
102. Simpson EH. Measurement of diversity. Nature. 1949;163:688
103. Chiarucci A, Bacaro G, Scheiner SM. Old and new challenges in using species diversity for assessing biodiversity. Philos Trans R Soc Lond Ser B Biol Sci. 2011;366(1576):2426–2437
104. Heip CHR, Herman PMJ, Soetart K. Indices of diversity and evenness. Oceanis. 1998;24(4):61–87
105. Chao A. Nonparametric estimation of the number of classes in a population. Scand J Stat. 1984;11(4):265–70
106. Chao A, Lee S-M. Estimating the number of classes via sample coverage. J Am Stat Assoc. 1992;84(417):210–217
107. Hughes JB, Hellmann JJ, Ricketts TH, Bohannan BJM. Counting the uncountable: statistical approaches to estimating microbial diversity. Appl Environ Microbiol. 2001;67(10):4399–406
108. Jaccard P. The distribution of the flora in the alpine zone. New Phytol. 1912; XI(2):37–50
109. Bray JR, Curtis JT. An ordination of the upland Forest communities of southern Wisconsin. Ecol Monogr. 1957;27(4):326–49
110. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol. 2005;71(12):8228–35
111. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, et al. A core gut microbiome in obese and lean twins. Nature. 2009;457(7228):480–4
112. Ley RE, Backhed F, Turnbaugh P, Lozupone CA, Knight RD, et al. Obesity alters gut microbial ecology. Proc Natl Acad Sci U S A. 2005;102(31)
113. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, et al. Human gut microbiome viewed across age and geography. Nature. 2012; 486(7402):222–7
114. Walker AW, Ince J, Duncan SH, Webster LM, Holtrop G, et al. Dominant and diet-responsive groups of bacteria within the human colonic microbiota. ISME J. 2011;5(2):220–30
115. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, et al. Linking long-term dietary patterns with gut microbial enterotypes. Science. 2011;334(6052):105–8
116. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486(7402):207–14
117. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, et al. A survey of best practices for RNA-seq data analysis. Genome Biol. 2016;17
118. Del Fabbro C, Scalabrin S, Morgante M, Giorgi FM. An extensive evaluation of read trimming effects on Illumina NGS data analysis. PLoS One. 2013;8(12)
119. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10
120. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12:59–60
121. Leimena MM, Ramiro-Garcia J, Davids M, Van den Bogert B, Smidt H, et al. A comprehensive metatranscriptome analysis pipeline and its validation using human small intestine microbiota datasets. BMC Genomics. 2013;14
122. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, et al. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods. 2012;9(8):811–4
123. Sharon I, Banfield JF. Genomes from metagenomics. Science. 2013; 342(1057):1057–8
124. Pop M. Genome assembly reborn: recent computational challenges. Brief Bioinform. 2009;10(4):354–66
125. Davids M, Hugenholtz F, Martins Dos Santos V, Smidt H, Kleerebezem M, et al. Functional profiling of unfamiliar microbial communities using a validated de novo assembly metatranscriptome pipeline. PLoS One. 2016;11(1)
126. Namiki T, Hachiya T, Tanaka H, Sakakibara Y. MetaVelvet: an extension of velvet assembler to de novo metagenome assembly from short sequence reads. Nucleic Acids Res. 2012;40(20)
127. Wu Y-W, Tsang Y-H, Tringe SG, Simmons BA, Singer SW. Maxbin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome. 2014;2(26)

128. Yang B, Peng Y, Leung HC-M, Yiu S-M, Chen J-C, et al. Unsupervised binning of environmental genomic fragments based on an error robust selection of l-mers. BMC Bioinformatics. 2010;(11)
129. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25(7):1043–55
130. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. Genome Res. 2007;17(3):377–86
131. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15(46)
132. McHardy AC, Martin HG, Tsirigos A, Hugenholtz P, Rigoutsos I. Accurate phylogenetic classification of variable-length DNA fragments. Nat Methods. 2007;4(1):63–72
133. Lindgreen S, Adair KL, Gardner PP. An evaluation of the accuracy and speed of metagenome analysis tools. Sci Rep. 2016;6
134. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11
135. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000;28(1):33–36
136. Claudel-Renard C, Chevalet C, Faraut T. Khan D: enzyme-specific profiles for genome annotation - PRIAM. Nucleic Acids Res. 2003;31(22):6633–9
137. Karp PD, Paley S, Altman T. Data mining in the MetaCyc family of pathway databases. In: Data mining for systems biology: methods and protocols. Vol. 939. New York: Springer Science+Business Media; 2013. p. 183–200.
138. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42 Database issue:D490–5
139. Yin Y, Mao X, Yang J, Chen X, Mao F, et al. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2012; 40 Web Server issue:w445–51
140. The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. Nat Genet. 2000;(25):25–9
141. Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, et al. InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res. 2012;40 Database issue:D306–12
142. Radivojac P, Clark WT, Oron TR, Schnoes AM, Wittkop T, et al. A large-scale evaluation of computational protein function prediction. Nat Methods. 2013;10(3):221–7
143. Kantor RS, Wrighton KC, Handley KM, Sharon I, Hug LA, et al. Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. MBio. 2013;4(5)
144. Probst AJ, Weinmaier T, Raymann K, Perras A, Emerson JB, et al. Biology of a widespread uncultivated archaeon that contributes to carbon fixation in the subsurface. Nat Commun. 2014;5
145. Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, et al. Fermentation, hydrogen, and sulfur metabolism in mulitple uncultivated bacterial phyla. Science. 2012;(337):1661–5
146. Go Enrichment Analysis [http://geneontology.org/page/go-enrichment-analysis]. Accessed 12 Feb 2017.
147. Prosser JI. Replicate or lie. Environ Microbiol. 2010;12(7):1806–10
148. Markowitz VM, Ivanova NN, Szeto E, Palaniappan K, Chu K, et al. IMG/M: a data management and analysis system for metagenomes. Nucleic Acids Res. 2008;36 Database issue:D534–8
149. Mitchell A, Bucchini F, Cochrane G, Denise H, ten Hoopen P, et al. EBI metagenomics in 2016—an expanding and evolving resource for the analysis and archiving of metagenomic data. Nucleic Acids Res. 2016;44(D1):D595–603
150. Shi Y, Tyson GW, Eppley JM, DeLong EF. Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. ISME J. 2011;5(6):999–1013
151. Nocker A, Richter-Heitmann T, Montijn R, Schuren F, Kort R. Discrimination between live and dead cellsin bacterial communities from environmental water samples analyzed by 454 pyrosequencing. Int Microbiol. 2010;13(2):59–65
152. Kopylova E, Noe L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28(24):3211–7
153. Celaj A, Markle J, Danska J, Parkinson J. Comparison of assembly algorithms for improving rate of metatranacriptomic functional annotation. Microbiome. 2014;2(39)
154. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40
155. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12)
156. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, et al. Diet rapidly and reproducibly alters the human gut microbiome. Nature. 2014; 505(7484):7537–41
157. Turnbaugh PJ, Quince C, Faith JJ, McHardy AC, Yatsunenko T, et al. Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. Proc Natl Acad Sci U S A. 2010;107(16):7503–8
158. Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, et al. Shotgun metaproteomics of the human distal gut microbiota. ISME J. 2009;3(2):179–89
159. Fisher CK, Mehta P. Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. PLoS One. 2014;9(7)
160. Schubert AM, Rogers MA, Ring C, Mogle J, Petrosino JP, et al. Microbiome data distinguish patients with Clostridium Difficile infection and non-C. Difficile-associated diarrhea from healthy controls. MBio. 2014;5(3)
161. Cui H, Zhang Y. Alignment-free supervised classification of metagenomes by recursive SVM. BMC Genomics. 2013;14(641)
162. Knights D, Costello EK, Knight R. Supervised classification of human microbiota. FEMS Microbiol Rev. 2011;35(2):343–59
163. Statnikov A, Henaff M, Narendra V, Konganti K, Li Z, et al. A comprehensive evaluation of multicategory classification methods for microbiomic data. Microbiome. 2013;1(11)
