Summer 2011

Molecular Analyses of Changes Induced in the Microbial Populations of Murine Colon After As(III) Exposure

Rishu Dheer

MOLECULAR ANALYSES OF CHANGES INDUCED IN THE MICROBIAL POPULATIONS OF MURINE COLON AFTER As (III) EXPOSURE

A Dissertation

Submitted to Bayer School of Natural and Environmental Sciences

Duquesne University

In partial fulfillment of the requirements for

the degree of Doctor of Philosophy

By

Rishu Dheer

August, 2011

MOLECULAR ANALYSES OF CHANGES INDUCED IN THE MICROBIAL POPULATIONS OF MURINE COLON AFTER As (III) EXPOSURE

By

Rishu Dheer

Approved June 14, 2011

Dr. John F. Stolz Director, Center of Environmental Research and Education Professor of Biology (Committee Chair)	Dr. Aaron Barchowsky Associate Professor of Biology (Committee Member)
Dr. David Lampe Associate Professor of Biology (Committee Member)	Dr. Nancy Trun Interim Chair, Department of Biological Sciences Associate Professor of Biology (Committee Member)
Dr. David W. Syebert Dean, Bayer School of Natural and Environmental Sciences Professor of Chemistry and Biochemistry	

ABSTRACT

MOLECULAR ANALYSES OF CHANGES INDUCED IN THE MICROBIAL POPULATIONS OF MURINE COLON AFTER As (III) EXPOSURE

By

Rishu Dheer

August, 2011

Dissertation supervised by Dr. John F. Stolz

The gut microbiota is essential for mammalian health and metabolism. Thus identifying factors that influence the host microbiota is key to understanding the dynamic interplay between the host and its microbiota. Recent studies have shown the effect of chronic exposures of trivalent arsenic [As(III)] in environmentally relevant concentrations on host physiology, however, little is known of how it impacts the gut microbiota.

This study examined the hypothesis that environmentally relevant concentrations of As(III) in drinking water will directly affect the murine colon microbial composition and physiology. The colon microbial communities from 10 and 250 ppb of As(III) exposed mice were compared to the control mice after 2, 5 and 10 weeks of As(III) exposure. Molecular analysis based on 16S rRNA gene and 16S- 23S intergenic region indicated a time and dose dependent shift in microbial community composition. Analysis

of 16S rDNA clone libraries demonstrated an increase in Bacteroidetes and proportionally fewer Firmicutes in colon microbiota in response to As(III) exposure.

Microbes have developed mechanisms to tolerate arsenic present in the environment. This study is the first to show that the gut microbes express arsenic resistant genes (*arsA* and *arsB*) in the colon. However, exposure to ppb concentrations of As(III) did not induce the expression of these genes in colon microbes. These data suggest that the selective effect of As(III) on colon microbiota was not due to direct exposure of colon microbes to As(III), but rather a response to changes in the host physiology.

Since arsenic affects the NO levels in human cell lines, it was hypothesized that As(III) will affect the pathways that are linked to NO levels in human body. Many microbes present in human body have nitrogen metabolizing genes (*nrfA*) that contribute to NO levels, thus the expression levels of *nrfA* gene in colon microbes was examined. Real time RT-qPCR studies showed a time and dose dependent increase in *nrfA* expression in response to As(III) exposure. Together, the results presented in this study demonstrated an indirect effect of As(III) on the composition and physiology of murine colon microbiota that may further impact the host health.

DEDICATION

This dissertation is dedicated to my Grandpa who had great belief in me.

ACKNOWLEDGEMENT

I would like to thank my dissertation advisor Dr. John F. Stolz for his guidance, patience, encouragement and later on independence to move this project forward. Dr. Stolz's mentorship has helped me immensely to grow into a confident researcher. I would like to thank my dissertation committee members Dr. Nancy Trun, Dr. David Lampe and Dr. Aaron Barchowsky for giving me valuable suggestions that helped me in broadening my scientific knowledge. I am thankful to Dr. Peter Castric for his time and suggestions given to me for this project.

I would like to thank Dr. Aaron Barchowsky, Dr. Donna stolz and their lab members Linda Klei, Yessica Garcia and Mark Ross for conducting mouse exposure experiments. Many thanks to Bruce Nemzi for helping me with the sequencing of clone libraries. A special thanks to Dr. Nancy Trun, Dr. Michael Seaman, Dr. Elizabeth Dakin, Dr. Brady Porter and Tiffany Czapski for helping me with technical aspects of this project. This work could not have completed without the assistance of various lab aides who have given me continuous supply of autoclaved media and supplies.

I am thankful to Dr. Partha Basu for his useful comments given during our lab group meetings. I am thankful to the esteem faculty at Department of Biological Sciences. Their course work and guidance has helped me in completion of this program. Thanks to Phil, Pam and Judy for their help during all these years. I appreciate the help I have received from all the past and present members of Stolz and Basu group during my stay at Duquesne.

I would like to thank my parents, sister, brother, in laws and friends for their continuous support. Words cannot justify the support I have received from my husband in day to day life. I take this as an opportunity to thank him for all the love, understanding and encouragement without which I would not have been able to reach this platform.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

As(V)	Arsenate
As(III)	Arsenite
ARISA	Automated Ribosomal Intergenic Spacer Analysis
ΔΔCt	Comparative Ct
Ct	Cycle threshold
DGGE	Denaturing Gradient Gel Electrophoresis
EtBr	Ethidium bromide
NO	Nitric oxide
OTUs	Operational Taxonomic Units
PCA	Principle Component Analysis
PPB	Parts Per Billion
RDPII	Ribosomal Database Project II
RISA	Ribosomal Intergenic Spacer Analysis
Rn	Relative fluorescence
RT	Reverse transcriptase
SAM	S-adenosyl methionine
SCFA	Short Chain Fatty Acids
wk	Week

Chapter 1 : INTRODUCTION

1.1 Arsenic Background

Arsenic (Atomic number 33 and mass 74.92) is the 20th most abundant element in the earth's crust with a concentration of 3.4 parts per million. Arsenic in combination with elements such as iron, copper, nickel, cobalt etc. forms more than 245 minerals including realgar (As_4S_4), arsenopyrite (FeAsS), nickel arsenide (NiAs), cobaltite (CoAsS) and scorodite ($FeAsO_4$) (Wedepohl, 1991; Nriagu, 2002). Used since antiquity, arsenic was first discovered by Albert Magnus in 1250. Being colorless, odorless and tasteless in nature, arsenic in the form of As_2O_3 is considered as the perfect poisoning agent (NRC, 1997). Along with some famous cases of homicidal death due to arsenic poisoning, many cases of accidental arsenic poisoning have been reported throughout history (Nriagu, 2002). Despite its toxicity, arsenic has a long history for its use as a curative for diseases such as psoriasis, syphilis and trypanosomiasis (Leonard, 1991). In 1905, Paul Ehrlich received the Nobel Prize in medicine for the discovery of salvarsan to combat syphilis (Azcue & Nriagu, 1994). Arsenic in the form of arsenic trioxide is used till today for the treatment of patients with acute promyleocytic leukemia due to its ability to induce apoptosis in cancer cells (Antman, 2001). Both inorganic and organic forms of arsenic have been used at a large scale in the manufacture of chemical warfare agent, semiconductors, glass manufacturing, paints, dyes, wood preservation, feed additives, pesticides and herbicides (Nriagu, 2002; Azcue & Nriagu, 1994). Due to concerns about the risk posed by arsenic on human health, use of arsenic in certain household applications has been completely banned (ATSDR, 2007). Even though the production of

arsenic in United States has been ceased since 1985, it is still the largest consumer of arsenic. Most of the imported arsenic is used for commercial purposes such as manufacturing glass and semiconductor devices. It is also used in metal and agricultural industry sectors (ATSDR, 2007).

1.2 Arsenic in the environment

In the environment, inorganic arsenic occurs in sedimentary rocks, fossil fuel deposits and minerals while organic arsenic in the form of arsenobetaine and arsenocholine is found in marine animals and algae (Friberg et al, 1986; Lau et al, 1987). In nature, arsenic exists in four different oxidation states: +5, +3, 0 and -3. Of these, elemental arsenic [As(0)] is rare whereas arsine [As(-III)] is found at very low levels in gases from oxygen deficient environments (Cullen & Reimer, 1989). As(III) and As(V) are the most toxic and abundant forms of inorganic and organic arsenic, primarily existing as oxyanions in the environment. Depending upon pH and oxygen, arsenate usually exists as arsenic acid in 4 different states (H_3AsO_4, $H_2AsO_4^-$, $HAsO_3^{2-}$ and AsO_4^{3-}) whereas arsenite forms arsenous acid species (H_3AsO_3, $H_2AsO_3^-$ and $HAsO_3^{2-}$) under reducing conditions (Ni & Sadler, 1991). Compared to arsenite, arsenate is the less toxic and more oxidized form of arsenic, occurring mainly in aqueous aerobic environments whereas arsenite is prevalent in anaerobic environments (Cullen & Reimer, 1989). Arsenate due to its ability to adsorb readily to minerals such as ferrihydrite, iron oxide and alumina, is less soluble and relatively immobile as opposed to mobile arsenite (Smedley & Kinniburgh, 2002). Changes in pH, oxidation-reduction potential, ion concentration of iron, nitrate or sulfide and composition of microbiota in the surrounding environment influences the release of arsenic from rocks or sediments (Smedley &

Kinniburgh, 2002; Oremland & Stolz, 2005). Arsenic species can be transformed from one form to another but cannot be removed from the environment. Arsenic is released into the environment from natural as well as anthropogenic sources. Natural sources of arsenic release include volcanic disruption, dissolution from rocks, leaching from soil and minerals into ground water whereas anthropogenic sources include mining, fossil fuel combustion, waste incineration, wood treatment and irrigation runoff after pesticide and herbicide application (Pacyna, 1987; Welch et al, 2000). In addition, aquatic and terrestrial animals also introduce some methylated organoarsenicals in the environment as excretory products (Andreae & Klumpp, 1979; Aposhian et al, 2000). Arsenic is one of the most common health hazards in the environment and affects millions of people across the world. Due to its toxicity and ease of exposure to humans, arsenic ranks first among the hazardous elements on Superfund sites (HazDat, 2006).

1.3 Microbes mediated arsenic cycling

Arsenate enters into prokaryotic cells via phosphate transporters (Pst and Pit) while arsenite enters through glycerol transporter (GlpF) (Rosen & Liu, 2009; Meng et al, 2004; Harold & Baarda, 1966). To deal with arsenic toxicity, a wide variety of microorganisms have developed mechanism such as detoxification and methylation to remove toxic arsenic species from the cell or to convert it into less toxic forms. In addition, an increasing number of microorganisms are being discovered that can use arsenic for energy and growth (Stolz & Oremland, 1999; Silver & Phung, 2005). Many microorganisms have the ability to both oxidize and reduce arsenic and possess machinery to perform a combination of the above mentioned processes (Richey et al, 2009; Saltikov et al, 2005; Macur et al, 2004). Such microbial properties have led to the

establishment of a rigorous microbe mediated arsenic biogeochemical cycle in certain extreme arsenic rich environments like Mono Lake CA (Oremland et al, 2004). Recently GFAJ-1 a γ proteobacteria isolated from this lake has been shown to replace phosphate with arsenic in its genomic material and was able to grow actively after this substitution (Wolfe-Simon et al, 2010). Figure 1.1 shows the various pathways adopted by microorganism to combat arsenic toxicity.

Methylation

Microbe mediated methylation reaction is proposed to involve an alternating sequence of reduction and oxidation of As(V) and As(III) respectively with methylation (Bentley & Chasteen, 2002; Challenger, 1945), resulting in the production of volatile methylated arsine gases and nonvolatile arsenic acids (McBride & Wolfe, 1971). Methylation was originally considered as a detoxification method however recent evidence suggests that some of these methylated compounds are more toxic then inorganic arsenic species (Styblo et al, 2000; Petrick et al, 2001). In the case of Bacteria and Archaea transfer of a methyl group from S-adenosylmethionine (SAM) to As (III) is catalyzed by ArsM, a methyltransferase (Qin et al, 2006). Methyl cobalamin also serves as a methyl donor in some bacterial species but the mechanism by which methyl group is transferred to arsenic species is not clear yet (Gadd, 1993; Krautler, 1990).

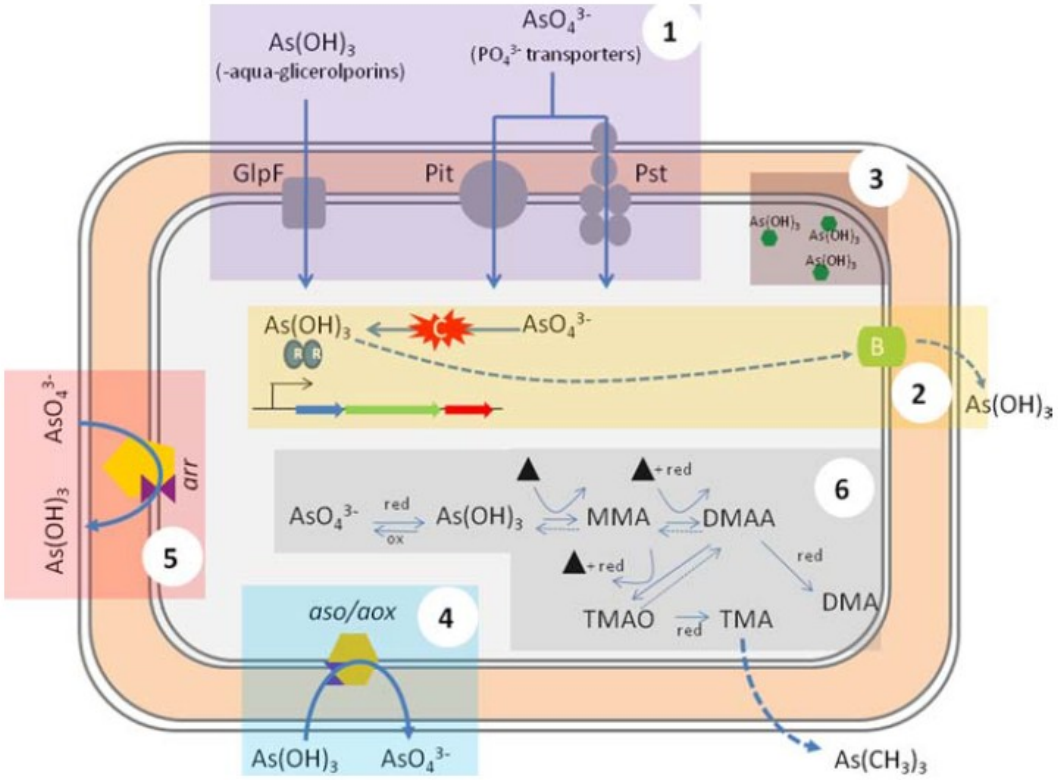


Figure 1.1: Pathways adopted by microorganism to combat arsenic toxicity. (1) Entry of arsenite and arsenate via glycerolporin and phosphate transporters. (2) As(V) reduction followed by expulsion of As(III) (3) Detoxification of As(III) by sequestration (4) arsenite oxidation (5) Dissimilatory As(V) respiration and (6) methylation of arsenic species. Figure reproduced from Espino et al; 2009 with permission from the author.

Arsenite oxidation

Arsenite oxidation by bacteria can be used either as a detoxification process or an energy generation process (Tamaki & Frankenberger, 1992; Oremland & Stolz, 2003). Bacteria from α, β and γ proteobacteria and other phyla representing more than 9 genera have been reported to have arsenite oxidizing abilities (Santini et al, 2002; Salmassi et al, 2002; Gihring & Banfield, 2001). The detoxification reaction, mediated by heterotrophic arsenite oxidizers (HAOs) involves oxidation of As (III) to As (V) by periplasmic enzyme arsenite oxidase and no energy is generated in the process (Muller et al, 2003). In Chemolithotrophic arsenite oxidizers (CAOs), oxidation of arsenite is an energy generating process and the process is coupled to reduction of oxygen, nitrate or nitrite. The energy generated in the process is used for fixing CO_2 into organic material (Oremland et al, 2002; Rhine et al, 2006). Arsenite oxidase belongs to the DMSO reductase family and consists of *aoxA* and *aoxB* genes that encode, respectively, the large 90 KDa Mo containing and the small 14KDa Rieske (2Fe-2S) subunits of arsenite oxidase (Ellis et al, 2001; Santini & vanden, 2004).

Arsenate reduction

Dissimilatory As(V)-respiring prokaryotes (DARPs) reduce As(V) to As(III) and energy is generated in the process (Oremland & Stolz, 2005; Stolz & Oremland, 1999). In DARPs, arsenate reduction is coupled to the oxidation of either organic or inorganic compounds (Niggemyer et al, 2001; Hoeft et al, 2004; Liu et al, 2004; Takai et al, 2003). The process was first discovered in *Sulfurospirillum arsenophilum* and has been identified in many other bacterial species since then (Oremland & Stolz, 2005; Ahmann et al, 1994). Arsenic requirement is not obligate for DARPs as they are capable of using

other terminal electron acceptors such as nitrate, selenate, selenite, fumarate, and sulfur compounds for energy generation (Oremland & Stolz, 2003). Respiratory arsenate reductases belong to DMSO family of mononuclear molybdenum-containing enzymes consisting of 87 and 29 kDa subunits. Both periplasmic and membrane bound respiratory arsenate reductases have been found. The large catalytic subunit, ArrA, contains motifs for binding an iron-sulfur cluster and Mo containing pyranopterin cofactor whereas ArrB, the smaller electron transfer subunit, contains three [4Fe-4S] and one [3Fe-4S] iron-sulfur clusters (Saltikov & Newman, 2003; Afkar et al, 2003).

Arsenate reduction via the *ars* operon is one of the most well studied systems for arsenic detoxification and has been identified in wide variety of Bacteria, Archea and yeast (Jackson & Dugas, 2003). In bacteria the genes for *ars* operon are located on plasmids or chromosome. The genes that typically form an *ars* operon include: *arsB*, *arsC* and *arsR* (Xu et al, 1998). *arsA*, *arsD* and *arsH* are some other genes that may occur in the *ars* operon of some bacteria. The arrangement and configuration of genes involved in arsenic tolerance vary from species to species (Figure 1.2).

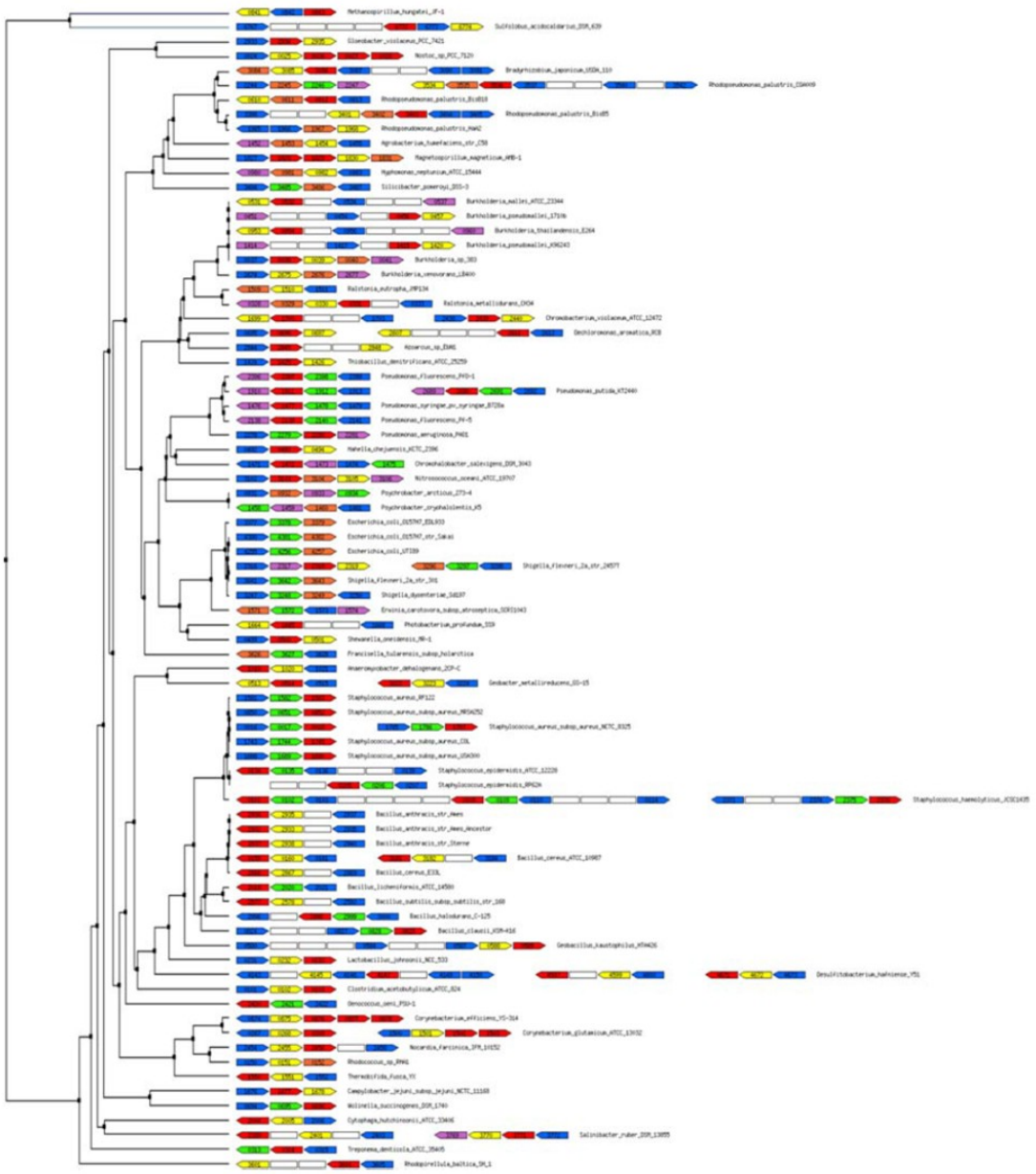


Figure 1.2: Arrangement of *ars* operon in bacterial species. *arsR* (blue), *arsB* (green), *ACR3* (yellow), glutaredoxin coupled *arsC* (orange), thioredoxin coupled *arsC* (red), *arsH* (purple) and other genes (white). Figure modified from Espino et al, 2009 with permission from the author.

ArsC, a 12- 16 KDa cytoplasmic arsenate reductase, reduces As(V) to As(III) (Ji & Silver, 1992). As (III) is exported out of the cell via ArsB, an arsenite specific transmembrane protein (Rosen, 1999). In some organisms the process of arsenite expulsion is coupled to ATP hydrolysis to form an oxyanion- translocating complex and the ATPase in these microbes is encoded by *arsA* (Dey et al, 1994). In bacteria where ArsA is not present, the proton motive force across the cell membrane facilitates arsenite expulsion (Figure 2A). ArsD is a 13KDa cytosolic metallochaperone that exists as homodimer and facilitates the transfer of As(III) to ArsA thus helping in As(III) expulsion (Lin et al, 2007). ArsR is a trans acting transcriptional repressor which in the absence of As(III) represses the expression of ars operon by binding to the operator region within the promoter of the operon (Xu et al, 1996). ArsH is a NADPH-flavin mononucleotide oxidoreductase that is required for arsenic resistance in *Yersinia enterocolitica*, *Shigella flexneri* and *Sinorhizobium meliloti* but its role in arsenic detoxification is not clear yet (Vorontsov et al, 2007; Ye et al, 2007).

On the basis of structure, reducing power and the location of catalytic cysteine residues the arsenate reductases are classified into 3 protein families: (i) *E.coli* plasmid R773 type glutaredoxin coupled arsenate reductase (Gladysheva et al, 1994) (ii) *Staphylococcus aureus* plasmid pI258 type thioredoxin coupled arsenate reductase (Ji et al, 1994) (iii) Eukaryotic glutaredoxin coupled arsenate reductase from *Saccharomyces cerevisiae* (Bobrowicz et al, 1997).

Similar to arsenate reductase, the As(III) membrane efflux protein is currently divided into 2 distinct protein families (Rosen, 1999) (i) the typical ArsB protein type found in Firmicutes and γ proteobacteria and (ii) ACR3 type present in Actinobacteria, α

proteobacteria, Archaea and *S. cerevisiae*. The 2 protein types although similar in size, differ in structure and specificity for metal transport (Wu et al, 1992; Wysocki et al, 1997). While ACR3 with 10 membrane spanning segments is specific for As(III) only, ArsB with 12 membrane spanning segments can transport antimonite in addition to arsenite (Ghosh et al, 1999).

The *ars* genes are transcribed from a single promoter to form a polycistronic mRNA with a very short half-life of less than 4 minutes. ArsR is a 13 KDa protein that belongs to winged-helix metalloregulatory proteins and forms a dimer (Busenlehner et al, 2003). In *E.coli*, arsenite binds to sulfur thiolates of 3 conserved Cys residues at position 32, 34 and 37 of the DNA binding domain of ArsR. The binding of As(III) or Sb(III) in the DNA binding domain induces a conformational change in ArsR leading to disruption of DNA-ArsR interaction. This disruption results in release of ArsR from the promoter region, inducing the transcription of *ars* genes (Shi et al, 1996).

1.4 Routes of arsenic exposure

Water

Humans are exposed to inorganic forms of arsenic mainly through drinking water and food and to a lesser extent via dermal absorption and by inhalation of dust (WHO, 2001). Natural level of arsenic in surface and ground water is 1-2 µg/L. However the levels may reach upto 3400 µg/L in water bodies close to natural or anthropogenic sources of arsenic (Matschullat, 2000). Approximately 100 million people across the globe are exposed to high levels of inorganic arsenic mainly through drinking water contaminated by natural mineral deposits. Besides natural sources, anthropogenic sources including industrial waste, mining and agricultural manure release arsenic into the water

(Nriagu & Pacyna, 1988). Countries like Taiwan, Japan, Chile, China, Finland, Denmark, Bangladesh and India have regions where arsenic concentration in water supplies have reached ppm levels (WHO, 2001). In United States, approximately 13 million people consume water from public supplies that have arsenic concentration >10 µg/L (Welch et al, 2000). High concentrations of arsenic in the US are associated with mining, geothermal activities and natural deposits. Due to its toxicity, EPA has reduced the permissible exposure limits of As in drinking water supplies from 50 µg/L to 10 µg/L in 2006, to match that of WHO levels (EPA, 2001). According to the data collected by US geological survey from 1973 − 2000 in 31,500 ground water supplies across the US, various regions of US including New England, Maine, Nevada, Utah, New Mexico and Arizona have arsenic concentrations above the recommended 10 ppb or 10 µg/L (USGS, 2007) (Figure 1.3).

Air

In the atmosphere arsenic is primarily released as arsenic trioxide or less frequently as volatile arsine (Pacyna, 1987; Matschullat, 2000). Concentration of arsenic in atmosphere ranges from 1-3 ng/m^3 in remote locations but may reach upto 2000 ng/m^3 in areas close to anthropogenic sources of arsenic. In the US about 355 tons of arsenic compounds are released into the atmosphere every year (EPA, 2005). Exposure to higher levels of arsenic via dust mainly occurs in people involved in production or use of arsenic compounds such as workers at pesticide industry, glass production and metal smelters. Also people living in vicinity of these industry inhale more arsenic as compared to people living in clean environments (Pacyna, 1987).

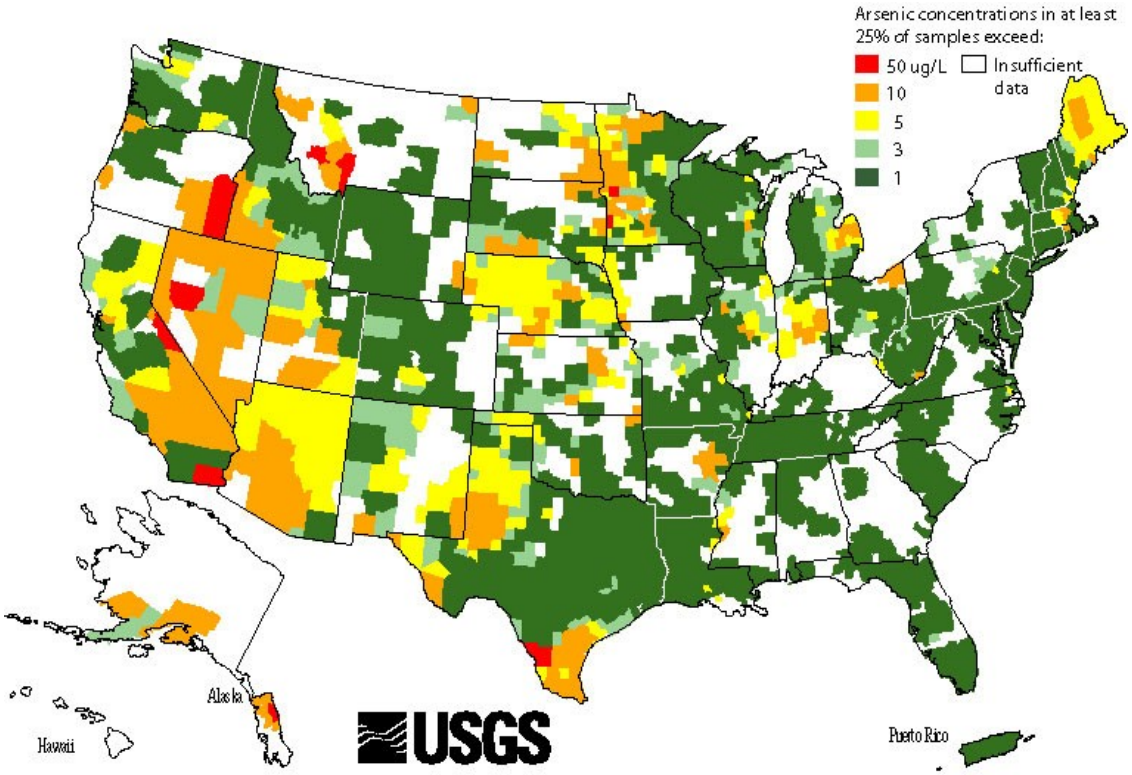


Figure 1.3: Arsenic concentrations in drinking water supplies across United States. http://water.usgs.gov/nawqa/trace/pubs/geo_v46n11/fig2.html.

Food

The arsenic concentration in food ranges from 20 to 140 ppb. Rice, apple, grapes, lettuce, tomatoes and potatoes may have arsenic concentrations in the ppm range (Schoof et al, 1999). Also the food grown in arsenic contaminated soils may accumulate higher levels of arsenic (Abedin et al, 2002; Carbonell-Barrachina et al, 1997). Exposure to arsenic also occurs via poultry consumption because of widespread use of arsenic as a feed additive in poultry industry (Lasky et al, 2003). Exposure to organoarsenicals (arsenobetaine and arsenocholine) occurs by consuming seafood but is considered nontoxic as organoarsenicals are excreted unchanged (Brown et al, 1990; Yamauchi et al, 1986). However, in vitro studies with intestinal microbes from humans have demonstrated the transformation of arsenobetaine to DMA(V), TMAO and dimethylarsinoylacetate under aerobic conditions (Harrington et al, 2008). It is estimated that the average dietary intake of arsenic in US ranges from 0.31 – 1.14 μg/kg/day (MacIntosh et al, 1997). According to Food and Agriculture Organization (FAO), the maximum tolerable daily intake of total arsenic is 2-7 μg/kg body weight/day (WHO, 2010).

Soil and dermal absorption

Levels of arsenic in soil range from 1 to 40 mg/kg but may reach higher levels in areas associated with arsenic use such as golf fields, agricultural lands, mining areas and waste disposal sites (Mandal & Suzuki, 2002). Arsenic from such sites leaches into the sub-surface aquifers and may enter into drinking water supplies. Although dermal absorption of arsenic from soils is negligible, it occurs more commonly due to occupational exposures (Lowney et al, 2007). In vitro studies with human skin have

shown 1.9% absorption of inorganic arsenic over a period of 24 hours (Wester et al, 1993).

1.5 Fate of arsenic in humans

Absorption of As(III)

Ingested inorganic arsenic is readily absorbed in the GI tract (Vahter & Norin, 1980) wherein it binds to proteins and other low molecular weight compounds present in the blood and is transported to other organs in the body (NRC, 1999). The amount of arsenic absorbed by GI tract, its retention and distribution to other body parts is dependent on the type of arsenic species, oxidation state and dose level; with As(III) being more efficiently retained and transported (Thomas et al, 2001). Arsenic is also absorbed via lungs after inhalation and this inhaled arsenic can subsequently be absorbed in the GI tract after clearance from the respiratory tract (Vahter et al, 1986; Yamauchi et al, 1989). Like microbes, As(III) is transported in the mammalian cells via aquaglycoporins (APQ7 and APQ9) (Liu et al, 2002) whereas As(V) is transported by phosphate transporters (Huang & Lee, 1996).

Arsenic metabolism

After absorption, inorganic arsenic is transformed in the human body by reduction, oxidation, conjugation and methylation reactions (Carter et al, 2003). The liver is considered as the main organ where methylation of inorganic As(III) occurs but the role of other organs or tissues such as kidneys, lungs, skin and RBCs in arsenic metabolism has also been reported (Buchet et al, 1984; Cohen et al, 2006). Two models have been proposed for the biotransformation of arsenic in mammalian systems (i) Generally accepted pathway suggested by Challenger (Challenger, 1945) (ii) alternate

pathway proposed by Hayakawa (Hayakawa et al, 2005). The first step is common to both the pathways wherein As(V) is reduced to As(III) in the blood with thiols working as electron donors. Glutathione-S-transferase omega (GSTO1) is the arsenate reductase enzyme that has been identified to play a role in this transformation (Zakharyan et al, 2001). In addition several other enzymes including purine nucleoside phosphorylase (Radabaugh et al, 2002), glycogen phosphorylase (Gregus & Németi, 2007), glyceraldehyde dehydrogenase (Gregus & Németi, 2005) that can reduce arsenate in mammalian cells have been reported but their physiological role has not been established so far. More recently the catalytic domains of human Cdc25B and Cdc25C phosphatases have been proposed to play a role in arsenate reduction. These enzymes are structurally similar to ArsC found in yeast (Bhattacharjee et al, 2010). In the generally accepted pathway suggested by Challenger, arsenic undergoes alternate oxidation and reduction along with methylation to produce methylated arsenic species in a specific order with +5 oxidation state of an arsenic species generated before the +3 state of the same arsenic species (Figure 1.4). Both As3MT (arsenic methyltransferase) and CYT19 have been proposed to methylate arsenic species (Zakharyan et al, 1999; Lin et al, 2002). According to the Hayakawa pathway, As(V) is reduced to As(III) followed by formation of arsenic triglutathione (ATG) in the presence of glutathione. ATG is further converted into monomethyl arsenic diglutathione (MADG) and dimethyl arsinic glutathione (DMAG) by transfer of methyl group from S-adenosyl methionine As3MT. Both MADG and DMAG form equilibrium with MMA(III) and DMA(III) respectively. MMA and DMA are readily converted to their pentavalent forms by oxidation (Figure 1.5).

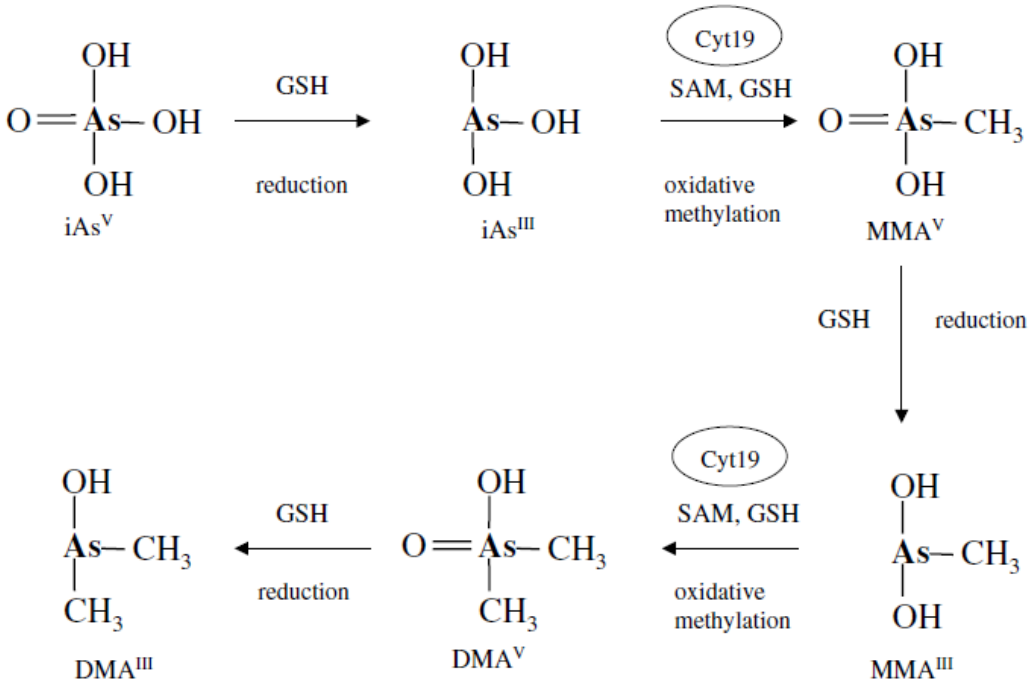


Figure 1.4: Arsenic biotransformation pathway suggested by Challenger. Figure reproduced from Hayakawa et al, 2005 with permission from the author.

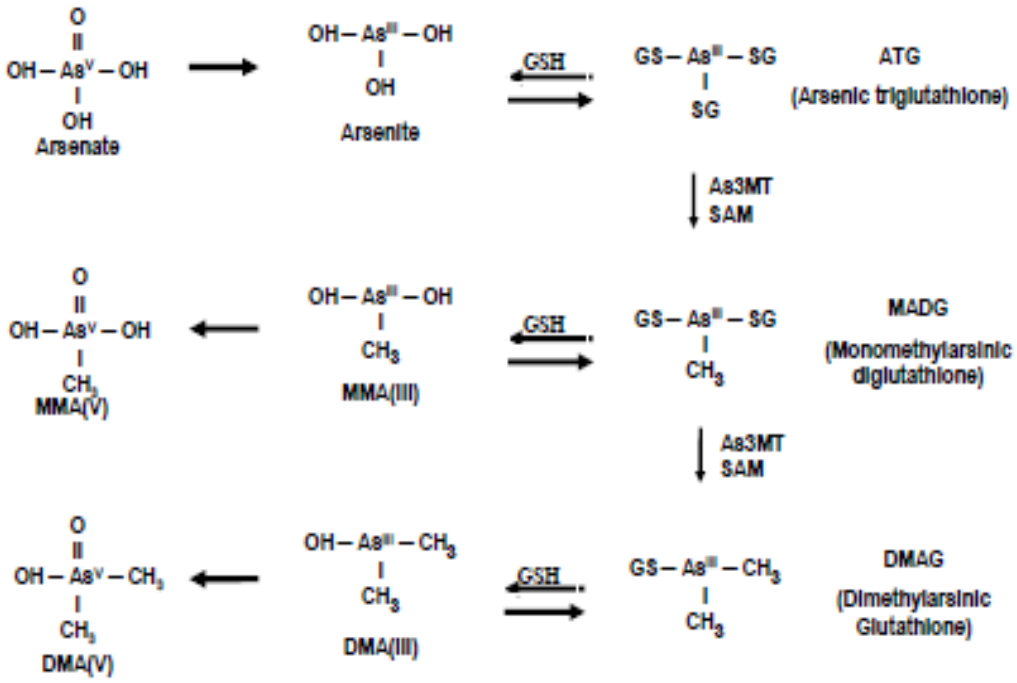


Figure 1.5: Arsenic biotransformation pathway proposed by Hayakawa. Figure reproduced from Hayakawa et al, 2005 with permission from the author.

Arsenic excretion

From liver inorganic arsenic and its methylated metabolites are distributed to other organs of the body. The retention time for inorganic arsenic is longer in skin, hair, upper GI tract, skeleton and lens of the eye (Lindgren et al, 1982; Yamauchi & Yamamura, 1985). The main products of methylation of inorganic As in humans are monomethylarsonic acid [MMA(V)], monomethylarsonous acid [MMA(III)], dimethylarsinic acid [DMA(V)] and dimethylarsinous acid [DMA(V)] ((Vahter, 2002; Le et al, 2000). Of these, MMA(III) and DMA(III) are the most toxic and reactive forms but are formed as intermediates only (Styblo et al, 2000; Petrick et al, 2001). MMA(V) and DMA(V) are less toxic forms of arsenic and do not react readily with tissues. These forms are readily excreted in the urine along with untransformed inorganic arsenic (Vahter, 2002). In addition, relatively nontoxic trimethylarsine oxide (TMAO) and highly toxic sulfur containing monomethylmonothioarsonic acid [MMMTA(V)] and dimethylmonothioarsinic acid [DMMTA(V)] have also been observed in human urine (Yoshida et al, 1998; Naranmandura et al, 2007). Elimination of arsenic from the human body occurs in three phases: (i) 66% removal within two days (ii) 30% elimination in ten days and (iii) 4% removed in 38 days (Mealey et al, 1959; Marafante et al, 1985). In humans typically 10-30% of the arsenic is removed in its inorganic form, 10-20% as MMA and 55-75% as DMA. However these values may differ in individuals depending on genotype, sex, exposure dose and route of exposure (Loffredo et al, 2003). Although the majority of absorbed arsenic is removed from the body via urine (80%), a small amount of it is also removed via stool and sweat (10%). A small percentage of ingested arsenic accumulates in the hair, nails, bones and muscles (ATSDR, 2007). Multi-drug

resistance proteins (MRP1 and MRP2) have been documented to be involved in the expulsion of As(III) and possibly arsenic glutathione from mammalian cells (Leslie et al, 2004).

Arsenic transformation by gut microbes

In vivo studies with intravenous injection of sodium arsenate in rats demonstrated rapid formation of methylated arsenicals in a pattern similar to those found after oral administration suggesting that intestinal microbes do not play any significant role in methylation of arsenic under in vivo conditions (Rowland & Davies, 1982). However studies have shown that metabolism of arsenic in rats differ significantly from human and mice (Vahter, 1994). In all mammals, the rate of absorption of arsenic after oral consumption via the GI tract depends not only on the oxidation state and speciation but also on the bioavailability of the ingested arsenic species (Laparra et al, 2006; Juhasz et al, 2006). The reentry of arsenic or its metabolites into the gut via enterohepatic circulation also exposes the intestinal microbiota to this toxin (Csanaky & Gregus, 2002). An in vivo study with rats showed toxic effects of arsenic (1mg/L) on the total bacterial count in stool along with reduced levels of arsenic in the stool samples (Choudhry et al, 2010). In vitro studies have shown that the gut microflora of arsenic exposed individuals can also play a significant role in detoxification and elimination of arsenic from the body. Rat and mice cecal microbiota were shown to biotransform As(V) to As(III) with the formation of MMA and DMA (Rowland & Davies, 1981; Hall et al, 1997). In addition to As(V) reduction, dissimilatory arsenate reduction was observed in hamster fecal samples after being fed with 100 mg/L of As(V) via drinking water (Herbel et al, 2002). More recently it was demonstrated that the in vitro human intestinal microbial community has

the ability for presystemic arsenic metabolism resulting in the formation of MMMTA along with MMA and DMA (Van de Wiele et al, 2010). Moreover, formation of thiolated methylarsenicals is observed in the presence of intestinal microbiota (Kuroda et al, 2004). Together these studies indicate that arsenic transformation by intestinal microbes is common at least at high doses of arsenic exposure.

1.6 Mechanism of arsenic toxicity and human health

Twenty four years ago, arsenic was classified as group I carcinogen by International Agency for Research on Cancer (IARC, 1987). An acute exposure to arsenic (1-3 mg/kg body weight/day) leads to immediate effects such as vomiting, diarrhea, GI hemorrhage, multiple organ damage and death. The symptoms may appear within few minutes to hours after ingestion (Levin-Scherz et al, 1987; Armstrong et al, 1984). Regardless of the route of exposure, chronic exposure to arsenic in the range of 0.01–0.04 mg/kg/day results in arsenicosis/arseniasis, skin lesions, skin, lung, liver, kidney, prostate and bladder cancer in humans (Wu et al, 1989). Additionally various epidemiological studies have linked chronic arsenic exposure to reproductive defects, developmental defects, neurological defects, respiratory effects, type II diabetes, weight loss and several cardiovascular diseases including atherosclerosis, portal cirrhosis and blackfoot disease (WHO, 2001). Transplacental studies have shown that arsenic crosses the placenta and acts as a carcinogen inducing tumors at various sites in the progeny (Waalkes et al, 2004). Recently a number of studies have shown that drinking water containing environmentally relevant concentrations of As(III) in parts per billion range have significant effect on various organs of mammalian model systems (Table 1.1).

Dose	Exposure time	Organ studied	Effect	Reference
50 ppb	5&8 weeks	Lung and heart	Disruptive Extra Cellular Matrix	Hays et al, 2008
50 ppb	4 weeks	Lungs	Altered protein expression (RAGE, GST omega 1, enolase 1, peroxyredoxin 6)	Lantz et al, 2007
10, 50, 200 ppb	9 weeks	Lungs	Increased growth of implanted tumors	Kamat et al, 2005
50 ppb*	8 weeks	Liver	Increased glutathione and early hepatocyte degeneration	Bashir et al, 2006
50 ppb	16 weeks	Brain	Spontaneous locomotor activity, decreased dopamine and antidoxidants	Bardullas et al, 2009
10 &50 ppb	3 weeks	Kidney	Increased hexokinase II	Pysher et al, 2007
100 ppb	5 weeks	Lung	Altered expression of genes involved in cell adhesion, migration	Kozul et al, 2009b
100 ppb	5 weeks	Immune system	Reduced immunity to Influenza virus	Kozul et al, 2009a
50, 250, 500 ppb	5, 10, 20 weeks	Heart	Differential expression of angiogenic and tissue remodeling genes	Soucy et al, 2003
10&250 ppb	2&5 weeks	Liver	Angiogenesis and vascular remodeling	Straub et al, 2007b; Straub et al, 2007a; Straub et al, 2008
50, 100, 300 ppb*	~6 weeks	Brain	Increased glutathione	Chaudhuri et al, 1999

Table 1.1: Changes induced in different organs of mammalian biological model sytems in response to As(III) exposure.* indicates studies done in rats. All other studies were done in mice.

The understanding of the molecular mechanisms behind the etiology of these diseases and organ dysfunction is very limited. However, various mechanisms for arsenic toxicity have been proposed including protein binding of arsenic, chromosome abnormalities, altered DNA repair, mitochondrial damage, DNA hypo and hyper methylation, oxidative stress and altered signal transduction pathways that lead to cell proliferation and tumorogenesis (Kitchin & Wallace, 2008; Liu et al, 2005; Hughes, 2002).

Arsenate exhibits structural similarity to phosphate and enters into the cell via phosphate transporters. Within the cell arsenate interferes with phosphate uptake and affects the processes that involve phosphate such as oxidative phosphorylation (Tamaki & Frankenberger, 1992). Arsenite is neutral and structurally similar to non-ionized glycerol at a pH less than 9.2. Arsenite can enter into the cell either by diffusion or through glyceroporin membrane channels (Ramírez-Solís et al, 2004; Mukhopadhyay et al, 2002). AQP9, a mammalian aquaglycerporin, has been shown to transport arsenite into the cells ((Liu et al, 2002). Arsenite has the ability to bind to the sulfhydryl and thiol groups of amino acids and disrupt protein structure and function (NRC, 1999). It is estimated that arsenic affects more than 200 enzymes just by binding to them (Abernathy et al, 1999).

Studies with vascular endothelial cells have shown that arsenite interacts extracellularly with sphingosine-1-phosphate ($S1P_1$), a G protein coupled receptor and affects Rac1 GTPase activity (Straub et al, 2009). Rac1 GTPase is involved the in cell survival signal transduction pathway (Sun et al, 2006). At sublethal doses, arsenite alters the expression of stress response transcription factors (NFkB, Nrf2 and AP 1) and mitogen activated protein kinases. These proteins play an important role in cell cycle

control but the mechanism by which arsenite affect these pathways is not clear (Druwe & Vaillancourt, 2010).

Altered DNA methylation has been associated with reduced levels of SAM due to its use in biomethylation of arsenic in the body. Change in the expression of enzymes involved in DNA methylation is considered as another factor involved in altered DNA methylation (Ren et al, 2010). Aberrant chromosomes occur due to arsenic induced deletion mutations, altered DNA repair enzymes, endoreduplication, aneuploidy and micronuclei formation (Moore et al, 1997; Zhang et al, 2007).

Arsenite exposure at low levels has been shown to be associated with reactive oxygen species (ROS) and OH- free radical production by various mechanisms, resulting in oxidative damage and inactivation of proteins (Samikkannu et al, 2003; Shi et al, 2004). In human vascular smooth muscle cells, arsenite has been shown to activate NADPH oxidase which reduces O_2 to superoxide anion (Lynn et al, 2000). ROS alters the redox potential of the cell by affecting GSH reductase and thioredoxin reductase thus reducing the levels of thiols in the cell (Lin et al, 1999). ROS can directly damage DNA or can alter signal transduction pathways (Hwang & Kim, 2007).

Arsenite has been reported to exhibit its toxicity via regulation of eNOS (endothelial NO synthase) and iNOS (inducible NO synthase), thus directly affecting the production of nitric oxide (NO) (de et al, 1996; Souza et al, 2001). Depending upon the cell type and dose, the effect of arsenite on NO levels can vary from increased to decreased or have no effect on NO production (Gurr et al, 2003). In addition, arsenite has been demonstrated to have an indirect effect on cellular NO levels. In vascular endothelial cells arsenite induces the expression of NOX, a superoxide generating

enzyme (Smith et al, 2001). The superoxide anion generated in the process combines with NO generated from the eNOS thus reducing the availability of NO for its normal physiological functions (Pryor & Squadrito, 1995). NO causes DNA damage either by directly binding to the amino groups of DNA, by forming N-nitroso compounds that alkylate DNA or by forming reactive oxidative species such as peroxynitrite which causes oxidative DNA damage (Tamir & Tannenbaum, 1996).

1.7 Mammalian GI tract microbiome

The GI tract of mammals hosts a variety of microorganisms including bacteria, archaea, viruses, fungi, yeast and protozoa. Of these, bacteria are the predominant forms with the number of bacterial cells in the GI tract far exceeding the number of eukaryotic cells in human body (Savage, 1977). The GI tract of the neonate is usually sterile but bacterial colonization starts during birth with intestinal microbiota of infants showing similarity to their mothers (Palmer et al, 2007; Turnbaugh et al, 2009). The microbial profile of infants delivered vaginally differed significantly from those delivered by cesarean section and is also affected by infant diet, hygiene and medication (Penders et al, 2006). Bacterial profiling by molecular techniques showed that infant microflora is in constant flux but becomes more stable and resembles to adult towards the end of infancy (Palmer et al, 2007). Throughout life, a number of factors have been shown to influence the microbiota of the host including age, diet, antibiotic use and genotype (Mariat et al, 2009; Turnbaugh et al, 2008; Dethlefsen et al, 2008; Zoetendal, 2001). Beside these external and genetic factors, environmental factors within the gut such as oxygen, pH, nutrient availability, redox potential and presence of other microbes also play a key role in determining the microflora of the gut (Hill, 1995). A large number of commensal,

symbiotic and opportunistic bacteria inhabit the GI tract of mammals, with density of bacteria increasing from stomach to large intestine (O'Hara & Shanahan, 2006). The diversity and type of bacteria differs not only longitudinally from stomach to the distal portion of the large intestine but also radially from mucosa to luminal (fecal) regions (Frank et al, 2007; Zoetendal et al, 2002). Culture independent high throughput 16S rRNA gene sequencing of human and mice intestinal microbiota have shown that gram negative Bacteroidetes and low GC gram positive Firmicutes are the major phyla in the GI tract, together constituting about 90% of the total sequenced population. Within Firmicutes, genera belonging to clusters XIVa and IV of the clostridial group predominate. Proteobacteria, Verrucomicrobia, Deferribacteres, Actinobacteria, Fusobacteria and other gram negative bacteria are also found in the GI tract but in very low abundance (Nava et al, 2011; Eckburg et al, 2005). Although the number of bacterial phyla in mammalian gut is very low, there are approximately 1800 different genera found in the gut with the number of species reaching close to 15,000. Out of these only about 2000 different phylotypes are harbored by any particular individual at any point (Frank et al, 2007; Frank & Pace, 2008). The predominating bacterial genera found in the mammalian gut include *Clostridium*, *Bacteroides*, *Ruminococcus*, *Lactobacillus*, *Acinetobacter*, *Enterobacter*, *Bifidobacterium* with more anaerobes as compared to aerobes (Eckburg et al, 2005; Gordon & Dubos, 1970). The bacterial composition is generally conserved at the phylum level; it is the genera and species that vary greatly from individual to individual. Molecular 16S rRNA gene based studies of gut microbiota have shown sequences that do not show identity to the previously described microbes indicating the uniqueness of gut microbiota for each individual (Nava et al, 2011;

Eckburg et al, 2005; Frank & Pace, 2008). In addition, the proportion of different bacteria can vary from individual to individual at all levels of bacterial taxonomy. A recent metagenomic study has shown that although the bacterial populations differ widely between individuals the overall metabolic pathways encoded by these microbes tend to remain the same (Qin et al, 2010). A reciprocal gut microbiota transplant study done by using germ free mice and zebrafish showed that the gut of a given host can be colonized by the microbiota of another host but eventually the recipient host modulates the colonized microbiota such that it resembles to its own natural microbiota (Rawls et al, 2006). These emergent properties of the intestinal microbiota while providing valuable information regarding evolutionary history of gut residing microbes also helps in discerning the mechanistic aspect of differential patterns of microbial localization within the gut in response to various factors.

