2008

Validation Of A Custom-made Microarray To Study Human Intestinal Microflora

Harshavardhan Kenche
Wright State University

VALIDATION OF A CUSTOM-MADE MICROARRAY TO STUDY HUMAN INTESTINAL MICROFLORA

A thesis submitted in the partial fulfillment

of the requirements for the degree of

Master of Science

By

HARSHAVARDHAN KENCHE
B.Tech. Biotechnology Jawaharlal Nehru Technological University, 2006

2008

Wright State University

WRIGHT STATE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

DATE _____

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Harshavardhan Kenche **ENTITLED** Validation of a custom-made microarray to study human intestinal microflora **BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF** Master of Science.

_____
Oleg Paliy, Ph.D.
Thesis Director

_____
John V. Paietta, Ph.D.
Program Director

Committee on
Final Examination

_____
Oleg Paliy, Ph.D.

_____
John V. Paietta, Ph.D.

_____
Michael P. Markey, Ph.D.

_____
Joseph F. Thomas, Jr, Ph.D.
Dean, School of Graduate Studies

ABSTRACT

Kenche, Harshavardhan. M.S., Biochemistry and Molecular Biology, Wright State University, 2008. Validation of a custom-made microarray to study human intestinal microflora

Intestinal microflora refers to all the different species of bacteria that reside in the human gut and is an important 'organ' of the human body because almost all the digestive reactions of the host occur in the intestine. The bacteria of the intestine play a key role in this process by supplementing the intestine with various enzymes and proteins that are required for the digestive process. At the same time, these bacteria were shown to be implicated in a variety of gastrointestinal disorders like Irritable Bowel Syndrome, Inflammatory Bowel Disorder and Gastrointestinal Cancer, but with the current knowledge about the microflora it is difficult to determine which exact species is responsible for a particular disease caused. The knowledge about the composition of the typical intestinal microflora is very limited, the cause at large being the lack of proper culture techniques to isolate and study the microfloral species in artificial media. Majority of the species of the microflora are obligate anaerobes and selective culturing techniques provide very limited knowledge about the composition of such complex microflora. Phylogenetic microarrays are one such approach to study various members of the microflora because they contain probes for numerous species of bacteria on a single glass slide and are also known to provide robust and high throughput analysis.

ENTREZ nucleotide database was used to compile a list of 16S ribosomal DNA (rDNA) sequences of bacterial species isolated from the human intestine and they were grouped into various phylo-species. Representative sequences for each phylo-species were extracted and the probes on the microarray were designed based on these representative sequences. 16 different bacterial species were used for validation experiments, which represented bacteria from various groups. The results showed that the microarray correctly identified 15 of a total 16 bacterial species. The detection sensitivity of the microarray was at least 1pg. As a test, fecal samples from adults and children were analyzed by the microarray. *Clostridia* were the dominant group of the microflora followed by *Bacteroidetes* in both adults and children. The analysis of the fecal samples showed clear differences between the microflora composition of adults and children.

TABLE OF CONTENTS

List of Figures

Page

List of Tables

Page

ACKNOWLEDGEMENTS

My thesis work is the outcome of long hours of study, struggle, research and commitment in the lab which will further help me to achieve my future goals. I would like to thank many people who have made this possible for me.

I would like to thank my advisor Dr. Oleg Paliy for giving me an opportunity to carry out my thesis work in his laboratory and for his constant supervision and guidance through my research and development of my thesis. I would like to specifically thank him for being available all the time to answer my doubts and questions. I would like to thank Dr. Paliy for giving me more than one opportunity to improve myself which was his way of inspiring me to work harder to bring out the best in me. I would also like to thank the previous members of Paliy Lab, Dr. Thusitha Gunasekhara, Richard Read, and John Lynch for their support and guidance. My sincere thanks to my labmate Ben Withman for all his invaluable help. I would also like to thank my thesis committee members Dr. Paietta and Dr. Markey for making time for me to answer all my questions and for helping me with my thesis.

I sincerely thank my parents Ravi Prakash Kenche and Arundhathi Kenche without whom I would not have been in the United States and for their unconditional love, prayer, encouragement and support. Finally, I would like to thank my current family in the United States, Suman Niranjan, Vishnu Kesaraju, Uday Chandupatla, Amit Bathraj, Rakesh B and Rakesh V for their all-time support and encouragement and for providing me with a beautiful home and my best buddies Vijay Sambaturu, Satwik

Palagiri, Prudhvidhar Perati and Divya Sunkara for believing in me most, for their words of encouragement and support when I needed it.

I would like to express my sincere thanks to Ramakrishna Kommagani, who has been a great friend of mine and a good guide for the past two years. Finally, my thanks to all the members of Center for Genomics Research for their help and support without which this work would have not been possible.

Dedicated to my parents Ravi Prakash Kenche and Arundhathi Kenche and

To Ravi Krishna Rajanala, my dearest pal ever

1. Introduction

1.1 Microflora: Introduction

All the various groups of bacterial species that inhabit the intestine of humans are collectively referred to as microflora. The field of intestinal microecology is as old as medical microbiology. Much effort has been devoted towards investigation of the indigenous intestinal flora since the beginning of the century and considerable progress has been made in characterizing the intestinal microflora, especially in recent years with modern methods of molecular biology, but several problems are confronted in defining the normal intestinal microflora. The composition of the flora is quite complex, particularly in areas where there are higher counts of bacteria such as colon ($>10^{11}$ organisms per gram dry weight). Many of these organisms have fastidious growth requirements and they require a very selective medium for. Detailed studies of the bowel flora are extremely time-consuming. However, shortcuts lead to significant inaccuracies. As the laboratory techniques get improvised, new species continue to be discovered.

Savage (1977) has observed that about 90% of the 10^{14} cells associated with human body are microorganisms, and that the vast majority of these bacteria reside in the large intestine. Direct microscopic estimations of bacterial numbers in the gut contents indicate that considerably more cells are present, and that total counts increase by an order of magnitude from the proximal end to distal end of the colon. A vast majority of the microbes of the intestine are anaerobes but they exhibit varying degrees of tolerance to oxygen. Anaerobic bacteria appear to outnumber the aerobic species by a factor of about 1000.

Research suggests that the relationship between the gut flora and host is not merely commensal, but rather a mutualistic/symbiotic relationship. Though people can survive with no gut flora, the latter perform a lot of useful functions like fermentation of unused metabolites, inhibition of growth of harmful species, production of vitamins for host and production o hormones to direct the host to store fats. However, in some conditions, some species are thought to cause disease by causing infection or increasing cancer risk for the host.

Normal microflora of the Stomach

It is believed that the intraluminal environment of the normal human stomach is relatively sterile, with only few counts of organisms. Studies have shown low counts (<= 10^3 colonies per milliliter gastric contents) of α-hemolytic streptococci, anaerobic cocci, lactobacilli, *Staphylococcus epidermis*, and *Candida albicans* (Giannella *et al*., 1972; Franklin and Skoryna, 1966). These counts may represent oral and ingested organisms since counts tend to decrease to zero gradually within several hours after eating. Giannella *et al*. (1972) showed that the gastric pH plays a significant role in controlling the growth of organisms in the stomach. When a marker organism (*Serratia marcescens*) was introduced into the normal stomach, it was totally eliminated within half hour. Patients with hypochlorhydria have higher bacterial counts. Patients who have undergone surgical procedures for gastric disorders also show higher bacterial counts in the stomach, which suggests the importance of lower gastric pH in maintaining low microbial counts.

Normal microflora of the Small Intestine

The amount of flora encountered in the small intestine is strictly dependent of the location of sampling. The upper small intestine is usually sterile (or has a low microbial count) when compared to that of the lower small intestine. Gorbach *et al.* (1967) used a long polyvinyl tube to sample locations throughout the small intestine and stomach. Multiple samples taken from healthy human volunteers showed the upper small intestine has low counts ($0\text{-}10^4$ colonies per milliliter) of both aerobic and anaerobic organisms. An important aspect here is the complete absence of coliforms and *Bacteroides*. As we go towards the distal end of the small intestine, the counts were higher ($10^3 - 10^6$ colonies per milliliter), and the flora had higher counts of coliforms and *Bacteroides*.

A potential criticism of this type of study is that the presence of the tube itself might alter the physiology of the bowel and change the nature of the flora. Another important observation was the relationship between the gastric pH and the microfloral count. Subjects with higher gastric pH had higher counts of both anaerobes and aerobes in the midintestinal aspirates. This supports the concept that gastric acid acts as a barrier to the swallowed organisms (Drasar *et al.* 1969).

Normal microflora of the Large Intestine

Several problems are encountered when trying to study the composition of the microflora in large intestine. Bacterial counts vary throughout the large bowel, and the numbers found in fecal specimens may not accurately represent the counts found in other locations of the colon. Bentley *et al.* (1972) studied patients undergoing elective cholecystectomy and compared the microflora of the transverse colon, cecum and

terminal ileum with the microflora of the stool specimens. The highest bacterial counts were obtained from stool specimens; the mean anaerobic count was 10^9 colonies per gram, and the mean coliform count was 10^7 colonies per gram. Bacterial counts in the terminal ileum were even lower with relatively fewer anaerobes. Although there were substantial numerical differences in the counts of bacteria between stool samples and samples from various locations in the large intestine, there did not appear marked qualitative differences in the flora. Dominant species found in this part of the human gut are *Clostridia* and *Bacteroides*.

Effect of diet and age on normal intestinal microflora

Many factors influence the composition of normal intestinal microflora. Normal individuals, when on a chemically defined diet, show a marked reduction in the number of organisms in their stool. Fewer studies were performed on the relationship between the age and microflora of an individual. The changes that occur in the intestinal microflora when a newborn is weaned and introduced to solid food are most likely due to the effect of change in diet rather than change in the age. The fecal flora of children appears to closely resemble that of adult fecal flora by the age of 1 year. Ellis-Pegler *et al*. (1975) noted that the concentration of aerobes (particularly streptococci and gram-negative facultative bacilli) decreased during the first year of life and anaerobes (particularly *Bacteroides*) became the predominant members of the fecal flora.

The effect of advanced age on the intestinal microflora in the adult is another area where few data exists. Gorbach *et al.* (1975) found that elderly subjects harbored fewer bifidobacteria but larger number of fungi and coliforms than younger subjects. These

results were in agreement with studies done by other investigators. However, relatively small number of patients have been studied and considerable variation existed among individuals of the same age.

1.2 Metabolic activities of the Intestinal Microflora

Low molecular weight carbohydrates

Most of the simple sugars and disaccharides that are consumed by the host do not reach the colon because they are absorbed as they pass through the small intestine. However, small amounts of few simple sugars may reach the colon. Moreover, the amount of simple sugars that escape digestion in the small intestine may depend on whether the sugars are ingested in a mixture along with complex carbohydrates. For example, certain types of polysaccharides can decrease the rate of glucose absorption from the small intestine (Holt *et al.*, 1979; Schwarz and Levine, 1980). Thus, it is possible that small amounts of few simple sugars or disaccharides in foods reach the colon.

Dietary polysaccharides

Dietary fiber, which are plant cell wall polysaccharides, comprise a significant portion of many human foods. Nutritional studies have indicated that most of the dietary fiber is not excreted. As the human intestinal enzymes cannot degrade plant cell wall polysaccharides, the degradation of dietary fiber has been attributed to fermentation by colonic bacteria. Since at least 50% of dietary fiber is digestible by colonic bacteria (Van

Soest, 1978), the carbohydrate in dietary fiber could provide a substantial portion of the resources needed to maintain the bacterial mass in the colon.

Previous studies have shown that members of *Bacteroides* are most active fermenters of plant polysaccharides. Members of the genus *Bacteroides* account for about 20% of all fecal isolates (Moore and Holdeman, 1974; Holdeman *et al.,* 1976). But it is incorrect to conclude that most of the catabolism of dietary fiber is done by *Bacteroides* as the studies have been done using pure cultures and the isolated polysaccharides in artificial medium, and thus they may not take into account many of the factors that are inherent in the actual gut environment.

Polysaccharides produced by the host

The host itself produces complex carbohydrates that can be utilized by the colon bacteria. In rats and presumably in humans, the intestinal mucosa is completely replaced once in every 4-5 days (Lipkin, 1973). In addition to the mucosal cells which are constantly given out into the intestinal lumen, the host produces large amounts of saliva, gastric juice and mucinous secretions, all of these products contain glycoproteins, and there is some evidence that host products are utilized by microflora *in vivo*. Hoskins and Boulding (1976) have shown that human fecal homogenates can degrade blood glycoproteins. The organisms responsible for this have not been isolated and identified, but Miller and Hoskins (1981) have estimated that these organisms account for 1% of the total population of bacteria in the colon.

