Molecular Methods for the Analysis of Gut Microbiota

Pawel Namsolleck¹, Ralph Thiel¹, Paul Lawson², Kim Holmström³, Mirjana Rajilic⁴, Elaine E. Vaughan⁴, Lionel Rigottier-Gois⁵, Matthew D. Collins², Willem M. de Vos⁴ and Michael Blaut¹

From the ¹German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany, ²School of Food Biosciences, University of Reading, UK, ³Department of Molecular Characterization, Bioneer A/S, Horsholm, Denmark, ⁴Laboratory of Microbiology, Wageningen University, The Netherlands and ⁵Institut National de la Recherche Agronomique, Jouy en Josas, France

Correspondence to: Michael Blaut, German Institute of Human Nutrition Potsdam-Rehbruecke, Arthur-Scheunert-Allee 114-116, 14458 Nuthetal, Germany. E-mail: blaut@mail.dife.de

Microbial Ecology in Health and Disease 2004; 16: 71–85

This review focuses on methodological approaches used to study the composition of human faecal microbiota. Gene sequencing is the most accurate tool for revealing the phylogenetic relationships between bacteria. The main application of fluorescence in situ hybridization (FISH) in both microscopy and flow cytometry is to enumerate faecal bacteria. While flow cytometry is a very fast method, FISH microscopy still has a considerably lower detection limit. Key words: DGGE, TGGE, dot blot, FISH, flow cytometry, gut ecology, in situ gene expression, micro-array, sequencing.

INTRODUCTION

The microbial community resident in the human intestinal tract has a considerable effect on health and well-being and has been implicated in both beneficial and adverse health effects. Diet is a major factor that affects the composition and the activity of the gut microorganisms. In order to influence the gut microbiota in a targeted way it is paramount to better characterize this microbial community. In the last decades, culture-independent methods based on 16S rRNA gene analysis have been applied to various microbial habitats. Easy-to-use methods are urgently required so as to be able to study the influence of diet on gut microbiota composition in large numbers of subjects. As the conventional methods for analysing the intestinal microbiota are time-consuming and tedious, high-throughput methods for the automated detection of fluorescently labelled cells based on microscopic image analysis, flow cytometry and DNA arrays are being developed.

PHYLOGENETIC MARKERS

In the 1970s, Carl Woese and co-workers (1, 2) identified 16S rRNA as an extremely useful phylogenetic marker, which has found a wide range of applications in microbial taxonomy and microbial ecology. Ribosomal RNA is the preferred molecule for bacterial identification and systematics. The information content of a marker molecule is defined as the log (base 2) of the number of possible character states (4 for nucleotides, 20 for amino acids) times the length of the sequence. Table I shows the information content of several marker molecules. The number of variable, and therefore informative, residues is the most important parameter for the applicability of the marker. The most widely used marker, the 16S rRNA, contains 974 variable residues for the domain Bacteria. The 23S rRNA contains more than twice as much information as the 16S rRNA and can be used as an additional marker for bacterial phylogeny.

It should be noted, however, that the discriminative power of the 16S rRNA has its limitations (3, 4). It has been demonstrated that the use of protein-coding gene sequences may be more effective for bacterial identification (5). Owing to the low evolutionary rate of the 16S rRNA genes, some ecologically distinct bacterial taxa cannot be distinguished. The 16S rRNA marker molecule is very useful for distinguishing moderately divergent populations. However, very closely related populations can be better distinguished by other means.

One of the phylogenetically interesting protein-coding genes is gyrB (encoding the subunit B of the bacterial DNA gyrase). The rate of horizontal gene transfer of this gene is low and it evolves at a higher rate than the rRNA gene (6). Phylogenetic trees based on the 16S rRNA sequences diverge less than gyrB-based trees. Other phylogenetic markers of interest are the 16S–23S intergenic spacer region (7, 8), the infB gene (encoding the translation
initiation factor 2) (9), the ATPase β-subunit (10) and the elongation factor Tu (11). Table II shows examples of sequence databases of various phylogenetic markers. In the future, classification of bacteria will be based on combinations of two or more unlinked phylogenetic markers. This could improve our understanding of the diversity and complexity of bacterial ecosystems.

PHYLOGENETIC DIVERSITY OF THE HUMAN GUT MICROBIOTA

The gastrointestinal tract (GIT) of humans is colonized by an extremely complex and diverse assemblage of microorganisms. Estimations indicate that it harbours approximately $10^{14}$ bacterial cells, which is approximately 10 times more than all tissue cells of the human body taken together. The GIT is inhabited by several hundred prokaryotic species, the majority of which are strict anaerobes belonging to the bacterial domain.