1.8 Role of intestinal microbiota in health and disease

In a healthy adult the host bacterial interaction is primarily a symbiotic relationship. Intestinal bacteria play a variety of protective, structural and metabolic functions that provide benefits to the host (McCracken & Lorenz, 2001). The protective function of intestinal bacteria helps in immunity development, GI tract homeostasis and prevention against pathogen colonization (Hand & Belkaid, 2010; Rakoff-Nahoum et al, 2004; Medellin-Pena & Griffiths, 2009). The metabolic function of microbes helps in release of key nutrients and energy from the diet, which is used by host and microbes for growth and sustenance (Hooper et al, 2002). However information about the role of individual bacterial species in host related physiological function is limited due to the difficulty in culturing these strict anaerobes in laboratory environments (Duncan et al,

2007). Section 1.9 discusses the general properties of a few bacterial genera that have been characterized and are commonly found in the mammalian colon.

Within the gut environment the metabolic function of microbes produce short chain fatty acids (acetate, butyrate, propionate), CO_2, H_2 and CH_4. In addition to generating energy from complex dietary sources, intestinal microbes also account for uptake and storage of this energy (Hooper et al, 2002; Bäckhed et al, 2004). Intestinal microbes such as *B. thetaiotamicron* have been shown to adapt their metabolism according to nutritional availability (Bjursell et al, 2006). The metabolic characters of different bacterial groups differ significantly with some species being more efficient in one pathway of energy generation than others (Mahowald et al, 2009). Studies have shown that obesity is related to the relative distribution of *Bacteroidetes* and *Firmicutes* in the gut, with obese mice and humans having less *Bacteroidetes* and more *Firmicutes* (Ley et al, 2005; Ley et al, 2006). Within the Firmicutes, mollicutes were found in increased proportions in high fat/ high sugar diet related obese mice. In these mice the diet induced the selection of microbes that were more efficient in metabolizing the readily available carbon source (Turnbaugh et al, 2008). The intestinal microbiota has also been found to be involved in detoxification and biotransformation of toxic metals, modulation of host metabolic phenotypes, metabolism of otherwise indigestible dietary compounds and metabolism of xenobiotics that can have profound effect on host health (Diaz-Bone & Van de Wiele, 2010a; Li et al, 2008; Sidhu et al, 2001; Nicholson et al, 2005). Until recently, a number of diseases were found to be associated with a single type of microorganism such as *C. difficile* caused antibiotic associated diarrhea and *H. pylori* induced gastric carcinoma (McFarland, 2008; Correa & Houghton, 2007). However some

of the recent studies on human and mice GI tract have shown a link between changes in multiple microbial populations of the host and diseases such as autism, inflammatory bowel diseases, type I diabetes and atopy (Table 1.2). In most of these diseases the patient shows altered microbial composition and diversity of the gut microbiota when compared to healthy individuals. However it is not known whether this dysbiosis (change in microbiota) is responsible for the disease, plays a role in the progression of the disease, or is merely a result of the disease. The GI microbiota is known to be resilient but bacterial profiling after antibiotic treatment has shown that certain species may take anywhere from few weeks or months to years to reappear at normal levels (Dethlefsen et al, 2008; De La Cochetiere et al, 2005; Jernberg et al, 0000). Additionally comparative studies between germ free and pathogen free animals have shown that gut microbes contribute to the development and maintenance of other organ systems in the host by releasing metabolites that influence gene regulation (Wikoff et al, 2009). These studies have linked gut microbiota with cardiac output, immune system development, appetite control, development of nervous system, host response to stress and mood disorders (Gordan et al, 1963; Macpherson & Harris, 2004; Fetissov et al, 2008; Rhee et al, 2009; Sudo, 2006; Forsythe et al, 2010).

Manipulation of host microflora by using pro and prebiotics is rapidly increasing as a treatment of diseases such as atopic eczema, neurological defects and various gastrointestinal diseases (Kalliomäki et al, 2003; Forsythe & Bienenstock, 2008; Preidis & Versalovic, 2009). However a complete understanding of the mechanisms lying behind host microbe interaction is needed for the development of individual based treatment

strategies. Nevertheless, the microbial community residing in the GI tract contributes to host health and disease and can be a useful tool in the prognosis of the diseased state.

Host	**Disease**	**Part of GI tract studied**	**Type of bacteria affected**	**Reference**
Mice	Obesity	Caecum	Bacteroidetes(↓) Firmicutes(↑)	Ley et al, 2005
Human	Obesity	Stool	Bacteroidetes(↓) Firmicutes(↑)	Ley et al, 2006
Children (0-7yrs)[@]	Obesity	Stool	Bifidobacteria(↓) *Staphylococcus aureus*(↑)	Kalliomäki et al, 2008
Children (2-13 yrs)	Autism	feces	Bacteroidetes(↑) Firmicutes(↓) Actinobacteria(↓) Proteobacteria(↑) Overall diversity(↑)	Finegold et al, 2010
Human*	Liver Cirrhosis	feces	*Bifidobacterium*(↓) *Enterobacter*(↑) *Enterococcus*(↑) *Clostridia*(↑)	Zhao et al, 2004
IL-10 deficient 129 Sv/Ev mice [$]	Colitis	Large intestine	*Bacteroides*(↑) *Bifidobacterium*(↑) *Clostridium*(↑) *Eubacterium*(↓) *Acidophilus*(↓)	Bibiloni et al, 2005
Human	Crohn's disease and Ulcerative Colitis	colon	Bacteroidetes(↓) Lachnospiraceae(↓) Proteobacteria(↑) Bacillus(↑)	Frank et al, 2007
Human	Colorectal cancer	Colorectal Mucosal biopsies	Bacteroides(↓) Proteobacteria(↑) *Dorea*(↑) *Faecalibacterium*(↑) *Coprococcus*(↓) *Bacteroides*(↑) Overall diversity(↑)	Shen et al, 2010
Children	Type I diabetes	Stool	Bacteroidetes(↑) Firmicutes(↓) Ruminococceae (↓) Lachnospiraceae(↓) Eubacteriaceae(↓) Veillonellaceae(↑)	Giongo et al, 2011

			Bacteroideaceae(↑) Porphyromonadaceae(↓) Rikenellaceae(↓)	
Infants^	Atopic eczema	Feces	*E.coli*(↑) *C.difficile*(↑)	Penders et al, 2007
Infants*	Wheezing/ Allergy	Feces	*Bifidobacterium*(↓) *Clostridium*(↑)	Verhulst et al, 2008
Human	Type II diabetes	Stool	Bacteroidetes(↑) Firmicutes(↓) Proteobacteria(↑) Bacteroidia(↑) Prevotella(↑) Clostridia(↓) Roseburia(↓) Bacilli(↑)	Larsen et al, 2010
Preterm Infants	Nectrotizing enterocolitis	Stool	Bacteroides(↓) Firmicutes(↓) Fusobacteria(↓) Proteobacteria(↑) Overall diversity(↓)	Wang et al, 2009

Table 1.2: Changes in the composition of gut microbiota associated with various diseases. * indicates culture dependent studies, $ represents DGGE analysis, @ indicates results based on RT-qPCR and FISH, ^ represents RT-qPCR based study whereas all other studies were based on clone library. (↑) indicates increase in population and (↓) represents decrease in microbial population.

1.9 Predominant microbes in mammalian colon

Bacteroidetes

The phylum Bacteroidetes includes gram negative bacteria found in aqueous environments, soil, sediments and the intestines of animals. The phylum consists of three classes Bacteroidia, Flavobacteria and Sphingobacteria (Krieg et al, 2010). Of these, Bacteroidia is the predominant class found in the mammalian GI tract (Frank et al, 2007). Within class Bacteroidia, members of Bacteroidaceae, Porphyromonadaceae and Prevotellaceae are routinely found in the mammals (Nava et al, 2011).

Bacteroides

Members of the genus *Bacteroides* are rod shaped non-motile, gram-negative cells with rounded ends that occur singly or in pairs. The % mole GC of DNA varies from 39 to 49%. These strict anaerobes form smooth, white to gray colonies that are 1-3 mm in diameter and are nonhemolytic. These bacteria are characterized by their ability to grow on media containing 20% bile. Most of the species are chemoorganotrophic and saccharolytic but weakly proteolytic (Song et al, 2010). *Bacteroides* are notorious for antibiotic resistance and have an ability to transfer antibiotic resistant genes via transposons and plasmids. Members of this genus can utilize a wide variety of complex polysaccharides as energy sources (Xu et al, 2003). The main members of this genus are *B. thetaiotamicron*, *B. vulgaris*, *B. intestinalis*, *B. distasonis, B. ovatus* and *B. fragilis*.

Porphyromonas

Members of the *Porphyromonas* are sensitive to bile, asaccharolytic and pigmented. The % mole GC of DNA is 40−55. The species belonging to this genus are short, non-motile, non-spore forming rods or coccobacilli. Most species yield n-butyric

acid, acetic acid, propionic acid and other organic acids upon fermentation (Summanen & Finegold, 2010).

Prevotella

Members of the *Prevotella* are sensitive to bile, moderately saccharolytic and may have pigmented or non-pigmented species. The % mole GC of DNA is 39.2. The species from this genus are short, non-motile, non-spore forming anaerobes that are rod shaped (Shah et al, 2010).

Firmicutes

The phylum Firmicutes includes both low and high GC gram positive bacteria. Members of this phylum vary in shape ranging from cocci to straight, helical or curved rods. Most of the members are chemoorganotrophic, form endospores and may be motile or non-motile. The % mole GC ratio is generally < 50%. Some members can grow under thermophilic and halophilic conditions (Schleifer, 2009). The phylum consists of 26 families and 223 genera. Bacilli, Clostridia and Erysipelotrichia are the three classes that are classified under Firmicutes. Class Bacilli includes two orders: Bacillales and Lactobacillales. Class Clostridia consists of three orders: Clostridiales, Halanaerobiales and Thermoanerobacterales. Within each order there are several families however only one family has been assigned to Erysipelotrichia (Ludwig et al, 2009). The important genera that are found in the colon of mammalian hosts from this phylum include *Lactobacillus, Ruminococcus, Coprococcus, Clostridium, Dorea,* and *Roseburia.*

Lactobacillus

Lactobacillus is taxonomically classified under the order Lactobacillales and family Lactobacillaceae. Most of the species belonging to this genus are non-motile, non-

spore forming, obligate saccharoclastic and facultative anaerobes with a % mole GC content ranging from 32−55%. Optimal growth conditions are 30-40^{o}C with pH ranging from 5.5-6.2. Members of this genus cannot reduce nitrate but have the ability to use nitrite as final electron acceptor during lactate oxidation. Cells vary in shape and may exist in chains and form convex, smooth, non-pigmented opaque colonies that are 2-5 mm in diameter. These bacteria metabolize carbohydrates via Embden Meyerhof pathway and phosphogluconate pathway. Both homo and hetero fermentative species are found in this genus (Hammes & Hertel, 2009).

Ruminococcus

Member of the *Ruminococcus* are gram positive, strict anaerobes isolated from the GI tract of animals and classified under order Clostridiales and family Ruminococceae. Cells are coccids that exist in pairs or chains and are motile with 1−3 flagella. Most species are chemoorganotrophic and yield small chain fatty acids (SCFA) such as acetate, formate, lactate, succinate and ethanol upon carbohydrate fermentation. The % mole GC content of the DNA is 39-47%. Species from this genus do not have the ability to ferment amino acids or peptides (Tayayuki, 2009a).

Coprococcus

The *Coprococcus* genus is assigned to the order Clostridiales and family Lachnospiraceae. *Coprococcus* species are gram positive, non-motile, cocci that may occur in pairs or chains and are obligate anaerobes with 39-42% mole GC content in the DNA. Members are chemoorganotrophic and yield butyric, acetic, formic and propionic acid on fermentation (Tayayuki, 2009b).

Clostridia

Classified under order Clostridiales and family Clostridiaceae, *Clostridia* consists of 168 species assigned to ten clusters. With 22−53% mole GC content in DNA, these gram positive, endospore forming rods are either motile or non-motile. Most members are obligate anaerobes and chemoorganotrophs, however, some species are aerotolerant and chemoautotrophic or chemolithotrophic. The species vary in their metabolic properties from being saccharolytic to proteolytic or neither or both (Rainey et al, 2009).

Roseburia

The *Roseburia* genus is taxonomically assigned under order Clostridiales and family Lachnospiraceae. *Roseburia* species including are gram negative, anaerobic, chemoorganotrophic, motile, curved rod shaped cells that occur singly or in pairs. The % mole GC content of DNA varies from 29−42%. Members of this genus can hydrolyze and ferment starch and can use oligosaccharides such as maltose and xylose as energy and carbon source (Stanton et al, 2009).

Dorea

Members of the *Dorea* genus are gram positive, non-spore former, non-motile, obligate anaerobes that belong to order Clostridiales and family Lachnospiraceae. Cells are rod shaped that occur singly or in pairs (0.5-1.0 x 1.0-4.5 μm). Species from this genus cannot hydrolyze complex polysaccharides such as starch and cellulose. Ethanol, formate, acetate, H_2 and O_2 are produced after fermentation but no butyrate production occurs. The % mole GC content of the DNA is 40-45.6% (Blaut et al, 2009).

1.10 Role of gut microbes in nitric oxide (NO) generation

NO is a key signaling molecule that plays an important role in numerous physiological processes such as neurotransmission, vasodilation, signaling, respiration and immunity (Moncada, 1999). Within blood NO is rapidly oxidised to nitrite or nitrate by heamoglobin (Hb) (Doyle & Hoekstra, 1981). Due to a very short half life, NO mainly acts as a paracrine signaling molecule and affects the sites or cells in its close proximity (Lundberg & Weitzberg, 2005). Classically it was assumed that in humans NO is generated by eukaryotic oxygen dependent NO synthases (NOS) wherein L-arginine combines with O_2 to produce NO and citrulline (Palmer et al, 1988). However in the past decade or so, evidence for the role of host commensal bacteria as a source of NO has emerged.

In humans, diet is the major source of nitrate mainly derived from green leafy vegetables (Ysart et al, 1999). Nitrate present in the diet is reduced to nitrite by bacteria residing in the oral cavity (Lundberg & Govoni, 2004). The enzymes involved in reducing nitrate to nitrite are cytoplasmic Nas, periplasmic Nap and membrane associated Nar (Potter et al, 1999; Stolz & Basu, 2002). No mammalian homologues of these enzymes are found indicating the process is exclusive to bacteria residing in the host. Following ingestion of a nitrate rich meal the levels of nitrite in saliva increase about 10−20 fold from 50−150 μm to 1−2 mM range (Spiegelhalder et al, 1976). Nitrite obtained from any source (diet, oxidized product of NO or reduced product of nitrate) under acidic conditions of the stomach is non-enzymatically protonated to form nitrous acid (Benjamin et al, 1994). The nitrous acid thus formed is converted to various compounds including NO. In addition, xanthine oxidase (XO), aldehyde oxidase and

deoxygenated Hb have been shown to convert nitrite to NO in various cells and tissues under hypoxic conditions (Zhang et al, 1997; Zweier et al, 2010; Huang et al, 2005). In the stomach NO has been demonstrated to have a gastroprotective effect and plays a role in mucosal blood flow, mucus production and host defense against pathogen (Bjoerne et al, 2004; Petersson et al, 2007). Typical concentrations of NO in stomach lumen is 20-400 ppm but is found in very low levels in the lower GI tract (Benjamin et al, 1994; Lundberg et al, 1994).

Recent studies with human feces showed generation of NO in the presence of nitrate or nitrite at pH equivalent to those found in the lower GI tract under anaerobic conditions (Sobko et al, 2005). In vivo studies with germ free and conventional animals showed increased NO levels in the latter indicating that bacteria in the lower GI tract play a role in NO generation (Sobko et al, 2004). In the lower GI tract besides eukaryotic NO generation, gut microbes can produce substantial amounts of NO by any of the 3 processes: bacterial NOS, denitrification or dissimilatory nitrate reduction to ammonia (DNRA). Bacterial NOS have been identified in many organisms and show similarity to eukaryotic NOS (Choi et al, 1997; Yarullina et al, 2006; Gusarov et al, 2008). Species of *Streptomyces, Bacillus* and *Deinococcus* have been shown to generate NO from arginine under in vivo conditions. (Johnson et al, 2008a; Adak et al, 2002a; Adak et al, 2002b) A variety of environmental bacteria have been shown to possess either the copper containing NirK or the cytochrome *cd* nitrite reductase NirS in the periplasm and convert nitrite to NO during denitrification (Berks et al, 1995; Potter et al, 2001). Another set of bacteria possess nitrite reductase (pentaheme NrfA or siro heme NirBD) that converts nitrite to ammonia but are known to produce enzyme bound NO as an intermediate (Cole

& Brown, 1980; Einsle et al, 2002). In vitro study with *E.coli* has demonstrated that NrfA can reduce exogenously supplied NO to ammonia, an important mechanism for cellular detoxification (Poock et al, 2002). Studies to date have showed that most of the commensal bacteria reduce nitrate mainly via the DNRA pathway under conditions similar to those found in GI tract (Allison & Macfarlane, 1988; Parham & Gibson, 2000). At low concentrations of nitrate and nitrite, similar to those found in human body the periplasmic enzymes for nitrate (Nap) and nitrite reduction (Nrf) are activated. At high concentrations of nitrate and nitrite the cytoplasmic enzymes (Nar and NirBD) are expressed (Potter et al, 1999). In vitro studies involving N15 tracer experiments with human fecal microbiota showed that the most plausible explanation for the formation of NO by gut bacteria is enzymatic reduction of nitrate to ammonia wherein NO is produced along the pathway (Joan Vermeiren et al, 2009). Despite an increasing amount of evidence showing a role of gut microbes in NO generation, the significance of this alternate pathway of microbes mediated NO generation in lower GI tract is not well understood.

1.11 Overview of specific aims and hypothesis

Chronic exposure to arsenic via drinking water is a global health problem (NRC, 2001). Although, the effect of environmentally relevant chronic exposures to arsenic via drinking water on various organs has been well established, its effect on the microbiota of the GI tract is not well studied. There is evidence that suggests arsenic exposure via drinking water may affect the host microbiota either by being toxic or by affecting physiology. An in vivo study with rats has showed toxic effects of arsenic (1 mg/L) on the total bacterial count in stool samples (Choudhry et al, 2010). A wide variety of microorganisms have developed detoxification mechanism to deal with As toxicity and arsenate respiring bacteria have been isolated from the stools of Syrian Hampsters exposed to As(V) (Oremland & Stolz, 2003; Herbel et al, 2002). In vitro studies have shown that the gut microflora of the arsenic exposed individual can play a role in detoxification and elimination of the arsenic from the host body (Diaz-Bone & Van de Wiele, 2010). Thus the purpose of this work was to investigate the in vivo effect of chronic ingestion of environmentally relevant concentrations of arsenic via drinking water on the composition and physiology of the microbial community of the lower GI tract by molecular techniques. To accomplish this aim, microbial community composition and gene expression (*arsA, arsB* and *nrfA*) in colon samples from 10 and 250 ppb of As(III) exposed mice were compared to the control mice after 2, 5 and 10 weeks of As(III) exposure via drinking water. Four specific aims were proposed to test the following hypotheses:

Hypothesis 1: As(III) exposure will change the microbial community profile of mouse colon in dose and time dependent manner.

Specific Aim 1: Compare the microbial community profile of mice colon exposed to 10 and 250 ppb of As(III) for a period of 2, 5 and 10 weeks to that of control mice by RISA and ARISA.

Hypothesis 2: As exposure will promote the selection of arsenic metabolizing microbes.

Specific Aim 2: Determine the microbial community structure and composition of control and As(III) exposed mice using molecular techniques (DGGE and 16S rRNA gene based clone libraries) and perform statistical methods to determine the populations that are significantly affected by As(III) exposure.

Hypothesis 3: As (III) associated microbiota will have more widespread distribution of *ars* operon and arsenic resistant genes (*arsA* and *arsB*) will be induced in time and dose dependent manner in the presence of As(III).

Specific Aim 3: Bioinformatically determine the distribution of *ars* operon in colon microbes and design species specific primers and perform comparative Ct to compare the expression levels of *arsA* and *arsB* gene.

Hypothesis 4: Increased Nitric oxide production in endothelial cells is one of the many pathways that play a role in arsenite induced toxicity. As(III) induced altered NO level in the host may affect the physiology of colon microbes by altering the transcription of genes such as *nrfA* that are involved in NO generation or are regulated by NO levels. Thus it was hypothesized that As(III) exposure will change the physiology of colon microbes.

Specific Aim 4: To design species specific primers and perform comparative Ct for determining the expression levels of *nrfA*.

Chapter 2 : MATERIALS AND METHODS

2.1 Mice exposure

Sodium arsenite at a concentration of 0, 10 and 250 ppb was added to the drinking water and was fed for a period of 2, 5 and 10 weeks to groups of five C57BL/6NTac male mice that weighed approximately 20 g and aged 6-8 weeks. However, only four mice were exposed to As(III) for 2 week 0 ppb group to obtain colon samples. The 0 ppb group for each time point was considered as the control group. Sodium arsenite solution was prepared using the regular tap water supplied by the Municipality of Pittsburgh, PA. The mice from each group were euthanized by IP injection of sodium pentobarbital and their colons were removed. All mouse exposure experiments were done at the University of Pittsburgh by members of Dr. Aaron Barchowsky's group in accordance to the institution's guidelines for animal safety. Freshly removed colon samples were flash frozen in liquid nitrogen and stored at -80^{o}C until used. The microbial composition and structure of these mice was characterized using molecular approaches and compared to those of control mice (0 ppb group). These exposure experiments were performed twice on two different sets of mice (study 1 and 2). Due to the fewer number of mice and absence of 10 ppb As(III) exposed mice, the results of study 1 are described in appendix 1. From here on things relevant to study 2 are mentioned.

2.2 DNA extraction

Microbial community DNA was isolated from 2, 5 and 10 week mouse colon samples (control and As(III) exposed) using the QIAamp DNA stool mini kit according to the manufacturer's instructions (Qiagen Inc., Valencia, CA, USA). Flash frozen

samples were thawed on ice and about 40 mg of each colon sample was homogenized in a 2.0 ml eppendorf tube using a pipette tip followed by addition of 1.4 ml of buffer ASL (provided with the kit) for cell lysis. The samples were vortexed for 1 min to obtain uniform suspension followed by heating at 80^{o}C for 5 minutes. The tubes were centrifuged at 10,000 *x* g for 2 minutes. The supernatant was transferred to a new eppendorf tube and 1 inhibit EX tablet was added to remove PCR inhibitors from the colon samples. The tubes were vortexed for 1 minute, incubated at room temperature for one minute and centrifuged at 10,000 *x* g for 6 minutes. The supernatant was transferred to a fresh tube and 30 µl of proteinase K and buffer AL (equal to the volume of supernatant) was added. After vortexing the tubes for few seconds the samples were incubated at 70^{o}C for 10 minutes. DNA was precipitated from the above lysate by adding equal volumes of 100% ethanol. DNA was purified from the above lysate by passing the lysate through a QIAamp spin column and spinning the column at 10,000 *x* g for 1 min. The column was washed twice with 500 µl of buffer AW1 and once with 500 µl of buffer AW2. Each washing was followed by centrifugation at 10,000 *x* g for the removal of filtrate. After all the washings, DNA was eluted into a 1.5 ml eppendorf tube with 200 µl of buffer AE by centrifuging the column at 10,000 *x* g for 1 minute. The quantity of nucleic acid was estimated by using Perkin Elmer UV/Vis spectrophotometer lambda 2S (260/280 nm, 50 µg DNA = O.D. of 1). A 0.8% agarose gel was used to check for high MW genomic DNA. The extracted DNA was then used for the downstream processes.

2.3 Ribosomal intergenic spacer analysis (RISA)

Bacterial community estimation by RISA and ARISA technique were based on the principle that the length of the spacer region between 16S and 23S rRNA gene varies

from species to species (Borneman, J. and Triplett, E. W., 1997). However RISA and ARISA estimation of bacterial community in a sample can differ from the actual estimation as the same bacterial species can have a different spacer length or different bacterial species can have same spacer length (Kovacs et al, 2010). For this study differences in RISA and ARISA profiles were assumed to reflect the differences in bacterial community of the samples and were used for comparison of samples and bacterial diversity estimations. Although both RISA and ARISA are based on the same principle, the output of the results differs between the two techniques. RISA results are interpreted from a polyacrylamide gel whereas ARISA results are obtained from a genetic analyzer. In this study different primer pairs were used for the two techniques.

The spacer region between the 16S and 23S rRNA gene was amplified from the microbial community DNA of each of the colon sample using ITSF (binds to 3' end of 16S rRNA gene) and ITSReub primer pairs (specific for 23S rRNA gene) (Table 2.1) (Cardinale et al, 2004). The reaction mixture consisted of 150 ng of genomic DNA, 10 µl of 5X buffer, 35 pmol each of forward and reverse primers, 1 µl dNTPs from a 200 µM stock, 1.25U of GoTaq DNA polymerase (Promega, Madison, WI, USA) and sterile water to a final volume of 50 µl. The touchdown program was used for amplification to increase the diversity of microbial estimation (Hecker & Roux, 1996). The reaction was performed on Techgene thermal cycler (Techne incorporated, Princeton, NJ, USA) with initial denaturation at 94^{o}C for 5 minutes followed by 20 cycles of denaturation at 94^{o}C for 1 minute, annealing temperature decreasing from 60^{o}C to 50^{o}C at a rate of 0.5^{o}C decrease/cycle for 1 minute, extension at 72^{o}C for 1 minute then 15 cycles of denaturing (1 min), annealing (1 min) and extension (3 min) at 94^{o}C, 50^{o}C and 72^{o}C and a final

extension of 7 minutes at 72^{o}C. The absence of contaminants in the reagents was confirmed by using no template (water) as a negative control. After PCR the products were checked for amplification by visualizing on a 1.8% agarose gel after ethidium bromide staining.

Amplified products of the 16S-23S intergenic region from 2, 5 and 10 weeks mouse colon samples were loaded separately on a 5% polyacrylamide gel (Appendix 2). The gel was run at 70V for 105 minutes in 1X TBE buffer on a Mini-PROTEAN gel system (Biorad, Hercules, CA, USA) and the products were separated on the basis of their length. The resulting gels were stained in ethidium bromide solution (0.5 µg/ml EtBr in DI water) and were photographed using a Kodak gel documentation and analysis system 120.

2.4 Dendrogram analysis

The digital images of RISA and DGGE gels from 2, 5 and 10 weeks were loaded into the Quantity one software (Biorad, Hercules, CA, USA). The background was subtracted from the images, bands were detected automatically as well as manually and matching of bands between the lanes was done with 1% matching tolerance. Similarity values for the community was obtained with the Quantity one software using a similarity matrix calculated using the Dice coefficient based on the presence and absence of bands. Dendrograms were constructed by the unweighted pair group method (UPGMA) using similarity matrix. The dice coefficient is a similarity coefficient calculated by the following formula:

$$D_{sc} = [2j / (a+b)], \quad \text{---------- (2.4.1)}$$

Where a = number of DGGE bands in lane 1,

b = number of DGGE bands in lane 2, and

j = number of DGGE bands that are common in lane 1 and lane 2.

A D_{sc} value of 1 indicates that the DGGE profiles are identical (Dice, 1945). UPGMA is a hierarchical clustering method in which the two most similar lanes are grouped together based on the highest similarity index to form a composite group and sequentially more such groups are determined until only two groups are left that are then joined to form a rooted dendrogram (Sokal, R. and Michener, C., 1958).

2.5 Automated ribosomal intergenic spacer analysis (ARISA)

PCR for ARISA was done following the protocol used by Fisher and Triplett (1999). The reaction mixture consisted of 150 ng of genomic DNA, 5 µl of 5X buffer, 2.5 mM $MgCl_2$, 0.5ug of BSA per µl, 12.5 pmol each of forward and reverse primers, 0.5 µl dNTPs from a 200 µM stock, 1.25U of GoTaq DNA polymerase (Promega, Madison, WI, USA) and sterile water to a total volume of 25 µl. The primers used were 1406F and 23Sr, with 1406F being labeled at the 5'end with the phosphoramidite dye 5-FAM (Table 2.1). The reaction mixture was denatured at 94^{o}C for 2 minutes, followed by 30 cycles of denaturation at 94^{o}C for 15 sec, annealing at 55^{o}C for 15 sec, extension at 72^{o}C for 45 sec and a final extension at 72^{o}C for 2 min. Water was used as negative control to check for the absence of contamination in the reagents.

Equal amounts of PCR product along with 0.75 µl of the internal size standard LIZ-1200 (Applied Biosystems, Foster City, CA, USA) and 8 ul of DI formamide were added to final volume of 10 µl and the reaction mixture was denatured at 95^{o}C for 5 minutes followed by cooling on ice. LIZ-1200 is a size standard containing 68 LIZ dye labeled single stranded DNA fragments that range in size from 20 to 1200 base pairs. The

fragments were separated using a 3100-Avant Genetic Analyzer under gene scan mode and the results were analyzed by using the peak scanner software (Applied Biosystems, Foster City, CA, USA). An electropherogram with a series of peaks was obtained for each sample wherein each peak represents a different operational taxonomic unit (OTU). The program provides the sizes of these peaks based on the fragment sizes of the internal standard along with the fluorescence of each peak in the form of peak height and peak area. In order to remove the background noise, ARISA peaks exceeding the peak height of 50 fluorescence units were considered for analyses. All the experiments were done in triplicates and the peaks for individual sample replicates were averaged for relative peak height and relative peak area using TREX software (Culman et al, 2009). The heat maps for the relative peak height dataset were generated using permut matrix software (Caraux & Pinloche, 2005). The matrices obtained for presence/absence of peaks, relative peak height and area were used for Principle Component Analysis (PCA) using Ade 4 software package (Chessel et al, 2004). PCA is an ordination method that determines two major principle components from multivariate datasets under the constraint that these principle components are linear combination of all variables and account for most of the variation observed in the variables (Pearson K., 1901). In the present study different OTUs (presence/absence matrix) and their abundance (peak height and area matrix) were considered as variables.

2.6 Denaturing gradient gel electrophoresis (DGGE)

DGGE is one the most widely used molecular technique to determine the microbial community of environmental samples. This technique is based on the principle that under denaturing conditions DNA fragments of equal sizes denature differently on

the basis of their % mole GC content. Such fragments when run on a denaturing gel will move differently. The high % mole GC fragments, due to less denaturation, migrate towards the bottom of the gel whereas the low % mole GC fragments stay towards the top of the gel due to high denaturation (Fischer & Lerman, 1983). In order to protect the two strands of DNA fragments from falling apart, a GC clamp (stretch of about 40 GC rich nucleotides) is added to one of the primers used during amplification. In this study, similar sized PCR products amplified from same region of the 16S rRNA gene were used for comparison of microbial community among colon samples. Since GC content of the same 16S rRNA region varies from species to species the profile obtained from DGGE is an indicator of microbial diversity. However like RISA and ARISA, DGGE estimation of bacterial community in a sample can differ from the actual estimation as the same bacterial species can have multiple 16S rRNA genes and each gene within a species can have a different GC content or different bacterial species can have the same GC content (Muyzer & Smalla, 1998).

DGGE was performed according to protocol followed by Muyzer et al (Muyzer et al, 1993) with some modifications. Variable region 3 (V3) of the 16S rRNA gene was amplified from the colon bacteria using HDA1-GC and HDA2 primers (Table 1) (Walter et al, 2001). The reaction mixture consisted of 150 ng of genomic DNA, 10 µl of 5X buffer, 25 pmol each of forward and reverse primers, 1 µl dNTPs from a 200 µM stock, 1.25U of GoTaq DNA polymerase (Promega, Madison, WI, USA) and sterile water to a total volume of 50 µl. The touchdown program was used for amplification with annealing temperature decreasing from 60^{o}C to 50^{o}C at a rate of 0.5^{o}C decrease/cycle. Absence of contamination in reagents was confirmed by using sterile water as negative control. After

PCR, the amplification was confirmed by visualizing the products on a 1.8% agarose gel after EtBr staining.

Amplified products of 16S rRNA gene from 2, 5 and 10 weeks mouse colon samples were loaded separately on a 10% acrylamide-bis acrylamide denaturing gel containing 40% - 60% Urea- formamide gradient (Appendix 3). A non denaturing 10% acrylamide spacer gel was applied on top of the denaturing gel and was allowed to polymerize for 2 hours. The gel was run at 70V for 990 minutes at 60^{o}C in 1X TAE buffer using a DCode universal mutation detection system (Biorad, Hercules, CA, USA). The resulting gels were stained in a 1: 10,000 dilution of SYBR green in 1X TAE buffer and were photographed using Kodak gel documentation and analysis system 120.

2.7 Clone library Construction

Near full length 16S rRNA genes were amplified from the microbial community DNA isolated from each of the 2, 5 and 10 week colon sample using universal primers 8F (Edwards et al, 1989) and 1492R (Stackebrandt & Liesack, 1993). The reaction mixture consisted of 50 ng of genomic DNA, 4 µl of 5X buffer, 10 pmol each of forward and reverse primers, 0.4 µl dNTPs from a 200 µM stock, 0.5U of GoTaq DNA polymerase (Promega, Madison, WI, USA) and sterile water to a total volume of 20 µl. The touchdown program was used for amplification with annealing temperature decreasing from 60^{o}C to 50^{o}C at a rate of 0.5^{o}C decrease/cycle. The PCR products obtained from each of the sample were ligated into the pCR2.1®-TOPO vector. The ligation reaction mixture consisting of 4 µl of PCR product, 1 µl of salt solution and 1 µl of TOPO vector provided with the kit was incubated for 15 minutes at room temperature and was used immediately or stored at −20^{o}C until transformation was done.

Transformation

Ligation reaction (2 µl) described above was added to a tube of OneShot top10 chemically competent *E.coli* cells provided with the kit. The reaction was incubated for 15 to 30 minutes on ice, and then heat shocked in a water-bath at 42^{o}C for 30 seconds. To the transformed cells, 250 µl of SOC media (provided with the kit) was added and the tube was incubated horizontally on a shaker at 37^{o}C and 200 rpm for one hour. Transformed cells (50 µl) were then spread on prewarmed ampicillin selective, X-gal spread LB plates and 50 µl of untransformed cells were spread on an LB plate to verify cell viability (Appendix 4). The plates were incubated at 37^{o}C overnight. Approximately 24 positive clones were picked randomly from each sample and transferred with a sterile toothpick to a fresh ampicillin selective plate and 5 ml of LB broth containing ampicillin at a concentration of 100 µl/ml. The plates and the LB broth inoculated with single colony were then incubated at 37^{o}C overnight.

Plasmid extraction

Plasmid extraction was performed using the Promega Wizard SV DNA column purification kit (Promega, Madison, WI, USA). The culture obtained from overnight grown colonies in LB-amp media was spun down at 7,000 rpm for 5 minutes. The pellet was then resuspended in 250 µl of cell resuspension buffer and the cells were lysed with 250 µl of cell lysis buffer. The mixture was incubated for 3-5 minutes at room temperature and 10 µl of alkaline protease was added, followed by 5 minute incubation. Neutralization solution (350 µl) was then added, and the tubes were spun at 12,000 rpm for 15 minutes. The supernatant was removed to a fresh tube and spun a second time under the same conditions. The supernatant from the second spin was added to the

column. The column was washed twice with wash buffer, and the plasmid was eluted with 100 µl nuclease free water and stored at −20°C.

Sequencing from plasmid DNA

The purified plasmid DNA was sequenced using M13 forward and reverse primers at the Genomics and Proteomics Core Laboratory, University of Pittsburgh, PA using ABI 3730 sequencer (Applied Biosystems, Foster City, CA, USA).

2.8 Sequence analysis

Sequences obtained from the clone libraries were analyzed in Chromas version 2.01 (http://www.technelysium.com.au/chromas_lite.html). Sequences obtained from the reverse primer were converted into forward reading sequence and were aligned to the forward complementary strand using CLUSTAL W (Larkin et al, 2007). Low quality sequences with ambiguous bases (N) were not analyzed further. The consensus sequences were edited to exclude the primer binding sites and were compared to those in the GenBank databases by using the BLAST (Johnson et al, 2008b). The sequences were assigned to taxonomic classification using the Ribosomal Database Project II (RDPII) classifier with a confidence threshold of 90% (Wang et al, 2007). Multiple sequence alignment of nearly full length 16S rRNA sequences obtained from clonal libraries was done by Clustal X (Larkin et al, 2007) and NAST alignment (DeSantis et al, 2006). Sequences aligned by Clustal X were edited manually using seaview program (Gouy et al, 2010) to check for any ambiguity in the aligned sequences. All sequences were checked for possible chimeric artifacts using the Bellerophon (DeSantis et al, 2006), mallard (Ashelford et al, 2006) and chimera slayer programs (Schloss et al, 2009). Significant difference in the composition of control, 10 ppb and 250 ppb As(III) exposed

clone libraries was determined using LIBSHUFF program of Mothur software (Schloss et al, 2009). Comparison between libraries at the taxonomic level was also done by libcompare tool from RDP II database (Cole et al, 2009).

2.9 Phylogenetic analysis and diversity estimation

The distance matrix of nearly full length 16S sequences was prepared with DNADIST program from PHYLIP package (Felsenstein, 1989) based on Jukes-Cantor model (a substitution model based on the assumption that the rate of transition and transversion is same) (Jukes & Cantor, 1969). A neighbor-joining tree was generated using MEGA 4.0 software (Tamura et al, 2007) and the evolutionary distances were computed using Jukes-Cantor model. Based on distance matrix generated with Phylip package, clones with 97% (species level), 95% (genus level) and 90% (family level) sequence similarity were grouped into the same phylotype/operational taxonomic units (OTUs) using Mothur program (Schloss et al, 2009). Good's coverage (C) of the clone libraries was calculated by the formula:

$$C = (1 - n/N)*100, \quad \text{---------- (2.10.1)}$$

Where n= no. of sequences that are found only once in a library and

N= Total number of sequences in a library (Good, 1953).

Microbial richness and diversity in each of the clone library and the combined datasets was estimated by rarefaction curve, Chao I estimator, Shannon and Simpson diversity index using the Mothur program (Schloss et al, 2009). Rarefaction curve is based on species richness (total no. of species in a sample) and gives information about sampling efforts in a clone library and can be used to compare species richness among different clonal libraries for a given level of sampling effort. Rarefaction curve represents

the average number of OTUs/species observed when **n** clone are sequenced from the same library (Gotelli & Colwell, 2001). Steep slope of the rarefaction curve indicates that a large fraction of the species diversity remains to be discovered whereas flattening of the curve towards end of the sampling indicates that enough clones have been sequenced and additional sequencing will only yield fewer new OTUs. Chao I is a measure of species richness and is based on the concept that the number of rare species in a clone library are the best indicator of diversity. Chao I is calculated by the formula:

$$S = S_{obs} + (a^2/2b), \quad \text{---------- (2.10.2)}$$

Where S_{obs} = number of OTUs observed,

a = number of OTUs observed only once,

b = number of OTUs observed only twice and

S = estimated number of OTUs in a clone library (Chao, 1984).

Microbial diversity in a clone library is measured by taking into account two main factors: richness and abundance. Shannon and Simpson diversity index are the two most commonly used measures of diversity but differ due to the weightage given by the two methods to species richness and abundance in the mathematical calculations. Simpson diversity index gives more weightage to abundance whereas Shannon diversity index gives equal weight to richness and abundance. Simpson diversity index is calculated by the formula:

$$D = 1 - \Sigma(n_i/N)^2, \quad \text{---------- (2.10.3)}$$

Where n_i = proportion of OTU *i* relative to the total number of OTUs (N).

D ranges from 0 to 1 and decreases with increasing diversity (Simpson, 1949).

Shannon diversity index is calculated by the formula:

$$H = -\Sigma\, p_i \ln p_i, \qquad \text{---------- (2.10.4)}$$

Where p_i = proportion of sequences found in the ith OTU.

H ranges from 1.5 to 5 and increases with increasing richness and evenness (Weaver & Shannon, 1949).

Data deposition

The bacterial 16S rRNA gene sequences obtained in this study are deposited in Genbank database with accession numbers HQ681318 - HQ681412 (2 week control), JN012608 - JN012688 (2 week 10 ppb), HQ681413 - HQ681511 (2 week 250 ppb), JN012882 - JN012991 (5 week control), JN012689 - JN012791 (5 week 10 ppb), JN012792 - JN012881 (5 week 250 ppb), JN001204 - JN001304 (10 week control), JN012992 - JN013087 (10 week 10 ppb) and JN013088 - JN013183 (10 week 250 ppb).

2.10 RNA extraction and cDNA preparation

For each time (2, 5 and 10 weeks) and dose combination (0, 10 and 250 ppb) set of four mice were euthanized. The colon was removed and stored in 1 ml of RNA later solution (Qiagen Inc., Valencia, CA, USA) at 4^{o}C for 1 day and then transferred to −20^{o}C until used. For extraction of total RNA the sample was removed from RNA later solution right before processing. Total RNA was extracted from 250 mg of colon sample using trizol (Invitrogen, Carlsbad, CA, USA) according to manufacturer's instruction with some modifications. Each colon sample was homogenized using an electrical hand held homogenizer in the presence of 250 µl of trizol. After homogenization, additional 1.25 ml of trizol was added and the samples were vortexed for 2-3 minutes to achieve uniform suspension. The samples were centrifuged at 12,000 *x* g for 10 minutes and the supernatant was transferred to new 2.0 ml centrifuge tube. Chloroform (0.3 ml) was

added to the supernatant and the tubes were shaken vigorously. The samples were incubated at room temperature for 2-3 minutes followed by centrifugation at 12,000 *x* g for 15 min at 4^{o}C. Upper aqueous phase that contains the RNA was transferred to a new 2.0 ml centrifuge tube. RNA was precipitated by adding 675 µl isopropanol and few grains of glycogen. The tubes were incubated at room temperature for 10 minutes and subsequently centrifuged at 10,000 *x* g. The supernatant was removed and the RNA pellet was washed with 1 ml of 75% ethanol prepared from RNase free water and centrifuged at 7,500 *x* g for 5 minutes. The RNA pellet was dried in air and then resuspended in 35 µl of RNase free water. Extracted RNA was quantified by reading the absorbance at 260 and 280 nm using Perkin Elmer UV/Vis spectrophotometer lambda 2S (260/280 nm, 40 µg RNA = O.D. of 1). To remove any residual DNA in the extracted RNA, 15 µg of RNA was treated with 6 Units of DNase I (Qiagen Inc. Valencia, CA, USA) in a total volume of 20 µl and incubated at 37^{o}C for 20 minutes. DNase activity was stopped by adding 2 µl of 25 mM EDTA to a total volume of 20 µl and incubating at 65^{o}C for 10 minutes.

Equal amounts of DNase treated RNA (~ 4µg) was used as starting material for the conversion to cDNA. Superscript III first strand synthesis super mix (Invitrogen, Carlsbad, CA, USA) was used for obtaining cDNA following the manufacturer's instructions. To ~ 4 µg of DNase treated RNA, 50 ng of random hexamer, 1 µl of 10 mM dNTP mix and RNase free water was added to a total volume of 10 µl. Annealing of the random hexamers to the RNA was facilitated by incubating the reaction at 65^{o}C for 5 minutes. cDNA synthesis mix containing reverse transcriptase (10 µl) was added to the reaction and the tubes were incubated at room temperature for 10 minutes, at 50^{o}C for 50

minutes and finally at 85^{o}C for 5 minutes to stop the reaction. To verify the absence of DNA, RNA samples without reverse transcriptase (-RT) were used as negative controls.

2.11 Primer design and optimization

Primers for amplification of target genes (*arsA, arsB* and *nrfA*) and internal control/reference genes (*rpoB*) from *Bacteroidetes thetaiotamicron* were designed using the primer3 program (Table 2.1) (Rozen & Skaletsky, 2000). The sequences for these genes were obtained from the DOE Joint Genome Institute (http://img.jgi.doe.gov/). The specificity of each primer pair was tested by amplifying the gene from microbial community DNA from murine colon and confirmed by: (i) visualizing the expected product size on 1.8% agarose gel and (ii) sequencing the amplified product and performing a BLAST search. Optimal concentration of the primer pairs was determined by comparing the Ct and relative fluorescence (Rn) values of both forward and reverse primers at 4 different concentrations: 2.5, 5.0, 7.5 and 10 pmol/µl. The primer concentration that resulted in low Ct and high Rn value was used for relative quantification. Efficiency of each primer pair was tested by preparing a 10 fold dilution series of the template and amplifying the template with its specific primer pair. The cycle threshold (Ct) value for each dilution was determined by real time PCR system and plotted against log template dilution. The slope of the line was determined and PCR efficiency of the primer pair was determined by the formula:

$$m = -(1/\log E) \quad \text{---------- (2.12.1)}$$

where m is slope and E is efficiency.

The efficiency of the internal control gene and target genes were within 10% of each other. Expression of target genes and internal control genes in colon cDNA samples

was tested by regular PCR and visualizing the PCR product on a 1.8% agarose gel. Specificity of the primer pair was again checked by sequencing the amplified product. Relative quantification of expression levels of target genes among different samples was done by comparative Ct method (ΔΔCt) on step one real time PCR system.

2.12 Relative quantification by Comparative Ct

Real-time PCR reactions were prepared in 96 well, thin walled microplates and carried out on step one real time PCR system (Applied Biosystems, Foster City, CA, USA). The reaction mixture consisted of 7.5 µl of power SYBR green PCR master mix (Applied Biosystems, Foster City, CA, USA), 5 pmol each of forward and reverse primer, 3 µl of cDNA and sterile water to a final volume of 15 µl. Each reaction was performed in triplicate under thermocycling conditions consisting of initial step at 95^{o}C for 5 minutes followed by 50 cycles of 95^{o}C for 15 sec, 57^{o}C for 25 sec and 72^{o}C for 30 sec with data collection point. To check for the specificity of each reaction a final melt curve analysis was performed at the end of the amplification cycles with fluorescence data collected at 0.3^{o}C intervals between 60 and 100^{o}C. The negative controls included the no template (water) and samples without any reverse transcriptase added. The Ct values of target genes (*arsA, arsB* and *nrfA*) were normalized to internal control gene *rpoB* (ΔCt) and were subsequently normalized to the control (0 ppb) mouse sample (ΔΔCt). The ΔΔCt values were then transformed to $2^{-\Delta\Delta Ct}$ which was used for statistical analysis. The value of $2^{-\Delta\Delta Ct}$ was obtained from the software provided with the step one real time PCR system (Applied Biosystems, Foster City, CA, USA).

2.13 Statistical analysis of real time data

Dose dependent changes in *arsA*, *arsB* and *nrfA* expression in mouse colon were compared by one way analysis of variance (ANOVA) followed by Tukey's correction. The statistical analysis was performed using Graphpad Prism 5.0 software (Graphpad, SanDiego, CA, USA).

Gene	Primer	Sequence (5' to 3' end)	Tm (^{o}C)	Reference
16S-23S rRNA	ITSF ITSReub	GTCGTAACAAGGTAGCCGTA GCCAAGGCATCCACC	54.1 53.8	Cardinale et al, 2004
	1406F 23Sr	TGYACACACCGCCCGT GGGTTBCCCCATTCRG	59.1 53.6	Fisher & Triplett, 1999
16S rRNA	HDA1* HDA2	ACTCCTACGGGAGGCAGCAGT GTATTACCGCGGCTGCTGGCAC	62.6 63.1	Walter et al, 2001
	8F 1492R	AGAGTTTGATCCTGGCTCAG GGTTACCTTGTTACGACTT	52 47	Edwards et al, 1989 Stackebrandt & Liesack, 1993
arsA	BTarsAF-1089 BTarsAR-1211	TGACCCGGCAAATCACCTGAACTA GTTTCGGCAGCTTTGCTACGAACT	60.4 60.1	This study
arsB	BTarsBF-684 BTarsBR-827	GGTTGGCGGATTGCTATTAACGCT GCCCAACCGTAGGCAACAAAGAAT	59.9 60.2	This study
nrfA	BTnrfAF-376 BTnrfAR-538	TCACTTCGTACCGGTTCTCC TCGGATTCACGATTTCATCA	56.4 51.6	This study
rpoB	BTrpoBF-477 BTrpoBR-661	CAAATTTCACGCCCAAAGTT AGGTCGTCACGGTAACAAGG	52.2 56.9	This study

Table 2.1: List of primers used in this study. * indicates addition of GC clamp (CGCCCG GGGCGCGCCCCGGGCGGGGCGGGGGCACGGGGGG) towards 5' end for DGGE. F and R stand for forward and reverse primer respectively.

Chapter 3 : RESULTS

3.1 Estimation of bacterial community profile by RISA

The spacer regions between the 16S and 23S rRNA genes were amplified from 2, 5 and 10 week mouse colon community DNA to rapidly compare the bacterial profile of control and As(III) exposed mouse colon. PCR products thus obtained were separated on a 5% polyacrylamide gel on the basis of their size. The product size from spacer region amplification ranged from 300 bps to > 1500 bps. In addition to the spacer region, the primers used for RISA study also amplify approximately 120 bps from the 16S and 23S rRNA genes (Cardinale et al, 2004). Therefore the product range from 300 to 1500 bps on the polyacrylamide gel corresponds to an intergenic region ranging from 150 to 1350 bps. Variations in RISA profiles in the form of absence/presence of bands and band intensity were observed in As(III) treated groups and control mice groups and also between individuals of the same group. However, these alterations in band pattern and intensity were more consistent between the groups rather than within the same group. Due to a high number of bands and low resolution of these bands on the gel, dendrogram analysis of RISA profiles was done to confirm that the variations observed in band patterns and intensity were linked to arsenic exposure and not a result of inherent variation between the individual mice (Figure 3.1).

Dendrogram analysis of RISA gels showed separate clustering of control and 250 ppb As(III) exposed groups after 2, 5 and 10 weeks of arsenite exposure. However, samples from 10 ppb As(III) exposed groups either clustered with both control and 250 ppb exposed samples (as in case of 2 and 5 week samples) or formed a cluster with 250

ppb group as in 10 weeks of As(III) exposed mice (Figure 3.1). At each time point (2, 5 and 10 week) three separate clusters were formed in each of the dendrograms.

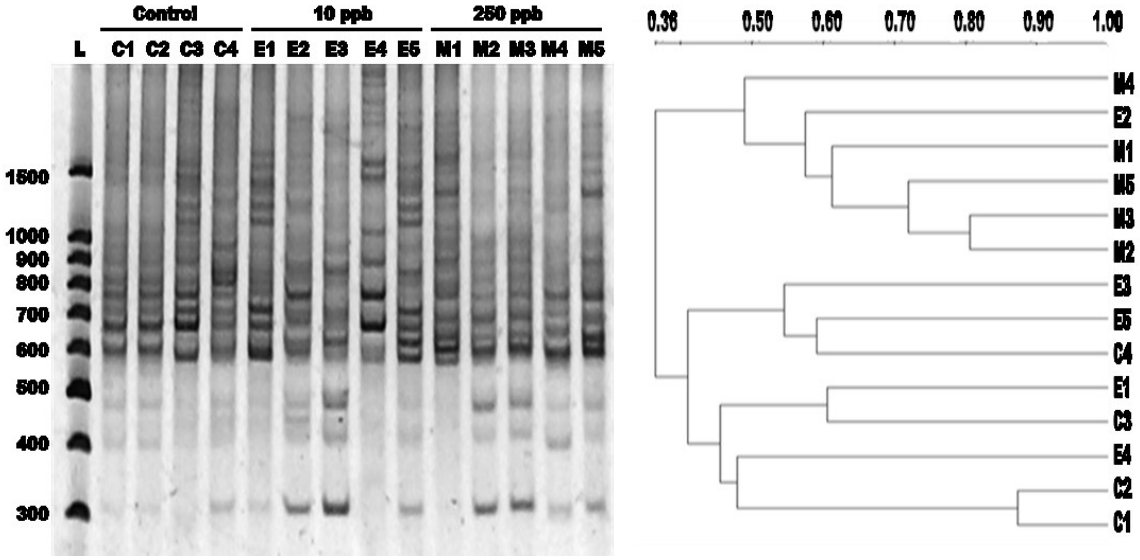
The mice from the 2 weeks of 250 ppb of As(III) exposed group formed a separate cluster with a similarity of 50% between the individuals and showed only 36% similarity with the rest of the control and 10 ppb group. However one mouse from the 10 ppb group clustered with 250 ppb group. This was expected from the 10 ppb group as they are in transition phase from normal control group to a group that has been affected by arsenic exposure.