Biotransformation of Bile acids and Cholesterol

The intestinal microflora of man and animals can transform bile acids and cholesterol into a variety of metabolites (Hayakawa, 1973). Known microbial biotransformations include hydrolysis of conjugated bile acids to yield free acids (Hylemon *et al*., 2006). The extent of degradation is limited by the constraints inherent in the strictly anaerobic environments of the colon. In man, the microflora can generate at least 15-20 different bile acids from the primary bile acids. The biotransformation markedly alters the physical characteristics as well as the physiological effects of steroid molecules.

1.3 Gastrointestinal disorders

Metchnikoff (1907) suggested that bacteria inhabiting the human intestine affected health and longevity of the host. Most of the diseases that occur in the human intestine are of unknown origin, but bacteria have been shown either as causative agents or maintenance factors involved in many colonic disorders. A number of species are able to upset the normal gut homeostasis and cause an acute inflammatory response. The principal organisms involved are enterotoxigenic strains of *Escherichia coli*, as well as species belonging to genera *Salmonella, Shigella, Campylobacter*, and *Yersinia*. Inflammatory Bowel Disease (IBD) and Irritable Bowel Syndrome (IBS) are thought to have an origin connected with the activities of gut flora.

1.3.1 Inflammatory Bowel Disease

Two major instances of Inflammatory Bowel Disease are Ulcerative Colitis (UC) and Crohn's disease (CD). Both conditions involve an inflammatory reaction and share many clinical features, which make individual diagnosis difficult. Marked difference is that CD affects primarily the small intestine and all regions of large intestine, whereas UC usually affects the distal colon (Whitehead 1989).

Ulcerative Colitis

The inflammatory response of UC is primarily located in the colonic mucosa and sub mucosa. The distal colon is always affected with the condition expressing itself in acute attacks followed by periods of symptom free remission. Bacterial involvement has been suggested in both initiation and maintenance stages of UC (Hill 1986). *Streptococcus mobilis, Fusobacteria* and S*higella* have been attributed as specific causative agents (Onderdonk 1983, Campieri 2001), largely because these organisms are either able to penetrate the gut mucosal epithelium or cause similar disease symptoms in animals. More direct and convincing evidence exists for a bacterial role in disease maintenance (Cummings *et al*., 2003).

Crohn's Disease

According to Chadwick (1991), *Eubacterium, Peptostreptococcus, Pseudomonas, Bacteroides vulgates* and *Clostridium difficile* are associated with the onset of Crohn's disease. Because this particular disease involves a granulomatous reaction, it is more likely that a persistent stimulus is involved. Mycobacteria are thoroughly studied in this respect and *M. paratuberculosis* has

been isolated from a number of CD patients (Chidoni *et al.,* 1984; Gitnick *et al.,* 1985, Graham *et al.,* 1987, Greenstein 2003).

1.3.2 Irritable Bowel Syndrome

Irritable bowel syndrome (IBS) is a functional bowel disorder characterized by abdominal pain, discomfort, bloating and cramping, relieved by defecation and alteration of bowel habits. IBS may also be predominated by diarrhea (IBS-D) or constipation (IBS-C) or both may alternate (IBS-A). However, the symptoms may vary from person to person. Some people have constipation, and report straining and cramping when trying to have a bowel movement but cannot eliminate any stool or able to eliminate only a small amount. People with diarrhea frequently feel an urgent and uncontrollable need to have a bowel movement. Other people with IBS alternate between constipation and diarrhea. Some people find that their symptoms subside for a few months and reoccur after a while whereas other people report a constant worsening of the symptoms over time.

The specific cause of IBS is yet to be discovered. One theory states that people who suffer from IBS have a large intestine that is particularly sensitive and reactive to certain foods and stress. The immune system may also be involved. Normal motility may not be present in the colon of a person suffering from IBS. It can be spasmodic or even stop working temporarily. The epithelial lining of the colon regulates the flow of fluids in and out of the colon. In IBS, the function of the epithelium appears normal but the contents inside the colon move too quickly for the colon to absorb the fluids. The result is too much fluid in the stool. In other people, the movement occurs slower than the normal

rate which results in extra fluids being absorbed from the contents passing through the colon which results in the person developing constipation.

Research publications from the later 1990s began identifying the biochemical changes present in the tissue and serum samples from IBS patients (Talley *et al*., 1999, Thompson *et al* 1999, Saito *et al* 2002). These studies identified cytokines and secretory products in tissues taken from IBS patients. A study done on the biopsy samples from constipation predominant IBS patients showed elevated levels of serotonin-a neurotransmitter. Ninety five percent of the serotonin in the body is located in the GI tract and rest is found in brain. Cells that line the inside of the bowel work as transporters of serotonin and carry it in and out of the GI tract. People with IBS have diminished receptor activity, causing abnormal levels of serotonin in the GI tract. As a result, they experience problems with bowel movement, motility and sensation--having more sensitive pain receptors in their GI tract.

A study on the rectal biopsy tissues from IBS patients showed increased levels of cellular structures involved in the production of Interleukin 1-β (K-A Gwee *et al.,* 2003). Studies on blood samples from IBS patients showed increased levels of TNF-α, Interleukin 1 and Interleukin 6.

1.4 Existing methods to diagnose the GI disorders

GI disorders are extremely difficult to diagnose because of the characteristic symptom overlap among common ailments (like diarrhea). The patient may show no symptoms, even if the bowel becomes increasingly damaged for years. Once the

symptoms start to show up, they often resemble those of other conditions, making the diagnosis difficult.

The doctor may go through the 'medical history'- consisting of the patient's past health, family's health, any medications he/she is taking, any allergies the patient may have and other related issues. Blood tests may be done to determine the signs, if any, of the inflammation in the body which are often present with the disease. Analysis of the stool sample may be done. Colonoscopy may be done to see inflammation, bleeding or ulcers on the wall of the colon.

1.5 New and efficient ways to study Microflora

Previous studies of the microflora were usually done by culturing the fecal samples or the biopsy samples on various kinds of defined media which are specific for certain bacterial species. But the drawback in such kind of study is that only limited number of microfloral species can be studied. Detailed studies of the microflora using such an approach are extremely time consuming and difficult. However, shortcuts lead to significant inaccuracies.

The advent of new techniques in PCR amplifications, particularly those which use 16S rDNA as a phylogenetic classifier (Wang Q *et al*., 2007) helped to great extent to identify most of the unclassified species of the microflora. 16S rDNA is believed to be an important phylogenetic classifier because it is believed to be conserved from an evolutionary point of view. Many of the members of the microflora are believed to exist

only on the basis of their 16S rDNA sequence and were never cultured in a laboratory due to the lack of proper culture techniques.

Microarray technology is another powerful tool that can be used to detect thousands of genes/target sequences in a large population. Oligonucleotide probes that are complimentary to the 16S rDNA sequences of various species can be synthesized directly on the glass slide. The sample population is hybridized to the microarrays to interrogate the presence of species of interest. Microarrays provide advantages over the PCR amplification studies because of the robust and high throughput analysis.

The advantage of oligonucleotide microarrays is the use of photolithographic process, by which dense arrays are produced containing numerous copies of a large number of different probes in a small area. This allows each array to contain considerable probe redundancy and internal standards to evaluate hybridization efficiency (Graves, 1999). The oligonucleotide array also allows for discrimination based on single base pair differences (Nuwaysir *et al*., 1999). This allows the oligonucleotide arrays to be applied in fields of medical diagnosis, pharmacogenetics and sequencing due to their hybridization and gene expression analysis capabilities.

Various kinds of microarrays have been used to study the fecal microbial composition including community genome arrays, functional genome arrays, and phylogenetic oligonucleotide arrays. Of these, the phylogenetic oligonucleotide arrays are best suited to study the microflora because their probe sequences are based upon the ribosomal RNA sequences and are ideal for microbial community composition studies.

For our study purpose, we have chosen microarrays manufactured by Affymetrix Inc. (Santa Clara, CA) (Affymetrix Technical Note 2001). The advantages of the Affymetrix based microarrays include:

- Constructed with a very high precision and accuracy.
- Have a high probe density, sensitivity and specificity.
- The use of Perfect Match (PM) and Mismatch (MM) probe pairs offers higher selectivity at low target concentrations.
- Use of multiple probes per sequence or gene allows statistical algorithms to provide confidence in microarray results.
- Probes for many different rDNA genes can be synthesized on a single array and thus hundreds of species can be tested in each microfloral sample.

1.6 Design of custom-made Microarray

The design of the custom-made microarray was carried out by Dr. Oleg Paliy in collaboration with Dr. Qiong Wang and Dr. Jim Cole of Ribosomal Database Project at Michigan State University. The probes of the microarray were based on the 16S rDNA sequences of various bacterial species that are believed to inhabit the human gut. A list of 16S rDNA sequences of intestinal microflora was compiled by performing a search of Entrez Nucleotide Database by using the following search string:

"(SSU OR 16S OR small subunit) AND (rRNA OR rDNA OR ribosomal RNA OR ribosomal DNA) AND (bacteri* OR prokaryot* OR eubacteri*) AND (human OR sapiens OR humam) AND (GI OR colon* OR mucous OR intestin* OR fecal OR feces OR faec* OR stool) NOT (archae* OR oral OR esophag*) AND 1200:1700[SLEN]"

The length of sequences retrieved was limited to 1700 base pairs. The search returned a total of 15735 microbial sequences which are reported to inhabit the human intestine. Manual examination of all the sequences revealed that the compiled list was highly redundant and hence the task of narrowing down the list into smaller groups was done. The initial dataset of 15735 sequences was grouped into various "Phylo-species"- which share a 98% sequences similarity among them. This grouping was done so that any newly discovered sequence(s) would fall into any one of the phylo-species.

The representative sequences of all the phylo-species were constructed and were truncated in such a way that the 16S rDNA sequence would fall between nucleotide positions 28 and 1491 (*E.coli* 16S rDNA positions) because this region can be amplified using universal 16S rDNA PCR primers. The truncated file was saved and supplied to Affymetrix design team as in input for their algorithm. The probe length was restricted to 25 nucleotides. The minimum number of probes per probeset was 5 and the maximum was 11. As controls, standard human, rat and mouse were included on the array. The human controls serve to estimate the amount of contaminating human DNA and the mouse and rat controls serve as negative controls.

Table 1.6.1 Distribution of bacterial 16S rDNA sequences into various phylo-species

Class	No. of phylo-species
Cyanobacteria	1
Alphaproteobacteria	9
Betaproteobacteria	17
Gammaproteobacteria	11
Deltaproteobacteria	4
Epsilonproteobacteria	6
Clostridia	527
Mollicutes	12
Bacilli	24
Actinobacteria	29
Spirochaetes	4
Bacteroides	126
Fusobacteria	3
Verrucomicrobiae	1
Lentisphaerae	1
Total	**775**

In table 1.6.1, the number of phylo-species represents the number of bacterial species 16S rDNA sequences that belong to a particular class and have been known to reside in the human gut. *Clostridia* are the dominant members of the human microflora followed by *Bacteroides*.

1.7 Thesis overview

The work in this manuscript describes the validation of the custom made microarray. The validation experiments were done using 16 different pure bacterial cultures that were obtained from ATCC as frozen stocks. Nucleic acids were isolated from each bacterial species and PCR amplification of the 16S rDNA was carried out. The use of total genomic DNA, PCR-amplified 16S rDNA and total RNA as starting material

for hybridization was validated. The detection limit of the microarray when total gDNA and amplified 16S rDNA was used was determined. Fold change experiments were done to establish a relationship between the expected and observed signal ratios. Finally, as confirmatory tests, 4 fecal DNA samples from healthy volunteers-two from children and two from adults-were analyzed by the microarray to examine the quantitative differences between adults and child microflora composition.

2. MATERIALS AND METHODS

2.1 Choice of bacterial strains

Before the microarray can be put to actual experimental usage, its ability to correctly identify different bacterial species has to be validated. For this purpose, a total of 16 different bacterial species have been chosen so that they-

1. Are available as frozen stocks from ATCC (*American Type Culture Collection)*
2. Are culturable
3. Represent bacteria with different GC content
4. Represent bacteria from various classes.

