Applications of Phylogenetic Microarrays to Profiling of Human Microbiomes

11

Oleg Paliy and Vijay Shankar

Abstract

Human-associated microbial communities are known to be highly diverse, comprising between hundreds and a thousand species, depending on the body area. The sheer numbers of species as well as the fastidious nature of most of these organisms make culture-based techniques both inefficient and challenging to study these communities. As a result, analyses of such communities are best accomplished by the use of high-throughput molecular methods such as phylogenetic microarrays and next generation sequencing. Phylogenetic microarrays have recently become a popular tool for the compositional analysis of complex microbial communities, owing to their ability to provide simultaneous quantitative data for many community members. This chapter will focus on currently available phylogenetic microarrays for the interrogation of human-associated microbiota, the technologies used to construct the arrays, as well as several key features that distinguish them from other approaches. We will also discuss optimization strategies for the development and usage of phylogenetic microarrays as well as opportunities to complement microarray analysis with other techniques.

Introduction

Microbes inhabit diverse environments. Some of these environments include the human intestinal tract and skin, soil, roots, leaf and bark surfaces of plants, ocean waters, deep see vents, and air. The ecosystems of such environments are populated by communities of microorganisms, rather than by individual species, and often contain hundreds and even thousands of different microbial members. Many of these communities play pivotal roles in ecosystem processes such as energy flow, elemental cycling, and biomass production. Energy and nutrients in these systems are processed by intricate networks of metabolic pathways through multiple community members (Belenguer *et al.*, 2006; De Vuyst and Leroy, 2011; Duncan *et al.*, 2004; Flint *et al.*, 2008). The sheer complexity of such networks and the difficulty involved in culturing the members of these communities have challenged researchers who have tried to gain a clearer understanding of these interactions.

Significant progress has been achieved recently in the study of microbial communities that inhabit the various niches of the human body. These microbial consortia, usually referred collectively as the human microbiota, have been estimated to harbour microbial cells ten times the number of human cells in the body, along with a cumulative microbial gene count that is approximately 100-fold larger than the human genome (Gill *et al.*, 2006). Every exposed surface of the human body is colonized by a unique microbiota, including the mouth, skin, respiratory, gastrointestinal, and genitourinary tracts (Costello *et al.*, 2009). Renewed interest in the human microbiota is associated with the recognition of the important relationships these microbes form with our bodies. For example, microbiota of the gut participate in host energy metabolism by breaking down complex polysaccharides in the diet. Human associated microbes also protect the host from pathogen invasion through competition for resources or directly by inhibiting pathogen growth. Moreover, microbiota modulate the proper development and

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

functioning of the human immune system and help maintain epithelial homeostasis. At the same time, microbiota dysbiosis, defined as the perturbation of the normal microbial profile, has been linked to a number of human diseases including dental plaque, bacterial vaginosis, psoriasis, atopic dermatitis, inflammatory bowel disease, obesity, and colon cancer (Gao *et al*., 2008; Grice and Segre, 2011; Larsen and Monif, 2001; Neish, 2009; Sartor, 2008).

Recent advancements in molecular technologies have significantly simplified the analysis of these microbial communities because they remove the need to culture and grow community members individually. Some of the currently available molecular techniques include high-throughput sequencing, terminal restriction fragment length polymorphism, chequerboard DNA–DNA hybridization, quantitative real-time PCR, fluorescent *in situ* hybridization, and phylogenetic microarrays. Phylogenetic interrogation of small subunit (SSU) rRNA molecules using these techniques has led to considerable progress in our understanding of community structure and dynamics of various microbial ecosystems (Sekirov *et al*., 2010; Suau, 2003). Phylogenetic microarrays, one of the more popular choices among these techniques, have been successfully used to quantitatively profile a variety of human–associated microbial communities, including the gastrointestinal tract, oral cavity, and vaginal canal.

Although gene expression analysis was the original motivation behind the development of microarrays, their versatility has allowed researchers to adapt this technology for other uses, including phylogenetic analysis. Several types of microarrays have been developed to characterize the composition and function of microbial communities, including community genome arrays, functional gene arrays, and phylogenetic microarrays. Community genome arrays are constructed using whole-genomic DNA isolated from species in pure culture. They allow detection of individual species and strains in simple and complex communities. Functional gene arrays include probes to genes that encode important enzymes involved in various metabolic processes and are useful for monitoring physiological changes in microbial communities. A good example of a functional gene array is the GeoChip, which is described in more detail in previous chapters of this book. Phylogenetic microarrays (phyloarrays) contain probes complementary to ubiquitous gene sequences (usually the small subunit rRNA gene) and are primarily used for the analysis of microbial community composition and variability. Among different array types applied to the interrogation of human-associated microbial communities, phyloarrays are currently the most popular owing to the availability of a large set of near-full length SSU rRNA sequences deposited in NCBI, EMBL, RDP, and Greengenes databases.

The first recognized phylogenetic microarray, developed by Guschin *et al*. (1997), was capable of detecting select genera of nitrifying bacteria. One of the first microarrays developed for the analysis of human-associated microbiota communities was described by Wang *et al*. (2002) and was able to recognize 20 different species of gut microbiota. Since then, significant advancements in the breadth of detection (total number of different groups detected) have been achieved with phylogenetic microarrays. Progress has also been made to increase the sensitivity and specificity of phylogenetic microarrays. In this chapter, we will discuss the current developments in the technology, optimization of usage, applications, and potential future trends in the use of phylogenetic microarrays.

Current phylogenetic microarrays

The high-throughput and quantitative nature of phylogenetic microarrays makes them an excellent solution for researchers who seek to determine the composition of their microbial community of interest. Some key features that distinguish different phylogenetic microarrays are the choices of phylogenetic markers utilized for probe design and the experimental platform used to host these probes. A gene or a group of genes that are ubiquitously present among all or at least majority of species of interest often make the best target for phylogenetic analysis. A few already utilized examples that fit the above criteria include the small ribosomal subunit RNA gene (16S in prokaryotes and 18S in eukaryotes), the

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

large ribosomal subunit RNA gene (23S and 28S, respectively), genes coding for the heat shock proteins GroEL and GroES and for ribosomal proteins such as protein S1 (Martens *et al.*, 2007), and in the case of methanogens, the *mcrA* gene which encodes for methyl coenzyme-M reductase (Luton *et al.*, 2002). The SSU rRNA gene is currently the most popular choice in part because it can be fully and selectively amplified from total genomic DNA with a set of primers complementary to the conserved regions at the beginning and the end of the gene (Fig. 11.1).

In recent years, significant improvements have been achieved in the design of phylogenetic microarrays, including improvements in the breadth of detection, sensitivity, and specificity. Table 11.1 lists some of the currently available phylogenetic microarrays including the employed technology and targeted communities. The original phylogenetic microarray designed by Guschin *et al.* (1997) was capable of detecting a few genera of nitrifying bacteria. As we mentioned previously, one of the first microarrays developed for the analysis of human-associated microbiota communities was described by Wang *et al.* (2002). The initial array was able to recognize 20 different species of gut microbiota, and was later updated to include probes to 40 predominant members of human gut microbiota (Wang *et al.*, 2004). The current leader in the total number of potentially detectable groups, the third generation (G3) PhyloChip array, has been designed to detect as many prokaryotic phylotypes as possible (Brodie *et al.*, 2006; Hazen *et al.*, 2010). This microarray is based on the Affymetrix GeneChip technology and contains 1.1 million 25-mer probes with an approximate probe density of 10,000 molecules per μm^2. The array is capable of detecting approximately 50,000 phylotypes; the previous version of the array, G2, contained 500,000 probes and was able to detect approximately 9000 phylotypes. This increase in the breadth of detection allows for a wide range of applications, evidenced by the recent use of PhyloChip in profiling not only human-associated microbial communities but also microbes of soil, coastal salt marsh, and coral reef (Cox *et al.*, 2010a; Deangelis *et al.*, 2011; Lemon *et al.*, 2010; Mendes *et al.*, 2011).

Another broad detection microarray was developed by Palmer *et al.* (2007) This array, based on the Agilent microarray construction technology, is capable of detecting nearly 1600 bacterial and archaeal species from a variety of environments and niches. The array contains 10,500 40-nt probes that include 9121 microbial taxonomic group specific probes and 1379 control probes. The oligonucleotide probes are tethered to the array surface using a 10-nt poly-T linker. This microarray was used to profile faecal microbial communities of human infants (Palmer *et al.*, 2007).

A number of phylogenetic microarrays have been constructed specifically to assess microbial communities associated with humans. The Microbiota Array, also based on the Affymetrix photolithographic technology, was designed to profile microbiota of the human gastrointestinal tract. The array contains 16,223 probes, with multiple probe sets allowing detection and quantification of 775 different microbial phylotypes from the human intestine. Each probe set detects a single phylotype (also called operational taxonomic unit or phylogenetic species) and contains between 5 and 11 different probes to that phylotype's 16S rRNA sequences. The design of the probe sequences allows an interrogation of either genomic DNA or total RNA isolated from the gut microbial communities. The Microbiota Array also takes advantage of the Affymetrix microarray construction design to contain both perfect match

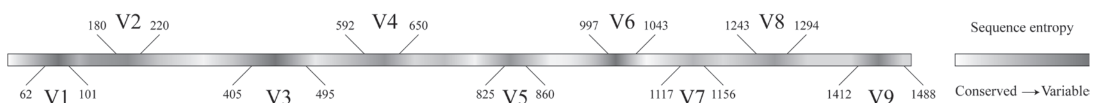

Figure 11.1 Sequence conservation and variability of 16S ribosomal RNA gene in prokaryotes. Sequence entropy is displayed using a gradient scale as shown in the legend. Positions of the variable regions (V1–V9, nucleotide positions are displayed for *Escherichia coli* 16S rRNA sequence) and sequence entropy values are based on the information from Ashelford *et al*. (2005).