In the 5 week RISA dendrogram, the three clusters within the dendrogram were separately predominated by the three groups (control, 10 ppb and 250 ppb). However some level of mixed clustering was observed between the groups. Four out of five mice from the 250 ppb As(III) exposed group clustered together and showed > 50% similarity whereas the fifth mouse clustered with the 10 ppb As(III) exposed mice.

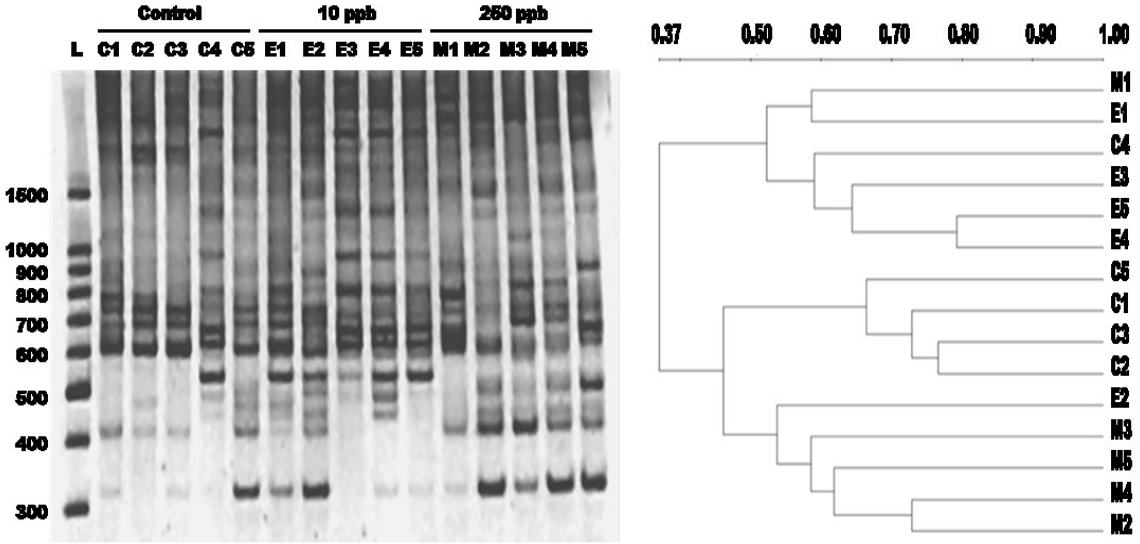
The dendrogram from the 10 week RISA showed separate clusters for control and 10 ppb group and a third cluster with all the mice from 250 ppb exposed group and two mice from 10 ppb exposed group. The control group shared > 40% similarity to each other but only 20% similarity with the 10 and 250 ppb groups.

These results suggest that changes in microbial profile as determined by RISA in response to 250 ppb of As(III) are dramatic after exposures for 2 weeks and continue till 10 weeks of 250 ppb exposure. Microbial changes in response to 10 ppb of As(III) exposure although started appearing at week two but became distinct and consistent only after 10 weeks of As(III) exposure.

A. 2 week



B. 5 week



C. 10 week

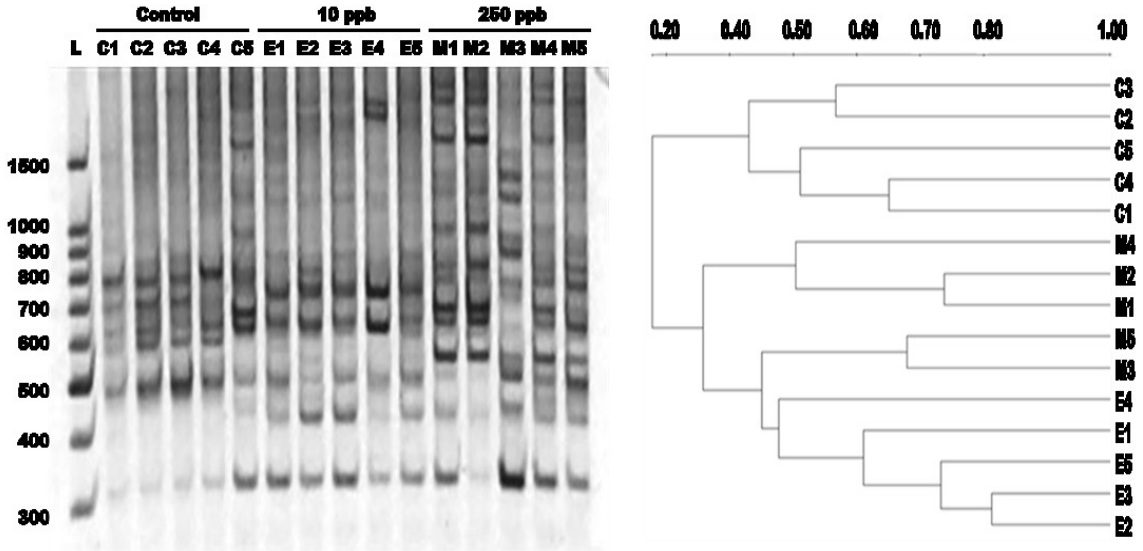


Figure 3.1: RISA gels and dendrogram analysis of negative image of RISA fingerprints generated from 16S-23S intergenic region of control (C1-C5), 10 ppb (E1-E5) and 250 ppb (M1-M5) arsenic exposed mice at A) 2 week, B) 5 week, and C) 10 week. L= 100 bp DNA ladder. Numbers on left side of the DNA ladder represent size of the marker band in base pairs. Scale bar on dendrograms represents similarity index values.

3.2 Bacterial community composition of mouse colon as determined by ARISA

The spacer region between the 16S and 23S rRNA genes was amplified from 2, 5 and 10 week mouse colon community DNA. The PCR products thus obtained were separated on a genetic analyzer to obtain a community profile for each colon sample. A representative ARISA electrophoregram showing bacterial peaks obtained from mouse colon is depicted in Figure 3.2 .

Regardless of time and dose exposures, ARISA electrophoregrams of the mouse colon showed a total of 117 OTUs (peaks) with fragment size ranging from 287 bps to 1155 bps representing intergenic spacer region of approximately 37 bps to 905 bps. The majority of the peaks ranged from 500 – 800 bps in size. Peaks shorter than 287 bps were observed in a small number of samples. These peaks were attributed to nucleotides, primers, primer dimers and chimeras and were excluded from further analyses. The reproducibility of ARISA profiles was high among the replicates of the same sample. In general, peaks with high fluorescence intensity were present constantly across all PCR replicates of a sample, however, a very small number of less intense peaks (< 50 fluorescence units) varied in between the replicates. To overcome this variability and to reduce the background noise, peaks less than 50 fluorescence units were not included in the analyses. For all time points, a number of peaks were identified that were present only in control samples and not in experimental mice and vice versa. In addition some of the peaks that were present in all three dose groups differed in intensity between the three sets of mice groups (0, 10 ppb and 250 ppb). Figure 3.3 shows heat maps (based on relative peak height) obtained from ARISA profiles of colon samples.

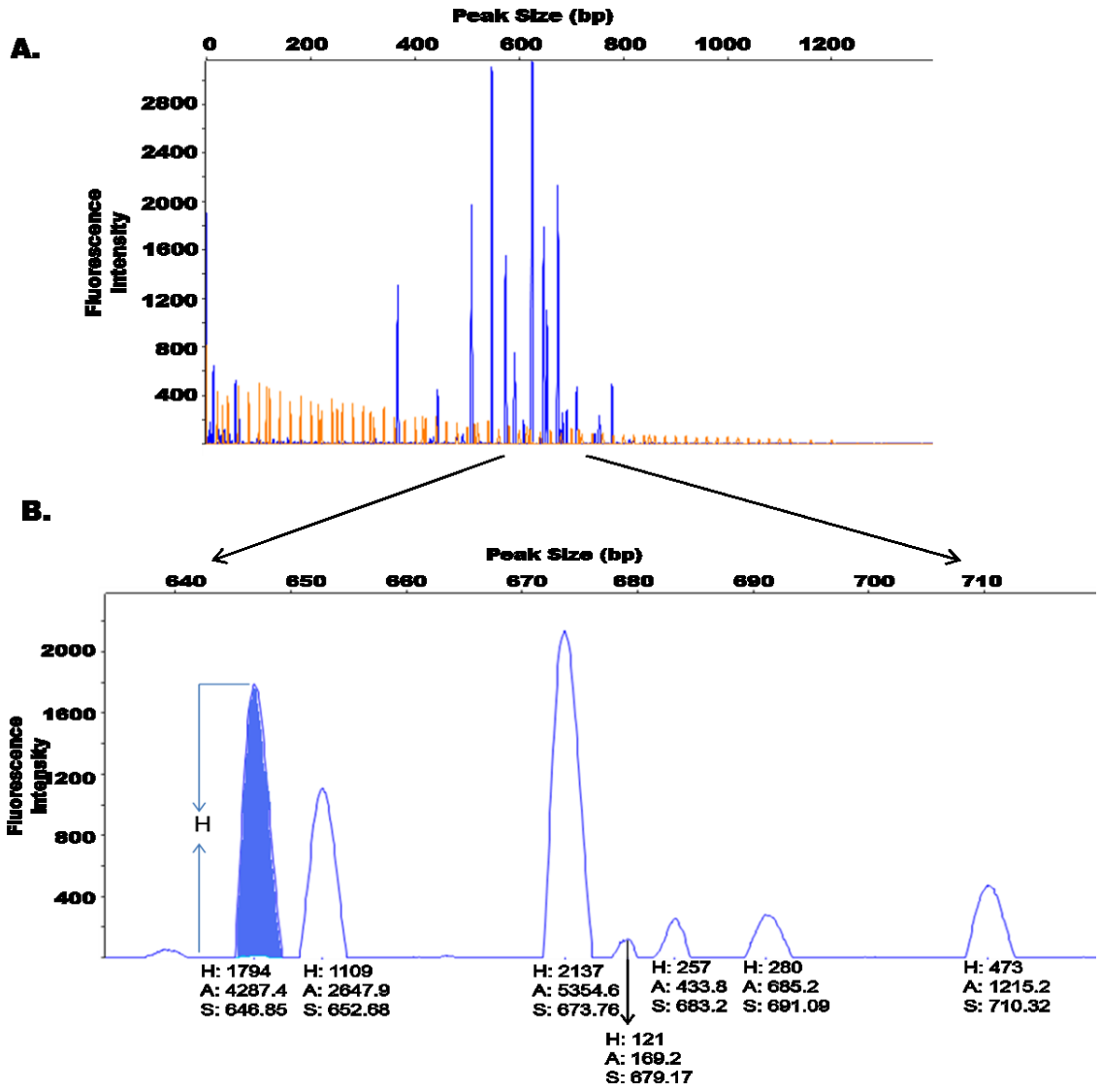


Figure 3.2: A typical bacterial community ARISA electropherogram obtained from mouse colon. X axis at the top shows size of the intergenic region in bps and Y axis corresponds to fluorescence intensity. (A) Full view of ARISA profile obtained from mouse colon. Orange lines are internal size standard peaks and blue lines are intergenic region peaks (B) Zoomed view of ARISA profile ranging from 640 to 710 bps. H= Peak height, A= Peak area (shown with filled blue) and S=Peak size. Values for H, A and S are shown at the bottom of each peak.

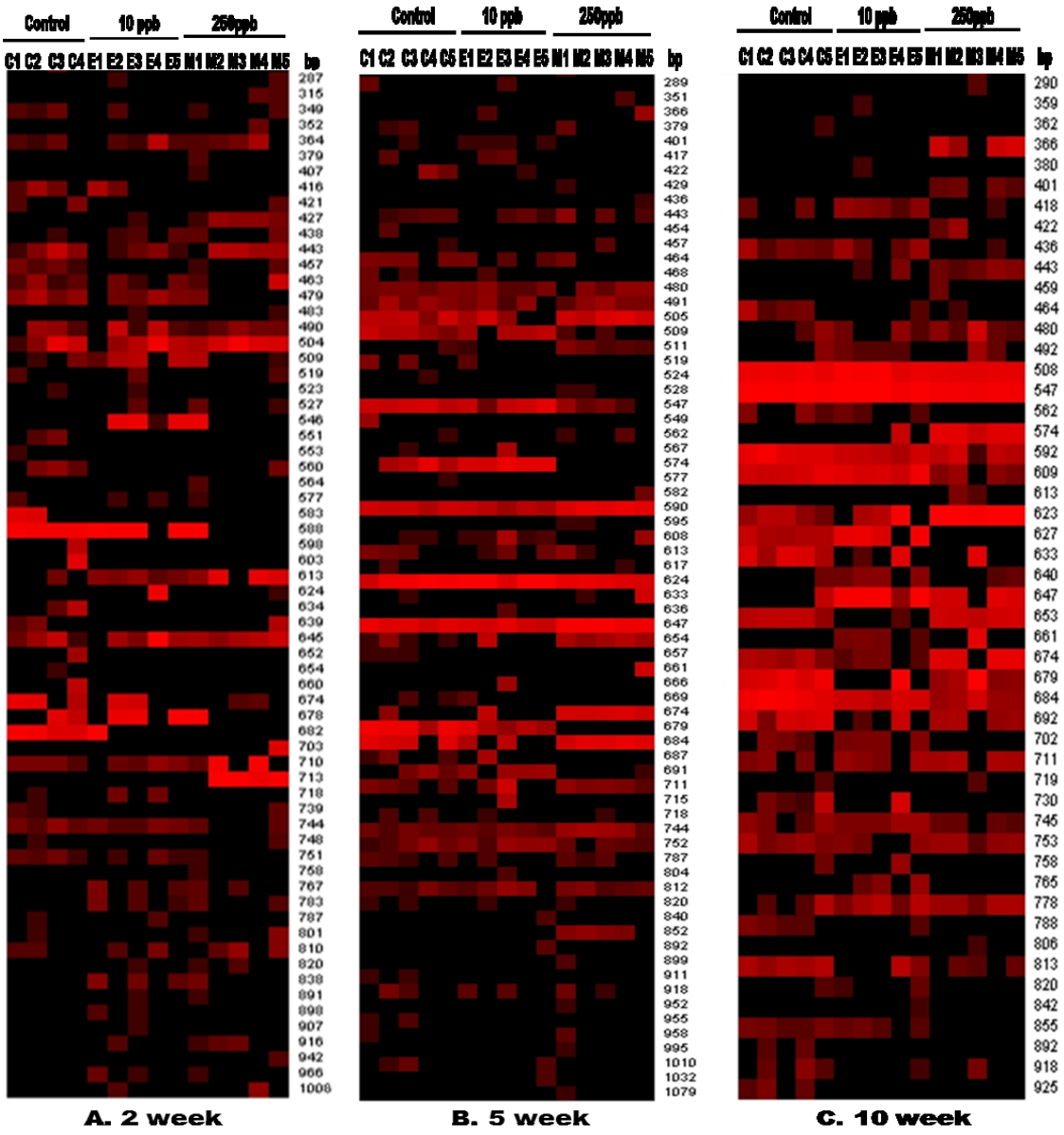


Figure 3.3: Heat map showing ARISA peaks obtained from the colon of control (C1-C5), 10 ppb (E1-E5) and 250 ppb (M1-M5) arsenite exposed mice at (A) 2 weeks (B) 5 weeks and (C) 10 weeks. Numbers towards the right of the image are equivalent to the length of 16S- 23S ribosomal intergenic region amplified by primers 1406F and 23Sr. Black indicates absence of peak and red indicates presence of peak. The color intensity indicates differences in the relative fluorescence intensity of the peak.

ARISA diversity

Not considering dose as a factor, the total number of peaks/OTUs detected in 2, 5 and 10 week colon samples were 66, 70 and 50 respectively. Approximately 22 peaks were common between 2, 5 and 10 week samples. The number of OTUs shared between different time points varied, with maximum number of OTUs (34 OTUs) shared by 10 and 5 week samples and least OTUs were shared between 10 and 2 week ARISA profile (25 OTUs). Two and 5 week samples shared 32 OTUs (Figure 3.4). For all time points, no significant differences were observed between control and As(III) exposed groups with respect to the total number of peaks generated by ARISA. The number of OTUs ranged from 43, 50, 36 in control samples, 41, 45, 39 in 10 ppb samples and 50, 44, 35 in 250 ppb As(III) exposed group in 2, 5 and 10 week samples respectively. These results indicate that As(III) treatment did not have any detectable effect on peak diversity. In addition, the results from ARISA profiles also indicate that arsenic exposure affects certain bacterial communities in the colon. Disappearance of some peaks and appearance of others in the As(III) exposed group along with variation in the intensity of peaks indicates that arsenic exposure disturbs the microbial balance in the mouse colon (Figure 3.3).

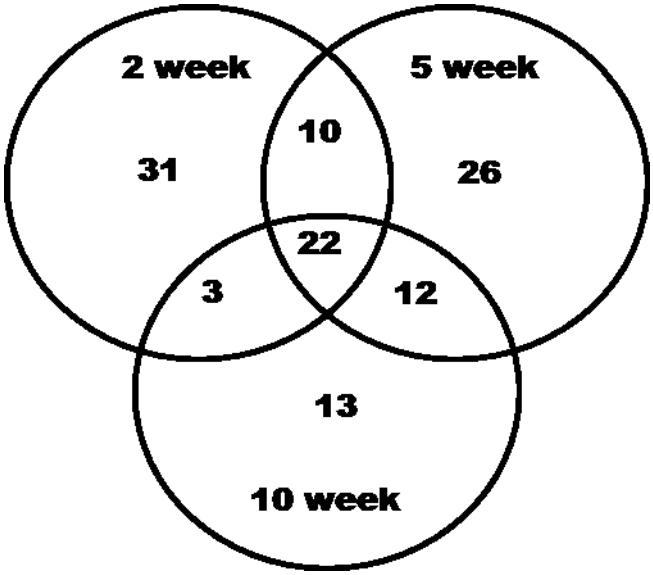
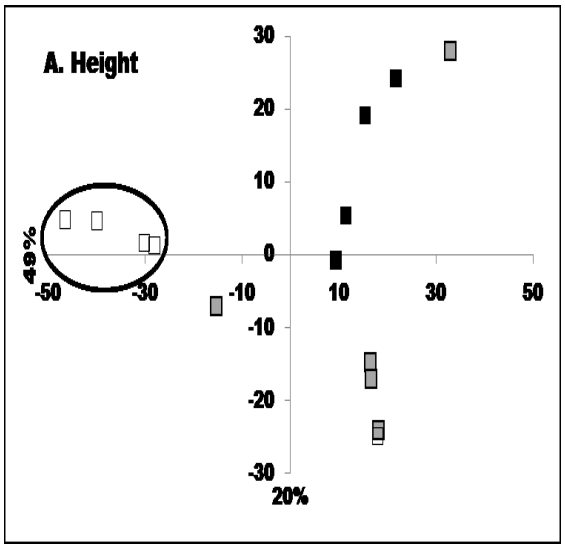


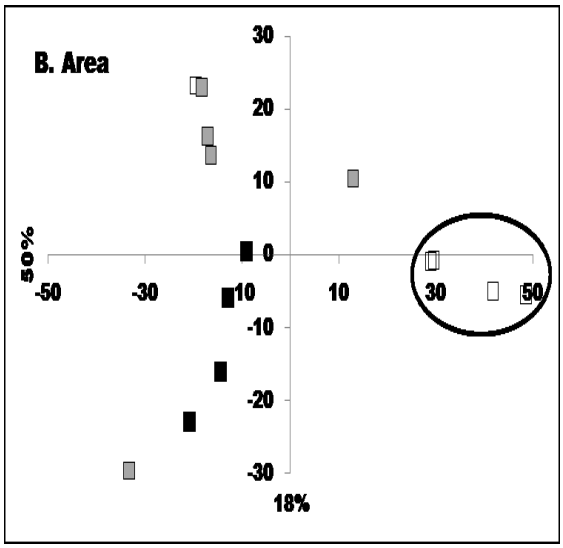
Figure 3.4: Venn diagram showing number of peaks/OTUs shared between ARISA profiles of 2, 5 and 10 week mouse colon samples irrespective of the As(III) exposure dose.

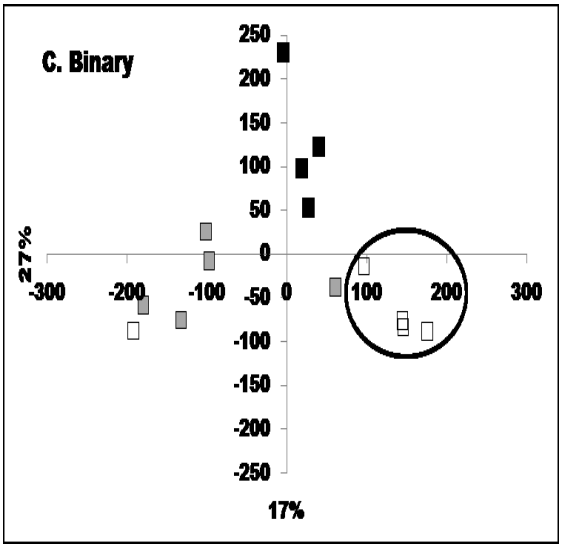
Principle Component Analysis (PCA)

The tables containing the information about the peaks present in all samples and their fluorescent intensity in terms of height and area were used for matrix generation. In this matrix the height and area for each peak was relativized by dividing the fluorescence of an individual peak by total fluorescent intensity of the sample to which that particular peak belongs. Also each data set was averaged across its replicates. The matrices obtained for presence/absence of peaks, relative peak height and relative peak area from each time point were used separately for PCA. PCA ordination plots with all three variables presence/absence of peaks (binary), relative peak height and area showed clear separation of control and 250 ppb As(III) treated samples at all three time points. However for the 10 ppb dosage level, separation of control and 10 ppb samples was clear in case of 10 week samples but not for 2 and 5 week samples (Figure 3.5). The peaks that contributed to the Principle Component 1(PC1) and PC2 above the mean % contribution of 1.51, 1.43 and 2.0 for 2, 5 and 10 weeks respectively are shown in Table 3.1. The peaks contributing significantly to PC1 and PC2 were almost identical when peak height and peak area matrices were used for PCA. However the peaks contributing to PC1 and PC2 differed in number as well as size from the above in the case of the binary matrix. Irrespective of the type of matrix used for PCA analysis, the overall relatedness of all the samples was similar in all the ordination plots (Figure 3.5).

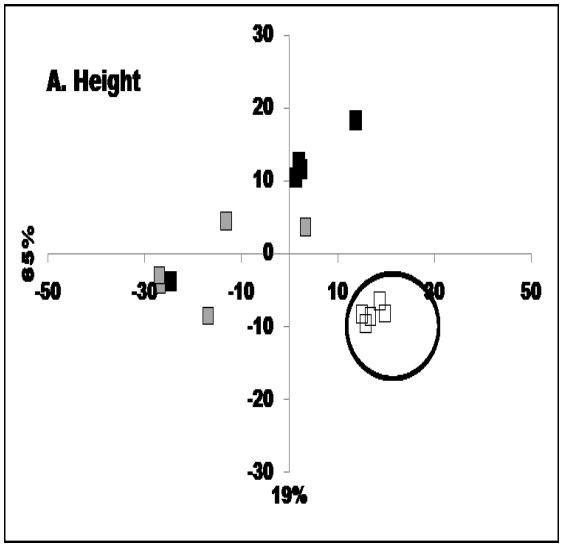
I. 2 week

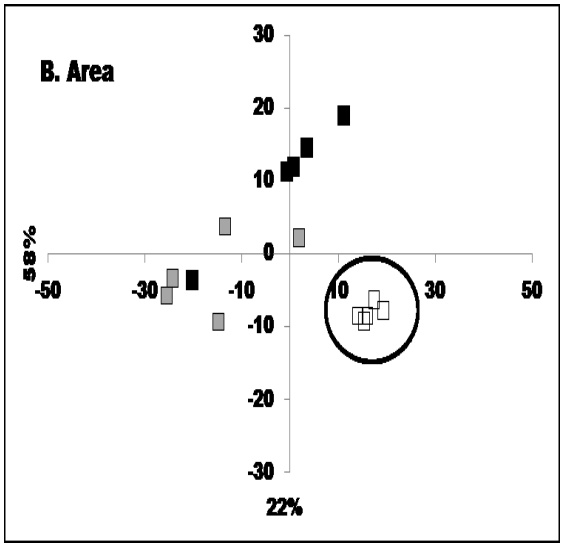

A. Height
30
20
10
0
-10
-20
-30
49%
-50
-30
-10
10
30
50
20%

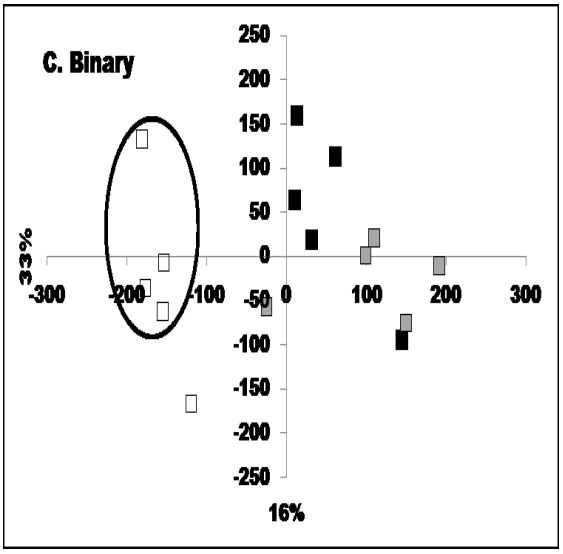

B. Area
30
20
10
0
-10
-20
-30
50%
-50
-30
10
10
30
50
18%


C. Binary
250
200
150
100
50
0
-50
-100
-150
-200
-250
27%
-300
-200
-100
0
100
200
300
17%

II. 5 week


A. Height
30
20
10
0
-10
-20
-30
-50
-30
-10
10
30
50
65%
19%


B. Area
30
20
10
0
-10
-20
-30
-50
-30
-10
10
30
50
58%
22%


C. Binary
250
200
150
100
50
0
-50
-100
-150
-200
-250
-300
-200
-100
0
100
200
300
33%
16%

III. 10 week

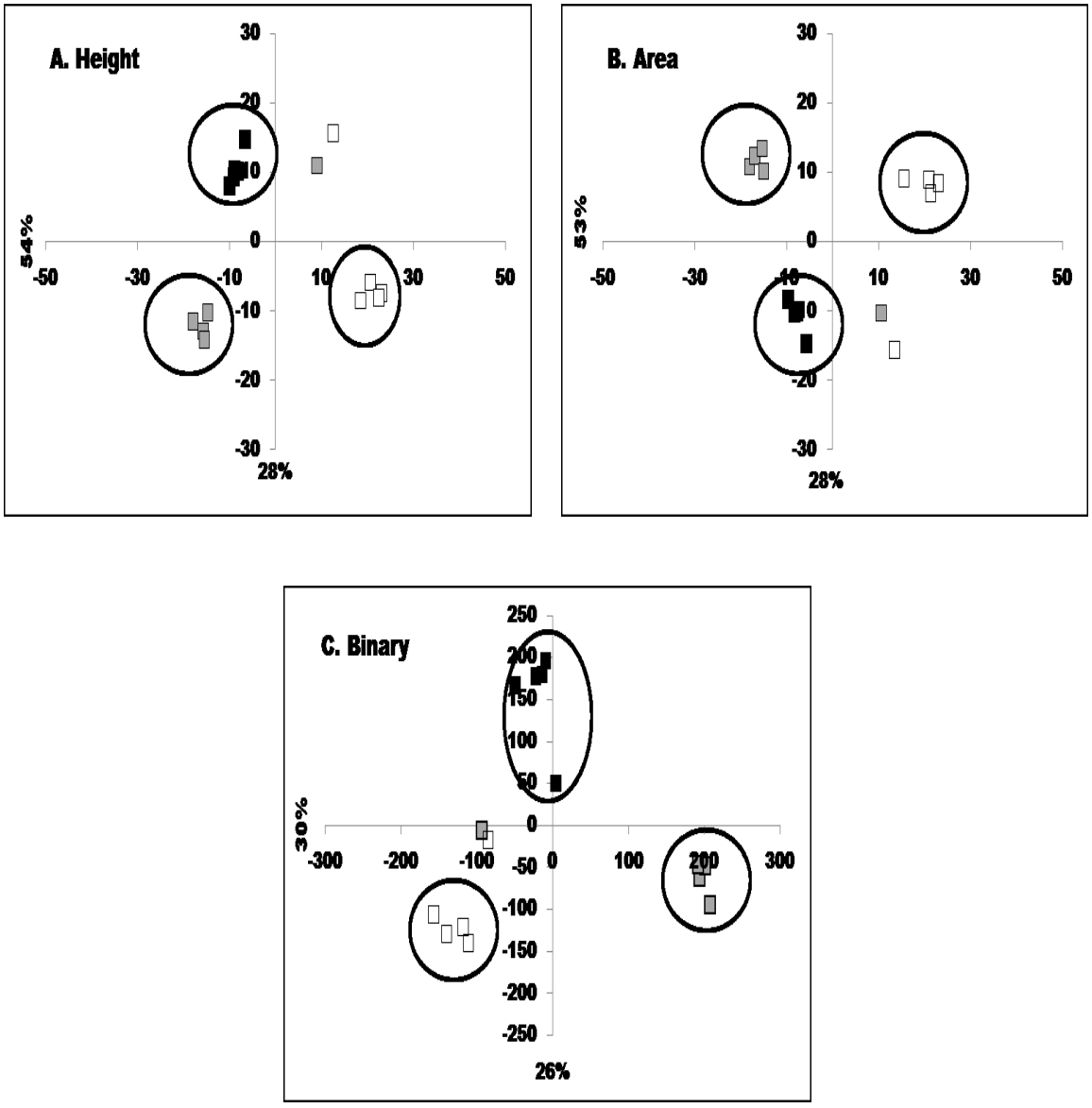


Figure 3.5: PCA plots generated from ARISA profiles of bacterial communities from (I) 2, (II) 5 and (III) 10week samples. Control (black squares), 10ppb (grey squares) and 250ppb (open squares). Percents shown along the axis represent the proportion of variation by that axis. PCA plots based on (A) relative peak height, (B) relative peak area and (C) presence/absence of peak (binary). Circles represent the groups that cluster together.

		Peaks (bps) contributing to PC1 and PC2
2 week	Height	504, 546, 583, 588, 613, 678, 682, 710, 713
	Binary	379, 421, 427, 438, 445, 457, 463, 479, 483, 509, 546, 562, 588, 590, 603, 613, 634, 645, 652, 660, 674, 678, 682, 713, 744, 767, 783, 810, 838
	Area	504, 546, 583, 588, 645, 678, 682, 710, 713
5 week	Height	505, 509, 547, 574, 590, 624, 647, 654, 674, 679, 684
	Binary	366, 379, 505, 509, 511, 519, 547, 574, 582, 608, 613, 633, 654, 661, 674, 679, 684, 691, 711, 787, 812, 852, 918, 958
	Area	505, 509, 547, 574, 590, 624, 647, 654, 674, 679, 684
10 week	Height	508, 547, 574, 609, 623, 627, 633, 647, 674, 679, 684, 692
	Binary	366, 401, 418, 436, 443, 464, 562, 574, 623, 627, 633, 640, 647, 653, 661, 674, 692, 702, 730, 765, 778, 813, 855, 925
	Area	508, 547, 574, 609, 623, 627, 633, 647, 674, 679, 684, 692

Table 3.1: Peaks contributing above the mean % contribution values for PC1 and PC2 of 2, 5 and 10 week PCA plots based on relative peak height, presence or absence of peaks (binary) and relative peak area.

3.3 Microbial community as determined by DGGE

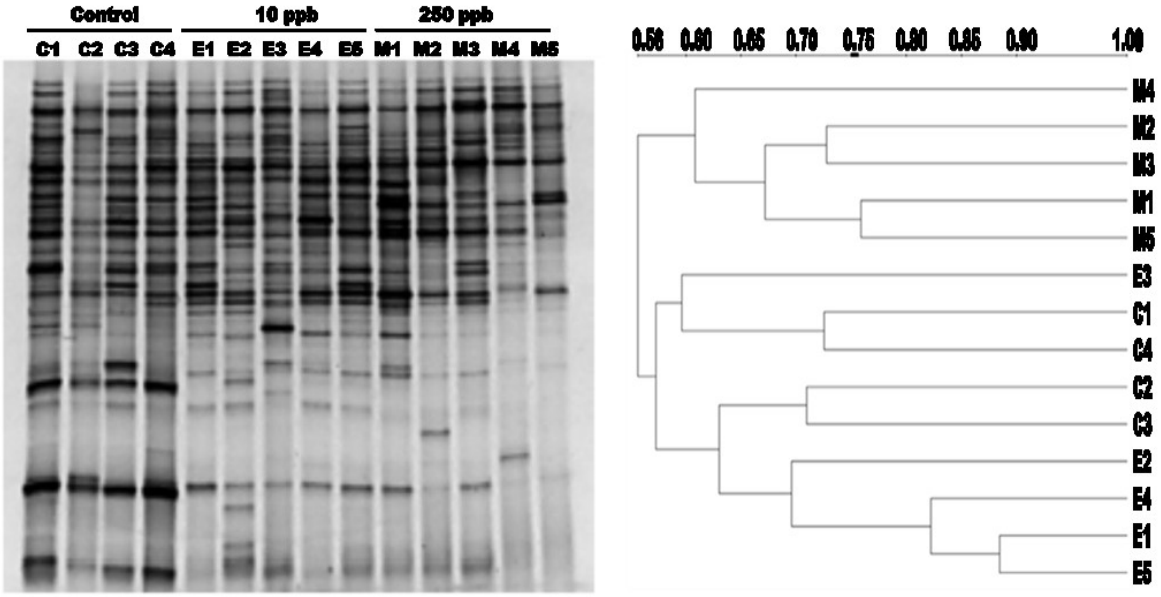
A small segment from the V3 region of the 16S rDNA was amplified from 2, 5 and 10 week mice colon bacterial community DNA. The expected product size after PCR amplification was ~ 177 bp. PCR products thus obtained were separated on the basis of their GC content on a 40 − 60% DGGE gel where the low GC fragments moved towards the bottom of the gel. Alterations in DGGE profiles in the form of band intensity and presence and absence of bands were observed when As(III) treated mice were compared to control mice. These variations were observed between the individuals from the same group as well. The differences in community composition were more pronounced for low GC bacteria, with a visible shift in low % Mole GC fragments at all three time points. However, careful observation of the band patterns indicated differences among high % Mole GC species too at the different dose exposure levels. These changes were more consistent and dramatic between the three groups with mice exposed to 250 ppb showing more variation as opposed to 10 ppb exposed mice when compared with control group (Figure 3.6). Similar to RISA, the total number of bands did not vary significantly between control and arsenic exposed groups. This suggested that arsenic exposure did not have any visible effect on microbial diversity of the murine colon. Although separated on different gels, a shift in banding pattern was observed in between the control groups at different time points indicating that microbial community of mouse is changing with time.

Similar to RISA gels, dendrogram analysis of DGGE gels indicated that even at 2 weeks, the 250 ppb exposed mice clustered into separate clades (Figure 3.6). Although visible differences in band patterns were found between control and 10 ppb As(III)

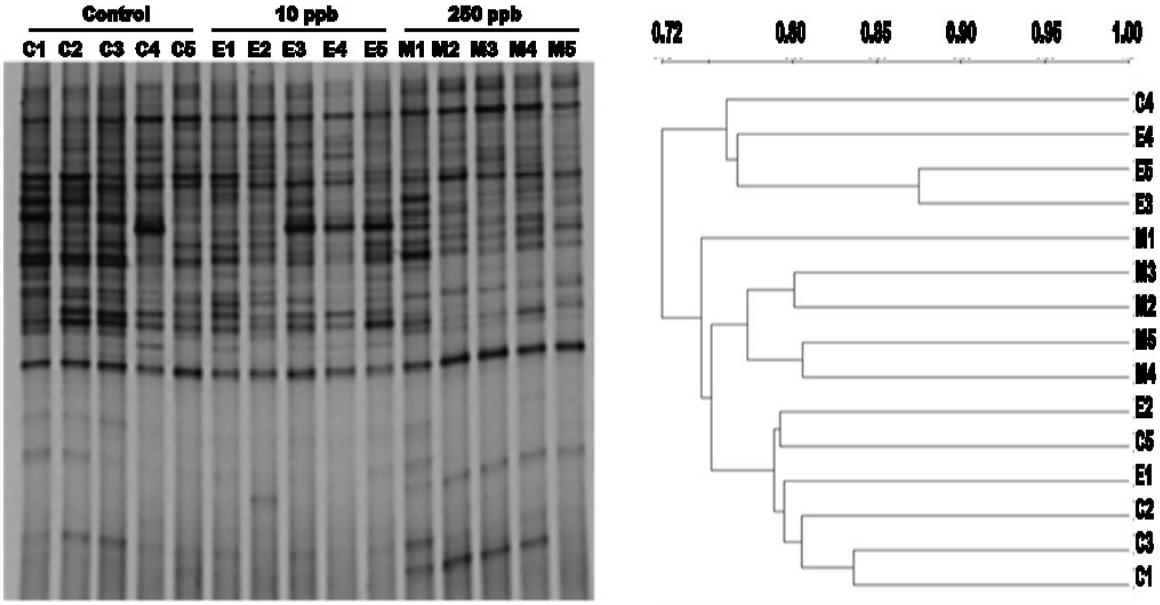
exposed groups at 2, 5 and 10 weeks but dendrogram analysis showed clustering of 10 ppb group with control groups. These results indicate that variations observed in 10 ppb group are so subtle that it is difficult to link these changes to As(III) exposure alone. Dendrogram analysis of DGGE microbial profiles for 2 week control and As(III) treated groups resulted in three main groups. Mice exposed to 250 ppb of As(III) clustered together whereas control and 10 ppb groups showed mixed clustering and formed other two groups. Mice exposed to 250 ppb of As(III) showed 55% similarity in band pattern to the other two groups whereas in between the 250 ppb As(III) group the similarity values ranged from 76 – 60%. Similarly, dendrogram analysis of 5 and 10 week DGGE gels resulted in three different clusters. Mice exposed to 250 ppb of As(III) formed separate cluster whereas control and 10 ppb groups formed mixed clustering.

Results from this study suggest that the observed differences between the three dose groups (0, 10 and 250 ppb) are more pronounced than the differences observed between the individuals of the same group. DGGE and dendrogram analysis clearly demonstrated that As(III) exposure affected the microbial community of mouse colon. The As(III) induced effect at 250 ppb exposure levels began at 2 weeks and continued until 10 weeks. Despite visible changes in the band pattern of As(III) 10 ppb group from the control group, the As(III) 10 ppb group did not form separate cluster from the control group at all three time points.

A. 2 week



B. 5 week



C. 10 week

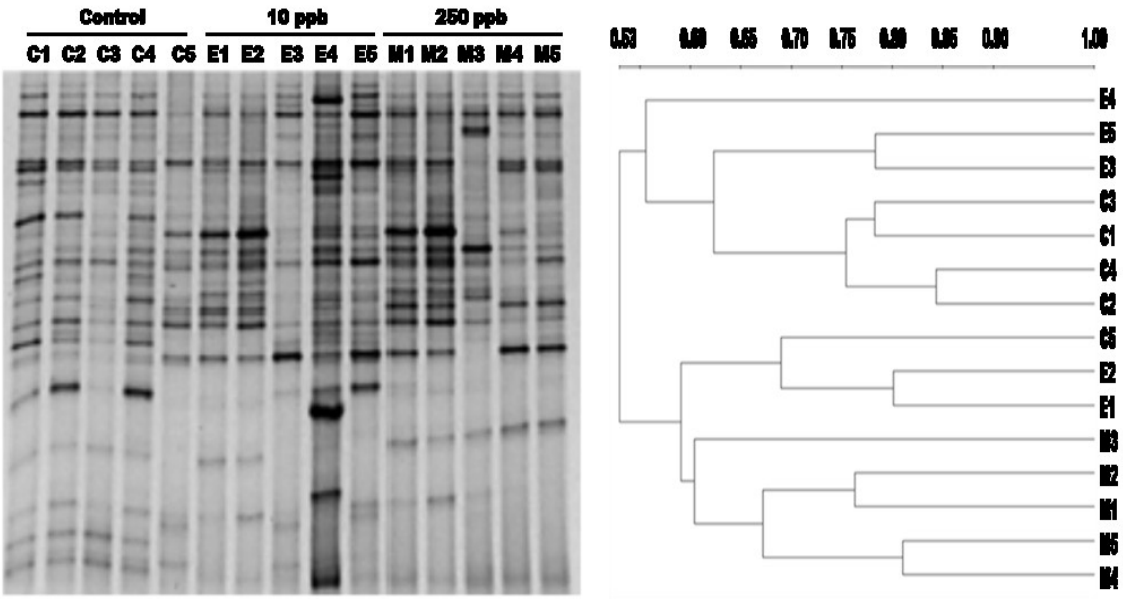


Figure 3.6: DGGE and dendrogram analysis of negative image of DGGE fingerprints generated from 16S rDNA PCR of control (C1-C5), 10 ppb (E1-E5) and 250 ppb (M1-M5) arsenic exposed mice at A) 2 week, B) 5 week, and C) 10 week. Scale bar on dendrograms represents similarity index values.

3.4 Firmicutes and Bacteroidetes are the major phyla in mouse colon

For studying the microbial community structure of mice colon, bacterial clone libraries generated from nearly full length 16S rRNA gene for each dose and time combination were sequenced using the Sanger method. A total of 947 good quality sequences were obtained, of which 123 sequences were found to be chimeras and removed from the analysis. Table 3.2 shows the number of clones sequenced and the number of good quality and non-chimeric sequences obtained from each library.

Based on RDP II database results, the mice colon microbial community contained sequences from 5 bacterial phyla. Regardless of time and dose of As(III) exposure, majority of the sequences belonged to the Bacteroidetes (37% of the total sequenced population) and Firmicutes (60% of all sequences). Sequences from Deferribacteres, Verrucomicrobia and Proteobacteria were observed in very low number, which was not surprising given the fact that these phyla are only a minor component of the anaerobic gut microbial community. Furthermore, only a few sequences (< 1%) were identified as uncultured bacterial clone but showed identity with 16S sequences obtained from other mammalian GI tract microbes. At the genus level, 12 different genera including *Barnesiella*, *Bacteroides*, *Ruminococcus*, *Oscillibacter*, *Anaerotruncus*, *Butyricicoccus*, *Roseburia*, *Dorea*, *Robinsoniella, Lactobacillus*, *Akkermansia* and *Mucispirillum* were identified in the mice colon representing only 34.6% of the total sequenced population. The sequences that were not classified at the genus level were classified to the next higher level in the order of family, order, class, phylum or unclassified bacteria.

	Total no. of Clones sequenced	Bad quality sequence*	No. of putative chimeric sequences	No. of sequences used for analysis
2week Control	96	9	20	67
2week 10ppb	120	24	15	81
2week 250ppb	120	22	18	80
5week Control	120	4	6	110
5week 10ppb	120	3	14	103
5week 250ppb	120	12	18	90
10week Control	120	11	8	101
10week 10ppb	120	15	9	96
10week 250ppb	120	9	15	96
Total	1056	109	123	824

Table 3.2: Sequencing information for the bacterial clone libraries obtained from colon samples. * indicates sequences with ambiguous bases (N).

Bacteroidaceae and Porphyromonadaceae were the only families observed in phylum Bacteroidetes along with a few sequences that were identified as unclassified Bacteroidales. Similarly, Lachnospiraceae was the most abundant family in phylum Firmicutes followed by Ruminococcaceae, Lactobacillaceae, Peptostreptococcaceae and Peptococcaceae. However unclassified Clostridiales (23.5%) formed a major part of the Firmicutes population. Deferribacteres and Verrucomicrobia were exclusively formed of a single genus namely *Mucispirillum* and *Akkermansia*, respectively.

When arsenite exposure was not considered as a factor, Firmicutes population decreased with time from 2 to 10 week clone libraries. Within Firmicutes, family Ruminococcaceae increased with time whereas Lachnospiraceae populations decreased from 2 to 10 week clone libraries. Bacteroidetes population showed an increase with time and a similar trend was observed for phylum Deferribacteres and Verrucomicrobia. Within the Bacteroidetes, populations belonging to the family Porphyromonadaceae increased with time from 2 to 10 week libraries whereas an opposite trend was observed for the family Bacteroideaceae. Table 3.3 shows the classification of the sequences obtained from these clone libraries to various levels of taxonomic hierarchy.

Taxonomic group	2 week	5 week	10 week	Over all
Bacteroidetes (P)	18.4%	41.6%	46.7%	37%
Bacteroidia (C)	18.4%	41.6%	46.7%	37%
Bacteroidales (O)	18.4%	41.6%	46.7%	37%
Unclassified Bacteroidales (O)	N	0.3%	2.7%	1%
Bacteroidaceae (F)	4.8%	4.9%	2.7%	4%
Bacteroides (G)	4.8%	4.9%	2.7%	4%
Porphyromonadaceae (F)	13.6%	36.3%	41%	32%
Unclassified Porphyromonadaceae (F)	3.9%	7.26%	12.3%	8.1%
Barnesiella (G)	9.6%	29%	28.6%	23.5%
Firmicutes (P)	78.9%	56.4%	49.8%	60%
Unclassified Firmicutes (P)	N	1%	0.3%	0.9%
Clostridia (C)	75.9%	55.1%	46.7%	57.9%
Clostridiales (O)	75.9%	55.1%	46.7%	57.9%
Unclassified Clostridiales (O)	34.2%	24.1%	14.7%	23.5%
Ruminococcaceae (F)	8.34%	11.9%	14.3%	11.8%
Unclassified Ruminococcaceae (F)	5.26%	8.58%	11.6%	8.73%
Butyricicoccus (G)	N	1%	N	0.4%
Oscillibacter (G)	1.7%	1.65%	2.7%	2%
Ruminococcus (G)	0.4%	0.3%	N	0.2%
Anaerotruncus (G)	0.9%	0.3%	N	0.4%
Lachnospiraceae (F)	33.3%	18.8%	17.4%	22.3%
Unclassified Lachnospiraceae (F)	32%	18.8%	16.4%	21.6%
Roseburia (G)	0.4%	N	1%	0.5%
Dorea (G)	0.4%	N	N	0.1%
Robinsoniella (G)	0.4%	N	N	0.1%
Unclassified Peptostreptococcaceae (F)	N	N	0.3%	0.1%
Unclassified Peptococcaceae (F)	N	0.3%	N	0.1%

Taxonomic group	2 week	5 week	10 week	Over all
Bacilli (C)	3%	0.3%	2.7%	1.9%
Lactobacillales (O)	3%	0.3%	2.7%	1.9%
Lactobacillaceae (F)	2.6%	0.3%	2.7%	1.8%
Lactobacillus (G)	2.6%	0.3%	2.4%	1.7%
Unclassified Lactobacillaceae (F)	N	N	0.3%	0.1%
Unclassified Lactobacillales (O)	0.4%	N	N	0.1%
Deferribacteres (P)	N	0.3%	1.3%	0.6%
Deferribacteres (C)	N	0.3%	1.3%	0.6%
Deferribacterales (O)	N	0.3%	1.3%	0.6%
Deferribacteraceae (F)	N	0.3%	1.3%	0.6%
Mucispirillum (G)	N	0.3%	1.3%	0.6%
Verrucomicrobia (P)	0.4%	0.6%	1.7%	0.9%
Verrucomicrobiae (C)	0.4%	0.6%	1.7%	0.9%
Verrucomicrobiales (O)	0.4%	0.6%	1.7%	0.9%
Verrucomicrobiaceae (F)	0.4%	0.6%	1.7%	0.9%
Akkermansia (G)	0.4%	0.6%	1.7%	0.9%
Proteobacteria (P)	0.4%	N	N	0.1%
Deltaproteobacteria (C)	0.4%	N	N	0.1%
Desulfovibrionales (O)	0.4%	N	N	0.1%
Unclassified Desulfovibrioniaceae (F)	0.4%	N	N	0.1%
Unclassified bacteria	1.7%	1%	0.3%	0.9%

Table 3.3: Percent distribution of bacterial 16S clone libraries obtained from mouse colon to bacterial taxonomy. N = not found, P = phylum (highlighted), C = class, O = order, F = family and G = genus.

3.5 As exposure is associated with increased Bacteroidetes

Despite interindividual variations, arsenite exposures clearly affected the microbial community of mouse colon. Compared with the control group, for all time points 10 ppb and 250 ppb As(III) exposed groups had an increase in the relative abundance of Bacteroidetes and proportionally fewer Firmicutes at the phylum level of taxonomy. Even at the class level, more Bacteroidia and fewer Clostridia were observed in As(III) exposed groups (Figure 3.7). However based on RDP II library comparison results, this change was significant ($p<0.01$) only for the 250 ppb groups at 2, 5 and 10 week interval whereas for the 10 ppb As(III) exposed groups the change was significant only at 10 week interval. Within Bacteroidia, the number of sequences classified under the family Porphyromonadaceae increased in response to As(III) exposure whereas no such trend was observed for Bacteroidaceae (Figure 3.8). With respect to phylum Firmicutes, sequences classified under family Lachnospiraceae and Ruminococcaceae decreased after two and five weeks of 250 ppb of As(III) exposure. In the case of 10 weeks of 10 and 250 ppb of As(III) exposure, only the sequences identified as Ruminococcaceae decreased significantly after As(III) exposure whereas no trend was observed for Lachnospiraceae. Bacilli were generally abundant in the As(III) exposed groups when compared to the control groups. Since many sequences from Firmicutes could not be classified beyond the family level it is difficult to determine whether arsenic treatment is affecting a single or more than one genus from these families. Paired comparison of the control and As(III) treated libraries using libshuff analysis yielded a value of <0.0001 which is less than a p value of 0.01 after Bonferroni correction, indicating that the libraries are significantly different from each other (Table 3.4).

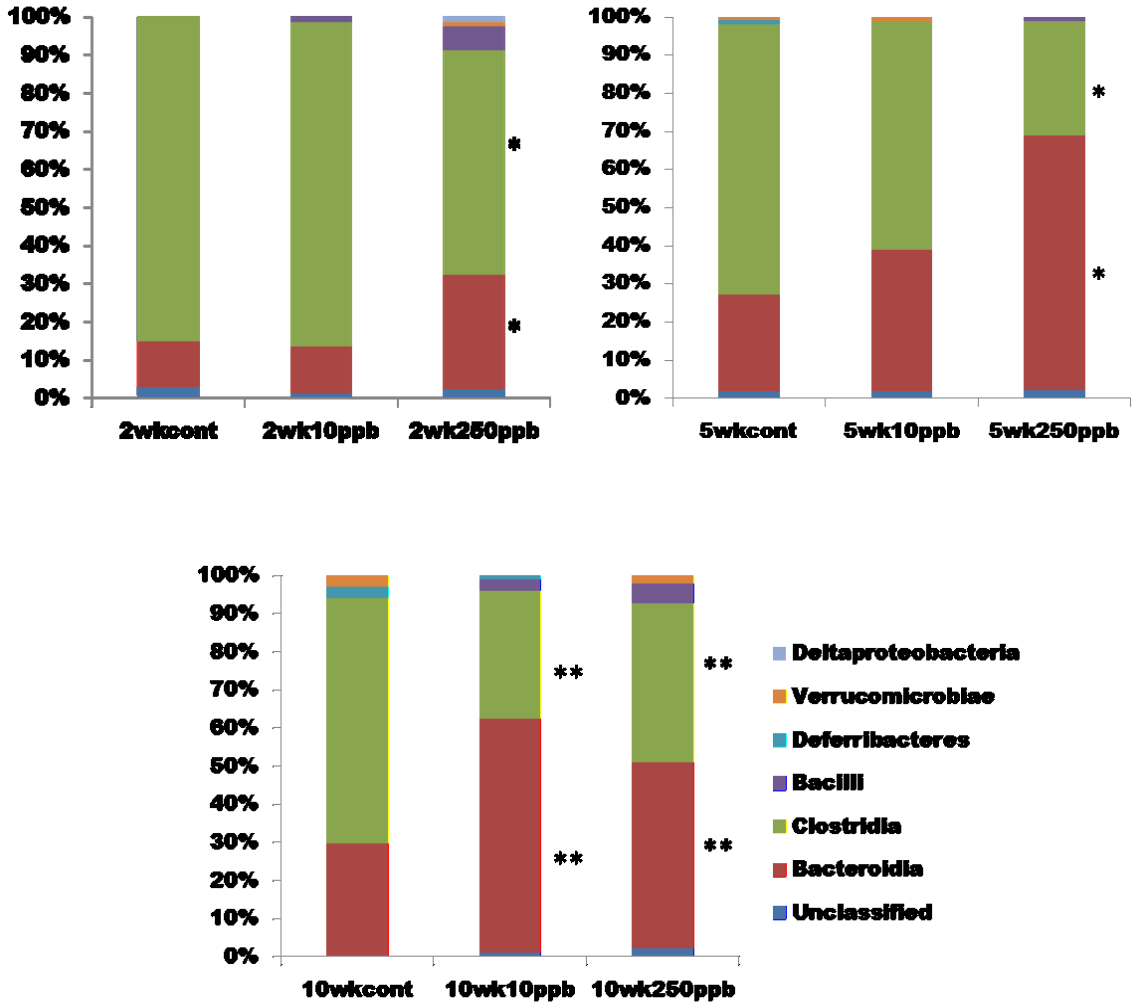


Figure 3.7: Class level taxonomic distribution of clones obtained from clone libraries. Stacked bar graphs show relative % distribution of sequenced clones to different taxonomic classes. * indicates significantly different from control and 10 ppb for that particular class ($p<0.01$) where ** represents significantly different from control group only ($p<0.01$)

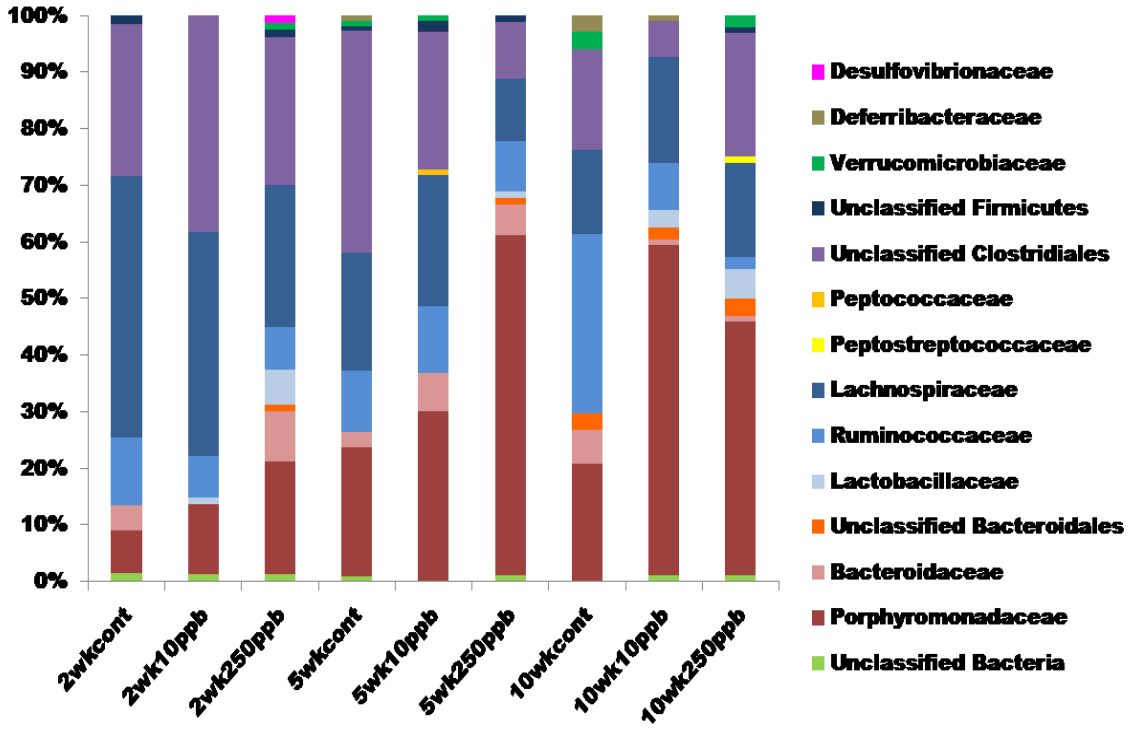


Figure 3.8: Family level taxonomic distribution of clones obtained from clone libraries. Stacked bar graphs show relative % distribution of sequenced clones to different taxonomic families.

