RESEARCH ARTICLE

Optimizing the analysis of human intestinal microbiota with phylogenetic microarray

Laura Rigsbee[1], Richard Agans[1], Brent D. Foy[2] & Oleg Paliy[1]

[1]Department of Biochemistry and Molecular Biology, Wright State University, Dayton, OH, USA; and [2]Department of Physics, Wright State University, Dayton, OH, USA

Correspondence: Oleg Paliy, Department of Biochemistry and Molecular Biology, Wright State University, 260 Diggs, 3640 Col. Glenn Hwy, Dayton, OH 45435, USA. Tel.: +1 937 775 3714; fax: +1 937 775 3730; e-mail: oleg.paliy@wright.edu

Received 23 June 2010; revised 2 November 2010; accepted 8 November 2010.
Final version published online 13 December 2010.

DOI:10.1111/j.1574-6941.2010.01009.x

Editor: Julian Marchesi

Keywords
phylogenetic; microarray; intestinal; microbiota; microbial community; microflora.

Abstract

Phylogenetic microarrays present an attractive strategy to high-throughput interrogation of complex microbial communities. In this work, we present several approaches to optimize the analysis of intestinal microbiota with the recently developed Microbiota Array. First, we determined how 16S rDNA-specific PCR amplification influenced bacterial detection and the consistency of measured abundance values. Bacterial detection improved with an increase in the number of PCR amplification cycles, but 25 cycles were sufficient to achieve the maximum possible detection. A PCR-caused deviation in the measured abundance values was also observed. We also developed two mathematical algorithms that aimed to account for a predicted cross-hybridization of 16S rDNA fragments among different species, and to adjust the measured hybridization signal based on the number of 16S rRNA gene copies per species genome. The 16S rRNA gene copy adjustment indicated that the presence of members of the class *Clostridia* might be overestimated in some 16S rDNA-based studies. Finally, we show that the examination of total community RNA with phylogenetic microarray can provide estimates of the relative metabolic activity of individual community members. Complementary profiling of genomic DNA and total RNA isolated from the same sample presents an opportunity to assess population structure and activity in the same microbial community.

Introduction

Many bacteria and other microorganisms coexist in nature as complex microbial communities. Such communities are found in soil, water, inside sewage pipes, on the bark and leaves of trees, and on the skin and other epithelial surfaces of animals (Kent & Triplett, 2002; Gao *et al.*, 2007; McLellan *et al.*, 2009). Among these, human-associated microbiota is of particular interest because of its connection to human health and disease. Indeed, many studies have been published recently that aimed to catalog the diversity of human-associated bacteria and the role that these microorganisms might play in human well-being (Eckburg *et al.*, 2005; Gill *et al.*, 2006; Palmer *et al.*, 2007). Among such communities, intestinal bacteria have received special attention due to their hypothesized role in several gastrointestinal disorders such as inflammatory bowel disease, obesity, and colon cancer (van Nuenen *et al.*, 2004; Turnbaugh *et al.*, 2006; Scanlan *et al.*, 2008).

The scarcity of previous research into the microbial community structure and function is explained by the difficulty of using standard microbiological techniques to interrogate community members. In many cases, complex nutritional relationships exist among the members of the community, and hence it is difficult to cultivate individual members (Belenguer *et al.*, 2006). In the case of human intestinal microbiota, the community is dominated by obligate anaerobic species, which presents additional problems for the isolation and culturing of these bacteria (Thompson-Chagoyan *et al.*, 2007). To circumvent the need to culture community members, new molecular tools have been developed utilizing the knowledge and interrogation of genomic DNA (gDNA) sequences derived from the microbial communities. Gene sequences encoding small

ribosomal subunit 16S rRNA have been used extensively for this purpose owing to the ubiquitous nature of this molecule and to our ability to selectively isolate and amplify the gene with phylogenetically conserved primers (Cannone *et al.*, 2002). A variety of methods have been used for 16S rDNA interrogation including restriction polymorphism analysis, different types of denaturing gel electrophoresis, FISH, quantitative PCR, high-throughput DNA sequencing, and microarrays. Among these methodologies, microarrays proved to present a good compromise between high sensitivity of detection of individual species, high-throughput simultaneous interrogation of all community members, and most importantly, an ability to quantitatively compare community structure among different samples (Carey *et al.*, 2007).

Unfortunately, the use of any technology based on the interrogation of DNA sequences presents a number of challenges. Most of these stem from the fact that DNA-manipulating reactions such as isolation, amplification, fragmentation, and annealing can proceed stochastically and dissimilarly for different DNA sequences and samples (Sachse, 2004). Differences among bacteria in the thermodynamic properties of gDNA, due to, for example, variations in the genome GC content, as well as vast dissimilarities among bacterial species in the number of 16S rRNA gene copies per genome [can vary between 1 and 15 (Lee *et al.*, 2009)] also create a challenge to link the measured DNA amount to population structure and function. Many strides have been made to reduce some of these problems (Polz & Cavanaugh, 1998; Kurata *et al.*, 2004); however, the optimal use of the technology remains a difficult task.

We have recently designed a phylogenetic Microbiota Array able to quantitatively measure the relative abundance of 775 different bacterial species found in the human intestine (Paliy *et al.*, 2009). The microarray has been extensively validated on pure bacterial cultures and was later used to profile intestinal microbiota of children and adults. This article describes a number of approaches that we developed in our laboratory to optimize the use of oligo microarrays for the analysis of microbial community composition and activity. We also show that by interrogating total community RNA, the phylogenetic microarray can provide estimates of metabolic activities of each community member.

Materials and methods

Fecal sample collection and processing

Fresh fecal samples were obtained from two healthy adults as approved by Wright State University IRB committee. Samples were cooled (4 °C) and homogenized immediately after collection. An aliquot of each sample was placed into 10 mL of 20 mM sodium phosphate buffer containing 10% ethanol, 1% acidic phenol (pH 4.5), and 1% β-mercaptoethanol for future RNA extraction. All samples were frozen within 30 min of collection on dry ice and kept at −80 °C.

Isolation of gDNA and total RNA

Total gDNA was isolated from fecal samples using the ZR Fecal DNA Isolation kit (Zymo Research Corporation) according to the manufacturer's protocol.