The taxonomic diversity of intestinal bacteria has been the subject of intense investigations over the last decades. In the 1970s and the 1980s, analysis of the microbial community in the human gut was largely dependent on the use of powerful enrichment procedures and the ability to grow strict anaerobes. The isolates were identified and characterized by various phenotypic methods. In a pioneering study, Moore and Holdeman (12) analysed the faecal flora of 20 male Japanese-Hawaiians. The total number of different bacterial species was estimated to exceed 400 or 500, although the actual number of identified species was only 113. Subsequent studies confirmed the great diversity of intestinal bacteria. The genera Bacteroides, Bifidobacterium, Clostridium, Eubacterium, Fusobacterium, Lactobacillus, Peptostreptococcus, Ruminococcus and Streptococcus were the first dominant taxa identified.

16S rRNA GENE SEQUENCING AND PHYLOGENETIC ANALYSIS

During the last decade, it has become clear that the taxonomic and phylogenetic diversity of the human gut microbiota has been grossly underestimated. This was caused by our inability to: a) reliably identify bacteria and recognize new species diversity and b) isolate and cultivate organisms by conventional microbiological methods.

The 16S rRNA gene has revolutionized the way taxonomists classify and identify bacteria. Through this approach, evolutionary relationships between organisms can be determined objectively by comparison of their rRNA gene sequences. Differences in compared sequences can be used to measure evolutionary distances, and phylogenetic relationships can be presented in the form of phylogenetic trees. Because 16S rRNA molecules contain regions with different degrees of variability (varying from conserved to highly variable regions) it is possible to distinguish organisms at different phylogenetic levels (from species to domain). Furthermore, many tens of thousands of bacterial 16S rRNA gene sequences are now available, including virtually all validly described species, which enables newly determined sequences to be compared to existing sequences and/or organisms. In collaboration with various laboratories, 16S rRNA gene sequencing has been used as a tool to search for new human intestinal bacteria. By sequencing short amplified rDNA fragments (approximately 500 bases proximal to the 5′ end of rRNA, which include diagnostic variable regions V1, V2 and V3) this approach was used to

Table I

Information content of phylogenetic marker molecules for the domain Bacteria (4, modified)

<table>
<thead>
<tr>
<th>Molecule</th>
<th>16S rRNA</th>
<th>23S rRNA</th>
<th>EF-Tu</th>
<th>ATPase β-subunit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (E. coli) n</td>
<td>1542</td>
<td>2904</td>
<td>394</td>
<td>460</td>
</tr>
<tr>
<td>Information (bits)</td>
<td>3084</td>
<td>5808</td>
<td>1706</td>
<td>1992</td>
</tr>
<tr>
<td>Number of conserved (Cons.) and variable (Var.) positions</td>
<td>568 974 934 1970 83 311 111 359</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information (bits)</td>
<td>1948</td>
<td>3940</td>
<td>1347</td>
<td>1555</td>
</tr>
</tbody>
</table>

n, nucleotides; aa, amino acids.

*Logarithm (base 2) of the number of possible character states (4 n; 20 aa) times the number of (E. coli) positions.

†Number of conserved (Cons.) and variable (Var.) positions.

Logarithm (base 2) of the number of possible character states (4 n; 20 aa) times the number of variable positions.

Table II

Sequence databases based on different phylogenetic markers

<table>
<thead>
<tr>
<th>Phylogenetic marker</th>
<th>Databases</th>
<th>Entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>16S rRNA</td>
<td>RDP <a href="http://rdp.cme.msu.edu/html/">http://rdp.cme.msu.edu/html/</a></td>
<td>79 900 aligned sequences</td>
</tr>
<tr>
<td>gyrB</td>
<td>ICB <a href="http://www.mbio.co.jp">http://www.mbio.co.jp</a></td>
<td>1295</td>
</tr>
</tbody>
</table>

n, nucleotides; aa, amino acids.

*Logarithm (base 2) of the number of possible character states (4 n; 20 aa) times the number of (E. coli) positions.
perform rapid phylogenetic identification on several hundred strictly anaerobic cultures. Although the great majority of isolates (> 90%) were readily assigned to established species, a large number of organisms possessed sequences which did not correspond to sequences available in any of the public sequence databases. These organisms have subsequently been subjected to detailed phylogenetic analysis based on full 16S rRNA gene sequences, and shown to correspond to hitherto unknown genera and/or species. Although detailed phenotypic characterization studies are ongoing on many of the unknown isolates, a plethora of new human intestinal species have now been formally described (Table III) and several others are pending.

