



## High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota

E G Zoetendal, M Rajilic-Stojanovic and W M de Vos

*Gut* 2008;57;1605-1615  
doi:10.1136/gut.2007.133603

---

Updated information and services can be found at:  
<http://gut.bmj.com/cgi/content/full/57/11/1605>

---

*These include:*

### References

This article cites 105 articles, 58 of which can be accessed free at:  
<http://gut.bmj.com/cgi/content/full/57/11/1605#BIBL>

### Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

---

### Notes

---

To order reprints of this article go to:  
<http://journals.bmj.com/cgi/reprintform>

To subscribe to *Gut* go to:  
<http://journals.bmj.com/subscriptions/>

# High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota

E G Zoetendal,<sup>1,2</sup> M Rajilić-Stojanović,<sup>1</sup> W M de Vos<sup>1,2,3</sup>

<sup>1</sup> Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands; <sup>2</sup> TI Food and Nutrition, Wageningen, The Netherlands; <sup>3</sup> Department of Basic Veterinary Medicine, University of Helsinki, Helsinki, Finland

Correspondence to:  
Dr E G Zoetendal, Laboratory of Microbiology, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, The Netherlands; [erwin.zoetendal@wur.nl](mailto:erwin.zoetendal@wur.nl)

## ABSTRACT

The human gastrointestinal (GI) tract microbiota plays a pivotal role in our health. For more than a decade a major input for describing the diversity of the GI tract microbiota has been derived from the application of small subunit ribosomal RNA (SSU rRNA)-based technologies. These not only provided a phylogenetic framework of the GI tract microbiota, the majority of which has not yet been cultured, but also advanced insights into the impact of host and environmental factors on the microbiota community structure and dynamics. In addition, it emerged that GI tract microbial communities are host and GI tract location-specific. This complicates establishing relevant links between the host's health and the presence or abundance of specific microbial populations and argues for the implementation of novel high-throughput technologies in studying the diversity and functionality of the GI tract microbiota. Here, we focus on the recent developments and applications of phylogenetic microarrays based on SSU rRNA sequences and metagenomics approaches exploiting rapid sequencing technologies in unravelling the secrets of our GI tract microbiota.

The human gastrointestinal (GI) tract consists of different and connected organs that are involved in supplying the human body with nutrients and energy sources by the conversion and absorption of food. The human GI tract has a well-known anatomical architecture and is approximately 7 m long with a surface area of approximately 300 m<sup>2</sup> in adults.<sup>1</sup> Since it is continuously exposed to the outside environment, the GI tract has several protection systems. These include the low pH in the stomach, the coverage of the complete GI tract with a mucus layer, an enormous army of immune cells that lie beneath this mucus layer, and the presence of commensal microbes that abundantly colonise the GI tract. These microbial communities are collectively called commensal microbiota or GI tract microbiota. This GI tract microbiota is distributed along the entire GI tract, with the density and diversity increasing from the stomach to the colon. The acknowledged functions of the GI tract microbiota include the conversion of indigestible food components<sup>2-3</sup> and the production of essential vitamins and co-factors. However, our understanding of the GI tract microbiota is fragmented, which is due to the limited accessibility of the different parts of the GI tract and the immensely complex and diverse community structure of the GI tract microbiota

which differs between individuals, intestinal location and age groups.<sup>4</sup>

## THE GASTROINTESTINAL TRACT MICROBIOTA

The myriad of microbial cells in the GI tract, which outnumber our body cells by a factor of at least 10, has a large species diversity and consists of cultivated species as well as those that have not yet been cultured. A recent estimation of the cultivable fraction of the GI tract microbiota includes 442 bacterial, three archaeal, and 17 eukaryotic species.<sup>5</sup> However, it is evident that any figures that report the known diversity of the GI tract microbiota are continuously outdated due to the steady reporting of novel intestinal inhabitants. The majority of the currently available cultivated representatives were discovered following the introduction of anaerobic cultivation technologies that are still being used as standard tools.<sup>6</sup> Novel bacteria that are still being isolated from the human GI tract include a wide variety of butyrate-producing bacteria belonging to the phylum Firmicutes<sup>7-8</sup> and bacteria belonging to the phylum Bacteroidetes,<sup>9-12</sup> which are abundant GI tract phyla. In addition, the use of alternative carbon sources brought the discovery of the first GI tract representatives of the phyla Verrucomicrobia and Lentisphaerae, *Akkermansia muciniphila* and *Vectivallis vadensis*, respectively.<sup>13-14</sup> *A muciniphila*, especially, is a relevant isolate since it is a specialist in mucin degradation and a very common and abundant inhabitant of the GI tract which has previously not been noted because of its small size and specific carbon source requirements.<sup>13-15-16</sup>

Recently, novel culturing methods have been developed which exploit microbeads or multiplexed solid surfaces and allow unconventional and high-throughput culturing approaches.<sup>17-18</sup> Such methods enable the simultaneous single-cell cultivation of thousands of microbes, which is essential for an appropriate study of the complex and dense GI tract microbiota. Application of high-throughput culturing is expected to expand our knowledge of this neglected yet important element of the human body. However, in the course of evolution many GI tract microbes have developed intimate relations with the host and with each other, which makes microbes dependent on the metabolic activity of another member of the ecosystem and, therefore, almost impossible to grow into pure culture.<sup>19</sup> As an alternative, the use

**Table 1** Overview of phylogenetic distribution of gastrointestinal tract inhabitants (A–P, see table footnote) that have been detected by culturing and by small subunit ribosomal RNA gene library sequence analysis. The cut-off value of SSU rRNA sequence similarity that is used for phylotype definition is 97%.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Reference (first author)
Cultured	33	31	0	208	13	41	11	1	0	6	67	4	18	2	1	0	Rajilić-Stojanović <sup>5</sup>
Healthy total microbiota	0	1	0	15	3	12	0	0	0	0	0	0	0	0	0	0	Zoetendal <sup>22</sup>
	0	15	0	40	13	24	0	0	0	0	0	0	0	0	0	0	Suau <sup>23</sup>
	0	5	0	17	4	13	0	0	0	0	0	0	0	0	0	0	Bonnet <sup>24</sup>
	9	1	0	8	0	5	0	0	0	0	2	0	0	0	0	0	Favier <sup>25</sup>
	0	0	0	15	3	9	0	0	0	0	0	0	0	0	0	0	Wang <sup>26</sup>
	0	8	0	25	4	15	0	2	6	6	5	0	0	0	0	0	Wang <sup>27</sup>
	1	12	0	43	7	28	2	0	3	3	6	0	0	0	1	0	Wang <sup>28</sup>
	8	18	0	115	26	58	0	0	0	2	2	1	0	0	0	0	Hayashi <sup>29</sup>
	4	6	0	45	12	21	0	0	0	0	0	0	0	0	1	0	Hayashi <sup>30</sup>
	0	16	0	55	15	17	0	0	0	1	4	2	0	0	0	0	Hayashi <sup>31</sup>
	0	9	0	37	9	24	0	0	0	2	1	0	0	0	0	0	Hold <sup>32</sup>
	10	62	1	301	81	93	1	0	4	5	2	4	1	0	1	0	Eckburg <sup>33</sup> *
	16	59	0	251	100	84	0	0	0	4	5	4	0	0	2	0	Saunier, unpublished
Ulcerative colitis	0	1	0	29	9	14	0	0	0	0	0	0	0	0	0	0	Hutson, unpublished
Crown	4	23	0	111	34	49	1	0	0	1	3	0	0	0	1	0	Mangin <sup>34</sup>
Specific primers	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	Heilig <sup>35</sup>
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Satokari <sup>36</sup>
	0	0	0	14	2	9	0	0	0	0	0	7	0	0	0	2	Rajilić-Stojanović <sup>5</sup>
Specific substrates/activity	1	0	0	24	3	20	0	0	0	0	0	0	0	0	0	0	Barcenilla <sup>7</sup> ; Duncan <sup>37</sup>
	0	0	0	22	1	19	0	0	0	0	2	1	0	0	1	0	Derrien <sup>38</sup>
	6	9	0	59	11	37	0	0	1	0	3	0	0	0	0	0	Leitch <sup>39</sup>

\*The cut-off value of small sub-unit rRNA sequence similarity that is used for phylotype definition is 99%.