Table 2.1.1 List of bacterial species used for validation experiments

ATCC Number	Bacterial species	Group	GC content	Gram +ve/-ve
27539	*Bifidobacterium catenulatum*	*Actinobacteria*	~60	positive
15707	*Bifidobacterium longum*	*Actinobacteria*	60	positive
25559	*Eggerthella lenta*	*Actinobacteria*	30-40	positive
27274	*Enterococcus faecalis*	*Bacilli*	37	positive
4356	*Lactobacillus acidophilus*	*Bacilli*	35	positive
8483	*Bacteroides ovatus*	*Bacteroidetes*	~43	negative
8492	*Bacteroides uniformis*	*Bacteroidetes*	~43	negative
9689	*Clostridium difficile*	*Clostridia*	29	positive
638	*Clostridium paraputrificum*	*Clostridia*	~32	positive
9714	*Clostridium sordellii*	*Clostridia*	~33	positive
19403	*Clostridium sphenoides*	*Clostridia*	~34	positive
8486	*Eubacterium limosum*	*Clostridia*	30-40	positive
27210	*Ruminococcus albus*	*Clostridia*	~42	negative
51649	*Holdemania filiformis*	*Mollicutes*	38	positive
25586	*Fusobacterium nucleatum*	*Fusobacteria*	27	negative
25922	*Escherichia coli*	*Gammaproteobacteria*	51	negative

The strains were obtained as frozen cultures from ATCC and stored at -80^{0}C until final use. The cells were grown either in 15ml centrifuge flasks or T-25 tissue culture flasks. Aerobic cultures were grown in water shaker at 37^{0}C until desired OD_{600} is reached. Anaerobic were grown in sealed GasPak bag with a gas generator packet (BD GasPak EZ) until sufficient density of cells is reached. Then the cells were spun down, washed in ice-cold PBS (4^{0}C), centrifuged, and then frozen at -80^{0}C.

2.2 Isolation of nucleic acids

The nucleic acids were obtained from all the cell cultures using ZR Fungal/Bacterial DNA Isolation Kit (Zymo Research) as per the following protocol:

a. Weigh about 100 mg of pellet and resuspend in 200ul of PBS.
b. Add 750 µl of lysis buffer supplied with the kit.
c. Transfer the contents into a ZR BashingBead™ Lysis Tube and vortex at maximum speed for 5 min.
d. Centrifuge the ZR BashingBead™ Lysis Tube in a micro centrifuge at ≥ 10,000 x g for 1min.
e. Transfer 400 µl of supernatant onto Zymo-Spin™ IV Spin Filter in a collection tube and centrifuge at 7,000 rpm for 1min.
f. Add 1,200 µl of Fungal/Bacterial DNA binding buffer to activate the filtrate in the collection tube.

g. Transfer 800 µl of the filtrate in above step to Zymo-Spin™ IIC Column in a collection tube and centrifuge at 10,000 x g for 1 min. Discard the flow through and repeat the step.

h. Add 200 µl of DNA prewash buffer to the Zymo-Spin™ IIC column in a new collection tube and centrifuge at 10,000 x g for 1 min.

i. Add 500 µl of Fungal/Bacterial DNA Wash Buffer to the column and centrifuge at 10,000 x g for 1min.

j. Transfer the Zymo-Spin™ IIC column to a clean 1.5 ml centrifuge tube and elute the DNA in 100 µl of nuclease-free H_2O by centrifuging at 10,000 x g for 30 seconds.

The nucleic acids were isolated following the above protocol. The DNA was resuspended in nuclease-free water and the absorbance was measured using a spectrophotometer to estimate the amount of DNA obtained from the pellets. A 1% agarose gel was run to check the purity and integrity of the nucleic acids isolated from all the bacterial species.

The above two steps, i.e., culturing of the species and isolation of nucleic acids from the cell pellets, were performed by Dr. Frank Abernathy at Dayton Children's Hospital.

2.3 PCR amplification of 16S rDNA

The amplification of 16S rDNA was carried out from all species by using two primers Amp_27F [AGAGTTTATC(C/A)TGGCTCAG] and Bact_1492R

[TACGG(C/T)TACCTTGTTACGACTT] , which are considered universal for most bacterial species. The reaction was carried out in 50 µl volume using Takara PrimeStar HotStart DNA Polymerase. Seventy five nanograms of pure bacterial DNA was used as starting material and the reaction was carried for 25 cycles.

Table 2.3.1 Reaction mixture for PCR amplification of 16S rDNA

Component	Amount
DNA template	75ng
Primer-1 (100 µM)	1 µl
Primer-2 (100 µM)	1 µl
2.5 mM dNTP Mix	4 µl
DNA Polymerase	1 µl
Nuclease free H_2O	up to 50 µl
Total reaction volume	50 µl

The amplified DNA was purified using Qiagen QIAquick PCR purification kit. The purified DNA was resuspended in nuclease-free water and the absorbance was measured using a spectrophotometer to estimate the nucleic acid yield. A 1% agarose gel was run to check the presence of a single band at 1500 basepairs position and confirm the amplification was correct and also to check the integrity of the sample.

2.4 Fragmentation of the amplified DNA

For hybridization onto the microarrays, fragmentation of the nucleic acids has to be performed to reduce the size of the DNA fragments to 100-300 basepairs (bp). A series of fragmentation experiments were performed to check for the concentration of DNase I enzyme that would give us the desired range of fragment length. It has been

determined that 0.075 U enzyme/µg of DNA produced the optimum range of fragments. The fragmentation was performed by incubating the DNA with reaction buffer and DNase I enzyme at 37°C for 10 min, followed by inactivation of the enzyme at 98°C for another 10 min. The fragmentation was verified for proper fragment size by running the fragmented sample on a 10% polyacrylamide gel.

2.5 Terminal labeling of the fragmented DNA

The fragmented DNA product was end-labeled with biotin in a terminal transferase reaction following standard Affymetrix protocol.

Table 2.5.1 Labeling reaction mixture

Component	Amount
fragmented DNA	as required
10X Reaction buffer	10 µl
Genechip Labeling reagent	2 µl
Terminal Transferase	2 µl
Nuclease-free H_2O	up to 50 µl
Total reaction volume	50 µl

The reaction mixture was prepared as described in Table 2.5.1 and was incubated at 37^{0}C for 60 min. The labeling was stopped by adding 2 µl of 0.5 M EDTA. The labeled product was ready to be hybridized onto the microarrays or alternately would be stored at -20^{0}C for future use.

2.6 Hybridization onto the microarrays

The hybridization solution was prepared as described below:

Table 2.6.1 Hybridization mix

Component	Amount
2X Hybridization buffer	65 µl
3 nM Control oligos	2.2 µl
10 mg/ml Herring Sperm DNA	1.3 µl
50 mg/ml BSA	1.3 µl
100 % DMSO	10.2 µl
Fragmented and labeled DNA	up to 50 µl
Total volume	130 µl

The probe array was equilibrated to room temperature immediately before use. The indicated amount of hybridization solution mixture was added to the probe array. The hybridization oven was preheated to 45^0C and the array was hybridized at the set temperature for 16 hours. After 16 hrs of hybridization, the hybridization cocktail from the probe array was removed and was replaced completely with appropriate volume of Non-Stringent Wash Buffer. The staining and washing solutions were prepared on the day of washing. Using the appropriate protocol for washing and staining of the probe array in the fluidics station, it was properly processed and scanned using a GeneChip Scanner 3000.

2.7 Isolation of DNA from fecal samples

DNA from fecal samples was isolated by Dr. Frank Abernathy at Dayton Children's Hospital using ZR Fecal DNA Kit (Zymo Research) following the protocol supplied by the manufacturer:

1. Add up to 150 mg of fecal sample to a ZR BashingBead™ Lysis Tube. Add 750 µl Lysis Buffer to the tube.

2. Secure in a bead beater fitted with a 2 ml tube holder assembly (e.g., Disruptor Genie™) and process at maximum speed for 5 minutes.

3. Centrifuge the ZR BashingBead™ Lysis Tube in a microcentrifuge at ≥10,000 x *g* for 1 minute.

4. Transfer up to 400 µl supernatant to a Zymo-Spin™ IV Spin Filter (orange top) in a Collection Tube and centrifuge at 7,000 rpm (~7,000 x *g*) for 1 minute.

5. Add 1,200 µl of Fecal DNA Binding Buffer to the filtrate in the Collection Tube from Step 4.

6. Transfer 800 µl of the mixture from Step 5 to a Zymo-Spin™ IIC Column in a Collection Tube and centrifuge at 10,000 x *g* for 1 minute.

7. Discard the flow through from the Collection Tube and repeat Step 6.

8. Add 200 µl DNA Pre-Wash Buffer to the Zymo-Spin™ IIC Column in a new Collection Tube and centrifuge at 10,000 x *g* for 1 minute.

9. Add 500 µl Fecal DNA Wash Buffer to the Zymo-Spin™ IIC Column and centrifuge at 10,000 x *g* for 1 minute.

10. Transfer the Zymo-Spin™ IIC Column to a clean 1.5 ml microcentrifuge tube and add 100 µl DNA Elution Buffer directly to the column matrix. Centrifuge at 10,000 x *g* for 30 seconds to elute the DNA.

11. Transfer the eluted DNA from Step 10 to a prepared Zymo-Spin™ IV-HRC Spin Filter (green top) (see above) in a clean 1.5 ml microcentrifuge tube and centrifuge at exactly 8,000 x *g* for 1 minute. The filtered DNA is now suitable for PCR and other downstream applications.

3. RESULTS

3.1 Isolation of total genomic DNA from the bacterial species

Genomic DNA was isolated from all the 16 bacterial species using ZR Fungal/Bacterial DNA Isolation Kit (Zymo Research). A 1% agarose gel was run to check the purity and integrity of the nucleic acids.

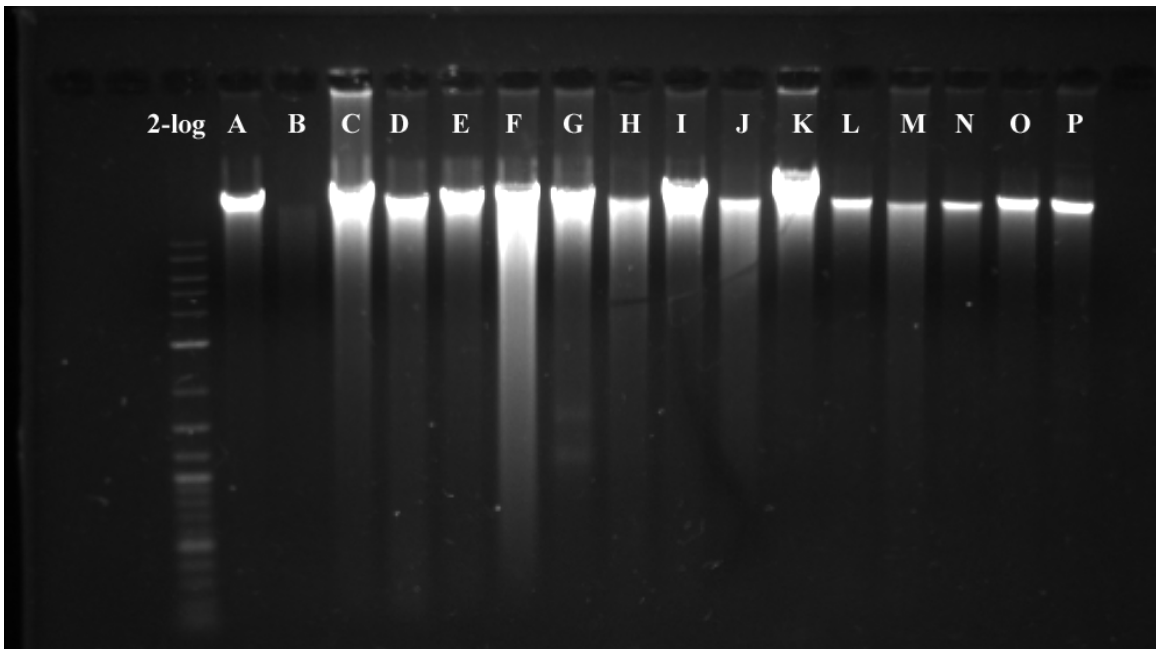


Figure 3.1.1: 1% Agarose gel of all bacterial genomic DNAs.

Each lane in the gel was loaded with 2 µl of genomic DNA sample. 1 µl of 2-log DNA ladder from NEB was used as size markers. The lanes on the gel correspond to:

A: *Holdemania filiformis*, B: *Ruminococcus albus*, C: *Eggerthella lenta*, D: *Bacteroides uniformis*, E: *Enterococcus faecalis*, F: *Lactobacillus acidophilus*, G: *Clostridium paraputrificum*, H: *Clostridium difficile*, I: *Clostridium sordellii*, J: *Clostridium sphenoides*, K: *Bacteroides ovatus*, L: *Eubacterium limosum*, M: *Bifidobacterium*

longum, N: *Bifidobacterium catenulatum*, O: *Fusobacterium nucleatum*, P: *Escherichia coli*.