Although the 16S rRNA gene sequencing of traditionally isolated and cultivated bacteria is adding to our knowledge of the diversity of the GIT microflora, many organisms cannot be cultivated and are therefore not yet included in conventional microbiological analysis. In recent years, polymerase chain reaction (PCR)-rDNA cloning and sequencing strategies have been used increasingly to directly access the phylogenetic diversity of bacteria within complex communities. By using universal primers, rRNA genes from potentially all bacteria in a sample can be amplified together, without cultivation and purification of organisms. Amplified rDNA products can then be cloned to obtain individual genes, which can be sequenced and subjected to phylogenetic analysis. Using this approach bacterial phylotypes can be readily identified by phylogenetic analysis of their sequences. Even if a sequence is novel, its position on the phylogenetic tree can be determined. This sequencing strategy was used to more comprehensively explore the phylogenetic diversity of faeces and colonic tissues. From earlier reports (13–15) and ongoing studies in our laboratories, it is now apparent that the great majority of bacterial species identified from the GIT using this culture-independent approach represent novel, previously undescribed species. Furthermore, it is now evident that the most prevalent sequence types fall into three main phylogenetic groupings referred to as the Bacteroides group, Clostridium coccoides group and Clostridium leptum group (Figs. 1–3).

Although our studies are still ongoing, they have already provided a wealth of information on species diversity, including the recognition of many new rDNA phylotypes and sequences, and the phylogenetic distribution of the predominant gut bacteria.

**Table III**

<table>
<thead>
<tr>
<th>Genus</th>
<th>Species</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anaerostipes</td>
<td>A. caccae</td>
<td>(70)</td>
</tr>
<tr>
<td>Anaerotricus</td>
<td>A. colhominis</td>
<td>(71)</td>
</tr>
<tr>
<td>'Anaerofustis</td>
<td>'A. stercorihominis'</td>
<td>Finegold et al., unpublished</td>
</tr>
<tr>
<td>'Stercorihominis'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alstipes</td>
<td>A. finegoldii</td>
<td>(72)</td>
</tr>
<tr>
<td></td>
<td>A. putridinis</td>
<td>(72)</td>
</tr>
<tr>
<td>Bryantella</td>
<td>B. formataxigens</td>
<td>(73)</td>
</tr>
<tr>
<td>Cetobacterium</td>
<td>C. somerae</td>
<td>(74)</td>
</tr>
<tr>
<td>Clostridium</td>
<td>C. bileae</td>
<td>(75)</td>
</tr>
<tr>
<td></td>
<td>C. hathewayi</td>
<td>(76)</td>
</tr>
<tr>
<td>Dorea</td>
<td>D. longicornis</td>
<td>(77)</td>
</tr>
<tr>
<td>Ruminococcus</td>
<td>R. litii</td>
<td>(78)</td>
</tr>
<tr>
<td>'Subdoligranulum'</td>
<td>'S. variabile'</td>
<td>Holmstrom et al., unpublished</td>
</tr>
</tbody>
</table>

**FLUORESCENCE IN SITU HYBRIDIZATION**

The principle of fluorescence in situ hybridization (FISH) is the detection of a target DNA or RNA site by a fluorescently labelled probe molecule. Based on the knowledge of the 16S rRNA sequence information it is possible to design a probe that specifically targets a given organism. Since each bacterial cell contains $10^3$–$10^5$ ribosomes, the rRNA probes find a sufficiently high number of targets to result in the illumination of the target cell.

FISH is a powerful method for the enumeration of bacteria in complex habitats such as the human gut. Most notably, it does not require cultivation of the target organisms. In previous years the number of probes available for the detection of intestinal bacteria has increased steadily. The developed probes target the bacteria at several levels of phylogenetic hierarchy. Table IV lists the most widely used probes for the characterization of the human gut microbiota. The list consists primarily of probes that target dominant phylogenetic groups.

Although FISH has been used widely in bacterial ecology, the method has to be adapted to the habitat under investigation and the type of samples analysed. The protocol of FISH has been optimized for the analysis of human faecal samples. A fresh faecal sample has to be thoroughly homogenized by rough shaking and the use of glass beads. This is a very important step, because only a small amount of the sample is analysed, therefore, it must be representative of the whole sample. Subsequently, large particles are removed by a brief centrifugation at low centrifugal force ($\leq 300$ g). Aliquots of the supernatant are subsequently fixed with paraformaldehyde and in parallel with ethanol (16). These treatments permeabilize the majority of cells. However, a number of primarily Gram-positive cells are not sufficiently permeabilized by this treatment and are therefore additionally treated with lysozyme or a lipase, which may help to render the cell envelope of the bacteria more permeable (17).

Although FISH samples for microscopy are usually prepared on glass slides, it is not necessary to use a transparent medium. Since the excitation light comes through the objective from above the sample, any surface can be used. Pernthaler et al. use filter membranes to concentrate the cells from a marine habitat by filtration. In
In this case, the bacterial cells are visualized on the membrane (18).