A, Actinobacteria; B, Bacteroidetes; C, Cyanobacteria; D, Firmicutes (all); E, Clostridium cluster IV; F, Clostridium cluster XIVa; G, Fucobacteria; H, Lentisphaerae; I, Alpha-Proteobacteria; J, Beta-Proteobacteria; K, Gamma-Proteobacteria; L, Delta-Proteobacteria; M, Epsilon-Proteobacteria; N, Spirochaetes; O, Verrucomicrobia; P, TM7 phylum.

of small subunit ribosomal RNA (SSU rRNA) and its corresponding gene has become established for the classification and phylogenetic analysis of microbes. The SSU rRNA gene includes approximately 1500 bp, which is a sufficient size for comparative sequence analysis. SSU rRNA is present in literally all organisms, and due to its conserved function its sequence has remained relatively conserved throughout evolution. Nevertheless, it contains nine variable regions that can be used for the identification and differentiation of specific microbial species. In fact, all life forms can be classified using the SSU rRNA sequences and this has resulted in the discovery of the Archaea as the third domain of life (Bacteria and Eucarya are the other two domains) and the construction of the revolutionary "tree of life".<sup>20 21</sup> SSU rRNA and its corresponding gene can be obtained directly from any environmental sample without cultivation procedures and this allows the detection of basically all members of an ecosystem, including those that cannot be cultured yet. Currently, over 400 000 SSU rRNA sequences are available in DNA databases, which is far more than for any other gene (<http://www.arb-silva.de> (accessed 11 July 2008)). The first breakthrough discovery of the SSU rRNA based analyses was that cultivation represents only a small fraction (estimated to be between 10 and 50%, depending on the study) of the true microbial diversity within the GI tract (table 1).

Exploring the diversity of an ecosystem by comparative sequence analysis is based on identification of species-level phylogenetic types, ie, phylotypes. Since variation of the SSU rRNA gene sequence is present within the members of the same species,<sup>40 41</sup> and sometimes even between different copies of the SSU rRNA gene within the same microbial genome,<sup>42</sup> phylotypes are defined as groups of SSU rRNA gene sequences with a certain level of similarity. However, the cut-off value of SSU rRNA sequence similarity that is used for phylotype definition is not consistent between different studies and varies between 97 and 99%. The higher the cut-off value, the higher is the number of distinct phylotypes that will be found in the same clone library.<sup>43</sup> For diversity studies the SSU rRNA sequences can be targeted using different, often complementary, approaches. Commonly applied approaches include cloning and subsequent sequencing of SSU rRNA genes, fingerprinting approaches such as denaturing gradient gel electrophoresis (DGGE) and fluorescence in situ hybridisation (FISH), which allows visualisation of specific microbes. These approaches are relatively low throughput and the choice of approach will mainly depend on the question to be answered. Cloning of SSU rRNA genes delivers the most detailed phylogenetic information, fingerprinting is most often used to compare microbial communities and monitor their dynamics, while FISH is the preferred approach to quantify specific microbial populations. The combination of these approaches has greatly advanced the

phylogenetic framework of the enormous microbial diversity and novel insights into its ecology. It has to be realised that neither of the methods for studying GI tract microbiota gives absolutely accurate information about the diversity of this ecosystem, including the "gold standard" cloning and sequencing-based studies. For example, the Actinobacteria, which belong to the high G+C Gram-positive bacteria, are frequently undetected or under-represented in the clone libraries, although they represent a dominant fraction of GI tract microbiota, as has been demonstrated by FISH using SSU rRNA-targeted oligonucleotide probes.<sup>44 45</sup>

The composition of GI tract microbiota is most often studied by analysing faecal samples, but the microbiota of other intestinal samples, such as specimens from the human colon and ileum, have also been characterised.<sup>27 32 33 46 47</sup> Fingerprinting studies based on DGGE analysis of SSU rRNA gene amplicons indicated that the dominant community in faecal samples in about half of the subjects studied does not necessarily represent that found in other parts of the GI tract, including the colonic mucosa.<sup>48 49</sup> This most likely reflects the fact that the number of microbes that are associated with the mucosa is small compared to the number of microbes present in the colon lumen. Although most likely primed by mucosa-associated microbes, the composition of the latter can be affected by the growth on specific substrates that reach the colon. If these substrates become limited at the end of the colon, this could also explain why certain abundant microbial species in faeces are not viable any more and this argues for studying samples derived from several GI tract regions.<sup>50</sup>

Studies that employ SSU rRNA gene sequence analysis are rapidly expanding our knowledge about the diversity of the GI tract microbiota, and only a decade after their introduction, the number of molecularly detected GI tract phylotypes has by far outnumbered the cultivated GI tract species (fig 1). An intensive survey of the publications focusing on the diversity of the human GI tract microbiota demonstrated that from more than 1200 microbes described, only 12% were recovered by application of both molecular and cultivation-based approaches, while the vast majority (~75%) was detected solely as an SSU rRNA sequence.<sup>51</sup> These values have to be taken with caution as it has also to be realised that most of the SSU rRNA sequences are retrieved from intestinal samples, mostly faeces, from only a dozen individuals. Considering the location and individual-specific microbial composition of the GI tract, this indicates that the 1200 species described only reflect a fraction of the true microbial diversity, which is estimated to consist of up to 1000 microbes per individual and more than 5000 microbes in total.<sup>23 33 51 52</sup> This implies that we are still at the beginning of describing the GI tract microbial diversity. Hence, there is a need to develop and apply high-throughput approaches in

further culturing studies, SSU rRNA sequencing and microbial diversity analysis, notably at different locations in the GI tract. These are essential to enable linking microbial populations to host-related factors, such as host genotype, age, health status, geographical location, ethnic origin, and diet.

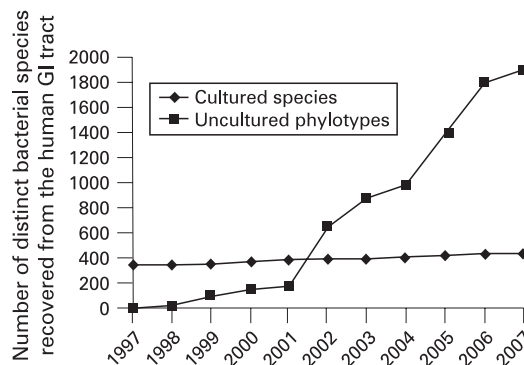
### PHYLOGENETIC MICROARRAY ANALYSIS OF GASTROINTESTINAL TRACT MICROBES

As indicated above, the human GI tract contains a microbiota, the diversity of which is beyond our imagination given the total number of microbes, and their location- and individual-specific composition. This complicates establishing links between members of the microbiota and GI tract disorders. At the moment, causal effects have been determined only for the well-studied and cultured pathogens such as *Helicobacter pylori*, *Listeria monocytogenes*, *Clostridium difficile* and members of the Enterobacteriaceae.<sup>53</sup> In addition, correlations have been suggested between the health status of individuals and the composition and activity of their microbiota. Irritable bowel syndrome (IBS) and intestinal bowel diseases (IBDs), such as Crohn's disease and ulcerative colitis, have frequently been associated with a rather unstable and disturbed microbiota composition in contrast to healthy individuals.<sup>54-57</sup> These differences have to be taken with caution since they can be influenced by several factors, including the use of medicine that may affect the microbiota. In addition, the unstable microbiota of patients with IBD can be caused by the disease status of the patients as differences in community structures were observed in patients who relapsed versus those who were in remission.<sup>54-57</sup> Last but not least, the stability of the microbiota in healthy individuals decreases when the time span between the sampling increases, and differs between microbial populations.<sup>5</sup> It was demonstrated that a significant fraction of microbial phylotypes is continuously present in the GI tract of a person over a 10 year time span, which indicates that the microbiota consists of a stable individual core of colonising microbes surrounded

by temporal visitors.<sup>5</sup> Despite these individual differences, there are indications that some microbial phylotypes are shared by different people and this led to the hypothesis that besides the individual core of microbes representing the stable colonisers in healthy individuals, humans also share a common core of microbes in their GI tract.<sup>58</sup> This hypothesis expands a previously proposed hypothesis that the human gastrointestinal microbiota is diverse but it is dominated by a limited number of bacterial species in everybody.<sup>59-61</sup>