For RNA isolation, an aliquot of stool/preservative mixture was thawed, centrifuged for 5 min, and the supernatant was removed. A total of 150 mg of stool pellet was used for RNA extraction. Stool pellets were resuspended in cold RNA Protect Bacterial Reagent (Qiagen Inc.) and allowed to incubate for 10 min on ice. After removing the RNA Protect reagent, stool samples were resuspended in 1.5 mL of cold Stool Resuspension Buffer (20 mM sodium phosphate buffer, pH 6.0, 50 mM NaCl, 20 mM EDTA, 0.5% sodium dodecyl sulfate, 10% buffered phenol, 1% β-mercaptoethanol). Tubes were vortexed for 5 min at maximum speed, and then kept in a 95 °C sandbath for 1 min. The bacteria were broken with ZR Bashing Beads (Zymo Research Corporation) by vortexing for 2 min at maximum speed. After centrifugation, the supernatant was collected, and nucleic acid extraction was carried out by the phenol–chloroform method (Gyaneshwar *et al.*, 2005). Extracted nucleic acids were passed through ZR-HRC-IV columns (Zymo Research Corporation) to remove enzyme inhibitors often present in fecal matter. DNA was removed by incubating samples with 10 U of DNase I for 40 min at 37 °C. RNA was purified with an RNeasy RNA purification kit (Qiagen Inc.). RNA quality was assessed by agarose gel electrophoresis.

Microarray experiments and data analysis

Taq 2× Master Mix (New England Biolabs) was used to amplify total gDNA with 16S rDNA-specific conserved primers Bact-27F (5′-AGRGTTYGATYMTGGCTCAG-3′) and Univ-1492R (5′-GGYTACCTTGTTACGACTT-3′). A total of 250 ng of starting material was used in each individual PCR reaction; the number of PCR amplification cycles varied as described in Results. At least four separate PCR reactions were run for each set of conditions; the reactions were pooled, and then purified with a QIAquick PCR purification kit (Qiagen Inc.). To obtain sufficient quantities of amplified DNA, a higher number of individual PCR reactions sometimes had to be carried out for experiments utilizing only 15 and 20 cycles of PCR amplification. A custom-developed Microbiota Array based on the AFFYMETRIX GENECHIP platform (Paliy *et al.*, 2009) was utilized for all microarray experiments. Microarray sample preparation and microarray hybridization experiments were carried out

as described (Paliy *et al.*, 2009). A total of 1500 ng of amplified and fragmented gDNA were loaded onto each microarray.

For RNA analysis, RNA was first reverse transcribed into cDNA essentially as described (Gunasekera *et al.*, 2008) with the following modifications: 2 μL of Superscript III reverse transcriptase (Invitrogen Corporation) was used; the amplification cycle consisted of 25 °C for 10 min, 37 °C for 30 min, 42 °C for 30 min, 50 °C for 1 h, and 70 °C for 15 min. The typical cDNA yield was about 10% of the starting RNA amount on average. cDNA was further processed and hybridized to microarrays as described (Gunasekera *et al.*, 2008). A total of 1500 ng of fragmented cDNA were loaded onto each microarray. Two replicate arrays were run for each test sample utilizing cDNA prepared in independent PCR reactions.

Comprehensive step-by-step microarray analysis procedure is shown in Fig. 1. Raw microarray data were analyzed in GCOS with a standard MAS5 detection algorithm (Affymetrix, Inc.) to obtain species detection calls (parameters used were $\alpha_1 = 0.03$, $\alpha_2 = 0.05$, and $\tau = 0.015$; this corresponds to 97% confidence in the positive detection call for each individual species) (Paliy *et al.*, 2009). When multiple replicate experiments were run for the same sample, a particular probeset had to be called present in each replicate in order to be classified as detected. To obtain hybridization signal estimates for each probeset, raw data files were imported into CARMAWEB, and MAS5-VSN-MAS5-MedianPolish procedure was utilized to normalize the data within and across experiments (Rainer *et al.*, 2006). A detailed description of this normalization procedure is available elsewhere (Choe *et al.*, 2005; Paliy *et al.*, 2009). Further data analysis was performed in MICROSOFT EXCEL.

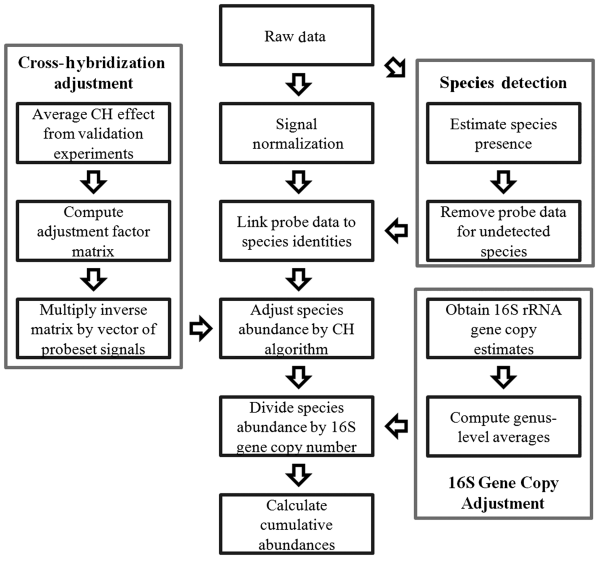

Fig. 1. Step-by-step procedure used for the analysis of microarray data obtained on Microbiota Array. CH, cross-hybridization.

The normalized signal values were adjusted by cross-hybridization and by 16S rRNA gene copy numbers as described below. If a particular phylo-species was represented by multiple probesets on the Microbiota Array, the consensus signal was obtained as described (Paliy *et al.*, 2009). Cumulative signal for higher level phylogenetic groups was calculated as a sum of adjusted signals of all detected species within that phylogenetic group. The estimated abundance of each phylogenetic group in the sample was expressed relative to the total adjusted signal for all detected phylo-species; signal from species called absent by GCOS detection algorithm was never used.

Signal adjustments

Cross-hybridization (CH) adjustment for each probeset was modeled as Total.Signal = True.Signal + ΣCH.Signal, where Total.Signal is a measured signal intensity for a particular probeset, True.Signal is the true hybridization signal due to correct binding of species-specific 16S rDNA fragment(s), and CH.Signal is cross-hybridization signal due to incorrect binding of 16S rDNA sequences corresponding to the other detected species. The CH.Signal contribution to Total.Signal by any given cross-hybridized 16S rDNA fragment is a product of the True.Signal intensity for this cross-hybridized segment and an adjustment factor reflecting the extent to which cross-hybridization occurs. Thus, to expand the above equation to a full dataset, it was written as a system of linear algebraic equations, $b = Ax$, where b is the vector of Total.Signals, x is the vector of True.Signals, and A is a square matrix (775×775) representing the contribution that True.Signal of each species makes to Total.Signal of all species, and thus includes the cross-hybridization effects. Solving for the vector of True.Signals is then carried out by $x = A^{-1}b$ (an additional explanation of the principle behind cross-hybridization adjustment is provided in the Supporting Information). An average cross-hybridization effect was calculated from the results of previously published validation experiments of the Microbiota Array by dividing the sum of measured signals for all cross-hybridizing probesets in each experiment by the total number of such probesets on the microarray (Paliy *et al.*, 2009). To avoid magnifying the ratio of signals among samples and probesets through division by a very small denominator, a small constant equal to the average noise level in microarray experiments (1.0) was added to each calculated True.Signal. Occasionally, cross-hybridization adjustment produced a probeset True.Signal value below 0; in these cases, that signal was set to 1. A custom MICROSOFT EXCEL template file was used to automatically compute the cross-hybridization-adjusted values on the imported normalized microarray data. The linear

algebraic manipulations were also executed in MATLAB (version 7.3, The Mathworks Inc.) to verify that the matrix inversion process was not close to singular.