**Probes**

Most probes targeting the bacterial 16S rRNA have a length of 15–23 nucleotides. To make them detectable the probes are labelled with fluorochromes at their 5'-end. The choice of the fluorochrome depends on the available light source of the microscope and the corresponding filter sets. The small bandwidth of the excitation wavelength principally allows the use of several probes, each labelled with a different fluorochrome, within the same sample. The use of several probes enables the simultaneous detection of several target organisms in a microbial ecosystem.
Prior to probe design and validation it is recommended to check whether a probe for the target group has already been described. To facilitate the search, an online database is available, which allows a search to be carried out for target organisms or probe names (probeBase; http://www.microbial-ecology.de/probebase) (19). This database also gives

Fig. 2. The distribution of representative rDNA lines within the Clostridium cocoides phylogenetic group derived from human colonic bacteria, which do not correspond to taxonomically defined species. HuCc rDNA, clone derived from human GIT flora.
more detailed information about the probes, including the optimal hybridization conditions and the specificity. Probes can be designed to detect microorganisms at different phylogenetic levels (from species to domain) using rRNA regions that differ in their extent of conservation. To design a probe that hybridizes only with the target organism(s), it is recommended that only complete 16S rRNA sequences are used (20). The most commonly used sequence databases are

![Fig. 3. The distribution of representative rDNA lines within the Clostridium leptum phylogenetic group derived from human colonic bacteria, which do not correspond to taxonomically defined species. HuCl rDNA, clone derived from human GIT flora.](image-url)

Probe design is a process that includes several steps: identification of short regions in a sequence alignment existing only in the target group, generation of several oligonucleotide probes targeting organisms of interest, modification of the probe sequence to optimize the hybridization conditions (such as melting temperature), in silico validation of the newly designed probes (21). There are a few parameters characterizing a ‘good’ oligonucleotide probe: a minimum of one mismatch to the same region in all non-target sequences, central position of the mismatch in the non-target sequences (it decreases the stability of non-target rRNA hybrids), G + C content in the range of 50–60% (it influences the hybrid stability), high accessibility to the target region leading to a high relative fluorescence intensity (see below), no self-complementarity which mean a high accessibility prior to use. Therefore, new probes developed by rRNA sequence alignments may increase the accessibility of the probe molecules. The addition of osmotically active compounds, such as dextran sulfate, to the hybridization buffer may increase the hybridization rate of the molecules (24). Another possibility to increase the accessibility of the probe molecule to its target is the addition of helper oligonucleotide probes targeting organisms of interest and shows no non-specific binding. However, especially for probes that target phylogenetically diverse groups, this is a difficult task.

The aim of a specificity test is to find hybridization and washing conditions that lead to a specific recognition of the target sites of the rRNA of Escherichia coli. They developed a set of 171 probes that hybridized to E. coli cells and the binding of the probes was subsequently measured by flow cytometry. The intensity of the resulting fluorescence was taken as a direct measure for the accessibility of the rRNA molecule. Bright probe signals – which mean a good accessibility – were reported, as well as totally blocked sites (23). Although these data were obtained for E. coli only, it is very likely that differences in the accessibility of potential target sites exist in other species, as well. Therefore, new probes developed by rRNA sequence alignments have to be tested for their in situ accessibility prior to use. The ideal target region is highly specific and accessible for a given probe.

The function of RDP-II or the ‘Basic Local Alignment Search Tool’ (BLAST, http://www.ncbi.nlm.nih.gov/blast). The ideal probe targets all organisms of interest and shows no non-specific binding. However, especially for probes that target phylogenetically diverse groups, this is a difficult task.

