METAGENOMIC ANALYSIS OF MICROBIAL
COMMUNITIES IN THE BOVINE RUMEN

By

BIYAGAMAGE RUCHIKA FERNANDO

Bachelor of Veterinary Medicine and Animal Science
University of Peradeniya
Peradeniya
Sri Lanka
2003

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of

DOCTOR OF PHILOSOPHY

December, 2012

METAGENOMIC ANALYSIS OF MICROBIAL

COMMUNITIES IN THE BOVINE RUMEN

Dissertation Approved:

Dr. Udaya DeSilva

Dissertation Adviser

Dr. Clint Krehbiel

Dr. Raluca Mateescu

Dr. Andrew Doust (Outside Committee Member)

Dr. Sheryl A. Tucker

Dean of the Graduate College

TABLE OF CONTENTS

Chapter I
General Introduction ... 1
Section I: The rumen and its microbes ... 1
Section II: Gut microbes and metagenomics ... 28
Section III: DNA methylation and epigenetic regulation of gene expression in viruses ... 33
References ... 39
Chapter II
Double stranded DNA virome of the bovine rumen ... 56
Introduction ... 56
Materials and methods ... 59
Results ... 63
Discussion ... 66
References ... 85
Chapter III
Double stranded viral DNA methylome of the bovine rumen ... 88
Introduction ... 88
Materials and methods ... 92
Results ... 94
Discussion ... 96
References ... 99
Chapter IV
Metagenomic analysis of anaerobic fungal communities in the bovine rumen ... 102
Introduction ... 102
Materials and methods ... 107
Results ... 109
Discussion ... 110
References ... 115

LIST OF TABLES

Table 1. Major cattle producing countries in the world and their production as of 2010...3

Table 2. Summary of small ruminant and buffalo production in selected countries……..4

LIST OF FIGURES

Figure 1. Live animal numbers of major livestock species 3

Figure 2. Schematic diagram showing the life cycle of anaerobic fungi 16

Figure 3. Structure of 5-methylcytosine 35

Figure 4. Schematic diagram showing the linker mediated PCR amplification process . 71

Figure 5. Agarose gel electropherograms with viral DNA amplicons............................ 73

Figure 6. Transmission electron micrographs of negatively stained virus particles........ 75

Figure 7. Similarity of rumen viral sequences to the domains of life and viruses.......... 77

Figure 8. Taxonomic classification of the putative rumen viral sequences.................... 79

Figure 9. Biological processes associated with the bovine rumen virome 81

Figure 10. Molecular functions associated with the bovine rumen virome.................... 83

Figure 11. Linker mediated PCR amplification of CpG methylated DNA..................... 98

Figure 12. Anaerobic fungi specific degenerate PCR primers 112

Figure 13. Genomic diversity of the ITS-1 region of anaerobic fungi 113

Figure 14. Community structure and diversity of anaerobic fungi............................... 114

CHAPTER I

GENERAL INTRODUCTION

SECTION I: THE RUMEN AND ITS MICROBES

THE RUMINANT ANIMAL

The relationship between humans and ruminants dates back to prehistoric times where hunter-gatherers realized they could increase their food supply by domesticating ruminant animals and subsequent farming communities resulting in stable social communities (Russell and Rychlik 2001). Indeed, a large portion of our meat and almost all the milk we consume comes from domestic ruminants. The word ruminant comes from the Latin verb “ruminare” with the meaning of “to chew the cud” (Collins et al. 1986). Cud chewing, even-toed, hooved mammals, with a four-chambered stomach and classified under the suborder Ruminantia are called true ruminants (Dobson and Dobson 1986). There are about 150 species of extant ruminants (Hofmann 1989). They make a diverse array of ruminants with the body weight ranging from about 3 kg to over 1,000 kg.

Diets that ruminants consume are diverse. Ruminants are classified into three morphophysiological groups based on their feeding type: concentrate selectors (40%), intermediate feeders (35%) and grass and roughage eaters (25%) (Hoffmann et al. 1989).

Concentrate selectors have small rumens, poorly developed omasums, and large livers (e.g. roe deer). Intermediate feeders are seasonally adaptive and their feeding preference could be either browsing (e.g. goats) or grazing (e.g. sheep). Grass and roughage grazing species generally have larger rumens and are less selective (e.g. cattle). Even though there are over 150 species of ruminants mainly four species are domesticated (cattle, sheep, goat and buffalos) throughout the world, outnumbering all other ruminants. Figure 1 gives a summary of the livestock live animal numbers as of 2010. At present, undoubtedly cattle are the most important livestock species throughout the world. Sheep and goats account for the second and fourth highest animal numbers in production agriculture. In industrialized countries, sheep farming is more important than goat production. During the last few decades, the overall small ruminant production (milk, meat and wool) grew significantly in developing countries, but either stagnated or decreased in the developed world (Morand-Fehr and Boyazoglu 1999). Tables 1 and 2 summarize the contribution of ruminant animals to livestock production in some selected countries as of 2010. The goat and buffalo production systems play an important role in the rural economy of the developing countries. Their contribution in providing high quality protein to children in the rural communities is worth mentioning because it reduces the incidence of malnutrition.

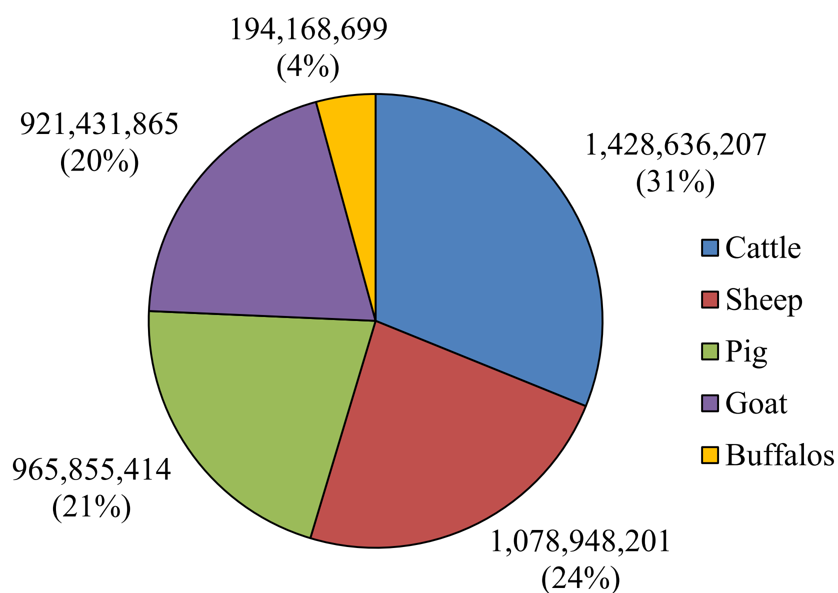

Figure 1. Live animal numbers of major livestock species

Source: FAOSTAT | © FAO Statistics Division 2012 | Accessed 28 March 2012

URL: http://faostat.fao.org/site/573/default.aspx#ancor

Production in 2010	Total cattle beginning stocks (in 1,000 head)	Calf crop (in 1,000 head)	Beef and veal (in 1,000 metric tons-carcass weight equivalent)	Liquid milk (in 1,000 metric tons)
India	316,400	61,700	2,842	50,300
Brazil	185,159	49,200	9,115	29,948
China	105,430	41,500	5,600	29,300
United States	93,881	35,685	12,047	87,461
EU-27	88,300	29,350	8,022	135,435
Argentina	49,057	11,600	2,620	10,600
Australia	27,907	8,040	2,087	9,327
Mexico	22,192	7,000	1,751	11,033
Russia	20,677	6,952	1,435	31,900
Canada	12,905	4,835	1,272	8,350

Table 1. Major cattle producing countries in the world and their production as of 2010

Source: United States Department of Agriculture, Foreign Agricultural Service

Country	Sheep meat (tons)	Sheep milk (tons)	Sheep skins (tons)	Goat meat (tons)	Goat milk (tons)	Buffalo meat (tons)
China	2,070,000	1,724,000	361,760	1,872,823	277,500	310,425
India	289,200		66,600	586,500	4,300,000	1,462,660
Nigeria	148,809		27,056	291,300		
Pakistan	158,000	36,000	44,079	275,000	739,000	760,000
Bangladesh	3,900	36,400	892	242,000	2,496,000	5,960
Sudan	349,000	452,100	64,150	159,900	1,601,900	
Iran (Islamic Republic of)	360,000	479,200	60,000	138,000	452,100	19,040
Turkey	259,000	816,832	45,920	36,600	272,811	990
South Africa	139,792		18,750	35,480		
Australia	555,616		115,466	25,000		
Morocco	139,000	37,700	19,440	23,000	57,600	
Kazakhstan	122,601	27,000	13,722	20,343	11,200	
Russian Federation	169,872	1,000	18,880	18,278	240,000	
Algeria	180,200	231,300	26,500	14,200	248,400	
France	121,861	259,240	11,800	12,053	645,176	
Turkmenistan	129,500		17,060	9,800		
Spain	131,231	585,190	19,200	9,000	602,000	
Syrian Arab Republic	198,000	643,000	22,743	8,800	139,000	330
Viet Nam				8,745		84,214
New Zealand	470,906		121,400	1,241		
United Kingdom	280,861		68,400			
U.S.A.	76,340		7,663			

Table 2. Summary of 2010 small ruminant and buffalo production in selected countries

Source: FAOSTAT | © FAO Statistics Division 2012 | Accessed 28 March 2012.

URL: http://faostat.fao.org/site/569/default.aspx#ancor

About 41% of the earth's land surface is covered by rangeland/dry land systems (Lund 2007). Ruminant animals beneficially digest and utilize this complex biomass by converting apparently indigestible carbohydrates and proteins to useful products with the help of rumen microbes. Rumen-inhabiting microorganisms produce energy and volatile

fatty acids (VFAs) from complex biomass such as pant cell walls. Further, they *de novo* synthesize amino acids using ammonia and convert them to microbial proteins, which are later available to the host animal. Compounds such as oxalates and mimosine are detoxified in the rumen with the help of the microorganisms (Russell and Wilson 1988). Certain vitamins such as components of the vitamin B complex are produced by rumen microorganisms (McElroy and Goss 1940).

Ruminants cannot be considered “primitive” even though there are inefficiencies in breaking down plant material inside the rumen (Hofmann 1989). Typically, cattle would waste about 6% of the ingested energy as methane, which is a greenhouse gas (Johnson and Johnson 1995). High producing ruminant animals (e.g. dairy) are susceptible to metabolic disorders such as acidosis and ketosis (Herdt 2000; Kleen et al. 2003). Some rumen bacteria can produce toxins such as 3-methylindole in the rumen which may be harmful to the host animal (Carlson et al. 1972).