	2 wk control	2 wk 10ppb	2 wk 250ppb	5 wk control	5 wk 10ppb	5 wk 250ppb	10 wk control	10 wk 10ppb
2 wk 10ppb	<0.0001							
2 wk 250ppb	<0.0001	<0.0001						
5 wk Control	<0.0001	<0.0001	<0.0001					
5 wk 10ppb	<0.0001	<0.0001	<0.0001	<0.0001				
5 wk 250ppb	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001			
10 wk Control	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001		
10 wk 10ppb	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	
10 wk 250ppb	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0002*	<0.0001	<0.0001

Table 3.4: Paired comparison of clone libraries by libshuff. The 2 libraries were considered significantly different if their p value generated by libshuff is lower than the corrected p value after Bonferroni correction. After Bonferroni correction the corrected p values of 0.00014 and 0.0007 are equivalent to the normal p values of 0.01 and 0.05. * indicates that the libraries are significantly different at p = 0.05 whereas all other libraries are significantly different at p = 0.01.

3.6 Microbial diversity estimation

In order to assess the diversity of bacteria present in the mice colons, 16S rRNA gene sequences obtained by sequencing were clustered into groups. Based on distance matrix, clones with 97% (species level), 95% (genus level) and 90% (family level) sequence identity were grouped into the same phylotype/operational taxonomic units (OTUs). The number of OTUs with only 1 sequence at a distance of 0.03, 0.05 and 0.1 were 236, 163 and 59 respectively. Regardless of As(III) exposure, a total of 313 unique phylotypes were identified including 159, 105 and 108 phylotypes from 2, 5 and 10 week clone libraries respectively at the distance level of 0.03. However at the genus level (95% sequence identity) only 136, 87 and 86 OTUs were found in 2, 5 and 10 week mice colon contributing to a total of 241 OTUs in mice colons. The number of OTUs determined in 2, 5 and 10 week clone libraries at family level were 74, 51 and 47 respectively, totaling towards 114 OTUs in mice colon. These results indicate that the diversity of mouse colon microbiota decreased after 2 weeks. Also, the number of OTUs shared between different libraries varied, with the maximum number of OTUs shared by 10 and 5 week clone libraries or 2 and 5 week clone libraries and least OTUs were shared between 10 and 2 week clone libraries. This pattern of shared OTUs was consistent for all distances. These results indicate that the microbial community from 5 week samples is intermediate in type and diversity to that of 2 and 10 week microbial populations. Figure 3.9 shows the number of OTUs shared between 2, 5 and 10 week clone libraries at a distance of 0.03, 0.05 and 0.1.

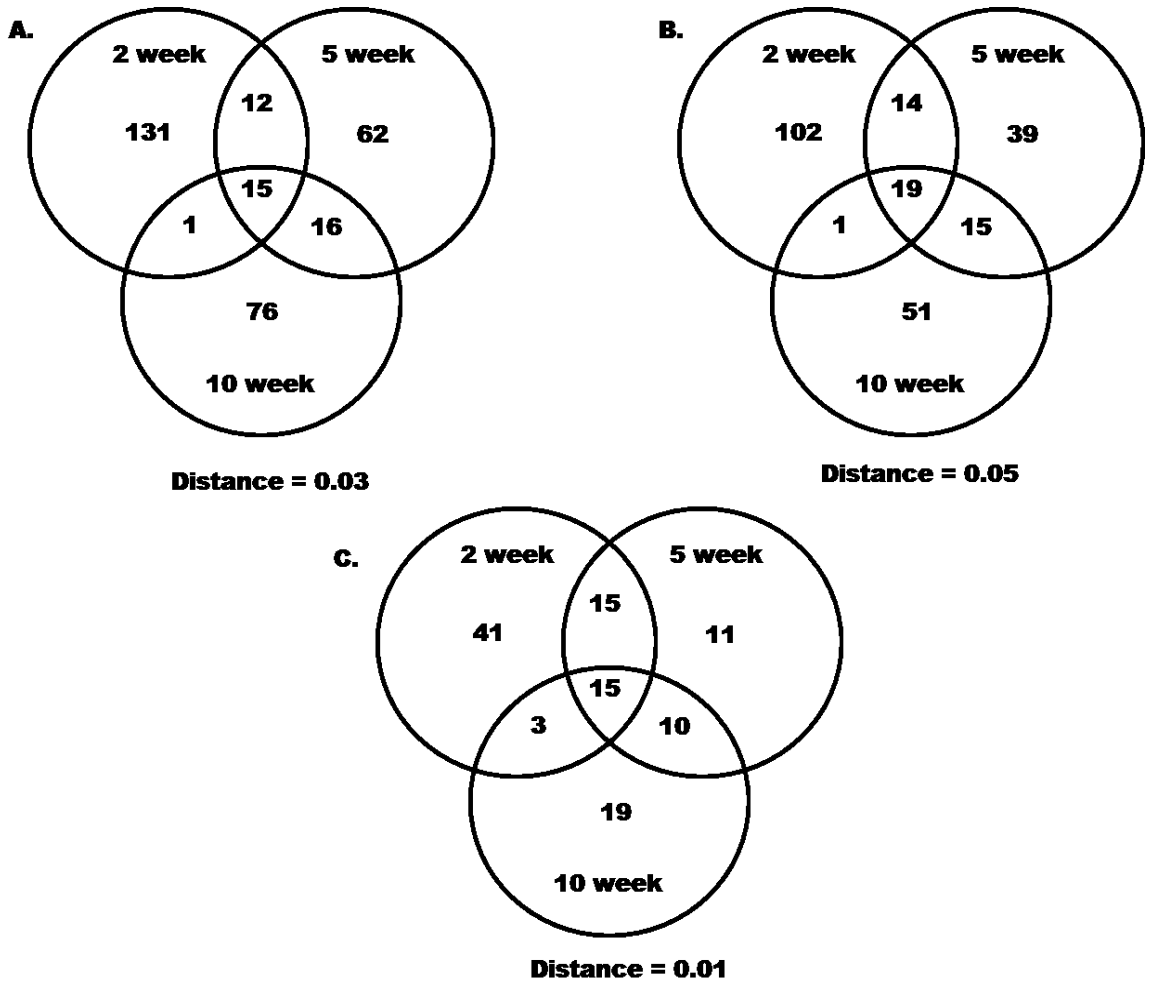


Figure 3.9: Venn diagrams showing the number of OTUs shared between 2, 5 and 10 week 16S rRNA clone libraries irrespective of As(III) exposure dose at a distance of (A) 0.03 (B) 0.05 and (C) 0.10.

Effect of arsenite on microbial diversity of the murine colon

Microbial diversity, richness and coverage were calculated for the combined data set as well as for all the sample subsets based on time (2, 5 and 10 week) and dose combinations (0, 10 and 250 ppb) (Table 3.5,Table 3.6). Arsenite exposure did not affect the overall richness of murine colon microbiota at the genus and family level as indicated by more or less the same number of observed OTUs in control and 250 ppb exposed groups. The number of observed OTUs was a little higher for 10 ppb at 0.05 and 0.1 distance levels in all the libraries except for the 5 week 10 ppb group where the number of observed OTUs were equal to the 5 week 250 ppb group.

Overall the Good's coverage was 93% for the combined libraries at a distance of 0.1 indicating that sequencing 100 additional clones will yield only seven new OTUs. When each library was considered individually, >83% coverage was observed for 5 and 10 week clone libraries at a distance of 0.10 but the coverage was lower (66 – 73%) for 2 week clone libraries. However the Good's coverage was generally low at the genus level for all the libraries except for 5 and 10 week control where 80 and 88% coverage was observed respectively.

To estimate the richness of the clone libraries, rarefaction curves were plotted (Figure 3.10). For the combined dataset at 0.03 and 0.05 distance levels the rarefaction curve did not reach saturation towards the end of the sampling. However at a distance of 0.1 the curve almost reached plateau towards the end of sampling for the combined dataset as well as for all individual types of clone libraries (Figure 3.10). Both rarefaction curves and Good's coverage estimates indicated that the majority of bacteria at the family level of taxonomy have been captured in the current sequencing efforts. However new

phylotypes would be detected with continued sampling from these colon samples at the genus and species level.

Similar to rarefaction curves, estimations of species richness by Chao's estimator predicted that sequencing more clones can reveal more unique sequences (Table 3.5, Table 3.6). Estimated numbers of OTUs as determined by Chao were relatively close to the observed numbers at higher distance level (0.1) whereas the difference between estimated and observed OTUs increased at a lower distance level. The percentage of observed/estimated OTUs was between 33−76% at cluster distance of 0.1 whereas this percentage was reduced to 14−69% at the genus level.

Bacterial diversity at the family and genus level as measured by Shannon and Simpson index was least for 250 ppb group and highest for control group when time was not considered as a factor (Table 3.5, Table 3.6). The highest diversity was observed for the 2 week libraries and least for the 5 week when As(III) exposure was not included as a factor. Within the individual library data sets, bacterial diversity was highest for 2 week control and lowest for 5 week 250 group. These results differed from the microbial richness wherein observed species richness as determined by rarefaction curve was highest in the 2 week 10 ppb group and lowest in 10 week control group. The Shannon index is prone to error and bias when the sampling is incomplete (Magurran, 1988). Therefore during any discrepancy between the Shannon and Simpson diversity indices more weight was given to the trends observed by the latter. Overall the results from microbial diversity estimations suggest that more sequencing efforts will generate more unique microbial species that in combination could have a profound effect on mouse health and disease.

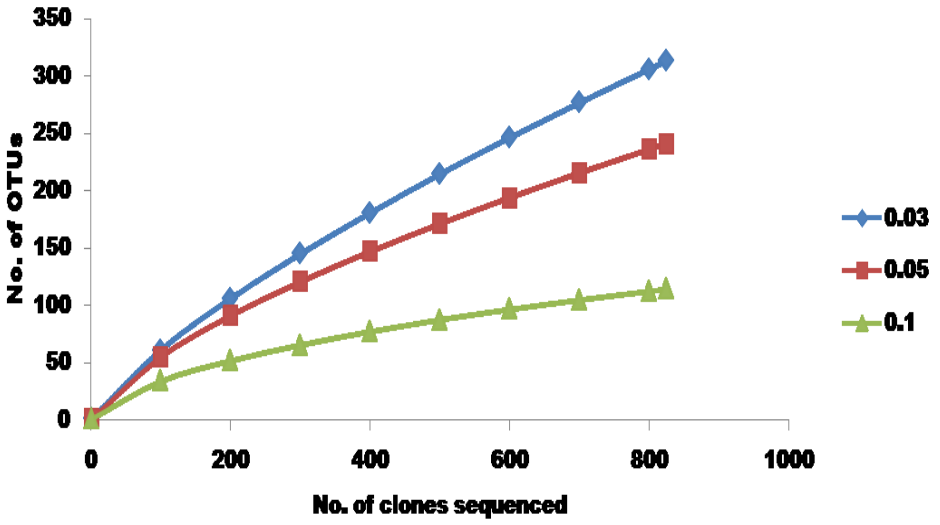
	Observed OTUs	Chao Estimator of species richness	Good's Coverage (C) (%)	Simpson Diversity Index (D)	Shannon Diversity Index (H)
2 wk Control	51	352	36	0.012212	3.797245
2 wk 10ppb	59	227	39.5	0.014815	3.889397
2 wk 250ppb	50	180	50	0.039873	3.539687
2wk All	136	533.5	53.5	0.013139	4.555065
5 wk Control	36	64.875	80	0.105755	2.806833
5 wk 10ppb	42	88.42857	75	0.059585	3.20864
5 wk 250ppb	37	112.6	69	0.209488	2.574984
5 wk All	87	185.076923	83	0.080104	3.270088
10 wk Control	29	42.2	88	0.077426	2.878059
10 wk 10ppb	42	108.4286	68	0.133991	2.90301
10 wk 250ppb	39	85.42857	73	0.093202	2.986581
10 wk All	86	164	82	0.053626	3.597955
All Control	96	243.5	78	0.046256	3.804828
All 10ppb	119	467	68.5	0.051152	3.974711
All 250ppb	97	275.75	75	0.06679	3.648178
All	241	696.275862	80	0.040893	4.328786

Table 3.5: Indices of richness, diversity and library coverage obtained for 16S clonal libraries at a distance of 0.05. D decreases with increasing diversity while H increases with increasing diversity.

	Observed OTUs	Chao Estimator of species richness	Good's Coverage (C) (%)	Simpson Diversity Index	Shannon Diversity Index
2 wk Control	32	95.25	66	0.069652	2.987348
2 wk 10ppb	34	80.2	73	0.074383	3.012234
2 wk 250ppb	31	94.25	71	0.089241	2.790469
2wk All	74	144.714286	80	0.066504	3.454768
5 wk Control	25	40.6	88	0.137114	2.423934
5 wk 10ppb	23	56	88	0.110603	2.508098
5 wk 250ppb	23	42.5	85.5	0.249438	2.145428
5 wk All	51	113.142857	90	0.116058	2.767992
10 wk Control	22	29	92	0.106337	2.546807
10 wk 10ppb	27	47	83	0.218421	2.286577
10 wk 250ppb	22	29.2	91	0.149123	2.371783
10 wk All	47	66.090909	93	0.09568	2.898701
All Control	55	86.90909	90	0.076669	3.091312
All 10ppb	66	156.3	85	0.102407	3.025548
All 250ppb	54	103.6	88	0.110427	2.8729
All	114	214.647059	93	0.079961	3.293593

Table 3.6: Indices of richness, diversity and library coverage obtained for 16S clonal libraries at a distance of 0.10. D decreases with increasing diversity while H increases with increasing diversity.

A.



B.

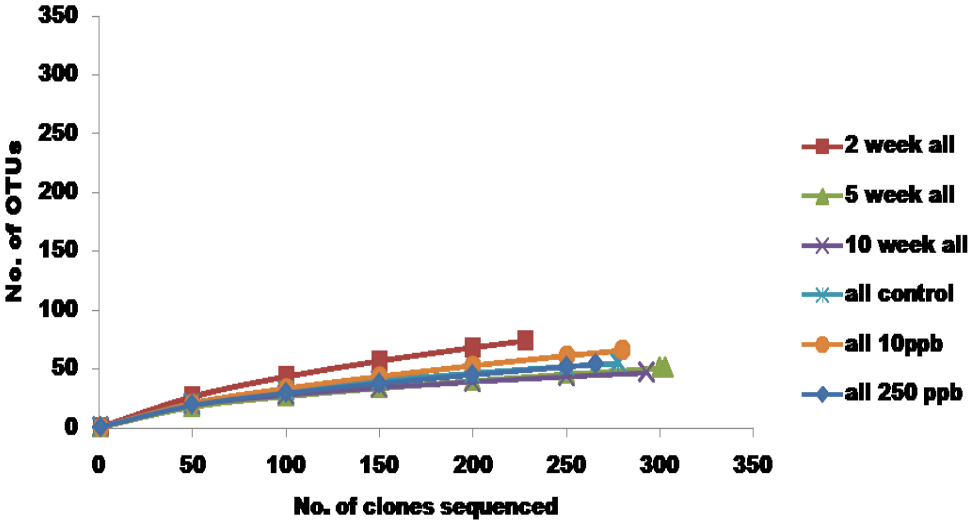


Figure 3.10: Rarefaction curves obtained from 16S rDNA clone libraries. (A) Rarefaction curves obtained from combined dataset at a distance of 0.03, 0.05 and 0.10. (B) Rarefaction curves obtained from 3 time points (2, 5 and 10 weeks) and 3 doses (0, 10 and 250 ppb) at a distance of 0.10.

3.7 Phylogenetic tree

To determine the phylogenetic relationship between the OTUs obtained from 16S rDNA clone libraries of the murine colon a neighbor joining phylogenetic tree of the representative OTU sequences (at 90% sequence similarity) from control, 10 ppb and 250 ppb group irrespective of the time point was constructed (Figure 3.11). A total of 174 clones were represented on the neighbor-joining tree. The phylogenetic tree indicated that for most of the OTUs identified from the control group, nearly identical OTUs were found in As(III) groups too. As expected, Clostridia formed the largest clade in the phylogenetic tree consisting of OTUs from all three libraries. However the number of sequences within certain phylotypes of Clostridia clade varied between the three groups. Similar to Clostridia, clades within Bacteroidia consisted of OTUs from all three libraries but the number of sequences within certain phylotypes were drastically different between the control and As(III) exposed libraries. For example one of the OTU in Bacterodia had 133 sequences from the As(III) groups but only 18 sequences from the control group. Bacilli formed a small clade in the phylogenetic tree and was exclusively formed by OTUs obtained from As(III) exposed groups.

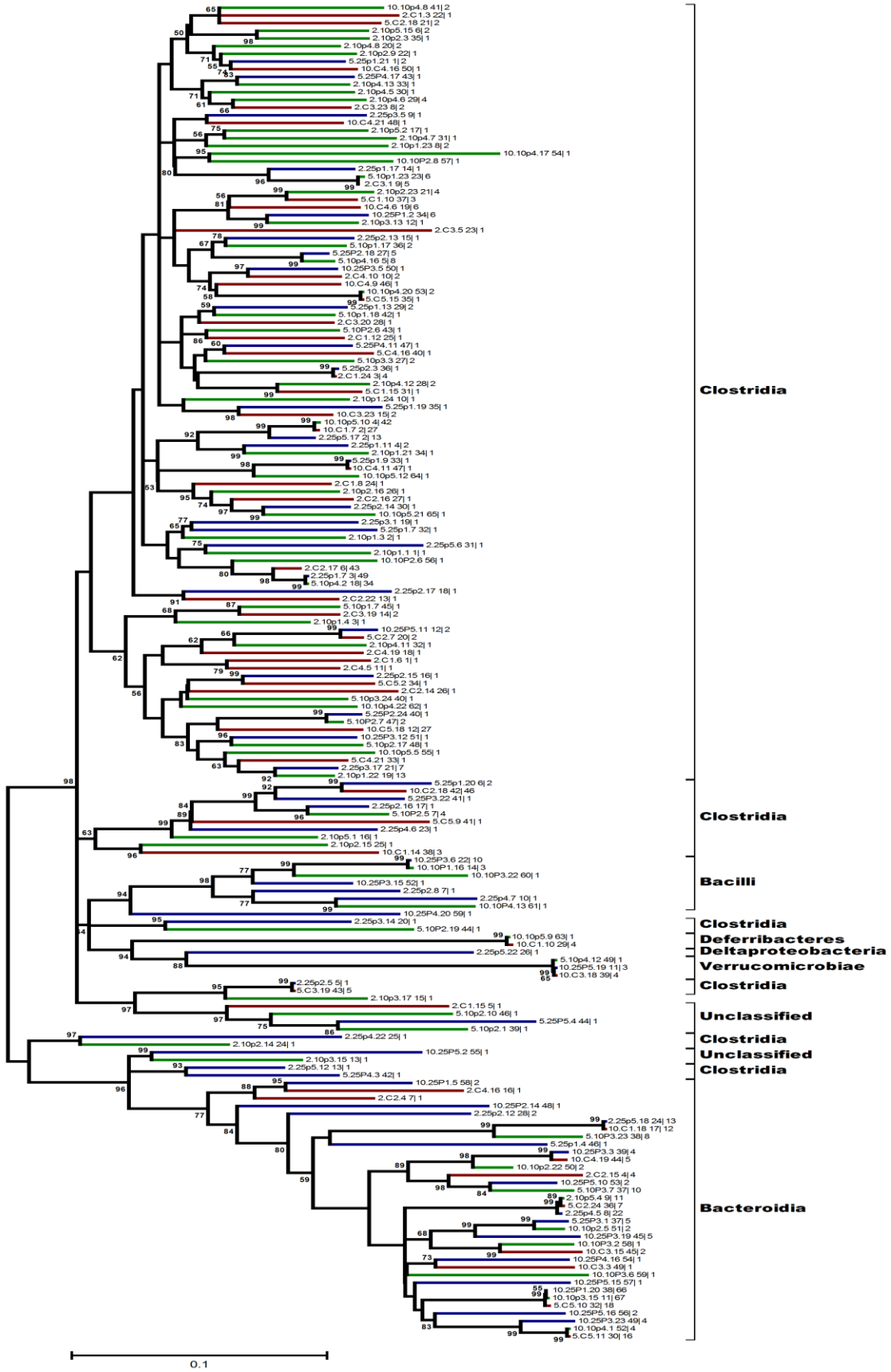

Clostridia
Clostridia
Bacilli
Clostridia
Deferribacteres
Deltaproteobacteria
Verrucomicrobiae
Clostridia
Unclassified
Clostridia
Unclassified
Clostridia
Bacteroidia
0.1

Figure 3.11: Neighbor joining tree showing the positions of 16S rDNA phylotypes obtained from mouse colon. Clones obtained from control mice are shown on red terminal branches, clones from 10 ppb exposed mice are shown on green branches and those obtained from 250 ppb As(III) exposed mice are shown on blue branches. Boot strap values obtained from 1000 replications and above 50% are shown at the nodes. Scale bar represents genetic distance based on 10 substitutions per 100 nucleotides. The number after " | " at the end of each of the clone name indicates the number of sequences that were found under the same representative OTU.

3.8 Distribution of *ars* operon

Distribution and occurrence of *ars* operons were determined bioinformatically in partially or completely sequenced microorganisms from the DOE Joint Genome Institute portal (http://img.jgi.doe.gov/). This was done to determine whether the mechanism for arsenic resistance and detoxification occurs in microbes that typically reside in the gut environment. The search was restricted to the microbial taxa that were identified from the sequencing data of the mouse colon from section 3.4. Due to the high ambiguity in gene annotation the search for *ars* operon in these genomes was done by three methods: COG (Cluster of Orthologous Groups) search, protein BLAST using ArsB sequence from *B. thetaiotamicron* and keyword search. The keywords used for the search for *ars* operon in different bacterial species include ars, arsenate reductase and arsenical pump.

In total, 182 different bacterial species and strains were found on JGI website that fall under the taxonomic classification found in the mouse colon. Of these, only 52 species had the whole genome sequenced. A total of 49 Bacteroidia, 39 Clostridia, 74 Bacilli, 14 Deltaproteobacteria, 4 Verrucomicrobia and 2 Deferribacters species were searched for the presence of *ars* operon in their genome. The *ars* operon was defined as sequences that occur as cluster of at least two *ars* genes and are separated by not more than 200 bps. Based on this definition, only 18% of the total observed bacteria were found to have *ars* operon. Table 3.7 lists the taxonomic classification used for search of *ars* operon and microbes that were found to have an *ars* operon along with the arrangement of *ars* genes within the operon.

Species (Genome status)	*ars* operon arrangement*
P= Bacteroidetes, C= Bacteroidia, F= Bacteroidiaceae	
Bacteroides dorei 5_1_36/D4(D)	*RBCxxxxxR*
B. thetaiotamicron VPI-5482(F)	*BCADxxxR, BADxxxR*
B. fragilis 3_1_12(D)	*RBCxxxxxR*
B. eggerthii DSM 20697(D)	*BCADx*
B. intestinalis 341 DSM 17393(D)	*BCADxxxR, BADxxxR*
*B. sp.*1_1_16 (D)	*BADxxxR*
B. sp. 2_1_33B (D)	*BCADxxxR*
B. sp. 4_3_47FAA (D)	*BCADxxxR*
B. stercoris ATCC 43183(D)	*BCADxxxRx*
B. vulgatus ATCC8482 (F)	*BCADxxxR*
B. vulgatus PC510 (D)	*BCADxxxR*
P= Firmicutes, C= Bacilli, F= Lactobacillaceae	
Lactobacillus antri DSM 16041 (D)	*CBR*
L. fermentum IFO 3956 (F)	*CBR*
L. johnsonii NCC 533 (F)	*CBR*
L. reuteri CF48-3A (D)	*BRCCBR, CBR*
L. reuteri SD2112 (D)	*BRCCBR, CBR*
L. plantarum WCFS1(F)	*xDxBADR*
P= Firmicutes, C= Clostridia, F= Lachnospiraceae	
Oribacterium sinus F0268 (D)	*BAR*
P= Firmicutes, C= Clostridia, F= Ruminococcaceae	
Ruminococcus sp. 5_1_39BFAA(D)	*CBxxxxxR*
Ethanoligenens harbinense YUAN-3T (D)	*BxxR*
P= Firmicutes, C= Clostridia, F= Peptococcaceae	
Desulfitobacterium hafniense Y51 (F)	*BRCxR, CABR, CBR*
D. hafniense DCB-2 (F)	*BRC*
Desulfotomaculum reducens MI-1 (F)	*BR*
Pelotomaculum thermopropionicum (F)	*BR*
Thermincola sp. JR (F)	*BR*
P= Deferribacteres, C= Deferribacteres, F= Deferribacteraceae	
Deferribacter desulfuricans SSM1(F)	*BCR*
Denitrovibrio acetiphilus N2460 (F)	*CB*
P= Verrucomicrobia, C= Verrucomicrobea, F= Verrucomicrobiaceae	
Verrucomicrobium spinosum DSM4136 (D)	*CBxRxxxRx*
Verrucomicrobiae bacterium DG1235 *(D)*	*RBCAD*
P= Proteobacteria, C= Deltaproteobacteria, F= Desulfovibrionaceae	
Desulfovibrio sp. 3_1_syn3 (D)	*BR*
D. aespoeensis Aspo-2 (D)	*BR*
D. desulfuricans G20 (F)	*CCB*
D. magneticus RS-1 (F)	*BR*
D. salexigens DSM 2638 (F)	*BR*
Reference sequences (used for phylogenetic tree)	

Species (Genome status)	*ars* operon arrangement*
P= Bacteroidetes, C= Bacteroidia, F= Bacteroidiaceae	
P. melaninogenica ATCC25845 (F)	*BCADxxxR*
P. bergensis DSM 17361 (D)	*BCAR*
P= Firmicutes, C= Clostridia, F= Clostridiaceae	
Clostridium kluyveri DSM 555 (F)	*CBR*
Cl. beijerinckii NCIMB 8052 (F)	*BCADR*
Cl. Phytofermentans ISDg (F)	*CDBAR*
P= Firmicutes, C= Bacilli, F= Staphylococcaceae	
Staphylococcus aureus RF122 (F)	*CBR*
P= Proteobacteria, C= Gammaproteobacteria, F= Enterobacteraceae	
Escherichia Coli FVEC 1412 (D)	*CBADR*
P= Proteobacteria, C= Gammaproteobacteria, F= Pseudomonadaceae	
Pseudomonas fluorescens PF-5 (F)	*HCBR*
P= Firmicutes, C= Bacilli, F= Bacillaceae	
Bacillus subtilis subtilis 168 (F)	*CBxR*

Table 3.7: Arrangement of *ars* operon in various bacterial species. For genome status F and D in parentheses indicates that the genome is completely sequenced (F) or in draft stage of sequencing (D). * The *ars* genes are written in the order they are found in the genome. "*x*" represents genes encoding proteins of unknown function.

Within the *ars* operon, *arsB* and *arsR* genes were found in almost all the bacteria whereas *arsA* occurred in 40% of the bacteria. *arsA* was mainly distributed in Bacteroidetes phylum mostly in association with *arsD*. The majority of the Firmicutes had *arsCBR* or *arsBR* gene composition. Some of the species were found to have two *ars* operons. The arrangement of genes constituting the *ars* operon varied from species to species. Based on current estimates, *ars* operon is present in far more Bacteroidetes species than Firmicutes indicating that the expression of *ars* operon could be the reason behind selection for Bacteroidetes populations in As(III) exposed samples. In some of the species, the *ars* operon had genes for transposition and/or integration in its vicinity leading to speculation that *ars* gene might be transposed to other species.

To determine the evolutionary relatedness of the *ars* operon in these microbes neighbor joining (NJ) phylogenetic trees were constructed separately for ArsA and ArsB protein sequences (Figure 3.12, Figure 3.13). In the NJ trees the ArsA and ArsB sequences formed three clusters. In both the trees Bacteroidetes were confined to a single cluster. All ArsA sequences from Firmicutes clustered together whereas some of the ArsB sequences formed clusters with Bacteroidia. Within Firmicutes, ArsB sequences from *Lactobacillus* species formed a separate sub cluster but were separated from other Bacilli. The ArsB sequences from Verrucomicrobiae, Delta proteobacteria and Deferribacteres grouped together. *Desulfitobacterium hafniense* consists of three ArsB sequences and each one of them was confined to separate sub cluster.

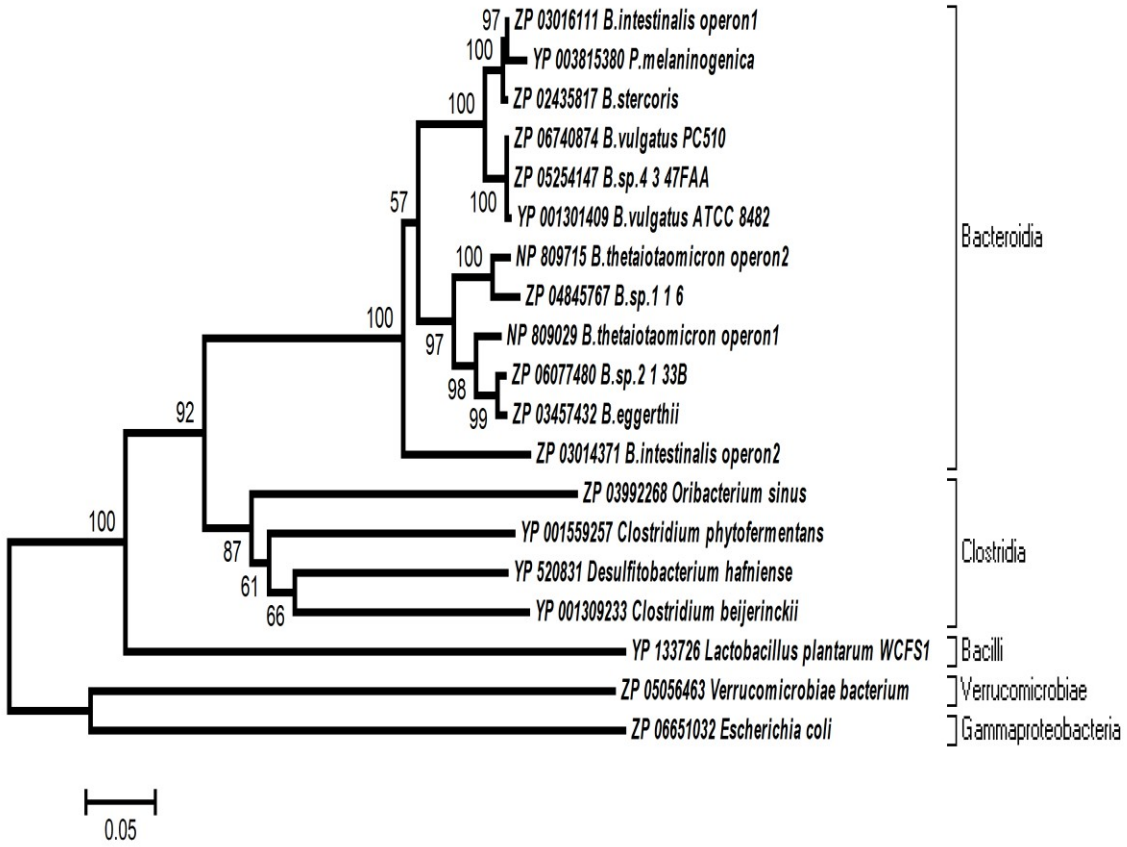


Figure 3.12: A Neighbor joining tree of ArsA protein sequences. Boot strap values obtained from 1000 replications are shown at the nodes. Scale bar represents evolutionary distance based on number of substitutions per site. The Genbank accession number for each sequence is given before the species name.

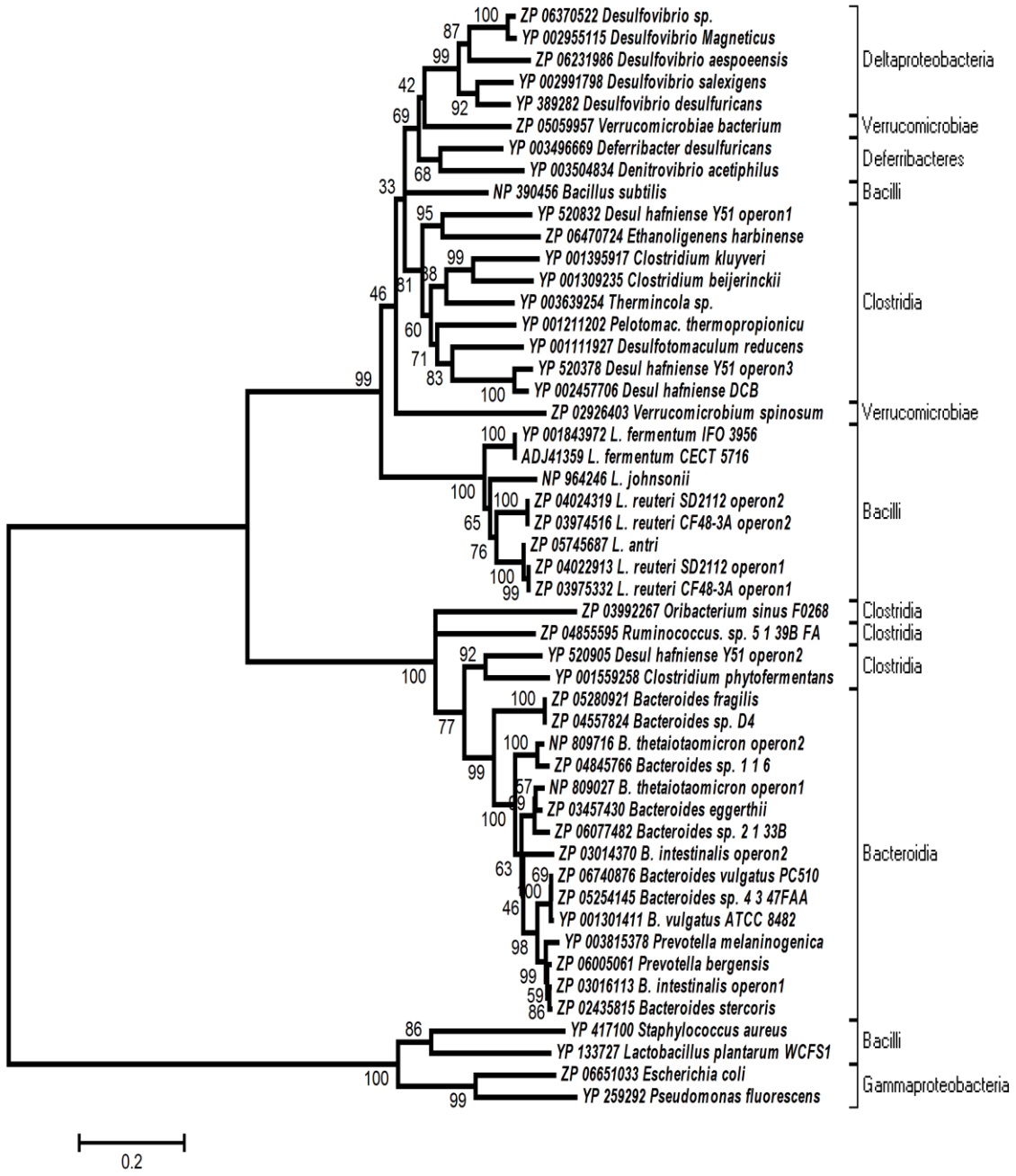


Figure 3.13: A Neighbor joining tree of ArsB protein sequences. Boot strap values obtained from 1000 replications are shown at the nodes. Scale bar represents evolutionary distance based on number of substitutions per site. The Genbank accession number for each sequence is given before the species name.

3.9 Effect of As(III) on microbial physiology of the host

Selection of model organism

In order to understand the in vivo effect of As(III) on gut microbiota physiology, *ars* and *nrf* operon genes were selected for comparing their expression at the mRNA level between control, 10 and 250 ppb As(III) exposed mice. Since enough variability exists between various species in the gene sequences from the two operons, a model organism was needed to accomplish the above aim. The selection of the model organism was based on the criteria that the selected organism should possess one or both of the operons (*ars* or *nrf* operon) being studied, should belong to the taxonomic group of significance in the gut and should be present in the mouse colon in detectable limits as determined by general PCR. Based on these criteria, seven microorganisms namely *Bacteroides thetaiotamicron*, *Bacteroides intestinalis, Prevotella melaninogenica, Lactobacillus planatarum*, *Dorea formicigenerans*, *Clostridium phytofermentans* and *Oribacterium sinus* were selected for further study. To screen for the presence of these species in the mouse colon microbial community, species specific primers were designed (Appendix 5). Table 3.8 shows the genes for which the species specific primers were designed and the results obtained with these primers after amplification by general PCR. The results indicated that *B. thetaiotamicron* was the microorganism that fulfilled all the above criteria. Thus, for the expression studies by real time qPCR, primers for reference gene (*rpoB*) and target genes (*arsA, arsB, arsC* and *nrfA*) were designed based on the sequence of these genes from *B. thetaiotamicron.*

Organism	Phyla	*arsA*	*rpoB*	*nrfA*
Bacteroides thetaiotamicron	Bacteroidetes	+	+	+
Bacteroides intestinalis	Bacteroidetes	−	+	−
Prevotella melaninogenica	Bacteroidetes	−	−	Ab
Oribacterium sinus	Firmicutes	−	+/−	Ab
Lactobacillus planatarum	Firmicutes	− (6)	NA	NA
Clostridium phytofermentans	Firmicutes	−(4)	NA	NA
Dorea formicigenerans	Firmicutes	Ab	+/−	−

Table 3.8: Genes used to determine the presence of specific microorganisms in the mouse colon microbiota. The symbols indicate amplification (+), no amplification (-), amplification in a few samples (+/−), gene sequence was not found in the genome (Ab) and no primers were designed (NA). The number in parentheses shown above indicates the number of primer pairs tested. For others only one primer pair was tested.

Specificity of the primers designed for expression studies

Due to non-availability of the pure culture, genomic DNA from mouse colon was used to test the specificity of each of the primer pair designed for amplifying reference gene (*rpoB*) and target genes (*arsA, arsB, arsC* and *nrfA*). A total of 30 different primer pairs were tested in three different reaction conditions for amplifying the *arsC* gene from microbial community DNA of the mouse colon. However no amplification was achieved with any of the primer pairs tested. These results indicate that either the *arsC* gene is present at a very low abundance in colon microbiota or the PCR conditions or the primers used were not suitable for the amplification of *arsC* gene from colon microbiota.

Figure 3.14 shows representative agarose gels with 185, 121, 144 and 163 bp PCR products obtained from mouse colon microbial community DNA with *rpoB*, *arsA, arsB* and *nrfA* primer pairs respectively. The PCR products thus obtained were free of any secondary products, as indicated by the formation of a single band of desired length on a regular agarose gel. The absence of any visible band in the negative control (water) indicated that the primers do not form dimers under the PCR conditions tested. To further confirm the specificity of the primers used, the PCR product obtained from each primer pair was used for direct sequencing. The results from NCBI BLAST search confirmed that the amplified product corresponds to the *rpoB*, *arsA, arsB* and *nrfA* genes of *B. thetaiotamicron*.

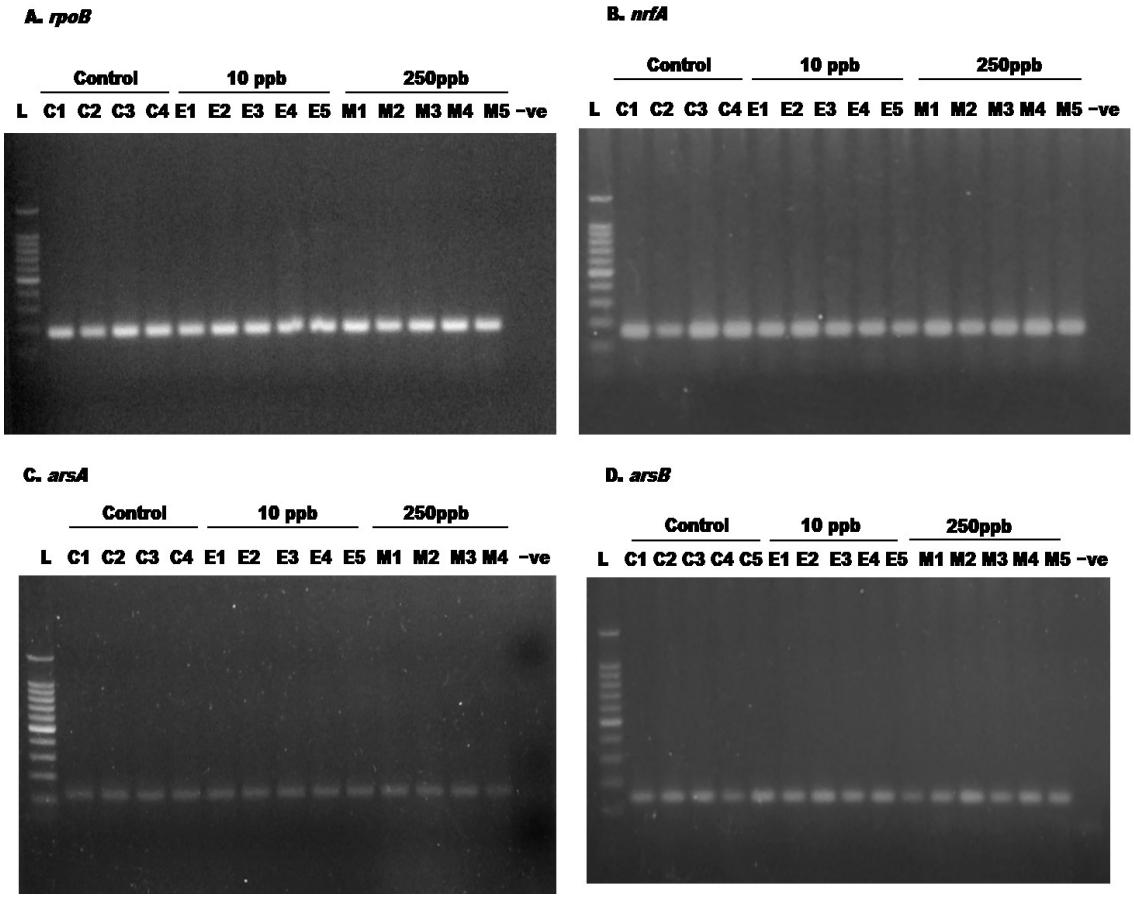


Figure 3.14: Representative 1.8% agarose gels showing amplification of reference [(A) *rpoB*] and target genes [(B) *nrfA* (C) *arsA* and (D) *arsB*] from mouse colon microbial community DNA. L = 100 bp DNA ladder, -ve = negative control.

As mentioned in table 3.7, *B. thetaiotamicron* had 2 *ars* operons with gene configurations *BCADxxxR* and *BADxxxR* respectively. Table 3.9 shows the nucleotide identity of the genes that constitute the two *ars* operons from *B. thetaiotamicron*. Although *arsA* and *arsB* genes were not 100% identical at the primer binding sites, the primers designed specifically for amplifying part of *arsA* and *arsB* gene from operon 1 (*BCADxxxR*) of *B. thetaiotamicron*, amplified the *arsA* and *arsB* regions from operon 2 (*BADxxxR*) as confirmed by direct sequencing of the PCR product. These results indicated that the operon 2 (without *arsC* gene) from *B. thetaiotamicron* or from a related strain was present in more copies in the mouse colon as compared to operon 1.

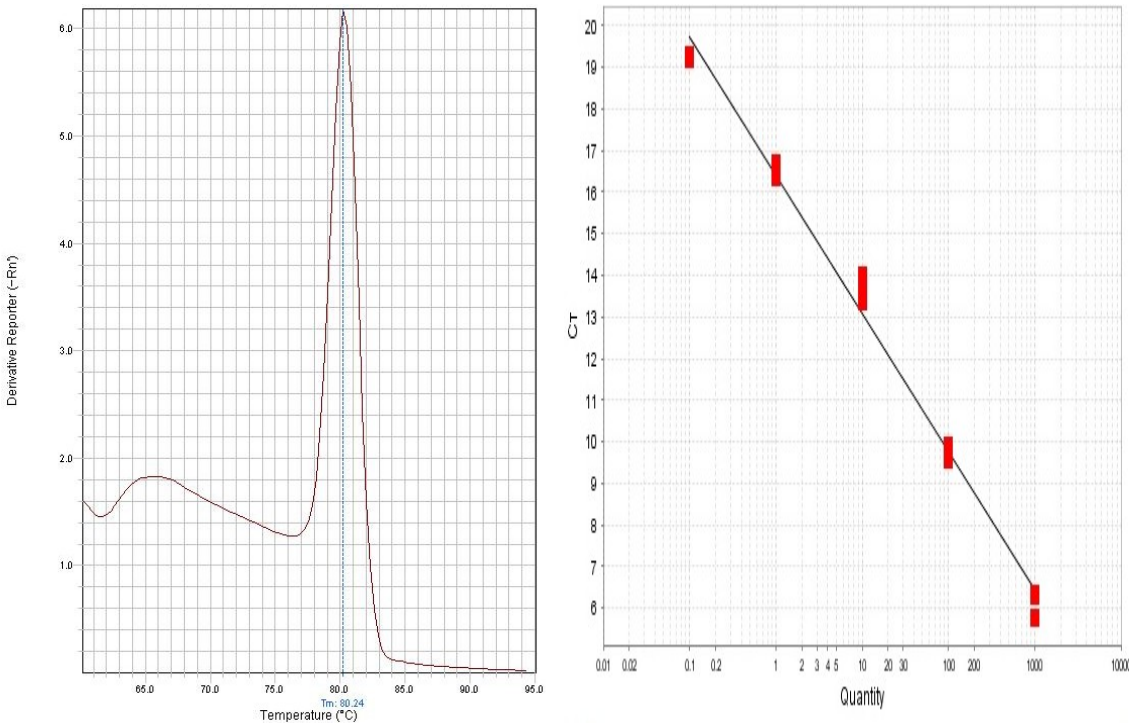
	arsA	***arsB***	***arsD***	***arsR***
Nucleotide similarity	86%	84%	87%	88%

Table 3.9: Nucleotide identity between the *ars* genes from the two *ars* operons of *B. thetaiotamicron.*

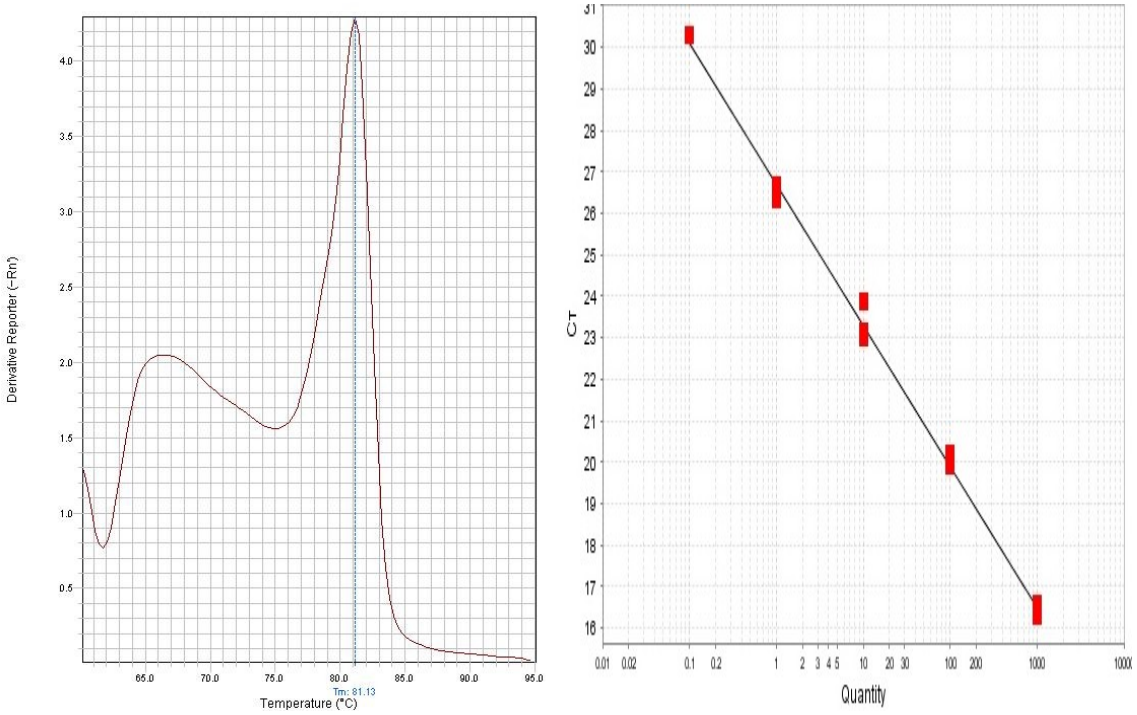
Accuracy of primer pairs for real time PCR

In order to obtain reliable results from comparative Ct method (ΔΔCt) the efficiency of reference gene and the target gene should be within 10% range. The efficiency of these genes was calculated by preparing a dilution series of the PCR product obtained from the general PCR performed on microbial community DNA extracted from mouse colon. Figure 3.15 shows the melt curve and the standard curve obtained for each of the primer pairs. The efficiency for *rpoB*, *nrfA, arsA* and *arsB* primer pairs was 99%, 96%, 99% and 90% respectively. A single Tm peak was observed in the melt curve for each of the primer pair indicating that the primers were suitable for real time PCR.