Exemplary list of FISH probes with sequences, names and references

<table>
<thead>
<tr>
<th>Probe</th>
<th>Sequence from 5’ to 3’ end</th>
<th>OPD code</th>
<th>Reference</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aca194</td>
<td>CTATACCTGAGGGCCCTT</td>
<td>S-S-Acoc-0194-a-A-18</td>
<td>(79)</td>
<td>Butyrate producer</td>
</tr>
<tr>
<td>Ato 291</td>
<td>GTCGCGTCTTCACACC</td>
<td>S-S-Ato-0291-a-A-17</td>
<td>(80)</td>
<td>0.2–7% (81)</td>
</tr>
<tr>
<td>Bac 303</td>
<td>CCAATGTTGGGGGACCTT</td>
<td>S-S-Bacto-0303-a-A-17</td>
<td>(83)</td>
<td>2–5.5% (82)</td>
</tr>
<tr>
<td>Bif 164</td>
<td>CATCCGGCATTACCAACC</td>
<td>S-S-Bif-0164-a-A-18</td>
<td>(26)</td>
<td>0.2 – 1.5% (26) 2.3 – 4.4% (82)</td>
</tr>
<tr>
<td>E.bif462</td>
<td>CCCCCTACTACTACTCAC</td>
<td>S-S-Ebif-0462-a-A-18</td>
<td>(84)</td>
<td>0.1–3.3% (84)</td>
</tr>
<tr>
<td>Ecl 387</td>
<td>CGGGGCATTGCTGTTCA</td>
<td>S-S-Ecl-0387-a-A-18</td>
<td>(81)</td>
<td>0.1–7% (81)</td>
</tr>
<tr>
<td>Enter 1432</td>
<td>CTTTTGGCAACCCACT</td>
<td>S-S-Ent-1432-a-A-15</td>
<td>(39)</td>
<td>Potentially pathogens</td>
</tr>
<tr>
<td>Erec 482</td>
<td>GCTTCTTAGTCTRTAGGCG</td>
<td>S-S-Erec-0482-a-A-19</td>
<td>(24)</td>
<td>25–35% (24) 9–27% (82)</td>
</tr>
<tr>
<td>Fprau 645</td>
<td>CCTCTGCACTACTCAAGAAACAC</td>
<td>S-S-Fprau-0645-a-A-23</td>
<td>(37)</td>
<td>10–17% (37)</td>
</tr>
<tr>
<td>Lab 158</td>
<td>GGTTAGACAYCTGTHTTCA</td>
<td>S-G-Lab-0158-a-A-20</td>
<td>(17)</td>
<td>Used as probiotics</td>
</tr>
<tr>
<td>Rbro 730</td>
<td>TAAAGCCAGYAGCCGCG</td>
<td>S-S-Rbro-0730-a-A-18</td>
<td>(81)</td>
<td>Together</td>
</tr>
<tr>
<td>Rfla 729</td>
<td>AAAGCCCAATGACCGGCC</td>
<td>S-S-Rfla-0729-a-A-18</td>
<td>(81)</td>
<td>Butyrate producer</td>
</tr>
<tr>
<td>Rint1102</td>
<td>GCTTACCGGTGCTACT</td>
<td>S-S-Rint-1102-a-A-18</td>
<td>(79)</td>
<td>Butyrate producer</td>
</tr>
<tr>
<td>Veil 223</td>
<td>AGACGAATCTCCCTCTT</td>
<td>S-S-Veil-0223-a-A-18</td>
<td>(81)</td>
<td>0–4.5% (81)</td>
</tr>
<tr>
<td>NON 338</td>
<td>ACATTCATCGGGAGGC</td>
<td>Not applicable</td>
<td>(85)</td>
<td>Negative control</td>
</tr>
</tbody>
</table>

1Standardized nomenclature of oligonucleotide probes (86).

2Percentages are relative to the total cell count that was also detected by FISH.
tides to the hybridization mixture. These unlabelled molecules are directed against flanking regions of the target. By binding to a nearby region of the target site, they can help to dissociate RNA secondary structures or RNA-protein binding and thus allow a better access of the labelled probe to its target. On the other hand, false-positive binding to non-target regions has to be avoided. In this case, the stringency of the hybridization has to be increased by the addition of formamide to the hybridization buffer or by increasing the hybridization temperature. Unlabelled competitor probes can be used specifically for species that are known to give a false-positive signal. The sequence of the competitor probe has a higher specificity to the rRNA of the false positive organisms than the labelled probe and therefore occupies the target site without giving a fluorescence signal.

To find optimal hybridization conditions that afford high specificity and signal intensity often means to find a compromise between these two effects. It should be realized that in most cases only close relatives are included in specificity testing. Therefore it cannot be excluded that a given probe sequence finds a target region that was not predicted by the previous database analyses.

**FISH microscopy**

In fluorescence microscopy, the detection limit of labelled cells is mainly determined by the magnification of the microscope, the dilution by the fixation and the number of microscopic fields analysed. Usually this limit lies at approximately $10^6$ cells per gram of sample (25).

Manual microscopic enumeration is a very tedious and time-consuming task. For studies involving the analysis of large numbers of samples and probes, the use of an automated enumeration method is highly recommended. Several systems suitable for the fully automatic detection and enumeration of FISH signals have been described. In all cases, a computer-controlled microscope is used to change the position of every microscopic field, wells and slides. A cooled, charge-coupled device (CCD) camera captures images of the microscopic fields that are sent to the analysis software for signal detection and counting. The most important advantage of an automated enumeration system is the speed of analysis, but additionally the resulting systematic error no longer varies from person to person (26). In 1999, Jansen and co-workers developed an automated system for the detection and enumeration of faecal bacteria (27). They were able to analyse 1200 images within 20 h. At 25 images per sample, this corresponded to 48 samples. Other automated systems have been described for marine (18) and oral microbiota (28, 29).

An important task for the automated system is focusing. Jansen et al. used a stack of images, which were taken at a distance of 1 μm in z-axis (on top of each other at the same microscopic field), to calculate the focus position (27). Pernthaler et al. used a software autofocusing routine, which was performed in fluorescence mode at each position before the image was captured. When this focusing routine failed three times in a row, a bright-field autofocus was performed. As the samples were applied to membrane filters, the focusing routine was now able to use the high contrast of the pores of the membrane filters to restore the focus, and subsequently continue in fluorescence mode (18).