ANATOMY OF THE BOVINE RUMEN

Cattle have a complex stomach with four compartments referred to as the rumen, reticulum, omasum, and abomasum. These divisions can be seen clearly on the external surface as furrows. The size of the four-chambered stomach varies depending on the age and the size of the animal. The capacity of the complex stomach in an adult of a large breed is about 40-60 gallons (Sisson 1910). The right side of the reticulo-rumen is irregular and rests on the liver, left kidney, omasum, abomasum, pancreas, aorta, and posterior vena cava. Peritoneum and connective tissues firmly attach this organ to the diaphragm and sublumbar muscles. The reticulum rests against the diaphragm on the left side of the median plane and is connected to the omasum by a short neck. The reticulo-ruminal fold is the tissue mass that divides the rumen and reticulum into two compartments. The main functions of the reticulo-ruminal fold are mixing of digesta, sorting of particles and preventing fluid reaching the cardia during eructation. The pillars are highly musculated areas of the rumen wall that correspond with the grooves on the external surface. The grooves contain blood vessels, lymphatics, and nerves, and the pillars are important for rumen motility. The longitudinal pillar divides the rumen into dorsal and ventral sacs. The dorsal blind sac is separated from the dorsal sac by the dorsal coronary pillar and similarly the ventral blind sac is separated by the ventral coronary pillar. The cranial blind sac is separated from the rest of the rumen by the cranial pillar. The omasum is the third compartment and is an ellipsoidal muscular structure. There are numerous mucosal folds on the internal surface of the omasum. The abomasum represents the glandular stomach of non-ruminants animals.

A non-secretory stratified squamous epithelium covers the internal surface of the rumen and reticulum. This epithelium consists of, from outside to inside, a basement membrane, and four cell layers called stratum basale, stratum spinosum, stratum granulosum, and stratum corneum (Frandson et al. 2009). Stratum basale consist of actively multiplying cuboidal cells. As they are pushed away from the basement membrane they become flattened, tough, lifeless, and constantly get sloughed off. The most superficial layer is the stratum corneum, which consists of keratinized dead cells. The internal surface of the reticulum has a honey comb (4-6 sided small epithelial folds) appearance (Sisson 1910). Small vestigial papillae are present in the honey comb structures. The reticular groove is a short passage from the esophagus to omasum lined by two lips. Inside the rumen, there are numerous papillae of various sizes and shapes. These papillae are well-developed in the ventral sac (about 1 cm in length), the floor of the caudodorsal blind sac, and the cranial sac. These papillae dramatically increase the absorptive surface of the rumen. There are numerous folds inside the omasum. The abomasum has a true glandular epithelium. This epithelium secretes pepsinogen (by chief cells), HCl (by parietal cells) and the hormones gastrin and histamine (Frandson et al. 2009).

Blood supply to the ruminant stomach is by the common hepatic, right and left ruminal, and omaso-abomasal arteries (Sisson 1910). The organ is drained by the reticular, right and left ruminal, and omaso-abomasal veins (Trotter and Lumb 1958).The vagus and splanchnic nerves innervate the complex stomach (Ruckebusch 1988).

RUMEN DEVELOPMENT

Differentiation of primordial gut tissue begins around the fourth week of embryo development (Sidney and J.R.Lyford 1988). As the gut enlarges, well-defined dorsal and ventral curvatures become visible in the stomach as in simple stomached animals. On about the eighth week, compartments of the digestive system are formed including the ventricular groove. Around the third month, four compartments of equal size are formed. Rumen sacs and pillars become apparent. By four to five months, the relative proportion of the stomach is similar to that of an adult. At this time ruminal papillae become detectable. Form here onwards until birth, the abomasum continues to grow. At birth, the abomasum is about 60% of the total weight of the complex stomach (Short 1964).

During the postnatal period, rumen growth continues and the relative size of the abomasum decreases. There is an innate potential for rumen to develop, but full development requires dry feed, which provides necessary physical and chemical stimulation. Roughage in the diet physically stimulates the muscle contractions in the rumen making them stronger (Tamate et al. 1962). Fermentation of concentrates provides VFAs especially butyrate, which is important for rumen epithelial tissue growth and papillae development (Coverdale et al. 2004). It takes about four months for a fully functional rumen to develop (Sidney and J.R.Lyford 1988).

At birth, the rumen is sterile and bacteria start colonizing the rumen as early as 38 hours after birth (Cheng et al. 1991). It takes approximately 8-13 weeks to establish the normal microbial flora (Bryant et al. 1958; Mueller et al. 1984). During the first week of life, *E.coli, Clostridium* and *Streptococcus* are predominant in the rumen. There are

reports of finding cellulolytic and methanogenic bacteria in the rumen during the first week of life (Bryant et al. 1958; Fonty et al. 1986) and it takes about six weeks to be fully established (Mueller et al. 1984). By two to three weeks, lactic acid producing and utilizing bacteria settle in the rumen (Ziolecki and Briggs 1961). Protozoa and anaerobic fungi begin colonizing in the rumen around two to three weeks (Bryant et al. 1958; Eadie et al. 1959; Fonty et al. 1987; Stewart et al. 1988). Animal to animal contact, a favorable rumen environment for microbes (CO_2, NH_3, H_2, branched chain VFAs, vitamin B, pH 6-7), and a good digesta turnover are vital for the establishment of a balanced and diverse microbiome (Church 1988).

THE RUMEN ECOSYSTEM

The rumen can be considered a large fermentation vat**.** The rumen is open to the external environment, and there is a continuous flow of material into and out of the rumen. However, the rumen environment is strictly anaerobic (-150 to -350 mv) and tightly regulated (Clarke 1977). Most studies have reported a dry matter content of 10-13% in the rumen (Dehority 2003). The rumen temperature is typically regulated between 38-40° C in normally fed animals (Church 1969; Clarke 1977). The physiological pH range is between 5.5 and 7.0 (Dehority 2003), and it is one of the most variable factors in the rumen environment (Church 1969; Clarke 1977). However, the buffering capacity of the rumen content is quite high (Hungate 1966; Church 1969). Osmotic pressure of the rumen contents before feeding is hypotonic (250 mOsm/kg) with respect to plasma (300 mOsm/kg), and after feeding it rises up to about 400 mOsm/kg, then gradually (in 8-10 hrs.) decreases to its hypotonic state (Warner and Stacy 1965).

Rumen contractions help continuous mixing of the rumen content. The end-products of rumen fermentation include VFAs (mainly acetate, propionate and butyrate) and gases [CO_2 (65%) and CH_4 (27%)] (Hungate 1966; Clarke 1977). Volatile fatty acids are the main source of carbon and energy to the animal. This fermentation process results in many intermediate products such as lactate, succinate, formate, H_2, and NH_3. The steady supply of food and constant removal of digested feed material and end-products along with the above mentioned rumen conditions allow a dense population of microorganisms to grow inside the rumen (Hungate 1966). There is a symbiotic relationship between the rumen inhabiting microorganisms and the host animal. The host animal provides the substrates to the microbes by consuming forage, which is then digested and fermented by the rumen microbes to form VFAs, CO_2 and CH_4 (Hungate 1966). The complex rumen ecosystem consists of bacteria, fungi, protozoa, and viruses.

Bacteria and other microorganisms can be classified based on their micro-environment within the rumen. Some bacteria (i.e. facultative anaerobes) and protozoa (i.e. holotrichs) are closely associated with the rumen wall (Ellis et al. 1989; McAllister et al. 1994; Dehority 2003). The rumen wall-associated bacteria scavenge oxygen and breakdown urea into CO_2 and NH_3 that enter into the rumen via feed, saliva and blood (McAllister et al. 1994). Further, some of them degrade the sloughing off epithelial cells. NH_3 is a source of nitrogen to the rumen bacteria, while CO_2 is an essential growth factor for fiber digesting bacteria. The extramural bacteria (99% of the rumen bacteria) can be either free-floating, attached to protozoa (symbiotic methanogens) or feed particles (Koike and Kobayashi 2009). About 75% of the bacteria are associated with feed particles (Craig et al. 1987). Out of the feed particle associated bacteria, approximately

80-90% are responsible for fiber digestion, 75% for protein digestion, and 70% for starch digestion (Brock et al. 1982; Minato et al. 1993). The free-floating bacteria or bacteria associated with liquid material inside the rumen are about 30% of the total bacteria inside the rumen (Legay-Carmier and Bauchart 1989). They can either become attached to the feed particles or simply washed away into the lower gastrointestinal tract (G.I.T) and then become available as a good source of protein for the host animal.

BACTERIA AND ARCHAEA

Bacterial role in fermentation of plant material became well-known after the work of Louis Pasteur in 1863. In 1879, Zuntz inferred that the rumen microorganisms ferment plant material anaerobically forming gas and acids (Hungate 1966). The rumen ecosystem allows the selection of specific and suitable microbial types (Hungate 1966). There is a dense and diverse population of bacteria inside the rumen. An estimated amount of 10^9-10^{11} bacterial cells are present in a gram of rumen content or milliliter of rumen fluid (Hobson et al. 1958; Hungate 1966; Miron et al. 2001). Ruminants consume complex feed material containing carbohydrates, proteins, lipids, organic compounds, and minerals. The rumen contains a variety of highly specialized bacteria: bacteria capable of utilizing specific substrates, bacteria capable of utilizing a variety of substrates, and intermediate ones (Hungate 1966). In an environment with a high turnover rate, the efficiency of converting the rumen plant material into cell yield determines the survival rate of microbes. Quite possibly, the fermentation of carbohydrates to VFAs, CO_2 and CH_4 gives the highest cell yields than any other biochemical pathways for carbohydrate fermentation (Hungate 1966). Among the theoretically possible innumerous biochemical pathways resulting in the above mentioned end products (different VFAs,

CO_2 and CH_4), some combinations yield more cells than others. These combined pathways need a diverse set of enzymes. A single bacterial cell has a limited biocatalytic capacity. Therefore, the maximum total growth requires a complex and a diverse population of cells. Rumen bacteria are fastidious and difficult to grow under laboratory conditions (Moir and Masson 1952; Hungate 1966). Traditionally rumen bacteria have been classified based on morphology, energy source, fermentation end products, sensitivity to and requirement of oxygen, and ecological niches occupied (Yokoyama and Johnson 1988). In forage-fed animals, a vast majority of the rumen bacteria are Gram-negative and obligate anaerobes (Hungate 1966). Among rumen bacteria, cocci and short rods of various sizes are typically most abundant (Hungate 1966).