To carry out 16S rRNA gene copy adjustment procedure, average and median 16S rRNA gene copy numbers were estimated for each genus based on the 16S rRNA gene numbers reported in rRNDB and NCBI microbial genomes databases (Lee *et al.*, 2009). When no species were listed for a particular genus, average 16S rRNA gene copy for a corresponding family or order was used as an estimate. The cross-hybridization-adjusted signal values for each probeset were divided by an average number of 16S rRNA gene copies for the corresponding genus.

Quantitative real-time PCR (qPCR)

qPCR was carried out on an ABI Prism 7000 Sequence Detection System using Platinum SYBR Green qPCR supermix (Invitrogen Corporation) essentially as described (Paliy *et al.*, 2009), including mathematical calculations taking into account unequal amplification efficiency for different primer pairs. Group-specific primers used to amplify selected 16S rDNA sequences are listed in the Supporting Information.

Statistical procedures

All statistical tests were carried out in SPSS v16 (SPSS Inc.). Nonparametric Spearman rank test was used for all correlation analyses because the datasets were not found to be normally distributed. Principle components analysis was performed in MATLAB.

Results

Detection of bacterial presence in human fecal samples as a function of the number of cycles of 16S rDNA-specific PCR amplification

Fecal gDNA isolated from two healthy adult volunteers was used to assess how the ability of Microbiota Array to detect bacterial presence depends on the number of 16S rDNA-specific PCR amplification cycles performed on fecal DNA. PCR amplification was carried out for 15, 20, 25, and 30 cycles, respectively, and the amplified DNA was loaded onto individual microarrays. The measurements obtained were normalized as described in Materials and methods and the data were compared among different amplifications. The level of species detection was similar between 25 and 30 cycles of PCR amplification (Table 1); however, a reduction in the number of species detected was observed for samples that were subjected to 20 and 15 cycles of PCR amplification. This finding is likely explained by the fact that each hybridization sample consisted of the original total gDNA and amplified 16S rDNA pool, and because we loaded the same amount of fragmented DNA onto each array, the relative fraction of the amplified pool in the sample becomes higher with an increasing number of 16S rDNA-specific amplification cycles. Interestingly, a few previous studies have reported that an increase in the number of amplification cycles used to construct rRNA clone libraries led to either reduced estimates of community diversity (Bonnet *et al.*, 2002) or no change in diversity estimates (Acinas *et al.*, 2005). The discrepancy between these findings and our results is likely (1) due to the different technologies utilized in the analyses (microarrays with defined set of probes vs. 16S rRNA gene sequencing) and the specifics of our hybridization approach, (2) because our comparisons were made among samples subjected to low number of PCR amplification cycles, or (3) they can be potentially explained by the generally low PCR bias experienced in our amplification experiments.

Table 1. Detection of bacteria by phylogenetic microarray as a function of 16S rDNA-specific PCR amplification of fecal genomic DNA

No. of PCR cycles	Species detected*
30	392 ± 3
25	389 ± 1
20	354 ± 29
15	322 ± 7

*Data are shown as arithmetic mean $\pm$ SEM.

Influence of PCR amplification bias on the measured relative abundance of bacterial species

Because of the differences among bacterial species in genome GC content, varied relative abundance of different bacteria in the sample, and variation in annealing affinity of 'universal' amplification primers to individual 16S rDNA sequences, multiple cycles of PCR amplification can introduce a noticeable bias into the measured ratios among bacteria (Polz & Cavanaugh, 1998; Kurata *et al.*, 2004; Sachse, 2004). To test whether such potential PCR bias was evident in our experiments, we compared relative abundances of each bacterial species, genus, and class among experiments that interrogated gDNA subjected to 15, 20, 25, and 30 cycles of PCR amplification.

Considering the measured signal across different bacterial species, good correlation was observed among different amplification cycle experiments of each sample (average Spearman rank correlation $R_S = 0.86$, $P < 0.01$). Comparisons of bacterial abundances among different amplification conditions at the class level indicated that in general the Microbiota Array produced stable estimates of the relative bacterial amounts (see Supporting Information, Table S1). The differences that were observed included a modest

decrease of *Actinobacteria* abundance and a modest increase in *Bacteroidetes* and *Mollicutes* abundances as the number of PCR amplification cycles increased. Principle components analysis revealed that experiments carried out on each sample clustered together and separately from the experiments performed on the other sample (see Fig. S1), indicating that sample-specific variation in bacterial abundance rather than the number of PCR amplification cycles contributed most to the differences observed among experiments.

To quantitatively assess any detected PCR bias, bacterial abundance at the genus level was evaluated: species-specific abundance tended to be less stable, whereas class and order-level signal often masked important differences in the abundance values evident at lower phylogenetic levels. A complete list of relative abundances of all genera among 15, 20, 25, and 30 PCR cycle experiments is available as Table S2. Among 115 different genera that can be interrogated by Microbiota Array, examination of genus abundances provided evidence of PCR bias for 23 of them. Seventeen genera had a tendency to increase in relative abundance when the number of amplification cycles was increased (positive bias), whereas six genera decreased their relative abundance with an increase in the number of PCR amplification steps (negative bias). Most of the genera with positive bias were present at a relatively low abundance level (average relative abundance across different experiments 0.83%), whereas genera with negative bias were well-represented in each sample (6.29% on average). Indeed, as can be seen from Table 2, genera that each were present at $< 1\%$ relative abundance on average showed a general boost in the measured abundance with an increase in the number of 16S rDNA-specific PCR amplification cycles. The opposite was observed for the genera present at $> 5\%$ relative abundance level. Interestingly, even though both genera *Bacteroides* (class *Bacteroidetes*) and *Papillibacter* (class *Clostridia*) were expressed at 4–5% abundance level, they displayed a consistent positive bias (3.6/4.0% relative abundance after 15 cycles, 5.2/5.8% after 30 cycles of PCR amplification, respectively). Apart from these two exceptions, no consistent bias was evident for genera with intermediate representation between 1% and 5%. The calculated bias did not reveal any relationship to the genome GC content: average GC content was 48.9% for genera with a positive bias, 45.5% for genera with no bias detected, and 47.5% for genera with a negative bias. Because positive PCR bias was associated with genera of low abundance, and because distribution of genus abundances in each sample displays hyperbolic shape with many low-abundance members (Paliy *et al.*, 2009), this can explain why we detected a significantly higher number of genera with positive bias. The noted distribution of the PCR biases is consistent with previous observations of low-abundance members becoming gradually over-represented in the mixture during PCR amplification due to preferential reannealing of highly abundant template species during the annealing step of PCR reaction effectively lowering their amplification rate (Kurata *et al.*, 2004; Acinas *et al.*, 2005).