When the captured image is stored in the computer system, the analysis software has to perform a series of operations to correctly detect and count bacterial signals. Initially, a threshold-based routine classifies every pixel of the image into signal or background, resulting in a binary image, containing only black pixels as a background and white pixels as potential signals. Positive signals from the faecal bacteria form areas of white pixels with roughly the same shape and size as the bacteria. With the naked eye, fluorescent signals of bacterial cells can be distinguished easily from artefacts such as autofluorescing plant cells and mineral particles or clumps of fluorescent dyes by virtue of their shape and fluorescence intensity. For the computer software, specific parameters have to be defined that are able to separate signals from artefacts. The most important parameter is the size of a fluorescent object. If the targeted cells are cocci with a projected size of 1 μm², for example, objects exceeding this size can be rejected. Moreover, it has to be defined which range of signal intensity an object may have relative to the background. Finally, these signals have to be counted and calculated back into cell numbers by multiplying them by the preparative dilution factors and the microscopic magnification factor.

**Enumeration of fluorescently labelled bacterial cells by flow cytometry**

Flow cytometry (FCM) is a commonly used method in medical diagnostics and immunology and the enumeration of blood and other cells is a predominant application. This high-throughput method has also been used in environmental microbiology. In 1990, Amann et al. (30) presented a combination of FISH with FCM for the analysis of defined mixtures of bacteria for the first time. Since then, FCM has been applied to the analysis of various microbial ecosystems (31–33).

In FCM, cells (Fig. 4) are passed through a capillary, where they are detected in a focused light beam. Using a differential pressure system, it is possible to produce a laminar flow effect which causes the sample fluid to flow in a central core and prevents mixing with the sheath fluid. Several detection methods are applied to bacterial cells: forward (FSC), 90° light scatter (SSC) and fluorescence emission (FL) at different wavelengths. The FCS detection indicates the cell size, while SSC gives a measure of the granularity. Plots of FSC versus FL or of SSC versus FL...
facilitate the distinction of different microbial populations in a given sample.

In principle, the same probes used for the microscopic detection of fluorescent cells can also be used for detection by FCM. In addition to the specific probe for the detection of the target organisms, the probe Eub338 is used to detect all bacteria and Non338 (Table IV) is used to detect unspecific background binding (31, 34). All preparation steps including permeabilization, hybridization and washing can be performed on standard 96-well plates. An autosampler also decreases the time necessary for sample analysis. The detection limit of the flow cytometric approach \(2 \times 10^8\) cells per gram of faeces) is somewhat higher than that of the microscopy-based approach \(10^6\) per gram of faeces).

In order to compare the two methods, Zoetendal and co-workers enumerated *Ruminococcus obeum*-like bacteria using FISH-FCM and FISH microscopy (32). Faecal samples obtained from three individuals were analysed over a period of 4 weeks using the probes Urobe63 (targeting uncultured *R. obeum*-like bacteria) and Erec 482 (targeting the *Clostridium cocoides-Eubacterium rectale* group). The numbers obtained with the two methods were similar (on average 2.5% for *R. obeum*-like bacteria and 16% for the *C. cocoides-E. rectale* group).

FISH-FCM is a powerful tool for fast, high-throughput enumeration of faecal bacteria. It has already been applied to large studies such as ‘Microbe diagnostics’ and ‘Crownalife’ (QLK1-2000-00108 and QLK1-2000-00067), in which over 400 samples are being analysed with approximately 20 oligonucleotide probes. FCM also allows cell sorting and thereby the enrichment of a defined bacterial population by up to 280-fold which can be used for subsequent investigations (35, 36).

**QUANTITATIVE DOT BLOT**

Quantitative dot blot hybridization was originally introduced for investigating bacterial diversity (16). This method requires the isolation of rRNA from environmental samples. The rRNA is immobilized on a nylon membrane and hybridized with \(^{32}\text{P}\) or fluorescently labelled probes. The signal intensity of each spot can be quantified. Dot blot hybridization has been widely used to determine the specificity of new oligonucleotide probes and the optimal hybridization conditions including hybridization temperature and formamide concentration (24, 37, 38). This method has also been used for a comparison of bacterial groups in caecal and faecal samples (39, 40). Using rRNA dot blot hybridization and FISH combined with FCM, Rigottier-Gois and co-workers (34) analysed the composition of...
faecal samples from 23 individuals. There were no statistically significant differences between the results obtained with the two methods for the groups of *C. coccoides* (mean 22%), *F. prausnitzii* (11.3%), *Bifidobacterium* spp. (3.9%) and enterobacteria (2.8%). However, differences between both methods for members of the *Bacteroides* (41.7% dot blot, 9.1% FCM) and *Atopobium* (0.3% dot blot, 2.8% FCM) groups were observed. These differences can be explained by the principal differences between the two methods – the rRNA dot blot hybridization reflects the status of metabolic activity (41, 42), while the FISH method enumerates the target cells.