Many rumen bacteria have been isolated over the last 60 years by means of enrichment and habitat-stimulating techniques. There are specialized plant cell wall degraders [*Ruminococcus flavefaciens* (Sijpesteijn 1951), *Fibrobacter* (formerly *Bacteroides*) *succinogenes* (Hungate 1950)], starch digesters [*Streptococcus bovis* (Hungate et al. 1952)*, Ruminobacter amylophilus* (Hamlin and Hungate 1956)], protein degraders [*Prevotella ruminicola* (Yokoyama and Johnson 1988), *Streptococcus bovis* (Russell et al. 1981), lipid digesters [*Anaerovibrio lipolytica* (Yokoyama and Johnson 1988) and lactate utilizers [*Megasphaera* (formerly *Peptostreptococcus*) *elsdenii* (Elsden et al. 1956), *Veillonella alcalesce* (Johns 1951)]. However, most rumen bacteria depend on carbohydrates as the main source of energy. Many species produce small amounts of proteolytic enzymes, which contribute to the extensive proteolytic activity in the rumen (Blackburn and Hobson 1960). The rumen environment is important for growth of specific bacteria. Most cellulolytic bacteria prefer a pH range of 6-7, NH_3, branched

chain VFAs (n-valeric, isobutyric, 2-methylbutyric acid or isovaleric acid), phenolic acids and CO_2 (Chesson et al. 1982; Yokoyama and Johnson 1988). The major fermentation end-products of cellulolytic bacteria are acetate, formate, succinate, CO_2, H_2 and ethanol (Hungate 1966). The amylolytic bacteria prefer a more acidic pH (5-6), and their major end-products are acetate and propionate (Owens and Goetsch 1988; Russell and Rychlik 2001). Very few bacteria only use protein as their sole energy source (Yokoyama and Johnson 1988). About 38% of the isolates are proteolytic. Lipolytic bacteria hydrolyze triglycerides and phospholipids releasing free fatty acids and glycerol (Yokoyama and Johnson 1988). Methane producing bacteria in the rumen regulate the overall rumen fermentation by efficient removal of H_2. This removal of H_2 by reducing CO_2 to CH_4 is important for the growth of other rumen microbes and efficient fermentation of substrates (Yokoyama and Johnson 1988). These methanogenic bacteria are either free-living or associated with protozoa.

Numerous experiments have shown the importance of bacterial adhesion in subsequent plant cell wall degradation (Miron et al. 2001). Initially, rumen bacteria attach to feed particles nonspecifically by hydrophobic, ionic, electrostatic, or by other types of bonds. Then, the bacteria start to colonize and bind specifically to the feed materials. The specific binding occurs through ligands or adhesins on the bacterial cell surface binding to the receptors on the substrate (Pell and Schofield 1993). These adhesins can be polycellulosome complexes, fimbriae, glycocalyx, cellulose binding proteins, and enzyme binding domains (Pell and Schofield 1993; Morrison and Miron 2000). Then, the bacteria get proliferated and make colonies on the potentially digestible sites of the feed particles. Therefore, the attached bacteria are very important for complex plant material

digestion inside the rumen (Gong and Forsberg 1989). Bacterial attachment enhances efficiency of degradation of plant material by the enzymes present on the cell wall or the surface, and facilitates the uptake of hydrolytic products. Further, bacteria attached to the larger feed particles resist passage to the lower G.I.T., and become less likely to be engulfed by protozoa.

It is generally accepted that there are about ten times more bacteriophages than bacteria in natural environments (Whitman et al. 1998). Therefore, the bacteria number and diversity is a key factor that governs the bacteriophages inside the rumen ecosystem and their diversity, which subsequently affects the bacterial diversity and numbers. This predator-prey relationship is in a fine balance where both the groups survive and continue to ferment complex biomass in the rumen. Bacteriophages are considered a critical factor that drives the evolution of bacteria and maintains the fine balance of bacterial community structure in the gut (Breitbart et al. 2003). However, very little or no information is available on the role of the rumen virome in shaping the evolution of bacterial populations and thereby rumen fermentation (Berg Miller et al. 2012).

ANAEROBIC FUNGI

Anaerobic fungi (AF) were first described by Liebetanz (1910), but were thought to be zooflagellates for many decades. Braune (1913) named such an organism as *Callimastix frontalis* due to its similarity to *Callimastix cyclopsis*. Later, Vavraj and Joyon (1966) found that *Callimastix cyclopsis* is likely a fungus related to the phylum Chytridiomycota. Therefore, *Callimastix frontalis* was assigned a new genus of zooflagellates called *Neocallimastix*. Warner (1966) observed that the flagellate protozoa

Neocallimastix frontalis increased rapidly even after one hour of feeding followed by a rapid death and lysis. Further, he estimated such a marked rise in cell numbers should result from 2-5 cycles of cell division within an hour of feeding. This result appeared unlikely to Warner and explained his observation as protozoa probably are sequestered on the rumen wall and are migrating back and forth due to chemotactic responses. In a subsequent experiment conducted by Orpin (1975), he found that this rapid rise in *Neocallimastix* following feeding is due to a release of flagellates (zoospores) by a reproductive body on a vegetative phase of the organism making the discovery of AF.

Rumen fungi are an integral part of the rumen microbiome, although very little is understood on their role and diversity. Rumen fungi can be broadly categorized into two groups based on their energy metabolism; facultative anaerobes and anaerobes. Anaerobic rumen fungi were originally classified as Chytridiomycetes, based on their thallus morphology and the presence of chitin in their cell walls (Orpin 1976; Orpin 1977). AF produce special flagellated structures called zoospores. They lack mitochondria, and energy is generated in an organelle called hydrogenosome. Hydrogenosomes are amorphous globules enclosed by a single membrane with an electron-dense matrix (Yarlett et al. 1986). Hydrogenosomes are in close association with the microtubules radiating into the cytoplasm kinetosome/basal body of the flagella in zoospores. *In vitro* studies have shown their growth is optimum between pH 6-7.5 (Orpin 1975). AF have now been identified from the alimentary tracts of a wide range of herbivores, and so far, they have not been isolated from any other anaerobic terrestrial or aquatic habitats (Bauchop 1989).

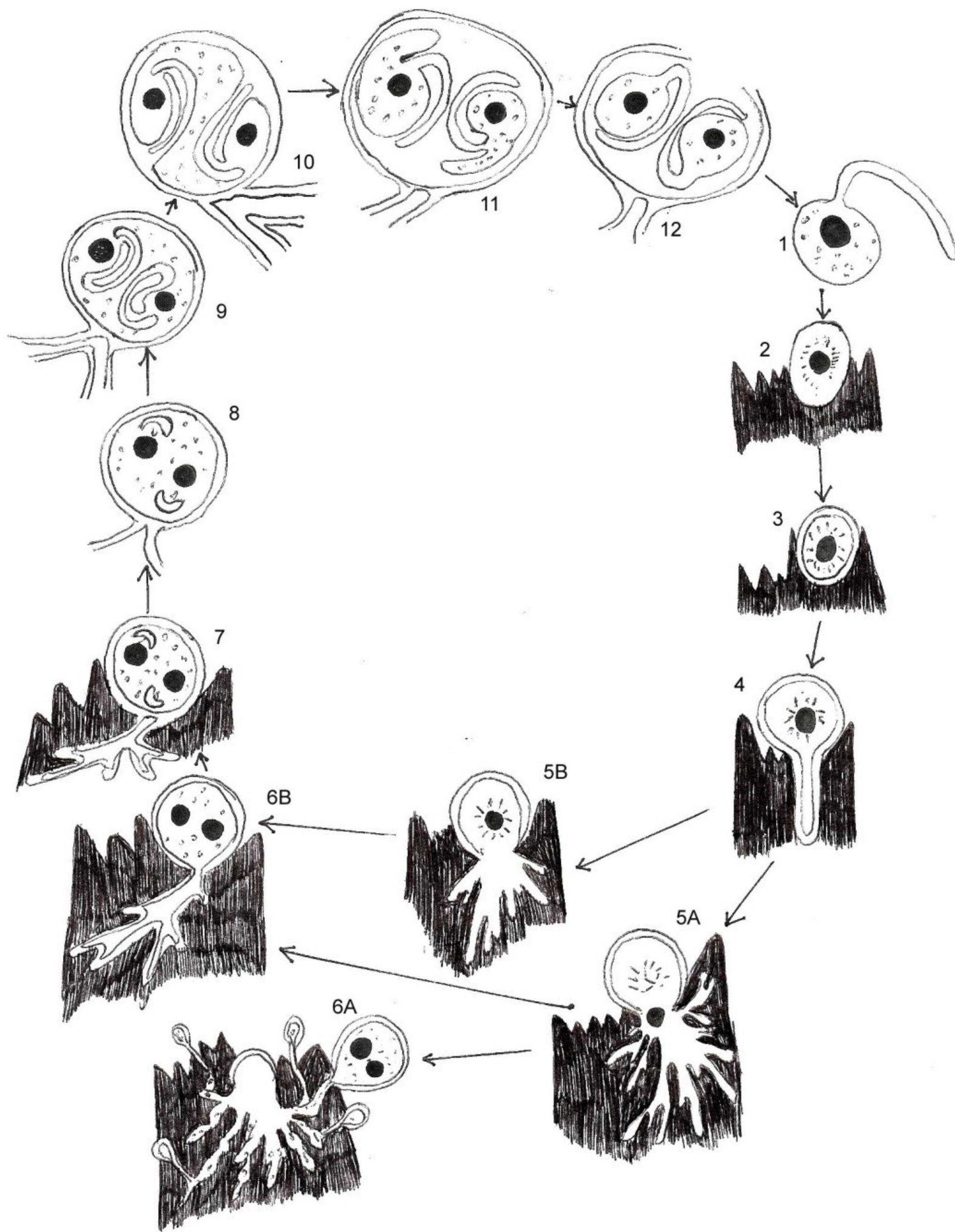

Figure 2. Schematic diagram showing the life cycle of anaerobic fungi

Figure 2. Schematic diagram showing the life cycle of anaerobic fungi. AF consist of two forms in their lifecycle; a motile zoospore (Fig.2-1) and a non-motile vegetative form called the thallus (thallus includes rhizoids and the reproductive body called sporangium with its developing zoospores) (Dehority 2003). In AF, the zoospores chemotactically attract to fresh feed particles and attach on to the feed particles while losing their flagella and becoming encysted (Fig.2-2 and 2-3). The encysted zoospore develops a rhizoid into the feed particle (Fig.2-4) initiating the vegetative growth phase with an extensive rhizoid system or rhizomycelium. In certain fungi, the nucleus migrates out of the cyst (Fig.2-5A) into the rhizoid system. Once the nucleus is migrated out from the cyst, it results in a single sporangium in the exogenous monocentric type (Fig.2-6B). In exogenous polycentric type (Fig.2-6A; one sporangium is drawn enlarged to show inside features), multiple sporangia are formed after the nuclear division. In endogenous monocentric fungi, the nucleus remains inside the cyst and develops into the future sporangium (Fig.2-5B). Two or more nuclei are produced by mitotic nuclear divisions (Fig.2-6B). The developing sporangium/sporangia get separated from the rest of the rhizomycelium (Fig.2-7). The precursor flagella cleavage vesicles appear near the kinetosomes, and subsequently become elongated forming flagella (Fig.2-8, 9). The surface layers coating the zoospore begin to polymerize. Zoospores detach from the sporangium wall and flagella vesicles fuse with the peripheral plasma membrane of the zoospore (Fig.2-10). Once the zoospore differentiation is completed, the sporangium wall is digested releasing zoospores (Fig.2-12). This diagram is adapted from several sources (Munn et al. 1988; Mountfort and Orpin 1994; Dehority 2003).