164. Parloff R. Why deep learning is suddenly changing your life. In: Fortune. New York: Time Inc.; 2016.
165. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, et al. Enterotypes of the human gut microbiome. Nature. 2011;473(7346):174–80
166. Jain AK. Data clustering: 50 years beyond k-means. Pattern Recogn Lett. 2010;(31):651–666
167. Desgraupes B. clusterCrit: clustering indices. In: R package version 1.2.7 edn; 2016.
168. Knights D, Ward TL, McKinlay CE, Miller H, Gonzalez A, et al. Rethinking "enterotypes". Cell Host Microbe. 2014;16(4):433–7
169. Schubert E, Koos A, Emrich T, Züfle A, Schmid KA, et al. A framework for clustering uncertain data. Proc VLDB Endowment. 2015;18(12):1976–9
170. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, et al. The WEKA data mining software—an update. SIGKDD Explorations. 2003;11(1):10–18
171. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 2012;40 Database issue:D109–14
172. May A, Brand BW, El-Kebir M, Klau GW, Zaura E, et al. metaModules identifies key functional subnetworks in microbiome-related disease. Bioinformatics. 2015;32(11):1678–85
173. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, et al. Population-level analysis of gut microbiome variation. Science. 2016;352(6285):560–4
174. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, et al. Human genetics shape the gut microbiome. Cell. 2014;159(4):789–99
175. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, et al. Personalized nutrition by prediction of glycemic responses. Cell. 2015;163(5):1079–94
176. Ondov BD, Bergman NH. Philippy AM: interactive metagenomic visualization in the web browser. BMC Bioinformatics. 2011;12
177. Gehlenborg N, O'Donoghue SI, Baliga NS, Goesmann A, Hibbs MA, et al. Visualization of omics data for systems biology. Nat Methods. 2010;7(3 Suppl):S56–68
178. Heer J, Bostock M, Ogievetsky V. A tour through the visualization zoo. Commun ACM. 2010;53(6):59–67
179. Kreft JU. Conflicts of interest in biofilms. Biofilms. 2004;1(4):265–276
180. Shou W, Ram S, Vilar JM. Synthetic cooperation in engineered yeast populations. Proc Natl Acad Sci U S A. 2007;104(6):1877–82
181. Hansen AK, Moran NA. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. Proc Natl Acad Sci. 2011; 108(7):2849–54
182. Van Leuven JT, Meister RC, Simon C, McCutcheon JP. Sympatric speciation in a bacterial endosymbiont results in two genomes with the functionality of one. Cell. 2014;158(6):1270–80
183. Morris JJ, Lenksi RE, Zinser ER. The black queen hypothesis: evolution of dependencies through adaptive gene loss. MBio. 2012;3(2)
184. Shetty SA, Hugenholtz F, Lahti L, Smidt H, de Vos WM. Intestinal microbiome landscaping: insight in community assemblage and implications for microbial modulation strategies. FEMS Microbiol Rev. 2017; 41(2):182–99

185. Borenstein E. Computational systems biology and in silico modeling of the human microbiome. Brief Bioinform. 2012;13(6):769–80
186. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, et al. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol. 2010;28(9):977–82