We calculated an average PCR bias per amplification cycle by comparing relative abundances of each genus at 20, 25, and 30 cycles (15 cycles were not used because fewer microbiota species were detected with those conditions). In these calculations, the bias was defined as a deviation of specific amplification efficiency from the average amplification efficiency of PCR reaction for that sample divided by this average amplification efficiency. The distribution of bias could be modeled by a hyperbolic function, where the majority of the genera had low PCR bias (see Fig. S2). A median bias per PCR amplification cycle was 1.59% among the genera that were detected in all samples. This bias amount is relatively small: it will lead to a 1.43-fold deviation of the relative amount of bacterial 16S rDNA after 25 cycles of PCR amplification.

Adjusting measured signal to account for DNA cross-hybridization

Because phylogenetic microarrays study microbial population composition based on the sequence variability of a single gene encoding 16S rRNA, there is a substantial potential for cross-hybridization of DNA fragments in the interrogated mixture to incorrect probes. To investigate the level of such cross-hybridization evident in the use of Microbiota Array, we evaluated experiments where pure bacterial gDNA isolated from 15 different bacterial species was hybridized to separate arrays (Paliy *et al.*, 2009). The microarray correctly identified the presence of each species in the corresponding sample, and the corresponding probeset had the highest measured signal. However, a number of additional probesets had a signal above the noise. Although we showed that such a cross-hybridization signal does not

Table 2. Bias in measured relative abundance of bacterial genera as a function of the number of 16S rDNA-specific PCR amplification cycles

Genus abundance	30 cycles	25 cycles	20 cycles	15 cycles
$< 1\%$ ($n = 66$)	$0.24 \pm 0.01\%$	$0.21 \pm 0.01\%$	$0.19 \pm 0.01\%$	$0.16 \pm 0.01\%$
1–5% ($n = 14$)	$2.78 \pm 0.09\%$	$2.78 \pm 0.09\%$	$2.75 \pm 0.08\%$	$2.70 \pm 0.08\%$
$> 5\%$ ($n = 6$)	$9.04 \pm 1.32\%$	$9.41 \pm 1.23\%$	$9.85 \pm 1.36\%$	$10.33 \pm 1.61\%$

Data are shown as arithmetic mean $\pm$ SEM.

influence the estimates of species presence when multiple replicate arrays are run for the same sample (Paliy *et al.*, 2009), the measured signals might be overestimated when many species are found in a fecal sample. We now carried out calculations of the measured signal for cross-hybridizing probesets in those microarray analyses of samples of defined composition. On average, cross-hybridization signal for probesets representing species within the same genus as the correctly identified species constituted 1.58% of that species signal (SD 3.22%); for probesets outside that genus, the average cross-hybridization signal was only 0.19% (SD 0.36%). Because it is not feasible to measure cross-hybridization pattern for each of the 775 bacterial phylo-species that can be interrogated by the Microbiota Array, we modeled the total measured signal of every probeset as a combination of true hybridization signal due to correct complementary binding of that species 16S rDNA fragments and a sum of cross-hybridization signals due to incorrectly bound rDNA. As described in Materials and methods, the problem of finding true signals for each probeset when we can only measure the total signals was solved through an inverse matrix calculation. To simplify the calculation process, cross-hybridization signal from DNA of species outside the same genus was considered to be insignificant and this cross-hybridization adjustment was set to zero.

Table 3 presents relative abundances of bacterial classes after cross-hybridization adjustment of the signal. Because of limiting the cross-hybridization to the species within the same genus, the adjustment results in a relative decrease in the overall signal of genera with many members (*Ruminococcus*, *Papillibacter*, *Roseburia*, *Bacteroides*) at the same time increasing the relative representation of genera with few members. For example, relative contribution of *Ruminococcus* (the genus with the highest number of species members) went down from 28% of the nonadjusted signal to 22% after cross-hybridization adjustment. Because several of the genera with a large number of members belong to class *Clostridia*, this results in a relative decrease in the abundance of this class.

Adjusting measured signal to account for multiple genomic 16S rRNA gene copies

Different species of bacteria are known to possess a wide range of rRNA-encoding gene copies per genome, with current estimates of between 1 and 15 genes (Lee *et al.*, 2009). The different copies of 16S rRNA genes can differ in their nucleotide sequence; however, such sequence divergence is rarely $>$ 1–2% nucleotide dissimilarity (Acinas *et al.*, 2004). As a result, most phylogenetic microarrays define all 16S rRNA sequences within 2% of sequence identity of each other as representing the same species. With all other variables being equal, a particular probeset interrogating 16S rDNA from a bacterial species with 10 16S rRNA gene copies in its genome would produce five times the signal measured by a probeset interrogating the genome of a species with only two 16S rRNA genes. This presents problems for translating the signal measured by a phylogenetic microarray into an estimate of species abundance.

We developed an adjustment algorithm for the measured probeset signal that takes into account an average 16S rRNA

Table 3. Influence of signal adjustments on relative abundance of bacterial classes

	Sample 1: signal adjustment			Sample 2: signal adjustment		
Class	None (%)*	CH (%)†	CH+16S copy (%)‡	None (%)*	CH (%)†	CH+16S copy (%)‡
Alphaproteobacteria	–	–	–	<0.1	<0.1	<0.1
Betaproteobacteria	1.1	1.3	3.1	1.2	1.4	3.5
Gammaproteobacteria	0.8	1.0	1.1	0.5	0.6	0.8
Deltaproteobacteria	0.5	0.6	0.9	0.5	0.6	1.0
Epsilonproteobacteria	–	–	–	–	–	–
Clostridia	79.8	77.1	64.4	80.9	78.4	67.7
Mollicutes	0.2	0.2	1.0	0.4	0.5	2.2
Bacilli	2.0	2.3	2.8	5.1	6.1	8.2
Actinobacteria	5.9	6.7	13.3	0.7	0.9	2.2
Spirochaetes	<0.1	<0.1	0.1	<0.1	<0.1	0.1
Bacteroidetes	9.0	9.8	11.6	10.3	11.1	13.7
Verrucomicrobiae	0.6	0.7	1.3	0.3	0.4	0.7
Lentisphaerae	0.1	0.1	0.5	<0.1	<0.1	<0.1

Comparisons were carried out for samples subjected to 25 cycles of PCR amplification.