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

Table 11.1 A selection of current phylogenetic microarrays

Array name	Target community	Resolution	Technology	Detectable groups	Reference
PhyloChip	All bacteria	Varied Species	Photolithography	9,000 phylotypes (G2) 50,000 phylotypes (G3)	Brodie *et al*. (2006), Hazen *et al*. (2010)
Custom array	All bacteria	Species	In situ chemical synthesis	1629 phylotypes	Palmer *et al*. (2007)
Microbiota Array	Human intestinal biota	Species	Photolithography	775 phylotypes	Paliy *et al*. (2009)
HITChip	Human intestinal biota	Species	In situ chemical synthesis	1140 phylotypes	Rajilic-Stojanovic *et al*. (2009)
AUS-HIT Chip	Human intestinal biota	Species	Electrochemistry	739 phylotypes	Kang *et al*. (2010)
Custom array	Human intestinal biota	Genus	Spotted array	310 genera	Manges *et al*. (2010)
HOMIM	Human oral biota	Species	Spotted array	272 phylotypes	Preza *et al*. (2009)
OC Chip	Human oral biota	Varied	Spotted array	350 groups	Crielaard *et al*. (2011)
V-Chip	Human vaginal biota	Varied	Spotted array	350 groups	Dols *et al*. (2011)

probes (provide target quantification) as well as mismatch probes (estimate cross-hybridization amount removed during normalization of probe signals) for each interrogated phylotype. This phyloarray has high sensitivity and can detect phylotypes that are present at an overall community abundance of less than 0.001% (Paliy *et al.*, 2009). To date the Microbiota Array has been used successfully to accurately profile the microbial communities of the distal gut in healthy and *Clostridium difficile* infected adults, healthy adolescents, and adolescents diagnosed with irritable bowel syndrome (Agans *et al.*, 2011; Rigsbee *et al.*, 2012).

The HITChip, based on Agilent technology, was also designed to profile the microbial communities of the human gastrointestinal tract. This glass slide based array consists of 4809 probes and is capable of detecting 1140 intestinal phylotypes (Rajilic-Stojanovic *et al.*, 2009). Each phylotype is represented on the array by 4–6 probes that are each 24-nt long. The antisense oligonucleotide probes on the array were selected to match sequences from two hypervariable regions of the 16S rRNA gene (V1 and V6). The probes are anchored onto the array surface using a 10-nt long poly-T spacer at the 3′ end of each probe, which alleviates steric hindrance during the hybridization process (see below). The ability to update array design with relative ease and the presence of two replicate arrays per glass slide provides flexibility and cost efficiency. The HITChip has been successfully used in a number of studies to profile microbial community structure of the gastrointestinal tract (Biagi *et al.*, 2010; Jalanka-Tuovinen *et al.*, 2011).

Two other phyloarrays for the analysis of human gut microbiota are currently available. The first, named the Aus-HIT Chip, is capable of detecting 739 microbial phylotypes and contains 2243 oligonucleotide probes (Kang *et al.*, 2010). In order to increase microarray sensitivity and specificity, a special strategy was used during probe design. Two separate 17- to 18-nt-long probes specific to two regions of the same target were bound together using a 4–6 nt long linker. This composite probe represents a specific phylotype. General probe design was accomplished using the GoArray software and the oligonucleotide probes were synthesized *in situ* on the array surface. Each phylotype is represented by three replicate probes that are randomly distributed on the array surface. The Aus-HIT Chip was used to assess microbial dysbiosis in patients with Crohn's disease (Kang

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

et al., 2010). The other array designed to profile the human microbiota from faecal samples was developed by Manges *et al*. (2010). This glass slide array contains 1,412 microbial group specific probes and 13 control probes. It is capable of detecting 310 microbial genera distributed across 128 families and 20 phyla. Each probe was spotted twice on the array surface for the purpose of signal validation and technical consistency. The oligonucleotide probes in this array are designed to be hybridized to RNA molecules (instead of DNA) derived from the transcription of PCR amplicons of the 16S rRNA gene. This array was successfully used to profile human faecal microbiota of patients who were at-risk for nosocomial *Clostridium difficile* associated disease (Manges *et al*., 2010).

The HOMIM (Human Oral Microbial Identification Microarray), an aldehyde-coated glass-slide microarray, was designed to detect 272 microbial phylotypes from human oral cavity through the interrogation of the 16S rRNA gene. The reverse capture probes in this array consist of 18–20 nucleotides complementary to the target sequence with a spacer sequence of eight thymidines and a 5′-(C6)-amine-modified base for attachment to the slide. The oligonucleotide probes are printed onto a 25 mm × 76 mm aldehyde slide. Each array is separated into five sections to facilitate the parallel processing of five samples, making the overall process more cost effective. This array has been an effective tool in detecting and profiling the oral microbiota in multiple studies, spanning several disease states as well as examining oral microbiota in healthy hosts (Docktor *et al*., 2012; Luo *et al*., 2012; Preza *et al*., 2009a). Another oral microbiota targeting phyloarray, OC chip, was developed by Crielaard *et al*. (2011). The array contains 350 probes that are 20–22 nucleotides in length. This oligonucleotide array is capable of detecting 350 microbial groups from human oral cavity and has been successfully used to validate oral microbial profiles of children with dentition obtained through next-generation sequencing technologies (Crielaard *et al*., 2011).

The V-Chip, also called the vaginal microbiota-representing microarray, is another spotted microarray that utilizes polymer-coated slides to house oligonucleotide probes. The array is constructed by employing a high precision robotic dispenser with fine-point quill pins to deliver oligonucleotide probes onto a slide surface. The probes contain a 5′-NH_2-C6 terminal region that is used in the probe attachment. The array surface is coated with a proprietary activated polymer that is responsible for the binding of the probes to the array. The V-Chip array contains a total of 459 probes allowing for the detection of 350 vaginal microbial groups that are spread across multiple taxonomic levels (from species to order level). This phylogenetic microarray was designed to profile human vaginal microbiota, and has demonstrated its effectiveness as a diagnostic tool for profiling changes in microbial communities in diseased states such as bacterial vaginosis (Dols *et al*., 2011).

There are several features to take into account when comparing different phylogenetic microarrays. As seen in Table 11.1, microarrays differ in the technology used. The Microbiota Array and the PhyloChip were developed using photolithographic synthesis, which has several advantages including the high degree of efficiency, uniformity, and probe density. The Affymetrix platform takes advantage of the high probe density to allow these arrays to contain multiple probes per target (phylotype) as well as to enable allocation of mismatch probes to each phylotype that provide an opportunity to adjust for target cross-hybridization. On the other hand, ink-jet and fine-point needle printing on glass slides allow for cost-effective production and modification of microarrays since expensive tools such as photolithographic masks are not required. Printing on glass slides is still considered the most cost-efficient method currently available. However, the drawback of this type of array manufacturing is the loss of uniformity; therefore, these arrays require more extensive validation tests before they are ready for application.

Phylogenetic microarrays are also distinguished based on their resolution. In order to achieve the degree of resolution seen with Sanger sequencing, a species or OTU (operational taxonomic unit) level specificity is required. Profiling communities at this depth allows researchers to understand species-level interactions such as metabolic interdependencies and co-pathogenicity.

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

Several of the currently available microarrays are capable of profiling microbial communities at the phylotype level (Table 11.1). Breadth of detection is yet another variable that differentiates phylogenetic microarrays. The PhyloChip is an excellent example of a phyloarray specifically designed to detect as many microbial phylotypes as possible across the bacterial and archaeal domains. Its detection breadth makes this phyloarray very versatile, enabling its usage in many environmental and clinical studies. The downside to this type of design strategy is a potential for the high number of false positives due to off-target hybridizations induced by the high number of probes. The issue of false positives and cross-hybridization can be ameliorated by optimizing the probe selection process and by assigning strict criteria for signal presence, though a complete resolution of the problem is very difficult. Opposite to such design, phylogenetic microarrays designed for the interrogation of specific communities, such as the Microbiota Array and HOMIM, benefit from the reduced cross-hybridization potential to provide robust estimates of community structure, while maintaining the ability to discriminate different communities with similar efficiency.

Phylogenetic microarrays based on non-traditional techniques have also been described in several reports. For example, fragment ligation reaction based DNA microarray has been developed by Candela *et al.* (2010). The microarray design involves the use of pairs of oligonucleotides complementary to the adjacent regions of each target sequence. One of the oligonucleotides contains a 5′-fluorescent label and the other has a unique 'zip-code' sequence. The oligonucleotide pair is ligated together only in the presence of the complementary target sequence binding to both oligos. Since the ligation is carried out by highly selective ligase enzyme, a high level of probe specificity can be achieved with the use of this approach. The quantification of the fluorescently-labelled ligated products is accomplished by the use of specially designed 'universal' detection array that houses probes complementary to the tag ('zip-code') sequences present within the ligated products. These universal arrays allow for uniform hybridization conditions and for the use of different ligation probe sets unique to each interrogated community, which enables flexible experimental design. A prototype ligation array developed by Candela and co-workers was capable of quantifying 30 groups of human intestinal microbiota, and the array was used to profile the faecal microbiota of several young adults (Candela *et al.*, 2010). A large subunit ribosomal RNA gene based phylogenetic microarrays have also been developed successfully (Mitterer *et al.*, 2004; Yoo *et al.*, 2009). For example, Mitterer *et al.* (2004) developed a custom glass-slide array that contained genus- and species-specific solid phase primers targeting a single variable region of the 23S rRNA gene. Using universal primers, genomic DNA from environmental samples was subjected to PCR amplification on the glass-slide. The generated PCR products were allowed to bind to the group-specific primers for subsequent elongation accompanied by the incorporation of biotin labelled nucleotides. Quantification was based on fluorescence scanning of the hybridized probe–target pairs. This array was successfully used to identify at a high-resolution bacterial communities in cervical swab samples (Mitterer *et al.*, 2004).

Phylogenetic microarray optimization

Phylogenetic microarrays provide several advantages over some of the other currently available techniques used to profile microbial communities. These advantages include a highly quantitative nature of the acquired data, an ability to analyse one sample at a time, a short processing time, and an opportunity for multi-probe interrogation of each community member. Phylogenetic microarrays can be used to identify taxa that vary in abundance by over five orders of magnitude (Roh *et al.*, 2010). Although these attractive features make phylogenetic microarrays a viable option for phylogenetic analysis, there are also some limitations to the technology that must be addressed. Firstly, phylogenetic microarrays typically do not allow for the detection of novel phylotypes. They are only capable of detecting and quantitating phylotypes to which they contain probes. Secondly, microarrays are technically demanding to design, use, and analyse, and thus require rigorous

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

testing, validation, and optimization (Hashsham *et al*., 2004). To help with the second limitation, a number of approaches that improve the robustness of microarray data have been developed and are discussed below.

Optimization of probe design and hybridization

A typical design process for a microarray specific to a particular ecosystem or community usually involves the acquisition of 16S rRNA genes from members of that community (through clone library sequencing, for example) and subsequent selection of regions within the genes for probe design. Region selection can either be done manually, based on the availability of unique fragments in the hypervariable regions of 16S rRNA sequence, or by using mathematical algorithms. Several software solutions such as ARB, GoArray, and PhylArray exist to facilitate this process and provide an optimized automated design of microarray probes. A lack of a rigorous probe selection process can lead to issues such as high level of fragment cross-hybridization, which can result in inaccurate or biased community profiles.