As it is not yet possible to define the microbiota of a healthy intestinal tract, it is equally difficult to define the microbiota associated with an intestinal disorder. Another factor that complicates such an association is the fact that these GI tract disorders have a largely undefined, complex and possibly heterogeneous aetiology in which, in addition to the microbiota, host genetics and environmental factors also play a role. Moreover, our inability to cultivate all members of the microbiota makes it impossible to formulate hypotheses about the role of uncultured microbes in health and disease as a (partial) SSU rRNA sequence provides no information about the function of an organism. Last but not least, the individual-specific composition of the GI tract microbiota is also an important factor that complicates the establishment of links between microbes and the health status of the host.<sup>22 62</sup> Since different people are recruited by the different research groups, for which each of them has a favourite target microbe or methodology for microbiota analysis, comparisons between different studies are basically impossible. Moreover, as the number of persons tested is still low, it is evident that high-throughput analyses of the microbiota using standardised methods are needed to make statistically relevant links between the presence or quantity of uncultured bacterial populations and GI tract disorders.

The most commonly used high-throughput analytical method is DNA microarrays. DNA microarrays are basically glass surfaces, each the size of a microscopic slide, that are spotted with thousands of covalently linked DNA probes. These DNA microarrays can be hybridised with DNA or RNA and the most current applications include monitoring gene expression (transcriptional profiling) and detecting DNA sequence polymorphisms or mutations in genomic DNA. However, it is also possible to use DNA microarrays for diversity analysis (fig 2). Guschin and colleagues<sup>63</sup> described the first phylogenetic microarray (or diversity microarray) in which oligonucleotides complementary to SSU rRNA gene sequences of nitrifying bacteria were used to detect and identify these bacteria in environmental samples. Thereafter, DNA microarray technology has been implemented in a variety of ecological studies in which not only SSU rRNA genes, but also antibiotics resistance genes were used as targets.<sup>64-69</sup> Recently, the great potential of using phylogenetic microarrays to detect thousands of microbes simultaneously was demonstrated by DeSantis and colleagues.<sup>70</sup>



**Figure 1** Cumulative number of specific gastrointestinal (GI) tract phylotypes detected with culture-dependent and culture-independent approaches based on the data provided in table 1.

**Table 2** Overview of small subunit ribosomal RNA (SSU rRNA)-based phylogenetic microarrays which are fruitful to gain insight into the microbial diversity of the human gastrointestinal (GI) tract microbiota

Target microbes	Number of target organisms	SSU rRNA database	Number of probes	Reference (first author)
Human commensal isolates	40 bacterial species	GenBank	120	Wang <sup>61</sup>
Sulfate-reducing bacteria	All recognised lineages	ARB*	132	Loy <sup>65</sup>
All microbes	842 prokaryotic subfamilies	Greengenes, 15 March 2002 version†	297851	DeSantis <sup>70</sup>
Human oral cavity microbes	NA	NA	NA	Smoot <sup>72</sup>
Known GI tract commensals and medically relevant microbes	1590 bacterial and 39 archaeal species	prokMSA SSU rDNA sequence database, 2004 version‡	10500	Palmer <sup>73</sup>
Human GI tract microbiota	1140 bacterial species	Human GI tract Microbiota database, 2006 version§	4809	Rajilić-Stojanović <sup>5</sup>

\*Described by Ludwig *et al.*<sup>74</sup> †Described by DeSantis *et al.*<sup>75</sup> ‡Described by DeSantis *et al.*<sup>76</sup> §Described by Rajilić-Stojanović *et al.*<sup>51</sup> NA, not available.

This study illustrated that the use of phylogenetic microarrays is more powerful for the analysis of microbial community structure than a canonical clone library approach.<sup>70</sup> Besides this general phylogenetic microarray, there are also specific microarrays that focus on the microbial communities in specific ecosystems<sup>5 71 72</sup> (table 2). These include a phylogenetic microarray that targets the microbiota of the oral cavity and two microarrays that focus on the microbiota of the human GI tract.

The first studies in which phylogenetic microarrays were used to characterise the GI tract microbiota have illustrated the power of such an approach to gain insights into the structure and population dynamics in the GI tract (box 1). In one study 14 newborn babies were monitored over a year and the results showed that the microbiota of these infants is relatively simple and individual-specific, but chaotic in the early months of life followed by a similar pattern of development towards a more complex adult-like microbiota.<sup>77</sup> These results confirm and expand earlier studies performed using DGGE analysis.<sup>25</sup> The direct phylogenetic identification enabled by the microarray brought to light another remarkable observation: bifidobacteria were found to be only a minor fraction of the total microbiota of the infants analysed. This contrasts with observations in several previous studies which showed that bifidobacteria are the most dominant group in infants.<sup>78 79</sup> Although such a surprising finding could be partially explained by technical biases (eg, inadequate primer sequence or inefficient cell lysis), a biological explanation is also possible as infants from different studies originated from different continents and received different paediatric practices and diets.<sup>77</sup> The disparity of findings of different studies illustrates our present inability to define the normal GI tract microbiota even of its simplified form, which at present is in early infancy.

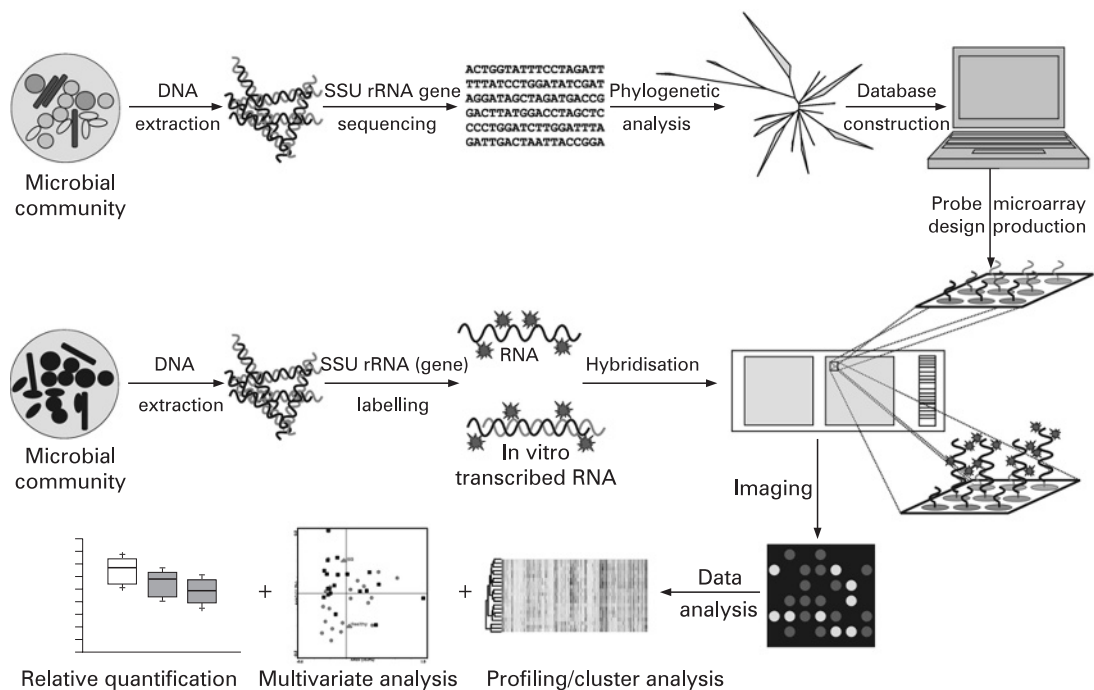
Besides focusing on the GI tract microbiota in healthy individuals, the first phylogenetic microarray studies on people with a GI tract disorder have also been performed.<sup>5</sup> As indicated before, the common GI tract disorders, such as IBD and IBS, are complex since they are influenced by the microbiota and host genetic and environmental factors<sup>80-83</sup> and, therefore, comprehensive and high-throughput approaches are needed to gain insight into this