*No adjustment was applied to normalized signal values.

†Signal was adjusted by the cross-hybridization algorithm.

‡Signal was first adjusted by the cross-hybridization algorithm, and then by the average number of 16S rRNA gene copies per bacterial genome.

–, none detected.

gene copy number per bacterial genome. rRNDB and NCBI microbial genome databases were searched for the number of 16S rRNA gene copies among species of each genus; if the data from multiple species of the same genus were available, an average gene copy number was calculated. The genus-level 16S rRNA gene copy table is available as Table S3. The average genus 16S rRNA gene copy number ranged from 1 for *Brachyspira* to 9.4 for *Clostridium*. Members of *Clostridia* had relatively high number of 16S rRNA gene copies (range 2–15, average between 8 and 9), whereas other prominent classes of intestinal bacteria had fewer 16S rRNA genes on average (*Bacteroidetes*, about 6; *Actinobacteria*, 3–4; different classes of *Proteobacteria*, between 3 and 7). The normalized signal for each probeset, first adjusted by the cross-hybridization algorithm as described above, was then divided by the average number of 16S rRNA gene copies calculated for the corresponding genus. As can be seen from Table 3, the adjustment led to a noticeable change in relative abundance values. *Clostridia* representation decreased by 1/5th–1/6th, whereas relative abundances of most other classes increased. Comparing our microarray results before and after 16S rRNA copy adjustment with several recent analyses of microbiota composition of healthy adults performed by counting bacterial cells visualized by FISH (Rigottier-Gois *et al.*, 2003; Lay *et al.*, 2005; Sokol *et al.*, 2006), we observe that our 16S gene copy and cross-hybridization-algorithm adjusted numbers are in better agreement with the FISH data. The ratio among three major intestinal bacterial classes *Clostridia*, *Actinobacteria*, and *Bacteroidetes* was 74%–11%–15% when bacteria were visualized with fluorescent probes (average among three studies, see Table S4 for all the relevant data); this is closer to our gene copy adjusted averages of 76%–9%–15% compared with the 85%–4%–11% relative signal distribution for these classes when no 16S rRNA gene copy and no cross-hybridization adjustments were applied.

Measuring fecal RNA to assess microbiota metabolic activity

The use of selective 16S rDNA-specific PCR amplification combined with the relative ease of obtaining total gDNA from community samples has led to a substantial number of publications elucidating the population structure of complex microbial communities (Eckburg *et al.*, 2005; Brodie *et al.*, 2007; Gao *et al.*, 2007; Palmer *et al.*, 2007). The relative metabolic activities of individual community members, on the other hand, have received little attention. Indeed, whereas profiling gDNA can help us understand which bacteria are present in the community, this approach cannot be used to differentiate between metabolically active and metabolically inert (or even dead) members. However, because the metabolic activity of the species is generally linked to that species' growth rate in the environment, which in turn in most cases is proportional to the number of ribosomes and thus to the amount of rRNA produced by each cell (Bremer & Dennis, 1996; Stevenson & Schmidt, 2004; Zaslaver *et al.*, 2009; also see Binder & Liu, 1998; Klappenbach *et al.*, 2000; Schmid *et al.*, 2001 for exceptions to this general correlation), interrogation of 16S rRNA can allow us to obtain such estimates. With this in mind, the original design of Microbiota Array was carried out to allow the microarray to be used for interrogation of either DNA or RNA (Paliy *et al.*, 2009), and we have now explored the use of this microarray to measure the relative metabolic activity of individual members of intestinal microbiota.

Total fecal RNA was isolated from two adult stool samples, processed, and subsequently hybridized to the microbiota array. Because gDNA was already profiled from the same samples, we could compare bacterial detection between gDNA and RNA samples. Comparable numbers of bacteria from the same sample were detected between RNA and gDNA sources even though no amplification of 16S rRNA was carried out (sample 1: 355 ± 29 detected species from gDNA, 336 ± 21 from RNA; sample 2: 374 ± 21 species from gDNA, 427 ± 8 from RNA). As expected, calculations of the Spearman rank correlations among experiments showed that bacterial RNA abundances matched strongly the distribution of bacterial DNA abundances from the same fecal sample (Fig. 2).

Comparing metabolic activities of different bacterial classes (Table 4), known starch and fiber degraders (*Clostridia* and *Bacteroidetes*) were relatively active in both samples, whereas *Mollicutes*, *Bacilli*, *Actinobacteria*, and *Spirochaetes* had lower rRNA counts, which could signify their relative metabolic quiescence (inactivity). Figure 3 shows a scatter plot comparing relative RNA and DNA abundances at the genus level. Most genera clustered around

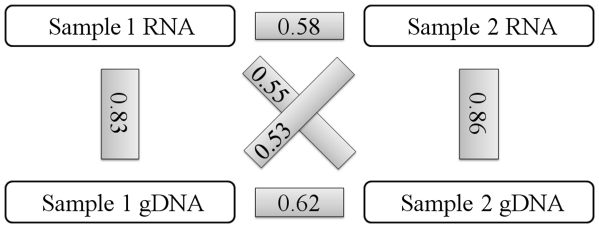

Fig. 2. Correlation among measured microarray signals obtained in the interrogation of total RNA and gDNA**.** gDNA signals for each sample were calculated as an average from experiments using gDNA subjected to 15, 20, 25, and 30 cycles of PCR amplification. Normalized DNA signals were cross-hybridization- and 16S copy-adjusted (see footnote to Table 3); normalized RNA signals were cross-hybridization-adjusted. Values shown are Spearman rank correlation coefficients for each pairwise comparison; the width of the box is proportional to the correlation coefficient.

Table 4. Relative measured amounts of genomic DNA and total RNA for different bacterial classes

	Sample 1		Sample 2	
Class	RNA (%)*	DNA (%)†	RNA (%)*	DNA (%)†
Alphaproteobacteria	–	–	–	< 0.1
Betaproteobacteria	1.8	3.0	1.8	2.9
Gammaproteobacteria	1.2	0.7	0.9	0.7
Deltaproteobacteria	1.0	0.7	1.3	0.7
Epsilonproteobacteria	–	–	–	–
Clostridia	75.5	65.9	75.5	70.6
Mollicutes	0.3	0.6	0.7	1.3
Bacilli	1.3	1.9	5.7	9.8
Actinobacteria	5.0	15.5	1.0	1.4
Spirochaetes	–	0.4	< 0.1	0.4
Bacteroidetes	12.5	9.6	12.4	11.6
Verrucomicrobiae	1.3	1.5	0.9	0.6
Lentisphaerae	0.1	0.3	–	< 0.1

*Measured signal was adjusted with cross-hybridization algorithm.
†Measured signal was adjusted with cross-hybridization and 16S gene copy algorithms. Values shown are averages of the experiments with 15, 20, 25, and 30 cycles of PCR amplification.
–, none detected.