There are several variables that control the probe–target hybridization process and the subsequent estimation of signal. One such variable, the size of the probe oligonucleotide or DNA fragment, has a large influence on the hybridization behaviour. In general, the length of the probe is positively correlated with hybridization chance (sensitivity) and is negatively correlated with hybridization specificity. Selecting probes that are small can lead to high specificity but at the cost of lower hybridization sensitivity. On the other hand, picking long probes can increase the sensitivity of detection, but risks hybridization of unrelated fragments to each probe. An ideal probe length provides a balance between a high sensitivity and high specificity. While oligonucleotides of lengths between 20 and 30 nucleotides are often selected in many microarray designs, other phyloarrays were successfully designed to contain longer probes (see above).

The melting temperature of each probe-target duplex (T_m) is another important variable that should be taken into consideration when designing probes. Since the hybridization efficiency at any given temperature depends on the sequence T_m, it is important to constrict the melting temperatures of all of the microarray probes to a relatively narrow range. The resulting consistency will reduce probe hybridization bias due to T_m variability, thereby increasing the validity of the acquired signals. While designing probes for phylogenetic microarrays, it is also important to consider the optimal choice of probe targets. Most phylogenetic microarrays use the small subunit ribosomal RNA gene for identification and taxonomic analysis of community members. While much of the 16S rRNA gene sequence is highly conserved, the gene contains nine sections commonly referred to as the 'hypervariable' (V) regions that display considerable sequence variability among different microbes (Fig. 11.1). Phylogenetic studies tend to exploit the variability within these regions for the detection and identification of microbial members within the analysed community. Most hypervariable regions are flanked by conserved sequences, allowing the use of 'universal' primers for the amplification of these regions from most microbial species. The degree of sequence variability varies among different V regions as shown in Fig. 11.1. As a result, the regions differ in their ability to distinguish among microbial phylotypes, and some regions (V3, V6) are slightly better suited to resolve closely related microbial species. This characteristic emphasizes the need for careful consideration of probe target sequence selection within the 16S rRNA gene. For example, community analysis using a microarray with probes to only a single hypervariable region has a potential to introduce a bias in the microbial community profile. It is generally considered a good practice to design probes to multiple hypervariable regions, since such design strategy can adjust for region specific level of variability and any potential hybridization biases.

General strategies for optimizing the design of probes have been previously considered. In the study by Letowski and co-workers (Letowski *et al*., 2004), the authors explored the effects of sequence mismatch on the destabilization of the probe–target hybridization at different fragment GC% and at different temperatures. One of the objectives of the study was to determine an optimal method for designing probes to closely

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

related target sequences. To obtain quantitative results, the authors designed probes that differed in the number and distribution of mismatches. The probe specificities were determined and compared at various hybridization temperatures. The main conclusion of the study was that the greatest destabilization effect was achieved when mismatches were distributed across the entire sequence of the probe. From that observation the authors inferred that in order to achieve optimal specificity when designing probes to closely related sequences, it is important to choose probes with sequence variability spread along the probe length. Conversely, variability concentrated towards the terminal regions of the probes showed greatly reduced specificity and therefore should be avoided. This study also confirmed previous reports of the dependence of the optimal hybridization temperature on the GC% of the probes. In general, optimal specificity was achieved when the hybridization temperature correlated positively with the probe GC%.

Hybridization specificity is also dependent on other parameters such as orientation of the immobilized probe, steric hindrance against binding, and secondary structure formation in target molecules. The influence of these parameters on the hybridization specificity as well as methods to curtail their negative impacts have been introduced and discussed by Peplies *et al.* (2003). Probe orientation was tested using variants of select probes immobilized by either their 5′ or 3′ ends. The hybridization of these probes to their target revealed a higher annealing efficiency for the 3′ immobilized probes. The reduction in the hybridization efficiency of the 5′ immobilized probes was likely due to the occurrence of steric hindrance as the target has to bind the probe with its 3′ end facing the array surface. Note that a potential presence of secondary and tertiary structures in the target molecules can complicate the interpretation of these results. The effects of such steric hindrance can be mitigated by the use in probes of spacer sequence that is positioned between the array surface and the target-specific sequence of the probe. Indeed, Peplies and co-workers determined that there was a linear positive relationship between hybridization signal intensity and the length of the spacer sequence, indicating that longer spacers significantly reduce steric hindrance (Peplies *et al.*, 2003). Lastly, the use of helper oligonucleotides can resolve secondary and tertiary structures of the target molecules. Helper oligonucleotides are unlabelled sequences designed to bind adjacent to the probe's binding site on the target molecule. By binding to the target molecule, the helper oligonucleotides prevent the target molecule from binding to itself, thereby increasing the efficiency of probe–target hybridization. Other optimization strategies such as selective calibration for particular probes to recover false-negatives and improving specificity through signal-limiting parameters can also be applied (Peplies *et al.*, 2003).

Optimization of sample preparation

Methods to improve the experimental procedures for the use of phylogenetic microarrays have been described. A study by Salonen *et al.* (2010) illustrated and compared several methods for the extraction of genomic DNA from faecal samples. Interestingly, the study found that the method used for the extraction of the genomic DNA from environmental samples had an effect on the compositional analysis of the community, and thus it is important to choose an extraction method that accurately reflects the actual community composition as well as provide efficient PCR amplification. This study proposed to use DNA quality, amount extracted, and community composition analysis as criteria for selecting and statistically authenticating an optimal method of genomic DNA extraction. The main conclusion from the comparison of methods was that the repeated bead beating approach to cell breakdown performed significantly better than the other methods, likely because it is generally more universal than alternative enzyme and chemical-based techniques. The bead beating method was capable of uncovering certain groups of microbes such as the methanogenic archaea and some Gram-positive bacteria that remained undetected when other commonly utilized extraction protocols were employed. As an alternative to bead beating protocol, a recently developed pressure cycling technology can be utilized. In this approach, microbial or tissue samples are sealed in high-density tubes and are subjected to repeated rounds of high-low pressure

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

fluctuations. This process not only leads to the breakdown of cells, but can also separate proteins, lipids, and DNA based on their hydrophobicity and ionic properties. Pressure cycling technology was shown to also reduce the effect of PCR inhibitors (see below), presumably because of the separation of the inhibitors and nucleic acids into different phases (Tao *et al.*, 2006).

A study of microbial community composition typically involves subjection of DNA collected from the community to rounds of target gene (such as the 16S rRNA) specific PCR amplification. The goal of this approach is to selectively enrich the DNA pool with the fragments of interest, since 16S rRNA genes, for example, constitute less than 0.5% of total genomic DNA in most microorganisms. In the case of 16S rRNA gene, primers that bind to universally conserved regions at the start and at the end of the gene or flanking one or several variable regions are usually used. Methods such as the phylogenetic microarrays and next-generation sequencing are then employed to determine the composition of the amplified library. It is important to keep in mind that environmental communities are composed of a large number of individual phylotypes with sequence differences in the interrogated target gene. Thus, any PCR amplification of such mixture of sequences is multi-template, and it has potential to introduce a skew in the composition of the amplified PCR library compared to that of the original DNA mixture. Several causes have been proposed to explain this often observed deviation, which include the difference in the template GC% leading to unequal denaturation of template–product pairs during the melting step of PCR reaction, the higher binding efficiency of the GC-rich variants of the degenerate primer mixtures used to amplify fragments, and the re-annealing of high abundance templates during the annealing step that results in the selection against major templates. In addition, carrying out successful PCR reaction is always difficult for the genomic DNA obtained from environmental samples due to the presence of PCR inhibitors extracted during DNA isolation process. Faecal material, for example, contains bile salts and complex polysaccharides that are known to inhibit DNA polymerase activity (Lantz *et al.*, 1997; Monteiro *et al.*, 1997). The problems with PCR inhibitors often necessitate the use of lower amounts of the starting DNA material in the amplification reactions in order to dilute the inhibitor concentration below critical level.

Possible approaches to mitigate such PCR bias have been recently considered by Paliy and Foy (Paliy and Foy, 2011). In this study, mathematical modelling of the multi-template PCR amplification of 16S ribosomal RNA genes as well as detection of the PCR products by phylogenetic microarray was used in conjunction with experimentally determined parameters to define optimal amplification conditions that lead to accurate estimations of phylotype levels. One of the important conclusions from that study was that both the detection and the accuracy of species abundance estimations depended heavily on the number of PCR amplification cycles used. The model predicted that the improvements in the detection and accuracy reached optima between 15 and 20 cycles of PCR amplification. Because of the unequal amplification rate for different templates in the mixture, the accuracy of community composition estimates was negatively affected when DNA was subjected to more than 20 cycles of amplification – at that point gradually increasing PCR bias outpaced any further improvements in phylotype detection. Modelling the presence of PCR inhibitors in the samples showed that the use of more than 50 ng of starting DNA was detrimental to the overall reaction yield and to the accuracy of phylotype detection and abundance estimates. With higher starting amounts, the higher levels of inhibitors caused a significant reduction in the amplification efficiency, and thus more amplification cycles were needed to reach an appropriate reaction yield, which in turn led to a higher PCR bias. Furthermore, the detection and accuracy of phylotype abundance estimates correlated positively with sample-wide PCR amplification rate but related negatively to the sample template-to-template PCR bias and community complexity (Paliy and Foy, 2011). Although this model was developed based on the simulated interrogation of human intestinal microbiota community and subsequent detection by the Microbiota Array, it can be easily modified to simulate the analysis of other communities, other available or novel

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

microarray designs as well as other PCR amplification protocols.

Optimization of data normalization

In order to draw accurate conclusions regarding microbial profiles, raw signal values measured by each microarray have to be normalized and adjusted, so that a valid comparison of signals among multiple samples and arrays can be performed. One goal of such signal normalization is to account for technical variability during sample preparation and microarray hybridization that can lead to systemic variations in measured signals. The objective of normalization is therefore to reduce the technical systemic variability among arrays so that it is easier to discern patterns or changes in microbial profiles across arrays. Many different methods of microarray data normalization have been developed over the years, and these approaches are generally applicable to the analysis of phylogenetic microarray data. The best choice of method often depends on the microarray technology used, the type of study, and the error or systemic variation present in the raw data. An interested reader is encouraged to refer to the study by Choe and colleagues who compared the efficiency of different methods of microarray data normalization (Choe *et al.*, 2005).