**MICRO-ARRAYS**

Micro-array technology allows the parallel analysis of RNA and DNA of thousands of genes, or of the same gene from thousands of organisms, in a single experiment. In microbiology, micro-arrays were initially used for exploring transcriptional profiles and genome differences for a variety of microorganisms (43). Several recent publications described micro-array systems for analysing the diversity of bacterial communities based on the 16S rRNA gene or other functional genes (44–46).

Figure 5 shows the principle of the micro-array technique. Total DNA or RNA is isolated from an environmental sample and fragmented. The fragments are subsequently amplified by PCR and simultaneously labelled by the use of labelled nucleotides (in the following referred to as DNA arrays) or the fragments are directly labelled chemically (in the following referred to as RNA arrays) and hybridized to oligonucleotide probes immobilized on a glass or a membrane surface. After a washing step, a probe–target duplex can be visualized using a fluorescence scanner.

The stringency of the hybridization is the key for the analysis. The detection of a single species in the presence of a mostly unknown genetic background is difficult, because a single mismatch may already lead to incorrect results (47). However, it is difficult to find a set of conditions suitable for all probes (48). Therefore, thermal dissociation analysis of all hybrids performed on a single micro-array can help to better distinguish between matched and mismatched probe–target duplexes (48). Melting (*T_d* – dissociation temperature) profiles are collected by simultaneously increasing the array temperature and measuring the fluorescence intensity of each spot. Since the effect of position and type of a mismatch on *T_d* of hybrids is non-linear, standard statistical methods are not suitable for data set analysis (49).

The common strategy, where DNA is amplified and labelled during PCR (DNA array), is a simple highly sensitive method. The usefulness of DNA micro-arrays in microbial ecology has been demonstrated for sulfate-reducing bacteria (50), cyanobacteria (51) and faecal bacteria (52). However, amplification steps often introduce biases that lead to false results (13, 48). Therefore, another strategy employs labelled rRNA, which is hybridized to immobilized DNA capture probes (RNA arrays) (53). Since the detection of RNA arrays is less sensitive, it is only possible to detect species present in relatively high numbers. Therefore, an approach was presented by Small et al. (47), who improved the specificity of the hybridization and the detection limit by using a chaperone-detector probe strategy. The principle of this approach is the application of a labelled detector probe that binds to a target near the capture region. It stabilizes the target RNA-capture hybrid and thus leads to higher signal intensity and specificity. The detection limit reported for this strategy was 0.5 μg of total RNA, representing approximately 7.5 × 10^6 cells, a value that is similar to that reported for DNA micro-arrays (47).

**DGGE/TGGE**

In 1993, Muyzer et al. (54) introduced denaturing gradient gel electrophoresis (DGGE) for the analysis of the diversity
of complex microbial ecosystems. In this method (Fig. 6), PCR-amplified fragments of 16S rDNA are separated in polyacrylamide gels, containing a gradient of denaturing agents (urea or formamide). Heteroduplexes of different amplicons (with different G/C content) are dissociating at different positions in the denaturing gradient, resulting in a hold of migration. The result is a pattern of bands, which is characteristic of the bacterial community present in the sample. In principle, each band represents one species. In a similar method – temperature gradient gel electrophoresis (TGGE), DNA fragments are separated in a temperature gradient instead of a denaturing gradient. Using universal 16S rDNA primers, the estimated detection limit is approximately 1% of the total bacterial content (54, 55). Fine analysis of bacterial diversity using T/DGGE is possible, but some modifications are required. Using group-specific primers, it is possible to analyse bacterial communities at a higher resolution and a lower limit of detection. In some cases, the blotting of DNA fragments separated by T/DGGE on a nylon membrane and hybridization using specific probes can be helpful. For correct identification of bacteria presented on T/DGGE gels, sequencing of the bands is required.

T/DGGE methods are mostly used for the investigation of the dynamics of bacterial compositions in various ecosystems (13, 56). Zoetendal and colleagues analysed faecal samples obtained from 16 healthy individuals by TGGE (55). Two of the individuals were investigated over a period of 6 months. They demonstrated that the TGGE band pattern is specific for each individual. However, some of the bands were present in all faecal samples. Patterns of the bands analysed in individuals over a period of 6 months were highly constant.

In the ‘Microbe Diagnostics’ project, DGGE was used for assessing the intestinal microbiota profiles of ulcerative colitis (UC) patients. Faecal samples from 33 volunteers in an active state of disease were analysed by DGGE. The profiles show that the total bacterial community is complex and differences between the patients can be observed (Fig. 7). These observations of high individual variability are similar to previous observations resulting from the analyses of faecal bacterial communities from healthy volunteers. Although each UC patient has a unique microbiota composition, there were also some common bands in a number of DGGE profiles. Preliminary sequencing results (data not shown) indicate that diseased subjects investigated in two study centres in Europe (Barcelona, Spain and Cork, Ireland) share some bacteria that have not yet been detected in faeces from healthy subjects.