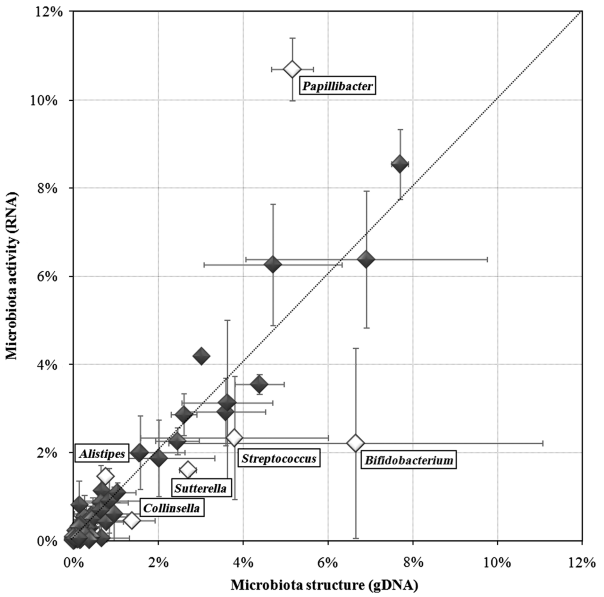

Fig. 3. Comparison of relative gDNA and rRNA sample compositions across bacterial genera. *X* axis displays gDNA abundance of each genus relative to the total signal (average from experiments using gDNA subjected to 15, 20, 25, and 30 cycles of PCR amplification); *Y* axis shows relative rRNA abundance. Each diamond represents an average abundance of a particular genus between two samples. Error bars show SEM. Genus *Ruminococcus* (21.6% gDNA abundance, 22.1% rRNA abundance) is not shown. Examples of genera with disproportional metabolic activity are shown with unfilled diamonds.

the diagonal line indicating that their total metabolic activity corresponded to their relative cell count as estimated by genomic rDNA amount. However, several genera had significantly larger RNA fraction than their DNA abundance would suggest. These included *Papillibacter* (10.7% relative RNA abundance vs. 5.2% DNA abundance), *Sporobacter* (0.8% RNA and 0.1% DNA), *Acidaminobacter* (all in class *Clostridia*, 0.3% RNA, and 0.1% DNA), and *Alistipes* (class *Bacteroidetes*, 1.5% RNA and 0.8% DNA). On the other hand, genera *Bifidobacterium* (2.2% RNA and 6.7% DNA), *Collinsella* (both *Actinobacteria*, 0.4% RNA and 1.4% DNA), *Streptococcus* (*Bacilli*, 2.3% RNA and 3.8% DNA), *Brachyspira* (*Spirochaetes*, < 0.1% RNA and 0.4% DNA), and *Holdemania* (*Mollicutes*, 0.4% RNA and 0.8% DNA) displayed decreased RNA abundance.

To obtain a relative activity index of each bacterial genus in fecal samples examined, we divided the relative abundance of RNA (expressed as percent of total measured signal) by the relative abundance of DNA for each genus. Because we used two separate samples in these experiments, we could ask whether relative metabolic activity of different genera is sample/person specific by comparing the activity indexes of each genus between samples. Although the majority of genera clustered around 1 : 1 ratio (indicating 'average' metabolic index in both samples), we found several interesting examples (see Fig. 3 and Fig. S3). Two genera of *Actinobacteria* – *Bifidobacterium* and *Collinsella* – had low metabolic activity in both samples (defined as activity index of 0.5 or lower in each sample); at the same time, genera *Papillibacter*, *Sporobacter*, and *Alistipes* were highly metabolically active in both samples (activity index ≥ 2). Comparison of samples also revealed four genera that differed in their metabolic activity between these two samples: *Turicibacter* (class *Bacilli*), *Parasporobacterium*, *Coprobacillus* (both *Clostridia*), and *Eggerthella* (class *Actinobacteria*). A simultaneous analysis of RNA and DNA also identified two bacterial genera – *Limnobacter* (class *Betaproteobacteria*) and *Bryantella* (class *Clostridia*) – which despite being present at the same cell abundance level in both samples had different metabolic activity between two samples. Note that because we only profiled two different samples in these analyses, any drawn conclusions were not tested for statistical significance.

We have used qPCR to validate these findings revealed by the Microbiota Array. As can be assessed in Table 5, qPCR results matched microarray data well, with only slight deviation observed for *Faecalibacterium* genus abundance estimates.

A recent study examined the metatranscriptome of intestinal microbiota in healthy adults utilizing the deep sequencing approach (Turnbaugh *et al.*, 2010). Similar to our findings, overall gene expression was found to be low for members of genera *Collinsella* and *Bifidobacterium*, whereas an abundant expression of genes was found for members of *Alistipes*. Many genera of class *Clostridia* had an 'average' gene expression proportionate to the numbers of these

Table 5. Validation of microarray results of RNA vs gDNA comparison with quantitative PCR

	RNA/gDNA signal ratio[†]	
Bacterial group[*]	Array[‡]	qPCR[§]
Faecalibacterium	0.84	1.39 ± 0.70
Papillibacter	1.83	1.82 ± 0.91
Bifidobacterium	0.61	0.56 ± 0.27
Collinsella	0.69	0.80 ± 0.36

[*]Primers were developed to cover most of the 16S rDNA sequences in the RDP database belonging to each group.
[†]Numbers represent ratios of total unadjusted signal between RNA and gDNA preparations of sample 1.
[‡]DNA signal was calculated as an average of the experiments with 15, 20, 25, and 30 cycles of PCR amplification.
[§]Data are shown as arithmetic mean ± SEM.

bacteria in the samples – this also matches well with our results of the RNA vs. gDNA comparisons. On the other hand, genus *Bacteroides* was highly active in the samples analyzed by Turnbaugh and colleagues, whereas we found only a slight increase in their metabolic activity over average in our samples.

Discussion

Phylogenetic microarrays are an attractive tool to obtain high-throughput quantitative measurements of microbial abundance in complex communities. However, in order to use any DNA interrogating technique to estimate community structure, one has to be certain of a good concordance between measured values and the true species distribution. The analyses described in this article were intended to improve the robustness of the measured estimates of microbial abundance as well as to provide a new approach to community interrogation.