AF are generally reported in lower numbers (10^3-10^5 per milliliter of rumen fluid) than prokaryotes and protozoa (Mountfort and Orpin 1994) based on the zoospore count. It is generally accepted that their contribution for providing nutrients to the ruminant host is less significant than other rumen microbes (Griffith et al. 2009). However, AF are important colonizers of the complex feed material and are active during the early phase of fermentation (Theodorou et al. 1992). AF play a significant role in fiber digestion in the rumen (Ho and Barr 1995; Orpin and Joblin 1997). They can penetrate the cuticle, the rigid structural barrier on the outside of the plant epidermis, as a result of their filamentous growth and cutinase activity. AF have been shown to digest and degrade lignified plant material (Lowe et al. 1987; McSweeney et al. 1994). These functions do not appear to occur with other rumen microorganisms, thus giving AF an additional advantage in degrading plant fiber. AF have a range of highly active enzymes (cellulases, xylanase, acetylesterases etc.), and they are the only known microorganisms that produce exo-acting cellulases (Forsberg et al. 1997). Also, AF act synergistically with rumen bacteria to digest complex plant material in the rumen (Dehority and Tirabasso 2000). Among the species of AF, *Neocallimastix* are the most studied organisms and are found to digest crystalline cellulose rapidly (by the action of endo-β-1, 4-Glucanase and exo-β-1, 4-Glucanase) (Mountfort and Orpin 1994). Furthermore, AF are prospective candidates to be used in biofuel production from lignocellulosic biomass.

It has been thought that these AF have evolved from free-living chytridiomycetes to survive the anaerobic conditions in the digestive tract of their herbivorous hosts (Heath 1988). Based on extensive molecular phylogenetic analyses, AF are now classified as Neocallimastigomycetes under the new phylum Neocallimastigomycota (Hibbett et al.

2007). AF have been formally classified into six genera; *Neocallimastix*, *Piromyces, Caecomyces*, *Anaeromyces, Orpinomyces* and *Cyllamyces* based on the morphology of their thalli, number of flagella present in zoospores, and development. So far, over twenty species have been named and described under these genera (Griffith et al. 2009). These species have been recognized according to the taxonomic convention by the ultrastructure of zoospores (Heath et al. 1983; Orpin and Munn 1986; Webb and Theodorou 1991). However, the validity of this method has been debated due to occurrence of natural morphological variants within the same species of AF (Ho and Barr 1995). Age, method, and quality of preparation for identification are some factors which could adversely affect the morphology based classification. Initially, zoospore counts were used to enumerate AF (Orpin 1975). In polycentric species, new rhizobia are formed from the old rhizomycelium in addition to zoospores. Therefore, zoospore count does not infer the number of thallus-forming units accurately (Hespell et al. 1997).

It has been almost forty years since the discovery of anaerobic fungi, but very little is understood on their community structure and role inside the gut due to unavailability of convenient and reliable methods for accurate identification, differentiation and enumeration of anaerobic fungi (Orpin 1994; Brookman et al. 2000). However, over the past few decades, various methods have been applied in identifying and/or quantifying rumen fungi with varying success. Most Probable Number (MPN) is a commonly used enumeration technique in microbiological work (McCrady 1915) and has been adapted to enumerate AF (Theodorou et al. 1990). In this method the samples are serially diluted until no thallus-forming units are found in the latter dilutions and incubate at 39°C under anaerobic conditions (Griffith et al. 2009). Later, the tubes with fungal

growth/acidity are enumerated. Since this is a culture based method, it has the tendency to under estimate the true population especially with very fastidious organisms. In order to overcome the limitation of culturing, chitin assays have been used to quantify rumen fungal biomass (Gordon and Phillips 1989; Gay 1991). This assay has been reported to have limited usage in *in vivo* experiments (Rezaeian et al. 2004) and lower extraction efficiencies with certain substrates (Davies 1991). The greatest promise lies with the metagenomic approaches in identifying and quantifying rumen AF (Griffith et al. 2009).

PROTOZOA

Protozoa are relative large unicellular organisms with a cuticle (skin) and a digestive tract with a mouth and an anus. Therefore, protozoa can be considered as the simplest form of animal life (Dehority 2003). They have a macronucleus, a micronucleus and contractile vacuoles. The contractile vacuoles may be functioning in excretion of gas and liquid waste products (Hungate 1978). They reproduce by binary fission (Warner 1966). Protozoa are the first microorganism discovered from the rumen in 1843 by Gurby and Delafond (Hungate 1966). Further, Gurby and Delafond have link the ability of the ruminant animal to survive exclusively on plant material to existence of protozoa in the rumen. There are about 10^4 to 10^6 protozoal cells per gram of rumen content (Dehority 2003). Animal to animal contact and favorable rumen conditions are important for establishment of ciliate protozoal population in the rumen (Hungate 1966; Dehority 2003). Protozoal numbers increase in the rumen as the amount of energy available in the diet increases (Nakamura and Kanegasaki 1969; Grubb and Dehority 1975). Also, a drop in rumen pH, as in acidosis resulted from high concentrate diets, would eliminate protozoa (Bryant and Small 1960; Eadie 1962). It is generally accepted that feed with 40-

60% concentrates will result in highest number of protozoa with the greatest diversity (Dehority 2003). The majority of the rumen protozoa are ciliates. However, flagellates species are also present in the rumen. The ciliate protozoa come under the subclass Trichostomata according to the latest classification (Lee et al. 1985). Four of the families of protozoa named Isotrichidae, Paraisotrichidae, Blepharocorythidae and Buetschliidae are classified under Trichostomata, which were originally classified under the subclass Holotricha, and were commonly referred to as “holotrichs” in scientific literature (Dehority 2003). The remaining ciliated protozoa are classified under the family Ophryoscolecidae according to Lee *et al.* (1985), and were earlier called “oligotrichs,” and more recently, called “entodiniomorphs.” Holotrichs and entodiniomorphs account for approximately 10% and 90% of rumen ciliated protozoa, respectively (Ankrah et al. 1990). *Entodinium* is the most common protozoal genus found in domestic ruminants (Ibrahim et al. 1970; Dearth et al. 1974). By evolution, rumen ciliates have become highly specialized for the rumen ecosystem (Dehority 2003). There have been numerous genera and species described under the previously mentioned families based on the cellular morphology. The morphological criteria used in classification include: number of skeletal plates, number of contractile vacuoles, the number and size of ciliary zones, location of the nuclear apparatus, variation in cytoalimentary system, and overall body size (Dogiel 1947; Lubinsky 1957a; Lubinsky 1957b; Furness and Butler 1988). Protozoa can ferment plant material for cellular energy synergistically with the billions of other microorganisms found in the rumen (Hungate 1966). Before the discovery of rumen AF in 1977 by Orpin, the reports on flagellate protozoa were confounded by the coexistence of fungal zoospores (Dehority 2003). So far, seven species of flagellate protozoa have

been isolated from the rumen. These species belong to five genera: *Monocercomonas, Monocercomonoides, Chilomastix, Tetratrichomonas,* and *Pentatrichomonas* (Ogimoto and Imai 1981).

Protozoa engulf bacteria and therefore control bacterial numbers in the rumen (Coleman and Hall 1969). Furthermore, protozoa appear to require the coexistence of live bacteria in the medium for survival (Fondevila and Dehority 2001). Entodiniomorph species are capable of digesting all major plant polysaccharides (i.e. starch, pectin, hemicellulose and cellulose), and isotrichid species primarily utilize soluble sugars as a source of energy (Dehority 2003). The major end-products of protozoal fermentation are acetate, butyrate, propionate, H_2, CO_2 and lactate. Protozoa play an important role in animals fed with a low-protein diet or undergoing short periods of starvation (Yokoyama and Johnson 1988). Protozoa convert bacterial proteins to protozoal proteins by engulfing bacteria, and they cannot efficiently utilize NH_3 as a source of nitrogen. Protozoa are relatively resistant to the passage through the G.I.T., and dead protozoa become a source of nitrogen to the rumen bacteria. This recycling of microbes may be advantageous for animals that are fed low-protein diets. However, they can be a source of microbial cell proteins for the animal at lower G.I.T. Protozoa also have storage polysaccharides, which stabilize rumen fermentation when the host animal is starving.

VIRUSES

An estimated number of 10^{31} virus-like particles are present on earth (Rohwer et al. 2009). Most of these viruses are thought to be bacteriophages that infect prokaryotes, which outcompete their host by at least one order of magnitude (Krupovic et al. 2011). Undoubtedly, bacteriophages are the most abundant form of life on earth (Regenmortel et al. 2000). Based on the half-life of free viruses (~48h), it has been estimated that about 10^{27} viruses are produced every minute eventually destroying 10^{25} microbes, or about 100 million metric tons of cell mass every minute (Rohwer et al. 2009). Therefore, these natural predators of bacteria have an immense impact on bacterial population densities and community structure (Wilhelm and Suttle 1999; Rohwer et al. 2009). Hence, these predators of bacteria have a profound effect on all bacteria mediated processes such as nutrient and biogeochemical cycling, horizontal gene transfer, antimicrobial resistance, and bacterial virulence (Suttle 1994; Wilhelm and Suttle 1999; Casas and Rohwer 2007; Roberts et al. 2008; Rohwer et al. 2009; Tap et al. 2009).Viral predation along with protist grazing keeps the bacterial numbers below the carrying capacity of any ecosystem (Rohwer et al. 2009).

The ecological role of bacteriophages and other viruses has been mainly studied in aquatic systems (Rohwer et al. 2009) and in the human gut (Kim et al. 2011b). The ocean and lake waters contain an estimated number of 10^6 phage particles per milliliter and 10^9 phage particles per gram of sediment (Maranger and Bird 1996; Danovaro et al. 2001; Hewson et al. 2001). Several studies have screened the viral diversity in the human gut using a limited number of human subjects by means of shotgun sequencing (Breitbart et al. 2003; Breitbart et al. 2008; Kim et al. 2011b). Phages are thought to be important in

homeostasis of the gut (Qin et al. 2010). In the recent past, it has been recognized that the animal gut inhabiting microbial populations play a significant role in pathogenesis of certain metabolic and inflammatory diseases such as diabetes, obesity, and inflammatory bowel disease (Turnbaugh et al. 2006; Frank et al. 2007; Wen et al. 2008). Based on the recent research conducted on gut associated viromes, the phage-microbial host relationship has been described as a complex association rather than a simple predator-prey relationship commonly observed in many other ecosystems (Reyes et al. 2010). Furthermore, the diversity and the unexplored component of the fecal virome are increasingly recognized in recent studies emphasizing the importance of such knowledge (Kim et al. 2011).