187. Karp PD, Paley S, Romero P. The pathway tools software. Bioinformatics. 2002;18 Suppl 1:S225-32
188. KBase - predictive biology [http://kbase.us]. Accessed 12 Feb 2017.
189. Thiele I, Palsson BO. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc. 2010;5(1):93–121
190. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc Natl Acad Sci U S A. 2007;104(6):1777–82
191. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, et al. A comprehensive genome-scale reconstruction of Escherichia coli metabolism—2011. Mol Syst Biol. 2011;7
192. van Heck RG, Ganter M, Martins Dos Santos VA, Stelling J. Efficient Reconstruction of Predictive Consensus Metabolic Network Models. PLoS Comput Biol. 2016;12(8)
193. Heinken A, Sahoo S, Fleming RM, Thiele I. Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut. Gut Microbes. 2013;4(1):28–40
194. El-Semman IE, Karlsson FH, Shoaie S, Nookaew I, Soliman TH, et al. Genome-scale metabolic reconstructions of *Bifidobacterium adolescentis* L2-32 and *Facebaclibacterium prausnitzii* A2-165 and their interaction. BMC Syst Biol. 2014;8(41)
195. Levy R, Borenstein E. Reverse ecology: from systems to enviornments and back. Adv Exp Med Biol. 2012;751:329–45
196. Feist AM, Palsson BO. The biomass objective function. Curr Opin Microbiol. 2010;13(3):344–9
197. Khandelwal RA, Olivier BG, Roling WF, Teusink B, Bruggeman FJ. Community flux balance analysis for microbial consortia at balanced growth. PLoS One. 2013;8(5)
198. Zomorrodi AR, Islam MM. Maranas CD: d-OptCom: dynamic multi-level and multi-objective metabolic modeling of microbial communities. ACS Synth Biol. 2014;3(4):247–57
199. Harcombe WR, Riehl WJ, Dukovski I, Granger BR, Betts A, et al. Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. Cell Rep. 2014;7(4):1104–15
200. Munoz-Tamayo R, Laroche B, Walter E, Dore J, Duncan SH, et al. Kinetic modelling of lactate utilization and butyrate production by key human colonic bacterial species. FEMS Microbiol Ecol. 2011;76(3):615–24
201. Munoz-Tamayo R, Laroche B, Walter E, Dore J, Leclerc M. Mathematical modelling of carbohydrate degradation by human colonic microbiota. J Theor Biol. 2010;266(1):189–201
202. Xavier JB, Foster KR. Cooperation and conflict in microbial biofilms. Proc Natl Acad Sci U S A. 2007;104(3):876–881
203. Bucci V, Bradde S, Biroli G, Xavier JB. Social interaction, noise and antibiotic-mediated switches in the intestinal microbiota. PLoS Comput Biol. 2012;8(4)
204. Marino S, Baxter NT, Huffnagle GB, Petrosino JF, Schloss PD. Mathematical modeling of primary succession of murine intestinal microbiota. Proc Natl Acad Sci U S A. 2014;111(1):439–44
205. Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, et al. A whole-cell computational model predicts phenotype from genotype. Cell. 2012;150(2):389–401
206. Castiglione F, Pappalardo F, Bianca C, Russo G, Motta S. Modeling biology spanning different scales: an open challenge. Biomed Res Int. 2014;2014
207. Williams CF, Walton GE, Jiang L, Plummer S, Garaiova I, et al. Comparative analysis of intestinal tract models. Annu Rev Food Sci Technol. 2015;6:329–50
208. Van den Abbeele P, Belzer C, Goossens M, Kleerebezem M, De Vos WM, et al. Butyrate-producing Clostridium cluster XIVa species specifically colonize mucins in an in vitro gut model. ISME J. 2013;7(5):949–61
209. Kim HJ, Li H, Collins JJ, Ingber DE. Contributions of microbiome and mechanical deformation to intestinal bacterial overgrowth and inflammation in a human gut-on-a-chip. Proc Natl Acad Sci U S A. 2016; 113(1):E7–15
210. Williams SC. Gnotobiotics. Proc Natl Acad Sci U S A. 2014;111(5):1661
211. Lee SM, Donaldson GP, Mikulski Z, Boyajian S, Ley K, et al. Bacterial colonization factors control specificity and stability of the gut microbiota. Nature. 2013;501(7467):426–429
212. Samuel BS, Hansen EE, Manchester JK, Coutinho PM, Henrissat B, et al. Genomic and metabolic adaptations of Methanobrevibacter smithii to the human gut. Proc Natl Acad Sci U S A. 2007;104(25):10643–8
213. Laycock G, Sait L, Inman C, Lewis M, Smidt H, et al. A defined intestinal colonization microbiota for gnotobiotic pigs. Vet Immunol Immunopathol. 2012;149:216–24.