By evaluating microarray hybridization signal in samples subjected to 15, 20, 25, or 30 cycles of 16S rDNA-specific PCR amplification, we observed a modest amount of measured signal bias as a function of the number of amplification cycles. Because of this PCR bias, several highly abundant genera decreased their representation as the number of amplification cycles increased, while the opposite was evident for some but not all low-abundance members. The protocol for PCR amplification that we used already incorporated features aimed to decrease the amplification bias [low annealing temperature (Sipos *et al.*, 2007), fast ramp down between denaturing and annealing steps (Kurata *et al.*, 2004), pooling together multiple PCR reactions (Polz & Cavanaugh, 1998; Acinas *et al.*, 2005), and low number of amplification cycles (Polz & Cavanaugh, 1998)]; this can explain why the amplification bias measured in our experiments was lower than that reported previously (Polz & Cavanaugh, 1998; Kurata *et al.*, 2004).

We developed two adjustment algorithms intended to improve the concordance between initial bacterial numbers in the community and the distribution of measured DNA hybridization signals. Both algorithms have been incorporated into our analysis workflow as shown in Fig. 1. Whereas cross-hybridization adjustment had an influence mostly on bacterial groups with large numbers of species, profound changes were observed after adjusting measured signals by the number of 16S rRNA gene copies per bacterial genome. Comparison of average 16S rRNA gene copies for different genera and classes indicated that the numbers of *Clostridia* might be overestimated in many studies of intestinal microbiota using DNA sequence interrogation techniques. Indeed, we found that there is a significant positive correlation between the average number of 16S rRNA gene copies per genome of a species in a particular genus, and the number of corresponding 16S rRNA sequences of human intestinal bacteria deposited into the NCBI database (Spearman rank correlation $R_S = 0.39$, $P < 0.01$; see Fig. S4).

The limitation in the use of both adjustment algorithms is their generalized estimation of adjustment factors for different species within the same genus. Although species of the same genus can have different numbers of rRNA operon copies in their genomes, and cross-hybridization by DNA fragments from a particular species is often limited to just a few other species within the same genus (Paliy *et al.*, 2009), the lack of species-specific data for the vast majority of intestinal bacteria prevented us from specifying adjustment factors uniquely for each species. Nevertheless, as shown in Results, both algorithms are likely to provide an improvement to the estimated association between measured microarray signals and actual community composition. Current large-scale sequencing efforts such as the one initiated by the Human Microbiome Project (Turnbaugh *et al.*, 2007) are expected to provide genome information necessary to eventually extend both algorithms to a species-level precision.

Finally, we also established a protocol to utilize the Microbiota Array for the interrogation of relative metabolic activities of community members. Whereas analysis of gDNA can provide us information on the number of bacterial cells present in a particular community, the measurements of the 16S rRNA content of the community can give us estimates of the relative metabolic activity of different community members (see Results for explanation of association and its limitations). 16S rRNA profiling can also reveal changes in the metabolic activity of individual microbial species over time even when the counts of bacteria do not change. A number of previous reports revealed that RNA profiles did not match DNA profiles in the same microbial community (Tannock *et al.*, 2004; Licht *et al.*, 2006); however, prior studies could not obtain relative DNA and RNA signal distributions for individual community

members. Complementary analysis of RNA and DNA abundances in our pilot experiments exposed a number of cases where a significant discrepancy between RNA and DNA amounts was detected. Apart from being an indication of bacterial metabolic activity, several other factors can cause the observed differences between RNA and DNA abundances. Because no amplification of RNA was necessary in our experiments due to inherently large numbers of 16S rRNA molecules per cell (Bremer & Dennis, 1996), the observed discrepancies can potentially be a function of the 16S rDNA amplification bias. However, the observed differences between RNA and DNA abundances were specific to particular genera rather than widespread, and those differences uncovered were significantly higher than an estimated average PCR bias can account for (up to eightfold RNA-vs.-DNA ratios calculated; median PCR bias produces 1.43-fold deviation after 25 amplification cycles). Alternatively, because gDNA is much more stable than RNA, low RNA abundance might potentially be attributed to the dead cells that lost viability either due to exposure to air, or because of the long transit time in human large intestine and the corresponding lack of nutrients in fecal matter. However, this is not sufficient to explain a specific high RNA abundance of members of *Papillibacter*, obligate anaerobic bacteria, in both of our samples. Further studies will be needed to evaluate potential influence of these mitigating factors on the RNA-vs.-DNA abundance relationships.

If we can assume that measured rRNA abundance is proportional to the bacterial metabolic activity, members of the genus *Papillibacter* were very metabolically active in the examined samples and might therefore play a more prominent than previously recognized role in defining the overall metabolism of intestinal microbiota. At the same time, genera *Limnobacter* and *Bryantella* had different metabolic activities in two interrogated samples, although their cell numbers were estimated to be the same. Because only few samples were used for these pilot analyses, at the moment we are unable to generalize our findings, because the observed differential metabolic activities might be sample-specific [note, however, that our data match well with another recent study (Turnbaugh *et al*., 2010) profiling bacterial activity and gene expression in the gut]. Nevertheless, such examples demonstrate the power of simultaneous profiling of gDNA and RNA isolated from the same sample to assess not only community structure but also relative metabolic activity. Several intestinal diseases are believed to be associated with a misbalanced metabolic state of intestinal microbiota (van Nuenen *et al*., 2004; Turnbaugh *et al*., 2006); therefore, examination of community RNA might reveal putative associations between bacterial activity and disease status that might be missed when only gDNA is profiled to reveal bacterial abundance.

Acknowledgements

Our thanks to Michael Markey and Sameep Naik for valuable comments on the manuscript and to the members of the WSU Center for Genomics Research for access to the facility. This work was supported by the National Institutes of Health grant AT003423 to O.P.

References

Acinas SG, Marcelino LA, Klepac-Ceraj V & Polz MF (2004) Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. *J Bacteriol* **186**: 2629–2635.

Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V & Polz MF (2005) PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microb* **71**: 8966–8969.

Belenguer A, Duncan SH, Calder AG, Holtrop G, Louis P, Lobley GE & Flint HJ (2006) Two routes of metabolic cross-feeding between *Bifidobacterium adolescentis* and butyrate-producing anaerobes from the human gut. *Appl Environ Microb* **72**: 3593–3599.

Binder BJ & Liu YC (1998) Growth rate regulation of rRNA content of a marine synechococcus (*Cyanobacterium*) strain. *Appl Environ Microb* **64**: 3346–3351.

Bonnet R, Suau A, Dore J, Gibson GR & Collins MD (2002) Differences in rDNA libraries of faecal bacteria derived from 10- and 25-cycle PCRs. *Int J Syst Evol Micr* **52**: 757–763.