Figure 3.1.1 shows that all the species gave good amounts of nucleic acids except *Ruminococcus albus*, which had a low yield. Hence, it was not used extensively in our experiments.

3.2 Identification of individual species by the microarray

Initially, we wanted to test whether the microarray can correctly identify individual species when pure gDNA from each species was hybridized to individual microarrays. For these experiments, 250 ng of gDNA from individual species were taken and were fragmented with 0.075 U/μg of DNase I. The fragmentation product was visualized by running on a 10% polyacrylamide gel to make sure that we obtained desired fragment length. The fragmented product was then end labeled with biotin in a terminal transferase reaction and the biotin-labeled product was hybridized onto microarrays. Only one bacterial species was used per microarray.

We wanted to determine the optimum values of concentration of the DNase I enzyme and the time of fragmentation required. Experiments were done with various concentrations of DNase I enzyme and various incubation times and the fragmented products were run on a 10% polyacrylamide gel to visualize the fragmented product. All of the fragmented product was loaded onto the gel. Five microlitres of Tri-dye ladder (100bp-1500bp) from NEB were loaded onto the gel and used as DNA fragment size markers. We wanted most of the fragments to fall in the range of 100-300bp.

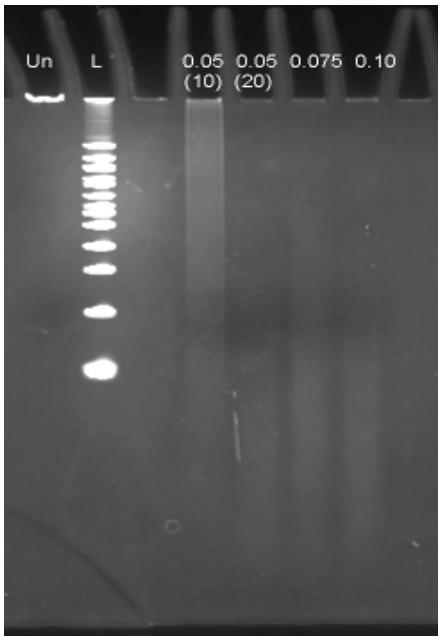


Figure 3.2.1 10% polyacrylamide gel of the experiments to determine optimum concentration of the DNase I enzyme and time of fragmentation. 'Un' represents the unfragmented total genomic DNA sample. L corresponds to the Tri-dye ladder from NEB. The samples to the right of the ladder represent the various concentrations of the DNase I enzyme in U/µg of sample and the numbers in parenthesis below the concentrations denote the time of fragmentation.

From the figure 3.2.1, 0.075 U/µg of DNA gave the ideal fragment size. Hence this concentration of the DNase I enzyme is used as standard concentration and is used in all further experiments which involve the fragmentation of the total genomic DNA sample for hybridization onto the microarrays. For the experiments where fragmentation of the PCR-amplified 16S rDNA is used, DNase concentration of 0.04 U/µg of DNA was used because the 16S rDNA is 1500 bp long and the use of higher concentrations may lead to overfragmentation of the nucleic acid sample.

Table 3.2.1 Identification of individual species by microarray when total gDNA was used

	Expt 1		Expt 2		Expt 3		Expt 4		Expt 5		Expt 6		Expt 7		Expt 8	
Bacterial species	Amt	Det	Amt	Det	Amt	Det	Amt	Det	Amt	Det	Amt	Det	Amt	Det	Amt	Det
Bifidobacterium longum	250ng	**P**	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Bifidobacterium catenulatum	-	A	250ng	**P**	-	A	-	A	-	A	-	A	-	A	-	A
Eggerthella lenta	-	A	-	A	250ng	**A**	-	A	-	A	-	A	-	A	-	A
Bacteroides uniformis	-	A	-	A	-	A	250ng	**P**	-	A	-	A	-	A	-	A
Bacteroides ovatus	-	A	-	A	-	A	-	A	250ng	**P**	-	A	-	A	-	A
Enterococcus faecalis	-	A	-	A	-	A	-	A	-	A	250ng	**P**	-	A	-	A
Lactobacillus acidophilus	-	A	-	A	-	A	-	A	-	A	-	A	250ng	**P**	-	A
Ruminococcus albus	-	A	-	A	-	A	-	A	-	A	-	A	-	A	250ng	**P**
Clostridium paraputrificum	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Clostridium difficile	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Clostridium sordellii	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Clostridium sphenoides	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Eubacterium limosum	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Holdemania filiformis	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Fusobacterium nucleatum	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Escherichia coli	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A

Amt: Amount of gDNA hybridized to chip

Det: Detection **P**- Present/**A**-Absent/**M**-Marginal

Continued on next page…

Table 3.2.1 Identification of individual species by microarray when total gDNA was used (continued)

	Expt 9		Expt 10		Expt 11		Expt 12		Expt 13		Expt 14		Expt 15		Expt 16	
Bacterial species	Amt	Det	Amt	Det	Amt	Det	Amt	Det	Amt	Det	Amt	Det	Amt	Det	Amt	Det
Bifidobacterium longum	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Bifidobacterium catenulatum	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Eggerthella lenta	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Bacteroides uniformis	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Bacteroides ovatus	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Enterococcus faecalis	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Lactobacillus acidophilus	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Ruminococcus albus	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Clostridium paraputrificum	250ng	**P**	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Clostridium difficile	-	A	250ng	**P**	-	A	-	A	-	A	-	A	-	A	-	A
Clostridium sordellii	-	A	-	A	250ng	**P**	-	A	-	A	-	A	-	A	-	A
Clostridium sphenoides	-	A	-	A	-	A	250ng	**P**	-	A	-	A	-	A	-	A
Eubacterium limosum	-	A	-	A	-	A	-	A	250ng	**P**	-	A	-	A	-	A
Holdemania filiformis	-	A	-	A	-	A	-	A	-	A	250ng	**P**	-	A	-	A
Fusobacterium nucleatum	-	A	-	A	-	A	-	A	-	A	-	A	250ng	**P**	-	A
Escherichia coli	-	A	-	A	-	A	-	A	-	A	-	A	-	A	250ng	**P**

Amt: Amount of gDNA hybridized to chip

Det: Detection **P**- Present/**A**-Absent/**M**-Marginal

From Tables 3.2.1, the microarray correctly identified 15 of the total 16 bacterial species used. *Eggerthella lenta* was called absent. We then wanted to see whether the microarray can detect individual species when all the nucleic acids were pooled together. 100ng of gDNA from each individual species (except *Ruminococcus albus*) were pooled into a mixture. The mixture was fragmented with 0.075U/µg of DNase I. The fragmented product was end labeled with biotin in a terminal transferase reaction and hybridized onto the microarray.

Figure 3.2.1 shows the scanned image of the microarray when all the samples were pooled together. Table 3.2.4 shows the detection of the individual species when all the gDNAs were pooled together. The microarray correctly identified 15 out of a total of 16 individual species when all the pure gDNAs were pooled together into a mixture. *Eggerthella lenta* was called absent.

Figure 3.2.2: Scanned image of the microarray experiment where all the gDNAs were pooled together and hybridized onto the microarray

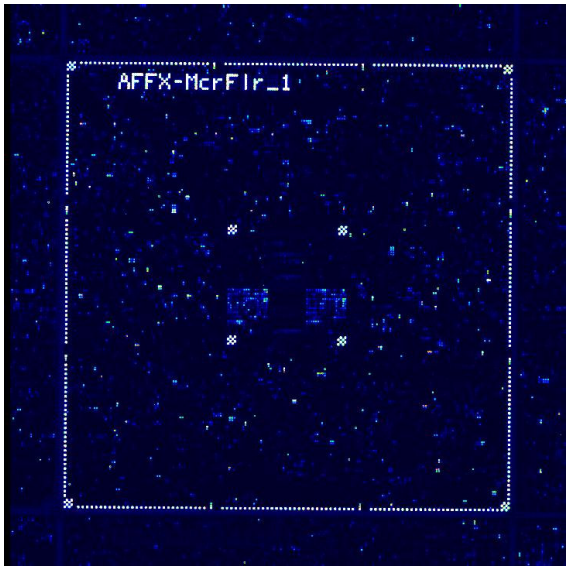


Table 3.2.2: Detection of individual species when all the gDNAs were pooled together

Bacterial species	Amt	Det
Bifidobacterium longum	100ng	**P**
Bifidobacterium catenulatum	100ng	**P**
Eggerthella lenta	100ng	**A**
Bacteroides uniformis	100ng	**P**
Bacteroides ovatus	100ng	**P**
Enterococcus faecalis	100ng	**P**
Lactobacillus acidophilus	100ng	**P**
Ruminococcus albus	-	**-**
Clostridium paraputrificum	100ng	**P**
Clostridium difficile	100ng	**P**
Clostridium sordellii	100ng	**P**
Clostridium sphenoides	100ng	**P**
Eubacterium limosum	100ng	**P**
Holdemania filiformis	100ng	**P**
Fusobacterium nucleatum	100ng	**P**
Escherichia coli	100ng	**P**

Amt: Amount of gDNA hybridized to chip

Det: Detection **P**- Present/**A**-Absent/**M**-Marginal

3.3 PCR amplification of 16S rDNA

The near full length of 16S rDNA was amplified from all the bacterial species using two universal (phylogenetically conserved) primers- Amp_27F and Bact_1492R, which bind to conserved regions of the 16S rDNA gene. 75ng of gDNA was used as starting material in each case and the amplification was carried out for 25 cycles. The reaction products were purified using Qiagen PCR Purification Kit and the purified product was run on a 1% agarose gel to check the size of the amplified band and also the

integrity of the samples. Two microlitres of amplified 16S rDNA sample was loaded onto the gel.

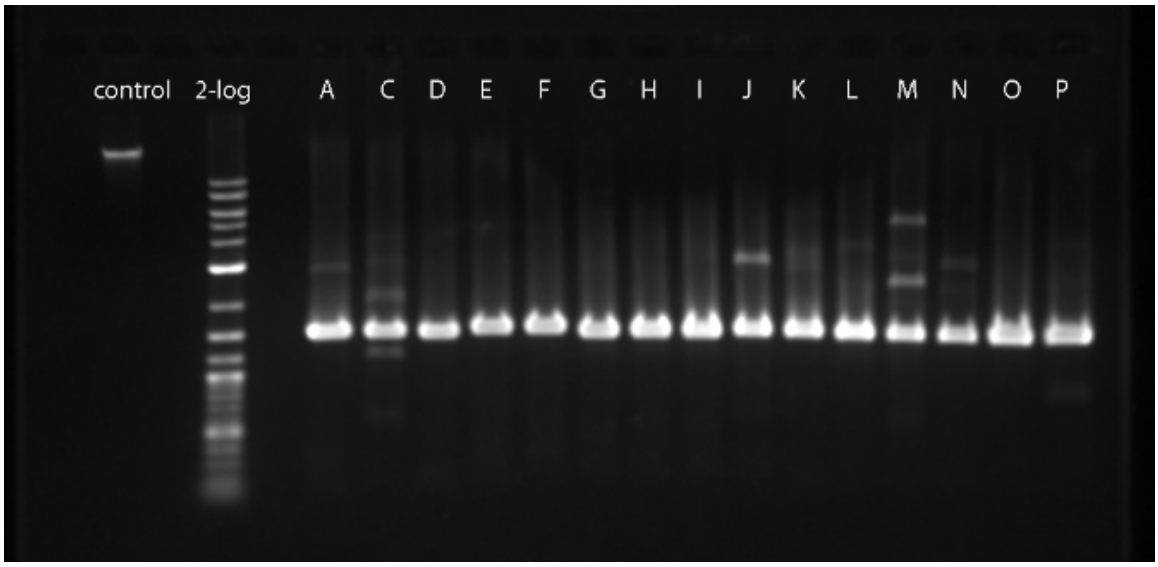


Figure 3.3.1: 1% Agarose gel of PCR amplification of 16S rDNA

A: *Holdemania filiformis*, C: *Eggerthella lenta*, D: *Bacteroides uniformis*, E: *Enterococcus faecalis*, F: *Lactobacillus acidophilus*, G: *Clostridium paraputrificum*, H: *Clostridium difficile*, I: *Clostridium sordellii*, J: *Clostridium sphenoides*, K: *Bacteroides ovatus*, L: *Eubacterium limosum*, M: *Bifidobacterium longum*, N: *Bifidobacterium catenulatum*, O: *Fusobacterium nucleatum*, P: *Escherichia coli*.

In the above figure, the control represents the unamplified total genomic DNA. 1 µl of 2-log DNA ladder (NEB) was used as DNA size markers and a nucleic acid band that corresponds to size of 1500 basepairs can be seen in all the lanes. This is the 16S rDNA from each species and it shows that the universal primers were effective in amplifying the 16S rDNA from all the bacterial species tested.