In terms of numbers and diversity, the bacteriophage populations inhabiting the rumen of cattle are not any lesser than the above mentioned environments. It was estimated that the bacteriophage count in a milliliter of rumen fluid is between 3×10^9 and 1.6×10^{10} particles (Klieve and Swain 1993). Rumen bacteriophages play an important role in population dynamics of rumen bacteria, and they cause a specific and considerable lysis of rumen bacteria (Klieve and Swain 1993). Lysis of crucial bacteria involved in metabolic pathways inside the rumen could result in altered foregut fermentation and a change in total animal performance. However, very little is known about the diversity of bacteriophages inside the rumen and their effects on bacteria. To the best of our knowledge, there is only one previous study reporting on the rumen virome and its diversity based on high throughput sequencing of virus enriched samples (Berg Miller et al. 2012).

According to present knowledge, bacteriophages make the largest group of viruses on earth (Regenmortel et al. 2000). As of now, viruses have been classified into three orders, 61 families, and 241 genera (King et al. 2012). In 1991, the International Committee on Taxonomy of Viruses (ICTV) endorsed that "a virus species is a polythetic class of viruses that constitute a replicating lineage and occupy a particular ecological niche" (Regenmortel et al. 2000). Therefore, phages are also classified based on a "polythetic species concept", meaning that each species is defined by a number of properties, all of which may be not found in every member in the group. A vast majority of the phages contain double stranded DNA. However, there are single stranded DNA, single stranded RNA, and double stranded RNA phages. Double stranded DNA phages consist of 13 families and the genus *Slaterprovirus*. These 13 families are: Myoviridae, Siphoviridae, Podoviridae, Lipothrixviridae, Plasmaviridae, Corticoviridae, Fuselloviridae, Ampullaviridae, Globuloviridae, Guttaviridae, Tectiviridae, Rudiviridae, and Bicaudaviridae (King et al. 2012). Single stranded DNA phages consist of two families; Microviridae and Inoviridae. Double stranded RNA phages belong to the family Cystoviridae and single stranded RNA phages belong to Leviviridae family.

The tailed bacteriophages are the largest group of bacteriophages, and account for 96% of the reported bacterial viruses. They are classified under the order of Caudovirales, which consists of three families; Myoviridae (24%), Siphoviridae (62%) and Podoviridae (14%) (King et al. 2012). Members of the family Myoviridae have long contractile tails, Siphoviridae have long non-contractile tails, and Podoviridae have short contractile tails. The virion contains a single molecule of linear, double stranded DNA. Tailed bacteriophages are known to be parasites of both bacteria and archaea. There are

enough phylogenetic and taxonomic features left among the members of Caudovirales, though many of the above features have been obscured by great evolutionary age, extremely large populations, and horizontal gene transfer. It has been suggested that the tailed phages constitute a monophyletic evolutionary group based on many morphological and physiological characteristics. However, bacteriophages have distinguishing characteristics: differences in DNA content and composition (ranging in size from 17 to over 700 kbp), tail length variations (ranging from 10 to 800 nm), and also differences in their fine structure, and physiology (Ackermann 2003).

The virion has a head/capsid and a tail. Virion heads can be icosahedral in shape, but some can be elongated derivatives of it (Regenmortel et al. 2000). Tails are either helical or consist of stacked disks. Tails often carry structures such as spikes, base plates or fibers, which aid in attachment. Phage DNA may contain modified bases such as 5-hydroxymethylcytosine, can be glycosylated, or associated with internal proteins. The linear genomes have an estimated number of genes between 27 to over 200. These genes seem to be arranged in interchangeable blocks or "modules". Replicating DNA forms concatemers (large branched intermediates). Then, the DNA is cut into unit size and combined with the bacteriophage heads. Virion particles are assembled through a complex process comprised of separate pathways for heads, tails, and tail fibers. Maturation completes with the addition of a matured, DNA-filled capsid to a protein tail.

A high genotypic diversity of single stranded DNA viruses, particularly microphages, have been found in many environments ranging from rice paddy soil (Kim et al. 2008) to seawater (Angly et al. 2006), to Antarctic lakes (López-Bueno et al. 2009)

using multiple-displacement amplification with phi29 polymerase (Edwards and Rohwer 2005). However, there is very little information available about their hosts (Rebecca Vega 2009). Single stranded DNA viruses, mainly the members of family Microviridae, have been identified in human feces. *Bacteroides* and *Prevotella* are thought to be the hosts of microphages in the human gut. Single stranded DNA viruses such as *nanovirus*, *circovirus* or *parvovirus* have been not detected in the human gut (Kim et al. 2011).

In the recent past, there has been an increase in the attention on the role of prophage elements in their host genomes as they were identified as having beneficial effects to the hosts' phenotypes (Canchaya et al. 2003; Wang et al. 2010; Roossinck 2011). The genes related to bacterial metabolism found in the bacteriophages might have originated from gut bacteria as lysogenic conversion genes, which could be potentially transferred to other bacteria present in the same niche. (Canchaya et al. 2003). Based on these findings, the bacteriophages in gut might facilitate transient gut bacteria to colonize and adapt to the gut microbial ecosystem (Kim et al. 2011). As a result, the conventional thinking of viruses and their microbial host association of predator-prey based on the lytic cycle is not valid presently.

SECTION II: GUT MICROBES AND METAGENOMICS

INTRODUCTION

Most rumen microbes are fastidious and extremely difficult to culture under laboratory conditions. There is also an ambiguity associated with phenotypic criteria used in defining taxa among the few gut microbes that have been cultured (Ho and Barr 1995; Whitford et al. 1998). Traditional techniques such as the roll-tube technique and most probable number technique used to study gut associated microbes tend to underestimate the microbial numbers and the diversity (Deng et al. 2008). Only about 1% of the microbes detected by direct microscopy can be cultured with existing knowledge and laboratory conditions (Pace 1997). It is generally accepted that good classification schemes utilize natural evolutionary relationships among organisms and any scheme entirely based on morphology fails to recognize evolutionary relationships (Raskin et al. 1997). As a result, an increasingly growing interest has developed in molecular techniques to overcome these limitations, and to rapidly identify and quantify gut microorganisms (Kobayashi et al. 2000). Zuckerkandl and Pauling (1965) have highlighted that macromolecule sequence comparison is the most reliable way of finding the phylogenetic relationships among living organisms. Macromolecule sequence data provides a quantitative interpretation of the microbial community structure along with accurate phylogenetic relationships provided that the sequences being used are sufficiently long (Pace et al. 1986). Such ideal macromolecules should not be transferred between organisms (resist horizontal gene transfer), have a strictly constant function, and have a sufficient number of residues that undergo change at a rate proportionate to the evolutionary distance considered. There is no one ideal molecule which

fulfill all of the above criteria. However, 16S and 18S ribosomal RNA (rRNA) sufficiently meet these criteria. Ribosomal RNA and their coding genes are evolutionarily conserved. Although 5S rRNA and transfer RNA (tRNA) have conserved sequences, they contain very few independently varying nucleotides. Microbial communities, except viruses, can be described by comparative analysis of rRNA gene sequences without having to culture them.

Using molecular techniques based on rRNA genes will not only provide a way of classifying rumen microbes, but will also aid in predicting the evolutionary relationships among them (DeLong et al. 1989). High throughput sequencing technologies such as Roche 454 pyrosequencing and Illumina sequencing, Quantitative Real Time PCR (qRT-PCR), and Terminal Restriction Fragment Length Polymorphism (T-RFLP) are some of the most widely used rRNA gene based methods to identify and quantify microorganisms (Stahl et al. 1988). Fluorescence in situ hybridization (FISH) or blot hybridization can be used to readily determine the absolute or relative abundance of a given sequence or organism (Dangler, 1996). There are well-supported public databases having rRNA gene sequence information available such as the Ribosomal Database Project (RDP) (http://www.cme.msu.edu) and GenBank (http://www.ncbi.nlm.nih.gov).

Sequence comparisons allow identification of new operational taxonomic units/species based on the sequence similarity (<97% for bacteria) to the already characterized species, and allows for the placement in existing phylogenetic classification (Deng et al. 2008). Universal probes targeting rRNA genes or Internal Transcribed Spacer (ITS) regions between rRNA genes quantify the total microbial populations in a sample and provide for the basis of normalizing the results obtained with group specific probes.

Generally 16S and 18S rRNA genes are preferred over 23S or 28S rRNA genes in microbial profiling due to their higher availability in public databases (Deng et al. 2008).

Comparative analysis of 16S rRNA gene sequences is a very useful tool for studying bacterial communities. Metagenomic studies based on rRNA genes have found many new microbial species from the animal gut, since studies based on conventional methods could not discover them (Pace et al. 1986). Comparative analysis of 16S rRNA gene sequences led to the discovery of the two fundamental types of the prokaryotes; archaea and eubacteria (Woese et al. 1990). A recent analysis of 16S rRNA gene libraries derived from bovine rumen fluid has led to the discovery of a new cluster of Euryarchaeota, which is not associated with the known methanogens (Tajima et al. 2001). More recent studies based on 16S rRNA gene sequence comparisons have reported a higher diversity in the rumen bacterial/archaeal communities (Whitford et al. 2001; An et al. 2005; Deng et al. 2007).

PCR methods have been successfully used to amplify 18S rRNA genes and ITS regions of rumen protozoa and fungi (Dore and Stahl 1991; Wright et al. 1997). However, it is quite challenging to design 18S rRNA gene or ITS region based primers, which are also specific to protozoa or fungi due to high conservation of regions of rRNA genes among many eukaryotic divisions. The specificity of the primers is a crucial factor when working with complex biomass such as rumen content since it contains multiple sources of 18S rRNA genes, which includes: sloughing off of host animal cells, ingested plant material, rumen protozoa, fungi, nematode parasites etc. Several studies using 18S rRNA gene sequences to classify rumen fungi (Dore and Stahl 1991; Bowman et al. 1992) have failed to discuss the inter-relationships at genera or species level, which is most likely due to insufficient

variability in the 18S rRNA genes. Li and Heath (1992) used the ITS 1 region sequences to compare anaerobic fungi (AF). Unlike the 18S rRNA genes, ITS 1 region (and also ITS 2 region) is not well conserved among species and strains of fungi due to less evolutionary pressure. This allows for a higher resolving power in classifying sequences at genus or species level.