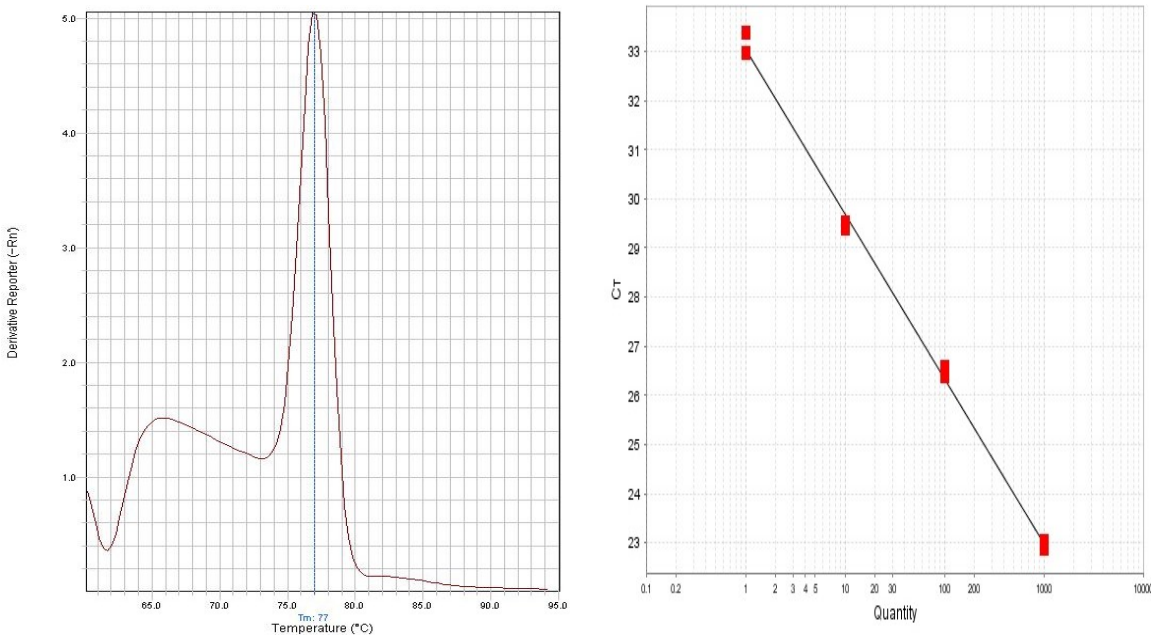
A. rpoB


Derivative Reporter (-Rn')
Tm: 80.24
Temperature (°C)
CT
Quantity

B. nrfA


Derivative Reporter (-Rn')
Tm: 81.13
Temperature (°C)
CT
Quantity

C. arsA



D. *arsB*

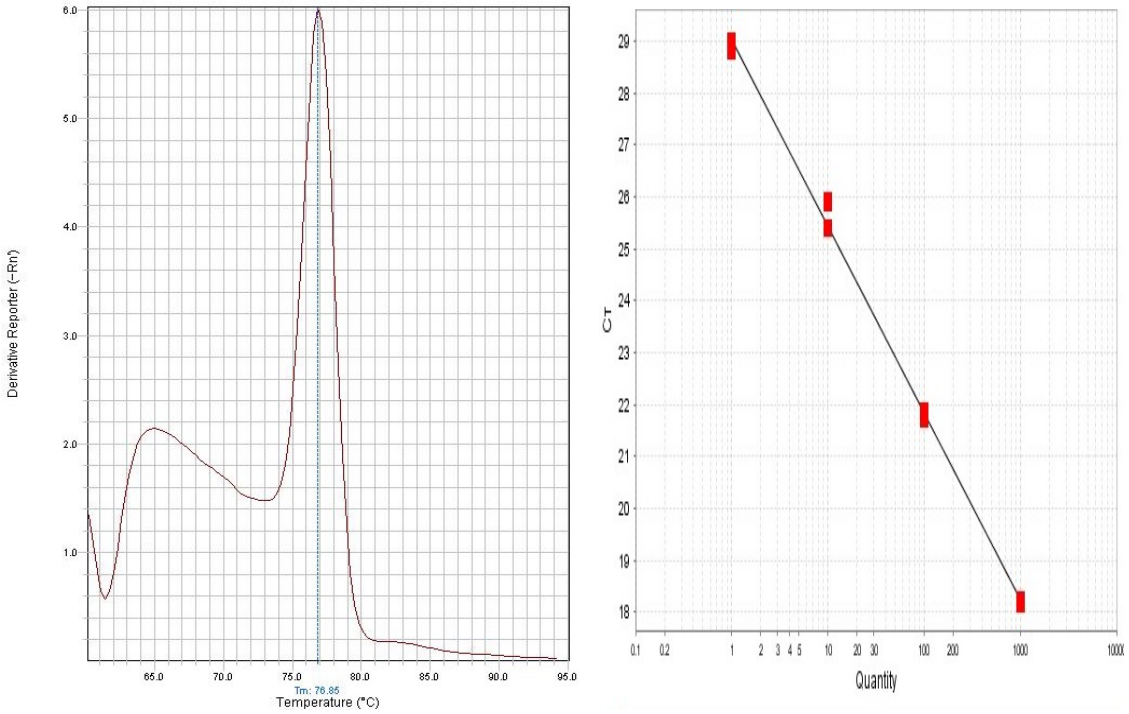
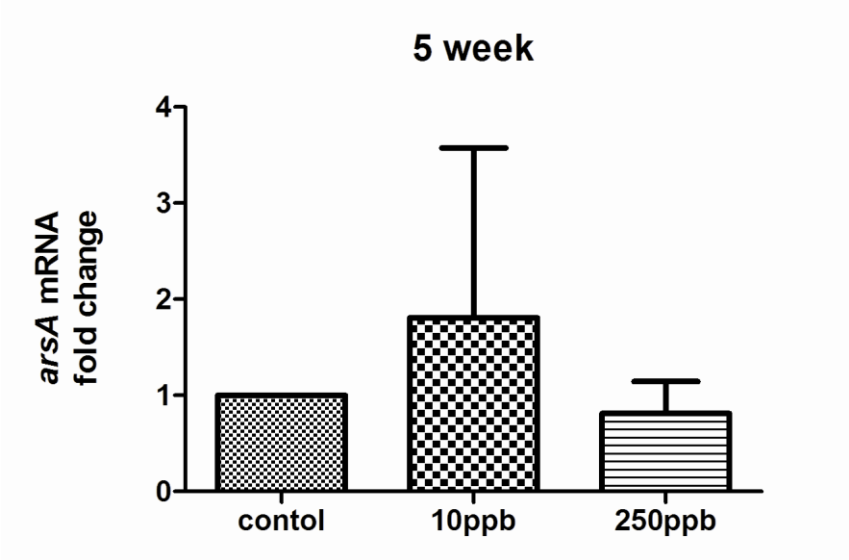


Figure 3.15: Melt curve and standard curve obtained for (A) *rpoB* (B) *nrfA* (C) *arsA* and (D) *arsB* gene.

Relative quantification of arsA and arsB expression in mouse colon

The mRNA expression levels of *arsA* and *arsB* genes between control, 10 ppb and 250 ppb mouse colon samples were compared by comparative Ct method ($2^{-\Delta\Delta Ct}$) of real time RT-qPCR. The *arsA* and *arsB* genes at 2 week time point yielded Ct values above 35 indicating that the genes were expressed at very low levels and could not be compared by comparative Ct method. However for 5 and 10 week mouse colon samples, the Ct values for *arsA* gene were lower than 35, indicating that the gene copy number of *arsA* is higher in 5 and 10 week colon samples. Figure 3.16 shows the dose dependent response of *arsA* expression in the 5 and 10 week mouse colon samples after As(III) exposure. These results indicate that As(III) exposure at a concentration of 10 and 250 ppb does not affect the expression of *arsA* in gut microbes.

With respect to the *arsB* gene, Ct values varied from sample to sample. For most of the 5 week samples at least one of the replicates of each sample yielded melt curve with desired Tm of 76^0C and a single peak, whereas for the 10 week colon samples multiple Tm peaks were observed with most of the samples indicating that the *arsB* gene was not expressed at levels to be compared by real time qPCR. Figure 3.17 shows the dose dependent response of *arsB* expression in 5 week mouse colon samples after As(III) exposure. No evidence for changes in expression of the *arsB* gene was observed in the 5 week samples in response to As(III) exposure. Together the results from expression studies of *arsA* and *arsB* genes indicated that As(III) exposure at a concentration of 10 and 250 ppb via drinking water does not alter the expression of *ars* operon in mouse gut microbes up to 10 weeks.



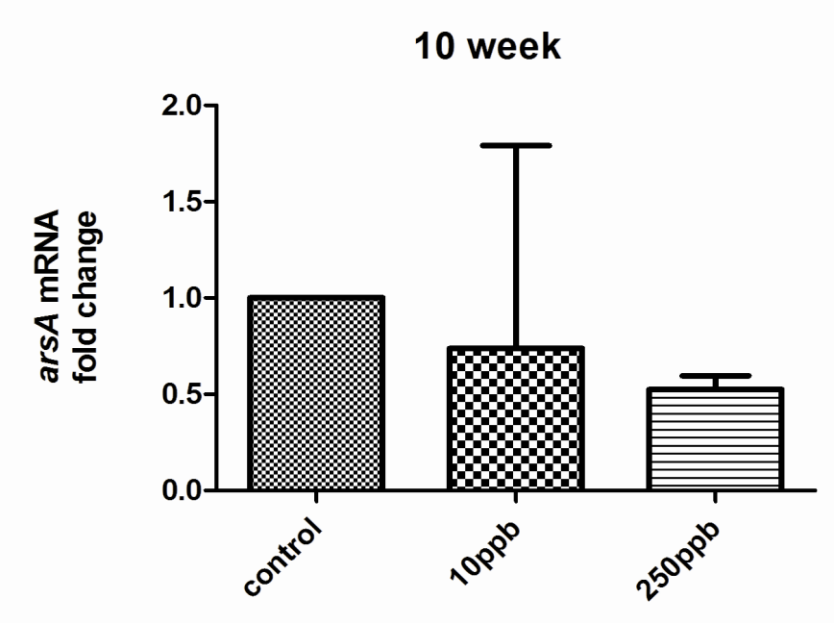


Figure 3.16: Relative expression levels of *arsA* gene in control, 10 ppb and 250 ppb mouse colons at (A) 5 week and (B) 10 week. Error bars indicate standard deviation.

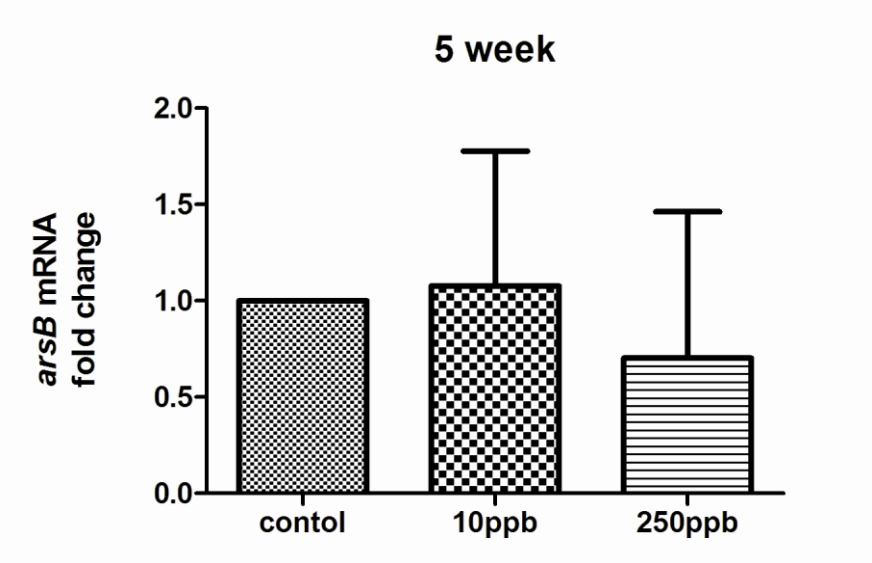


Figure 3.17: Relative expression levels of *arsB* gene in control, 10 ppb and 250 ppb mouse colons at 5 week time point. Error bars indicate standard deviation.

As(III) induced dose dependent alterations in nrfA expression

To determine the effect of As(III) exposure on physiology of the colon microbes, the mRNA expression levels of *nrfA* gene were compared between control, 10 ppb and 250 ppb mouse colon samples by comparative Ct method ($2^{-\Delta\Delta Ct}$). Figure 3.18 shows the time and dose dependent response of *nrfA* expression in 5 and 10 week mouse colon samples after As(III) exposure. These results indicated that As(III) exposure at a concentration of 250 ppb induced significant increase in the expression of *nrfA* in gut microbes. For the 10 ppb As(III) exposure dose, significant increase in the expression of *nrfA* was observed only after 10 weeks of As(III) exposure. Altered expression of the *nrfA* gene in time and dose dependent manner in response to As(III) exposure demonstrated that As(III) exposure impacts the physiology of the microbes residing in the mouse colon.

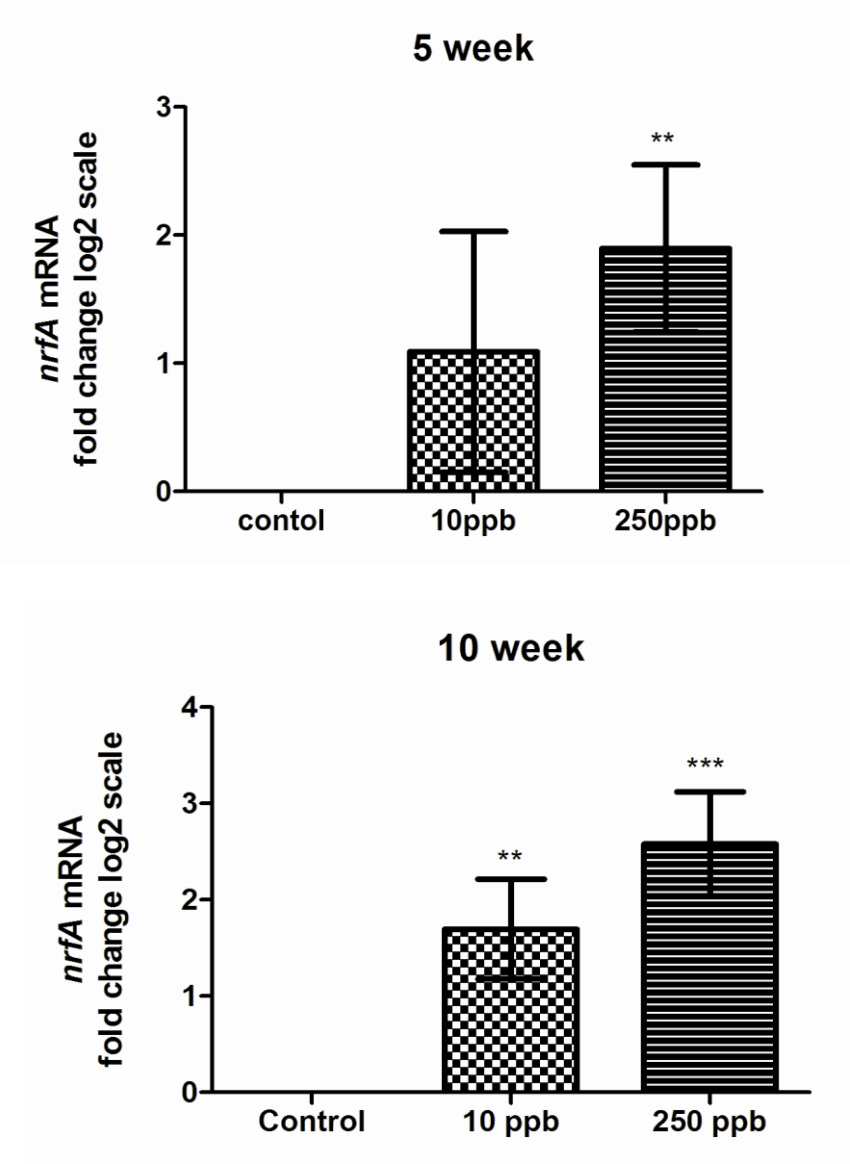


Figure 3.18: Relative expression levels of *nrfA* gene in control, 10 ppb and 250 ppb mouse colons at (A) 5 week and (B) 10 week time point. Error bars indicate standard deviation. ** indicates $p < 0.01$ and *** indicates $p < 0.001$ after Tukey's post-hoc test.

Chapter 4 : DISCUSSION

4.1 Molecular techniques as a tool for studying colon microbiota

In recent years, the use of culture independent techniques have grown at a fast pace and have extended our knowledge about the diversity and composition of the mammalian microbiota. This development has overcome the disadvantages associated with the traditional time consuming and labor intensive cultivation based approaches wherein the difficulty lies in differentiating related species/strains as well as in culturing anaerobic gut microbes under laboratory conditions (Duncan et al, 2007). Numerous studies have used 16S rRNA gene based sequencing (pyrosequencing, near full length 16S rDNA sequencing) and fingerprinting techniques (DGGE, RFLP) to study the gut microbiota (Zoetendal et al, 2002; Nava et al, 2011; Eckburg et al, 2005; Osborn et al, 2000). In addition some of the studies have used the variability in the spacer region between 16S and 23S rRNA gene (RISA, ARISA) of different bacterial species for comparative studies related to gut microbiota (Scanlan et al, 2008; Spencer et al, 2011). However the dynamic and complex nature of the mammalian microbiota along with the microbial variation between individuals pose a challenge in these studies as each of these technique comes with its own advantages and limitations (Sekirov et al, 2010). 16S rDNA clone libraries based results give high taxonomic resolution but is very time consuming and expensive approach. Pyrosequencing on the other hand is convenient and cost effective approach in case of multiple samples but provides shorter reads of <400 bp, thus giving only class or family level taxonomic resolution (Liu et al, 2008). DGGE and RFLP are rapid methods for comparing bacterial profiles from different groups but provide low taxonomic resolution due to smaller product size. With DGGE it is difficult

to achieve reproducible gels and the technique itself is associated with inherent biases since different bacterial species that have same GC content migrate to the same position in the gel (Muyzer & Smalla, 1998). Both RISA and ARISA are highly reproducible and less time consuming approaches. These provide low taxonomic resolution mainly because the intergenic region database is not extensively available, especially for gut microbes where majority of the bacteria have not been sequenced yet (Kovacs et al, 2010). ARISA, although more expensive than RISA, is advantageous as the resolution of PCR products on the capillary is very high and it provides an accurate estimation of intergenic length. However in both cases, the technique is biased as a single species can have several intergenic regions with each one differing in length. Additionally a number of bacteria have been identified where 16S and 23S rRNA genes are not organized in operons but are separated from each other with > 3000 bp distance between them. This size between 16S and 23S rRNA genes is far greater than the normal size separated on a gel or a capillary (Kovacs et al, 2010). All of the fingerprinting techniques suffer from the limitation of being unable to detect species that are present at a very low abundance (<1%) in a mixed community. Fluorescent in situ hybridization (FISH) and RT-qPCR are two other 16S rRNA gene based techniques that are used to identify and compare bacterial communities in the gut (Kalliomäki et al, 2008; Penders et al, 2007). For both of these techniques either species specific or group specific probes/primers need to be designed, thus limiting the identification of novel species. Depending upon the conditions used for hybridization and PCR, these probes/primers can bind to other closely related species producing biased results (Sekirov et al, 2010).

The primary goal of this study was to compare the bacterial profile of control and As(III) exposed murine colon and to identify the taxonomic groups that are mainly affected by arsenic exposure. In order to achieve this goal and to overcome the constraints associated with cultivation independent techniques, 4 different molecular approaches (DGGE, RISA, ARISA, clone library) based on principles of inter species differences of GC content, variable length of 16S and 23S intergenic regions among different bacterial species and variability within the conserved 16S rRNA gene were used in this study to determine the effect of As(III) on colon microbiota.

4.2 Biases associated with method of DNA extraction

Besides individual limitations, all molecular techniques are prone to biases associated with methods of DNA extraction and PCR. The silica and guanidinium thiocynate method as well as QIAamp DNA stool mini kit from Qiagen Inc. were used in this study for extracting microbial community genomic DNA from mouse colon. Both the methods have been used previously in numerous studies to extract microbial DNA from the stool samples of various species (Larsen et al, 2010; Boom et al, 1990; Barnard et al, 2011). Of these two methods, the silica method is inexpensive but time consuming procedure. The silica slurry used for isolating DNA requires a two day preparation time and requires preparation of buffers that are not regularly used in other procedures. Moreover when the same amount of DNA was used for PCR, some of the samples did not show any or minimal amplification indicating that PCR inhibitors were left in the final extracted genomic DNA (Figure A.1). QIAamp DNA stool mini kit protocol on the other hand gave more or less consistent yields of genomic DNA as well as PCR products for all the samples when equal amounts of starting materials were used. However in both

the methods, variations in the amount of DNA extracted due to incomplete homogenization of the sample cannot be ruled out. Also the 260:280 ratios for the extracted nucleic acid were better with QIAamp DNA stool mini kit protocol. Due to inconsistencies associated with silica protocol, only results obtained with QIAamp DNA stool mini kit from study 2 and part of study 1 are discussed below.

4.3 Microbial composition of the murine colon

Most of the studies related to the microbial communities in the colon of the host have used stool samples as their study source. However, within the gut the composition of microbial communities differs longitudinally between ascending and descending colon as well as radially between mucosal and luminal region (Frank et al, 2007; Zoetendal et al, 2002). To overcome these differences and to ensure that the bacterial community directly associated with the gut lining is accounted for, both the ascending and descending colon was used in this study to determine the in vivo impact of As(III) on colon microbiota.

In the present study, comparison of RISA and ARISA profiles obtained from microbial community DNA of murine colon were used to track the effect of As(III) exposure on murine colon microbiota. The number of detectable bands obtained from RISA profiles of an individual mouse colon ranged from 16 – 30 and the similarity value between individuals of the same group ranged from 40 – 85%. The similarity index was relatively low considering that the mice were genetically identical and housed under same conditions. An attempt was made to sequence some of the dominant phylotypes obtained from RISA both by making a clone library and by excising specific bands from the gel. Reamplification of excised bands generally yielded multiple bands. This was attributed to

low resolution of RISA products on the polyacrylamide gel. Of the 21 sequences obtained from RISA clone libraries, 18 sequences showed similarity with *Lactobacillus* (Appendix 6). The product length of the 18 sequences varied from 290 bps to 504 bps indicating that the spacer regions of *Lactobacillus* species and strains are distributed over a broad range. These results also showed bias towards preferential insertion of smaller products into the cloning vector. These results indicate that the inexpensive and rapid RISA approach is more suitable for comparative studies rather than the studies that require taxonomic identification of bacterial communities present in the host.

ARISA profiles from individual colon samples yielded only 15 − 35 peaks whereas about 113 peaks were obtained collectively from all the samples. The number of bands and peaks obtained from RISA and ARISA profiles were far less than the number of species typically found in a host indicating that the techniques were not suitable for diversity estimations. The ordination plots obtained after statistical analysis of ARISA data differed between peak height, area and presence/absence and so does the peaks that contributed to the variation in these plots. Similar differences have been observed between the three types of ordination plots in other studies too, wherein it was recommended that ordination plots obtained from relative peak height are best suited for ordination based analysis of microbial community (Culman et al, 2008).

Results from DGGE showed that the number of bands ranged from 20 − 40 in any individual mouse. Significant variation was observed between individuals of the same group with similarity index values ranging from 55 – 90% which was considerably low for the lower range. However other DGGE based studies on gut microbiota of mouse have showed comparable similarity index values (McCracken et al, 2001). Similar to

RISA and ARISA profiles, the number of bands obtained from DGGE fingerprints were far less than the number of species typically found in a host. These observations indicate that the DGGE based approach is more suitable for comparative studies that involve detecting shifts in microbial populations rather than diversity estimation studies. Excising specific bands from the DGGE gel and sequencing them or preparing a clone library from the same PCR products can give information about the type of bacteria present in the sample. However the PCR products resolved on DGGE gel in this study were ~177 bps in length and ≥50% of the obtained sequences were not classified to any phylum at 90% confidence threshold. When the classification was repeated at low confidence threshold of 50 – 60% most of the unclassified bacteria resolved into the phylum level of taxonomy. Moreover, the classification of these sequences to lower taxonomic levels was highly ambiguous. A similar pattern was observed with clone libraries obtained from the PCR products used for DGGE analysis (Figure A.5). These results showed that sequences obtained from DGGE are best classified at the phylum level of taxonomy or at the class level. Within each phylum the % GC ratio of species varies significantly. Thus sequences from Bacteroidetes and Firmicutes, the two predominant phyla found in mouse colon, were observed throughout the length of DGGE gel (Table A.1). Additionally no chimeric sequences were detected in a total of 418 sequences obtained from DGGE gel or DGGE-PCR product clone library. This was in concordance to the fact that it is difficult to detect chimeras within smaller sequences (Haas et al, 2011). A typical near full length 16S clone library generally yields 5−30% sequences that are chimeric (Ashelford et al, 2005). Inability to detect chimeras in a DGGE sequence library indicates a potential bias associated with this approach. Despite these limitations, DGGE is a relatively

inexpensive method for detecting shifts in microbial populations under different conditions.

The results from near full length 16S sequence analysis as well as the DGGE-PCR product based library showed Bacteroidetes and Firmicutes as the predominant phyla in the murine colon accounting for >90% of the total sequenced population. This is in concordance with pyrosequencing based studies obtained from the colon and cecal samples of the same strain of mice (Nava et al, 2011; Ley et al, 2005; Hoffmann et al, 2009). However those studies also found Actinobacteria, TM7, Acidobacteria, Tenericutes and Cyanobacteria in the murine colon and cecal regions comprising less than 1% of the total sequenced population. At the family level, 10 families were identified in this study with near full length 16S clone library based approach (Figure 3.8). However sequences belonging to Prevotellaceae, Rikenellaceae, Clostridiaceae and other rare families that have been observed in murine colon by other research groups were found only in the clone libraries obtained from DGGE-PCR products (Figure A.6). These differences in the identification of rare phyla and assignment of sequences to lower taxonomic levels could reflect bias associated with inadequate sampling, criteria used for the classification of sequences, the length of 16S region amplified and most importantly the different primer sets used in the DGGE-PCR clone library, pyrosequencing and near full length 16S based sequencing approaches. At the genus level ~65% of the sequences were not assigned to any genus level, a trend found by other researchers too indicating that the lower GI tract of mammals is largely colonized by uncultivated and uncharacterized bacteria (Ley et al, 2005). These results emphasize the need to cultivate and study the microbes that colonize the gut environment.

4.4 Microbial community of the host changes with time

An important observation found in this study was the gradual change in host microbiota with time (Figure 3.7). This observation was consistent with previous studies that have shown the microbial community of the host to be dynamic and changing with time (Tiihonen et al, 2010). However those time dependent studies estimated the microbial difference between infants, adult and elderly humans which is different from this study wherein differences in colon microbiota were observed within intervals of few weeks in adult mice. Other time dependent studies that had observed the effect of diet, antibiotic or any other factor on host microbiota have used only a 0 time point reference as the control (Dethlefsen et al, 2008; Ley et al, 2006). Only one study with four week old mice was found to reveal the differences in the microbial composition of the luminal and mucosal portions of caecum, proximal and distal colon within a five day interval (Hoffmann et al, 2009). No distinct pattern was observed regarding the relative proportions of Bacteroidetes and Fimicutes in the above study. However the data presented in the current study indicates a gradual increase in the ratio of Bacteroidetes and Firmicutes over time. The reason for this shift in microbial communities over such a short time span could be related to the developmental changes occurring in the mouse within this time frame. Thus the above observation underscores the importance of using separate control for each of the time points in microbiota related time course experiments.

4.5 Effect of arsenic on microbial composition of the murine colon

Exposure to arsenic via drinking water is a global health problem. The effect of acute and chronic exposures of arsenic on human health differs significantly. Acute exposure to arsenic at high doses generally leads to GI hemorrhage, multiple organ

damage and death (Armstrong et al, 1984). On the other hand, chronic exposures at environmentally relevant doses alter cell signaling pathways leading to cancers of various organs, cardiovascular disease and diabetes (WHO, 2001; Hughes, 2002). Therefore in this study the in vivo effect of As(III) on colon microbiota was determined after adding sodium arsenite to drinking water at concentrations that are commonly found in the environment. The results from this study reveal that As(III) in drinking water affects the microbial composition of the gut. The differences between the control and As(III) exposed mouse colon microbiota were more significant than the differences that could have been due to variability within each group. Results from all the molecular approaches used in this study showed shifts in host microbial populations after two weeks of 250 ppb of As(III) exposure and ten weeks of 10 ppb of As(III) exposure. Banding patterns and dendrogram analysis of RISA and DGGE gels showed that the molecular similarity between the 2, 5 and 10 week 250 ppb exposed mice is greater than the similarity between any of the 250 ppb mice to any of the control group mice from the same time point. This type of intra group similarity was observed for 10 ppb As(III) exposed mice only after 10 weeks of As(III) exposure. PCA plots obtained from ARISA showed scattering of control and to a certain extent of the 10 ppb groups (except in 10 week samples) across the plot whereas the 250 ppb groups tended to form clusters. The results obtained from the 16S clone libraries were analyzed at various levels of taxonomical hierarchy to identify the taxa that differed significantly between control and As(III) exposed groups. An increase in Bacteroidetes and a proportionate decrease in Firmicutes were observed with increasing dose of arsenic exposure versus control groups. As expected the taxonomic composition of the 10 ppb group was intermediate to the control

and the 250 ppb group but like the 250 ppb group, the average Bacteroidetes population was also more in the 10 ppb group as compared to the control group. The stacked bar graph obtained for relative taxonomic distribution of near full length 16S as well as DGGE-PCR product based clone libraries showed that the 10 week 250 ppb exposed group is more similar to the 10 week control as opposed to the 2 and 5 week 250 ppb group, which differed significantly from their respective control groups. Also the level of significant differences between the control and 250 ppb group of 10 week changed when different confidence thresholds were used for classification. However, the significant difference remained the same at all confidence thresholds for all the other libraries used for comparison. The reason for such observation is unknown. The phylogenetic analysis showed that the changes observed between control and As(III) groups were attributed to class Bacteroidia and Clostridia and families Porphyromonadaceae, Lachnospiraceae and Ruminococcaceae. At the genus level over 19% of the increase in Bacteroidetes in As(III) exposed groups was associated with *Barnesiella*. In addition, an increase in Bacilli and *Lactobacillus* was found in the As(III) exposed group. Since many sequences remain unclassified at the order level it is not possible to conclude whether As(III) affects different genera and families within a phyla differently, causing one subgroup to increase while causing other to decrease. The results presented in this study reveal that, due to inter individual variations in microbial community structure, it is necessary to use a combination of molecular and statistical approaches to demonstrate the effect of any variable on the microbial community of the host.

Human and mouse show considerable similarity in their distal gut microbiota at the higher levels of bacterial taxonomy (Ley et al, 2005). Several recent studies have used

murine models to reveal a relation between diseased and altered gut microbial communities (Bibiloni et al, 2005; Turnbaugh et al, 2006). Some of these studies as in the case of obesity showed similar results when repeated on human subjects (Ley et al, 2006; Turnbaugh et al, 2006). Thus it is plausible that the results from this work may be extended to humans.

4.6 Effect of arsenic on microbial diversity

The results from DGGE and RISA showed no significant difference in the total number of bands obtained from control and As(III) exposed groups. Similarly ARISA profiles and 16S clone libraries did not showed any difference in microbial richness between control and As(III) exposed groups. For DGGE-PCR product based clone libraries, the 250 ppb As(III) group was more diverse in the 5 week and less diverse in the 10 week libraries when compared to the control groups. However, the diversity index data from the near full length 16S based clone libraries showed low diversity in the 250 ppb As(III) groups when compared to control for 2, 5 and 10 week time points at genetic distances of 0.05 and 0.1. These differences in the diversity estimations of 5 week samples between 16S and DGGE based clone libraries could be related to the differences in DNA extraction procedures used for obtaining the microbial community DNA. In the 10 ppb group from 2, 5 and 10 week samples the diversity was always more than the 250 ppb group but no particular trend was observed with respect to the control group from the same time point. Although no firm conclusion about the effect of As(III) on the microbial diversity of the mouse colon can be made from this study, nevertheless, the reduced diversity in the 250 ppb group that is predominated by Bacteroidetes might be related to the fact that Bacteroidetes is a far less diverse phylum when compared to the

Firmicutes. Reduced diversity in the colon may result in a decreased number of microbe associated metabolic pathways in the colon that might affect the physiological processes of the host.

On the basis of diversity estimations and taxonomic classification, it can be concluded that in this study although the rare bacteria might not have been sampled efficiently, the effect of arsenic on predominant groups has been resolved.

4.7 Selective effect of arsenite on host microbiota: a direct effect

Bioinformatics analysis of distribution and occurrence of the *ars* operon in gut microbes showed that far more Bacteroidetes harbor this arsenic resistant mechanism as compared to the Firmicutes (Table 3.7). Grouping of ArsA and ArsB sequences from Bacteroidetes in the same cluster of the phylogenetic tree indicated common origin of these operons. However, the two proteins showed little similarity with any of the reference ArsA and ArsB sequences, which was surprising considering that most of the Bacteroidetes species had the *E.coli* type five gene *ars* operon configuration consisting of *arsA, arsB, arsC, arsD* and *arsR* genes. Most of the Firmicutes were devoid of *arsA* gene but their ArsB protein sequence clustered with *B. subtilis* indicating that these Firmicutes possess an Acr3 type As(III) transporter system.

Primers designed specifically for amplifying part of *arsA* and *arsB* DNA and cDNA from operon 1 of *B. thetaiotamicron*, in return amplified the regions from operon 2 as confirmed by sequencing. The reason for amplification of *arsA* and *arsB* cDNA from operon 2 could be related to the fact that in species that have more than one *ars* operon, the operon that lacks *arsC* gene is expressed at high levels in the presence of arsenite. However the amplification of 2nd operon with genomic DNA indicates that the sequence

for 2nd operon is present in more copies in the colon microbial community DNA. Presence of multiple copies of 2nd operon in community DNA could relate to either intra or inter genomic mobilization of the *ars* operon under selective effect of As(III) exposure or gene duplication or binding of primers to *ars* operon of other closely related Bacteroides species. Bacteroidetes populations have been documented to possess well developed mechanisms for transposition of antibiotic resistant genes to other species (Salyers et al, 2004). Occurrence of the *ars* operon on plasmids and the presence of transposition related genes in the vicinity of the *ars* operon of certain species coupled with the results above indicate that the *ars* operon can undergo horizontal transfer under As(III) pressure and needs further investigation.

No change in expression of *arsA* and *arsB* was observed in response to 10 and 250 ppb of As(III) exposure under in vivo conditions of the colon. The reason could be that neither As(III) nor its metabolic forms are present in enough quantities in the colon so as to induce the expression of these genes. Studies with *Shewanella* sp strain ANA 3 have demonstrated that the expression of *ars* operon requires ≤ 100 μM concentrations of As(III) in culture media and only 1.5 fold increase in expression was observed with 10 fold increase in As(III) in the media (Saltikov et al, 2005a). These values for induction and fold increase in *ars* expression varies from species to species with a 14 - 39 fold change in the expression of *ars* operon genes observed in the presence of equal amounts of As(III) with different *C. jejuni* strains (Wang et al, 2009). However, the culture conditions used in those studies were very different from the current study where the fold change in *arsA* and *arsB* expression was determined under in vivo conditions with 10 and 250 ppb of As(III) in drinking water. Although levels of arsenic tested in this study did

not appear to induce any expression of *ars* genes in colon bacteria, the basal level of expression of these genes in colon bacteria might still provide selective advantage to the Bacteroidetes populations that harbor arsenic resistance genes as opposed to majority of the Firmicutes generally found in the colon that do not possess *ars* operon.

Arsenic compounds have been used since ages as antimicrobial agents at high doses (Leonard, 1991). Arsenite levels in the range of 4 – 64 μg/ml of culture media have been found to be inhibitory to *C. jejuni* strains that possess *ars* operon ((Wang et al, 2009). However the effect of ppb levels of As(III) on microbes that do not possess mechanism for arsenic resistance is not known. Another reason for the decrease in Firmicutes population in As(III) exposed murine colon could be to due antimicrobial action by other gut residents that bloomed in the colon in response to As(III) exposure. For example *Lactobacillus* converts sugars to lactic acid resulting in the acidification of the colon environment (Corr et al, 2009). An increase in *Lactobacillus* in response to As(III) exposure can suppress the growth of other gut microbes by reducing pH. However further work is needed to determine the role of microbe-microbe interaction in development of As(III) associated microbiota.

This part of the study underscores the importance of measuring levels of arsenic and its metabolites in the colon as most of the ingested arsenic happens to be absorbed in the small intestine and any remaining amounts of arsenic or its metabolites may or may not have any direct effect on the viability of gut microbes per se.

4.8 Effect of arsenic on the microbial physiology of the host

Increased levels of nitrite, as an indicator used for estimating NO levels in the host, have been observed in mouse liver after As(III) exposure at low doses

(Barchowsky, unpublished data). Additionally, increased levels of nitrogen species such as peroxynitrite have been reported in mouse liver in response to the generation of superoxide anion from As(III) stimulated NADPH oxidase (Straub et al, 2008). Increased NO levels can affect the gut lining, cause inflammation and can have a bactericidal effect on gut bacteria (Endo et al, 2010; Kim et al, 2008; Sobko et al, 2006). An increase in NO levels in the host may result in selective increase of microbes that possess genes to combat high levels of NO. *E.coli* has been shown to combat NO toxicity by inducing the expression of *nrfA* (Poock et al, 2002). In this study, time and dose dependent increase in *nrfA* expression was observed in the mouse colon microbes after As(III) exposure under in vivo conditions. The increased expression of *nrfA* in colon microbes along with selective colonization of the colon by microbes that possess this gene may alter important metabolic pathways in the host that could result in development of pathogenic conditions in the host colon. The effect of As(III) on NO levels of the colon epithelia and microbial generation of NO is not known currently. However, based on the results above, increased NO level in the host in response to As(III) exposure seemed like a plausible explanation for the observed increase in the expression of *nrfA* in As(III) exposed mouse colon.

4.9 Arsenic induced changes in the host: An indirect effect on gut microbiota

Although evidence for angiogenesis and vascular remodeling in mouse liver indicated by defenestration of liver sinusoidal epithelial layer were seen by two weeks after exposure to 10 ppb of arsenite (Straub et al, 2008), the changes in the microbial flora of the murine colon started appearing at two weeks but these changes were not significant until ten weeks of arsenite exposure. These observations suggest that the effect of As(III) on gut community occurs in response to systemic changes of the host

and are not due to the direct exposure to arsenic itself. The interaction between colon epithelium and its microbiota is very complex with microbes present in the mucus layer of the epithelial lining playing a role in the regulation of epithelial cell regeneration and mucus layer production by goblet cells present in the epithelia. The outer mucus layer of the epithelia is the site where bacterial colonization occurs whereas the inner mucus layer acts as a shield and prevents microbial invasion into the host blood stream (Cherbuy et al, 2010; Johansson et al, 2011). Like respiratory epithelia, colon epithelia may secrete innate immune protein like PLUNC that can regulate epithelial hydration as well as biofilm formation (Gakhar et al, 2010). Thus the colon epithelial environment influences the colonization of commensal bacteria by immune regulation and by providing access to complex carbohydrates in the mucus layer. In situ observations using transmission electron microscopy revealed that the small 0.25 μm coccoids (e.g. *Ruminococcus*) disappeared by the 5 week time point in mice treated with 250 ppb of As(III) (Stolz, unpublished data). This population also disappeared from the As(III) exposed 16S rRNA based clone libraries. Dietary factors such as Vitamin A deficiency affect the proliferation rate and differentiation of epithelial cells in the intestine (Uni et al, 2000). It has been well established that arsenic at low levels alter signal transduction pathways involved in cell cycle regulation and differentiation and proteins involved in immune response and inflammation (Kozul et al, 2009b; Druwe & Vaillancourt, 2010). Although the effect of arsenic on gut epithelium is not known, the above evidence suggest that As(III) exposure can affect the colon epithelium differentiation, regeneration and function making it less favorable habitat for certain bacterial populations. In this way, As(III) associated changes in the gut microbiota can be more likely linked to arsenic

induced changes in the host and gut environment rather than the direct impact of toxic arsenic on gut community.

4.10 Implications of Arsenic induced changes in colon microbiota on host health

Gut microbes play a key role in various important processes in humans including nutrient processing, energy harvesting, immune system regulation and brain development (Hooper et al, 2002; Macpherson & Harris, 2004; Turnbaugh et al, 2006; Heijtz et al, 2011). In this study an increase in Bacteroides and lesser proportions of Firmicutes were observed in colon of As(III) exposed mice. Similar trends involving an increased ratio of Bacteroidetes to Firmicutes have also been associated with type I and type II diabetes (Giongo et al, 2011; Larsen et al, 2010) and an opposite trend was observed in obesity (Ley et al, 2006). Higher levels of Bacilli and *Lactobacillus* observed in this study were also observed in type 2 diabetic human adults (Larsen et al, 2010). Bacteroidetes are gram negative bacteria with LPS in their outer membrane which is an important virulence factor and may cause inflammation (Allcock et al, 2001). Additionally Bacteroidetes produce small chain fatty acids as end products of their metabolic pathways that have important neurological effect on rat brain (MacFabe et al, 2007). Moreover arsenic exposure in the environmental settings has been associated with neurological defects, type II diabetes and reduced innate immunity ((Bardullas et al, 2009; Kozul et al, 2009a; Navas-Acien et al, 2008)). Together they suggest that arsenic induced changes in microbial community composition may induce further changes in mouse physiology such as nutrient uptake, fat distribution, weight loss, diabetes and innate immune responses.

Chapter 5 : SUMMARY AND FUTURE DIRECTIONS

The overall hypothesis of this project was that the chronic exposure to environmentally relevant concentrations of As(III) will have direct impact on the composition and physiology of the gut microbiota. By using a combination of molecular approaches this study has shown that As(III) has time and dose dependent selective effect on the murine colon microbiota. These changes in the mouse colon microbiota started appearing at 2 weeks of 250 ppb of As(III) exposure and continued in 10 weeks. For 10 ppb of As(III) exposure levels these changes became significant only after 10 weeks of As(III) exposure. Results from the 16S clone libraries showed an increase in Bacteroidetes and a proportionate decrease in Firmicutes in 10 and 250 ppb As(III) exposed groups versus control groups. Additionally, time and dose dependent changes in the physiology of gut microbes was observed in the form of increased expression of *nrfA* from *Bacteroides* in response to As(III) exposure. However no effect of As(III) exposure was observed on the same organism with respect to the expression of genes involved in arsenic resistance and detoxification (*arsA* and *arsB*). Since evidence for As(III) induced changes in mouse liver were seen by two weeks after exposure to 10 ppb of As(III) (Straub et al, 2008), it was concluded that the observed changes in the colon microbiota occurred more likely in response to As(III) induced changes in mouse physiology rather than the direct impact of As(III) on gut microbes.

The data presented in this work is novel in the sense that it a first step towards understanding the in situ effect of chronic exposure to environmentally relevant concentrations of As(III) via drinking water on the colon microbiota. This study also

sheds light on the possible impact of arsenic on the physiology of microbes residing in the gut by using *B. thetaiotamicron* as the model organism and is the first study to show that arsenic resistant genes are expressed in colon microbes under normal conditions of the gut. However further work needs to be done to address specific questions such as the impact of As(III) associated altered microbiota on the host health and host-microbial interactions. The significance of this study is that it provides framework to study the effect of other environmental contaminants on the GI tract. In any case, further arsenic toxicological studies should include GI microbes as a possible affected organ or as an organ that may play a role in arsenic associated disease progression.

Since enough similarity exists in the colon microbiota of mouse and humans, the findings of this study can be investigated on human subjects. The results presented in this study do not in any way prove that the microbiota of As(III) exposed mice is unhealthy, yet this type of changed microbiota can have health effects in addition to those directly related to arsenic exposure such as weight loss, diabetes, colon cancer and other GI tract related problems. Metabolomics or proteomics approaches might help in deciphering the effect of changed microbiota on host health and will be able to resolve the pathways by which As(III) associated changes in colon microbiota may or may not affect the health status of its host.

Reversal studies involving removal of As(III) after few weeks of exposure and subsequent microbial studies will help in demonstrating whether As(III) associated changes in the microbial community of the host is reversible or not and to what extent. If the effect of As(III) is not reversible, then it should be analyzed whether introduction of prebiotics/probiotics aids in reestablishment of the normal microflora. Another question

worth investigating would be whether reestablishment of normal microflora reverts any of the host physiological conditions caused due to As(III) exposure. These studies will help to establish the link between As(III) associated microbiota and host physiology. Studies demonstrating the effect of chronic exposures of As(III) at low doses on the epithelial lining of the GI tract will help in understanding the mechanisms for selective effect of As(III) on gut microbiota. This study will help in enhancing the current view of the interactions that happen between the GI tract and its microbiota.

REFERENCES

- Abedin MJ, Cresser MS, Meharg AA, Feldmann J and Cotter-Howells J. (2002). Arsenic accumulation and metabolism in rice (*Oryza sativa* L.). Environ Sci Technol 36:962-968.
- Abernathy CO, Liu YP, Longfellow D, Aposhian HV, Beck B, Fowler B, Goyer R, Menzer R, Rossman T, Thompson C and Waalkes M. (1999). Arsenic: health effects, mechanisms of actions, and research issues. Environ Health Perspect 107:593-597.
- Adak S, Aulak KS and Stuehr DJ. (2002a). Direct evidence for nitric oxide production by a nitric-oxide synthase-like protein from *Bacillus subtilis*. J Biol Chem 277:16167-16171.
- Adak S, Bilwes AM, Panda K, Hosfield D, Aulak KS, McDonald JF, Tainer JA, Getzoff ED, Crane BR and Stuehr DJ. (2002b). Cloning, expression, and characterization of a nitric oxide synthase protein from *Deinococcus radiodurans*. Proc Natl Acad Sci 99:107-112.
- Afkar E, Lisak J, Saltikov C, Basu P, Oremland RS and Stolz JF. (2003). The respiratory arsenate reductase from *Bacillus selenitireducens* strain MLS10. FEMS Microbiol Lett 226:107-112.
- Agency for Toxic Substances and Disease Registry (ATSDR). (2007).Toxicological profile for Arsenic. Department of Health and Human Services, Public Health Service: Atlanta, GA.

- Ahmann D, Roberts AL, Krumholz LR and Morel FMM. (1994). Microbe grows by reducing arsenic. Nature 371:750-750.
- Allcock GH, Allegra M, Flower RJ and Perretti M. (2001). Neutrophil accumulation induced by bacterial lipopolysaccharide: effects of dexamethasone and annexin 1. Clin Exp Immunol 123:62-67.
- Allison C and Macfarlane GT. (1988). Effect of nitrate on methane production and fermentation by slurries of human fecal bacteria. J Gen Microbiol 134:1397-1405.
- Andreae MO and Klumpp D. (1979). Biosynthesis and release of organoarsenic compounds by marine algae. Environ Sci Technol 13:738-741.
- Antman KH. (2001). Introduction: The history of arsenic trioxide in cancer therapy. Oncologist 6:1-2.
- Aposhian HV, Gurzau ES, Le XC, Gurzau A, Healy SM, Lu X, Ma M, Yip L, Zakharyan RA, Maiorino RM, Dart RC, Tircus MG, Gonzalez-Ramirez D, Morgan DL, Avram D and Aposhian MM. (2000). Occurrence of monomethylarsonous acid in urine of humans exposed to inorganic arsenic. Chem Res Toxicol 13:693-697.
- Armstrong CW, Stroube RB, Rubio T, Siudyla EA and Miller GB. (1984). Outbreak of fatal arsenic poisoning caused by contaminated drinking water. Arch Environ Health 39:276-279.
- Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ and Weightman AJ. (2006). New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. Appl Environ Microbiol 72:5734-5741.

- Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ and Weightman AJ. (2005). At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Appl Environ Microbiol 71:7724-7736.
- Azcue JM and Nriagu JO (1994). Arsenic: Historical perspectives. Arsenic in the Environment, Part 1: Cycling and Characterization. John Wiley and Sons, Inc.: New York, NY, pp 1-15.
- Bäckhed F, Ding H, Wang T, Hooper LV, Koh GY, Nagy A, Semenkovich CF and Gordon JI. (2004). The gut microbiota as an environmental factor that regulates fat storage. Proc Natl Acad Sci 101:15718-15723.
- Bardullas U, Limón-Pacheco JH, Giordano M, Carrizales L, Mendoza-Trejo MS and Rodríguez VM. (2009). Chronic low-level arsenic exposure causes gender-specific alterations in locomotor activity, dopaminergic systems, and thioredoxin expression in mice. Toxicol Appl Pharmacol 239:169-177.
- Barnard TG, Robertson CA, Jagals P and Potgieter N. (2011). A rapid and low-cost DNA extraction method for isolating *Escherichia coli* DNA from animal stools. Afr J Biotechnol 10:1485-1490.
- Bashir S, Sharma Y, Irshad M, Gupta SD and Dogra TD. (2006). Arsenic-induced cell death in liver and brain of experimental rats. Basic Clin Pharmacol Toxicol 98:38-43.
- Benjamin N, O'Driscoll F, Dougall H, Duncan C, Smith L, Golden M and McKenzie H. (1994). Stomach NO synthesis. Nature 368:502.

- Bentley R andChasteen TG. (2002). Microbial methylation of metalloids: arsenic, antimony, and bismuth. Microbiol Mol Biol Rev 66:250-271.
- Berks BC, Ferguson SJ, Moir JWB and Richardson DJ. (1995). Enzymes and associated electron transport systems that catalyze the respiratory reduction of nitrogen oxides and oxyanions. Biochim Biophys Acta, Bioenerg 1232:97-173.
- Bhattacharjee H, Sheng J, Ajees AA, Mukhopadhyay R and Rosen BP. (2010). Adventitious arsenate reductase activity of the catalytic domain of the human Cdc25B and Cdc25C phosphatases. Biochemistry 49:802-809.
- Bibiloni R, Simon M, Albright C, Sartor B and Tannock G. (2005). Analysis of the large bowel microbiota of colitic mice using PCR/DGGE. Lett Appl Microbiol 41:45-51.
- Bjoerne H, Petersson J, Phillipson M, Weitzberg E, Holm L and Lundberg JO. (2004). Nitrite in saliva increases gastric mucosal blood flow and mucus thickness. J Clin Invest 113:106-114.
- Bjursell MK, Martens EC and Gordon JI. (2006). Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. J Biol Chem 281:36269-36279.
- Blaut M, Collins MD and Taras D (2009). Genus VIII *Dorea*. Bergey's Manual of Systematic Bacteriology: The Firmicutes. Springer: Newyork, pp 941.
- Bobrowicz P, Wysocki R, Owsianik G, Goffeau A and Ulaszewski S. (1997). Isolation of three contiguous genes, ACR1, ACR2 and ACR3, involved in

resistance to arsenic compounds in the yeast *Saccharomyces cerevisiae*. Yeast 13:819-828.

- Boom R, Sol CJ, Salimans MM, Jansen CL, Wertheim-van Dillen PM and van der Noordaa J. (1990). Rapid and simple method for purification of nucleic acids. J Clin Microbiol 28:495-503.
- Borneman, J. and Triplett, E. W. (1997). Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. Appl Environ Microbiol 63:2647-2653.
- Brown RM, Newton D, Pickford CJ and Sherlock JC. (1990). Human Metabolism of Arsenobetaine Ingested with Fish. Hum Exp Toxicol 9:41-46.
- Buchet JP, Geubel A, Pauwels S, Mahieu P and Lauwerys R. (1984). The influence of liver disease on the methylation of arsenite in humans. Arch Toxicol 55:151-154.
- Busenlehner LS, Pennella MA and Giedroc DP. (2003). The SmtB/ArsR family of metalloregulatory transcriptional repressors: structural insights into prokaryotic metal resistance. FEMS Microbiol Rev 27:131-143.
- Caraux G and Pinloche S. (2005). PermutMatrix: a graphical environment to arrange gene expression profiles in optimal linear order. Bioinformatics 21:1280-1281.
- Carbonell-Barrachina AA, Burló F, Burgos-Hernández A, López E and Mataix J. (1997). The influence of arsenite concentration on arsenic accumulation in tomato and bean plants. Sci Hort 71:167-176.

- Cardinale M, Brusetti L, Quatrini P, Borin S, Puglia AM, Rizzi A, Zanardini E, Sorlini C, Corselli C and Daffonchio D. (2004). Comparison of different primer sets for use in Automated Ribosomal Intergenic Spacer Analysis of complex bacterial communities. Appl Environ Microbiol 70:6147-6156.
- Carter DE, Aposhian HV and Gandolfi AJ. (2003). The metabolism of inorganic arsenic oxides, gallium arsenide, and arsine: a toxicochemical review. Toxicol Appl Pharmacol 193:309-334.
- Challenger F. (1945). Biological methylation. Chem Rev 36:315-361.
- Chao A. (1984). Nonparametric estimation of the number of classes in a population. Scand J Stat. Theory and Applications 11:265-270.
- Chaudhuri AN, Basu S, Chattopadhyay S and Das Gupta S. (1999). Effect of high arsenic content in drinking water on rat brain. Indian J Biochem Biophys 36:51-54.
- Cherbuy C, Honvo-Houeto E, Bruneau A, Bridonneau C, Mayeur C, Duée P, Langella P and Thomas M. (2010). Microbiota matures colonic epithelium through a coordinated induction of cell cycle-related proteins in gnotobiotic rat. Am J Physiol Gastrointest Liver Physiol 299:G348-G357.
- Chessel D, Dufour AB and Thioulouse J. (2004). The ade4 package-I- One-table methods. R News 4:5-10.
- Choi W, Chang M, Han J, Hong S and Lee H. (1997). Identification of nitric oxide synthase in *Staphylococcus aureus*. Biochem Biophys Res Commun 237:554-558.