The amplified DNA is smaller in length (1.5Kb) than average fragment size of isolated genomic DNA and as a result was fragmented with 0.04 U/µg of DNase I because prior experiments showed that this concentration of the enzyme gave the optimum fragment size for hybridization onto the microarrays. A part of the fragmented product was run on a 10% polyacrylamide gel to check whether the fragmentation worked or not.

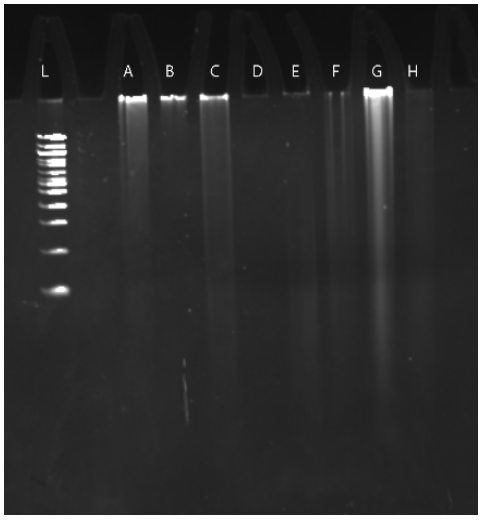


Figure 3.3.2: 10% Polyacrylamide gel for Fragmentation-I

1/10th of the fragmentation mixture was run on a 10% polyacrylamide gel to check the size of the fragments obtained. L: Tridye DNA ladder 100bp-1500bp (NEB) – used as DNA size markers

A: *Holdemania filiformis,* B: *Ruminococcus albus*, C: *Eggerthella lenta*, D: *Bacteroides uniformis*, E: *Enterococcus faecalis*, F: *Lactobacillus acidophilus*, G: *Clostridium paraputrificum*, H: *Clostridium difficile*

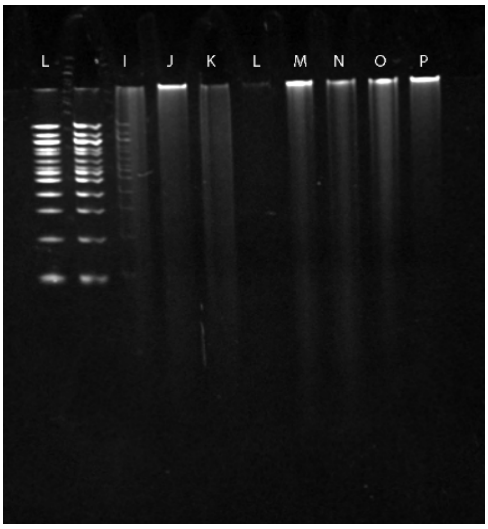


Figure 3.3.3: 10% Polyacrylamide gel for Fragmentation-II

L: Tri-dye DNA ladder 100bp-1500bp (NEB) – used as DNA size markers.

I: *Clostridium sordellii*, J: *Clostridium sphenoides*, K: *Bacteroides ovatus*, L: *Eubacterium limosum*, M: *Bifidobacterium longum*, N: *Bifidobacterium catenulatum*, O: *Fusobacterium nucleatum*, P: *Escherichia coli*

The fragmented products were end labeled with biotin in a terminal transferase reaction and hybridized onto microarrays. Only one sample was hybridized per array.

Table 3.3.1 Identification of individual species by microarray when 16S rDNA was used

	Expt 1		Expt 2		Expt 3		Expt 4		Expt 5		Expt 6		Expt 7		Expt 8	
Bacterial species	Amt	Det	Amt	Det	Amt	Det	Amt	Det	Amt	Det	Amt	Det	Amt	Det	Amt	Det
Bifidobacterium longum	250ng	**P**	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Bifidobacterium catenulatum	-	A	250ng	**P**	-	A	-	A	-	A	-	A	-	A	-	A
Eggerthella lenta	-	A	-	A	250ng	**A**	-	A	-	A	-	A	-	A	-	A
Bacteroides uniformis	-	P	-	A	-	A	250ng	**P**	-	A	-	P	-	A	-	A
Bacteroides ovatus	-	A	-	A	-	M	-	A	250ng	**P**	-	A	-	A	-	A
Enterococcus faecalis	-	A	-	M	-	A	-	A	-	A	250ng	**P**	-	A	-	A
Lactobacillus acidophilus	-	A	-	A	-	A	-	A	-	A	-	A	250ng	**P**	-	A
Ruminococcus albus	-	A	-	A	-	A	-	A	-	A	-	A	-	A	250ng	**P**
Clostridium paraputrificum	-	P	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Clostridium difficile	-	P	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Clostridium sordellii	-	M	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Clostridium sphenoides	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Eubacterium limosum	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Holdemania filiformis	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Fusobacterium nucleatum	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Escherichia coli	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A

Amt: Amount of gDNA hybridized to chip

Det: Detection **P**- Present/**A**-Absent/**M**-Marginal

Continued on next page

Table 3.3.1 Identification of individual species by microarray when 16S rDNA was used (continued)

	Expt 9		Expt 10		Expt 11		Expt 12		Expt 13		Expt 14		Expt 15		Expt 16	
Bacterial species	Amt	Det	Amt	Det	Amt	Det	Amt	Det	Amt	Det	Amt	Det	Amt	Det	Amt	Det
Bifidobacterium longum	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Bifidobacterium catenulatum	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	M
Eggerthella lenta	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Bacteroides uniformis	-	P	-	A	-	A	-	A	-	A	-	P	-	A	-	A
Bacteroides ovatus	-	A	-	A	-	A	-	A	-	M	-	A	-	A	-	A
Enterococcus faecalis	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Lactobacillus acidophilus	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A
Ruminococcus albus	-	A	-	A	-	A	-	A	-	M	-	A	-	A	-	A
Clostridium paraputrificum	250ng	**P**	-	A	-	P	-	P	-	A	-	A	-	A	-	A
Clostridium difficile	-	A	250ng	**P**	-	M	-	P	-	A	-	A	-	A	-	A
Clostridium sordellii	-	A	-	P	250ng	**P**	-	A	-	A	-	A	-	A	-	A
Clostridium sphenoides	-	A	-	P	-	A	250ng	**P**	-	A	-	A	-	A	-	A
Eubacterium limosum	-	A	-	A	-	A	-	A	250ng	**P**	-	A	-	A	-	A
Holdemania filiformis	-	A	-	A	-	A	-	A	-	A	250ng	**P**	-	A	-	A
Fusobacterium nucleatum	-	A	-	A	-	A	-	A	-	A	-	A	250ng	**P**	-	A
Escherichia coli	-	A	-	A	-	A	-	A	-	A	-	A	-	A	250ng	**P**

Amt: Amount of gDNA hybridized to chip

Det: Detection **P**- Present/**A**-Absent/**M**-Marginal

From the above results, it can be seen that 15 of a total of 16 species were correctly identified when amplified 16S rDNA was used, except in case of *Eggerthella lenta* which was again called absent. This will be discussed in the next section. It can be seen that there were a few cases where the target sequence cross-hybridized to probes of other species. There were cross hybridizations of the targets to probes of other species, for example, in Expt 1, *Bacteroides* and Clostridia were also called present. These cross hybridizations were shown to be reduced by the use of replicates. For example, the experiment in which amplified 16S rDNA from *Clostridium sphenoides* was hybridized onto the microarray, the DNA from that species cross hybridized to 21 other probes. The same experiment was repeated again exactly as it was done before.

Table 3.3.2 Experiment to test reduction of cross hybridization using replicates

Clostridium sphenoides cross hybridization to other probes		
original experiment	replicate	
S051_Clostridium_x_at	S051_Clostridium_x_at	
S052_Clostridium_at	S162_Acetivibrio_at	
S052_Clostridium_x_at	**S270_Ruminococcus_at**	
S054_Clostridium_at	S336_Ruminococcus_x_at	
S087_Anaerotruncus_x_at	S807_Prevotella_at	
S226_Anaerostipes_at	S827_Prevotella_at	
S233_Coprococcus_x_at		
S257_Roseburia_x_at		
S270_Ruminococcus_at		
S270_Ruminococcus_x_at		
S348_Ruminococcus_at		
S353_Roseburia_at		
S487_Clostridium_at		
S487_Clostridium_x_at		
S494_Roseburia_at		
S573_Papillibacter_at		
S579_Roseburia_at		
S579_Roseburia_x_at		
S581_Roseburia_x_at		
S599_Lachnobacterium_at		
S620_Holdemania_at		
S851_Victivallis_at		

Table 3.3.2 shows the results of the repeat of the experiment where 16S rDNA from *Clostridium sphenoides* was used as target. In the second experiment, the target cross hybridized to only 5 other probes.

We then wanted to see whether the microarray can identify individual species correctly when all the samples of 16S rDNA were pooled together. 100ng of 16S rDNA from each species was taken and pooled together. The mixture was fragmented using 0.04 U/μg of DNase I enzyme. The fragmentation product was end labeled with biotin and then hybridized onto the microarrays.

Table 3.3.3 shows the results of experiments. The microarray correctly identified 15 out of a total of 16 species correctly. *Eggerthella lenta* was called as absent. Figure 3.3.4 shows the scanned image of the microarray when all the PCR-amplified 16S rDNAs were pooled together, which shows increased detection sensitivity of the microarray in terms of the florescence of the probes when compared to the use of total genomic DNA (Figure 3.2.4).

Later, a negative control experiment was carried out where no DNA sample was added to hybridization mix. This experiment was important to ensure that the herring sperm DNA and the control oligonucleotides that are added to the hybridization mix do not interfere with the actual target DNA sample during the hybridization process.

Table 3.3.3: Identification of individual species by the microarray when 16S rDNA from individual species were pooled together

Bacterial species	Amt	Det
Bifidobacterium longum	100ng	**P**
Bifidobacterium catenulatum	100ng	**P**
Eggerthella lenta	100ng	**A**
Bacteroides uniformis	100ng	**P**
Bacteroides ovatus	100ng	**P**
Enterococcus faecalis	100ng	**P**
Lactobacillus acidophilus	100ng	**P**
Ruminococcus albus	-	**-**
Clostridium paraputrificum	100ng	**P**
Clostridium difficile	100ng	**P**
Clostridium sordellii	100ng	**P**
Clostridium sphenoides	100ng	**P**
Eubacterium limosum	100ng	**P**
Holdemania filiformis	100ng	**P**
Fusobacterium nucleatum	100ng	**P**
Escherichia coli	100ng	**P**

Amt: Amount of gDNA hybridized to chip

Det: Detection **P**- Present/**A**-Absent/**M**-Marginal

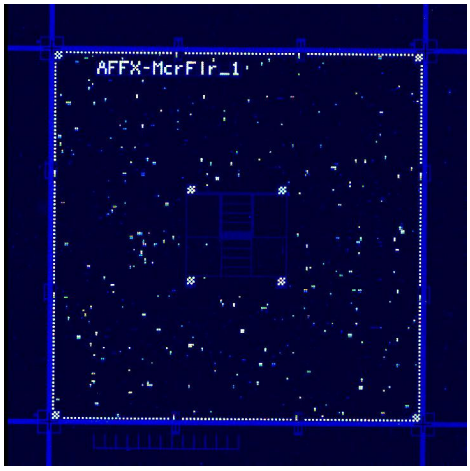


Figure 3.3.4: Scanned image of the microarray when all 16S rDNA samples are pooled together

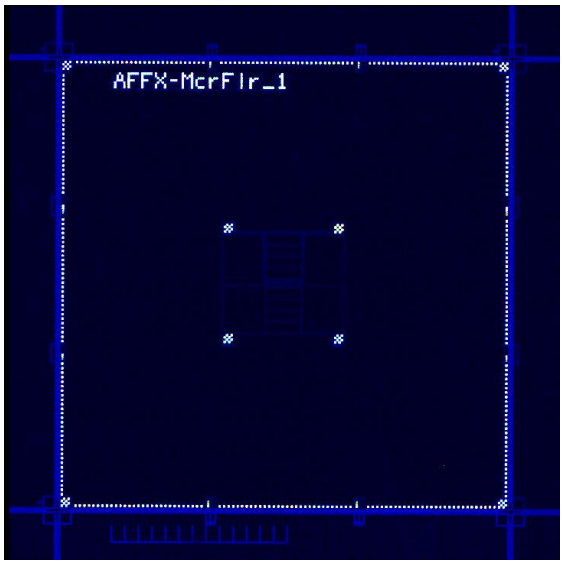


Figure 3.3.5: Negative control experiment where in no DNA sample is added to the hybridization mixture.