Another type of error that is often present in the phylogenetic microarray data is the occurrence of signal due to off-target fragment hybridization, i.e. cross-hybridization. This issue is especially problematic for 16S rRNA gene based phylogenetic analysis because most probes on such microarrays interrogate a single highly conserved molecule, and thus many fragments in the mixture are likely to possess significant sequence similarity, which leads to increased off-target hybridization and cross-hybridization signal. Without an appropriate method to adjust for cross-hybridization, acquiring accurate estimates of community members' abundances becomes challenging. To mitigate such potential cross-hybridization effect, microarrays based on Affymetrix GeneChip design (Microbiota Array, PhyloChip) include a mismatch probe for each interrogating probe. These mismatch probes provide an estimate of potential cross-hybridization that can be removed from the raw probe set signal during data processing. The situation is more difficult for the designs where such mismatch probes are not incorporated. Several methods have been explored recently to correct for such fragment cross-hybridization. One such approach, described by Rigsbee *et al.* (2011) involved the use of an algorithm for the correction of cross-hybridization of 16S rRNA gene targets among different phylotypes. In this method, the model was first built to estimate the measured total signal for each probe set as a combination of true signal from target–probe hybridization and false signal from cross-hybridizing fragments. To provide model parameters, the levels of cross-hybridization for different phylotypes were acquired from validation experiments for the Microbiota Array. The experimentally determined cross-hybridization was generally limited to phylotypes within the same genus. These cross-hybridization estimates were subsequently incorporated into an adjustment algorithm to calculate true signal from total signal. The resulting true signal was then used instead of the total signal for phylotype abundance calculations. This algorithm was successfully applied to phylogenetic data acquired with Microbiota Array, and the adjusted values were shown to be more consistent with other estimates of microbial community compositions acquired with alternative molecular techniques (Rigsbee *et al.*, 2011).

Rigsbee *et al.* (2011) also introduced a second algorithm to adjust the normalized signal values for the estimated number of 16S rRNA gene copies per phylotype genome. Since different microbial species are known to contain a broad range of ribosomal RNA-encoding gene copies per genome (between 1 and 15), the measured true signal of a phylotype represents both its abundance as well as the total number of 16S rRNA gene copies it contains (for most species, 16S rRNA genes within the same organism have nucleotide sequence identity of ≥98% and thus would be expected to bind to the same probe set on the microarray). The known numbers of 16S rRNA gene copies for the various microbial species can be acquired from publicly accessible databases such as rrnDB and NCBI. Adjusting the phylotype signal value by the estimated number of 16S rRNA gene copies allowed for a more

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

accurate inference of each phylotype abundance (Rigsbee *et al*., 2011).

Phylogenetic microarray applications

Phylogenetic microarrays have been utilized to carry out successfully many different studies that interrogated a diverse set of human associated microbial niches. Examples of such high-throughput analyses using phylogenetic microarrays are discussed in this section.

PhyloChip

The G2 version of the PhyloChip was used by Lemon *et al.* (2010) to investigate the microbial profile of the nostril and oropharynx niches in seven healthy adults. Microarray results indicated a higher diversity and greater stability in the microbial communities associated with the oropharynx compared with those of the nostril. All communities were dominated by members of the phyla Firmicutes, Proteobacteria, Actinobacteria, and Bacteroidetes. Interestingly, the nostril community was dominated by Firmicutes and Actinobacteria, while the oropharynx was dominated by Firmicutes and Proteobacteria. Microarray results displayed another interesting observation – the nostril microbiota was similar to that of the skin, while the oropharynx communities were akin to microbial profiles of the saliva. While both communities were dominated by Firmicutes, each region housed distinct families of Firmicutes, again suggesting existence of region specific microbial profiles. Finally, a stark inverse relationship was observed between the relative abundances of Firmicutes and Proteobacteria in the oropharynx and between Fimicutes and Actinobacteria in the nostril (Lemon *et al*., 2010).

Palmer *et al.* (2007) array

A custom phylogenetic microarray designed by Palmer *et al*. (2007) was employed to investigate the gut microbial communities of 14 full-term infants. In this study, the composition and the temporal changes of the intestinal microbiota in infants were explored. The results from the microarray indicated that the phylum level diversity of the infant faecal samples was limited. Most samples were dominated by three groups of bacteria (*Flexibacter–Cytophaga–Bacteroides*, Proteobacteria, and Firmicutes-Actinobacteria), and a high degree of microbiota variability was observed among individual infants. Each baby had a distinct set of microbial species that had been acquired and maintained in his gut. Surprisingly, this initial microbial profile was temporally stable in each infant over intervals of weeks to months. The timing of the first appearance of genus *Bacteroides* varied greatly between individuals, however, it was routinely found at varied abundance levels in almost all subjects by the end of the first year of age. As infants reached the age of one year, these distinct individual profiles began to converge to resemble that of an 'adult-like' microbiota community (Palmer *et al*., 2007). This study also confirmed prior reports that the earliest colonizers of the human gut are usually aerobes and facultative anaerobes, while late colonizers are obligate anaerobes.

The Microbiota Array

The faecal microbiomes of healthy adolescents and adolescents with diarrhoea-predominant irritable bowel syndrome (IBS) were profiled in a study by Rigsbee *et al.* (2012). The objective of the study was to assess the differences in the faecal microbiota profiles between the two groups and to potentially identify putative associations among different microbial members. This study took advantage of the quantitative nature of the Microbiota Array to compare relative abundances among the interrogated samples at several taxonomic levels; microarray data were validated with high-throughput 454-based pyrosequencing and fluorescence *in situ* hybridization (FISH). The study showed that the overall structure of the faecal microbiomes was generally similar between healthy adolescents and adolescents with IBS. In both groups, the phylum Firmicutes was the most abundant, followed by Actinobacteria and Bacteroidetes, with members of these three phyla cumulatively constituting 91% of the overall community composition on average. At the genus level, the relative fractions of the abundant genera in the microbial communities were also similar between the two groups; the polysaccharide-degrading

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

members of the genus *Ruminococcus* were the most abundant.

Some distinct differences in the microbial profiles were observed at lower taxonomic levels (genus and species). More specifically, the array detected lower levels of genus *Bifidobacterium* but higher levels of genera *Lactobacillus*, *Veillonella*, and *Prevotella* in adolescents with IBS, an observation that is consistent with several other reports. The array also allowed for the characterization of a set of phylotypes that were present in all or most samples. Such set of phylotypes can be referred to as the core microbiome of that niche, which is often thought to play important roles in the community functional capacity including inter-species and host–microbial interactions. In the combined set of adolescent faecal samples, the array identified a core microbiome of 55 phylotypes (see also Fig. 11.3). The microarray based findings were confirmed both by pyrosequencing and by FISH (Rigsbee *et al.*, 2012).

HITChip

The HITChip was used to study the diversity and the temporal stability of microbial communities in the ileum of patients with ileostomy (Booijink *et al.*, 2010b). Microarray results revealed the dominance of *Streptococcus*, *Veillonella*, and *Lactobacillus* in the ileal contents. In general, the microbial community associated with the ileum was found to be less complex than that of a typical distal gut community. Temporal changes in the microbial profiles were characterized by assessing ileal microbiota composition over a period of 28 days. Interestingly, substantial differences in the microbial profiles were seen even within 1 day for the same individual, as significant changes were detected between morning and evening samples of the same day. This instability of the ileal microbiota is in stark contrast with previous reports that indicate that the human distal gut microbiota is relatively stable over long periods of time (Claesson *et al.*, 2011; Costello *et al.*, 2009; Jalanka-Tuovinen *et al.*, 2011; Rajilic-Stojanovic *et al.*, 2009; Zoetendal *et al.*, 1998). The authors explained this relative instability in the ileal biota by the more significant fluctuations of lumenal contents in the small intestine compared with those of the colon (Booijink *et al.*, 2010b).

AUS-HIT Chip

The Aus-HIT Chip was used to identify the impact of gut region and gender on the microbiota composition in the human intestinal tract (Aguirre de Carcer *et al.*, 2011). Biopsy samples from different gut regions of ten human volunteers were analysed, and numerical ecology approach was utilized to uncover the influence of the biogeographic gut location on the microbial communities and to subtract subject-specific effect on the microbiota composition. The constrained ordination based analysis showed a gradual reduction in the abundance levels of *Streptococcus*, Comamonadaceae, *Enterococcus*, *Corynebacterium*, and *Lactobacillus* between proximal colon and the rectum. Enterobacteriaceae, on the other hand, increased in abundance along the colon. These collective differences were substantiated by the multivariate analysis of quantitative PCR data. The authors were also able to identify gender-based differences in the microarray profiles, especially for the streptococci and *Faecalibacterium prausnitzii* (Aguirre de Carcer *et al.*, 2011).

Manges *et al.* (2010) array

The custom phylogenetic microarray developed by Manges and colleagues was used to determine the link between intestinal microbiota, epidemiological cofactors such as nosocomial exposures, and the consequent risk of *Clostridium difficile* associated disease (CDAD) (Manges *et al.*, 2010). Array results indicated a higher abundance of phyla Firmicutes, Proteobacteria, Actinobacteria, and Tenericutes and a lower abundance of Bacteroidetes in patients with CDAD, based on the probe signal intensities prior to adjustments for confounding factors. However, after adjusting for antibiotic use, only a small group of organisms within the phyla Bacteroidetes and Firmicutes remained associated with CDAD risk. At lower taxonomic levels, consistent with previous reports, microarray results showed an association between CDAD risk and the high abundance of Lactobacillaceae and Enterococcaceae. Overall, the study indicated that exposure to confounding factors such as antibiotic use elicits substantial changes in the distal gut microbial profiles and contributes to the risk of nosocomial CDAD (Manges *et al.*, 2010).

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

HOMIM

HOMIM microarray was utilized to assess the microbiota composition in the saliva of healthy children and children with dental caries (Luo *et al.*, 2012). The objective of this research project was to determine microbial biomarkers for the onset of dental caries in mixed dentition and to characterize the community profile of the microbial disease. In total, the study identified 86 phylotypes as well as eight clusters of closely related OTUs. In agreement with several sequencing studies, the microbial community of the saliva was found to be dominated by the phyla Firmicutes and Proteobacteria. The overall relative contribution of different phyla to the total microbial abundance was similar in both sample groups with the exception of TM7 phylum, which was only detected in the caries-active group. A higher microbial diversity, with 89 detected species, was observed in communities from the caries-active group, compared to the caries-free healthy group that contained on average 59 species. This suggested a shift in microbial community structure in response to the change from a healthy to a diseased oral environment. Examining the genus level relative abundances revealed that genus *Streptococcus* was the most abundant, followed by *Prevotella* and *Selenomonas*.