Bremer H & Dennis PP (1996) Modulation of chemical composition and other parameters of the cell by growth rate. *Escherichia Coli and Salmonella: Cellular and Molecular Biology, Vol. 2* (Neidhardt FC, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M & Umbarger HE, eds), pp. 1553–1569. ASM Press, Washington, DC.

Brodie EL, DeSantis TZ, Parker JP, Zubietta IX, Piceno YM & Andersen GL (2007) Urban aerosols harbor diverse and dynamic bacterial populations. *P Natl Acad Sci USA* **104**: 299–304.

Cannone JJ, Subramanian S, Schnare MN *et al*. (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3**: 2.

Carey CM, Kirk JL, Ojha S & Kostrzynska M (2007) Current and future uses of real-time polymerase chain reaction and microarrays in the study of intestinal microbiota, and probiotic use and effectiveness. *Can J Microbiol* **53**: 537–550.

Choe SE, Boutros M, Michelson AM, Church GM & Halfon MS (2005) Preferred analysis methods for Affymetrix GeneChips revealed by a wholly defined control dataset. *Genome Biol* **6**: R16.

Eckburg PB, Bik EM, Bernstein CN *et al*. (2005) Diversity of the human intestinal microbial flora. *Science* **308**: 1635–1638.

Gao Z, Tseng CH, Pei Z & Blaser MJ (2007) Molecular analysis of human forearm superficial skin bacterial biota. *P Natl Acad Sci USA* **104**: 2927–2932.

Gill SR, Pop M, Deboy RT *et al.* (2006) Metagenomic analysis of the human distal gut microbiome. *Science* **312**: 1355–1359.

Gunasekera TS, Csonka LN & Paliy O (2008) Genome-wide transcriptional responses of *Escherichia coli* k-12 to continuous osmotic and heat stresses. *J Bacteriol* **190**: 3712–3720.

Gyaneshwar P, Paliy O, McAuliffe J, Popham DL, Jordan MI & Kustu S (2005) Sulfur and nitrogen limitation in *Escherichia coli* K-12: specific homeostatic responses. *J Bacteriol* **187**: 1074–1090.

Kent AD & Triplett EW (2002) Microbial communities and their interactions in soil and rhizosphere ecosystems. *Annu Rev Microbiol* **56**: 211–236.

Klappenbach JA, Dunbar JM & Schmidt TM (2000) rRNA operon copy number reflects ecological strategies of bacteria. *Appl Environ Microb* **66**: 1328–1333.

Kurata S, Kanagawa T, Magariyama Y, Takatsu K, Yamada K, Yokomaku T & Kamagata Y (2004) Reevaluation and reduction of a PCR bias caused by reannealing of templates. *Appl Environ Microb* **70**: 7545–7549.

Lay C, Rigottier-Gois L, Holmstrom K *et al.* (2005) Colonic microbiota signatures across five northern European countries. *Appl Environ Microb* **71**: 4153–4155.

Lee ZM, Bussema C III & Schmidt TM (2009) rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucleic Acids Res* **37**: D489–D493.

Licht TR, Hansen M, Poulsen M & Dragsted LO (2006) Dietary carbohydrate source influences molecular fingerprints of the rat faecal microbiota. *BMC Microbiol* **6**: 98.

McLellan SL, Huse SM, Mueller-Spitz SR, Andreishcheva EN & Sogin ML (2009) Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. *Environ Microbiol* **12**: 378–392.

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

Palmer C, Bik EM, Digiulio DB, Relman DA & Brown PO (2007) Development of the human infant intestinal microbiota. *PLoS Biol* **5**: e177.

Polz MF & Cavanaugh CM (1998) Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microb* **64**: 3724–3730.

Rainer J, Sanchez-Cabo F, Stocker G, Sturn A & Trajanoski Z (2006) CARMAweb: comprehensive R- and bioconductor-based web service for microarray data analysis. *Nucleic Acids Res* **34**: W498–W503.

Rigottier-Gois L, Bourhis AG, Gramet G, Rochet V & Dore J (2003) Fluorescent hybridisation combined with flow cytometry and hybridisation of total RNA to analyse the composition of microbial communities in human faeces using 16S rRNA probes. *FEMS Microbiol Ecol* **43**: 237–245.

Sachse K (2004) Specificity and performance of PCR detection assays for microbial pathogens. *Mol Biotechnol* **26**: 61–80.

Scanlan PD, Shanahan F, Clune Y *et al.* (2008) Culture-independent analysis of the gut microbiota in colorectal cancer and polyposis. *Environ Microbiol* **10**: 789–798.

Schmid M, Schmitz-Esser S, Jetten M & Wagner M (2001) 16S–23S rDNA intergenic spacer and 23S rDNA of anaerobic ammonium-oxidizing bacteria: implications for phylogeny and *in situ* detection. *Environ Microbiol* **3**: 450–459.

Sipos R, Szekely AJ, Palatinszky M, Revesz S, Marialigeti K & Nikolausz M (2007) Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. *FEMS Microbiol Ecol* **60**: 341–350.

Sokol H, Seksik P, Rigottier-Gois L *et al.* (2006) Specificities of the fecal microbiota in inflammatory bowel disease. *Inflamm Bowel Dis* **12**: 106–111.

Stevenson BS & Schmidt TM (2004) Life history implications of rRNA gene copy number in *Escherichia coli*. *Appl Environ Microb* **70**: 6670–6677.

Tannock GW, Munro K, Bibiloni R *et al.* (2004) Impact of consumption of oligosaccharide-containing biscuits on the fecal microbiota of humans. *Appl Environ Microb* **70**: 2129–2136.

Thompson-Chagoyan OC, Maldonado J & Gil A (2007) Colonization and impact of disease and other factors on intestinal microbiota. *Dig Dis Sci* **52**: 2069–2077.

Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER & Gordon JI (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**: 1027–1031.

Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R & Gordon JI (2007) The human microbiome project. *Nature* **449**: 804–810.

Turnbaugh PJ, Quince C, Faith JJ *et al.* (2010) Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *P Natl Acad Sci USA* **107**: 7503–7508.

van Nuenen MH, Venema K, van der Woude JC & Kuipers EJ (2004) The metabolic activity of fecal microbiota from healthy individuals and patients with inflammatory bowel disease. *Dig Dis Sci* **49**: 485–491.

Zaslaver A, Kaplan S, Bren A *et al.* (2009) Invariant distribution of promoter activities in *Escherichia coli*. *PLoS Comput Biol* **5**: e1000545.

Supporting Information

Additional figures, tables, and data are available as Supporting Information at http://www.wright.edu/∼oleg.paliy/Papers/MF_optimum/MF_optimum.html.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.