Figure 3.3.5 shows the image of the microarray scanned by the GeneChip 3000 Scanner. The negative control experiment produced significantly lower signal for a particular probe/species to be called present, indicating that neither the herring sperm DNA nor the control oligonucleotides that are added to the hybridization mixture interfere with the actual sample during the hybridization process.

3.4 Detection limit of the Microarray

3.4.1 In the absence of human gDNA

We wanted to find the lowest amount of the nucleic acid sample that can be detected when hybridized to the microarray. For this, we have chosen only four different bacterial species, *Bifidobacterium longum*, *Bacteroides uniformis*, *Lactobacillus*

acidophilus and *Clostridium sphenoides*. We have chosen only these four species because these samples had the highest nucleic acid yield.

Total genomic DNA from the above four species was pooled together and fragmented using 0.075 U/µg of DNase I. A part of the fragmented product was run on a 10% polyacrylamide gel to check that the fragment size was within the desired range. Then the fragmented product was end labeled with biotin in a terminal transferase reaction and then hybridized onto the microarray.

Table 3.4.1.1 Detection limit of the microarray when total gDNA was used in the absence of human gDNA

Bacterial species	Amt	Det
Bifidobacterium longum	10ng	**P**
Bifidobacterium catenulatum	-	A
Eggerthella lenta	-	A
Bacteroides uniformis	50ng	**P**
Bacteroides ovatus	-	A
Enterococcus faecalis	-	A
Lactobacillus acidophilus	1ng	**A**
Ruminococcus albus	-	A
Clostridium paraputrificum	-	A
Clostridium difficile	-	A
Clostridium sordellii	-	A
Clostridium sphenoides	200ng	**P**
Eubacterium limosum	-	A
Holdemania filiformis	-	A
Fusobacterium nucleatum	-	A
Escherichia coli	-	A

Amt: Amount of gDNA hybridized to chip

Det: Detection **P**- Present/**A**-Absent/**M**-Marginal

Above results shows that 10ng of gDNA could be detected by the microarray in the absence of human gDNA.

3.4.2 In the presence of human gDNA

The experiment was done in the same way as above, but this time in presence of human gDNA. The human gDNA isolated from HeLa cells was kindly donated by Dr. Leffak. All the DNAs were pooled together along with the human DNA. The mixture was fragmented using 0.075 U/µg of DNase I enzyme. The fragmented product was end-labeled with biotin in a terminal transferase reaction and hybridized onto the microarray.

Table 3.4.2.1 Detection limit of the microarray in presence of human gDNA when total gDNA was used

Bacterial species	Amt	Det
Bifidobacterium longum	10ng	**P**
Bifidobacterium catenulatum	-	A
Eggerthella lenta	-	A
Bacteroides uniformis	50ng	**P**
Bacteroides ovatus	-	A
Enterococcus faecalis	-	A
Lactobacillus acidophilus	1ng	**A**
Ruminococcus albus	-	A
Clostridium paraputrificum	-	A
Clostridium difficile	-	A
Clostridium sordellii	-	A
Clostridium sphenoides	200ng	**P**
Eubacterium limosum	-	A
Holdemania filiformis	-	A
Fusobacterium nucleatum	-	A
Escherichia coli	-	A
Human gDNA	3.74µg	**P**

Amt: Amount of gDNA hybridized to chip

Det: Detection **P**- Present/**A**-Absent/**M**-Marginal

From the Table 3.4.2.1, it can be seen that the lowest amount that could be detected by the microarray was 10ng total genomic DNA when used along with human gDNA. The signal obtained is higher in the experiment where no human gDNA was used

as compared to the experiment where human gDNA was used. This is thought to be due to the fact that the amount of human gDNA added is several fold higher in concentration which is near saturation and this somehow limits the accessibility of the actual target sequences towards their respective probes. In order to see if we could detect amounts greater than 1ng but smaller than 10ng when the bacterial total gDNA was used in presence of human gDNA, we did a new experiment where 4ng each of bacterial gDNAs were used in addition to 4.0µg of human gDNA. The DNA mixture was fragmented using 0.075 U/µg of DNase I.

Table 3.4.2.2 Detection limit of the microarray in presence of human gDNA

Bacterial species	Amt	Det
Bifidobacterium longum	4ng	**P**
Bifidobacterium catenulatum	-	A
Eggerthella lenta	-	A
Bacteroides uniformis	4ng	**A**
Bacteroides ovatus	-	A
Enterococcus faecalis	-	A
Lactobacillus acidophilus	4ng	**P**
Ruminococcus albus	-	A
Clostridium paraputrificum	-	A
Clostridium difficile	-	A
Clostridium sordellii	-	A
Clostridium sphenoides	4ng	**P**
Eubacterium limosum	-	A
Holdemania filiformis	-	A
Fusobacterium nucleatum	-	A
Escherichia coli	-	A
Human gDNA	4.0µg	**P**

Amt: Amount of gDNA hybridized to chip

Det: Detection **P**- Present/**A**-Absent/**M**-Marginal

From the Table 3.4.2.2, it can be seen that in 3 of 4 cases, the microarray could detect 4ng of sample. Hence, we established the detection limit of the microarray to be at least 4ng of total unamplified genomic DNA.

3.4.3 Detection limit of the microarray when PCR-amplified 16S rDNA was used

Total gDNA from the previously used four bacterial species was taken in similar amounts and subjected to 10 cycles of PCR amplification. The amplified product was fragmented using 0.04 U/µg of DNase I and hybridized onto the microarray.

Table 3.4.3.1 Detection limit when PCR-amplified 16S rDNA was used

	No human gDNA		Presence of human gDNA	
Bacterial species	Amt	Det	Amt	Det
Bifidobacterium longum	10ng	**P**	10ng	**A**
Bifidobacterium catenulatum	-	A	-	A
Eggerthella lenta	-	A	-	A
Bacteroides uniformis	50ng	**P**	50ng	**P**
Bacteroides ovatus	-	A	-	A
Enterococcus faecalis	-	A	-	A
Lactobacillus acidophilus	1ng	**P**	1ng	**P**
Ruminococcus albus	-	P	-	A
Clostridium paraputrificum	-	A	-	A
Clostridium difficile	-	A	-	A
Clostridium sordellii	-	A	-	A
Clostridium sphenoides	200ng	**P**	200ng	**P**
Eubacterium limosum	-	A	-	A
Holdemania filiformis	-	A	-	A
Fusobacterium nucleatum	-	A	-	A
Escherichia coli	-	A	-	A
Human gDNA	-	**A**	3.74µg	**P**

Amt: Amount of gDNA hybridized to chip

Det: Detection **P**- Present/**A**-Absent/**M**-Marginal

Table 3.4.3.1 summarizes the results of experiments. When amplified 16S rDNA was used as target, the microarray could detect even 1ng. So, we wanted to see if the microarray can detect even lower amounts when 16S rDNA was used. Due to problems with the TAKARA Hot Star DNA polymerase enzyme, all the subsequent amplifications were performed using Taq 2X MasterMix PCR reaction mixture from NEB. In order to reduce the variability among the experiments, we pooled the bacterial genomic DNA with human genomic DNA and carried out the PCR amplifications. Since the amounts of starting template are very low, we added gDNA of *Holdemania filiformis* to be able to visualize the amplified product on the agarose gel. In the first experiment, 10pg each of *Bacteroides uniformis*, *Lactobacillus acidophilus*, *Clostridium sphenoides* and *Escherichia coli* gDNA were pooled together along with 4.0µg of human gDNA. In another experiment, 1pg each of the above four bacterial gDNAs were pooled together but no human gDNA was added. In both the experiments, 50ng of *Holdemania filiformis* gDNA was used as a control to be able to visualize the amplified product on a gel.

Table 3.4.3.2 Detection limit when 10pg and 1pg of amplified 16S rDNA is used

	In presence of human gDNA		No human gDNA	
Bacterial species	Amt	Det	Amt	Det
Bacteroides uniformis	10pg	**P**	1pg	**P**
Lactobacillus acidophilus	10pg	**P**	1pg	**P**
Clostridium sphenoides	10pg	**P**	1pg	**P**
Holdemania filiformis	50ng	**P**	50ng	**P**
Escherichia coli	10pg	**P**	1pg	**M**
Human gDNA	4.0µg	**P**	-	A

Amt: Amt: Amount of gDNA hybridized to chip

Det: Detection **P**- Present/**A**-Absent/**M**-Marginal

From the Table 3.4.3.2, it can be seen that the lowest amounts that could be detected by the microarray were 10pg of amplified 16S rDNA in presence of human gDNA and 1pg of the PCR-amplified 16S rDNA in absence of human gDNA.

3.5 RNA as starting material

The microarray was designed as antisense type which allows both DNA and RNA targets to be interrogated and so we wanted to validate the use of total RNA as the starting material for hybridization. RNA from five bacterial species, *Bacteroides uniformis*, *Lactobacillus acidophilus*, *Clostridium sphenoides*, *Bifidobacterium longum* and *Escherichia coli* was used. Only these five species were used in particular because these are the species from which we had highest yields of total RNA.

cDNA synthesis from the RNA mixture was carried out as described in the standard Affymetrix GeneChip protocol (Appendix I). The resultant end-product was terminally labeled with biotin and hybridized onto the microarrays. Two replicates were performed to check for consistency in the detection by the microarray and reproducibility of the results.

Table 3.5.1 RNA as starting material

	1st replicate		2nd replicate	
Bacterial species	Amt	Det	Amt	Det
Bifidobacterium longum	100ng	**P**	100ng	**P**
Bacteroides uniformis	100ng	**P**	100ng	**P**
Lactobacillus acidophilus	100ng	**P**	100ng	**P**
Clostridium sphenoides	100ng	**P**	100ng	**P**
Escherichia coli	100ng	**P**	100ng	**P**

Amt: Amount of gDNA hybridized to chip

Det: Detection **P**- Present/**A**-Absent/**M**-Marginal

From the above table, it can be seen that all the bacterial species' RNA has been detected and the use of RNA as a starting material for hybridization has been successfully tested. Sensitivity of the microarray was even greater when RNA was used as target.

3.6 Fold change experiments

Fold change experiments were done to examine if the microarray can detect quantitative changes in the composition of the microflora and try to establish a relationship between the expected and observed signals. Six bacterial species, *Bifidobacterium longum, Bacteroides uniformis, Lactobacillus acidophilus, Clostridium sphenoides, Holdemania filiformis* and *Escherichia coli* were used. Of these, the first four species were used to assess actual fold change experiments and *Holdemania filiformis* was used in constant amount in all the experiments to normalize the signal intensities across all the experiments. *E.coli* gDNA was used in appropriate amounts to adjust the total amount of sample hybridized to the chips in each experiment to 1.0µg. 12.5ng to 200ng of the four bacterial species were used in order to provide a 16-fold difference in the amounts hybridized onto the microarray.

Table 3.6.1 Fold change experiments

	Amounts hybridized onto the microarray					
Bacterial species	Expt 1	Expt 2	Expt 3	Expt 4	Expt 5	Expt 6
Bifidobacterium longum	12.5ng	25ng	50ng	100ng	150ng	200ng
Bacteroides uniformis	12.5ng	25ng	50ng	100ng	150ng	200ng
Lactobacillus acidophilus	12.5ng	25ng	50ng	100ng	150ng	200ng
Clostridium sphenoides	12.5ng	25ng	50ng	100ng	150ng	200ng
Holdemania filiformis	100ng	100ng	100ng	100ng	100ng	100ng
Escherichia coli	850ng	800ng	700ng	500ng	300ng	100ng
Total amount on chip	1.0µg	1.0µg	1.0µg	1.0µg	1.0µg	1.0µg

For these experiments, 1.5 µg of each of the four bacterial total gDNA was pooled together and fragmented. Similarly, 1 µg of gDNA from *Holdemania filiformis* and 4 µg of *E.coli* gDNA were taken and fragmented separately.