At the phylotype level, cariogenic species such as *Streptococcus mutans* and members of the cariogenic genus *Lactobacillus* were surprisingly not highly prevalent in the caries-active group, to the contrary to several previous reports. Interestingly, these cariogenic groups were substituted by the high prevalence of other streptococci. Examples of phylotypes that were differentially abundant between the two groups included species of *Leptotrichia*, which were found only in caries-active patients, and *Granulicatella* sp. and *Rothia dentocariosa*, which were found at much higher abundance in healthy children. There was a much greater number of phylotypes unique to the caries-active group compared to those unique to the healthy group, likely due to the higher community diversity seen in the caries-active group. A member of genus *Fusobacterium*, *Fusobacterium nucleatum* was found to be prevalent in all oral samples, which the authors attributed to the key role this species plays in the establishment of microbial communities in naturally forming dental plaques (Luo *et al.*, 2012).

OC Chip

The oral microbiota of children at various stages of their dentition with respect to their oral health was investigated using the Oral Cavity Chip and 454-based pyrosequencing (Crielaard *et al.*, 2011). Microarray results revealed ubiquitous presence of members from the phyla Actinobacteria, Bacteroidetes, Firmicutes, and Fusobacteria in the oral samples. At the genus level, *Actinomyces*, *Rothia*, *Prevotella*, *Streptococcus*, *Veillonella*, *Lactobacillus*, *Granulicatella*, and *Fusobacterium* were widely present in all analysed samples. The abundances of *Veillonella* and *Prevotella* increased with age in studied children. Proteobacteria was found in all but one sample, with Betaproteobacteria being the most prevalent class. Significance Analysis of Microarrays (SAM) was used to determine associations of bacterial groups with the oral health status of the profiled individuals. SAM revealed that the healthy oral status was associated with the prevalence of *Porphyromonas catoniae* and *Neisseria flavescens*. Finally, using microbe abundance values, Principal Components Analysis was able to partially separate treated and carious oral samples from those of the healthy group (Crielaard *et al.*, 2011).

V-Chip

The vaginal microbiota of African women with or without bacterial vaginosis (BV) was examined by Dols *et al.* (2011) through the use of the vaginal microbiota-representing microarray (V-Chip). The goal of the study was to first test the ability of the microarray to successfully detect microbes found at high prevalence in BV, and to characterize the profiles of the vaginal microbial communities in women in the study group. The microarray results showed that women who were negative for BV had a high prevalence of various species of *Lactobacillus*, a genus that include many members considered beneficial to human health. BV-positive women harboured a much larger set of known microbial pathogens as well as a more complex microbiota than women from BV negative or intermediate groups. The microarray data also indicated that high prevalence of HIV

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

in many cases correlated with high prevalence of BV. At a species level, the study revealed that *Gardnerella vaginalis* and *Atopobium vaginae* co-occurred in nearly 70% of the women, suggesting potential microbial interaction(s) between these species towards pathogenesis. The presence of *Gardnerella* was also associated with the presence of *Leptotrichia* and *Prevotella* species. Interestingly, while previous reports found *Gardnerella vaginalis* to be generally associated with BV diagnosis, this species was also present in 24% of BV-negative women profiled in this study. Thus, the microarray data did not support the previous use of the presence of this organism as a diagnostic tool for BV. Instead, the authors proposed to employ the co-occurrence of *Gardnerella vaginalis* and other pathogens such as *Atopobium vaginae* as a criterion for the diagnosis of bacterial vaginosis.

Unique analyses enabled by phylogenetic microarrays

One of the major advantages of phylogenetic microarrays is their ability to measure *quantitatively* the levels of *all* interrogated phylotypes in *all* analysed samples. Thus, the presence and abundance of a particular phylotype can be assessed in each sample, and this assessment does not rely on the depth of sequencing that is the case with next generation sequencing methods. This unique phyloarray ability provides opportunities to ask and answer several important questions about microbial community composition and ecological characteristics.

1. Because the hybridization signal measured by phyloarray is proportional to the number of DNA fragments of each phylotype in the interrogated mixture, we can quantitatively compare the abundance of each phylotype between any two samples.
2. This quantitative nature of the data also allows us to compute any correlative associations in the levels of any number of phylogenetic groups among the analysed set of samples.
3. Finally, we can assess the ubiquitousness and, alternatively, uniqueness of each phylotype in different hosts by calculating the fraction of samples it is detected in. We present several examples of these analyses in the sections below.

Quantitative comparisons of phylotype levels

While many microbiota analysis studies employing alternative technologies assess collected data by comparing taxonomical group distribution in each sample (exemplified by the stacked column or bar graphs), microarray data sets also allow researchers to directly compare taxonomical group levels among interrogated samples. For example, Agans *et al.* (2011) has identified 14 genera that were present at different relative abundances in faecal samples collected from healthy adults and adolescent children. Among these, levels of *Sutterella*, *Enterobacter*, *Butyrivibrio*, *Peptococcus*, and *Slackia* were statistically significantly higher in stools of adults; children samples showed statistically higher levels of *Clostridium*, *Turicibacter*, and *Bifidobacterium* (see Table 11.1 in the original publication).

Rigsbee and colleagues have compared the levels of different species of genus *Bacteroides* in faecal samples isolated from healthy kids and those diagnosed with diarrhoea-predominant irritable bowel syndrome (Rigsbee *et al.*, 2012). Cumulatively, the members of this genus had similar abundances among samples of each group on average (5.7% and 6.1% relative abundance). However, species-level comparison revealed that while *Bacteroides caccae*, *B. fragilis*, *B. thetaiotaomicron*, and *B. uniformis* were equally abundant among healthy and IBS samples, *B. eggerthii* was 11-fold more abundant among healthy children. At the same time, both *B. ovatus* and *B. salyersiae* were more widespread among IBS patients (see Table 2 in the original publication).

Biagi and co-workers have employed HITChip to assess the gut microbiota of centenarians, elderly, and young adults (Biagi *et al.*, 2010). By comparing directly microarray signals among sample groups, several genus-like bacterial groups were found to differ significantly between centenarians and young adults. For example, centenarian guts contained statistically significantly higher levels of *Vibrio*, *Bacillus*, *Eggerthella*, and *Klebsiella*, while samples from young adults were comparatively enriched in *Faecalibacterium*,

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

Eubacterium, *Roseburia*, and *Ruminococcus* (Biagi *et al.*, 2010).

Quantitative associations among community members

The quantitative nature of the data from phylogenetic microarrays allows for statistical assessment of the relationships among the abundance levels of various microbial groups across the analysed samples. This type of information can be useful in predicting and testing putative metabolic or spatial interactions among community members. In the context of community metabolic function, a positive association can indicate synergistic metabolic interaction(s) in the overall flow of metabolites and energy through the community metabolic network. An example of such community correlation analysis is provided by Rigsbee *et al.* (2012). In this study, a Spearman's correlation matrix was calculated using the genus level abundances from 44 analysed samples obtained with the Microbiota Array (Fig. 11.2). Overall, a total of 53 interactions spanning 35 genera were identified in this analysis. The large number of identified associations is consistent with our current understanding of the intricate nature of metabolic networks among the community members in the intestinal ecosystem. One of the most striking observations from this analysis was the substantial number of statistically significant relationships of the genus *Veillonella*

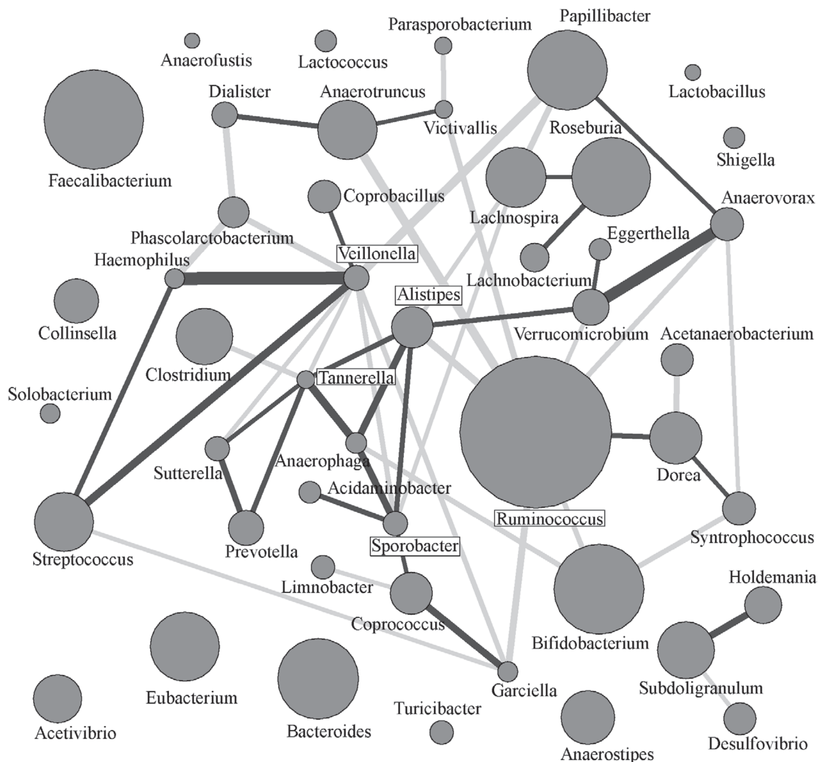

Figure 11.2 Associations of genus abundances among faecal samples collected from healthy and IBS children (Rigsbee *et al.*, 2012). Each node represents a particular bacterial genus as shown. The lines connecting the nodes denote the statistically significant relationships among the genera. The widths of the lines are proportional to the absolute value of the Spearman's rank correlation of the abundances of the corresponding genera among all samples. Positive relationships are designated by dark grey lines and negative – by light grey lines. Only genera with an average relative abundance of above 0.1% are shown; and only relationships with at least 95% confidence are displayed. The size of the node is proportional to the square root of the average genus abundance among all samples. The figure was first published in American Journal of Gastroenterology, issue 107, 2012 (Rigsbee *et al.*, 2012), produced by Nature Publishing Group, a division of Macmillan Publishers Limited.