complexity. The phylogenetic microarray analysis demonstrated that persons having a GI tract disorder had distinct microbiotas compared to healthy individuals and that the severity of the disease is correlated with the significance of the difference with the healthy group.<sup>5</sup> A striking observation was that patients with IBS had increased heterogeneity among the microbiota composition compared to healthy individuals.<sup>5</sup> This could be explained by the fact that IBS is known as a heterogeneous disorder based on varying clinical symptoms.<sup>84</sup> Despite this increased heterogeneity of the GI tract microbiota, a remarkable observation is that, in IBS, members of the Firmicutes were affected the most dramatically. This is in line with the previous phylogenetic microarray studies which indicated that the Firmicutes are more strongly affected by environmental changes than are Bacteroidetes and Actinobacteria.<sup>5</sup> This suggests that the alterations in the Firmicutes could be candidate biomarkers for IBS. The Firmicutes is the most diverse and abundant phylum within the GI tract microbiota consisting of a wide variety of uncultured organisms. As a result, the functionality of the majority of Firmicutes is not known and it is difficult to hypothesise what it might be. Therefore, explaining the observed trends is only possible for a few microbial groups, such as *Roseburia* spp., which are among the butyrate-producing isolates from the gut.<sup>85</sup> For instance, *Roseburia* spp. were found to be increased in patients suffering from diarrhoea-predominant IBS,<sup>5</sup> which adds to the controversy about the effect of butyrate on human health.<sup>86 87</sup> With respect to these GI tract disorders it is evident that obtaining more insight into the function of Firmicutes is crucial. To overcome the problem of the fact that the majority of Firmicutes will remain uncultured in the near future, culture-independent approaches to study functionality are needed.

The first applications of phylogenetic microarrays have already provided novel insights into the microbiota composition and dynamics in relation to health and disease and these approaches are promising for future research. However, the application of phylogenetic microarrays also has its limitations, as does any other microbiological approach. Phylogenetic microarray analyses are dependent on the isolation of nucleic acids and subsequent polymerase chain reaction (PCR) amplification of SSU rRNA genes, which are

## Recent advances in basic science

**Figure 2** Schematic representation of high-throughput analysis of human gastrointestinal (GI) tract microbiota via brute force sequencing and phylogenetic microarray analysis. SSU rRNA, small sub-unit ribosomal RNA.



vulnerable to technical biases. These are, unfortunately, general drawbacks of culture-independent technologies and should be minimised as much as possible. Furthermore, phylogenetic microarrays have a dynamic range that only covers the dominant microbes present in the GI tract. On the other hand, at present, there are numerous group- and species-specific quantitative PCR (qPCR) assays that are useful for quantification of different phylogenetic groups belonging to the human GI tract microbiota.<sup>88–90</sup> These specific PCR protocols can be combined with phylogenetic microarray analysis, which will allow determination of the diversity and relative abundance within low abundant groups. This could be of special interest for studying the population dynamics of pathogens or probiotics as they are usually present in low numbers.

In conclusion, the first phylogenetic microarray-based studies provided novel insights into the GI tract ecology by demonstrating correlations between certain microbial populations and host factors. To determine the significance of a certain correlation, the next questions to be addressed are whether this correlation reflects a causal relation and, if so, what mechanisms underlie the observed effects. Answering this type of question cannot be done by SSU rRNA gene-based approaches, but need the integration of functional-based approaches, including the application of the so-called meta-“omics” approaches in GI tract research, as will be discussed in the next section.

### METAGENOMICS AND OTHER COMMUNITY-BASED APPROACHES

SSU rRNA-based approaches are fruitful for describing the microbial diversity in the human GI tract and finding potential links between microbes and a certain health status. However,

the results obtained with such approaches cannot be interpreted beyond the description of the microbial diversity, since potential functions of these microbes cannot be extracted from SSU rRNA data. This means that a correlation between a disease status and the presence of a microbe cannot explain whether its presence is the result or the cause of the disorder. This means that other approaches are needed to gain insight into the potential roles of microbes in the human GI tract and how they are related to health and disease. In this respect, much insight has been gained from isolates that are known as pathogens and have been well-characterised in the laboratory. Animal and in vitro models have shed some light on the strategies these microbes have to interact with the host and, moreover, the availability of their genome sequences allows the discovery of genes involved in these interactions. Besides pathogens, commensal bacteria and their role in the GI tract have also been studied in a similar way. The pioneering work of Gordon and co-workers<sup>91–94</sup> and that of others showed that microbial colonisation improves nutritional and defensive functions of host. Moreover, the impact of the host on the microbe was also studied.<sup>94–95</sup> These so-called reductionist approaches provide insight into new genes and functions of individual organisms, which serve as models for community based studies.

Model systems have provided detailed insights into mechanisms that underlie the communication between host and microbe. However, there is still a huge gap between understanding these interactions in a model system and that in a complex ecosystem, such as the human GI tract. One way to gain insight into potential functions and activities of microbes without the need of cultivation is by performing metagenomics and other community approaches

(box 1). Metagenomics is a DNA-based approach to gain insights into the genetic potential of microbial communities, while the other meta-“omics” approaches focus on activity biomarkers, such as messenger RNA, proteins or metabolites (fig 3). Metagenomics is defined as the study of collected genomes from an ecosystem that can be used to study the phylogenetic, physical and functional properties of microbial communities.<sup>96</sup> Metagenomics is performed by extracting DNA from the microbial community followed by cloning of the DNA fragments in a suitable host (usually *E coli*) using a vector, such as fosmid or bacterial artificial chromosome vectors. This results in a metagenomic library that can be used for sequence-driven or function-driven analysis. Sequence-driven analyses are basically performed to obtain a snapshot of the genetic diversity of an ecosystem, while function-driven analyses are done to screen the library for novel enzymes or particular functions of interest.

Function-driven metagenomics offers great possibilities to discover new classes of genes with specific functions. The screening for these features requires functional transcription and translation of the genes, which are located in the metagenome clone, in the host that was used for the metagenome library construction. *E coli* is the most commonly used host for construction of the metagenomic library and it is predicted that it can express up to 40% of the functional potential from randomly cloned environmental DNA.<sup>97</sup> Function-driven metagenomic studies based on enzyme activity assays have already led to the discovery of novel activities, such as  $\beta$ -glucanases in the colon of mice,<sup>98</sup> bacterial hydrolases in rumen,<sup>99</sup> and antibiotic resistance in the oral cavity.<sup>100</sup> Recently, a phenotyping approach was used to screen for metagenomic clones that contain capacities to modulate the growth of epithelial cells in vitro.<sup>101</sup> This approach showed that the identification of potentially novel mechanisms of host-microbe interactions in the GI tract is possible. These types of screening methods can be very

useful to screen for beneficial or harmful functions that are present in the GI tract microbiota. Subsequently, the expression of the genes that are responsible for the function under study can be monitored in situ and serve as potential biomarkers for intervention studies.