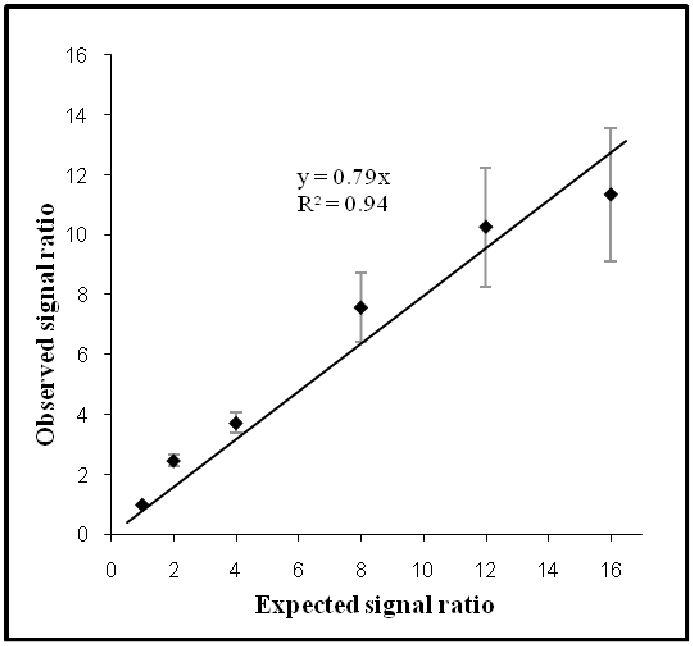


Figure 3.6.1 Relationship between expected signal and actual signal ratios

Figure 3.6.1 shows the results of the fold change experiments and the relationship between the expected signal ratio and the observed signal ratio. The dots in the graph correspond to the averages of the expected signal and observed signal ratios and it can be observed that there is a good correspondence between expected and observed signal values with the R^2 (coefficient of determination) value being equal to 0.94 and slope of the linear trendline equal to 0.79. It can be seen from the above graph that the observed signal ratios deviate from the expected ratios at higher fold changes. This is thought to occur because with the increasing fold change the hybridization mixture becomes saturated with higher amounts of nucleic acid fragments of the same species.

3.7 Microarray analysis of fecal samples

To know whether the microarray can identify and quantify the various bacterial species present in fecal samples, four fecal samples were obtained from healthy volunteers (two adults and two children). DNA was isolated from all the fecal samples using QIAamp DNA Stool Mini Kit (Qiagen). PCR amplification of 16S rDNA was carried out from all the samples using the Taq 2X MasterMix from NEB. Three individual PCR reactions were carried out per each sample so as to reduce the biases, if any that might occur during the PCR reaction (Polz *et al*., 1998).

The amplified DNA was fragmented with 0.04U/µg of DNase I enzyme. The fragmented product was run on a 10% polyacrylamide gel to check the size of the fragments obtained. The fragmented DNA was end labeled with biotin and then hybridized onto the microarray. Two replicates were done per each sample.

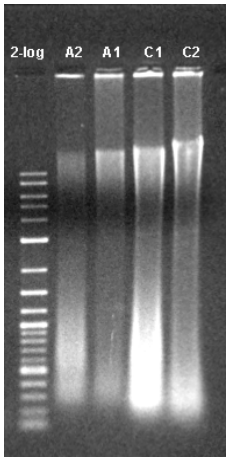


Figure 3.7.1 1% agarose gel of DNA isolated from two adults and two children

2 µl of each sample is loaded onto gel and A1 and A2 are adult fecal DNAs and C1 and C2 are child fecal DNA samples.

Table 3.7.1 Bacteria detected by the microarray in fecal samples at Order level

	Child 1		Child 2		Adult 1		Adult 2	
Class	# Phylo	Signal	# Phylo	Signal	# Phylo	Signal	# Phylo	Signal
Cyanobacteria	0	0.0%	0	0.0%	0	0.0%	1	0.1%
Alphaproteobacteria	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Betaproteobacteria	4	4.3%	0	0.0%	1	0.1%	1	0.9%
Gammaproteobacteria	2	2.1%	6	8.3%	0	0.0%	2	0.1%
Deltaproteobacteria	1	0.3%	1	0.4%	0	0.0%	1	0.3%
Epsilonproteobacteria	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Clostridia	201	79.0%	192	80.7%	168	90.7%	174	87.3%
Mollicutes	2	0.0%	3	0.1%	4	0.6%	2	0.1%
Bacilli	7	1.1%	2	0.1%	2	0.1%	3	0.3%
Actinobacteria	6	0.1%	0	0.0%	1	0.0%	0	0.0%
Spirochaetes	0	0.0%	0	0.0%	0	0.0%	1	0.0%
Bacteroidetes	26	13.0%	33	10.4%	18	7.8%	21	8.9%
Fusobacteria	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Verrucomicrobiae	1	0.1%	0	0.0%	1	0.7%	1	1.9%
Lentisphaerae	0	0.0%	1	0.0%	0	0.0%	1	0.0%

Phylo: Number of phylo-species detected for each bacterial class

Signal: % contribution to total signal by phylo-species in each bacterial class

The two microarray replicates per each fecal DNA sample had excellent correlation among them (Pearson correlation coefficient equal to 0.97, 0.99 respectively for the two child fecal samples and 0.99, 0.97 respectively for the two adult samples) which indicates excellent reproducibility of the results.

Analysis of the microarray results showed 238, 250 species were called present in the two child fecal samples and 195, 208 species were called present in the two adult fecal samples. The above table shows various bacterial classes identified by the microarray in the four fecal samples analyzed. All the four fecal samples studied were dominated by the members of *Clostridia*, which is in accordance with the literature that *Clostridia* are the dominant inhabitants of the human intestine. Second highest number of species of bacteria present next to *Clostridia* is *Bacteroidetes*. Adult fecal samples had higher percentages of *Clostridia* (87-90%) when compared to the children's samples (79-81%), whereas the two child samples had higher *Bacteroidetes* when compared to adult samples (10-13% in child samples, 8-9% in adults). Another important difference is that children had higher percentages of *Proteobacteria* when compared to adults (6.7-8.7% in children, 0.1-1.3% in adults). However, *Verrucomicrobiae* were present in adults and not in children. Diversity of the bacterial genera detected was quite similar among all the four samples- 40 and 43 different genera were detected in child fecal samples, 35-40 genera were detected amount the adult samples. Among the child fecal samples, at the order level, there were some differences observed. For example, *Burkholderiales* was present at 4.3% of the total signal in one child sample, whereas it was completely absent in another child sample. The two adult fecal samples had similar compositions at the order level.

At the genus level, the most abundant species in the child samples were *Faecalibacterium*, *Ruminococcus*, and *Bacteroides* with the contribution being 26, 18, and 8% to the total signal, respectively. Among the adult samples studied, the dominant species were *Ruminococcus*, *Papillibacter*, and *Faecalibacterium* with 24, 17 and 8% (respectively) contribution to the total signal. *Roseburia* was another genus that is present in all the four samples analyzed at a relatively similar level (~7% of total signal).

Table 3.7.2 Bacterial species identified in the fecal samples at Family level

Family	Adult 1		Adult 2		Child 1		Child 2	
	# Phylo	Signal	# Phylo	Signal	# Phylo	Signal	# Phylo	Signal
Cyanobacteria - F3.1	0	0.0%	1	0.1%	0	0.0%	0	0.0%
Rhodobacteraceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Sphingomonadaceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Phyllobacteriaceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Bradyrhizobiaceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Methylobacteriaceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Methylocystaceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Rhodobiaceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Burkholderiaceae	0	0.0%	0	0.0%	1	0.0%	0	0.0%
Oxalobacteraceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Alcaligenaceae	1	0.1%	1	0.9%	3	4.2%	0	0.0%
Neisseriaceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Xanthomonadaceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Moraxellaceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Succinivibrionaceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Enterobacteriaceae	0	0.0%	2	0.1%	1	0.0%	5	6.9%
Pasteurellaceae	0	0.0%	0	0.0%	1	2.1%	1	1.4%
Desulfovibrionaceae	0	0.0%	1	0.3%	1	0.3%	1	0.4%
Campylobacteraceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Helicobacteraceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Clostridiaceae	63	36.4%	54	26.9%	68	44.0%	54	36.8%
Lachnospiraceae	64	28.3%	83	44.7%	97	24.4%	100	38.0%
Peptostreptococcaceae	0	0.0%	0	0.0%	1	0.0%	0	0.0%
Eubacteriaceae	3	0.2%	5	0.6%	8	2.4%	9	1.7%
Peptococcaceae	3	1.0%	2	0.2%	0	0.0%	0	0.0%

Acidaminococcaceae	35	24.8%	30	14.8%	27	8.1%	27	4.1%
Thermoanaerobacteriaceae	0	0.0%	0	0.0%	0	0.0%	2	0.1%
Erysipelotrichaceae	4	0.6%	2	0.1%	2	0.0%	3	0.1%
Staphylococcaceae	0	0.0%	0	0.0%	1	0.0%	0	0.0%
Turicibacteraceae	0	0.0%	0	0.0%	1	0.0%	0	0.0%
Lactobacillaceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Enterococcaceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Streptococcaceae	2	0.1%	2	0.1%	5	1.0%	1	0.0%
Firmicutes - IS9	0	0.0%	1	0.2%	0	0.0%	1	0.1%
Coriobacteriaceae	0	0.0%	0	0.0%	3	0.1%	0	0.0%
Actinomycetaceae	0	0.0%	0	0.0%	1	0.0%	0	0.0%
Micrococcaceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Dermabacteraceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Corynebacteriaceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Bifidobacteriaceae	1	0.0%	0	0.0%	2	0.0%	0	0.0%
Serpulinaceae	0	0.0%	1	0.0%	0	0.0%	0	0.0%
Leptospiraceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Bacteroidaceae	11	4.8%	15	7.3%	17	10.8%	25	9.1%
Rikenellaceae	7	3.0%	6	1.6%	6	2.0%	6	1.2%
Porphyromonadaceae	0	0.0%	0	0.0%	3	0.2%	2	0.0%
Prevotellaceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Fusobacteriaceae	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Fusobacteria - IS11	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Verrucomicrobiaceae	1	0.7%	1	1.9%	1	0.1%	0	0.0%
Victivallaceae	0	0.0%	1	0.0%	0	0.0%	1	0.0%

At the family level, clear differences between the adult and fecal flora were observed. Members of *Clostridiaceae* and *Lachnospiraceae* were present at similar levels in both adults and children where the members of *Acidaminococcaceae* were present in higher amounts (14.8-18.8%) in adults when compared to children (4.1-8.1%). *Pasteurellaceae* were present in child fecal samples but were not detected in adult fecal samples. *Bacteroidaceae* were present in higher amounts in children (9.1-10.8%) when compared to adults (4.8-7.3%).

4. Discussion

This study describes the validation of a high-throughput, custom-made microarray designed to study the human intestinal microflora. It contains probes that can identify and quantify the numerous bacterial phylo-species that are believed to reside in the human intestine. The validation experiments were carried out using 16 different pure bacterial cultures representing various bacterial phylo-species. The microarray correctly identified 15 out of a total of 16 species. Only *Eggerthella lenta* was called absent in all the validation experiments. The 16S rDNA gene was amplified from *Eggerthella lenta* and DNA sequencing analysis of the fragment showed that the species belonged to genus *Propionibacterium*. The microarray did not contain probes for this bacterium and hence could not detect them. Members of *Propionibacterium* commonly live on skin of humans (Brüggemann *et al* 2004) and are most common contaminants of bacterial cultures due to the lack of proper care. Though the experiment turned out to be negative, it represents an important validation result because it did not randomly produce signal that might have resulted due to incorrect sequence hybridizations.

Use of PCR-amplified 16S rDNA led to increased sensitivity of the microarray when compared to the use of total genomic DNA because it produced higher signal. Cross hybridizations of the target to other probes were low and they have been shown to be reduced to significant level by the use of replicates.

Other community microarray studies like those performed by Huyghe *et al*, 2008, describe the development and validation of a custom-made community microarray, but we believe our study is better in terms of the quality of the validation experiments done.

The study by Huyghe *et al* used only three bacterial species for their validation purposes and our study used 16 species and hence is more reliable and accurate.

The detection limit of the microarray was 4ng when total genomic DNA was used in absence of human gDNA and 1ng when 10 cycles of PCR reaction were carried out to amplify the 16S rDNA. When the PCR cycles were increased to 30, as low as 10 pg of bacterial DNA was detected in presence of human gDNA (represents 0.00025% of total sample, which represents 4000-fold dynamic range of detection) and 1pg when no human gDNA was used. To ensure that the microarray can quantitatively compare the difference in amount of a particular species among different samples, a series of fold change experiments have been carried out. A good, close-to-linear correspondence was seen between observed and expected signal intensities.

The design of microarray was carried out as an antisense type allowing both DNA and RNA to be interrogated. The use of DNA as target has been established and probably this study is the first to establish that even RNA can be used as a starting material. The use of RNA as target led to increased sensitivity of the microarray.