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

with other microbiota members (Fig. 11.2). This is likely because the members of this genus participate in the metabolic cross-feeding pathways. Specifically, *Veillonella* species lack enzymes necessary for the degradation of complex as well as simple sugars that are commonly found in the colonic environment and thus rely on the use of intermediary end-products of carbohydrate fermentation (such as lactate, pyruvate, and fumarate) that are excreted by other members of the gut community (Gronow *et al.*, 2009). Indeed, a physical association between lactate-utilizing *Veillonella* and lactate-producing *Streptococcus* was observed in dental plaque (Chalmers *et al.*, 2008).

Correlation of the level of *Lactobacillus casei* LGG with other members of the microbiota was assessed in infant gut microbiome (Cox *et al.*, 2010b). Infants were fed daily supplements of either LGG or placebo, and the microbiota profile was determined with PhyloChip array. Daily supplementation with *L. casei* promoted higher abundance of other gut members with known beneficial properties, and resulted in increased community evenness. LGG was significantly correlated with 361 taxa, with vast majority of these correlations being positive. Positively associated taxa included other probiotic species such as *Lactobacillus fuchuensis* and *Bifidobacterium bifidum*, members of the Helicobacteraceae, and *Streptomyces coelicolor*, a known producer of antibiotics (Cox *et al.*, 2010b). The negative associations were formed with *Bacteroides uniformis*, *B. merdae*, and Lachnospiraceae clone.

Definition of a core microbiome

When a set of phylotypes is found to be present in most or all analysed samples, this set can be considered to form a 'core' of that microbiome. The members of the core set might serve as the primary degraders of available nutrients, or they can potentially play important roles in the inter-species and host–microbial interactions. Definition of a core phylotype group for diseased states can provide strong evidence of association of certain microbial groups with disease pathologies. Likewise, a core microbiome of healthy states can be used to understand the role of the core members in the maintenance of homeostasis. Several recent efforts have been described that sought to define core component of human-associated microbiomes. Rigsbee *et al.* (2012) defined gut core microbiome sets for healthy and IBS adolescent children. The study determined that the core microbiome of the IBS set (46 phylotypes) was smaller than that of the healthy samples (56 phylotypes). By allowing a particular phylotype to be detected in all but one sample, a combined core of 55 phylotypes was defined among all analysed samples. Consistent with previous reports that were based on 16S rRNA gene sequencing studies, this core set was dominated by members of Clostridia and by genus *Ruminococcus* specifically. The members of the core set contributed on average about 30% to the total microbial abundance, indicating that the core set factored considerably towards the overall community composition. There was less variability in the abundances of the members within this core, compared to the members of the 'shared' group which represented phylotypes that were present in multiple but not all samples. As expected, extending the analysis to 60 samples reduced the core size slightly (Fig. 11.3), while still maintaining representation of well-known members of the gut microbiota community in the core (members of genera *Anaerostipes*, *Bacteroides*, *Dorea*, *Eubacterium*, *Faecalibacterium*, *Peptostreptococcus*, *Roseburia*, *Ruminococcus*, and *Streptococcus*). Many microbial phylotypes were found to be uniquely present in one out of 60 analysed samples, emphasizing the unique personal nature of human microbial communities.

HITChip and PhyloChip were also utilized in several studies to define microbiota core among human samples (Huang *et al.*, 2010; Jalanka-Tuovinen *et al.*, 2011; Rajilic-Stojanovic *et al.*, 2009; Salonen *et al.*, 2012). Huang and colleagues have assessed the airway bacterial communities in patients with chronic obstructive pulmonary disease. A core community of 75 taxa was detected in all patients, many of these taxa contained known pathogens including several species of the genera *Helicobacter* and *Pseudomonas* as well as *Leptospira interrogans* and *Arcobacter cryaerophillus* (Huang *et al.*, 2010). Another example of core microbiome analysis was presented by Salonen and co-workers (Salonen *et al.*, 2012). The study addressed the effects of technical and biological variables on the determination of the core microbiome. The study

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

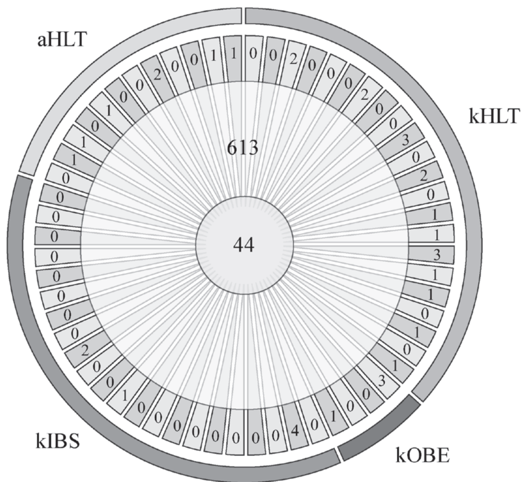

Figure 11.3 Definition of the core human distal gut microbiome using Microbiota Array. The figure displays the distribution of detected phylotypes among 60 samples of human distal gut microbiota obtained from four groups of participants. The outmost bands illustrate the group designation. The phylotypes unique to each analysed sample are shown in the individual outer segments. The inner circle displays core species detected in at least 59 samples. The shared set represents phylotypes present in more than 1 but not all analysed samples. Key: aHLT, healthy adults; kHLT, healthy children (kids); kOBE, obese children; kIBS, children diagnosed with IBS. The figure was first published in FEMS Microbiology Ecology, issue 79, 2012 (Paliy and Agans, 2012), published by Blackwell Publishing Ltd.

showed that when utilizing the 100% prevalence criterion and while including low abundance phylotypes, the estimated core was nearly one-third of the overall number of detected phylotypes among 115 analysed samples from healthy subjects. These estimates are considerably larger than those from previous sequencing studies that have used lower stringency criteria (Salonen *et al.*, 2012), possibly due to the higher sensitivity of detection achieved with phylogenetic microarrays. The core size depended heavily on the minimum abundance cutoff – a requirement for higher minimum phylotype abundance in each sample reduced the core size substantially. Similar to the study by Rigsbee *et al.* (2012), the core microbiomes of healthy individuals and patients with ulcerative colitis (UC) have been also calculated separately. The core microbiome of the patients with UC was smaller than that of healthy individuals, possibly implying a loss of potential health-promoting members of the gut community in these patients (Salonen *et al.*, 2012).

Future trends and outlook

High-throughput techniques such as phylogenetic microarrays and next-generation sequencing provide us with extensive knowledge regarding the composition of complex microbial communities. This knowledge enables us to understand which members are present in the community as well as to predict their potential role. Examples of the phyloarray applications that have been described in the previous section of this chapter highlight a multitude of questions that can be answered through the use of phylogenetic microarrays. Phyloarrays were also used to study gut microbiota development in infants (Cox *et al.*, 2010b; Palmer *et al.*, 2007); altered faecal microbiota in patients with IBD (Kang *et al.*, 2010), IBS (Kajander *et al.*, 2008) and *Clostridium difficile* infection (Manges *et al.*, 2010); oral microbiota in children (Crielaard *et al.*, 2011), adults (Olson *et al.*, 2011) and in the elderly (Preza *et al.*, 2009b); and the differences in airway microbiota of paediatric and adult cystic fibrosis patients (Cox *et al.*, 2010a). The intricate nature of the microarray design process and the extensive validation procedures have been limiting factors towards the wider use of phylogenetic microarrays. Nonetheless, there already exists an assortment of phylogenetic microarrays capable of analysing a variety of microbial ecosystems (Table 11.1). The improvements in cost efficiency and the highly quantitative nature of phyloarrays make them an excellent choice for high-throughput compositional analysis of microbial communities. A particularly attractive application is the use of both phylogenetic microarrays and next-generation sequencing for the analysis of the same microbial community (Ahn *et al.*, 2011; Crielaard *et al.*, 2011; Rigsbee *et al.*, 2012; van den Bogert *et al.*, 2011). In this approach, the phyloarrays can provide quantitative data for the comparison of abundances across groups and samples, while the 16S rRNA amplicon sequencing can allow the identification of novel members of the community.

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

Thanks to the advancements in technology and our knowledge of microbial communities, several enhancements to the design and use of phylogenetic microarrays can also be conceived. Programs such as the Human Microbiome Project (Peterson *et al.*, 2009) and the MetaHIT initiative (Qin *et al.*, 2010) have made available a substantial number of genome sequences of human-associated microbiota members. The availability of such resources has given rise to the possibility of designing phylogenetic detection arrays based on functionally conserved genes such as *groEL*, *rpoB*, *gyrA*, and *tuf* (Loy and Bodrossy, 2006). Specific pathogen detection arrays have a potential to play a vital role in the field of forensics for the rapid detection and identification of pathogens in the environment (Jin *et al.*, 2005). Furthermore, phylogenetic microarrays can also be designed to contain probes to functional genes to enable simultaneous analysis of community structure and function (Louis and Flint, 2007).

The future trends in the use of phylogenetic microarrays are likely to be defined by a shift towards integrative approaches to community analysis. Current studies have helped us understand the composition of microbial communities. Using this information in combination with new molecular tools, future studies will likely focus on the interactions among members of the microbial communities as well as between microbiota and the environment. There is also a growing interest towards understanding the link(s) between the function and the activity of microbiota in various environmental niches or disease states. In integrative approaches, the use of phylogenetic microarrays can be augmented with other high-throughput methods such as metabolomics, meta-genomics, meta-transcriptomics, and meta-proteomics to construct a more comprehensive model of the interrogated community (Booijink *et al.*, 2010a; Klaassens *et al.*, 2007; Martin *et al.*, 2010). A combination of these techniques would allow us to determine the profile of the community composition, total gene content, and the expression levels of many genes and proteins, and we would be able to relate this data to the metabolite profiles of the environment and community members. Such an approach will enable us to understand the intricate relationships and the roles the members of the microbiota play within different microbial ecosystems.

In the clinical setting, phylogenetic microarrays can be used as diagnostic tools, where their ability to detect human-associated microbiota members at a species level in a relatively short period of time can help in the diagnosis of various pathological states and rapid selection of treatment procedures that are most likely to succeed (Loy and Bodrossy, 2006). Since phyloarrays can be run one at a time, they offer great flexibility when individual patient samples need to be analysed. At the same time, the high-throughput nature of microarrays allows for the simultaneous species-level detection of hundreds of human microbiota members and pathogens, a feature that is vital in clinical diagnostics. Custom microarrays can be designed that are focused on the identification of a set of pathogens from a particular body organ or niche (gut, skin, oral cavity, etc.) (Jin *et al.*, 2005; Preza *et al.*, 2009a). At the same time, overall community disbalance can also be determined, which can indicate predisposition to a particular disease (obesity, diabetes) or the early stages of a non-symptomatic pathophysiological state [lack of *Faecalibacterium prausnitzii* as a marker of IBD (Swidsinski *et al.*, 2008)]. Based on the analysis of these microbiotas, a personalized treatment option can be designed for each individual patient, which is expected to increase the success rate of such treatment and to ultimately facilitate patient recovery.