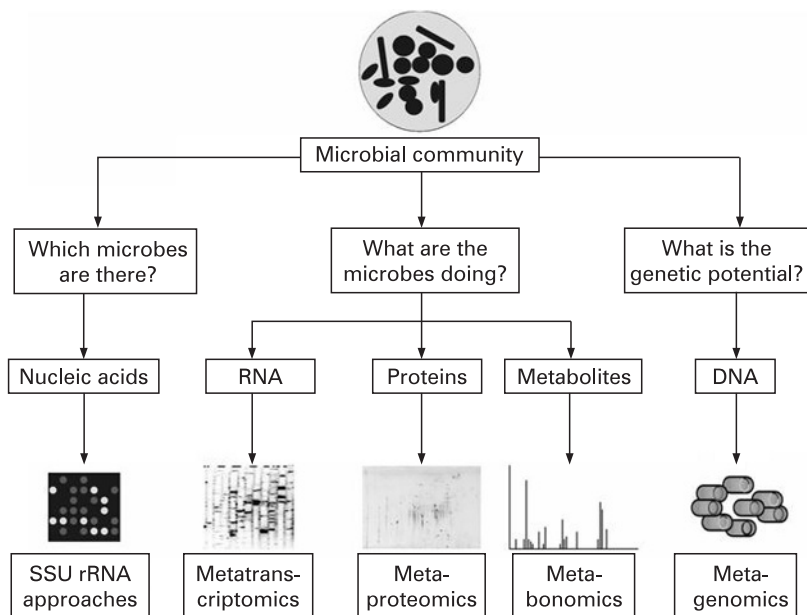
It has to be realised that functional screening is quite laborious since relatively large metagenomic libraries have to be screened to obtain a handful of positives per enzyme screen.<sup>99</sup> To overcome this drawback, a high-throughput screening approach, termed substrate-induced gene-expression screening was developed, which allowed fluorescence-activated cell sorting, in which catabolic genes were activated by various substrates.<sup>102</sup> Although this approach looks promising for high-throughput screening of human GI tract microbiota to discover novel enzyme-encoding genes, the positive identification relies on several host requirements, which include the recognition and transport of the inducing substrate and the ability of the host to express the corresponding genes located on the cloned insert.

Sequence-driven metagenomics approaches are used to create a catalogue of the genetic potential that is present in an ecosystem, which can be fruitful to gain insight into the functionality of the particular ecosystem. Sequence-driven metagenomics has already been applied to study a variety of ecosystems and the first metagenomic approach that was performed on the human GI tract described the diversity of the viral community in human faeces.<sup>103</sup> This study revealed that the viral community in faeces is very diverse consisting of approximately 1200 viral genotypes from which the majority of viral sequences had the highest sequence similarity to phages that are known to infect Gram-positive bacteria. As mentioned previously, Gram-positive bacteria, which include the phyla Firmicutes and Actinobacteria, were identified as dominant members of the GI tract microbiota in SSU rRNA gene-based studies. The first prokaryotic sequence-driven metagenomic approach focused on the SSU rRNA sequences that are represented in the metagenomic libraries from pooled faecal samples from healthy individuals and from patients with Crohn's disease.<sup>104</sup> This study demonstrated that Firmicutes are significantly reduced in complexity in patients with Crohn's disease compared to healthy subjects. This is in line with SSU rRNA gene-based observations, including those in which phylogenetic microarrays were used.<sup>5</sup> In another study the “mobile metagenome” of the human GI tract microbiota was investigated. In this study transposon-aided capture was developed and applied to the study of the plasmid-encoded genes that are present in the human GI tract.<sup>105</sup> In addition to genes involved in replication and mobilisation of the plasmids, genes encoding for phosphoesterase or phosphohydrolase enzymes could be detected. Besides these targeted sequence-driven metagenomic approaches, large-scale sequence-driven metagenomic approaches have also been applied

### Box 1 Novel insights that have been gained from high-throughput analysis of the human gastrointestinal (GI) tract microbiota

- ▶ The human GI tract microbiota is composed of numerous uncultured microbes and predominated by Firmicutes, Bacteroidetes and Actinobacteria
- ▶ In healthy adults the human GI tract microbiota fluctuates around a stable individual core of phylotypes that are affected by host genetics, environmental and stochastic factors
- ▶ In infants the GI tract microbiota is succeeding from an unstable chaotic community towards a stable adult community
- ▶ A reduction in the abundance and diversity of Firmicutes is frequently associated with intestinal bowel diseases and irritable bowel syndrome
- ▶ The human GI tract microbiome is enriched in functions that are essential to the human host
- ▶ The human GI tract microbiome from healthy human adults consists of a functionally uniform core and is a hotspot for gene transfer





**Figure 3** Schematic representation of the metagenomics and other community-based "omics" approaches. SSU rRNA, small subunit ribosomal RNA.

to investigate the genetic potential in the GI tract ecosystem.<sup>106 107</sup> The first study demonstrated that many genes in the colonic microbiota represent functions that are essential to the host, such as vitamin production, which have often been ascribed to the microbiota.<sup>106</sup> Remarkably, sequences affiliated to Bacteroidetes, one of the dominant microbial groups in the colon, were not detected in this library, which indicated that cell lysis, DNA extraction and cloning procedures can have a major impact on the genetic diversity that is represented in a library. Recently, a comparative metagenomics approach was described to compare and contrast the genetic diversity in the human colon of 13 subjects, including adults and infants.<sup>107</sup> This study indicated that the human microbiota is a hot spot for horizontal gene transfer. Moreover, this study demonstrated that the genomic features of infants were less complex and individual-specific, while those of adults were more complex with high functional uniformity between individuals. The latter finding confirms the proposal that all human GI tract microbiotas share a common core of microbes, despite their individual-specific composition.<sup>58</sup>

Sequence-driven approaches result in a wealth of scattered pieces of sequences and this means that special tools for post-metagenomic analysis are indispensable. Although software packages have been developed to re-assemble the genomic fragments that are sequenced, a major limitation of metagenomics is that the genetic diversity in an ecosystem is enormous. Despite the continuous development of novel, relatively cheap, sequencing technologies, such as 454 pyrosequencing,<sup>108</sup> which will result in increased numbers of metagenomic sequences, at the moment it is almost impossible to obtain reasonable coverage

for complex ecosystems such as the human GI tract microbiota. Nevertheless, this raw sequence information can be analysed by comparative metagenomics and these types of analyses can already reveal the identification of genes that are specific or enriched in a certain ecosystem or niche.<sup>109 110</sup>

It has to be realised that the detection of genes in a metagenomic library does not necessarily mean that these are functionally important. Therefore, metagenomics should be considered more as catalogues for activity-based approaches and that other meta-"omics" approaches, which use RNA, proteins and metabolites as targets, are better approaches to gain insight into the activity and functionality of the microbes in an ecosystem (fig 3). These meta-"omics" approaches are still in their infancy,<sup>111</sup> but we expect that they will be extensively used in the near future.

### CONCLUDING REMARKS AND FUTURE PERSPECTIVES

For more than a decade it has been recognised that a major part of the human GI tract microbiota has not been characterised by cultivation. This has led to the implementation of sequence analysis of SSU rRNA and its corresponding gene in studying the human GI tract microbiota diversity and this has provided tremendous expansion of our knowledge about the ecology of the GI tract. However, these approaches have indicated that the GI tract microbiota is individual- and location-specific, and that its diversity is enormous with thousands of novel microbial species to be discovered.<sup>51</sup> This argues for the introduction of novel high-throughput and comprehensive technologies for studying the microbiota in the human GI tract, as described in this paper.

The implementation of high-throughput phylogenetic microarrays allows the simultaneous analysis of thousands of microbes in a single experiment and, therefore, is very attractive for studying the population dynamics of the GI tract microbiota in health and disease. This resulted in the discovery that the development of the microbiota in infants is initially chaotic but stabilises towards an adult-like community after 1 year.<sup>77</sup> In addition, phylogenetic microarray analysis indicated that the human microbiota fluctuates around an individual core of stable colonisers.<sup>5</sup> Last, but not least, significant links have been established between the presence or abundance of specific groups of microbes and GI tract disorders such as IBS and IBD.<sup>5</sup> Therefore, it is already evident that the implementation of phylogenetic microarrays in GI tract research will increase our knowledge in the near future. Nevertheless, the up-to-date status of phylogenetic microarrays will always depend on the discovery of novel GI tract inhabitants and, therefore, ongoing sequencing of SSU rRNA gene libraries and cultivation of the novel GI tract inhabitants are indispensable.