Finally, as a test, four fecal samples have been analyzed by the microarray, two each from children and adults. The microarray identified several bacterial phylo-species from the fecal samples revealing interesting observations. However, Bifidobacteria, Lactobacilli, the two genera often used in several probiotic preparations were not significantly detected in all four samples. This was thought to be due to the universal primers not being efficient in amplifying the members of both genera. For this purpose, a modified forward primer Amp_27F_V4 was designed (AGRGTTYGATYMTGGCTCAG) which has 4 degenerate positions (the original

forward primer, Amp_27F had a single degenerate nucleotide position). PCR on fecal sample from one of the child volunteers was carried out with the new forward primers (Amp_27F_V4) and the original reverse primer (Univ_1492R) and the product was fragmented and hybridized onto microarray. It identified seven different phylo-species of *Bifidobacteria*. But the signal was not high when compared to other members of the microflora. This was consistent with the study of Palmer *et al*, who also showed lower amounts of bifidobacteria in adults and children. The detection of *Lactobacilli* was not improved even with the use of new primers indicating absence of species in the samples or that they are below the threshold of detection even though they are present.

Future work includes the use of this custom microarray to examine clinical samples of patients suffering from intestinal disorders like inflammatory bowel disease, irritable bowel syndrome, Crohn's disease and colon cancer to obtain information about the composition of the microflora and its changes during diseased condition. The information obtained from these studies can help in development of a suitable treatment strategy and also decrease chances of mis-diagnosis which happens quite often because of the symptom overlap with other common gastro-intestinal ailments.

APPENDIX I

Protocol for cDNA synthesis from total RNA

1. Prepare the following mixture

Components	Volume
Total RNA	As required
75 ng/µl Random primers	10 µl
Nuclease – free H_2O	up to 30 µl
Total Volume	30 µl

2. Incubate the RNA primer mix at the following temperatures
    a. 70^0C for 10 minutes
    b. 25^0C for 10 minutes
    c. Chill to 4^0C
3. Prepare the following cDNA synthesis reaction mixture

Components	Volume
RNA primer hybridization mix from previous step	30 µl
5X 1^{st} Strand buffer	12 µl
100mM DTT	6 µl
10mM dNTPs	3 µl
SUPERase (20U/ µl)	1.5 µl
SuperScript II (200U/ µl)	7.5 µl
Total volume	60 µl

4. Incubate the reaction mixture as follows

a. 25^0C for 10 minutes

b. 37^0C for 60 minutes

c. 42^0C for 60 minutes

d. Inactivate SuperScript II at 70^0C for 10 minutes

e. Chill to 4^0C

5. Remove the RNA by adding 20 µl of 1N NaOH and incubate for 65^0C for 30 minutes followed by addition of 20 µl of 1N HCl to neutralize.
6. Purify the cDNA by Qiagen MinElute PCR Purification Columns.

The resultant cDNA can now be fragmented using 0.6 U/µg of cDNA. The fragmented product is now ready to be terminally labeled with biotin in a terminal transferase reaction as follows-

Components	Volume
5X Reaction buffer	10 µl
GeneChip DNA labeling reagent	2 µl
Terminal deoxynucleotidyl transferase	2 µl
Fragmented cDNA product	up to 20 µl
Nuclease-free H_2O	16 µl
Total volume	50 µl

Incubate the above reaction mixture at 37^0C for 60 minutes. Add 2 µl of 0.5M EDTA to stop the labeling reaction. The labeled product is ready to be hybridized onto the microarrays. Prepare the hybridization mixture as follows-

Components	Volume
2X Hybridization buffer	65 µl
3nM B2 Control oligo	2.2 µl
10 mg/ml Herring Sperm DNA	1.3 µl
50 mg/ml BSA	1.3 µl
100% DMSO	10.2 µl
Fragmented and labeled cDNA product	up to 50 µl
Molecular biology grade water	-
Total volume	130 µl

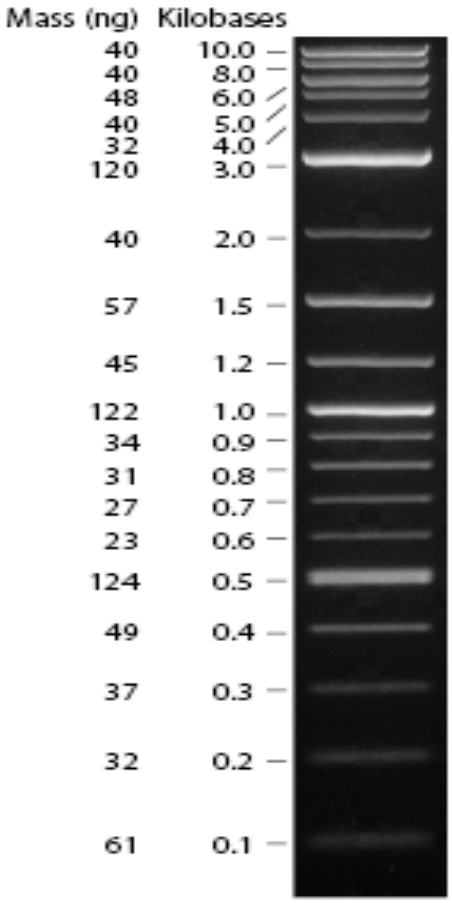
APPENDIX II

PCR Reaction protocol

1. 94^0C for 30 seconds
2. 98^0C for 10 seconds
3. 55^0C for 15 seconds
4. 72^0C for 90 seconds
5. Repeat Step 2 as per the number of cycles required
6. 4^0C for ever
7. End

APPENDIX III

2-log DNA ladder



References:

1. Campieri and Gionchetti, Bacteria as the cause of ulcerative colitis. Gut.2001; 48: 132-135

2. Cummings*, G. T. Macfarlane, and S. Macfarlane. Intestinal Bacteria and Ulcerative ColitisCurr. Issues Intest. Microbiol. (2003) 4: 9-20.

3. R . Greenstein. Is Crohn's disease caused by a mycobacterium? Comparisons with leprosy, tuberculosis, and Johne's disease . The Lancet Infectious Diseases , Volume 3 , Issue 8 , Pages 507 - 514

4. Yuri A. Saito, Philip Schoenfeld, G. Richard LockeIII.The epidemiology of irritable bowel syndrome in North America: a systematic review1 The American Journal of Gastroenterology (2004) Volume 97 Issue 8, Pages 1910 - 1915

5. Holger Brüggemann,1* Anke Henne,1 Frank Hoster,1 Heiko Liesegang,1 Arnim Wiezer,1 Axel Strittmatter,1 Sandra Hujer,2 Peter Dürre,2 Gerhard Gottschalk1.The Complete Genome Sequence of Propionibacterium Acnes, a Commensal of Human Skin. Science 2004:Vol. 305. no. 5684, pp. 671 – 673

6. Rodney D. Berg .The indigenous gastrointestinal microflora. Trends in Microbiology;Volume 4, Issue 11, November 1996, Pages 430-435

7. Sören Nobaek, Marie-Louise Johansson, , Siv Ahrné et al. Alteration of intestinal microflora is associated with reduction in abdominal bloating and pain in patients with irritable bowel syndrome ,The American Journal of Gastroenterology (2004); Volume 95 Issue 5, Pages 1231 - 1238

8. Gary L. Simon, Sherwood L. Gorbach. The Human Intestinal Microflora. Digestive Diseases and Sciences, Vol. 31, No. 9 (September 1986 Supplement), pp. 147S-162S

9. Daniel J. O'Sullivan. Analysis of Intestinal Microflora Curr. Issues Intest. Microbiol. (2000) 1(2): 39-50.

10. Eur. J. Clin, C. Tancrede. Role of Human Microflora in Health and Disease. Microbiol. Infect. Dis., November 1992, p. 1012-1015

11. G W Tannock .Analysis of the intestinal microflora using molecular methods European Journal of Clinical Nutrition December 2002, Volume 56,S44-S49

12. D C Savage. Gastrointestinal Microflora in Mammalian Nutrition. Annual Review of Nutrition Vol. 6: 155-178 ; 1986

13. Gerald W. Tannocka. Studies of the Intestinal Microflora: A Prerequisite for the Development of Probiotics. International Dairy Journal (1998);Volume 8, 527-533,

14. Probiotics and prebiotics : microflora management for improved gut health , FULLER R. (1) ; GIBSON G. R. (2) ; Clinical microbiology and infection, 1998, vol. 4, no9, pp. 477-480

15. McKay DM. Intestinal inflammation and the gut microflora. Can J Gastroenterology 1999;13(6):509-516.

16. Hooper et al. Molecular Analysis of Commensal Host-Microbial Relationships in the Intestine.

Science 2 February 2001: 881.

17. Rong-Fu Wang, Marjorie L Beggs , Latriana H Robertson , Carl E Cerniglia. Design and evaluation of oligonucleotide-microarray method for the detection of human intestinal bacteria in fecal samples. FEMS Microbiology Letters 2006;213;175 – 182

18. Lora V. Hooper, Jeffrey I. Gordon. Commensal Host-Bacterial Relationships in the Gut. Science 11 May 2001: Vol. 292. no. 5519, pp. 1115 – 1118

19. Daniel D Shoemaker and Peter S Linsley. Recent developments in DNA microarrays . June 2002 Elsevier Science

20. Sartor RB. Microbial influences in inflammatory bowel diseases. Gastroenterology 2008;134:577-94.

21. Eckburg PB, Bik EM, Bernstein CN, et al. Diversity of the human intestinal microbial flora. Science 2005;308:1635-8.

22. Frank DN, St Amand AL, Feldman RA, et al. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci USA 2007;104:13780-5.

23. Wilson KH, Blitchington RB. Human colonic biota studied by ribosomal DNA sequence analysis. Appl Environ Microbiol 1996;62:2273-8.

24. Suau A, Bonnet R, Sutren M, et al. Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. Appl Environ Microbiol 1999;65:4799-807.

25. Palmer C, Bik EM, Eisen MB, et al. Rapid quantitative profiling of complex microbial populations. Nucleic Acids Res 2006;34:e5.

26. Cole JR, Chai B, Farris RJ, et al. The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. Nucleic Acids Res 2005;33:D294-6.

27. Wang Q, Garrity GM, Tiedje JM, et al. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 2007;73:5261-7.

28. GeneChip arrays for gene expression analysis. 2001. Affymetrix, Santa Clara, Ca: Affymetrix Technical Note.

29. GeneChip arrays provide optimal sensitivity and specificity for microarray expression analysis. Affymetrix, Santa Clara, Ca: Affymetrix Technical Note.

30. Wilson WJ, Strout CL, DeSantis TZ, et al. Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. Mol Cell Probes 2002;16:119-127.

31. Huyghe A, Francois P, Charbonnier Y, et al. Novel microarray design strategy to study complex bacterial communities. Appl Environ Microbiol 2008;74:1876-85.

32. Attebery , H.r., Sutter, V.L., et al. Normal human intestinal flora. In anaerobic bacteria: Role in Disease, Chapter X, pp 81-97. Thomas, Springfield, Illinois.

33. Drasar, B.S., Goddard, P., Heaton, S., et al. Clostridia isolated from faeces. J. Med. Microbiol 1976; 9, 63-71.

34. Savage DC. Microbial ecology of the gastrointestinal tract. Ann Rev Microbiol (1977), 31; 107-133.

35. Bornside GH. Stability of the human fecal flora. Am J Clin Nutr (1978) 31; 5141-5144.

36. Chadwick VS. Etiology of chronic ulcerative colitis and Crohn's disease. Raven Press Ltd 1991, pp 445-463.

37. Drasar BS, Hill MJ. Human intestinal flora. Academic Press 1974, London.

38. Hill MJ. The possible role of bacteria in inflammatory bowel disease. Curr Con Gastroenterol 1986, 3:10-14.

39. Noble, W.C. The skin microflora and microbial skin disease, Cambridge Press 1993, Cambridge.

40. Bik EM, Eckburg PB, Gill SR, et al. Molecular analysis of the bacterial microbiota in the human stomach. Proc Natl Acad Sci U S A 2006;103:732-7.

41. K-a Gwee, S M Collins, N W Read, *et al*., Increased rectal mucosal expression of interleukin 1ß in recently acquired post-infectious irritable bowel syndrome. Gut 2003; 52: 523-526.

42. P.B. Hylemon , J. Harder., Biotransformation of monoterpenes, bile acids, and other isoprenoids in anaerobic ecosystems. FEMS Microbiology Reviews 2006; 22:475-488.

43. J. H. Cummings, G. T. Macfarlane, and S. Macfarlane. Intestinal Bacteria and Ulcerative Colitis. Curr. Issues Intest. Microbiol. 2003; 4:9-20.

44. Qiong Wang, George M. Garrity, James M. Tiedje and James R. Cole. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Applied and Environmental Microbiology 2007;73:5261-5267.

45. Polz MF, Cavanaugh CM. Bias in template-to-product ratios in multitemplate PCR. Appl Environ Microbiol 1998;64:3724-30.