214. Sonnenburg JL, Chen CT, Gordon JI. Genomic and metabolic studies of the impact of probiotics on a model gut symbiont and host. PLoS Biol. 2006;4(12)
215. Backhed F, Manchester JK, Semenkovich CF, Gordon JI. Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. Proc Natl Acad Sci U S A. 2007;104(3):979–84
216. Kong LC, Tap J, Aron-Wisnewsky J, Pelloux V, Basdevant A, et al. Gut microbiota after gastric bypass in human obesity: increased richness and associations of bacterial genera with adipose tissue genes. Am J Clin Nutr. 2013;98(1):16–24
217. Zhang H, DiBaise JK, Zuccolo A, Kudrna D, Braidotti M, et al. Human gut microbiota in obesity and after gastric bypass. Proc Natl Acad Sci U S A. 2009;106(7):2365–70
218. Tremaroli V, Karlsson F, Werling M, Stahlman M, Kovatcheva-Datchary P, et al. Roux-en-Y gastric bypass and vertical banded Gastroplasty induce long-term changes on the human gut microbiome contributing to fat mass regulation. Cell Metab. 2015;22(2):228–38
219. Graessler J, Qin Y, Zhong H, Zhang J, Licinio J, et al. Metagenomic sequencing of the human gut microbiome before and after bariatric surgery in obese patients with type 2 diabetes: correlation with inflammatory and metabolic parameters. Pharmacogenomics J. 2013;13(6):514–22
220. van Nood E, Vrieze A, Nieuwdorp M, Fuentes S, Zoetendal EG, et al. Duodenal infusion of donor feces for recurrent Clostridium difficile. N Engl J Med. 2013;368(5):407–15
221. Jalanka J, Mattila E, Jouhten H, Hartman J, de Vos WM, et al. Long-term effects on luminal and mucosal microbiota and commonly acquired taxa in faecal microbiota transplantation for recurrent Clostridium difficile infection. BMC Med. 2016;14(1)
222. Borody TJ, Warren EF, Leis S, Surace R, Ashman O. Treatment of ulcerative colitis using fecal bacteriotherapy. J Clin Gastroenterol. 2003;37(1):42–7
223. Bojanova DP, Bordenstein SR. Fecal transplants: what is being transferred? PLoS Biol. 2016;14(7)
224. Ott SJ, Waetzig GH, Rehman A, Moltzau-Anderson J, Bharti R, et al. Efficacy of sterile fecal filtrate transfer for treating patients with Clostridium difficile infection. Gastroenterology. 2017;152(4):799–811
225. Petrof EO, Khoruts A. From stool transplants to next-generation microbiota therapeutics. Gastroenterology. 2014;146(6):1573–1582
226. Smith MB, Kelly C, Alm EJ. How to regulate faecal transplants. Nature. 2014; 506:290–1
227. de Vos WM. Fame and future of faecal transplantations–developing next-generation therapies with synthetic microbiomes. Microb Biotechnol. 2013;6(4):316–25
228. Roberfroid M. Prebiotics: the concept Revisisted. J Nutr. 2007;137(3 Suppl 2): 830S-7S
229. Martin FP, Wang Y, Sprenger N, Yap IK, Lundstedt T, et al. Probiotic modulation of symbiotic gut microbial-host metabolic interactions in a humanized microbiome mouse model. Mol Syst Biol. 2008;4
230. Ventura M, O'Connell-Motherway M, Leahy S, Moreno-Munoz JA, Fitzgerald GF, et al. From bacterial genome to functionality; case bifidobacteria. Int J Food Microbiol. 2007;120(1-2):2–12