It is evident that phylogenetic microarrays will enable us to make correlations between microbial

groups and characteristics of the host. However, this will not lead to extrapolation of microbial functions as the majority of GI tract microbes are only known as a partial SSU rRNA gene. Therefore, metagenomics and other meta-“omics” approaches are needed to gain insight into the genetic potential and activity of GI tract microbiota. The field of these meta-“omics” is hardly 5 years old and their application in the study of the human GI tract is even younger. Therefore, the analysis and interpretation of data derived from meta-“omics” approaches are still in infancy. We expect many novel technological procedures to be developed and improved in the coming years, not only with respect to wetlab technologies, such as pyrosequencing and functional screening tools, but also in the field of bioinformatics to analyse the enormous mass of data that is obtained with such approaches. Nevertheless, the first applications of meta-“omics” approaches have already demonstrated their power as discussed in this paper and we expect this field to grow explosively in the near future.

Only an integration of all reductionist and meta-“omics” approaches in the near future will provide adequate understanding of GI tract microbiota, as these approaches complement each other by delivering different pieces of the GI tract puzzle. For example, the analysis of different cell cultures will result in the discovery of novel genes and functions of individual organisms and therefore, serve as milestones for meta-“omics” approaches. This integration is essential to explain data derived from meta-“omics” analysis, since the limitations in our predictive capacity was demonstrated in the first metaproteomics study of the human GI tract microbiota.<sup>112</sup>

Overall, it will be a challenging future for GI tract researchers. With the introduction of the novel high-throughput technologies that are described here, it will be possible, for the first time, to obtain statistically relevant links between microbial phylotypes and activities, and human health. Ultimately, this will lead to the discovery of biomarkers that will help us to understand and predict the microbial life in our intestine, which is the dream of a GI tract microbial ecologist.

**Competing interests:** None.

## REFERENCES

1. **Bengmark S.** Ecological control of the gastrointestinal tract. The role of probiotic flora. *Gut* 1998;**42**:2–7.
2. **Cummings JH, Macfarlane GT.** Colonic microflora: Nutrition and health. *Nutrition* 1997;**13**:476.
3. **MacFarlane GT, Cummings JH, Alison C.** Protein degradation by human intestinal bacteria. *J Gen Microbiol* 1986;**132**:1647–56.
4. **Zoetendal EG, Vaughan EE, de Vos WM.** A microbial world within us. *Mol Microbiol* 2006;**59**:1639–50.
5. **Rajilić-Stojanović M.** *Diversity of the human gastrointestinal microbiota: novel perspectives from high throughput analyses.* Laboratory of microbiology. [PhD thesis]. Wageningen, The Netherlands, Wageningen University, 2007:1–214.
6. **Aranki A, Syed SA, Kenney EB, et al.** Isolation of anaerobic bacteria from human gingival and mouse cecum by means of a simplified glove box procedure. *Appl Microbiol* 1969;**17**:568–76.
7. **Barcenilla A, Pryde SE, Martin JC, et al.** Phylogenetic relationships of butyrate-producing bacteria from the human gut. *Appl Environ Microbiol* 2000;**66**:1654–61.
8. **Pryde SE, Duncan SH, Hold GL, et al.** The microbiology of butyrate formation in the human colon. *FEMS Microbiol Lett* 2002;**217**:133.
9. **Hayashi H, Shibata K, Bakir MA, et al.** *Bacteroides coprophilus* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* 2007;**57**(Pt 6):1323–6.
10. **Hayashi H, Shibata K, Sakamoto M, et al.** *Prevotella copri* sp. nov. and *Prevotella stercorea* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* 2007;**57**(Pt 5):941–6.
11. **Bakir MA, Kitahara M, Sakamoto M, et al.** *Bacteroides intestinalis* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* 2006;**56**:151–4.
12. **Sakamoto M, Kitahara M, Benno Y.** *Parabacteroides johnsonii* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* 2007;**57**(Pt 2):293–6.
13. **Derrien M, Vaughan EE, Plugge CM, et al.** *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int J Syst Evol Microbiol* 2004;**54**:1469–76.
14. **Zoetendal EG, Plugge CM, Akkermans ADL, et al.** *Victivallis vadensis* gen. nov., sp. nov., a sugar-fermenting anaerobe from human faeces. *Int J Syst Evol Microbiol* 2003;**53**:211–5.
15. **Collado MC, Derrien M, Isolauri E, et al.** Intestinal integrity and *Akkermansia muciniphila*, a mucin-degrading member of the intestinal microbiota present in infants, adults, and the elderly. *Appl Environ Microbiol* 2007;**73**:7767–70.
16. **Derrien M, Collado MC, Ben-Amor K, et al.** The mucin-degrader *Akkermansia muciniphila* is an abundant member of the human intestinal tract. *J Appl Environ Microbiol* 2008;**74**:1646–8.
17. **Zengler K, Toledo G, Rappe M, et al.** Cultivating the uncultured. *Proc Natl Acad Sci U S A* 2005;**99**:15681–5.
18. **Ingham CJ, Sprengels A, Bomer J, et al.** The micro-Petri dish, a million-well growth chip for the culture and high-throughput screening of microorganisms. *Proc Natl Acad Sci U S A* 2007;**104**:18217–22.
19. **Stams AJM.** Metabolic interactions between anaerobic bacteria in methanogenic environments. *Antonie Van Leeuwenhoek* 1994;**66**:271–94.
20. **Woese CR.** Bacterial evolution. *Microbiol Rev* 1987;**51**:221–71.
21. **Woese CR, Kandler O, Wheelis ML.** Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* 1990;**87**:4576–9.
22. **Zoetendal EG, Akkermans AD, de Vos WM.** Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bacteria. *Appl Environ Microbiol* 1998;**64**:3854–9.
23. **Suau A, Bonnet R, Sutren M, et al.** Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Appl Environ Microbiol* 1999;**65**:4799–807.
24. **Bonnet R, Suau A, Doré J, et al.** Differences in rDNA libraries of faecal bacteria derived from 10- and 25-cycle PCRs. *Int J Syst Evol Microbiol* 2002;**52**:757–63.
25. **Favier CF, Vaughan EE, de Vos WM, et al.** Molecular monitoring of succession of bacterial communities in human neonates. *Appl Environ Microbiol* 2002;**68**:219–26.
26. **Wang M, Ahrne S, Antonsson M, et al.** T-RFLP combined with principal component analysis and 16S rRNA gene sequencing: an effective strategy for comparison of fecal microbiota in infants of different ages. *J Microbiol Meth* 2004;**59**:53.
27. **Wang X, Heazlewood SP, Krause DO, et al.** Molecular characterization of the microbial species that colonize human ileal and colonic mucosa by using 16S rDNA sequence analysis. *J Appl Microbiol* 2003;**95**:508–20.
28. **Wang M, Ahrne S, Jeppsson B, et al.** Comparison of bacterial diversity along the human intestinal tract by direct cloning and sequencing of 16S rRNA genes. *FEMS Microbiol Ecol* 2005;**54**:219.
29. **Hayashi H, Sakamoto M, Benno Y.** Phylogenetic analysis of the human gut microbiota using 16S rDNA clone libraries and strictly anaerobic culture-based methods. *Microbiol Immunol* 2002;**46**:535–48.
30. **Hayashi H, Sakamoto M, Benno Y.** Fecal microbial diversity in a strict vegetarian as determined by molecular analysis and cultivation. *Microbiol Immunol* 2002;**46**:819–31.
31. **Hayashi H, Sakamoto M, Benno Y.** Molecular analysis of fecal microbiota in elderly individuals using 16S rDNA library and T-RFLP. *Microbiol Immunol* 2003;**47**:557–70.
32. **Hold GL, Pryde SE, Russell VJ, et al.** Assessment of microbial diversity in human colonic samples by 16S rDNA sequence analysis. *FEMS Microbiol Ecol* 2002;**39**:33–9.
33. **Eckburg PB, Bik EM, Bernstein CN, et al.** Diversity of the human intestinal microbial flora. *Science* 2005;**308**:1635–8.