- Choudhry Z, Misbahuddin M, Hossain AKM and Saleh A. (2010). Inhibitory effect of arsenic on aerobic gut flora in rat. Bangladesh Med Res Counc Bull 35:79-83.
- Cohen SM, Arnold LL, Eldan M, Lewis AS and Beck BD. (2006). Methylated arsenicals: the implications of metabolism and carcinogenicity studies in rodents to human risk assessment. Crit Rev Toxicol 36:99-133.
- Cole JA and Brown CM. (1980). Nitrite reduction to ammonia by fermentative bacteria: a short circuit in the biological nitrogen cycle. FEMS Microbiol Lett 7:65-72.
- Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM and Tiedje JM. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucl Acids Res 37:D141-145.
- Corr SC, Hill C and Gahan CGM (2009). Understanding the Mechanisms by Which Probiotics Inhibit Gastrointestinal Pathogens. Advances in Food and Nutrition Research. Academic Press, pp 1-15.
- Correa P and Houghton J. (2007). Carcinogenesis of *Helicobacter pylori*. Gastroenterology 133:659-672.
- Csanaky I and Gregus Z. (2002). Species variations in the biliary and urinary excretion of arsenate, arsenite and their metabolites. Comp Biochem Physiol C 131:355-365.
- Cullen WR and Reimer KJ. (1989). Arsenic speciation in the environment. Chem Rev 89:713-764.

- Culman SW, Gauch HG, Blackwood CB and Thies JE. (2008). Analysis of T-RFLP data using analysis of variance and ordination methods: A comparative study. J Microbiol Methods 75:55-63.
- Culman S, Bukowski R, Gauch H, Cadillo-Quiroz H and Buckley D. (2009). T-REX: software for the processing and analysis of T-RFLP data. BMC Bioinformatics 10:171-180.
- De La Cochetiere MF, Durand T, Lepage P, Bourreille A, Galmiche JP and Dore J. (2005). Resilience of the dominant human fecal microbiota upon short-course antibiotic challenge. J Clin Microbiol 43:5588-5592.
- De VME, Wong JM, Zhou JY, Tzeng E, Wong HR, Billiar TR and Geller DA. (1996). Cytokine-induced nitric oxide synthase gene transcription is blocked by the heat shock response in human liver cells. Surgery 120:144-149.
- DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P and Andersen GL. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069-5072.
- Dethlefsen L, Huse S, Sogin ML and Relman DA. (2008). The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16s rRNA sequencing. PLoS Biol 6:2383-2400.
- Dey S, Dou D, Tisa LS and Rosen BP. (1994). Interaction of the catalytic and the membrane subunits of an oxyanion-translocating ATPase. Arch Biochem Biophys 311:418-424.

- Diaz-Bone RA and Van de Wiele T. (2010). Biotransformation of metal(loid)s by intestinal microorganisms. Pure Appl Chem 82:409-427.
- Dice LR. (1945). Measures of the amount of ecologic association between species. Ecology 26:297-302.
- Doyle MP and Hoekstra JW. (1981). Oxidation of nitrogen oxides by bound dioxygen in hemoproteins. J Inorg Biochem 14:351-358.
- Druwe IL and Vaillancourt RR. (2010). Influence of arsenate and arsenite on signal transduction pathways: an update. Arch Toxicol 84:585-596.
- Duncan SH, Louis P and Flint HJ. (2007). Cultivable bacterial diversity from the human colon. Lett Appl Microbiol 44:343-350.
- Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE and Relman DA. (2005). Diversity of the human intestinal microbial flora. Science 308:1635-1638.
- Edwards U, Rogall T, Blocker H, Emde M and Bottger EC. (1989). Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. Nucleic Acids Res 17:7843-7853.
- Einsle O, Messerschmidt A, Huber R, Kroneck PMH and Neese F. (2002). Mechanism of the six-electron reduction of nitrite to ammonia by cytochrome c nitrite reductase. J Am Chem Soc 124:11737-11745.
- Ellis PJ, Conrads T, Hille R and Kuhn P. (2001). Crystal structure of the 100 kDa arsenite oxidase from *Alcaligenes faecalis* in two crystal forms at 1.64 Å and 2.03 Å. Structure 9:125-132.

- Endo H, Iijima K, Asanuma K, Ara N, Ito H, Asano N, Uno K, Koike T, Imatani A and Shimosegawa T. (2010). Exogenous luminal nitric oxide exposure accelerates columnar transformation of rat esophagus. Int J Cancer 127:2009-2019.
- Environmental Protection Agency (2001). National primary drinking water regulations; arsenic and clarifications to compliance and new source contaminants monitoring; final rule. 66:6976-7066.
- EPA. (2005). Emissions of arsenic compounds. Technology Transfer Network. National Air Toxics Assessment. Pollutant-Specific Data Tables. U.S. Environmental Protection Agency.http://www.epa.gov/ttn/atw/nata/tablemis.html.
- Felsenstein J. (1989). PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 5:164-166.
- Fetissov SO, Hamze Sinno M, Coëffier M, Bole-Feysot C, Ducrotté P, Hökfelt T and Déchelotte P. (2008). Autoantibodies against appetite-regulating peptide hormones and neuropeptides: Putative modulation by gut microflora. Nutrition 24:348-359.
- Finegold SM, Dowd SE, Gontcharova V, Liu C, Henley KE, Wolcott RD, Youn E, Summanen PH, Granpeesheh D, Dixon D, Liu M, Molitoris DR and Green JA. (2010). Pyrosequencing study of fecal microflora of autistic and control children. Anaerobe 16:444-453.
- Fischer SG and Lerman LS. (1983). DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. Proc Natl Acad Sci 80:1579-1583.

- Fisher MM and Triplett EW. (1999). Automated approach for Ribosomal Intergenic Spacer Analysis of microbial diversity and its application to freshwater bacterial communities. Appl Environ Microbiol 65:4630-4636.
- Forsythe P, Sudo N, Dinan T, Taylor VH and Bienenstock J. (2010). Mood and gut feelings. Brain Behav Immun 24:9-16.
- Forsythe P and Bienenstock J (2008). Probiotics in neurology and psychiatry. Ed. Versalovic J and Wilson M. Therapeutic Microbiology: Probiotics and Related Strategies. ASM press: Washington, DC, pp 285-298.
- Frank DN and Pace NR. (2008). Gastrointestinal microbiology enters the metagenomics era. Curr Opin Gastroenterol 24:4-10.
- Frank DN, St. Amand AL, Feldman RA, Boedeker EC, Harpaz N and Pace NR. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci 104:13780-13785.
- Friberg L, Nordberg GF and Vouk VB. (1986). Handbook on the Toxicology of Metals. Specific Metals. Elsevier Science Publisher: Amsterdam, The Netherlands.
- Gadd GM. (1993). Microbial formation and transformation of organometallic and organometalloid compounds. FEMS Microbiol Rev 11:297-316.
- Gakhar L, Bartlett JA, Penterman J, Mizrachi D, Singh PK, Mallampalli RK, Ramaswamy S and McCray J and Paul B. (2010). PLUNC is a novel airway surfactant protein with anti-biofilm activity. PLoS ONE 5:e9098.

- Ghosh M, Shen J and Rosen BP. (1999). Pathways of As(III) detoxification in *Saccharomyces cerevisiae*. Proc Natl Acad Sci 96:5001-5006.
- Gihring TM andBanfield JF. (2001). Arsenite oxidation and arsenate respiration by a new *Thermus* isolate. FEMS Microbiol Lett 204:335-340.
- Giongo A, Gano KA, Crabb DB, Mukherjee N, Novelo LL, Casella G, Drew JC, Ilonen J, Knip M, Hyoty H, Veijola R, Simell T, Simell O, Neu J, Wasserfall CH, Schatz D, Atkinson MA and Triplett EW. (2011). Toward defining the autoimmune microbiome for type 1 diabetes. ISME J 5:82-91.
- Gladysheva TB, Oden KL and Rosen BP. (1994). Properties of the arsenate reductase of plasmid R773. Biochemistry 33:7288-93.
- Good IJ. (1953). The Population Frequencies of Species and the Estimation of Population Parameters. Biometrika 40: 237-264.
- Gordan HA, Wostmann BS and Bruckner-Kardoss E. (1963). Effects of microbial flora on cardiac output and other elements of blood circulation. Proc Soc Exp Biol Med 114:301-304.
- Gordon JH and Dubos R. (1970). The anaerobic bacterial flora of the mouse cecum. J Exp Med 132:251-260.
- Gotelli NJ and Colwell RK. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecol Lett 4:379-391.
- Gouy M, Guindon S and Gascuel O. (2010). SeaView Version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27:221-224.

- Gregus Z and Németi B. (2007). Glutathione-dependent reduction of arsenate by glycogen phosphorylase—responsiveness to endogenous and xenobiotic inhibitors. Toxicol Sci 100:44-53.
- Gregus Z and Németi B. (2005). The glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase works as an arsenate reductase in human red blood cells and rat liver cytosol. Toxicol Sci 85:859-869.
- Gurr J, Yih L, Samikkannu T, Bau D, Lin S and Jan K. (2003). Nitric oxide production by arsenite. Mutat Res Fund Mol Mech Mut 533:173-182.
- Gusarov I, Starodubtseva M, Wang Z, McQuade L, Lippard SJ, Stuehr DJ and Nudler E. (2008). Bacterial nitric-oxide synthases operate without a dedicated redox partner. J Biol Chem 283:13140-13147.
- Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E, Methé B, DeSantis TZ, The Human Microbiome Consortium, Petrosino JF, Knight R and Birren BW. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res 21:494-504.
- Hall LL, George SE, Kohan MJ, Styblo M and Thomas DJ. (1997). In vitro methylation of inorganic arsenic in mouse intestinal cecum. Toxicol Appl Pharmacol 147:101-109.
- Hammes WP and Hertel C (2009). Genus I *Lactobacillus*. Bergey's Manual of Systematic Bacteriology: The Firmicutes. Springer: Newyork, pp 465-511.
- Hand T and Belkaid Y. (2010). Microbial control of regulatory and effector T cell responses in the gut. Curr Opin Immunol 22:63-72.

- Harold FM and Baarda JR. (1966). Interaction of arsenate with phosphate-transport systems in wild-type and mutant *Streptococcus faecalis*. J Bacteriol 91:2257-62.
- Harrington CF, Brima EI and Jenkins RO. (2008). Biotransformation of arsenobetaine by microorganisms from the human gastrointestinal tract. Chem Spec Bioavail 20:173-180.
- Hayakawa T, Kobayashi Y, Cui X and Hirano S. (2005). A new metabolic pathway of arsenite: arsenic-glutathione complexes are substrates for human arsenic methyltransferase Cyt19. Arch Toxicol 79:183-191.
- Hays AM, Lantz RC, Rodgers LS, Sollome JJ, Vaillancourt RR, Andrew AS, Hamilton JW and Camenisch TD. (2008). Arsenic-induced decreases in the vascular matrix. Toxicol Pathol 36:805-817.
- HazDat. (2006). Arsenic. ATSDR's Hazardous substance release and health effects database. Agency for Toxic Substances and Disease Registry. http://www.atsdr.cdc.gov/hazdat.html.
- Hecker KH and Roux KH. (1996). High and low annealing temperatures increase both specificity and yield in touchdown and stepdown PCR. BioTechniques 20:478-485.
- Heijtz RD, Wang S, Anuar F, Qian Y, Björkholm B, Samuelsson A, Hibberd ML, Forssberg H and Pettersson S. (2011). Normal gut microbiota modulates brain development and behavior. Proc Natl Acad Sci 108:3047-3052.
- Herbel MJ, Switzer Blum J, Hoeft SE, Cohen SM, Arnold LL, Lisak J, Stolz JF and Oremland RS. (2002). Dissimilatory arsenate reductase activity and arsenate-

respiring bacteria in bovine rumen fluid, hamster feces, and the termite hindgut. FEMS Microbiol Ecol 41:59-67.

- Hill MJ (1995). The normal gut bacterial flora. Ed. M. J. Hill. Role of Gut Bacteria in Human Toxicology and Pharmacology. Taylor and Francis Inc.: London, WC, pp 3-17.
- Hoeft SE, Kulp TR, Stolz JF, Hollibaugh JT and Oremland RS. (2004). Dissimilatory arsenate reduction with sulfide as electron donor: Experiments with Mono Lake water and isolation of strain MLMS-1, a chemoautotrophic arsenate respirer. Appl Environ Microbiol 70:2741-2747.
- Hoffmann C, Hill DA, Minkah N, Kirn T, Troy A, Artis D and Bushman F. (2009). Community-wide response of the gut microbiota to enteropathogenic *Citrobacter rodentium* infection revealed by deep sequencing. Infect Immun 77:4668-4678.
- Hooper LV, Midtvedt T and Gordon JI. (2002). How host-microbial interactions shape the nutrient environment of the mammalian intestine. Annu Rev Nutr 22:283-307.
- Huang R and Lee T. (1996). Cellular uptake of trivalent arsenite and pentavalent arsenate in κB cells cultured in phosphate-free medium. Toxicol Appl Pharmacol 136:243-249.
- Huang Z, Shiva S, Kim-Shapiro DB, Patel RP, Ringwood LA, Irby CE, Huang KT, Ho C, Hogg N, Schechter AN and Gladwin MT. (2005). Enzymatic function of hemoglobin as a nitrite reductase that produces NO under allosteric control. J Clin Invest 115:2099-2107.

- Hughes MF. (2002). Arsenic toxicity and potential mechanisms of action. Toxicol Lett 133:1-16.
- Hwang E and Kim G. (2007). Biomarkers for oxidative stress status of DNA, lipids, and proteins in vitro and in vivo cancer research. Toxicology 229:1-10.
- IARC. (1987).Arsenic. International Agency for Research on Cancer.
- Jackson C and Dugas S. (2003). Phylogenetic analysis of bacterial and archaeal *arsC* gene sequences suggests an ancient, common origin for arsenate reductase. BMC Evol Biol 3:18.
- Jernberg C, Lofmark S, Edlund C and Jansson JK. (2007). Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. ISME J 1:56-66.
- Ji G, Garber EAE, Armes LG, Chen C, Fuchs JA and Silver S. (1994). Arsenate reductase of *Staphylococcus aureus* plasmid pI258. Biochemistry 33:7294-7299.
- Ji G andSilver S. (1992). Reduction of arsenate to arsenite by the ArsC protein of the arsenic resistance operon of *Staphylococcus aureus* plasmid pI258. Proc Natl Acad Sci 89:9474-9478.
- Joan Vermeiren, Tom Van de Wiele, Willy Verstraete, Pascal Boeckx and Nico Boon. (2009). Nitric oxide production by the human intestinal microbiota by dissimilatory nitrate reduction to ammonium. J Biomed Biotechnol 2009: 284718.
- Johansson MEV, Holmén Larsson JM and Hansson GC. (2011). The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host–microbial interactions. Proc Natl Acad Sci 108:4659-4665.

- Johnson EG, Sparks JP, Dzikovski B, Crane BR, Gibson DM and Loria R. (2008a). Plant-pathogenic *Streptomyces* species produce nitric oxide synthase-derived nitric oxide in response to host signals. Chem Biol 15:43-50.
- Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S and Madden TL. (2008b). NCBI BLAST: a better web interface. Nucleic Acids Res 36:W5-W9.
- Juhasz AL, Smith E, Weber J, Rees M, Rofe A, Kuchel T, Sansom L and Naidu R. (2006). In vivo assessment of arsenic bioavailability in rice and its significance for human health risk assessment. Environ Health Perspect 114:1826-1831.
- Jukes TH and Cantor CR (1969). Evolution of protein molecules Ed. H. N. Munro. Mammalian protein metabolism. Academic Press, New York., pp 21-32.
- Kalliomäki M, Carmen Collado M, Salminen S and Isolauri E. (2008). Early differences in fecal microbiota composition in children may predict overweight. Am J Clin Nutr 87:534-538.
- Kalliomäki M, Salminen S, Poussa T, Arvilommi H and Isolauri E. (2003). Probiotics and prevention of atopic disease: 4-year follow-up of a randomised placebo-controlled trial. The Lancet 361:1869-1871.
- Kamat CD, Green DE, Curilla S, Warnke L, Hamilton JW, Sturup S, Clark C and Ihnat MA. (2005). Role of HIF signaling on tumorigenesis in response to chronic low-dose arsenic administration. Toxicol Sci 86:248-257.
- Kim M, Park J, Mo J, Ann E, Han S, Baek S, Kim K, Im S, Park J, Choi E and Park H. (2008). Downregulation by lipopolysaccharide of Notch signaling, via nitric oxide. J Cell Sci 121:1466-1476.

- Kitchin KT and Wallace K. (2008). The role of protein binding of trivalent arsenicals in arsenic carcinogenesis and toxicity. J Inorg Biochem 102:532-539.
- Kovacs A, Yacoby K and Gophna U. (2010). A systematic assessment of Automated Ribosomal Intergenic Spacer Analysis (ARISA) as a tool for estimating bacterial richness. Res Microbiol 161:192-197.
- Kozul CD, Ely KH, Enelow RI and Hamilton JW. (2009a). Low-dose arsenic compromises the immune response to influenza A infection in vivo. Environ Health Perspect 117:1441-1447.
- Kozul CD, Hampton TH, Davey JC, Gosse JA, Nomikos AP, Eisenhauer PL, Weiss DJ, Thorpe JE, Ihnat MA and Hamilton JW. (2009b). Chronic exposure to arsenic in the drinking water alters the expression of immune response genes in mouse lung. Environ Health Perspect 117:1108-1115.
- Krautler B. (1990). Chemistry of methylcorrinoids related to their roles in bacterial C1 metabolism. FEMS Microbiol Rev 7:349-354.
- Krieg NR, Ludwig W, Euzéby J and Whitman WB (2010). Phylum XIV. Bacteroidetes phyl. nov. Bergey's Manual of Systematic Bacteriology. Springer: New York, pp 25.
- Kuroda K, Yoshida K, Yoshimura M, Endo Y, Wanibuchi H, Fukushima S and Endo G. (2004). Microbial metabolite of dimethylarsinic acid is highly toxic and genotoxic. Toxicol Appl Pharmacol 198:345-353.
- Lantz RC, Brandon J. Lynch, Boitano S, Poplin GS, Littau S, Tsaprailis G and Burgess JL. (2007). Pulmonary biomarkers based on alterations in protein expression after exposure to arsenic. Environ Health Perspect 115:586-591.

- Laparra JM, Vélez D, Barberá R, Granero L, Polache A, Montoro R and Farré R. (2006). Cytotoxic effect of As(III) in Caco-2 cells and evaluation of its human intestinal permeability. Toxicol in Vitro 20:658-663.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG. (2007). Clustal W and Clustal X version 2.0. Bioinformatics 23:2947-2948.
- Larsen N, Vogensen FK, van den Berg, Frans W. J., Nielsen DS, Andreasen AS, Pedersen BK, Al-Soud WA, Sørensen SJ, Hansen LH and Jakobsen M. (2010). Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. PLoS ONE 5:e9085.
- Lasky T, Sun W, Kadry A and Hoffman MK. (2003). Mean total arsenic concentrations in chicken 1989-2000 and estimated exposures for consumers of chicken. Environ Health Perspect 112:18-21.
- Lau BP-, Michalik P and Porter CJ. (1987). Identification and confirmation of arsenobetaine and arsenocholine in fish, lobster and shrimp by a combination of fast atom bombardment and tandem mass spectrometry. Biol Mass Spectrom 14:723-732.
- Le XC, Lu X, Ma M, Cullen WR, Aposhian HV and Zheng B. (2000). Speciation of key arsenic metabolic intermediates in human urine. Anal Chem 72:5172-5177.
- Leonard A (1991). Arsenic. Metals and their compunds in the environment. Ed. E. Merian. Metals and their compunds in the environment. Occurance, Analysis and Biological Relevance. VCH: NewYork NY, pp 751-774.

- Leslie EM, Haimeur A and Waalkes MP. (2004). Arsenic transport by the human Multidrug Resistance Protein 1 (MRP1/ABCC1). J Biol Chem 279:32700-32708.
- Levin-Scherz J, Patrick JD, Weber FH, Garabedian C (1987). Acute arsenic ingestion. Ann Emerg Med 16:702-704.
- Ley RE, Turnbaugh PJ, Klein S and Gordon JI. (2006). Microbial ecology: human gut microbes associated with obesity. Nature 444:1022-1023.
- Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD and Gordon JI. (2005). Obesity alters gut microbial ecology. Proc Natl Acad Sci 102:11070-11075.
- Li M, Wang B, Zhang M, Rantalainen M, Wang S, Zhou H, Zhang Y, Shen J, Pang X, Zhang M, Wei H, Chen Y, Lu H, Zuo J, Su M, Qiu Y, Jia W, Xiao C, Smith L, Yang S, Holmes E, Tang H, Zhao G, Nicholson J, Li L and Zhao L. (2008). Symbiotic gut microbes modulate human metabolic phenotypes. Proc Natl Acad Sci 105:2117-2122.
- Lin S, Shi Q, Nix FB, Styblo M, Beck MA, Herbin-Davis KM, Hall LL, Simeonsson JB and Thomas DJ. (2002). A novel S-adenosyl-L-methionine: Arsenic(III) methyltransferase from rat liver cytosol. J Biol Chem 277:10795-10803.
- Lin S, Cullen WR and Thomas DJ. (1999). Methylarsenicals and arsinothiols are potent inhibitors of mouse liver thioredoxin reductase. Chem Res Toxicol 12:924-930.
- Lin Y, Yang J and Rosen BP. (2007). ArsD: an As(III) metallochaperone for the ArsAB As(III)-translocating ATPase. J Bioenerg Biomembr 39:453-458.

- Lindgren A, Vahter M and Dencker L. (1982). Autoradiographic studies on the distribution of arsenic in mice and hamster administered 74As-arsenite or -arsenate. Acta Pharmacol Toxicol 51:253-266.
- Liu A, Garcia-Dominguez E, Rhine ED and Young LY. (2004). A novel arsenate respiring isolate that can utilize aromatic substrates. FEMS Microbiol Ecol 48:323-332.
- Liu S, Davidson MM, Tang X, Walker WF, Athar M, Ivanov V and Hei TK. (2005). Mitochondrial damage mediates genotoxicity of arsenic in mammalian cells. Cancer Res 65:3236-3242.
- Liu Z, Shen J, Carbrey JM, Mukhopadhyay R, Agre P and Rosen BP. (2002). Arsenite transport by mammalian aquaglyceroporins AQP7 and AQP9. Proc Natl Acad Sci 99:6053-6058.
- Liu Z, DeSantis TZ, Andersen GL and Knight R. (2008). Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. Nucleic Acids Res 36:e120-e120.
- Loffredo CA, Aposhian HV, Cebrian ME, Yamauchi H and Silbergeld EK. (2003). Variability in human metabolism of arsenic. Environ Res 92:85-91.
- Lowney YW, Wester RC, Schoof RA, Cushing CA, Edwards M and Ruby MV. (2007). Dermal absorption of arsenic from soils as measured in the rhesus monkey. Toxicol Sci 100:381-392.
- Ludwig W, Schleifer K and Whitman WB (2009). Revised road map to phylum Firmicutes. Bergey's Manual of Systematic Bacteriology: The Firmicutes. Springer: Newyork, pp 1-17.

- Lundberg JO and Weitzberg E. (2005). NO generation from nitrite and its role in vascular control. Arterioscler Thromb Vasc Biol 25:915-922.
- Lundberg JO and Govoni M. (2004). Inorganic nitrate is a possible source for systemic generation of nitric oxide. Free Radic Biol Med 37:395-400.
- Lundberg JO, Weitzberg E, Lundberg JM and Alving K. (1994). Intragastric nitric oxide production in humans: measurements in expelled air. Gut 35:1543-1546.
- Lynn S, Gurr J, Lai H and Jan K. (2000). NADH oxidase activation is involved in arsenite-induced oxidative dna damage in human vascular smooth muscle cells. Circ Res 86:514-519.
- MacFabe DF, Cain DP, Rodriguez-Capote K, Franklin AE, Hoffman JE, Boon F, Taylor AR, Kavaliers M and Ossenkopp K. (2007). Neurobiological effects of intraventricular propionic acid in rats: Possible role of short chain fatty acids on the pathogenesis and characteristics of autism spectrum disorders. Behav Brain Res 176:149-169.
- MacIntosh DL, Williams PL, Hunter DJ, Sampson LA, Morris SC, Willett WC and Rimm EB. (1997). Evaluation of a food frequency questionnaire-food composition approach for estimating dietary intake of inorganic arsenic and methylmercury. Cancer Epidemiol Biomarkers Prevent 6:1043-1050.
- Macpherson AJ and Harris NL. (2004). Interactions between commensal intestinal bacteria and the immune system. Nat Rev Immunol 4:478-485.
- Macur RE, Jackson CR, Botero LM, Mcdermott TR and Inskeep WP. (2004). Bacterial populations associated with the oxidation and reduction of arsenic in an unsaturated soil. Environ Sci Technol 38:104-211.

- Magurran AE. (1988). Ecological diversity and its measurement. Princeton University Press.
- Mahowald MA, Rey FE, Seedorf H, Turnbaugh PJ, Fulton RS, Wollam A, Shah N, Wang C, Magrini V, Wilson RK, Cantarel BL, Coutinho PM, Henrissat B, Crock LW, Russell A, Verberkmoes NC, Hettich RL and Gordon JI. (2009). Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. Proc Natl Acad Sci 106:5859-5864.
- Mandal BK and Suzuki KT. (2002). Arsenic round the world: a review. Talanta 58:201-235.
- Marafante E, Vahter M and Envall J. (1985). The role of the methylation in the detoxication of arsenate in the rabbit. Chem Biol Interact 56:225-238.
- Mariat D, Firmesse O, Levenez F, Guimaraes VD, Sokol H, Dore J, Corthier G and Furet J. (2009). The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age. BMC Microbiol 9:123.
- Matschullat J. (2000). Arsenic in the geosphere -- a review. Sci Total Environ 249:297-312.
- McBride BC and Wolfe RS. (1971). Biosynthesis of dimethylarsine by *Methanobacterium*. Biochemistry 10:4312-4317.
- McCracken VJ and Lorenz RG. (2001). The gastrointestinal ecosystem: a precarious alliance among epithelium, immunity and microbiota. Cell Microbiol 3:1-11.

- McCracken VJ, Simpson JM, Mackie RI and Gaskins HR. (2001). Molecular ecological analysis of dietary and antibiotic-induced alterations of the mouse intestinal microbiota. J Nutr 131:1862-1870.
- McFarland LV. (2008). Antibiotic-associated diarrhea: Epidemiology, trends and treatment. Future Microbiol 3:563-578.
- Mealey J, Brownell GL and Sweet WH. (1959). Radioarsenic in plasma, urine, normal tissues, and intracranial neoplasms: distribution and turnover after intravenous injection in man. Arch Neurol Psychiatry 81:310-320.
- Medellin-Pena M and Griffiths MW. (2009). Effect of molecules secreted by *Lactobacillus acidophilus* Strain La-5 on *Escherichia coli* O157:H7 colonization. Appl Environ Microbiol 75:1165-1172.
- Meng Y, Liu Z and Rosen BP. (2004). As(III) and Sb(III) Uptake by GlpF and Efflux by ArsB in *Escherichia coli*. J Biol Chem 279:18334-18341.
- Moncada S. (1999). Nitric oxide: discovery and impact on clinical medicine. J R Soc Med 92:164-169.
- Moore LE, Smith AH, Hopenhayn-Rich C, Biggs ML, Kalman DA and Smith MT. (1997). Micronuclei in exfoliated bladder cells among individuals chronically exposed to arsenic in drinking water. Cancer Epidemiol Biomarkers Prevent 6:31-36.
- Mukhopadhyay R, Rosen BP, Phung LT and Silver S. (2002). Microbial arsenic: from geocycles to genes and enzymes. FEMS Microbiol Rev 26:311-325.

- Muller D, Lievremont D, Simeonova DD, Hubert J and Lett M. (2003). Arsenite oxidase *aox* genes from a metal-resistant β-proteobacterium. J Bacteriol 185:135-141.
- Muyzer G, Waal ECd and Uitterlinden AG. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl Environ Microbiol 59:695-700.
- Muyzer G and Smalla K. (1998). Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. Antonie Van Leeuwenhoek 73:127-141.
- Naranmandura H, Suzuki N, Iwata K, Hirano S and Suzuki T. (2007). Arsenic metabolism and thioarsenicals in hamsters and rats. Chem Res Toxicol 20:616-624.
- National Research Council. (2001). Arsenic in Drinking Water: 2001 Update. National academy press: Washington DC.
- Nava GM, Friedrichsen HJ and Stappenbeck TS. (2011). Spatial organization of intestinal microbiota in the mouse ascending colon. ISME J 5:627-638.
- Navas-Acien A, Silbergeld EK, Pastor-Barriuso R and Guallar E. (2008). Arsenic exposure and prevalence of type 2 diabetes in US adults. J Am Med Assoc 300:814-822.
- Ni D, Orla M. and Sadler PJ. (1991). The structure and reactivity of arsenic compounds: biological activity and drug design. Struct Bonding 78:129-190.

- Nicholson JK, Holmes E and Wilson ID. (2005). Gut microorganisms, mammalian metabolism and personalized health care. Nat Rev Micro 3:431-438.
- Niggemyer A, Spring S, Stackebrandt E and Rosenzweig RF. (2001). Isolation and characterization of a novel As(V)-reducing bacterium: implications for arsenic mobilization and the genus *Desulfitobacterium.* Appl Environ Microbiol 67:5568-5580.
- NRC. (1999).Arsenic in drinking water. National Research Council. National Academy Press: Washington D.C.
- NRC. (1997).Building a Foundation for Sound Environmental Decisions. National Research Council. National Academy Press: Washington, DC.
- Nriagu JO and Pacyna JM. (1988). Quantitative assessment of worldwide contamination of air, water and soils by trace metals. Nature 333:134-139.
- Nriagu JO (2002). Arsenic poisoning through the ages. Environ Chem Arsenic: 1-26.
- O'Hara AM and Shanahan F. (2006). The gut flora as a forgotten organ. EMBO Rep 7:688-693.
- Oremland RS and Stolz JF. (2005). Arsenic, microbes and contaminated aquifers. Trends Microbiol 13:45-49.
- Oremland RS, Stolz JF and Hollibaugh JT. (2004). The microbial arsenic cycle in Mono Lake, California. FEMS Microbiol Ecol 48:15-27.
- Oremland RS and Stolz JF. (2003). The ecology of arsenic. Science 300:939-944.
- Oremland RS, Hoeft SE, Santini JM, Bano N, Hollibaugh RA and Hollibaugh JT. (2002). Anaerobic oxidation of arsenite in Mono Lake water and by a facultative,

arsenite-oxidizing chemoautotroph, strain MLHE-1. Appl Environ Microbiol 68:4795-4802.

- Osborn AM, Moore ERB and Timmis KN. (2000). An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. Environ Microbiol 2:39-50.
- Pacyna JM (1987). Atmospheric emissions of arsenic, cadmium, lead and mercury from high temperature processes in power generation and industry. Lead, Mercury, Cadmium and Arsenic in the environment. John Wiley & Sons Ltd., pp 69-87.
- Palmer C, Bik EM, DiGiulio DB, Relman DA and Brown PO. (2007). Development of the human infant intestinal microbiota. PLoS Biology 5:1556-1573.
- Palmer RMJ, Ashton DS and Moncada S. (1988). Vascular endothelial cells synthesize nitric oxide from L-arginine. Nature 333:664-666.
- Parham NJ and Gibson GR. (2000). Microbes involved in dissimilatory nitrate reduction in the human large intestine. FEMS Microbiol Ecol 31:21-28.
- Pearson K. (1901). On lines and planes of closest fit to systems of points in space. Philosophical Magazine 2:559-572.
- Penders J, Thijs C, van den Brandt P,A., Kummeling I, Snijders B, Stelma F, Adams H, van Ree R and Stobberingh EE. (2007). Gut microbiota composition and development of atopic manifestations in infancy: the KOALA Birth Cohort Study. Gut 56:661-667.

- Penders J, Thijs C, Vink C, Stelma FF, Snijders B, Kummeling I, van dB and Stobberingh EE. (2006). Factors influencing the composition of the intestinal microbiota in early infancy. Pediatrics 118:511-521.
- Petersson J, Phillipson M, Jansson EÅ, Patzak A, Lundberg JO and Holm L. (2007). Dietary nitrate increases gastric mucosal blood flow and mucosal defense. Am J Physiol Gastrointest Liver Physiol 292:G718-G724.
- Petrick JS, Jagadish B, Mash EA and Aposhian HV. (2001). Monomethylarsonous Acid (MMAIII) and arsenite: LD50 in hamsters and in vitro inhibition of pyruvate dehydrogenase. Chem Res Toxicol 14:651-656.
- Poock SR, Leach ER, Moir JWB, Cole JA and Richardson DJ. (2002). Respiratory detoxification of nitric oxide by the cytochrome c nitrite reductase of *Escherichia coli*. J Biol Chem 277:23664-23669.
- Potter L, Angove H, Richardson D and Cole J. (2001). Nitrate reduction in the periplasm of Gram-negative bacteria. Adv Microb Physiol 45:51-112.
- Potter LC, Millington P, Griffiths L, Thomas GH and Cole JA. (1999). Competition between *Escherichia coli* strains expressing either a periplasmic or a membrane-bound nitrate reductase: does Nap confer a selective advantage during nitrate-limited growth? Biochem J 344:77-84.
- Preidis GA and Versalovic J (2009). Targeting the human microbiome with antibiotics, probiotics, and prebiotics: gastroenterology enters the metagenomics era. Gastroenterology 136:2015-2031.

- Pryor WA and Squadrito GL. (1995). The chemistry of peroxynitrite: a product from the reaction of nitric oxide with superoxide. Am J Phys - Lung Cell and Molecular Physiol 268:L699-L722.
- Pysher MD, Sollome JJ, Regan S, Cardinal TR, Hoying JB, Brooks HL and Vaillancourt RR. (2007). Increased hexokinase II expression in the renal glomerulus of mice in response to arsenic. Toxicol Appl Pharmacol 224:39-48.
- Qin J, Li R, Raes J, Arumugam M, Burgdorf K, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende D, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto J, Hansen T, Paslier D, Linneberg A, Nielsen H, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Doré J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, Bork P, Ehrlich S and Wang J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464:59-65.
- Qin J, Rosen BP, Zhang Y, Wang G, Franke S and Rensing C. (2006). Arsenic detoxification and evolution of trimethylarsine gas by a microbial arsenite S-adenosylmethionine methyltransferase. Proc Natl Acad Sci 103:2075-2080.
- Radabaugh TR, Sampayo-Reyes A, Zakharyan RA and Aposhian HV. (2002). Arsenate reductase II. Purine nucleoside phosphorylase in the presence of dihydrolipoic acid is a route for reduction of arsenate to arsenite in mammalian systems. Chem Res Toxicol 15:692-698.
- Rainey FA, Hollen BJ and Small A (2009). Genus I *Clostridia*. Bergey's Manual of Systematic Bacteriology: The Firmicutes. Springer: Newyork, pp 738-828.

- Rakoff-Nahoum S, Paglino J, Eslami-Varzaneh F, Edberg S and Medzhitov R (2004). Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. 118:229-241.
- Ramírez-Solís A, Mukopadhyay R, Rosen BP and Stemmler TL. (2004). Experimental and theoretical characterization of arsenite in water: Insights into the coordination environment of As−O. Inorg Chem 43:2954-2959.
- Rawls JF, Mahowald MA, Ley RE and Gordon JI (2006). Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. Cell 127:423-433.
- Ren X, McHale CM, Skibola CF, Smith AH, Smith MT and Zhang L. (2010). An emerging role for epigenetic dysregulation in arsenic toxicity and carcinogenesis. Environ Health Perspect 119:11-19.
- Rhee SH, Pothoulakis C and Mayer EA. (2009). Principles and clinical implications of the brain-gut-enteric microbiota axis. Nat Rev Gastroenterol Hepatol 6:306-314.
- Rhine ED, Phelps CD and Young LY. (2006). Anaerobic arsenite oxidation by novel denitrifying isolates. Environ Microbiol 8:899-908.
- Richey C, Chovanec P, Hoeft SE, Oremland RS, Basu P and Stolz JF. (2009). Respiratory arsenate reductase as a bidirectional enzyme. Biochem Biophys Res Commun 382:298-302.
- Rosen BP. (1999). Families of arsenic transporters. Trends Microbiol 7:207-212.
- Rosen BP and Liu Z. (2009). Transport pathways for arsenic and selenium: A minireview. Environ Int 35:512-515.

- Rowland IR and Davies MJ. (1981). In vitro metabolism of inorganic arsenic by the gastrointestinal microflora of the rat. J Appl Toxicol 1:278-283.
- Rowland IR and Davies MJ. (1982). Reduction and methylation of sodium arsenate in the rat. J Appl Toxicol 2:294-299.
- Rozen S and Skaletsky H. (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365-386.
- Salmassi TM, Venkateswaren K, Satomi M, Nealson KH, Newman DK and Hering JG. (2002). Oxidation of arsenite by *Agrobacterium albertimagni*, AOL15, sp. nov., isolated from Hot Creek, California. Geomicrobiol J 19:53-66.
- Saltikov CW, Wildman RA, Jr. and Newman DK. (2005). Expression dynamics of arsenic respiration and detoxification in *Shewanella* sp. strain ANA-3. J Bacteriol 187:7390-7396.
- Saltikov CW and Newman DK. (2003). Genetic identification of a respiratory arsenate reductase. Proc Natl Acad Sci 100:10983-10988.
- Salyers AA, Gupta A and Wang Y. (2004). Human intestinal bacteria as reservoirs for antibiotic resistance genes. Trends Microbiol 12:412-416.
- Samikkannu T, Chen CH, Yih LH, Wang AS, Lin SY, Chen TC and Jan KY. (2003). Reactive oxygen species are involved in arsenic trioxide inhibition of pyruvate dehydrogenase activity. Chem Res Toxicol 16:409-414.
- Santini JM andvanden H and Rachel N. (2004). Molybdenum-containing arsenite oxidase of the chemolithoautotrophic arsenite oxidizer NT-26. J Bacteriol 186:1614-1619.

- Santini JM, Sly LI, Wen A, Comrie D, De W, Pascal and Macy JM. (2002). New arsenite-oxidizing bacteria isolated from Australian gold mining environments-phylogenetic relationships. Geomicrobiol J 19:67-76.
- Savage DC. (1977). Microbial ecology of the gastrointestinal tract. Annu Rev Microbiol 31:107-133.
- Scanlan PD, Shanahan F, Clune Y, Collins JK, O'Sullivan GC, O'Riordan M, Holmes E, Wang Y and Marchesi JR. (2008). Culture-independent analysis of the gut microbiota in colorectal cancer and polyposis. Environ Microbiol 10:789-798.
- Schleifer K (2009). Phylum XIII. Firmicutes. Bergey's Manual of Systematic Bacteriology: The Firmicutes. Springer: Newyork, pp 19.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ and Weber CF. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537-7541.
- Schoof RA, Yost LJ, Eickhoff J, Crecelius EA, Cragin DW, Meacher DM and Menzel DB. (1999). A Market Basket Survey of Inorganic Arsenic in Food. Food Chem Toxicol 37:839-846.
- Sekirov I, Russell SL, Antunes LCM and Finlay BB. (2010). Gut Microbiota in Health and Disease. Physiol Rev 90:859-904.
- Shah HN, Chattaway MA, Rajakurana LandGharbia SE (2010). Family V. Prevotellaceae. Bergey's Manual of Systematic Bacteriology Springer: Newyork, pp 86-102.

- Shen XJ, Rawls JF, Randall T, Burcal L, Mpande CN, Jenkins N, Jovov B, Abdo Z, Sandler RS and Keku TO. (2010). Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas. Gut Microbes 1:138-147.
- Shi H, Shi X and Liu KJ. (2004). Oxidative mechanism of arsenic toxicity and carcinogenesis. Mol Cell Biochem 255:67-78.
- Shi W, Dong J, Scott RA, Ksenzenko MY andRosen BP. (1996). The role of arsenic-thiol interactions in metalloregulation of the *ars* operon. J Biol Chem 271:9291-9297.
- Sidhu H, Allison Mj, Chow JOM, Clark A and Peck AB (2001). Rapid reversal of hyperoxaluria in a rat model after probiotic administration of *Oxalobacter formigenes*. J Urol 166:1487-1491.
- Silver S and Phung LT. (2005). Genes and enzymes involved in bacterial oxidation and reduction of inorganic arsenic. Appl Environ Microbiol 71:599-608.
- Simpson EH. (1949). Measurement of diversity. Nature 163:688.
- Smedley PL and Kinniburgh DG. (2002). A review of the source, behaviour and distribution of arsenic in natural waters. Appl Geochem 17:517-568.
- Smith KR, Klei LR and Barchowsky A. (2001). Arsenite stimulates plasma membrane NADPH oxidase in vascular endothelial cells. Am J Physiol 280:L442-L449.

- Sobko T, Huang L, Midtvedt T, Norin E, Gustafsson LE, Norman M, Jansson EÅ and Lundberg JO. (2006). Generation of NO by probiotic bacteria in the gastrointestinal tract. Free Radic Biol Med 41:985-991.
- Sobko T, Reinders CI, Jansson EÅ, Norin E, Midtvedt T andLundberg JO. (2005). Gastrointestinal bacteria generate nitric oxide from nitrate and nitrite. Nitric Oxide - Biology and Chemistry 13:272-278.
- Sobko T, Reinders C, Norin E, Midtvedt T, Gustafsson LE and Lundberg JO. (2004). Gastrointestinal nitric oxide generation in germ-free and conventional rats. Am J Physiol - Gastrointestinal and Liver Physiology 287:G993-G997.
- Sokal, R. and Michener, C. (1958). A statistical method for evaluating systematic relationships. Univ Kans Sci Bull 38:1409-1438.
- Song Y, Liu CandFinegold SM (2010). Genus I. *Bacteroides*. Bergey's Manual of Systematic Bacteriology. Springer: NewYork, pp 27-40.
- Soucy NV, Ihnat MA, Kamat CD, Hess L, Post MJ, Klei LR, Clark C and Barchowsky A. (2003). Arsenic stimulates angiogenesis and tumorigenesis in vivo. Toxicol Sci 76:271-279.
- Souza K, Maddock DA, Zhang Q, Chen J, Chiu C, Mehta S and Wan Y. (2001). Arsenite activation of PI3K/AKT cell survival pathway is mediated by p38 in cultured human keratinocytes. Mol Med 7:767-772.
- Spencer MD, Hamp TJ, Reid RW, Fischer LM, Zeisel SH, Fodor AA (2011). Association between composition of the human gastrointestinal microbiome and development of fatty liver with choline deficiency. Gastroenterology 140:976-986.

- Spiegelhalder B, Eisenbrand G and Preussmann R. (1976). Influence of dietary nitrate on nitrite content of human saliva: possible relevance to in vivo formation of N-nitroso compounds. Food Cosmet Toxicol 14:545-8.
- Stackebrandt E and Liesack W (1993). Nucleic acids and classification. Ed. M. Goodfellow and A. G. O'Donnell. Handbook of New Bacterial Systematics. Academic Press: London, England, pp 152-189.
- Stanton TB, Duncan SH and Flint HJ (2009). Genus XVI *Roseburia*. Bergey's Manual of Systematic Bacteriology: The Firmicutes. Springer: Newyork, pp 954-955.
- Stolz JF and Oremland RS. (1999). Bacterial respiration of arsenic and selenium. FEMS Microbiol Rev 23:615-627.
- Stolz JF and Basu P. (2002). Evolution of nitrate reductase: molecular and structural variations on a common function. Chem BioChem 3:198-206.
- Straub AC, Klei LR, Stolz DB and Barchowsky A. (2009). Arsenic requires sphingosine-1-phosphate type 1 receptors to induce angiogenic genes and endothelial cell remodeling. Am J Pathol 174:1949-1958.
- Straub AC, Clark KA, Ross MA, Chandra AG, Li S, Gao X, Pagano PJ, Stolz DB and Barchowsky A. (2008). Arsenic-stimulated liver sinusoidal capillarization in mice requires NADPH oxidase-generated superoxide. J Clin Invest 118:3980-3989.
- Straub AC, Stolz DB, Ross MA, Hernandez-Zavala A, Soucy NV, Klei LR and Barchowsky A. (2007a). Arsenic stimulates sinusoidal endothelial cell capillarization and vessel remodeling in mouse liver. Hepatology 45:205-212.

- Straub AC, Stolz DB, Vin H, Ross MA, Soucy NV, Klei LR and Barchowsky A. (2007b). Low level arsenic promotes progressive inflammatory angiogenesis and liver blood vessel remodeling in mice. Toxicol Appl Pharmacol 222:327-336.
- Styblo M, Del Razo LM, Vega L, Germolec DR, LeCluyse EL, Hamilton GA, Reed W, Wang C, Cullen WR and Thomas DJ. (2000). Comparative toxicity of trivalent and pentavalent inorganic and methylated arsenicals in rat and human cells. Arch Toxicol 74:289-299.
- Sudo N. (2006). Stress and gut microbiota: Does postnatal microbial colonization programs the hypothalamic-pituitary-adrenal system for stress response? Int Congr Ser 1287:350-354.
- Summanen PH and Finegold SM (2010). Family III. Porphyromonadaceae. Bergey's Manual of Systematic Bacteriology Springer: NewYork, pp 62-69.
- Sun D, Xu D andZhang B. (2006). Rac signaling in tumorigenesis and as target for anticancer drug development. Drug Resist Updat 9:274-287.
- Takai K, Kobayashi H, Nealson KH andHorikoshi K. (2003). *Deferribacter desulfuricans* sp. nov., a novel sulfur-, nitrate- and arsenate-reducing thermophile isolated from a deep-sea hydrothermal vent. Int J Syst Evol Microbiol 53:839-846.
- Tamaki S and Frankenberger WT, Jr. (1992). Environmental biochemistry of arsenic. Rev Environ Contam Toxicol 124:79-110.
- Tamir S and Tannenbaum SR. (1996). The role of nitric oxide ($NO^{\cdot}$) in the carcinogenic process. Biochimica et Biophysica Acta - Reviews on Cancer 1288:F31-F36.

- Tamura K, Dudley J, Nei M and Kumar S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596-1599.
- Tayayuki E (2009a). Genus I *Ruminococcus*. Bergey's Manual of Systematic Bacteriology: The Firmicutes. Springer: Newyork, pp 1016-1018.
- Tayayuki E (2009b). Genus VII *Coprococcus*. Bergey's Manual of Systematic Bacteriology: The Firmicutes. Springer: Newyork, pp 940.
- Thomas DJ, Styblo M and Lin S. (2001). The cellular metabolism and systemic toxicity of arsenic. Toxicol Appl Pharmacol 176:127-144.
- Tiihonen K, Ouwehand AC and Rautonen N. (2010). Human intestinal microbiota and healthy ageing. Ageing Res Rev 9:107-116.
- Turnbaugh PJ, Bäckhed F, Fulton L andGordon JI. (2008). Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. Cell Host and Microbe 3:213-223.
- Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R and Gordon JI. (2009). A core gut microbiome in obese and lean twins. Nature 457:480-484.
- Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER and Gordon JI. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 444:1027-131.

- Uni Z, Zaiger G, Gal-Garber O, Pines M, Rozenboim I and Reifen R. (2000). Vitamin A deficiency interferes with proliferation and maturation of cells in the chicken small intestine. Br Poult Sci 41:410-415.
- USGS. (2007).Arsenic in ground water in the United States.U.S. Geological Survey. http://water.usgs.gov/nawqa/trace/pubs/geo_v46n11/fig2.html.
- Vahter M. (2002). Mechanisms of arsenic biotransformation. Toxicology 181-182:211.
- Vahter M. (1994). Species differences in the metabolism of arsenic compounds. Appl Organometallic Chem 8:175-182.
- Vahter M, Friberg L, Rahnster B, Nygren Å and Nolinder P. (1986). Airborne arsenic and urinary excretion of metabolites of inorganic arsenic among smelter workers. Int Arch Occup Environ Health 57:79-91.
- Vahter M and Norin H. (1980). Metabolism of 74As-labeled trivalent and pentavalent inorganic arsenic in mice. Environ Res 21:446-457.
- Van de Wiele T, Gallawa CM, Kubachka KM, Creed JT, Basta N, Dayton EA, Whitacre S, Du Laing G and Bradham K. (2010). Arsenic metabolism by human gut microbiota upon in vitro digestion of contaminated soils. Environ Health Perspect 118:1004-1009.
- Verhulst SL, Vael C, Beunckens C, Nelen V, Goossens H and Desager K. (2008). A longitudinal analysis on the association between antibiotic use, intestinal microflora, and wheezing during the first year of life. J Asthma 45:828-832.
- Vorontsov II, Minasov G, Brunzelle JS, Shuvalova L, Kiryukhina O, Collart FR and Anderson WF. (2007). Crystal structure of an apo form of *Shigella flexneri*

ArsH protein with an NADPH-dependent FMN reductase activity. Protein Sci 16:2483-2490.

- Waalkes MP, Liu J, Ward JM and Diwan BA. (2004). Animal models for arsenic carcinogenesis: inorganic arsenic is a transplacental carcinogen in mice. Toxicol Appl Pharmacol 198:377-384.
- Walter J, Hertel C, Tannock GW, Lis CM, Munro K and Hammes WP. (2001). Detection of *Lactobacillus, Pediococcus, Leuconostoc* and *Weissella* species in human feces by using group-specific PCR primers and Denaturing Gradient Gel Electrophoresis. Appl Environ Microbiol 67:2578-2585.
- Wang L, Jeon B, Sahin O and Zhang Q. (2009). Identification of an arsenic resistance and arsenic-sensing system in *Campylobacter jejuni*. Appl Environ Microbiol 75:5064-5073.
- Wang Q, Garrity GM, Tiedje JM and Cole JR. (2007). Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73:5261-5267.
- Wang Y, Hoenig JD, Malin KJ, Qamar S, Petrof EO, Sun J, Antonopoulos DA, Chang EB and Claud EC. (2009). 16S rRNA gene-based analysis of fecal microbiota from preterm infants with and without necrotizing enterocolitis. ISME J 3:944-954.
- Weaver W and Shannon CE (1949). The Mathematical Theory of Communication. Univ. Illinois Press, Urbana.
- Wedepohl KH (1991). The compunds of the upper earth's crust and the natural cycles of selected metals. Ed. E. Merian. Metals and their Compounds in the

Environment. Occurance, Analysis and Biological Relevance. VCH: NewYork NY, pp 3-17.

- Welch AH, Westjohn DB, Helsel DR and Wanty RB. (2000). Arsenic in ground water of the United States: occurrence and geochemistry. Ground Water 38:589-604.
- Wester RC, Maibach HI, Sedik L, Melendres J and Wade M. (1993). In vivo and in vitro percutaneous absorption and skin decontamination of arsenic from water and soil. Toxicol Sci 20:336-340.
- WHO. (2010).WHO Food Additives Series No. 63/ FAO JECFA Monographs 8.FAO/WHO publication.
- WHO. (2001).Arsenic and arsenic compounds. Environmental Health Criteria.United Nations Environment Programme. World Health Organization: Geneva http://www.inchem.org/documents/ehc/ehc/ehc224.htm.
- Wikoff W, Anfora A, Liu J, Schultz P, Lesley S, Peters E and Siuzdak G. (2009). Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. Proc Natl Acad Sci 106:3698:3703.
- Wolfe-Simon F, Blum JS, Kulp TR, Gordon GW, Hoeft SE, Pett-Ridge J, Stolz JF, Webb SM, Weber PK, Davies PCW, Anbar AD and Oremland RS. (2010). A bacterium that can grow by using arsenic instead of phosphorus. Science. DOI:10.1126/science.1197258.
- Wu MM, Kuo TL, Hwang YH and Chen CJ. (1989). Dose–response relation between arsenic concentration in well water and mortality from cancers and vascular disease. Am J Epidemiol 130:1123-1132.

- Wu J, Tisa LS and Rosen BP. (1992). Membrane topology of the ArsB protein, the membrane subunit of an anion-translocating ATPase. J Biol Chem 267:12570-12576.
- Wysocki R, Bobrowicz P and Ulaszewski S. (1997). The *Saccharomyces cerevisiae* ACR3 gene encodes a putative membrane protein involved in arsenite transport. J Biol Chem 272:30061-30066.
- Xu C, Zhou T, Kuroda M and Rosen BP. (1998). Metalloid resistance mechanisms in prokaryotes. J Biochem 123:16-23.
- Xu C, Shi W and Rosen BP. (1996). The chromosomal *arsR* gene of *Escherichia coli* encodes a trans-acting metalloregulatory protein. J Biol Chem 271:2427-2432.
- Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, Hooper LV and Gordon JI. (2003). A genomic view of the human-*Bacteroides thetaiotaomicron* Symbiosis. Science 299:2074-2076.
- Yamauchi H, Takahashi K, Mashiko M and Yamamura Y (1989). Biological monitoring of arsenic exposure of gallium arsenide and inorganic arsenic exposed workers by determination of inorganic arsenic and its metabolites in urine and hair. Am Ind Hyg Assoc J 50:606-612.
- Yamauchi H, Kaise T andYamamura Y. (1986). Metabolism and excretion of orally administered arsenobetaine in the hamster. Bull Environ Contam Toxicol 36:350-355.
- Yamauchi H and Yamamura Y. (1985). Metabolism and excretion of orally administrated arsenic trioxide in the hamster. Toxicology 34:113:121.