Besides this ambiguity in classification of gut microbes, the laborious nature and requirement of a greater level of expertise hinders the popularity of classical microbiological methods in most present day studies. Also, rumen microbes are extremely resistant to culture. High throughput sequencing and PCR based technologies are becoming available and affordable to many laboratories worldwide. Therefore, metagenomic techniques provide invaluable tools to explore the role of gut microbes and their diversity along with how they respond to various environmental conditions inside the mammalian gut.

So far, except few phages infecting Bacteroides and Clostridium, no viruses have been isolated that infects other dominant gut bacteria (Kim et al. 2011b) due to the fastidious nature of their bacterial hosts (Lorenzi et al. 2011). Also, it may be difficult to induce phage lytic cycles under experimental conditions. The absence of universally conserved genes in viruses, analogous to rRNA genes in cellular organisms, has hindered studying of viral diversity (Lorenzi et al. 2011). Additionally, standard metagenomic techniques cannot be effectively used to study bacteriophages due to the presence of modified nucleotide bases in their genomes. Consequently, phage DNA is lethal to bacteria and poses a barrier to standard cloning techniques (Warren 1980). The high throughput shotgun sequence analysis is the most widely used approach in studying the viral diversity, alleviating the above mentioned

limitations (Lorenzi et al. 2011). Metagenomic analyses of natural communities and man-made niches have shown that viruses are extremely diverse and novel (Rohwer et al. 2009). Metagenomic studies have been conducted in many extreme ecosystems ranging from hot springs (Rachel et al. 2002; Redder et al. 2009) to marine sediments (Breitbart et al. 2004) in which viruses are the only known microbial predators. However, with increasing number of comparative genomic studies being conducted on viromes isolated from the same host under different conditions have found striking similarities between them (Angly et al. 2009; Redder et al. 2009; Weynberg et al. 2009). This indicates that if the mammalian gut is sampled adequately, there is a possibility of gaining an understanding of the viral diversity inside the gut ecosystem.

SECTION III: DNA METHYLATION AND EPIGENETIC REGULATION OF GENE EXPRESSION IN VIRUSES

INTRODUCTION

The phenotype of a cell or an organism is primarily determined by its gene expression profile. Epigenetics provides an additional stability and diversity to the phenotype of a cell by modifying the process of transcription through chromatin marks that are maintained or regenerated during cellular division (Laird 2010). In eukaryotic organisms, methylation of cytosine base at the fifth carbon atom (C^{5m}) is a common epigenetic landmark found in the sequence formats CpG or CpHpG (H = A, T, C). Methylated cytosine bases were first discovered in a calf thymus about 60 years ago (Hotchkiss 1948). In mammals, two to seven percent of the total cytosine is converted to 5-methylcytosine (Fig. 3) after the new DNA is synthesized by the action of DNA methyltransferases (e.g. Dnmt1, Dnmt2 and Dnmt3) (Vanyushin et al. 1970). Methylated cytosine bases (C^{5m}) tend to mutate to thymine bases (Bird, 1980). Therefore, CpG dinucleotides are mutation hot spots and evolution tends to eliminate them unless they are positively selected (Salser et al. 1979). This supports the idea that non-coding regions should be deficient in CpG dinucleotides.

Bacteria and archaea are known to carry 5-methylcytosine residues, N-4-mmethylcytosine and N-6-methyladenine (Laird 2010). A majority of the N-6-methyladenine residues in bacteria are formed by the action of dam methylase/methyltransferase (Marinus and Morris 1973; Herman and Modrich 1982). Dam methyltransferases recognize a specific tetranucleotide sequence, which is 'GATC'

and methylates second adenine base in the sequence. There is another methyltransferase called dcm methylase/methyltransferase, which methylates the fifth carbon atom of the second cytosine base in specific sequences that have been identified as ‘CCAGG’ and ‘CCTGG’ (Marinus and Morris 1973; May and Hattman 1975). These methylated residues in bacteria are primarily important in restriction-modification systems and mismatch repair strand discrimination (Laird 2010). Unmethylated restriction sites have been shown to be very vulnerable to bacterial restriction enzymes, while methylated restriction sites either have a reduced affinity or no affinity to them (Yuan and Meselson 1970; Smith 1979). There are reports of bacteriophages carrying methylated bases (May and Hattman 1975). For example, one study reported that about 50% of the bacteriophage lambda DNA is dam methylated (Pirotta 1976), and the level of methylation is most likely decided by the available period of time between DNA replication and packaging into the phage head. Therefore, restriction enzymes provide a line of defense for bacteria against bacteriophage infections by cleaving non-methylated recognition sites in DNA.

Methylation of cytosine residues plays an important role in gene expression and differentiation (Razin and Riggs 1980) in higher eukaryotes. The conversion from cytosine to 5-methylcytosine introduces a methyl group to an exposed position in the major groove in the DNA double helix, and it makes a change in the binding of proteins to DNA (Riggs 1975). These DNA-protein interactions have an effect on secondary and tertiary folding of DNA (Comings and Riggs 1971). DNA-mediated cell transfection studies have given convincing evidence that methylated residues are responsible for a closed chromatin structure (Keshet et al. 1986). The precise mechanism by which the chromatin structure changes by DNA methylation is not fully elucidated, but many

studies have focused on the nucleosome structure to investigate how methylation affects the chromatin packaging (Cedar and Bergman 2012). The central core of the nucleosomes contains most of the methylated CpGs, and it has been predicted that there is an intrinsic effect of methylation on nucleosome position in DNA (Razin and Cedar 1977; Chodavarapu et al. 2010). Methylation mediates down regulation of genes through the action of methyl-binding proteins (Klose and Bird 2006). These methyl-binding proteins such as MeCP2, MBD2 and MBD3 can specifically bind to the methylated bases of DNA, and possibly initiate a modification of local chromatin structure by histone deacetylation (Jones et al. 1998; Nan et al. 1998) or through methylation of specific lysine residues (Fuks et al. 2003). Conversely, recent studies have shown that the regulation of *de novo* methylation of cytosine is dependent on many factors such as the local state of histone lysine methylation, the interaction with repeated sequences, small RNAs, components of the RNAi pathway, and divergent and catalytically inert cytosine methyltransferase homologues (Goll and Bestor 2005).

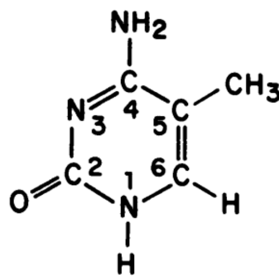

Figure 3. Structure of 5-methylcytosine

Over the past years, many sensitive and reliable methods have been developed to identify and quantify 5-methylcytosine residues in DNA including: immunological methods, High Performance Liquid Chromatography (HPLC), and Gas Chromatography-

Mass Spectrometry (GC-MS) (Miller et al. 1974; Singer et al. 1977; Singer et al. 1979). These methods cannot distinguish the function of DNA methylation since they do not relate the methylated residues to a particular region of chromosomes or specific sequence. However, they provide valuable information about the methylation status of a piece of DNA or a genome (Razin and Riggs 1980). This method limitation was partially fulfilled by using bacterial restriction endonucleases, which cleave DNA at specific sites. About 90% of the methylated cytosine residues in eukaryotic DNA are found in the dinucleotide combination CpG (Doskočil and Šorm 1962). Several restriction enzymes include CpG in their recognition sequences. Some of these restriction enzymes (e.g. *Sau3AI*) fail to cleave DNA when cytosine residues of CpG are methylated (New England Biolabs Inc., Ipswich, MA). Hence, these enzymes can be used to study DNA methylation patterns when compared to unmethylated control DNA. Such control DNA can be obtained by cloning the DNA fragment of interest to *E.coli*, since CpG sites of cloned DNA are not methylated (Rasin and Riggs 1980). It may be possible to obtain such DNA by using PCR since DNA methyltransferases are not present in the PCR reaction mix (Laird 2010). Important findings related to CpG methylation have been revealed by digesting total genomic DNA with *HpaII* (blocked by CpG methylation) or *MspI* (not sensitive to methylation) and by subsequent Southern blotting and visualization with a labeled probe (Waalwijk and Flavell 1978). Maxam–Gilbert sequencing technique has also been used to identify methylated cytosine residues (Maxam and Gilbert 1977). Almost all of the studies involved with working on sequence specific DNA methylation analyses have used a methylation dependent treatment of DNA before amplifying DNA (Laird 2010). The main steps in such a protocol would be restriction digestion, affinity

enrichment, and bisulphite conversion. At present, there have been methylation studies conducted on a whole genome scale with advancing sequence technologies (Cokus et al. 2008; Lister et al. 2009). Sodium bisulphite deaminates unmethylated cytosine residues rapidly compared to methylated cytosines (Wang et al. 1980; Hayatsu 2008). Therefore, bisulphite conversion turns the epigenetic difference into a genetic difference. Unmethylated cytosines appear as thymine residues (through uracil) in a DNA sequence after bisulphite conversion and sequencing (Laird 2003; Pomraning et al. 2009). The ultimate and comprehensive DNA methylation analysis is bisulfite sequencing, which produces a resolution at a single-base (Laird 2010).

Almost all the invertebrates, fungal species, and most common bacteria have been found to contain a ‘normal’ relative abundance of CpG dinucleotides, based on a statistical method, presumably due to the absence of the specific methyltransferases (Karlin et al. 1994a). Nearly all of the viral genomes (except most of the herpes viruses and vaccinia virus) are deficient in CpG dinucleotides. Additionally, there are no other dinucleotide extremes in these viral genomes. In contrast, the relative abundance of CpG dinucleotides in most bacteriophages studied is in the normal expected range. The gammaherpes viruses are deficient in CpG dinucleotides and have a higher relative abundance of TpG/CpA dinucleotides (Karlin et al. 1994a). It is not yet fully understood whether CpG dinucleotides are methylated in gammaherpes viruses during replication. Methylation of the cytosine bases in CpG dinucleotides of viral genomes has been shown to have an association with the latent infections of herpes simplex virus (Youssoufian et al. 1982). Varying degrees of methylation of Epstein-Barr virus (EBV) in tumorigenic cell lines have been detected and shown that there is an association of latent gene

expression with the level of methylation (Hu et al. 1991; Minarovits et al. 1991). Honess et al. (1989) proposed that herpes viruses in highly dividing cells, such as lymphocytes, follow the process of methylation deamination mutation as described above. However, human herpes virus 6 does not show any suppression of CpG dinucleotides (Karlin et al. 1994b).