34. **Mangin I**, Bonnet R, Seksik P, *et al.* Molecular inventory of faecal microflora in patients with Crohn's disease. *FEMS Microbiol Ecol* 2004;**50**:25–36.
35. **Heilig HG**, Zoetendal EG, Vaughan EE, *et al.* Molecular diversity of *Lactobacillus* spp. and other lactic acid bacteria in the human intestine as determined by specific amplification of 16S ribosomal DNA. *Appl Environ Microbiol* 2002;**68**:114–23.
36. **Satokari RM**, Vaughan EE, Akkermans ADL, *et al.* Bifidobacterial diversity in human feces detected by genus-specific PCR and denaturing gradient gel electrophoresis. *Appl Environ Microbiol* 2001;**67**:504–13.
37. **Duncan SH**, Louis P, Flint HJ. Lactate-utilizing bacteria, isolated from human feces that produce butyrate as a major fermentation product. *Appl Environ Microbiol* 2004;**70**:5810–7.
38. **Derrien M**, Adawi D, Ahrné S, *et al.* The intestinal mucosa as a habitat of the gut microbiota and a rational target for probiotic functionality and safety. *Microbiol Ecol Health Dis* 2004;**16**:137–44.
39. **Leitch EC**, Walker AW, Duncan SH, *et al.* Selective colonization of insoluble substrates by human faecal bacteria. *Environ Microbiol* 2007;**9**:667–9.
40. **Klappenbach JA**, Saxman PR, Cole JR, *et al.* rrndb: the ribosomal RNA operon copy number database. *Nucl Acids Res* 2001;**29**:181–4.
41. **Sacchi CT**, Alber D, Dull P, *et al.* High level of sequence diversity in the 16S rRNA genes of *Haemophilus influenzae* isolates is useful for molecular subtyping. *J Clin Microbiol* 2005;**43**:3734–42.
42. **Acinas SG**, Marcelino LA, Klepac-Ceraj V, *et al.* Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. *J Bacteriol* 2004;**186**:2629–35.
43. **Venter JC**, Remington K, Heidelberg JF, *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 2004;**304**:66–74.
44. **Harmsen HJM**, Raangs GC, He T, *et al.* Extensive set of 16S rRNA-based probes for detection of bacteria in human feces. *Appl Environ Microbiol* 2002;**68**:2982–90.
45. **Lay C**, Rigottier-Gois L, Holmstrom K, *et al.* Colonic microbiota signatures across five Northern European countries. *Appl Environ Microbiol* 2005;**71**:4153–5.
46. **Marteau P**, Pochart P, Doré J, *et al.* Comparative study of bacterial groups within the human cecal and fecal microbiota. *Appl Environ Microbiol* 2001;**67**:4939–42.
47. **Ahmed S**, Macfarlane GT, Fite A, *et al.* Mucosa-associated bacterial diversity in relation to human terminal ileum and colonic biopsy samples. *Appl Environ Microbiol* 2007;**73**:7435–42.
48. **Zoetendal EG**, von Wright A, Vilpponen-Salmela T, *et al.* Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Appl Environ Microbiol* 2002;**68**:3401–7.
49. **Lepage P**, Seksik P, Sutren M, *et al.* Biodiversity of the mucosa-associated microbiota is stable along the distal digestive tract in healthy individuals and patients with IBD. *Inflamm Bowel Dis* 2005;**11**:473–80.
50. **Ben-Amor K**, Heilig HG, Smidt H, *et al.* Genetic diversity of viable, injured, and dead fecal bacteria assessed by fluorescence-activated cell sorting and 16S rRNA gene analysis. *Appl Environ Microbiol* 2005;**71**:4679–89.
51. **Rajilić-Stojanović M**, Smidt H, de Vos WM. Diversity of the human gastrointestinal tract microbiota revisited. *Environ Microbiol* 2007;**9**:2125–36.
52. **Moore WEC**, Holdeman LV. Human fecal flora: the normal flora of 20 Japanese-Hawaiians. *Appl Microbiol* 1974;**27**:961–79.
53. **Acheson DW**, Luccioli S. Microbial–gut interactions in health and disease. Mucosal immune responses. *Best Pract Res Clin Gastroenterol* 2004;**18**:387–404.
54. **Seksik P**, Rigottier-Gois L, Gramet G, *et al.* Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. *Gut* 2003;**52**:237–42.
55. **Andoh A**, Sakata S, Koizumi Y, *et al.* Terminal restriction fragment length polymorphism analysis of the diversity of fecal microbiota in patients with ulcerative colitis. *Inflamm Bowel Dis* 2007;**13**:955–62.
56. **Maukonen J**, Satokari R, Mättö J, *et al.* Prevalence and temporal stability of selected clostridial groups in irritable bowel syndrome in relation to predominant faecal bacteria. *J Med Microbiol* 2006;**55**:625–33.
57. **Mättö J**, Maunukela L, Kajander K, *et al.* Composition and temporal stability of gastrointestinal microbiota in irritable bowel syndrome – a longitudinal study in IBS and control subjects. *FEMS Immunol Med Microbiol* 2005;**43**:213–22.
58. **Doré J**. A functional core of the human intestinal microbiota. In: *11th International Symposium on Microbial Ecology – ISME-11; The hidden powers – microbial communities in action*. 20–25 August 2006, Vienna, Austria. Further information from [events@isme-microbes.org](mailto:events@isme-microbes.org)
59. **Drasar BS**, Barrow PA. *Intestinal microbiology*. Washington, DC: American Society for Microbiology, 1985.
60. **Kleessen B**. Culture-based knowledge on biodiversity, development and stability of human gastrointestinal microflora. *Microb Ecol Health Dis* 2000;**12**(Suppl 2):53–63.
61. **Wang RF**, Beggs ML, Robertson LH, *et al.* Design and evaluation of oligonucleotide-microarray method for the detection of human intestinal bacteria in fecal samples. *FEMS Microbiol Lett* 2002;**213**:175–82.
62. **Ley RE**, Turnbaugh PJ, Klein S, *et al.* Microbial ecology: Human gut microbes associated with obesity. *Nature* 2006;**444**:1022–23.
63. **Guschin DY**, Mobarry BK, Proudnikov D, *et al.* Oligonucleotide microchips as biosensors for determinative and environmental studies in microbiology. *Appl Environ Microbiol* 1997;**63**:2397–402.
64. **Small J**, Call DR, Brockman FJ, *et al.* Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays. *Appl Environ Microbiol* 2001;**67**:4708–16.
65. **Loy A**, Lehner A, Lee N, *et al.* Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. *Appl Environ Microbiol* 2002;**68**:5064–81.
66. **El Fantroussi S**, Urakawa H, Bernhard AE, *et al.* Direct profiling of environmental microbial populations by thermal dissociation analysis of native rRNAs hybridized to oligonucleotide microarrays. *Appl Environ Microbiol* 2003;**69**:2377–82.
67. **Peplies J**, Glöckner FO, Amann R. Optimization strategies for DNA microarray-based detection of bacteria with 16S rRNA-targeting oligonucleotide probes. *Appl Environ Microbiol* 2003;**69**:1397–407.
68. **Call DR**, Bakko MK, Krug MJ, *et al.* Identifying antimicrobial resistance genes with DNA microarrays. *Antimicrob Agents Chemother* 2003;**47**:3290–5.
69. **Volokov D**, Chizhikov V, Chumakov K, *et al.* Microarray analysis of erythromycin resistance determinants. *J Appl Microbiol* 2003;**95**:787–98.
70. **DeSantis TZ**, Brodie EL, Moberg JP, *et al.* High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microb Ecol* 2007;**53**:371–83.
71. **Palmer C**, Bik EM, Eisen MB, *et al.* Rapid quantitative profiling of complex microbial populations. *Nucl Acids Res* 2006;**34**:e5.
72. **Smoot LM**, Smoot JC, Smidt H, *et al.* DNA microarrays as salivary diagnostic tools for characterizing the oral cavity's microbial community. *Adv Dent Res* 2005;**18**:6–11.
73. **Palmer C**, Bik EM, Digiulio DB, *et al.* Development of the human infant intestinal microbiota. *PLoS Biol* 2007;**5**:e177.
74. **Ludwig W**, Strunk O, Westram R, *et al.* ARB: a software environment for sequence data. *Nucl Acids Res* 2004;**32**:1363–71.
75. **DeSantis TZ**, Hugenholtz P, Larsen N, *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006;**72**:5069–72.
76. **DeSantis TZ**, Dubosarskiy I, Murray SR, *et al.* Comprehensive aligned sequence construction for automated design of effective probes (CASCADE-P) using 16S rDNA *Bioinformatics* 2003;**19**:1461–8.
77. **Palmer C**, Bik EM, Digiulio DB, *et al.* Development of the human infant intestinal microbiota. *PLoS Biol* 2007;**5**:e177.
78. **Harmsen HJM**, Wildeboer-Veloo ACM, Raangs GC, *et al.* Analysis of intestinal flora development in breast-fed and formula-fed infants by using molecular identification and detection methods. *J Pediatr Gastroenterol Nutr* 2000;**30**:61–7.
79. **Benno Y**, Sawada K, Mitsuoka T. The intestinal microflora of infants: composition of fecal flora in breast-fed and bottle-fed infants. *Microbiol Immunol* 1984;**28**:975–86.
80. **Gwee KA**. Irritable bowel syndrome in developing countries – a disorder of civilization or colonization? *Neurogastroent Motil* 2005;**27**:317–24.
81. **Sood A**, Midha V, Sood N, *et al.* Incidence and prevalence of ulcerative colitis in Punjab, North India. *Gut* 2003;**52**:1587–90.
82. **Bengtson MB**, Ronning T, Vatn MH, *et al.* Irritable bowel syndrome in twins: genes and environment. *Gut* 2006;**55**:1754–9.
83. **Halfvarson J**, Bodin L, Tysk C, *et al.* Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics. *Gastroenterology* 2003;**124**:1767.
84. **Sartor RB**. Mechanisms of disease: pathogenesis of Crohn's disease and ulcerative colitis. *Nat Clin Pract Gastroenterol Hepatol* 2006;**3**:390–407.
85. **Duncan SH**, Hold GL, Barcenilla A, *et al.* *Roseburia intestinalis* sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. *Int J Syst Evol Microbiol* 2002;**52**:1615–20.