- Yarullina DR, Il'inskaya ON, Aganov AV, Silkin NI and Zverev DG. (2006). Alternative pathways of nitric oxide formation in *Lactobacilli*: Evidence for nitric oxide synthase activity by EPR. Microbiol 75:634-638.
- Ye J, Yang H, Rosen BP and Bhattacharjee H. (2007). Crystal structure of the flavoprotein ArsH from *Sinorhizobium meliloti*. FEBS Lett 581:3996-4000.
- Yoshida K, Inoue Y, Kuroda K, Chen H and Wanibuchi H. (1998). Urinary excretion of arsenic metabolites after long term oral administration of various arsenic compounds to rats. J Toxicol Environ Health 54:179-192.
- Ysart G, Miller P, Barrett G, Farrington D, Lawrance P and Harrison N. (1999). Dietary exposures to nitrate in the UK. Food Addit Contam 16:521-532.
- Zakharyan RA, Sampayo-Reyes A, Healy SM, Tsaprailis G, Board PG, Liebler DC and Aposhian HV. (2001). Human Monomethylarsonic Acid (MMAV) reductase is a member of the glutathione-s-transferase superfamily. Chem Res Toxicol 14:1051-1057.
- Zakharyan RA, Ayala-Fierro F, Cullen WR, Carter DM and Aposhian HV. (1999). Enzymatic methylation of arsenic compounds. Monomethylarsonous acid (MMAIII) is the substrate for MMA methyltransferase of rabbit liver and human hepatocytes. Toxicol Appl Pharmacol 158:9-15.
- Zhang A, Feng H, Yang G, Pan X, Jiang X, Huang X, Dong X, Yang D, Xie Y, Peng L, Jun L, Hu C, Jian L and Wang X. (2007). Unventilated indoor coal-fired stoves in Guizhou province, China: cellular and genetic damage in villagers exposed to arsenic in food and air. Environ Health Perspect 115:653-658.

- Zhang Z, Naughton DP, Blake DR, Bejamin N, Stevens CR, Winyard PG, Symons MCR and Harrison R. (1997). Human xanthine oxidase converts nitrite ions into nitric oxide (NO). Biochem Soc Trans 25:524S.
- Zhao HY, Wang HJ, Lu Z and Xu SZ. (2004). Intestinal microflora in patients with liver cirrhosis. Chin J Dig Dis 5:64-67.
- Zoetendal EG, von Wright A, Vilpponen-Salmela T, Ben-Amor K, Akkermans ADL and de Vos WM. (2002). Mucosa-associated Bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. Appl Environ Microbiol 68:3401-3407.
- Zoetendal EG. (2001). The host genotype affects the bacterial community in the human gastronintestinal tract. Microb Ecol Health Dis 13:129-134.
- Zweier JL, Li H, Samouilov A and Liu X. (2010). Mechanisms of nitrite reduction to nitric oxide in the heart and vessel wall. Nitric Oxide 22:83-90.

APPENDICES

A.1 Materials and methods and results from Study 1

Materials and methods: Similar to section 2.1 to 2.9 except for the following:

Mice Exposure: Sodium arsenite at a concentration of 0 and 250 ppb in drinking water was fed to groups of six mice for a period of 2, 5 and 10 weeks. However, only three mice were exposed to As(III) for 10 week 0 ppb and 250 ppb group to obtain colon samples.

DNA extraction: Microbial community DNA extraction from the colon of 2 and 5 week samples was done by silica and guanidinium thiocynate (GuSCN) method as described by Boom et al, 1990. Silica suspension was prepared by adding DI water to 6 g of SiO_2 (spectrum cat.no.S1388) to a total volume of 50 ml and was sedimented for 24 hrs at unit gravity. Forty three ml of the supernatant was removed from the silica gradient and the volume was again raised to 50 ml with DI water. The silica mix was again allowed to settle for 5 hrs at unit gravity. Forty four ml of the supernatant was removed from the silica gradient and 60 µl of 32% HCl was added and mixed by vortexing. The silica suspension thus obtained was aliquoted in 1.5ml eppendorf tubes and autoclaved. Nine hundred µl of lysis buffer L6 [24 g GuSCN, 20 ml of 0.1 M tris HCl (pH 6.4), 4.4 ml 0.2 M EDTA (pH 8.0) and 0.5 ml Triton X] was added to 40 µl of silica suspension. Homogenized mouse colon sample was added to the above mix and vortexed for 5 seconds. The tubes were incubated at room temperature for 10 minutes, vortexed and centrifuged at 12,000 x g for 15 seconds. The supernatant was removed and the pellet was washed twice with 1 ml of washing buffer L2 [24 g GuSCN in 20 ml of 0.1 M tris HCl

(pH 6.4)] followed by washing with 1 ml of 70% ethanol and 1 ml acetone. The pellet was dried after the washings and then resuspended in 50 µl of TE buffer (10 mM tris HCl and 1 mM EDTA pH 8.0) by incubating at 56^0C for 10 minutes in a water bath. Any leftover debris was removed by centrifuging at 12,000 x g for 2 minutes and the supernatant containing the nucleic acid was transferred to a sterile eppendorf tube and quantified. DNA extraction from 10 week samples was done by QIAamp DNA stool mini kit as described in section 2.2.

Reamplification of bands obtained from DGGE gels

Specific bands were excised from the DGGE gels by using a razor blade and were kept in 10 µl of sterile water at 4°C overnight and later transferred to −20°C until further use. The excised bands were reamplified using primers HDA1 (without GC clamp) and HDA2 (Table 2.1) (Walter et al, 2001). The reaction mixture consisted of 10 µl of water eluted from DGGE bands, 4 µl of 5X buffer, 10 pmol each of forward and reverse primers, 0.4 µl dNTPs from a 200 µM stock, 0.5U of GoTaq DNA polymerase (Promega, Madison, WI, USA) and sterile water to a total volume of 20 µl. The reaction mixture was denatured at 94°C for 5 minutes, followed by 30 cycles of denaturation at 94°C for 30 sec, annealing at 50°C for 30 sec, extension at 72°C for 30 sec and a final extension at 72°C for 5 min. After PCR, the amplification was confirmed by visualizing the products on a 1.8% agarose gel after EtBr staining.

Cloning and sequencing of reamplified bands and DGGE-PCR products

The reamplified PCR products and DGGE-PCR products were ligated into the pCR2.1®-TOPO vector following the manufacturer's instructions (Invitrogen TOPO TA cloning kit, Invitrogen, Carlsbad, CA, USA). The ligation, transformation and plasmid

DNA extraction steps were done as mentioned in section 2.7. Sequencing was done using M13 forward and reverse primers provided with the Invitrogen TOPO TA cloning kit. The reaction mixture for sequencing consisted of 300 ng plasmid DNA, 3.2 pmol primer, 4 µl BigDye solution, 2 µl 5X sequencing buffer and sterile water to a total volume of 20 µl. The reaction conditions involved initial denaturation at 96^{o}C for 1 min, 40 cycles of denaturation at 96^{o}C for 30 sec, annealing at 50^{o}C for 5 sec and extension at 60^{o}C for 4 min followed by a final hold at 4^{o}C. The completed reaction was then purified using Centricep column refilled with 800 µl Sephadex G-50, dried in speed vac and resuspended in 10 µl of deionized formamide and sequenced using a 3100-Avant Genetic Analyzer (Applied Biosystems, Foster City, CA, USA).

Data deposition: The bacterial 16S rRNA gene sequences obtained from study 1 are deposited in Genbank database with accession numbers HQ681318-HQ681412 (10 week control), HQ681413 - HQ681511 (10 week 250 ppb), HQ681512 - HQ681605 (5 week 250 ppb) and HQ681606 - HQ681697 (5 week Control).

Results:

DNA extraction: Representative 0.8% agarose gels showing the quality of genomic DNA extracted from mouse colon using silica and guanidinium thiocynate method and QIAamp DNA stool mini kit are presented in Figure A.1. The 260:280 ratio for the extracted nucleic acids obtained with silica and guanidinium thiocynate method was generally ≤1.4 whereas with QIAamp DNA stool mini kit the 260:280 ratio varied between 1.6-1.8.

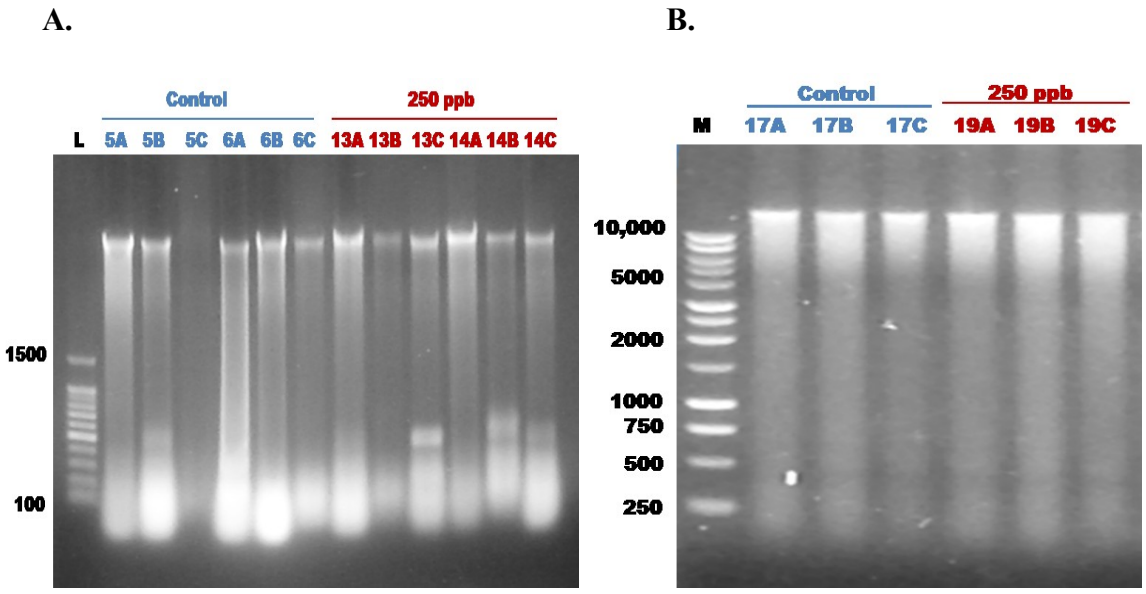
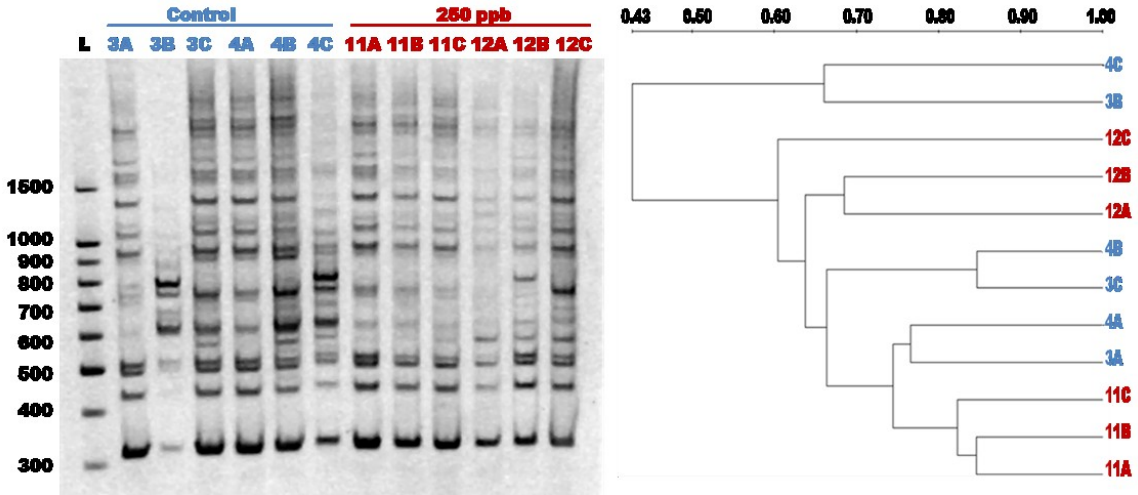


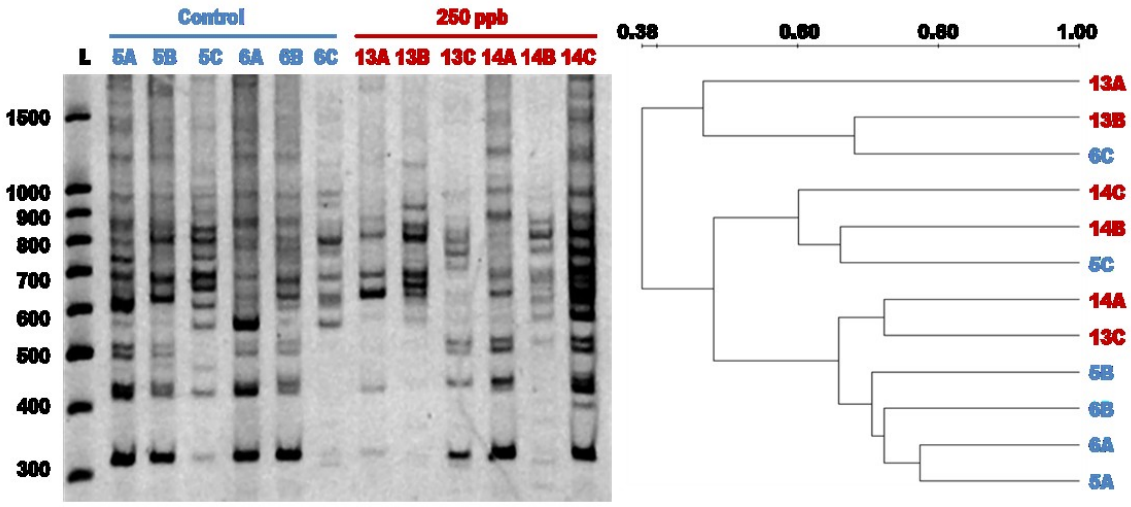
Figure A.1: Representative 0.8% agarose gels showing high MW genomic DNA obtained from mouse colons by using (A) Silica and guanidinium thiocynate method and (B) QIAamp DNA stool mini kit. L= 100 bp DNA ladder, M= 1Kb DNA ladder.

RISA: The spacer region between the 16S and 23S rRNA gene was amplified from 2, 5 and 10 week mouse colon community DNA. PCR products thus obtained were separated on a 5% polyacrylamide gel. Similar to study 2, the product size for the spacer region ranged from 300 bps to >1500 bps. Alterations in the RISA profiles of As(III) treated mice were observed when compared to the control mice in the form of absence/presence of bands and their intensity. These alterations between the control and 250 ppb As(III) exposed groups were more consistent and dramatic in 10 week samples than 2 and 5 week samples (Figure A.2). Dendrogram analysis of the 10 week RISA gel showed separate clustering of the control and 250 ppb As(III) exposed groups. However, mixed clustering of the control and 250 ppb As(III) exposed groups was observed for 2 and 5 week samples (Figure A.2).

A. 2 week



B. 5 week



C. 10 week

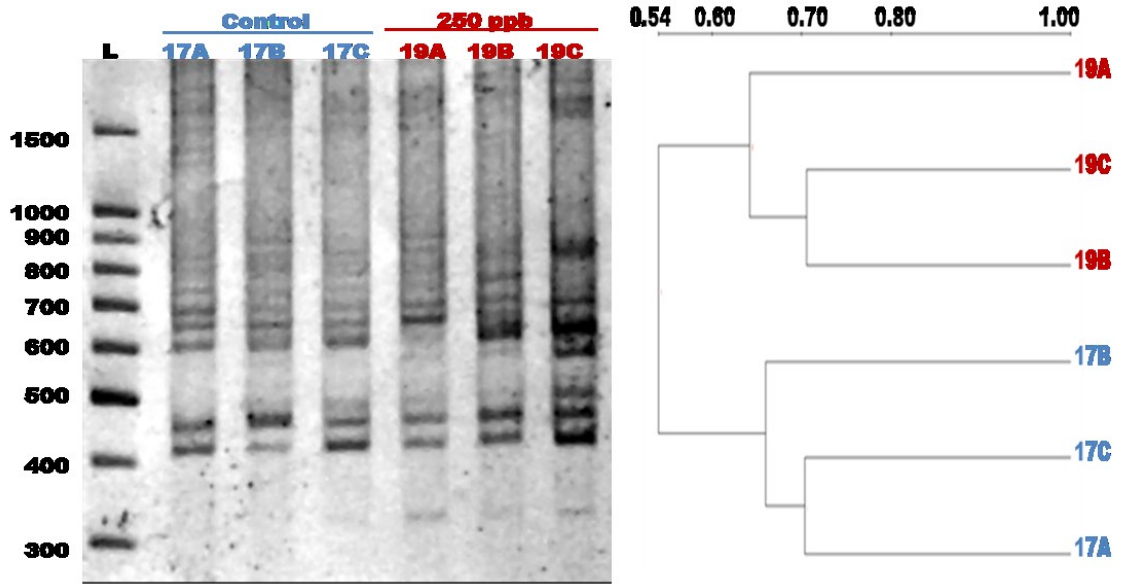
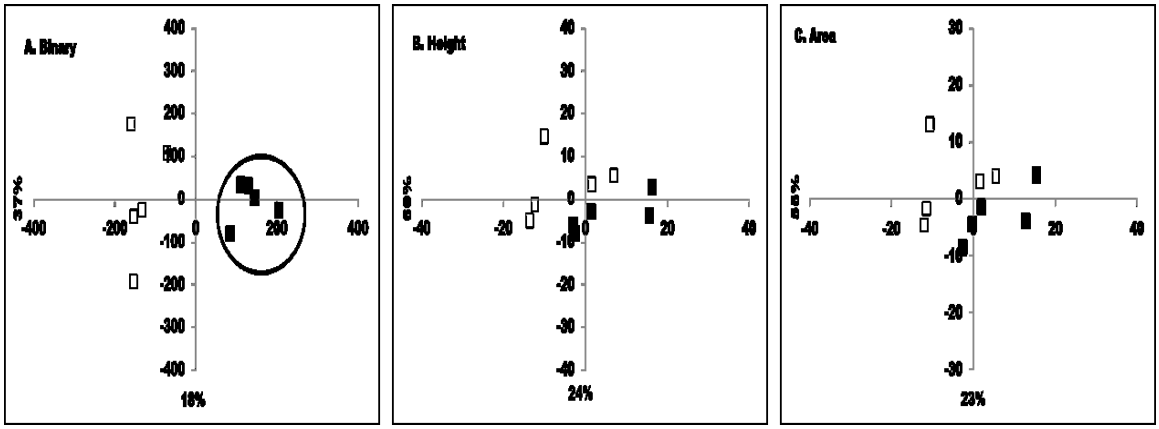


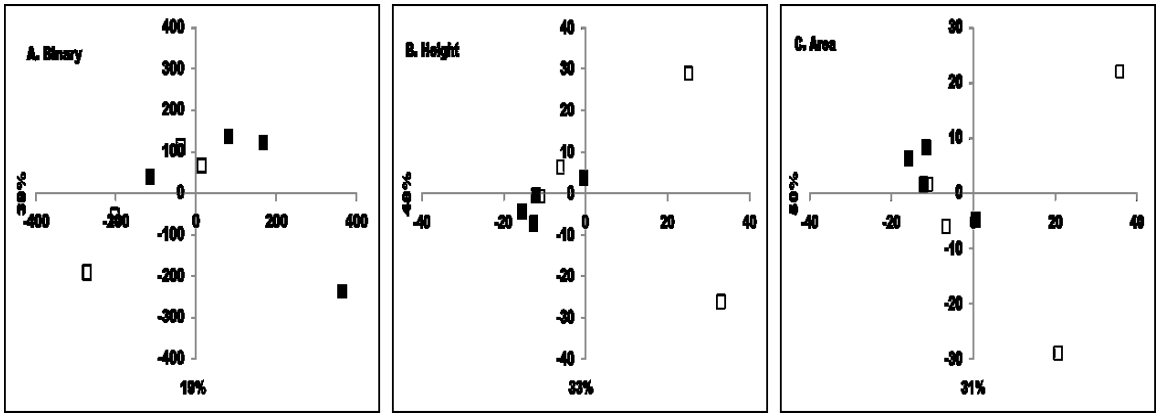
Figure A.2: RISA gels and dendrogram analysis of negative image of RISA fingerprints generated from 16S-23S intergenic region of control (blue) and 250 ppb (red) arsenic exposed mice at A) 2 week, B) 5 week, and C) 10 week. L= 100 bp DNA ladder. Numbers on left side of the DNA ladder represent size of the marker band in base pairs. Scale bar on dendrograms represents similarity index values.

ARISA: ARISA was performed on 2, 5 and 10 samples using 5'end labeled 1406F and 23Sr primers. The peaks obtained from ARISA electrophoregrams were analyzed in the same way as mentioned in section 3.2. The peak sizes obtained from ARISA profiles of this study ranged from 277 bps to 1155 bps representing intergenic spacer region of 27 bps to 905 bps. The matrices obtained for presence/absence of peaks, relative peak height and area were used for Principle Component Analysis (PCA). PCA ordination plots obtained from all three matrices showed clear separation of control and 250 ppb As(III) exposed samples in case of 10 week samples but not for the 2 and 5 week samples (Figure A.3). An exception to this observation was separate clustering of the control samples from the 250 ppb As(III) exposed samples in the 2 week binary matrix. However similar clustering pattern was not found in the other two matrices (height and area) from the 2 week samples.

2 week



5 week



10 week

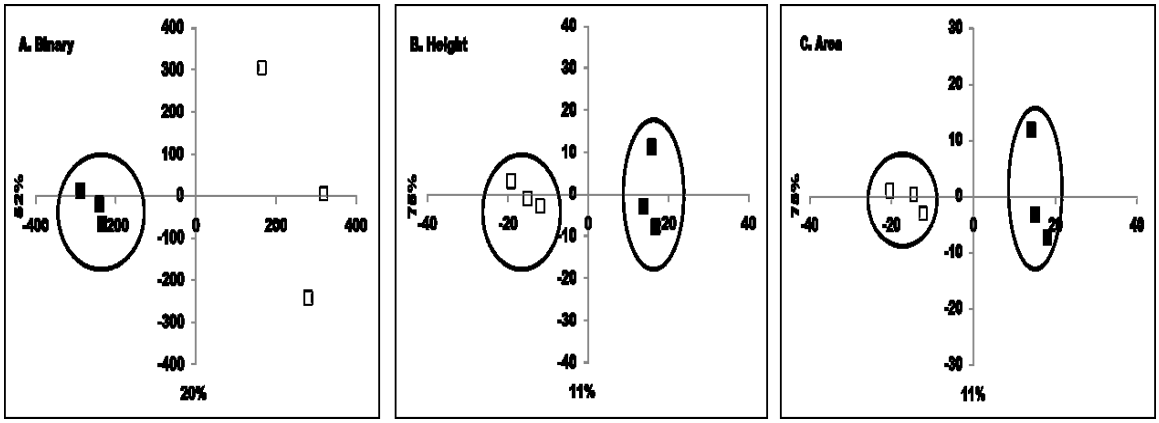
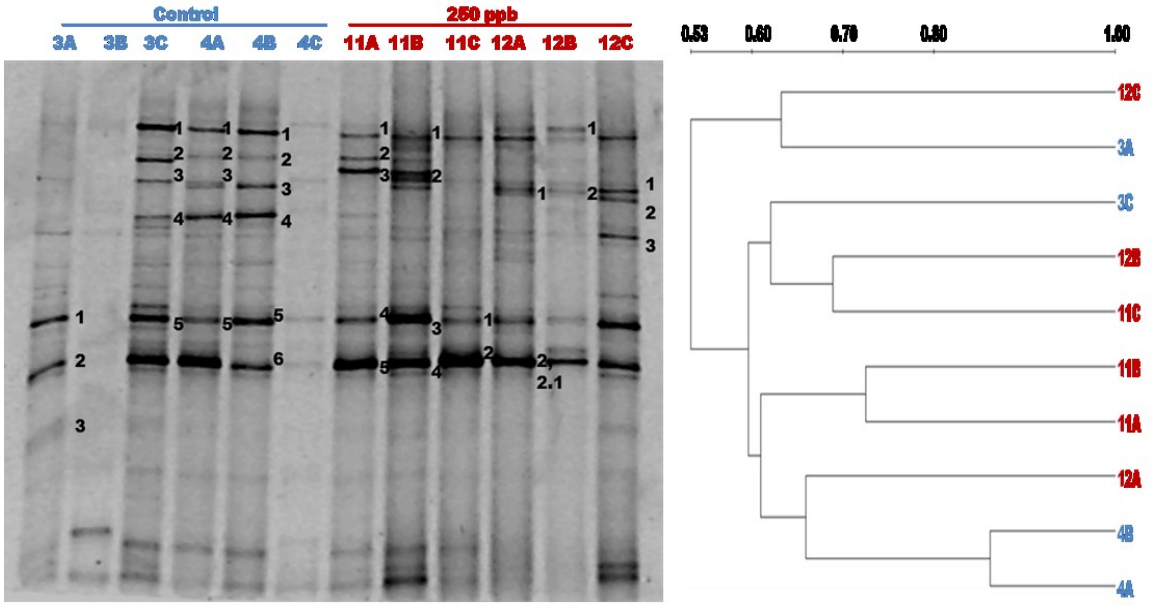


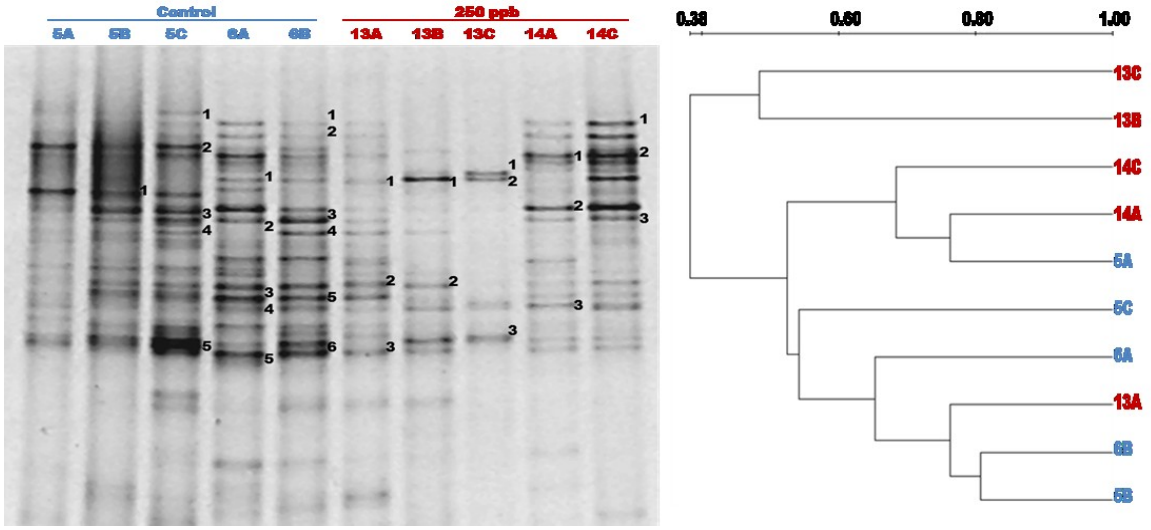
Figure A.3: PCA plots generated from ARISA profiles of bacterial communities from control (black squares) and 250 ppb As(III) exposed colons (open squares) of (I) 2, (II) 5 and (III) 10 week samples. The percentages shown along the axis represent the proportion of variation by that axis. PCA plots based on (A) presence/absence of peak (binary), (B) relative peak height and (C) relative peak area. Circles represent the groups that cluster together.

DGGE: A 177 bp long segment of the 16S rRNA gene was amplified from 2, 5 and 10 week mice colon community DNA. PCR products thus obtained were separated on a 40-60% DGGE gel and specific bands were excised and sequenced bidirectionally. Alterations in DGGE profiles of mice treated with 250 ppb of As(III) were observed when compared to the control mice. However, these changes were more consistent and dramatic in 10 week samples as compared to 2 and 5 week samples (Figure A.4). Dendrogram analysis of the DGGE gel showed clustering of control and 250 ppb As(III) exposed groups in two different clades in 10 week samples. Two and 5 week samples showed mixed clustering of the control and 250 ppb As(III) exposed groups (Figure A.4). Table A.1 shows the taxonomic classification of the sequences obtained from the DGGE gels of 2, 5 and 10 week colon samples.

A. 2 week



B. 5 week



C. 10 week

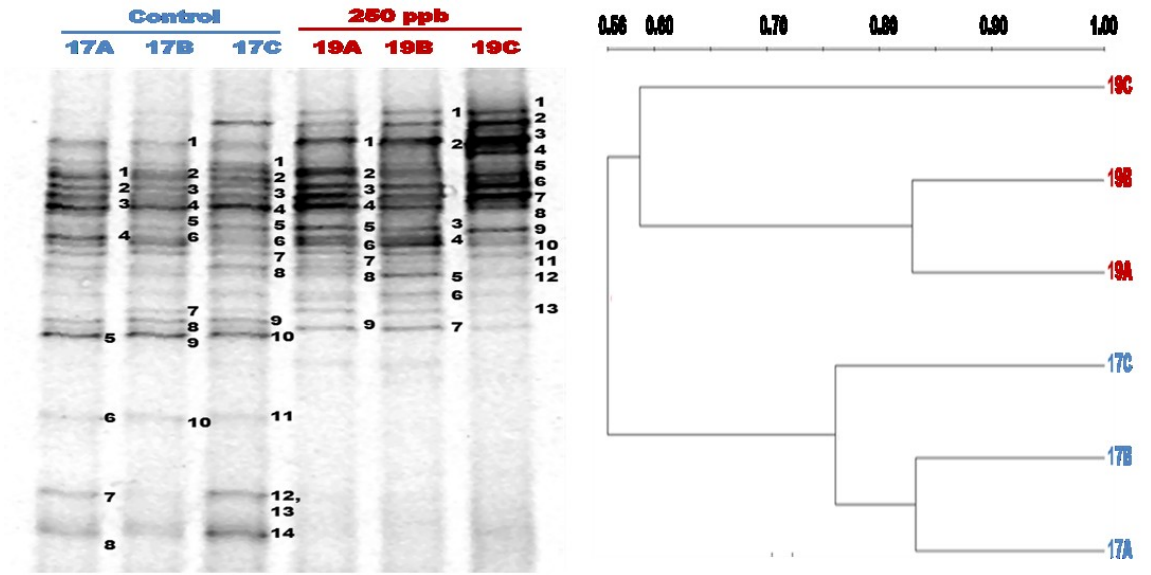


Figure A.4: DGGE and dendrogram analysis of negative image of DGGE fingerprints generated from 16S rDNA PCR of control (blue) and 250 ppb (red) As(III) exposed mice at A) 2 week, B) 5 week, and C) 10 week. Numbers on right side of each lane represent the sequenced band number. Scale bar on dendrograms represents similarity index values.

Taxonomic classification	Time point	Band name
Unclassified bacteria	5 week	dgge14C1, dgge13C3, dgge5C3, dgge5C4
	10 week	dgge19A9, dgge19B5
Unclassified Firmicutes	2 week	3C3
	10 week	dgge19B4, dgge17B1, dgge17B8
Unclassified Bacteroidetes	2 week	11B1
Unclassified Clostridia	5 week	dgge14A1
Unclassified Clostridiales	5 week	dgge6A1
Unclassified Lachnospiraceae	2 week	11A4, 11A5, 11B4, 11B2, 4B1, 3A3, 4A1, 3C2, 3C5, 4B4
	5 week	dgge13B1, dgge13A3, dgge14A2, dgge6B3, dgge6B2, dgge6B5, dgge6B4, dgge5B1
	10 week	dgge19C2, dgge19C12, dgge19C5, dgge19B3, dgge19A8, dgge17A4, dgge17B6, dgge17C6, dgge17B7, dgge17B5, dgge17A1
Unclassified Porphyromonadaceae	2 week	12C3, 11C1, 11C2, 3A2, 3C1, 4B3, 4B6
	10 week	dgge17C11, dgge17B10, dgge17A2, dgge17C12, dgge17C13, dgge17B2, dgge17C5, dgge17A3
Unclassified Prevotellaceae	10 week	dgge17C4
Unclassified Rikenellaceae	2 week	12A2
Unclassified Ruminococcaceae	10 week	dgge19A5, dgge19B7, dgge19C4
Lachnospiraceae	2 week	4B2, 12A2.1
	5 week	dgge5C1, dgge6A3, dgge6B1, dgge6A2, dgge6A5, dgge5C5, dgge13A1
	10 week	dgge17C8, dgge17B4, dgge19C13, dgge19C1
Porphyromonadaceae	2 week	12C1, 12B1, 11A1
	10 week	dgge17C14, dgge17C7, dgge17A7, dgge19A2, dgge19C6, dgge19C1
Prevotellaceae	2 week	3A1
	10 week	dgge17B9, dgge17C10, dgge17A5, dgge17A8
Rikenellaceae	2 week	11A3, 12C2, 12B2, 11B3, 4A2, 4A4, 4A3, 4A5, 4B5
	10 week	dgge19C3

Taxonomic classification	Time point	Band name
Ruminococcaceae	5 week	dgge13A2
	10 week	dgge19B6, dgge17B3, dgge17A6, dgge19B2
Anaeroplasmataceae	5 week	dgge14C3
Bacteroidaceae	2 week	11A2, 3C4
	5 week	dgge5C2, dgge14C2, dgge14A3
	10 week	dgge19C7, dgge17C2, dgge19A1,dgge17C9, dgge19C10,dgge17C3, dgge19C8, dgge19C9 dgge19A3, dgge19A4, dgge19A6, dgge19A7
Deferribacteraceae	5 week	dgge6A4, dgge13B2
	10 week	dgge17C1, dgge19B1
Lactobacillaceae	2 week	12A
	5 week	dgge13C2, dgge13C1, dgge6B6

Table A.1: Taxonomic classification of the bands sequenced from 2, 5 and 10 week DGGE gels. Each sequenced band is assigned to the lowest taxonomic level achieved at 50% confidence threshold. The time point refers to the DGGE gel from which the bands were sequenced. Nomenclature of the bands is such that it can be associated with the band numbers on the DGGE gels shown in figure A.3. For example, dgge13C2 from 5 week time point in the last row of the above table refers to band labeled as 2 from 13C lane of 5 week DGGE gel.

Clone library: A total of 194 clones from 10 week colon samples were analyzed, 32 from DGGE bands of control mice, 29 from DGGE bands of As(III) treated mice, 62 from control and 70 from 250 ppb of As(III) treated DGGE-PCR product clone library. In case of 5 week colon samples, 186 clones were analyzed, 17 from DGGE bands of control mice, 14 from DGGE bands of As(III) treated mice, 75 from 16S rDNA control clone library and 80 from 250 ppb of As(III) DGGE-PCR product clone library. The majority of the cloned sequences were assigned to the phylum Bacteroidetes and Firmicutes. Very few sequences belonging to the phylum Deferribacteres, Verrucomicrobia, Proteobacteria and Tenericutes were obtained. Most of the sequences classified under Firmicutes phylum were members of the class Clostridia and family Lachnospiraceae. Ruminococcaceae, Lactobacillaceae, Clostridiaceae and Erysipelotricheaceae were other families observed under the phylum Firmicutes. Within phylum Bacteroidetes, Bacteroidia was the predominant class constituted by families Bacteroidaceae, Prevotellaceae, Rikenellaceae and Porphyromonadaceae. Only one sequence was observed from the class Sphingobacteria, a Bacteroidetes. Furthermore, a number of clones were identified as uncultured bacterial clone when the sequences were classified at 90% confidence threshold. However the number of unclassified bacteria dropped immensely when classification was performed at 50% confidence threshold (Figure A.5).

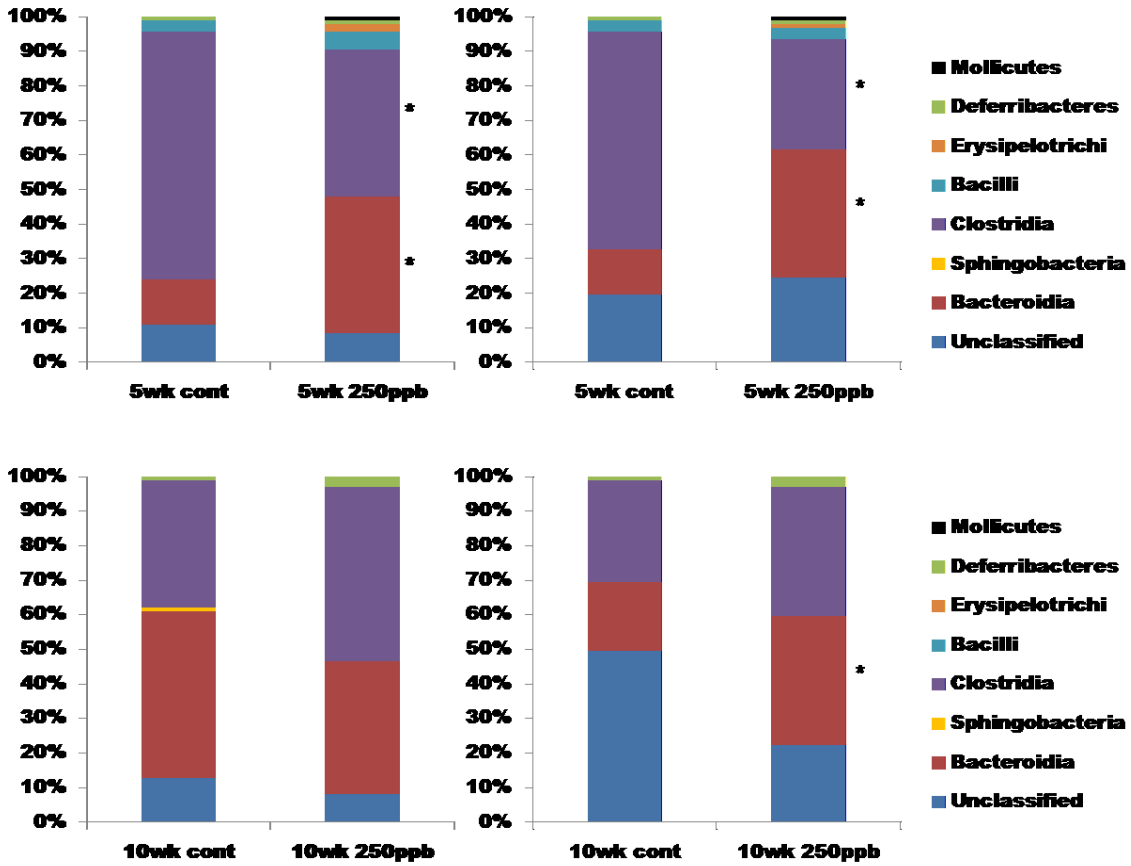


Figure A.5: Class level taxonomic classification of sequences obtained from DGGE-PCR product clone library. Left side classification was obtained at confidence threshold of 50% while right side taxonomic classification was obtained at a confidence threshold of 90%. * indicates significantly different from the control group for the same class ($p<0.01$).

Comparison of the control and 250 ppb libraries at 50% confidence threshold by using RDP II library compare tool showed that the two libraries differ with respect to the number of sequences that belong to phylum Bacteroidetes at both 5 and 10 weeks. At the same time, the control and 250 ppb libraries were different for Firmicutes only at the 5 week time point. Significant increase in Bacteroidaceae was observed after 250 ppb of As(III) exposure at 5 and 10 week intervals. In the 10 weeks samples, presence of

Prevotellaceae in the control clone library along with significant increase in Bacteroidaceae and decrease in Porphyromonadaceae after arsenic treatment kept the Bacteroidetes population constant in both control and 250 ppb of As(III) exposed libraries. Additionally sequences belonging to the Rikenellaceae family were observed only in the 5 and 10 week 250 ppb As(III) exposed clone libraries. The number of sequences belonging to class Clostridia and family Lachnospiraceae decreased significantly in 5 week samples after 250 ppb of As(III) exposure but did not had any effect at 10 weeks. Since most of the sequences that belong to family Lachnospiraceae could not be classified any further it was difficult to detect whether arsenic treatment was affecting a single or more than one genus from this family.

Paired comparison of 5 and 10 week control and As(III) treated libraries using Libshuff yielded a p-value <0.001, indicating that the two libraries are significantly different from each other. Regardless of the As(III) exposure it was found that the microbial community of mouse changed over time. The ratio of Firmicutes to Bacteroidetes decreased with time from 5 to 10 week in DGGE-PCR products clone libraries (Figure A.5, A.6).

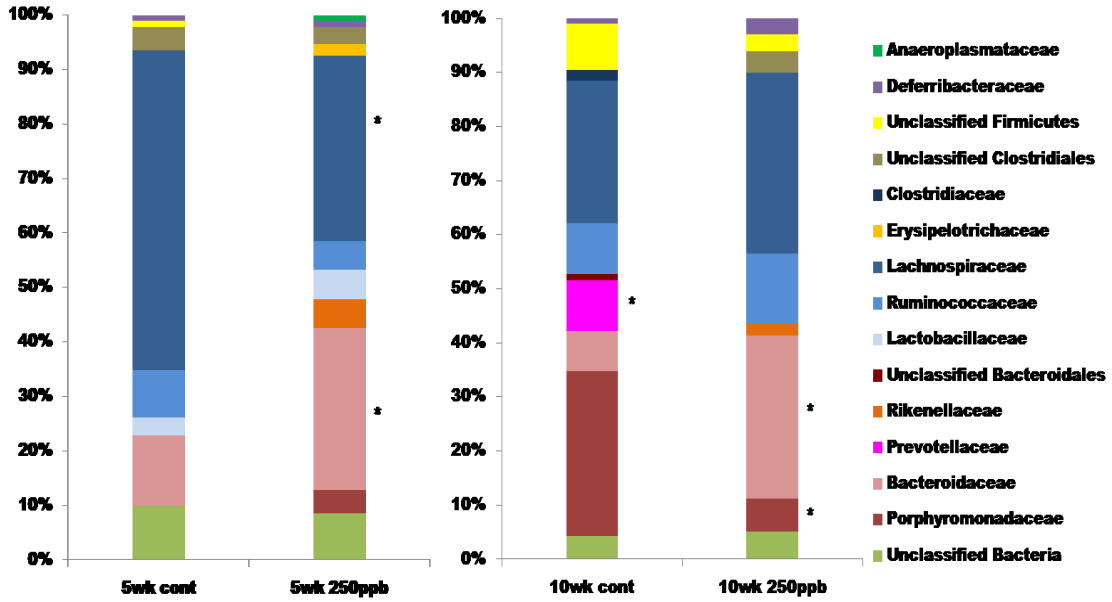


Figure A.6: Family level taxonomic classification of sequences obtained from DGGE-PCR product clone library at confidence threshold of 50%. * indicates significant difference between the control and 250 ppb As(III) exposed group for the same family ($p<0.01$).

Microbial diversity: Based on distance matrix, sequences obtained from the 5 and 10 week DGGE gels and DGGE PCR product based clone library were grouped into the same phylotype when they were 97%, 95% or 90% identical. For the combined dataset, the number of OTUs with only one sequence at a distance of 0.03, 0.05 and 0.1 were 71, 45 and 23 respectively. Regardless of As(III) exposure, a total of 121 unique phylotypes were identified including 72 and 71 phylotypes from 5 and 10 week clone libraries respectively at the distance level of 0.03. The number of unique phylotypes observed in the DGGE based clone libraries decreased with an increase in distance. However no significant difference in the number of observed OTUs was found in response to the As(III) exposure or time. Figure A.7 shows the number of OTUs shared between the

control and 250 ppb As(III) exposed clone libraries obtained from the 5 and 10 week mouse colon samples at a distance of 0.05.

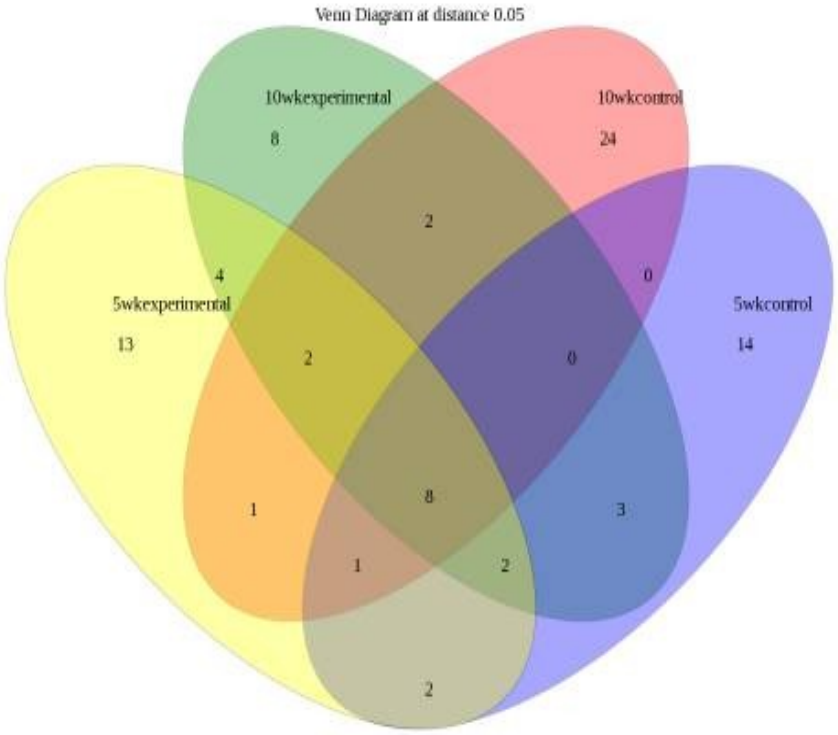


Figure A.7: Venn diagram showing the number of OTUs shared between control and 250 ppb As(III) treated DGGE-PCR product based clone libraries obtained from 5 and 10 week colon samples at a distance of 0.05. 250 ppb As(III) exposed groups are indicated by experimental in the figure.

The values obtained from Good's coverage indicate that sequencing additional clones will add only a few OTUs at the family and genus level. For each individual library, >80% coverage was observed at distances of 0.05 and 0.10. Table A.2 shows the number of OTUs, estimated richness, Good's coverage and diversity indices obtained

from each library and from the combined data set. Overall the Good's coverage was 88% and 94% for the combined libraries at a distance of 0.05 and 0.1 respectively.

Rarefaction curves obtained for the combined dataset at 0.03 and 0.05 distance levels did not reach saturation towards the end of the sampling. However for the 0.05 distance level, the curve rise was less steep towards the end of the sampling. At 0.1 distance level the curve almost reached plateau towards the end of sampling (Figure A.8). Both rarefaction curves and Good's coverage estimates indicated that the majority of bacteria at the family level of taxonomy have been captured in the current sequencing efforts. Similarly, estimation of species richness by Chao's estimator predicted that sequencing more clones could reveal more unique sequences in all the clone libraries (Table A.2). The percentage of observed/estimated OTUs was almost similar for both the family and genus level and ranged between 43−73%.

Within the individual library data sets, bacterial diversity at the family and genus level as measured by Shannon and Simpson index was highest for the 10 week control and lowest for the 5 week control group. These results differ slightly from the microbial richness wherein observed number of OTUs as determined by rarefaction curve was lowest in the 10 week 250 ppb As(III) exposed group. Overall, no effect of As(III) exposure was observed on the microbial diversity of mouse colon at any time point with any of the diversity estimations used in this study (Table A.2).

	Observed OTUs	Chao Estimator of species richness	Good's Coverage (C)	Shannon Diversity Index (H)	Simpson Diversity Index (D)
Distance 0.05					
5wk control	30	60.6	80%	2.733323	0.105829
5wk 250 ppb	33	57.42857	80%	2.971714	0.067948
5wk all	53	111.6667	82%	3.063833	0.098925
10wk control	38	63.5	81%	3.392611	0.030459
10wk 250 ppb	29	40	88%	3.009824	0.053597
10wkall	53	120.6667	85%	3.215441	0.051493
All	84	154.7143	88%	3.584084	0.049188
Distance 0.01					
5wk control	18	29.25	89%	1.766123	0.325609
5wk 250 ppb	23	53.33333	85%	2.231012	0.180508
5wk all	31	69.25	90%	2.153226	0.229468
10wk control	26	48.75	85%	2.687578	0.096305
10wk 250 ppb	18	27.33333	92%	2.112844	0.184086
10wk	36	83.5	90%	2.655157	0.128946
All	52	77.3	94%	2.619702	0.169657

Table A.2: Indices of richness, diversity and library coverage obtained for DGGE-PCR product based clonal libraries at distances of 0.05 and 0.10.

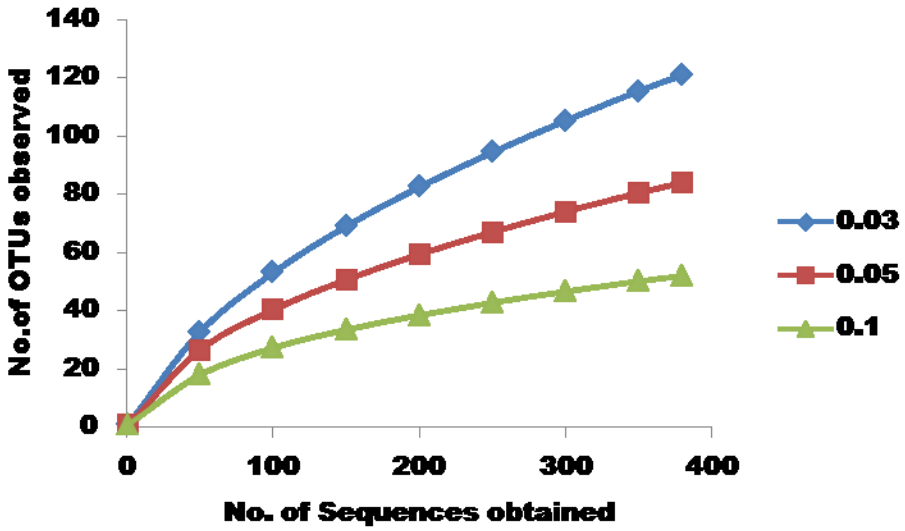


Figure A.8: Rarefaction curves obtained from all the sequences obtained from the DGGE-PCR product based clone library at a distance of 0.03, 0.05 and 0.10.

Phylogenetic tree: A neighbor joining tree of the representative OTU sequences obtained at a distance of 0.05 from the control and 250 ppb DGGE-PCR product based clone libraries of 5 and 10 week mouse colon samples was constructed (Figure A.9). A total of 130 OTUs were represented on the neighbor-joining tree. The phylogenetic tree indicated that most of the OTUs identified in the DGGE-PCR product based clone libraries are present in the control as well as 250 ppb As(III) exposed groups. The Lachospiraceae family formed the largest clade in the phylogenetic tree followed by Bacteroideaceae and Porphyromonadaceae. The number of sequences within certain OTUs was drastically different between the control and As(III) exposed libraries. For e.g. OTUs represented by Rikenellaceae were exclusively obtained from 5 and 10 week 250 ppb of As(III) exposed mouse colons.