Acknowledgements

The work in the Paliy laboratory is supported by the National Institutes of Health grants AT003423 and HD065575.

Web resources

Additional information about studies carried out in Paliy laboratory is available at
http://www.wright.edu/~oleg.paliy/papers.html.
Sequence databases:
http://rdp.cme.msu.edu/
http://www.ncbi.nlm.nih.gov/genbank/
Microarray design:
http://www.arb-home.de/
http://g2im.u-clermont1.fr/serimour/phylarray/
http://rdp.cme.msu.edu/
Microarray analysis:
https://carmaweb.genome.tugraz.at/carma/

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

References

Agans, R., Rigsbee, L., Kenche, H., Michail, S., Khamis, H.J., and Paliy, O. (2011). Distal gut microbiota of adolescent children is different from that of adults. FEMS Microbiol. Ecol. *77*, 404–412.

Aguirre de Carcer, D., Cuiv, P.O., Wang, T., Kang, S., Worthley, D., Whitehall, V., Gordon, I., McSweeney, C., Leggett, B., and Morrison, M. (2011). Numerical ecology validates a biogeographical distribution and gender-based effect on mucosa-associated bacteria along the human colon. ISME J. *5*, 801–809.

Ahn, J., Yang, L., Paster, B.J., Ganly, I., Morris, L., Pei, Z., and Hayes, R.B. (2011). Oral microbiome profiles: 16S rRNA pyrosequencing and microarray assay comparison. PLoS One *6*, e22788.

Ashelford, K.E., Chuzhanova, N.A., Fry, J.C., Jones, A.J., and Weightman, A.J. (2005). At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Appl. Environ. Microbiol. *71*, 7724–7736.

Belenguer, A., Duncan, S.H., Calder, A.G., Holtrop, G., Louis, P., Lobley, G.E., and Flint, H.J. (2006). Two routes of metabolic cross-feeding between Bifidobacterium adolescentis and butyrate-producing anaerobes from the human gut. Appl. Environ. Microbiol. *72*, 3593–3599.

Biagi, E., Nylund, L., Candela, M., Ostan, R., Bucci, L., Pini, E., Nikkila, J., Monti, D., Satokari, R., Franceschi, C., *et al.* (2010). Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. PLoS ONE *5*, e10667.

van den Bogert, B., de Vos, W.M., Zoetendal, E.G., and Kleerebezem, M. (2011). Microarray analysis and barcoded pyrosequencing provide consistent microbial profiles depending on the source of human intestinal samples. Appl. Environ. Microbiol. *77*, 2071–2080.

Booijink, C.C., Boekhorst, J., Zoetendal, E.G., Smidt, H., Kleerebezem, M., and de Vos, W.M. (2010a). Metatranscriptome analysis of the human fecal microbiota reveals subject-specific expression profiles, with genes encoding proteins involved in carbohydrate metabolism being dominantly expressed. Appl. Environ. Microbiol. *76*, 5533–5540.

Booijink, C.C., El-Aidy, S., Rajilic-Stojanovic, M., Heilig, H.G., Troost, F.J., Smidt, H., Kleerebezem, M., De Vos, W.M., and Zoetendal, E.G. (2010b). High temporal and inter-individual variation detected in the human ileal microbiota. Environ. Microbiol. *12*, 3213–3227.

Brodie, E.L., Desantis, T.Z., Joyner, D.C., Baek, S.M., Larsen, J.T., Andersen, G.L., Hazen, T.C., Richardson, P.M., Herman, D.J., Tokunaga, T.K., *et al.* (2006). Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. Appl. Environ. Microbiol. *72*, 6288–6298.

Candela, M., Consolandi, C., Severgnini, M., Biagi, E., Castiglioni, B., Vitali, B., De Bellis, G., and Brigidi, P. (2010). High taxonomic level fingerprint of the human intestinal microbiota by ligase detection reaction – universal array approach. BMC Microbiol. *10*, 116.

Chalmers, N.I., Palmer, R.J., Jr., Cisar, J.O., and Kolenbrander, P.E. (2008). Characterization of a Streptococcus sp.-Veillonella sp. community micromanipulated from dental plaque. J. Bacteriol. *190*, 8145–8154.

Choe, S.E., Boutros, M., Michelson, A.M., Church, G.M., and Halfon, M.S. (2005). Preferred analysis methods for Affymetrix GeneChips revealed by a wholly defined control dataset. Genome Biol. *6*, R16.

Claesson, M.J., Cusack, S., O'Sullivan, O., Greene-Diniz, R., de Weerd, H., Flannery, E., Marchesi, J.R., Falush, D., Dinan, T., Fitzgerald, G., *et al.* (2011). Composition, variability, and temporal stability of the intestinal microbiota of the elderly. Proc. Natl. Acad. Sci. U.S.A. *108 Suppl 1*, 4586–4591.

Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I., and Knight, R. (2009). Bacterial community variation in human body habitats across space and time. Science *326*, 1694–1697.

Cox, M.J., Allgaier, M., Taylor, B., Baek, M.S., Huang, Y.J., Daly, R.A., Karaoz, U., Andersen, G.L., Brown, R., Fujimura, K.E., *et al.* (2010a). Airway microbiota and pathogen abundance in age-stratified cystic fibrosis patients. PLoS ONE *5*, e11044.

Cox, M.J., Huang, Y.J., Fujimura, K.E., Liu, J.T., McKean, M., Boushey, H.A., Segal, M.R., Brodie, E.L., Cabana, M.D., and Lynch, S.V. (2010b). Lactobacillus casei abundance is associated with profound shifts in the infant gut microbiome. PLoS ONE *5*, e8745.

Crielaard, W., Zaura, E., Schuller, A.A., Huse, S.M., Montijn, R.C., and Keijser, B.J. (2011). Exploring the oral microbiota of children at various developmental stages of their dentition in the relation to their oral health. BMC Med. Genomics *4*, 22.

Deangelis, K.M., Allgaier, M., Chavarria, Y., Fortney, J.L., Hugenholtz, P., Simmons, B., Sublette, K., Silver, W.L., and Hazen, T.C. (2011). Characterization of trapped lignin-degrading microbes in tropical forest soil. PLoS ONE *6*, e19306.

De Vuyst, L., and Leroy, F. (2011). Cross-feeding between bifidobacteria and butyrate-producing colon bacteria explains bifdobacterial competitiveness, butyrate production, and gas production. Int. J. Food Microbiol. *149*, 73–80.

Docktor, M.J., Paster, B.J., Abramowicz, S., Ingram, J., Wang, Y.E., Correll, M., Jiang, H., Cotton, S.L., Kokaras, A.S., and Bousvaros, A. (2012). Alterations in diversity of the oral microbiome in pediatric inflammatory bowel disease. Inflamm. Bowel Dis. *18*, 935–942.

Dols, J.A., Smit, P.W., Kort, R., Reid, G., Schuren, F.H., Tempelman, H., Bontekoe, T.R., Korporaal, H., and Boon, M.E. (2011). Microarray-based identification of clinically relevant vaginal bacteria in relation to bacterial vaginosis. Am. J. Obstet. Gynecol *204*, 305 e301–307.

Duncan, S.H., Louis, P., and Flint, H.J. (2004). Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. Appl. Environ. Microbiol. *70*, 5810–5817.

Flint, H.J., Bayer, E.A., Rincon, M.T., Lamed, R., and White, B.A. (2008). Polysaccharide utilization by

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

gut bacteria: potential for new insights from genomic analysis. Nat. Rev. *6*, 121–131.

Gao, Z., Tseng, C.H., Strober, B.E., Pei, Z., and Blaser, M.J. (2008). Substantial alterations of the cutaneous bacterial biota in psoriatic lesions. PLoS ONE *3*, e2719.

Gill, S.R., Pop, M., Deboy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S., Gordon, J.I., Relman, D.A., Fraser-Liggett, C.M., and Nelson, K.E. (2006). Metagenomic analysis of the human distal gut microbiome. Science *312*, 1355–1359.

Grice, E.A., and Segre, J.A. (2011). The skin microbiome. Nat. Rev. *9*, 244–253.

Gronow, S., Welnitz, S., Lapidus, A., Nolan, M., Ivanova, N., Glavina Del Rio, T., Copeland, A., Chen, F., Tice, H., Pitluck, S., *et al.* (2009). Complete genome sequence of Veillonella parvula type strain (Te3). Stand. Genomic Sci. *2*, 57–65.

Guschin, D.Y., Mobarry, B.K., Proudnikov, D., Stahl, D.A., Rittmann, B.E., and Mirzabekov, A.D. (1997). Oligonucleotide microchips as genosensors for determinative and environmental studies in microbiology. Appl. Environ. Microbiol. *63*, 2397–2402.

Hashsham, S.A., Wick, L.M., Rouillard, J.M., Gulari, E., and Tiedje, J.M. (2004). Potential of DNA microarrays for developing parallel detection tools (PDTs) for microorganisms relevant to biodefense and related research needs. Biosens. Bioelectron. *20*, 668–683.

Hazen, T.C., Dubinsky, E.A., DeSantis, T.Z., Andersen, G.L., Piceno, Y.M., Singh, N., Jansson, J.K., Probst, A., Borglin, S.E., Fortney, J.L., *et al.* (2010). Deep-sea oil plume enriches indigenous oil-degrading bacteria. Science *330*, 204–208.

Huang, Y.J., Kim, E., Cox, M.J., Brodie, E.L., Brown, R., Wiener-Kronish, J.P., and Lynch, S.V. (2010). A persistent and diverse airway microbiota present during chronic obstructive pulmonary disease exacerbations. Omics *14*, 9–59.

Jalanka-Tuovinen, J., Salonen, A., Nikkila, J., Immonen, O., Kekkonen, R., Lahti, L., Palva, A., and de Vos, W.M. (2011). Intestinal microbiota in healthy adults: temporal analysis reveals individual and common core and relation to intestinal symptoms. PLoS ONE *6*, e23035.

Jin, L.Q., Li, J.W., Wang, S.Q., Chao, F.H., Wang, X.W., and Yuan, Z.Q. (2005). Detection and identification of intestinal pathogenic bacteria by hybridization to oligonucleotide microarrays. World J. Gastroenterol. *11*, 7615–7619.