231. Veiga P, Pons N, Agrawal A, Oozeer R, Guyonnet D, et al. Changes of the human gut microbiome induced by a fermented milk product. Sci Rep. 2014;4
232. Nami Y, Abdullah N, Haghshenas B, Radiah D, Rosli R, et al. Probiotic assessment of enterococcus durans 6HL and Lactococcus lactis 2HL isolated from vaginal microflora. J Med Microbiol. 2014;63(Pt 8):1044–51
233. Sakata T, Kojima T, Fujieda M, Takahashi M, Michibata T. Influences of probiotic bacteria on organic acid production by pig caecal bacteria in vitro. Proc Nutr Soc. 2003;62(1):73–80
234. Hosseini E, Grootaert C, Verstraete W, Van de Wiele T. Propionate as a health-promoting microbial metabolite in the human gut. Nutr Rev. 2011; 69(5):245–58
235. De Keersmaecker SC, Verhoeven TL, Desair J, Marchal K, Vanderleyden J, et al. Strong antimicrobial activity of lactobacillus rhamnosus GG against salmonella typhimurium is due to accumulation of lactic acid. FEMS Microbiol Lett. 2006;259(1):89–96
236. Gilad O, Jacobsen S, Stuer-Lauridsen B, Pedersen MB, Garrigues C, et al. Combined transcriptome and proteome analysis of Bifidobacterium animalis subsp. lactis BB-12 grown on xylo-oligosaccharides and a model of their utilization. Appl Environ Microbiol. 2010;76(21):7285–91

237. Everard A, Belzer C, Geurts L, Ouwerkerk J, Druart C, et al. Cross-talk betweeen *Akkermansia muciniphila* and intestinal epithelium controls diet-induced obesity. Proc Natl Acad Sci U S A. 2013;110(22):9066–71
238. Etxeberria U, Arias N, Boque N, Macarulla MT, Portillo MP, et al. Reshaping faecal gut microbiota composition by the intake of trans-resveratrol and quercetin in high-fat sucrose diet-fed rats. J Nutr Biochem. 2015;26(6):651–60
239. Selma MV, Espin JC, Tomas-Barberan FA. Interaction between phenolics and gut microbiota: role in human health. J Agric Food Chem. 2009;57(15):6485–6501
240. Carrington D. Reading the book of life. In: BBC news. London, UK; 2000.
241. Weiss R, Gillis J. Teams finish mapping human DNA. In: The Washington post. Washington: WP Company LLC; 2000.
242. Drmanac R. The advent of personal genome sequencing. Genet Med. 2011; 13(3):188–90
243. Maher B. Personal genomes: The case of the missing heritability. Nature. 2008;456(6):18–21
244. Offit K. Personalized medicine: new genomics, old lessons. Hum Genet. 2011;130(1):3–14
245. Snyder M, Du J, Gerstein M. Personal genome sequencing: current approaches and challenges. Genes Dev. 2010;24(5):423–31
246. Flores M, Glusman G, Brogaard K, Price ND, Hood L. P4 medicine: how systems medicine will transform the healthcare sector and society. Future Med. 2013;10(6):565–576
247. Hood L, Balling R, Auffray C. Revolutionizing medicine in the 21st century through systems approaches. Biotechnol J. 2012;7(8):992–1001
248. Robbins MJ. I got my personal genome mapped and it was bullshit. In: vicecom. New York: VICE Media LLC; 2013.
249. Hanage WP. Microbiology: microbiome science needs a healthy dose of scepticism. Nature. 2014;512(7514):247–8
250. Carroll AE. Exciting microbe research? Temper that giddy feeling in your gut. In: The New York times. New York: The New York Times Company; 2017.