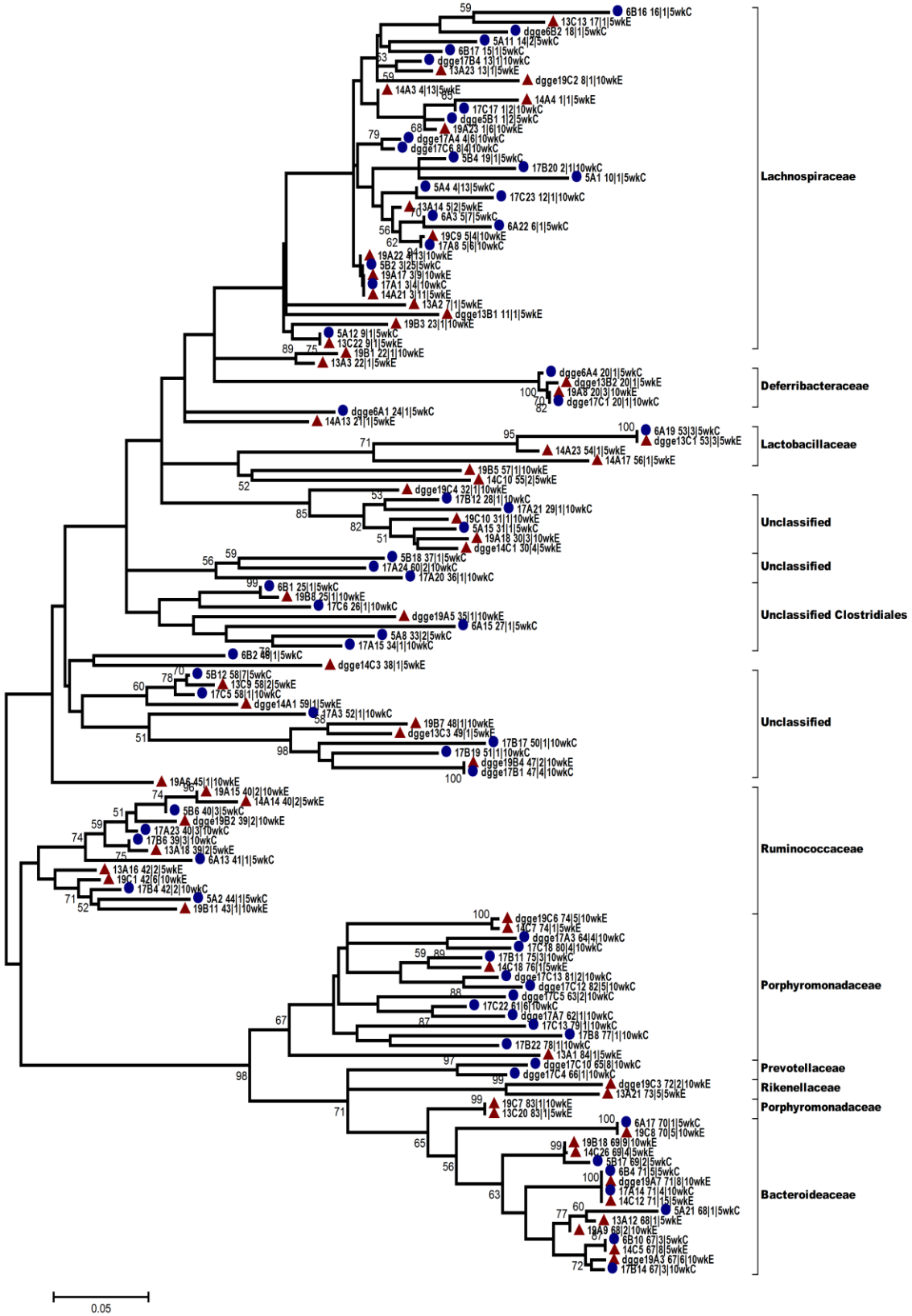

Lachnospiraceae
Deferribacteraceae
Lactobacillaceae
Unclassified
Unclassified
Unclassified Clostridiales
Unclassified
Ruminococcaceae
Porphyromonadaceae
Prevotellaceae
Rikenellaceae
Porphyromonadaceae
Bacteroideaceae
0.05

Figure A.9: Neighbor joining tree showing the positions of 16S rDNA phylotypes obtained from the mouse colon. Clones obtained from control samples are shown as blue circles and those obtained from 250 ppb As(III) exposed samples are shown as red triangles. Boot strap values obtained from 1000 replications and above 50% are shown at the nodes. Scale bar represents genetic distance based on 10 substitutions per 100 nucleotides. The number between "| |" indicates the number of sequences that were found under the same representative OTU.

A.2 Protocol for 5% polyacrylamide gel

5X TBE buffer

54 g Tris base
27.5 g Boric acid
20 ml 0.5M EDTA pH (8.0)
Make up the volume to 1000 ml with nanopure water

40% Acrylamide/Bis

38.93 g acrylamide
1.07 g bis-acrylamide
Dissolve in 100 ml nanopure water
Filter and store in dark at 4^{o}C

10% APS

Dissolve 1g APS in 10 ml nanopure water.

5% polyacrylamide gel

3 ml 5X TBE buffer
1.87 ml 40% Acrylamide/Bis
150 µl 10% APS
15 µl TEMED
Make up the volume to 15 ml with nanopure water

A.3 Protocol for DGGE

50X TAE buffer

242 g Tris base
57.1 ml glacial acetic acid
100 ml 0.5M EDTA pH (8.0)
Make up the volume to 1000 ml with nanopure water

40% Acrylamide/Bis

38.93 g acrylamide
1.07 g bis-acrylamide
Dissolve in 100 ml nanopure water
Filter and store in dark at 4^{o}C

10% APS

Dissolve 1g APS in 10 ml nanopure water.

0% Urea-formamide(UF) stock solution for 10% arylamide gels

12.5 ml 40% acrylamide
1 ml 50X TAE
Make up the volume to 50 ml with nanopure water

80% Urea-formamide stock solution for 10% arylamide gels

12.5 ml 40% acrylamide
1 ml 50X TAE
16 ml ultrapure formamide
16.8 g Urea
Make up the volume to 50 ml using pure water.

Working urea-formamide gradient solutions
40% UF: Mix 5 ml 0% UF stock solution and 5 ml 80% UF stock solution
60% UF: Mix 2.5 ml 0% UF stock solution and 7.5 ml 80% UF stock solution

Add 100 µl 10% APS and 10 µl TEMED to working gradient solutions immediately before pouring the gel.

A.4 LB media preparation

LB liquid broth (pH 7)

10 g tryptone
5 g yeast extract
10 g NaCl
Make up the volume to 1 L with nanopure water and autoclave.

LB Agarose plates

10 g tryptone
5 g yeast extract
10 g NaCl
15 g agar
Make up the volume to 1 L with nanopure water and autoclave.

Ampicillin selective media

Prepare stock solution (100 mg/ml) by dissolving 100 mg ampicillin in 1 ml of 50% ethanol. Add 1000 µl of ampicillin stock solution to 1 L of LB media.

X- Gal

Add 40 mg of X-gal to 1ml of dimethylformamide. Use 40 µl of X-gal for each LB plate.

A.5 Primers used in section 3.9

Organism	Gene	IMG Gene Object Identifier	Primer Sequence (5' to 3' end)
Bacteroides intestinalis	*rpoB*	642809876	F2168-GGCGTCGTACAAACCAGAAT R2368-CCACACGTTCGTTCAATACG
	arsA	642811825	F59-AGGGCGGTGTAGGAAAAACT R253-CTTGTTCCGGGTCGAGATTA
	nrfA	642811669	F1023-CAATGTGTACGAACGCCAAC R1221-ATGGAAAGAACCACCGTGAG
Prevotella melaninogenica	*rpoB*	645585915	F2269- AACGGTGAGTTGGCATTAGG R2579-TCGCTTTCACCCTTAGGAGA
	arsA	645586155	F121-GGCTTGGCAGATAATGGAAA R298-TAACGCCTTCCCTGTATTCG
Oribacterium sinus	*rpoB*	644432171	F627-TGTCTTTATTCGTGCGCTTG R867-GTATCGTCCCACCTTGGCTA
	arsA	644434364	F89-GTGCAACAGCAGTGGCTTTA R371-AAGGCAGCAATCTCCACTGT
Lactobacillus planatarum	*arsA*	637590361	F274-AGTGTTGTTGGCCCTTATCG R734-GGTGTTTTAAGCGTGCCATT
			F715-AATGGCACGCTTAAAACACC R1164-CAATACCTGCTGCTCGTCAA
			F1063-ACCGTTCATCTAGCCACCAC R1594- GTGTAGCCAGCATGCTTTGA
			F274-AGTGTTGTTGGCCCTTATCG R831- CGGCACTTCATACTGAGCAA
			F274-AGTGTTGTTGGCCCTTATCG R1164-CAATACCTGCTGCTCGTCAA
			F715-AATGGCACGCTTAAAACACC R1594- GTGTAGCCAGCATGCTTTGA
Clostridium phytofermentans	*arsA*	641293247	F254-AGGCTGCTGCTGAATATCGT R1017-TTTTCCAACACCACCCTTTC
			F254-AGGCTGCTGCTGAATATCGT R1158-ATGCTCATCAATGTGGCTCA
			F998-GAAAGGGTGGTGTTGGAAAA R1320-GTTATCTGCACGCTCCACAA
			F254-AGGCTGCTGCTGAATATCGT R1320-GTTATCTGCACGCTCCACAA
Dorea formicigenerans	*rpoB*	641806964	F2765-AGATGTCACCGGGAGTGAAC R3051-CGCACCATCAAATACAGGTG
	nirB	641808481	F1026-AGATAACCGCCTGACTGGTG R1210-CCGGTGTAAATCTGCAAGGT

Table- 0.3: The sequences for the above genes were retrieved from DOE Joint Genome Institute portal (http://img.jgi.doe.gov/) and the primers were designed using primer 3 program.

A.6 Sequences obtained from RISA clone library and their top BLAST result

>12A3 (472 bp)
GCCAGGCATCCACCATGCGCCCTTAATAACTTAACCTATATTATCTTATGATA
ATGGTTATGAGTTTAGCGATCAAACATTAAATGTTTTAAACTCTTTAAAACGC
GGTGTTCTCGGTTATTTTAATTAACAAAGAAATAAAAGATACTATCTAGTTTT
CAAAGAACAAGTTTGATAACTAATGTTATCAATGGAGCTTAACGGGATCGAA
CCGTTGACCTCCTGCGTGCAAAGCAGGCGCTCTCCCAGCTGAGCTAAAGCCC
CGTATAATGGGCCTAAGGTGGACTCGAACCACCGACCTCACGCTTATCAGGC
GTGCGCTCTAACCAGCTGAGCTATAGGCCCATTTAGCTCTGAGAGTAGACCTC
TCAAAACTAAACAAAGTTTACGATAAAAGTGCAGGTTTCCGAAATTATCCTT
AGAAAGGAGGTGATCCAGCCGCAGGTTCTCCTACGGCTACCTTGTTACGACA

gb|AY526616.1| *Lactobacillus animalis* strain LA51 16S ribosomal RNA gene, partial sequence; 16S/23S rRNA intergenic spacer, tRNA-Ile and tRNA-Ala genes, complete sequence; and 23S ribosomal RNA gene, partial sequence Length=762
Score = 819 bits (443), Expect = 0.0, Identities = 463/472 (98%), Gaps = 4/472 (1%)

>12A4 (486 bp)
GCCAGGCATCCACCATGCGCCCTTAATAACTTAACCTATATTATCTTACGATA
ATGGTTATGAGTTTAGCGATCAAACATTAAATGTTTTAAACTCTTTAAAACGC
GGTGTTCTCGGTTATTTTAATTAACAAAGAAATAAAAGATATTATCTAGTTTT
CAAAGAACAAGTTTGATAACTAACGTTATCAATGGAGCTTAACGGGATCGAA
CCGTTGACCTCCTGCGTGCAAAGCAGGCGCTCTCCCAGCTGAGCTAAAGCCC
CAAAATATTTATTTAATGGGCCTAAGTGGACTCGAACCACCGACCTCACGCTT
ATCAGGCGTGCGCTCTAACCAGCTGAGCTATAGGCCCAAAATAAGCAACCTT
GAGAAAAGTAAATCTCTCAAAACTAAACAAAGTTTACGATAAAAGTGCAGGT
TTCCAAATTATCCTTAGAAGGGGGGGATCCAGCCGCAGGTTCTCCTACGGCT
ACCTGGTTACGACA

gb|AY526614.1| *Lactobacillus animalis* strain LA51 16S ribosomal RNA gene, partial sequence; 16S/23S rRNA intergenic spacer, tRNA-Ile and tRNA-Ala genes, complete sequence; and 23S ribosomal RNA gene, partial sequence Length=776
Score = 822 bits (445), Expect = 0.0, Identities = 475/488 (97%), Gaps = 8/488 (2%)

>12A6 (395 bp)
GCCAAGGGCATCCACCATGGCGCCCTTAATAACTTAACCTATATTATCTTACG
ATAATGGTTATGAGTTTAGCGATCAAACATTAAATGTTTTAAACTCTTTAAAA
CGCGGTGTTCTCGGTTATTTTAATTAACAAAGAAATAAAAGATATTATCTAGT
TTTCAAAGAACAAGTTTGATAACTATTGTTATCAATGGGAGCTTAACGGGATC
GAACCGTTGACCTCCTGCGTGCAAAGCAGGCGCTCTCCCAGCTGAGCTAAAG
CCCCATGTTTTAAGAGTAAACCTCTCAAAACTAAACAAAGTTTACGATAAAA
AGTGCAGGTTTTCCGAAATTATCCTTAGAAAGGAGGTGATCCCAGCCGCAGG
TTCTCCTACGGCTACCTTGTTACGACA

gb|AY526613.1| *Lactobacillus animalis* strain LA51 16S ribosomal RNA gene, partial sequence; 16S/23S rRNA intergenic spacer and tRNA-Ala gene, complete sequence; and 23S ribosomal RNA gene, partial sequence Length=682
Score = 669 bits (362), Expect = 0.0, Identities = 386/396 (97%), Gaps = 8/396 (2%)

>12A7 (542 bp)
GTCGTACAGGGTAGCCGTACCGGAAGGTGCGGCTGGAACACCTCCTTTCTGG
GAGCGCANAGTTCGTTATCAAGTTGACTCAGAGGTATTAGTTAACTTGTACTA
CGGTTGAATATGTATAAAATATAGATCTACCGGCAATAAAGTGTCGGCAAGA
GAGAAAAATGATGCTGAGGGAAACCAAGGCAAAGTTGACAGTCCTATAGCT
CAGTTGGTTAGAGCGCTACACTGATAATGTAGAGGTCGGCAGTTCANCTCTG
CCTGGGACTACAGAATCTCTGAGAGAGAATTTTGGGGGATTAGCTCAGCTGG
CTAGAGCATCTGCCTTGCACGCAGAGGGTCAACGGTTCGAATCCGTTATTCTC
CACAAAAAAGTTACCGAGACATCAGAAACGTAAAGTAACGACAAGATCTTTG
ACNTGATGGACAACGTAAATAAAGTAACCAGGAGCAAGCTGAANATTTATCA
ATCCGATTTACCCCTGTGGGTANCCGGAAATAAAAAAGTAAGCCAGGGGCAG
ACGGGTGGATGCCNTTGGNCA

gb|EU136697.1| *Parabacteroides goldsteinii* strain JCM13446 16S ribosomal RNA gene, partial sequence Length=2002
Score = 889 bits (481), Expect = 0.0, Identities = 521/541 (96%), Gaps = 9/541 (2%)

>4A23 (291 bp)
GCCAGGCATCCACCATGGCGCCCTTAATAACTTAACCTATATTATCTTACGAT
AAATGGTTATGAGTTTAGCGATCANACATTAAATGTTTTAACTCTTTAAAACG
CGGTGTTCTCGGTTATTTTAATTAACANAGAAATAAAAGATATTATCTAGTTT
TCAAAGAACAAGTTTGAGAGTAGACCTCTCAAAACTAAACAAAGTTTACGAT
AAAAGTGCNGGTTTCCGAAATTATCCTTAGAAAGGAGGTGATCCAGCCGCAG
GTTCTCCTACGGCTACCTTGTTACGACA

gb|AY526615.1| *Lactobacillus animalis* strain LA51 16S ribosomal RNA gene, partial sequence; 16S/23S rRNA intergenic spacer, complete sequence; and 23S ribosomal RNA gene, partial sequence Length=581
Score = 473 bits (256), Expect = 2e-130, Identities = 280/292 (96%), Gaps = 6/292 (2%)

>4A7 (292 bp)
GCCAGGCATCCACCATGCGCCCTTAATAACTTAACCTATATTATCTTACGATA
ATGGTTATGAGTTTAGCGATCAAACATTAAATGTTTTAAACTCTTTAAAACGC
GGTGTTCTCGGTTATTTTAATTAACAAAGAAATAAAAGATATTATCTAGTTTT
CAAAGAACAAGTTTGAGAGTAGACCTCTCAAAACTAAACAAAGTTTACGATA
AAAGTGCAGGTTTCCGAAATTATCCTTAGAANGGAGGTGATCCAGCCGCAGG
TTCTCCTACGGGCTACCTTGTTACGACNA

gb|AY526615.1| *Lactobacillus animalis* strain LA51 16S ribosomal RNA gene, partial sequence; 16S/23S rRNA intergenic spacer, complete sequence; and 23S ribosomal RNA gene, partial sequence Length=581
Score = 492 bits (266), Expect = 6e-136, Identities = 283/291 (97%), Gaps = 4/291 (1%)

>4A17 (293 bp)
GCCAGGGCATCCACCATGGCGCCCTTAATAACTTAACCTATATTATCTTACGA
TAATGGTTATGAGTTTAGCGATCAAACATTAAATGTTTTAAACTCTTTAAAAC
GCGGTGTTCTCGGTTATTTTAATTAACAAAGAAATAAAAGATATTATCTAGTT
TTCAAAGAACANGTTTGAGAGTAGACCTCTCAAAACTAAACAAAGTTTACGA
TAAAAGTGCAGGTTTCCGAAATCATCCTTAGAAAGGAGGTGATCCAGCCGCA
GGTTCTCCTACGGCTACCTTGTTACGACNA

gb|AY526615.1| *Lactobacillus animalis* strain LA51 16S ribosomal RNA gene, partial sequence; 16S/23S rRNA intergenic spacer, complete sequence; and 23S ribosomal RNA gene, partial sequence Length=581
Score = 486 bits (263), Expect = 3e-134, Identities = 282/291 (97%), Gaps = 3/291 (1%)

>4A18 (389 bp)
GCCAGGCATCCACCATGCGCCCTTAATANCTTAACCTATATTATCTTACGATA
ATGGTTATGAGTTTAGCGATCAAACATTAAATGTTTTAAACTCTTTAAAACGC
GGTGTTCTCGGTTATTTTAATTAACAAAGAAATAAAAGATATTATCTAGTTTT
CAAAGAACAAGTTTGATAACTATTGTTATCAATGGAGCTTAACGGGATCGAA
CCGTTGACCTCCTGCGTGCAAAGCAGGCGCTCTCCCAGCTGAGCTAAAGCCC
CATGTTTTAAGAGTAAACCTCTCAAAACTAAACCAAAGTTTACGATAAAAAG
NGCNGGTTTCCGNAAATTATCCTTAGAAAGGAGGTGATCCAGCCGCGGTTCT
CCTACGGCTACCTTGTTACGAC

gb|AY526613.1| *Lactobacillus animalis* strain LA51 16S ribosomal RNA gene, partial sequence; 16S/23S rRNA intergenic spacer and tRNA-Ala gene, complete sequence; and 23S ribosomal RNA gene, partial sequence Length=682
Score = 658 bits (356), Expect = 0.0, Identities = 381/393 (97%), Gaps = 7/393 (2%)

>4A3 (392 bp)
GCCAGGCATCCACCATGGCGCCCTTAATAACTTAACCTATATTATCTTACGAT
AATGGTTATGAGTTTAGCGATCAAACATTAAATGTTTTAAACTCTTTAAAACG
CGGTGTTCTCGGTTATTTTAATTANCAAAGAAATAAAAGATATTATCTAGTTT
TCAAAGAACAAGTTTGATAACTATTGTTATCAATGGAGCTTAACGGGATCGA
ACCGTTGACCTCCTGCGTGCAAAGCAGGCGCTCTCCCAGCTGAGCTAAAGCC
CCATGTTTTAAGAGTAAACCTCTCAAAACTAAACAAAGTTTACGATAAAAAG
TGCNGGTTTCCGAAATTATCCTTAGAAAGGAGGTGATCCAGCCGCAGGTTTCT
CCTANCGGCTACCTNGGTTACGAC

gb|AY526613.1| *Lactobacillus animalis* strain LA51 16S ribosomal RNA gene, partial sequence; 16S/23S rRNA intergenic spacer and tRNA-Ala gene, complete sequence; and 23S ribosomal RNA gene, partial sequence Length=682
Score = 656 bits (355), Expect = 0.0, Identities = 382/395 (97%), Gaps = 8/395 (2%)

>4A19 (291 bp)
GCCAGGCATCCACCATGCGCCCTTAATAACTTAACCTATATTATCTTACGATA
ATGGTTATGAGTTTAGCGATCAAACATTAAATGTTTTAAACTCTTTAAAACGC
GGTGTTCTCGGTTATTTTAATTAACAAAGAAATAAAAGATATTATCTAGTTTT
CAAAGAACAAGTTTGAGAGTAGATCTCTCAAAACTAAACAAAGTTTACGATA
AANGTGCAGGGTTTCCGAAATTATCCTTAGAAAGGAGGTGATCCAGCCGCAG
GTCCTCCTACGGCTACCTTGTTACGACA

gb|AY526615.1| *Lactobacillus animalis* strain LA51 16S ribosomal RNA gene, partial sequence; 16S/23S rRNA intergenic spacer, complete sequence; and 23S ribosomal RNA gene, partial sequence Length=581
Score = 484 bits (262), Expect = 1e-133, Identities = 282/291 (97%), Gaps = 4/291 (1%)

>4A4 (~470 bp)
GCCAAGGCATCCACCATGCGCCCTTATAACTTAACCTATATTATCTTATGATA
TGGTTATGAGTTTAGCGATCAAACATTAAATGTTTTAACTCTTTAAAACGCGG
TGTTCTCGGTTATTTTAATTAACAAAGAAATAAAAGATACTATCTAGTTTTCA
AAGAACAAGTTTGATAACTAATGTTATCAATGGAGCTTAACGGGATCGAACC
GTTGACCTCCTGCGTGCAAAGCAGGCGCTCTCCCAGCTGAGCTAAAGCCCCG
TATANTGGGCCTAAGTGNACTCCNAACCACCGACCTCACGCTTATCAGGCGT
GCGCTCTAACCAGCTGAGCTATAGGCCCATTTACTCCTGAGNGTAANCCNCN
CCAAACTAACCAANGTTANNATAANGGNGCGGGTTTCCGAAATTATCCTTAA
AAAGGNGGGTGANCCCCCCCCCGGTTNCCCTACGGGTACCTTGGTACNACA

gb|AY526616.1| *Lactobacillus animalis* strain LA51 16S ribosomal RNA gene, partial sequence; 16S/23S rRNA intergenic spacer, tRNA-Ile and tRNA-Ala genes, complete sequence; and 23S ribosomal RNA gene, partial sequence Length=762
Score = 649 bits (351), Expect = 0.0, Identities = 431/473 (91%), Gaps = 14/473 (3%

>4A12 (389 bp)
GCCAGGCATCCACCATGGCGCCCTTAATAACTTAACCTATATTATCTTACGAT
NATGGTTATGAGTTTAGCGATCAAACATTAAATGTTTTAAACTCTTTAAAACG
CGGTGTTCTCGGTTATTTTAATTAACAAAGAATAAAAGATATTATCTAGTTTT
CAAAGAACAAGTTTGATAACTATTGTTATCAATGGAGCTTAACGGGATCGAA
CCGTTGACCTCCTGCGTGCAAAGCAGGCGCTCTCCCAGCTGAGCTAAGCCCC
ATGTTTTAAGAGTAACCCTCTCAAAACTAAACAAAGTTTACGATAAAAGGTG
CAGGTTTCCGAAATTATCCTTAGAAAGGAGGTGATCCAGCCGCAGGGTTCTC
CTACGGCTACCTTGTTACGACA

gb|AY526613.1| *Lactobacillus animalis* strain LA51 16S ribosomal RNA gene, partial sequence; 16S/23S rRNA intergenic spacer and tRNA-Ala gene, complete sequence; and 23S ribosomal RNA gene, partial sequence Length=682
Score = 649 bits (351), Expect = 0.0, Identities = 380/393 (97%), Gaps = 8/393 (2%)

>4A20 (290 bp)
GCCAGGCATCCACCATGCGCCCTTAATAACTTAACCTATATTATCTTACGATA
ATGGTTATGAGTTTAGCGATCAAACATTAAATGTTTTAAACTCTTTAAAACGC
GGTGTTCTCGGTTATTTTAATTAACAAAGAAATAAAAGATATTATCTAGTTTT
CAAAGAACAAGTTTGAGAGTAGACCTCTCAAAACTAAACAAAGTTTACGATA
AAAGTGCGGGTTTCCGAAATTATCCTTAGAAAGGAGGTGATCCAGCCGCAGG
TTCTCCTACGGCTACCTTGTTACGACA

gb|AY526615.1| *Lactobacillus animalis* strain LA51 16S ribosomal RNA gene, partial sequence; 16S/23S rRNA intergenic spacer, complete sequence; and 23S ribosomal RNA gene, partial sequence Length=581
Score = 494 bits (267), Expect = 2e-136, Identities = 283/290 (98%), Gaps = 3/290 (1%)

>4A6 (289 bp)
GCCAGGCATCCACCATGCGCCCTTAATAACTTAACCTATATTATCTTACGATA
ATGGTTATGAGTTTAGCGATCAAACATTAAATGTTTTAACTCTTTAAAACGCG
GTGTTCTCGGTTATTTTAATTAACAAAGAAATAAAAGATATTATCTAGTTTTC
AAAGAACAAGTTTGAGAGTAGACCTCTCAAAACTAAACAAAGTTTACGATAA
AAGNGCNGGTTTCCGAAATTATCCTTAGAAAGGAGGTGATCCAGCCGCAGGT
TCTCCTACGGCTACCTTGTTACGACA

gb|AY526615.1| *Lactobacillus animalis* strain LA51 16S ribosomal RNA gene, partial sequence; 16S/23S rRNA intergenic spacer, complete sequence; and 23S ribosomal RNA gene, partial sequence Length=581
Score = 486 bits (263), Expect = 3e-134, Identities = 281/290 (97%), Gaps = 4/290 (1%)

>4A15 (292 bp)
GCCAGGCATCCACCATGGCGCCCTTNAATAACTTAACCTATATTATCTTACGA
TAATGGTTATGAGTTTAGCGATCAAACATTAAATGTTTTAAACTCTTTAAAAC
GCGGTGTTCTCGGTTATTTTAATTAACAAAGAAATAAAAGATATTATCTAGTT
TTCAAAGAACAAGTTTGAGAGTAGACCTCTCAAAACTAAACAAAGTTTACGA
TAAAAGTGCAGGTTTCCGAAATTATCCTTAGAAAGGAGGTGATCCAGCCGCA
GGGTTCTCCTACGGCTACCTTGTTACGAC

gb|AY526615.1| *Lactobacillus animalis* strain LA51 16S ribosomal RNA gene, partial sequence; 16S/23S rRNA intergenic spacer, complete sequence; and 23S ribosomal RNA gene, partial sequence Length=581
Score = 486 bits (263), Expect = 3e-134, Identities = 284/293 (97%), Gaps = 6/293 (2%)

>4A5 (388 bp)
GCCAGGCATCCACCATGCGCCCTTATANCTTAACCTATATTATCTTACGATAA
TGGTTATGAGTTTAGCGATCAAACATTAAATGTTTTAAACTCTTTAAAACGCG
GTGTTCTCGGTTATTTTAATTAACAAAGAAATAAAAGATATTATCTAGTTTTC
AAAGAACAAGTTTGATAACTATTGTTATCAATGGAGCTTAACGGGATCGAAC
CGTTGACCTCCTGCGTGCAAAGCAGGCGCTCTCCCAGCTGAGCTAAAGCCCC
ATGTTTTAAGAGTAAACCTCTCAAAACTAAACANAGTTTACGATAANAAGTG
CAGGTTTCCGAAATTATCCTTAGAAAGGNGGTGATCCNGCCGCNGGTTCTCCT
ACGGCTACCTTNGTTACAAC

gb|AY526613.1| *Lactobacillus animalis* strain LA51 16S ribosomal RNA gene, partial sequence; 16S/23S rRNA intergenic spacer and tRNA-Ala gene, complete sequence; and 23S ribosomal RNA gene, partial sequence Length=682
Score = 647 bits (350), Expect = 0.0, Identities = 375/389 (96%), Gaps = 6/389 (2%)

>4A21 (~482 bp)
GCCAGGCATCCACCATGGCGCCCTTAATAACTTAACCTATATTATCTTAACGA
TAATGGTTATGAGTTTAGCGATCAAACATTAAATGTTTTAAACTCTTTAAAAC
GCGGTGTTCTCGGTTATTTTAATTAACAAAGAAATAAAAGATATTATCTAGTT
TTCAAAGAACAAGTTTGATAACTAACGTTATCAATGGAGCTTAACGGGATCG
AACCGTTGACCTCCTGCGTGCAAAGCAGGCGCTCTCCCAGCTGAGCTAAAGC
CCCGTATANTGGGCCTAAGTGGACTCGAACCNCCGACCTCACGCTTATCNGG
CGTGCGCTCTAACCAGCTGAGCTATAGGNCCNTTTAGCTCTGAAGAGTAGAC
CTCTCAAAACTAAACANAGGTTTACGATAANNGTGGCNGGTTTCCCNAAATT
ATCCTTAGAAAGGGAGGNGATCCNNNCCCGNNGGTTCNCCCACGGGCTACNT
TGGTTCNAACG

gb|AY526616.1| *Lactobacillus animalis* strain LA51 16S ribosomal RNA gene, partial sequence; 16S/23S rRNA intergenic spacer, tRNA-Ile and tRNA-Ala genes, complete sequence; and 23S ribosomal RNA gene, partial sequence Length=762
Score = 693 bits (375), Expect = 0.0, Identities = 419/443 (95%), Gaps = 10/443 (2%)

>4A1 (~486 bp)
GCCAGGCATCCACCATGCGCCCTTATACTTAACCTATATTATCTTACGATAAT
GGTTATGAGTTTAGCGATCAAACATTAAATGTTTTAAACTCTTTAAAACGCGG
TGTTCTCGGTTATTTTAATTAACAAAGAAATAAAAGATATTATCTAGTTTTCA
AAGAACAAGCTTGATAACTAACGTTATCAATGGAGCTTAACGGGATCGAACC
GTTGACCTCCTGCGTGCAAAGCAGGCGCTCTCCCAGCTGAGCTAANGCCCCA
AAATATTTATTTAANGGGGCCTAGGNGGACTCGAACCCCCGACCTCACGCTT
ATCNGGCGTGCGCTCTAACCAGNTGAGCTATNNGCCCAAATAAGCAACCTTN
GAANAAAAGTANANCTCTCAAACTAACAAAGTTTACGANNANGGTGCNGGN
TCCNAAATTACCCTTNGAANGAGGGGNTCCCNCCCCCNGGTNCNCNTACNGG
CTACCTTGGTANNAAC

gb|AY526614.1| *Lactobacillus animalis* strain LA51 16S ribosomal RNA gene, partial sequence; 16S/23S rRNA intergenic spacer, tRNA-Ile and tRNA-Ala genes, complete sequence; and 23S ribosomal RNA gene, partial sequence Length=776
Score = 651 bits (352), Expect = 0.0, Identities = 436/485 (90%), Gaps = 16/485 (3%)

>4A2 (~290 bp)
GCCAAGGCATCCACCATGCGCCCTTAATACTTAACCTATATTATCTTACGATA
ATGGNNATGGAGTTTAGCGATCAAACNTTAAATGTTTTTAAACTCTTTGNNNC
GCGGTGTTCTCGGTTATTTTAATTAACAANGAATAAAAGATTTTATCTAGTTT
TTCAAGAACAAGTTTGAGAGTAGGCCTCTCAAACCTANGCAAAGTTTNCATG
AGGGGGNNGGGNTCCNAAATTATCCTTAAAAGGTAGGGGACCCANCCCCNA
GTTCCCCTNCGGNNACNTTNNAACNANA

gb|AY526615.1| *Lactobacillus animalis* strain LA51 16S ribosomal RNA gene, partial sequence; 16S/23S rRNA intergenic spacer, complete sequence; and 23S ribosomal RNA gene, partial sequence Length=581
Score = 292 bits (158), Expect = 6e-76, Identities = 237/280 (85%), Gaps = 13/280 (5%)

>4A22 (Incomplete sequence)
GTCGTACAGGTAGCCGTATCGGAAGGTGCGGCTGGATCACCTCCTTTCTAGG
GAAAAAGTAGGGGTTGGATTACTGTCTAGTTGTCAAGGAGTCCGTGCGGCGG
AAGACATATAGGAACAGCTGCTGTGCTTGCACAGGCAGATGTGGATATATGG
ATTCCGGCATAAGGCNAATGACCGAGCGCGAGGTATTGCAAANCATTACCCG
AGAGTGCGGGGATTCGAGNCACGGGGCNGAGTAAATAAAANACNACNCAAN
NANANGNNTTCNCGNTGGCCAATGCCCCTTATGGGNACACCCCCCGTTTCNC
NTCCCAAACACGNNGGTTAANCCCGTANGCGGGCCGAA

gb|GU208684.1| *Alicyclobacillus tolerans* strain DSM 16297 16S-23S ribosomal RNA intergenic spacer and 23S ribosomal RNA gene, partial sequence Length=366
Score = 99.0 bits (53), Expect = 2e-17 Identities = 65/70 (93%), Gaps = 4/70 (6%)

>4A13 (Incomplete sequence)
GGTCGTAACAAGGGTAGCCGTATCGGNAAGGTGCGGCTGGATCACCTCCTTT
CTAGGGAAGAAGAAGTAGGGTTGTATTGCTGTTTAAAAGTGAAGGATTCCGG
TGGTGATGCNCTTAGGGGANACACACGTACCCTTCCCGAACACGATCGTTAA
GACTTAAGCGGCCGATGGTACTATACTGGAGACGGTATGGGAGAGCAGGTGT
NGTGCCGGAACNTAAAAAACATTTAGAATTGCTGGTTTTACCNGGGATCAGA
GACNGGGACANAGATTCTTTGNATCTGATGGCTGATAAAGCCTNCACCCATG
GCATGTGAAAACTTGTCCGNANCCCAAAGAGGNCCCCCCGTACCTTTAAAAA
CCAAATATTGNANANTTCTTTTTTACCNGNTTTGCCANCGAAGATNCC

gb|CP001107.1| *Eubacterium rectale* ATCC 33656, Features in this part of subject sequence: rRNA-16S ribosomal RNA rRNA-5S ribosomal RNA, complete genome Length=3449685
Score = 163 bits (88), Expect = 8e-37 Identities = 183/227 (81%), Gaps = 19/227 (8%)

A.7 Clustal alignment of ars genes from the two operons of *B. thetaiotamicron*

arsA

```
operon2           ATGAAAGCATTCAATTTATCCGATATAGAACTGACTAAATACCTGTTTTTTTACAGGTAAA 60
operon1           ATGAAAGCATTTAATTTATCCGATATAGAACTAACAAAATACTTATTCTTCACAGGAAAA 60
                  ********** ******************** ** ****** * ** ** ***** ***

operon2           GGTGGAGTAGGAAAGACTTCTATTGCTTGTGCCACAGCAGTTGGTTTGGCTGATAAGGGA 120
operon1           GGTGGAGTAGGAAAAACCTCTATTGCTTGCGCTACGGCAGTGGGATTAGCTGATAAAGGG 120
                  ************** ** ********** ** ** ***** ** ** ******** **

operon2           AAGAAAATACTTCTTATCAGTACAGATCCGGCTTCTAACTTACAAGATGTTTTCGATCAA 180
operon1           GAGAAAATTCTTCTTATCAGTACAGACCCGGCTTCTAACCTGCAAGATGTTTTCAATCAA 180
                   ******* ***************** ************ * ************ *****

operon2           TCCTTAAACGGACACGGTACAGCTATTTCAGAAGTGCCGGGACTGACTGTTGTAAACCTT 240
operon1           ACCCTAAACGGACATGGTACAGCTATTTCAGAAGTACCGGGGCTGACGGTCGTAAACCTT 240
                   ** ********** ******************** ***** ***** ** *********

operon2           GATCCTGAGCAGGCAGCAGCAGAATACAGAGAAAGTGTTATTGCACCTTTCAGGGGAAAA 300
operon1           GACCCTGAACAGGCAGCAGCAGAGTACAGGGAAAGTGTGATTGCACCTTTTAGAGGACAA 300
                  ** ***** ************** ***** ******** *********** ** *** **

operon2           TTACCCGAAAGCGTTATTCAGAATATGGAAGAACAGCTTTCAGGTTCCTGCACAGTAGAA 360
operon1           TTACCTGAAAGCGTTATTCAAAATATGGAAGAACAACTTTCAGGCTCTTGTACGGTAGAA 360
                  ***** ************** ************** ******** ** ** ** ******

operon2           ATCGCAGCTTTTAATGAGTTCTCGGACTTTATTACCGATGCTGATAAAGCAAAGGAGTAT 420
operon1           ATCGCTGCCTTTAATGAGTTTTCAGACTTTATCACCGATGCTGATAAAGCGAAGGAATAC 420
                  ***** ** *********** ** ******** ***************** ***** **

operon2           GACCATATAATTTTTGATACAGCACCTACCGGACATACATTGCGAATGTTGCAGCTTCCA 480
operon1           GACCACATTATTTTTGATACAGCACCTACCGGACACACATTGCGAATGTTGCAGCTTCCA 480
                  ***** ** ************************** ************************

operon2           TCTGCATGGAGTACATTTATTAGTGAAAGTACACATGGTGCATCCTGTTTAGGGCAGTTA 540
operon1           TCGGCATGGAGTACATTTATTAGTGAAAGTACACATGGCGCATCTTGTTTAGGACAATTA 540
                  ** ********************************** ***** ******** ** ***

operon2           TCAGGCTTGGAGGAACGGAAAGGAATTTATAAACAGGCGGTAGAAACCCTATCTAATACA 600
operon1           TCAGGCTTGGAAGAAAGGAAAGGTATCTATAAACAGGCGGTTGATACATTGTCTGATACA 600
                  *********** *** ******* ** ************** ** **  * *** *****

operon2           AGCGCAACTCGGTTAGTACTGGTCAGTCGTCCTGAAATTTCTCCATTGAAAGAAGCTGCC 660
operon1           AGCGCAACCCGGTTAGTATTGGTAAGCCGTCCTGAAATAGCACCGTTGAAAGAAGCTGCC 660
                  ******** ********* **** ** ***********  * ** ***************

operon2           CGCTCTTCATCTGAATTACAGCTATTGGGAATTAAAAATCAGTTGTTGGTGATAAATGGT 720
operon1           CGTTCTTCGCATGAATTACAACTATTGGGAATTAAAAATCAGCTATTGGTGATAAATGGT 720
                  ** *****   ********* ********************* * ***************

operon2           ATTTTACAGCAACTGAATGAAGCGGATGATGTGTCACGACAACTGCATAACAGGCAGCAA 780
operon1           GTTTTACAGCAATTAGATGAAGCAGACAATGTTTCACAACAACTGTATAACAGGCAGCAA 780
                  *********** *  ******* **  **** **** ******* **************

operon2           AAGGCATTACAGGGTATGCCTGCTGAACTATCTGAATATCCTATGTACAGCGTTCCTCTA 840
operon1           AAGGCGTTACAAAGTATGCCTATTGCATTATCTGAATATCCCATGTATAGCGTTCCTCTG 840
                  ***** *****  ********  ** * ************** ***** ************

operon2           CGTTCTTATAATTTATCTGATATAGCTAATATACGCCGTATGTTATACAGCGACAGCCTT 900
operon1           CGTTCTTACAATTTATCCAATATTGCCAATATCCGCCGTATGCTATACAGTGATAGTATT 900
                  ******** ********  **** ** ***** ********* ******* ** **  **

operon2           GCGGATGATATTTGTTATCAGCCTGTAAGTGGTGCTAAAAGTATAGACGACTTGGTTAAT 960
operon1           ACGAATGAAATTCGTTATCAGCCGATAACTGATAGCAAAAGTATAGATGAATTGGTAAAT 960
                  ** **** *** **********  *** ** *   *********** ** ***** ***
```

```
operon2           GACCTCTATACTTCCGGTAAGCGGGTAGTGTTCACAATGGGAAAAGGCGGTGTAGGTAAA 1020
operon1           GACCTCTATACTTCCGGTAAACGAGTAGTGTTCACAATGGGAAAAGGTGGCGTAGGAAAA 1020
                  ******************** ** ********************** ** ***** ***


operon2           ACTACCTTAGCAACAGAAATAGCTCTGAAATTAACAAAACTCGGTGCAAAGGTACATCTT 1080
operon1           ACGACCTTAGCCACAGAAATAGCCTTGAAATTAACAAAGCTCGGTGCAAAGGTTCATCTT 1080
                  ** ******** ***********  ************* ************** ******

operon2           ACCACCACTGATCCGGCAAACCATCTGAACTATGATCTCGCTATAAAGTCGGGTATTACA 1140
operon1           ACCACTACTGACCCGGCAAATCACCTGAACTATGATTTTGCCATCAAATCAGGTATTACG 1140
                  ***** ***** ******** ** ************ * ** ** ** ** ********

operon2           GTAAGTCATATTGATGAAGCGGAAGTATTGGAAAACTACAAGAATGAAGTTCGTAGCAAA 1200
operon1           GTAAGCCATATAGACGAAGCGGAAGTGCTGGAAAAATACAAGAATGAAGTTCGTAGCAAA 1200
                  ***** ***** ** ***********  ******* ************************

operon2           GCGGCAGAAACCATGACTGCCGAAGATATGGAATATATTGAAGAAGATTTACGTTCGCCA 1260
operon1           GCTGCCGAAACCATGACTGCCGAAGATATGGAATATATAGAGGAAGATTTACGTTCACCA 1260
                  ** ** ********************************* ** ************** ***

operon2           TGCACACAAGAAATCGCAGTTTTCAAGGCATTTGCCGAAATTGTAGATAAAGCGGACAAT 1320
operon1           TGTACGCAAGAAATCGCCGTATTTAAGGCTTTTGCTGAAATTGTAGATAAAGCGGAGAAT 1320
                  ** ** *********** ** ** ***** ***** ******************** ***

operon2           GAAATTGTGGTGATTGATACTGCACCGACAGGTCATACATTGTTGCTTTTGGATGCTACC 1380
operon1           GAAATCGTAGTGATTGATACTGCACCAACAGGACATACTTTGTTGCTTTTGGATGCTACC 1380
                  ***** ** ***************** ***** ***** **********************

operon2           CAAAGCTACCATAAAGAAGTAGAACGTACCCAAGGAGCAGTTACGGGAGCGGTAGCCAAT 1440
operon1           CAAAGCTATCACAAAGAAGTAGAACGTACACAGGGAGAAGTAACCGGAGCAGTAGCCAAT 1440
                  ******** ** ***************** ** **** *** ** ***** *********

operon2           TTATTGCCCCGTTTGCGTAATCCAAAGGAAACGGAAGTCGTAATTGTTACCTTGCCTGAA 1500
operon1           TTATTACCCCGTTTACGCAATCCACAGGAAACGGAAGTAGTAATAGTCACTTTACCCGAA 1500
                  ***** ******** ** ****** ************* ***** ** ** ** ** ***

operon2           GCGACACCTGTATTTGAAGCTGAACGCCTGCAAATGGATTTGCAACGCGCCGGAATTAAT 1560
operon1           GCAACACCTGTTTTTGAAGCAGAACGCCTACAAATGGATTTGCAGCGTGCCGGGATTAAC 1560
                  ** ******** ******** ******** ************** ** ***** *****

operon2           AACAAATGGTGGGTAGTAAATGCCTGCCTGTCAATGACAAATACGGAAAATTCATTTTTG 1620
operon1           AACAAATGGTGGGTAGTAAATGCCTGTTTGTCATTGACAAATACAGCAAATTCATTTTTG 1620
                  *************************  ***** ********** * *************

operon2           CAGGCGAAAGCACAAAATGAAGTAAATTGGATTAAGAAAGTAGAACAATTGTCAAAGGGT 1680
operon1           CAAGCAAAAGCGCAAAGTGAATTGACTTGGATTAAGAAAGTAGAAGAATTGTCAAAGGGC 1680
                  ** ** ***** **** **** * * ******************** *************

operon2           AATGCTGCCTTGATTGGATGGAAAAATATATAG 1713
operon1           AATACTGCCCTGATTGAATGGAAAAACTTATAA 1713
                  *** ***** ****** *********  ****
```

arsB

```
operon2           ATGGAAAAATATATAGGAATTGGATTTTTTGAAAGATACCTGACTGTTTGGGTAACTTTG 60
operon1           ATGGAAAAGAAACAAGGAATTGGATTTTTTGAAAGATACCTGACTATTTGGGTAGCCTTG 60
                  ********  *   ****************************** ******** * ***

operon2           TGCATTGTTGCTGGTATTGCTATCGGACAATGGTTTCCGGCAATATCGCAAACATTGAGT 120
operon1           TGTATCGTTATCGGGATAGCCATCGGACAATATCTTCCGGCAATACCGCAAACATTGAGC 120
                  ** ** ***   ** ** ** **********   ************ *************

operon2           AAATTAGAATACGCCAATGTGTCTATACCTGTGGCTATCCTGATTTGGTTAATGATTTAC 180
operon1           AAATTTGAATATGCCAACGTGTCTATACCTGTGGCTATCCTGATTTGGTTAATGATATAT 180
                  ***** ***** ***** ************************************ **
```

```
operon2           CCCATGATGCTGAAAGTAGATTTTCAGAGTATTAAAAACGTTGGTAAGCGTCCGAAAGGA 240
operon1           CCCATGATGCTGAAAGTAGATTTTCAGAGCATCAAGAACGTAGGCAAACGTCCGAAAGGG 240
                  ****************************** ** ** ***** ** ** ***********

operon2           ATTATCATCACTTGTATTACAAATTGGCTTATAAAGCCTTTTACAATGTTTGGAATAGCT 300
operon1           ATTGTCATTACTTGTGTGACGAACTGGGTAATAAAGCCATTCACCATGTTTGGAATAGCT 300
                  *** **** ****** * ** ** *** * ******** ** ** ***************

operon2           TATCTATTCTTCTACGTTGTTTTCAAGTCTCTAATATCTGCCGAATTGGCCGAAGAATAT 360
operon1           TATCTATTCTTCTACGTTATTTTCAAGTCATTAATCCCTGCCGGATTAGCCGAAGAATAT 360
                  ****************** **********  ****  ****** *** ************

operon2           TTAGCAGGGGCAGTCTTATTGGGTGCTGCACCTTGTACGGCAATGGTATTTGTATGGAGT 420
operon1           TTAGCTGGGGCAGTTCTGTTAGGTGCTGCACCCTGTACGGCGATGGTGTTTGTATGGAGT 420
                  ***** ********  * ** ********** ******** ***** ************

operon2           TATCTGACCAAAGGGGATGCTGCCTATACATTGGTGCAGGTTGCAGTGAATGATTTAATC 480
operon1           CATTTAACCAAAGGGGATGCAGCTTACACATTGGTGCAAGTGGCAGTCAATGATTTAATA 480
                   ** * ************** ** ** ********** ** ***** ***********

operon2           ATATTGGTTGCTTTCGCTCCTATTGTTGCTTTTTTATTGGGCGTAGGTGGCGTTACGATT 540
operon1           ATATTGGTTGCTTTCGCTCCTATCGTTGCGTTTTTGTTGGGTGTCGGTGGCGTTTCTATC 540
                  ********************** ***** ***** ***** ** ********* * **

operon2           CCGTGGGATACATTATTGCTTTCAGTTGTATTGTTTGTTGTCATTCCTCTTTCTGCCGGA 600
operon1           CCGTGGGACACGCTTATACTCTCGGTCGTACTGTTTGTTGTCATTCCACTTTCTGCCGGA 600
                  ******** **  *  * ** ** ** *** *************** ************

operon2           ATAATAACCCGCATACTGGTTATCCGGCGAAAAGGGATAGAATATTTTAATACTGTATTT 660
operon1           ATAGTAACCCGTGTAACGGTTATCCGGCGAAAAGGAATAGACTATTTCAATACCGTCTTT 660
                  *** *******  **  ****************** ***** ***** ***** ** ***

operon2           ATCCGTAAGTTTGATAATTATACGACTGGCGGGTTACTATTAACGCTTATCATTCTGTTT 720
operon1           GTCCGCAAGTTCGATAACTATACGGTTGGCGGATTGCTATTAACGCTTATCATCCTGTTT 720
                   **** ***** ***** ******  ****** ** ***************** ******

operon2           TCATTTCAAGGAGAAACCATATTGAACAATCCGTTGCACATCGTATTGATTGCCGTTCCG 780
operon1           TCGTTTCAGGGGGAAACGATACTGAACAATCCGTTACATATTGTCTTGATTGCCGTCCCG 780
                  ** ***** ** ***** *** ************** ** ** ** ************ ***

operon2           CTTGTGTTGCAAACAGTTCTGATATTCTTTGTAGCCTATGGTTGGGCGAAATGGTGGAAA 840
operon1           CTTGTCCTGCAAACGGTTCTGATATTCTTTGTTGCCTACGGTTGGGCAAAGTGGTGGAAG 840
                  *****  ******* ***************** ***** ******** ** ********

operon2           TTACCTCACAATATAGCTGCACCTGCCGGAATGATTGGTGCAAGCAATTTTTTTGAATTA 900
operon1           TTACCCCACAATGTCGCCGCACCTGCCGGAATGATTGGTGCGAGTAACTTCTTTGAGTTG 900
                  ***** ****** * ** ************************ ** ** ** ***** **

operon2           TCGGTGGCTGTGGCTATATCGCTTTTCGGTTTGCAATCCGGTGCTGCATTGGCTACAGTG 960
operon1           GCGGTGGCGGTGGCTATTTCTCTTTTCGGCTTACAGTCCGGTGCTGCACTGGCTACCGTA 960
                   ******* ******** ** ******** ** ** ************ ******* **

operon2           GTAGGGGTGTTGGTTGAAGTTCCGGTTATGCTGATGTTAGTGAGAGTTGCAAATAATACA 1020
operon1           GTAGGCGTATTGGTTGAAGTTCCAGTAATGCTTATGTTAGTAAAAATAGCCAATAATACA 1020
                  ***** ** ************** ** ***** ******** * * * ** **********

operon2           TACAGGTGGTTTACAATAAAATCAAGTATGAGTAAATAG 1059
operon1           AGAAGTTGGTTTCC--TAAAGTAAAATAA---------- 1047
                     ** ****** *  **** * ** **
```

arsR

```
Operon1       ATGGGTCACCCGGCACGAATGGCAATTCTTAGTTTCTTAGCTAAACAAGAAAGTTGTTTC
Operon2       ATGGGACACCCGGCACGAATGGCTATTCTTAGTTTCTTGGCTAAACAGGAAAGTTGTTTC
              ***** ***************** ************** ******** ************

Operon1       TTCGGAGATATTCACGAAGTACTACCCATTGCTAAAGCAACCGTTTCGCAGCATTTAAAA
Operon2       TTCGGAGATATTCACGAAGAATTGCCTATTGCCAAAGCAACCGTTTCACAGCACTTGAAA
              ******************* * * ** ***** ************** ***** ** ***
```

```
Operon1       GAACTGAAAGATGCAGGGTTAATTCAGGGGGAAATAGAAACACCAAAAGTTCGGTATTGT
Operon2       GAGCTGAAAGATGCAGGATTAATTCAGGGAGAAATAGAAACGCCCAAGGTGCGATATTGT
              ** ************** ********** ********** ** ** ** ** ******

Operon1       ATCAACCGAGAGAATTGGGAACTTGCCCGTAAATTATTTGCTGCATTTTTGGGTGATTGT
Operon2       ATTAATAAGGAAAATTGGGAACTTGCCCGCAAATTATTTGCTGCATTTTTGGGAGATTGT
              ** **    ** ***************** ********************** ******

Operon1       AAATGTACGGGTACATCGTGCTGTGGATAA
Operon2       AAATGTACAGGAACATCGTGCTGTGGATAA
              ******** ** ******************
```

arsD

```
Operon1       ATGAAAACAATAGAAATTTTTGACCCGGCAATGTGTTGCCCTACGGGTTTGTGTGGAACT
Operon2       ATGAAAAAGATAGAAATTTTTGATCCGGCAATGTGTTGCCCTACGGGTTTGTGTGGAACA
              *******  ************** ********************************

Operon1       AATATTAATCCTGAATTAATGAGGATAGCCGTTGTTATCGAAACATTGAAAAGACAGGGT
Operon2       AATATCAATCCTGAATTGATGAGGGTTGCAGTTGTTGTCGAAACATTGAAAAGACAGGGT
              ***** ********** ****** * ** ****** **********************

Operon1       GTTGTTGTTACCCGTCATAATTTGCGGGACGAACCGCAAGTTTATGTAAGTAACAAGACT
Operon2       GTTATTGTTACCCGTCACAATTTACGGGATGAACCACAGGTGTATGTGAGCAATAAAACA
              *** ************* ***** ***** ***** ** ** ***** ** ** ** **

Operon1       GTAAACCAATATCTTCAAAAGAATGGTGCAGAAGCACTGCCTATTACTTTGGTGGATGGT
Operon2       GTAAATGAATATCTTCAAAAGAATGGTGCAGAAGCTTTACCTATTACTTTAGTAGATGGT
              *****  ***************************  * ********** ** ******

Operon1       GAAATTGCAGTTTCTAAAGATTATCCCACAACTAAACAAATGAGTGAATGGACAGGAATT
Operon2       GAAATTGCAGTTTCTAAAGTTTATCCTACCACTAAACAAATGAGTGAATGGACGGGAGTG
              ****************** ****** ** ********************** *** *

Operon1       AATTTAGACTTGATGCCCGTTAAATAA
Operon2       AATTTAGATTTAATGCCCGCTAAATAA
              ******** ** ******* *******
```