It is not clear why a dsDNA viral genome present in a nucleus would not be methylated (Karlin et al. 1994a), and it may require a special DNA-protein structure. It is plausible to think that small viral genomes can escape methylation due to their short replication times. In general, mechanisms leading to CpG suppression in small viruses may be different from methylation driven CpG suppression. A good example for the existence of such mechanisms is the metazoan species mitochondrial genomes, where CpG dinucleotides are highly suppressed. CpG methyltransferases are absent in mitochondria, and in the nucleus of lower eukaryotes. It has been demonstrated that CpG dinucleotides have the highest thermodynamic stacking energy. Consequently, avoidance of CpG dinucleotides enhances the rate of DNA replication, which is beneficial for viruses (Breslauer et al. 1986). Beside these suggestions, having low CpG relative abundances in viral genomes reduces the incidence of methylation and subsequent transcription inhibition (Karlin et al. 1994a). There are reports of down regulation of gene expression in several vertebrate retroviral proviruses having methylated DNA (Cooper 1983). Based on these facts, bacteriophages carrying normal CpG dinucleotide relative abundance can be justified due to the unavailability of the required methyltransferases in the host cell and smaller genome sizes.

REFERENCES

Ackermann HW. 2003. Bacteriophage observations and evolution. *Research in Microbiology* **154**(4): 245-251.

An D, Dong X, Dong Z. 2005. Prokaryote diversity in the rumen of yak (*Bos grunniens*) and Jinnan cattle (*Bos taurus*) estimated by 16S rDNA homology analyses. *Anaerobe* **11**(4): 207-215.

Angly F, Youle M, Nosrat B, Srinagesh S, Rodriguez-Brito B, McNairnie P, Deyanat-Yazdi G, Breitbart M, Rohwer F. 2009. Genomic analysis of multiple Roseophage SIO1 strains. *Environmental Microbiology* **11**(11): 2863-2873.

Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes M, Kelley S, Liu H et al. 2006. The marine viromes of four oceanic regions. *Plos Biology* **4**: 2121-2131.

Ankrah P, Loerch SC, Dehority BA. 1990. Sequestration, migration and lysis of protozoa in the rumen. *Journal of General Microbiology* **136**(9): 1869-1875.

Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. 2005. Host-bacterial mutualism in the human intestine. *Science* **307**(5717): 1915-1920.

Bauchop T. 1989. Biology of gut anaerobic fungi. *Biosystems* **23**(1): 53-64.

Berg Miller ME, Yeoman CJ, Chia N, Tringe SG, Angly FE, Edwards RA, Flint HJ, Lamed R, Bayer EA, White BA. 2012. Phage–bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome. *Environmental Microbiology* **14**(1): 207-227.

Blackburn TH, Hobson PN. 1960. The degradation of protein in the rumen of the sheep and redistribution of the protein nitrogen after feeding. *Br J Nutr* **14**(4): 445-456.

Bowman BH, Taylor JW, Brownlee AG, Lee J, Lu SD, White TJ. 1992. Molecular evolution of the fungi: relationship of the Basidiomycetes, Ascomycetes, and Chytridiomycetes. *Mol Biol Evol* **9**(2): 285-296.

Braune R. 1913. Untersuchungen uber die im Wiederkauermagen vorkommenden Protozoen. *Archiv fur Protistenkunde* **32**: 111-170.

Breitbart M, Felts B, Kelley S, Mahaffy JM, Nulton J, Salamon P, Rohwer F. 2004. Diversity and population structure of a near-shore marine-sediment viral community. *Proc Biol Sci* **271**(1539): 565-574.

Breitbart M, Haynes M, Kelley S, Angly F, Edwards RA, Felts B, Mahaffy JM, Mueller J, Nulton J, Rayhawk S et al. 2008. Viral diversity and dynamics in an infant gut. *Research in Microbiology* **159**(5): 367-373.

Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P, Rohwer F. 2003. Metagenomic Analyses of an Uncultured viral community from human feces. *J Bacteriol* **185**(20): 6220-6223.

Breslauer KJ, Frank R, Blöcker H, Marky LA. 1986. Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences* **83**(11): 3746-3750.

Brock FM, Forsberg CW, Buchanan-Smith JG. 1982. Proteolytic activity of rumen microorganisms and effects of proteinase inhibitors. *Applied and Environmental Microbiology* **44**(3): 561-569.

Brookman JL, Mennim G, Trinci APJ, Theodorou MK, Tuckwell DS. 2000. Identification and characterization of anaerobic gut fungi using molecular methodologies based on ribosomal ITS1 and 18S rRNA. *Microbiology-Uk* **146**: 393-403.

Bryant MP, Small N. 1960. Observations on the ruminal microorganisms of isolated and inoculated calves. *J Dairy Sci* **43**(5): 654-667.

Bryant MP, Small N, Bouma C, Robinson I. 1958. Studies on the composition of the ruminal flora and fauna of young calves. *J Dairy Sci* **41**(12): 1747-1767.

Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann M-L, Brüssow H. 2003. Phage as agents of lateral gene transfer. *Current Opinion in Microbiology* **6**(4): 417-424.

Carlson JR, Yokoyama MT, Dickinson EO. 1972. Induction of pulmonary edema and emphysema in cattle and goats with 3-methylindole. *Science* **176**(4032): 298-299.

Casas V, Rohwer F. 2007. Phage metagenomics. *Advanced Bacterial Genetics: Use of Transposons and Phage for Genomic Engieering* **421**: 259-268.

Cedar H, Bergman Y. 2012. Programming of DNA methylation patterns. *Annu Rev Biochem* **81**(4): 1-21.

Cheng K-J, Forsberg CW, Minato H, Costerton JW. 1991. Microbial ecology and physiology of feed degradation with in the rumen. In *Physiological Aspects of*

Digestion and Metabolism in Ruminants, (ed. T Tsuda, Y Sasaki, R Kawashima). Academic Press, Toronto.

Chesson A, Stewart CS, Wallace RJ. 1982. Influence of plant phenolic acids on growth and cellulolytic activity of rumen bacteria. *Applied and Environmental Microbiology* **44**(3): 597-603.

Chiu Y-W, Suh S, Pfister S, Hellweg S, Koehler A. 2012. Measuring ecological impact of water consumption by bioethanol using life cycle impact assessment. *The International Journal of Life Cycle Assessment* **17**(1): 16-24.

Chodavarapu RK, Feng S, Bernatavichute YV, Chen P-Y, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ et al. 2010. Relationship between nucleosome positioning and DNA methylation. *Nature* **466**(7304): 388-392.

Church DC, ed. 1969. *Digestive Physiology and Nutrition of Ruminants*. O.S.U. Book Stores, Inc., Corvallis, OR.

Church DC. 1988. Salivary function and production. In *The ruminant animal digestive physiology and nutrition*, (ed. DC Church). Prentice Hall, Englewood Cliffs, NJ.

Clarke RTJ. 1977. The gut and its micro-organisms. In *Microbial ecology of the gut*, (ed. RTJ Clarke, T Bouchop). Academic Press New York.

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE. 2008. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**(7184): 215-219.

Coleman GS, Hall FJ. 1969. Electron microscopy of the rumen cilate *Entodinium caudatum*, with special referense to the engulfment of bacteria and other particulate matter. *Tissue and Cell* **1**(4): 607-618.

Collins JS, Brennan FN, Lee RJ, Love AH. 1986. Rumination in adults--a rare cause of gastro-oesophageal regurgitation in two patients. *Ulster Med J* **55**(2): 178-180.

Comings DE, Riggs AD. 1971. Molecular mechanisms of chromosome pairing, folding and function. *Nature* **233**(5314): 48-50.

Cooper DN. 1983. Eukaryotic DNA methylation. *Hum Genet* **64**(4): 315-333.

Coverdale JA, Tyler HD, Quigley Iii JD, Brumm JA. 2004. Effect of various levels of forage and form of diet on rumen development and growth in calves. *J Dairy Sci* **87**(8): 2554-2562.

Craig WM, Broderick GA, Ricker DB. 1987. Quantitation of microorganisms associated with the particulate phase of ruminal ingesta. *The Journal of Nutrition* **117**(1): 56-62.

Danovaro R, Dell'Anno A, Trucco A, Serresi M, Vanucci S. 2001. Determination of virus abundance in marine sediments. *Appl Environ Microbiol* **67**(3): 1384-1387.

Davies DR. 1991. *Growth and survival of anaerobic fungi in batch culture and in the digestive tract of ruminants*. University of Manchester, Manchester.

De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, Collini S, Pieraccini G, Lionetti P. 2010. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences* **107**(33): 14691-14696.

Dearth RN, Dehority BA, Potter EL. 1974. Rumen microbial numbers in lambs as affected by level of feed intake and dietary diethylstilestrol. *Journal of Animal Science* **38**(5): 991-996.

Dehority BA, ed. 2003. *Rumen Microbiology*. Nottingham University Press, Nottingham, UK.

Dehority BA, Tirabasso PA. 2000. Antibiosis between ruminal bacteria and ruminal fungi. *Applied and Environmental Microbiology* **66**(7): 2921-2927.

DeLong EF, Wickham GS, Pace NR. 1989. Phylogenetic stains: ribosomal RNA-based probes for the identification of single cells. *Science* **243**(4896): 1360-1363.

Deng W, Wanapat M, Ma S, Chen J, Xi D, He T, Yang Z, Mao H. 2007. Phylogenetic analysis of 16S rDNA sequences manifest rumen bacterial diversity in gayals (*Bos frontalis*) fed fresh bamboo leaves and twigs (*Sinarumdinaria*). *Asian-Aust J Anim Sci* **20**(7): 1057-1066.

Deng W, Xi D, Mao H, Wanapat M. 2008. The use of molecular techniques based on ribosomal RNA and DNA for rumen microbial ecosystem studies: a review. *Molecular Biology Reports* **35**(2): 265-274.

Dibaise JK, Husen Z, Crowell MD, Krajmalnik-Brown R, Decker AG, Rittmann BE. 2008. Gut microbiota and its possible relationship with obesity. *Mayo Clinic Proceedings* **83**(4): 460-469.

Dobson A, Dobson MJ, ed. 1986. *Aspects of Digestive Physiology in Ruminants*. Cornell University Press, Ithaca.

Dogiel V. 1947. The phylogeny of the stomach-infusorians of ruminants in the light of palaeontological and parasitological data. *Quarterly Journal of Microscopical Science* **s3-88**(3): 337-344.

Dore J, Stahl DA. 1991. Phylogeny of anaerobic rumen Chytridiomycetes inferred from small subunit ribosomal RNA sequence comparisons. *Canadian Journal of Botany* **69**(9): 1964–1971.

Doskočil J, Šorm F. 1962. Distribution of 5-methylcytosine in pyrimidine sequences of deoxyribonucleic acids. *Biochimica et Biophysica Acta (BBA) - Specialized Section on Nucleic Acids and Related Subjects* **55**(6): 953-959.

Eadie JM. 1962. The development of rumen microbial populations in lambs and calves under various conditions of management. *Journal of General Microbiology* **29**(4): 563-578.

Eadie JM, Hobson PN, Mann SO. 1959. A Relationship between some bacteria, protozoa and diet in early weaned calves. *Nature* **183**(4661): 624-625.