**Fig. 6.** Principle of the T/DGGE technique. Fragments of 16S rRNA genes (usually 300–500 bp) are amplified. One of the primers used in the PCR reaction has a GC-clamp (GC-rich 5’ end), which prevents complete denaturation of the heteroduplexes separated in a gradient gel. Different amplicons migrate to different positions (so-called melting domains), where the denaturing conditions are specific for the nucleotide sequence.

**Fig. 7.** Total bacterial DGGE profiles from 33 UC patients (33 faecal samples used in this study were collected as part of the PROGID project – QLK1-2000-108). The subjects who donated faecal samples had active ulcerative colitis and they originated from two geographic locations: (a) Cork, Ireland (13 samples) and (b) Barcelona, Spain (20 samples). DNA isolation, PCR and DGGE analysis of the V6–V8 region of the 16S rDNA were performed as described previously (55, 69).
IN SITU MONITORING OF GENE EXPRESSION

There is an urgent need not only to be able to identify the individual bacterial constituents of the human intestinal microbiota, but also to assess the metabolic activity of the bacteria in the gut. One approach to meet this goal is directed at monitoring gene expression at the cellular level. The degree of expression of an enzyme is usually correlated with its activity. This leads to the assumption that the level of mRNA is a measure for the activity of the corresponding gene.

The number of mRNA molecules in an individual cell usually is not high enough to allow their direct detection by in situ hybridization analogously to the well-established FISH technique (57). In order to circumvent this problem a number of studies employed reverse transcription and PCR (58) to amplify intracellular mRNA. Through various detection schemes cell-associated PCR products have been visualized in different applications as reporters for the presence of specific genes or microbial activity (59–65).

The visualization of mRNA has already been successfully demonstrated in single Salmonella cells (66, 67). Attempts to transfer the developed methodology from the Salmonella system to a bacterial species more relevant for the gut microbial ecosystem focused on bifidobacteria.

With bifidobacteria as a representative genus of the intestinal microbiota the development of an in situ PCR-based method for monitoring mRNA was initiated. The most important step in a procedure for intracellular PCR is the permeabilization of cells. The cell-wall structure needs to be penetrated to allow the components of the PCR reaction mixture (i.e. primers, DNA-polymerase and deoxyribonucleotides) to enter the interior of the cell. However, the permeabilization must not be too excessive as it may result in cell lysis and/or in diffusion of the intracellularly generated PCR products out of the cell. Hence, optimal permeabilization conditions have to be worked out. Extensive work has been devoted to optimizing permeabilization conditions for different Bifidobacterium spp.

Using an rRNA-based FISH assay based on detecting a biotinylated rRNA-targeting oligonucleotide probe and visualizing the biotin moiety by fluorophore-labelled streptavidin (Mw = 53 kDa) the optimum permeabilization conditions were assessed. A large number of different kinds of treatments including lysozyme, proteinase K, mutanolysin, CTAB (cetyl trimethyl ammonium bromide) and combinations of these have been tested on bifidobacteria fixed with either paraformaldehyde or ethanol. These tests led to the identification of a combined treatment with lysozyme and mutanolysin that rendered B. longum permeable. However, to illustrate the complexity in determining the optimal permeabilization conditions, the same treatment did not enable the uptake of large molecules in B. bifidum to the same extent.

In spite of these difficulties, the objective of visualizing metabolic activity in individual bacteria has not been abandoned. A method has been developed to monitor β-galactosidase activity at the single-cell level (56, 68). The procedure has successfully been implemented using the X-gal substrate (producing a coloured precipitate) and C12-fluorescein-digalactoside substrate (C12-FDG) (producing a fluorescent compound). Together with population-level techniques, activity staining of native PAGE and iso-electric focusing of crude protein lysates from cultures of bifidobacteria as well as DNA micro-array-based gene expression analyses, the single-cell β-galactosidase activity method is used to decipher regulatory patterns of β-galactosidase expression in different types of bifidobacteria.

CONCLUSIONS

The days when scientists spent weeks with Petri dishes trying to isolate, purify and characterize the phenotypes of new species of bacteria are history. Today, as a result of molecular biological methods the dominant human gut bacteria have been identified. However, many faecal bacterial species present in small numbers are still unexplored. Moreover, many species have not yet been cultured and, hence, their specific role in the ecosystem is still unknown. In spite of these shortcomings, the developed methods are valuable tools for the more detailed analysis of bacterial communities. Unfortunately, their widespread use is still limited by several factors including insufficient automation of sample preparation and high costs. Nonetheless these methods have been increasingly applied to studies exploring the role of diet in influencing the composition and activity of the human gut microbiota.

REFERENCES