86. **Macfarlane GT**, Gibson GR, Macfarlane S. Short chain fatty acid and lactate production by human intestinal bacteria grown in batch and continuous culture. In: Binder HJ, Cummings JH, Soergel KH, eds. *Short chain fatty acids*. London: Kluwer Publishing, 1994.
87. **Scheppach W**, Weiler F. The butyrate story: old wine in new bottles? *Curr Opin Clin Nutr Metab Care* 2004;**7**:563–7.
88. **Matsuki T**, Watanabe K, Fujimoto J, *et al*. Development of 16S rRNA-gene-targeted group-specific primers for the detection and identification of predominant bacteria in human feces. *Appl Environ Microbiol* 2002;**68**:5445–51.
89. **Matsuki T**, Watanabe K, Fujimoto J, *et al*. Use of 16S rRNA gene-targeted group-specific primers for real-time PCR analysis of predominant bacteria in human feces. *Appl Environ Microbiol* 2004;**70**:7220–8.
90. **Rinttilä T**, Kassinen A, Malinen E, *et al*. Development of an extensive set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal samples by real-time PCR. *J Appl Microbiol* 2004;**97**:1166–77.
91. **Hooper LV**, Gordon JI. Commensal host–bacterial relationships in the gut. *Science* 2001;**292**:1115–8.
92. **Hooper LV**, Xu J, Falk PG, *et al*. A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem. *Proc Natl Acad Sci U S A* 1999;**96**:9833–8.
93. **Derrien M**. *Mucin utilisation and host interactions of the novel intestinal microbe Akkermansia muciniphila*. *Laboratory of Microbiology*. [PhD thesis]. Wageningen, The Netherlands, Wageningen University, 2007:1–174.
94. **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;**12**:e413.
95. **de Vries MC**. *Analyzing global gene expression of Lactobacillus plantarum in the human gastrointestinal tract*. *Laboratory of Microbiology*. [PhD thesis]. Wageningen, The Netherlands, Wageningen University, 2006:1–160.
96. **Handelsman J**. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 2004;**68**:669–85.
97. **Gabor EM**, de Vries EJ, Janssen DB. Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. *Environ Microbiol* 2004;**6**:948–58.
98. **Walter J**, Mangold M, Tannock GW. Construction, analysis, and beta-glucanase screening of a bacterial artificial chromosome library from the large-bowel microbiota of mice. *Appl Environ Microbiol* 2005;**71**:2347–54.
99. **Ferrer M**, Golyshina OV, Chernikova TN, *et al*. Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environ Microbiol* 2005;**7**:1996–2010.
100. **Diaz-Torres ML**, Villedieu A, Hunt N, *et al*. Determining the antibiotic resistance potential of the indigenous oral microbiota of humans using a metagenomic approach. *FEMS Microbiol Lett* 2006;**258**:257–62.
101. **Gloux K**, Leclerc M, Iliozer H, *et al*. Development of high-throughput phenotyping of metagenomic clones from the human gut microbiome for modulation of eukaryotic cell growth. *Appl Environ Microbiol* 2007;**73**:3734–7.
102. **Uchiyama T**, Abe T, Ikemura T, *et al*. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat Biotechnol* 2005;**23**:88–93.
103. **Breitbart M**, Hewson I, Felts B, *et al*. Metagenomic analyses of an uncultured viral community from human feces. *J Bacteriol* 2003;**185**:6220–3.
104. **Manichanh C**, Rigottier-Gois L, Bonnaud E, *et al*. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 2006;**55**:205–11.
105. **Jones BV**, Marchesi J. Transposon-aided capture (TRACA) of plasmids resident in the human gut mobile metagenome. *Nat Methods* 2007;**4**:55–61.
106. **Gill SR**, Pop M, Deboy RT, *et al*. Metagenomic analysis of the human distal gut microbiome. *Science* 2006;**312**:1355–9.
107. **Kurokawa K**, Itoh T, Kuwahara T, *et al*. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res* 2007;**14**:169–81.
108. **Margulies M**, Egholm M, Altman WE, *et al*. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;**437**:376–80.
109. **Raes J**, Korbel JO, Lercher MJ, *et al*. Prediction of effective genome size in metagenomic samples. *Genome Biol* 2007;**8**:R10.
110. **Foerstner KU**, von Mering C, Bork P. Comparative analysis of environmental sequences: potential and challenges. *Phil Trans R Soc Lond, B Biol Sci* 2006;**361**:519–23.
111. **Booijink CC**, Zoetendal EG, Kleerebezem M, *et al*. Microbial communities in the human small intestine: coupling diversity to metagenomics. *Future Microbiol* 2007;**2**:285–95.
112. **Klaassens ES**, de Vos WM, Vaughan EE. A metaproteomics approach to study the functionality of the microbiota in the human infant gastrointestinal tract. *Appl Environ Microbiol* 2007;**73**:1388–92.