Kajander, K., Myllyluoma, E., Rajilic-Stojanovic, M., Kyronpalo, S., Rasmussen, M., Jarvenpaa, S., Zoetendal, E.G., de Vos, W.M., Vapaatalo, H., and Korpela, R. (2008). Clinical trial: multispecies probiotic supplementation alleviates the symptoms of irritable bowel syndrome and stabilizes intestinal microbiota. Aliment. Pharmacol. Ther. *27*, 48–57.

Kang, S., Denman, S.E., Morrison, M., Yu, Z., Dore, J., Leclerc, M., and McSweeney, C.S. (2010). Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. Inflamm. Bowel Dis. *16*, 2034–2042.

Klaassens, E.S., de Vos, W.M., and Vaughan, E.E. (2007). Metaproteomics approach to study the functionality of the microbiota in the human infant gastrointestinal tract. Appl. Environ. Microbiol. *73*, 1388–1392.

Lantz, P., Matsson, M., Wadstrom, T., and Radstrom, P. (1997). Removal of PCR inhibitors from human faecal samples through the use of an aqueous two-phase system for sample preparation prior to PCR. J. Microbiol. Meth. *28*, 159–167.

Larsen, B., and Monif, G.R. (2001). Understanding the bacterial flora of the female genital tract. Clin. Infect. Dis. *32*, e69–77.

Lemon, K.P., Klepac-Ceraj, V., Schiffer, H.K., Brodie, E.L., Lynch, S.V., and Kolter, R. (2010). Comparative analyses of the bacterial microbiota of the human nostril and oropharynx. mBio *1*, e00129–00110.

Letowski, J., Brousseau, R., and Masson, L. (2004). Designing better probes: effect of probe size, mismatch position and number on hybridization in DNA oligonucleotide microarrays. J. Microbiol. Meth. *57*, 269–278.

Louis, P., and Flint, H.J. (2007). Development of a semiquantitative degenerate real-time pcr-based assay for estimation of numbers of butyryl-coenzyme A (CoA) CoA transferase genes in complex bacterial samples. Appl. Environ. Microbiol. *73*, 2009–2012.

Loy, A., and Bodrossy, L. (2006). Highly parallel microbial diagnostics using oligonucleotide microarrays. Clin. Chim. Acta *363*, 106–119.

Luo, A.H., Yang, D.Q., Xin, B.C., Paster, B.J., and Qin, J. (2012). Microbial profiles in saliva from children with and without caries in mixed dentition. Oral Dis. *18*, 595–601.

Luton, P.E., Wayne, J.M., Sharp, R.J., and Riley, P.W. (2002). The mcrA gene as an alternative to 16S rRNA in the phylogenetic analysis of methanogen populations in landfill. Microbiol. *148*, 3521–3530.

Manges, A.R., Labbe, A., Loo, V.G., Atherton, J.K., Behr, M.A., Masson, L., Tellis, P.A., and Brousseau, R. (2010). Comparative metagenomic study of alterations to the intestinal microbiota and risk of nosocomial Clostridum difficile-associated disease. J. Infect. Dis. *202*, 1877–1884.

Martens, M., Weidner, S., Linke, B., de Vos, P., Gillis, M., and Willems, A. (2007). A prototype taxonomic microarray targeting the rpsA housekeeping gene permits species identification within the rhizobial genus Ensifer. Syst. Appl. Microbiol. *30*, 390–400.

Martin, F.P., Sprenger, N., Montoliu, I., Rezzi, S., Kochhar, S., and Nicholson, J.K. (2010). Dietary modulation of gut functional ecology studied by fecal metabonomics. J. Proteome Res. *9*, 5284–5295.

Mendes, R., Kruijt, M., de Bruijn, I., Dekkers, E., van der Voort, M., Schneider, J.H., Piceno, Y.M., DeSantis, T.Z., Andersen, G.L., Bakker, P.A., *et al.* (2011). Deciphering the rhizosphere microbiome for disease-suppressive bacteria. Science *332*, 1097–1100.

Mitterer, G., Huber, M., Leidinger, E., Kirisits, C., Lubitz, W., Mueller, M.W., and Schmidt, W.M. (2004). Microarray-based identification of bacteria in clinical samples by solid-phase PCR amplification of 23S

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

ribosomal DNA sequences. J. Clin. Microbiol. *42*, 1048–1057.

Monteiro, L., Bonnemaison, D., Vekris, A., Petry, K.G., Bonnet, J., Vidal, R., Cabrita, J., and Megraud, F. (1997). Complex polysaccharides as PCR inhibitors in feces: Helicobacter pylori model. J. Clin. Microbiol. *35*, 995–998.

Neish, A.S. (2009). Microbes in gastrointestinal health and disease. Gastroenterology *136*, 65–80.

Olson, J.C., Cuff, C.F., Lukomski, S., Lukomska, E., Canizales, Y., Wu, B., Crout, R.J., Thomas, J.G., McNeil, D.W., Weyant, R.J., *et al.* (2011). Use of 16S ribosomal RNA gene analyses to characterize the bacterial signature associated with poor oral health in West Virginia. BMC Oral Health *11*, 7.

Paliy, O., and Agans, R. (2012). Application of phylogenetic microarrays to interrogation of human microbiota. FEMS Microbiol. Ecol. *79*, 2–11.

Paliy, O., and Foy, B. (2011). Mathematical modeling of 16S ribosomal DNA amplification reveals optimal conditions for the interrogation of complex microbial communities with phylogenetic microarrays. Bioinformatics *27*, 2134–2140.

Paliy, O., Kenche, H., Abernathy, F., and Michail, S. (2009). High-throughput quantitative analysis of the human intestinal microbiota with a phylogenetic microarray. Appl. Environ. Microbiol. *75*, 3572–3579.

Palmer, C., Bik, E.M., Digiulio, D.B., Relman, D.A., and Brown, P.O. (2007). Development of the Human Infant Intestinal Microbiota. PLoS Biol. *5*, e177.

Peplies, J., Glockner, F.O., and Amann, R. (2003). Optimization strategies for DNA microarray-based detection of bacteria with 16S rRNA-targeting oligonucleotide probes. Appl. Environ. Microbiol. *69*, 1397–1407.

Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., Schloss, J.A., Bonazzi, V., McEwen, J.E., Wetterstrand, K.A., Deal, C., *et al.* (2009). The NIH Human Microbiome Project. Genome Res. *19*, 2317–2323.

Preza, D., Olsen, I., Willumsen, T., Boches, S.K., Cotton, S.L., Grinde, B., and Paster, B.J. (2009a). Microarray analysis of the microflora of root caries in elderly. Eur. J. Clin. Microbiol. Infect. Dis. *28*, 509–517.

Preza, D., Olsen, I., Willumsen, T., Grinde, B., and Paster, B.J. (2009b). Diversity and site-specificity of the oral microflora in the elderly. Eur. J. Clin. Microbiol. Infect. Dis. *28*, 1033–1040.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., *et al.* (2010). A human gut microbial gene catalogue established by metagenomic sequencing. Nature *464*, 59–65.

Rajilic-Stojanovic, M., Heilig, H.G., Molenaar, D., Kajander, K., Surakka, A., Smidt, H., and de Vos, W.M. (2009). Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. Environ. Microbiol. *11*, 1736–1751.

Rigsbee, L., Agans, R., Foy, B.D., and Paliy, O. (2011). Optimizing the analysis of human intestinal microbiota with phylogenetic microarray. FEMS Microbiol. Ecol. *75*, 332–342.

Rigsbee, L., Agans, R., Shankar, V., Kenche, H., Khamis, H.J., Michail, S., and Paliy, O. (2012). Quantitative profiling of gut microbiota of children with diarrhea-predominant Irritable Bowel Syndrome. Am. J. Gastroenterol. *107*, 1740–1751.

Roh, S.W., Abell, G.C., Kim, K.H., Nam, Y.D., and Bae, J.W. (2010). Comparing microarrays and next-generation sequencing technologies for microbial ecology research. Trends Biotechnol. *28*, 291–299.

Salonen, A., Nikkila, J., Jalanka-Tuovinen, J., Immonen, O., Rajilic-Stojanovic, M., Kekkonen, R.A., Palva, A., and de Vos, W.M. (2010). Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. J. Microbiol. Meth. *81*, 127–134.

Salonen, A., Salojarvi, J., Lahti, L., and de Vos, W.M. (2012). The adult intestinal core microbiota is determined by analysis depth and health status. Clin. Microbiol. Infect. *18 Suppl 4*, 16–20.

Sartor, R.B. (2008). Microbial influences in inflammatory bowel diseases. Gastroenterology *134*, 577–594.

Sekirov, I., Russell, S.L., Antunes, L.C., and Finlay, B.B. (2010). Gut microbiota in health and disease. Physiol. Rev. *90*, 859–904.

Suau, A. (2003). Molecular tools to investigate intestinal bacterial communities. J. Pediatr. Gastroenterol. Nutr. *37*, 222–224.

Swidsinski, A., Loening-Baucke, V., Vaneechoutte, M., and Doerffel, Y. (2008). Active Crohn's disease and ulcerative colitis can be specifically diagnosed and monitored based on the biostructure of the fecal flora. Inflamm. Bowel Dis. *14*, 147–161.

Tao, F., Li, C., Smejkal, G., Lazarev, A., Lawerence, N., and Schumacher, R. (2006). Pressure Cycling Technology (PCT) Applications in Extraction of Biomolecules from Challenging Biological Samples. High Pressure Biosci. Biotech. *1*, 166–173.

Wang, R.F., Beggs, M.L., Robertson, L.H., and Cerniglia, C.E. (2002). Design and evaluation of oligonucleotide-microarray method for the detection of human intestinal bacteria in fecal samples. FEMS Microbiol. Lett. *213*, 175–182.

Wang, R.F., Beggs, M.L., Erickson, B.D., and Cerniglia, C.E. (2004). DNA microarray analysis of predominant human intestinal bacteria in fecal samples. Mol. Cell Probes *18*, 223–234.

Yoo, S.M., Lee, S.Y., Chang, K.H., Yoo, S.Y., Yoo, N.C., Keum, K.C., Yoo, W.M., Kim, J.M., and Choi, J.Y. (2009). High-throughput identification of clinically important bacterial pathogens using DNA microarray. Mol. Cell Probes *23*, 171–177.

Zoetendal, E.G., Akkermans, A.D., and De Vos, W.M. (1998). Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bacteria. Appl. Environ. Microbiol. *64*, 3854–3859.

UNCORRECTED PROOF

Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P

UNCORRECTED PROOF
Date: 15:01 Wednesday 21 May 2014
File: Microarrays 2P