Edwards RA, Rohwer F. 2005. Viral metagenomics. *Nat Rev Micro* **3**(6): 504-510.

Ellis JE, Williams AG, Lloyd D. 1989. Oxygen consumption by ruminal microorganisms: protozoal and bacterial contributions. *Applied and Environmental Microbiology* **55**(10): 2583-2587.

Elsden SR, Volcani BE, Gilchrist FMC, Lewis D. 1956. Properties of fatty acid forming orgamism isllated from the rumen of sheep. *Journal of Bacteriology* **72**(5): 681-689.

Floch MH, Montrose DC. 2005. Use of probiotics in humans: an analysis of the literature. *Gastroenterology clinics of North America* **34**(3): 547-570, x.

Fondevila M, Dehority BA. 2001. In vitro growth and starch digestion by *Entodinium exiguum* as influenced by the presence or absence of live bacteria. *Journal of Animal Science* **79**(9): 2465-2471.

Fonty G, Gouet P, Jouany JP, Rieu F, Senaud J, Citron A, Breton A. 1986. Microbial colonization of the rumen in young lambs. In *Biology of Anaerobic Bacteria*, (ed. HC Dubourguier, G Albagnac, J Montreuil, C Romond, P Sautier, J Guillaume), pp. 40-46. Elsevier, Amsterdam.

Fonty G, Gouet P, Jouany JP, Senaud J. 1987. Role de la microflore sur la cinetique d'apparition des protozoaires cilies dans le rumen de l'agneau. *Reprod Nutr Dev* **26**: 285-286.

Forsberg CW, Cheng K-J, White BA. 1997. Polysaccharide degradation in the rumen and large intestine. In *Gastrointestinal Ecosystems and Fermentations*, Vol 1 (ed. RI Mackie, BA White). Capman and Hall, New York.

Frandson RD, Wilke WL, Fails AD. 2009. *Anatomy and Physiology of Farm Animals*. Blackwell Publishing, Ames, Iowa.

Frank DN, St. Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. 2007. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences* **104**(34): 13780-13785.

Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Kouzarides T. 2003. The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J Biol Chem* **278**(6): 4035-4040.

Furness DN, Butler RD. 1988. The functional and evolutionary significance of the ultrastructure of the Ophryoscolecidae (Order Entodiniomorphida)1. *Journal of Eukaryotic Microbiology* **35**(1): 34-38.

Gay L. 1991. Chitin content and chitin synthase activity as indicators of the growth of three different anaerobic rumen fungi. *FEMS Microbiology Letters* **80**(1): 99-102.

Gibson GR, Roberfroid MB. 1995. Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. . *Journal of Nutrition* **125**: 1401-1412.

Goll MG, Bestor TH. 2005. Eukaryotic cytosine methyltransferases. *Annual Review of Biochemistry* **74**(1): 481-514.

Gong J, Forsberg CW. 1989. Factors affecting adhesion of *Fibrobacter succinogenes* subsp. succinogenes S85 and adherence-defective mutants to cellulose. *Applied and Environmental Microbiology* **55**(12): 3039-3044.

Gordon GL, Phillips MW. 1989. Degradation and utilization of cellulose and straw by three different anaerobic fungi from the ovine rumen. *Applied and Environmental Microbiology* **55**(7): 1703-1710.

Griffith GW, Ozkose E, Theodorou MK, Davies DR. 2009. Diversity of anaerobic fungal populations in cattle revealed by selective enrichment culture using different carbon sources. *Fungal Ecol* **2**(2): 87-97.

Grubb JA, Dehority BA. 1975. Effects of an abrupt change in ration from all roughage to high concentrate upon rumen microbial numbers in sheep. *Applied Microbiology* **30**(3): 404-412.

Hamlin LJ, Hungate RE. 1956. Culture and physiology of a starch digsting bacterium (*Bacteroides amylophilus* n. sp.) from the bovine rumen. *Journal of Bacteriology* **72**(4): 548-554.

Hashizume K, Tsukahara T, Yamada K, Koyama H, Ushida K. 2003. *Megasphaera elsdenii* JCM1772T normalizes hyperlactate production in the large intestine of fructooligosaccharide-fed rats by stimulating butyrate production. *The Journal of Nutrition* **133**(10): 3187-3190.

Hayatsu H. 2008. Discovery of bisulfite-mediated cytosine conversion to uracil, the key reaction for DNA methylation analysis - A personal account. *Proceedings of the Japan Academy, Series B* **84**(8): 321-330.

Heath B, Bauchop T, Skipp RA. 1983. Assignment of the rumen anaerobe *Neocallimastix frontalis* to the Spizellomycetales (Chytridiomycetes) on the basis of its

polyflagellate zoospore ultrastructure. *Canadian Journal of Botany* **61**(1): 295–307.

Heath IB. 1988. Gut fungi. *Trends in Ecology and Evolution* **3**(7): 167-171.

Herdt TH. 2000. Ruminant adaptation to negative energy balance. Influences on the etiology of ketosis and fatty liver. *Vet Clin North Am Food Anim Pract* **16**(2): 215-230.

Herman GE, Modrich P. 1982. *Escherichia coli* dam methylase. Physical and catalytic properties of the homogeneous enzyme. *J Biol Chem* **257**(5): 2605-2612.

Hespell RB, Akin DE, Dehority BA. 1997. Bacteria, fungi, and protozoa of the rumen. In *Gastrointestinal Microbiology*, Vol 2 (ed. RI Mackie, BA White, RE Isaacson). Chapman and Hall, New York.

Hewson I, O'Neil JM, Fuhrman JA, Dennison WC. 2001. Virus-like particle distribution and abundance in sediments and overlying waters along eutrophication gradients in two subtropical estuaries. *Limnology and Oceanography* **46**(7): 1734-1746.

Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF, Eriksson OE, Huhndorf S, James T, Kirk PM, Lücking R et al. 2007. A higher-level phylogenetic classification of the fungi. *Mycological Research* **111**(5): 509-547.

Ho YW, Barr DJS. 1995. Classification of anaerobic gut fungi from herbivores with emphasis on rumen fungi from malaysia. *Mycologia* **87**(5): 655-677.

Hobson PN, Mann SO, Oxford AE. 1958. Some studies on the occurrence and properties of a large gram-negative coccus from the rumen. *Journal of General Microbiology* **19**(3): 462-472.

Hoffmann AM, Bergh A, Olivecrona T. 1989. Changes of testicular cholesteryl ester hydrolase activity in experimentally cryptorchid rats. *J Reprod Fertil* **86**(1): 11-18.

Hofmann RR. 1989. Evolutionary steps of ecophysiological adaptation and diversification of ruminants: a comparative view of their digestive system. *Oecologia* **78**(4): 443-457.

Honess RW, Gompels UA, Barrell BG, Craxton M, Cameron KR, Staden R, Chang Y-N, Hayward GS. 1989. Deviations from expected frequencies of CpG dinucleotides in herpesvirus DNAs may be diagnostic of differences in the states of their latent genomes. *Journal of General Virology* **70**(4): 837-855.

Hotchkiss RD. 1948. The quantitative seperation of purines, pyrimidines, and nucleosides by paper chromatography. *J Biol Chem* **175**(1): 315-332.

Hu LF, Minarovits J, Cao SL, Contreras-Salazar B, Rymo L, Falk K, Klein G, Ernberg I. 1991. Variable expression of latent membrane protein in nasopharyngeal carcinoma can be related to methylation status of the Epstein-Barr virus BNLF-1 5'-flanking region. *Journal of Virology* **65**(3): 1558-1567.

Hungate RE. 1950. The anaerobic mesophilic cellulolytic bacteria. *Bacteriological Reviews* **14**(1): 1-49.

Hungate RE, ed. 1966. *The Rumen and its Microbes*. Academic Press, Inc., New York.

Hungate RE. 1978. The rumen protozoa. In *Parasitic Protozoa*, Vol II (ed. JP Kreier). Academic Press, Inc., New York.

Hungate RE, Dougherty RW, Bryant MP, Cello RM. 1952. Microbiological and physiological changes associated with acute indigestion in sheep. *Cornell Vet* **42**: 423-449.

Ibrahim EA, Ingalls JR, Stanger NE. 1970. Effect of dietary diethylstilbesterol on populations and concentrations of ciliate protozoa in dairy cattle. *Can J Anim Sci* **50**(1): 101-106.

Johns AT. 1951. Isolation of a bacterium, producing propionic acid, from the rumen of sheep. *Journal of General Microbiology* **5**(2): 317-325.

Johnson KA, Johnson DE. 1995. Methane emissions from cattle. *Journal of Animal Science* **73**(8): 2483-2492.

Jones PL, Jan Veenstra GC, Wade PA, Vermaak D, Kass SU, Landsberger N, Strouboulis J, Wolffe AP. 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* **19**(2): 187-191.

Karlin S, Doerfler W, Cardon LR. 1994a. Why is CpG suppressed in the genomes of virtually all small eukaryotic viruses but not in those of large eukaryotic viruses? *Journal of Virology* **68**(5): 2889-2897.

Karlin S, Mocarski ES, Schachtel GA. 1994b. Molecular evolution of herpesviruses: genomic and protein sequence comparisons. *Journal of Virology* **68**(3): 1886-1902.

Keshet I, Lieman-Hurwitz J, Cedar H. 1986. DNA methylation affects the formation of active chromatin. *Cell* **44**(4): 535-543.

Kim K-H, Chang H-W, Nam Y-D, Roh SW, Kim M-S, Sung Y, Jeon CO, Oh H-M, Bae J-W. 2008. Amplification of uncultured single-stranded DNA viruses from rice paddy soil. *Applied and Environmental Microbiology* **74**(19): 5975-5985.

Kim MS, Park EJ, Roh SW, Bae JW. 2011b. Diversity and abundance of single-stranded DNA viruses in human feces. *Appl Environ Microbiol* **77**(22): 8062-8070.

King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ. 2012. Virus Taxonomy. Ninth report of the International Committee on Taxonomy of Viruses, San Diego.

Kleen JL, Hooijer GA, Rehage J, Noordhuizen JPTM. 2003. Subacute ruminal acidosis (SARA): a review. *Journal of Veterinary Medicine Series A* **50**(8): 406-414.

Klieve AV, Swain RA. 1993. Estimation of ruminal bacteriophage numbers by pulsed-field gel electrophoresis and laser densitometry. *Appl Environ Microbiol* **59**(7): 2299-2303.

Klose RJ, Bird AP. 2006. Genomic DNA methylation: the mark and its mediators. *Trends in Biochemical Sciences* **31**(2): 89-97.

Kobayashi Y, Forster RJ, Teather RM. 2000. Development of a competitive polymerase chain reaction assay for the ruminal bacterium *Butyrivibrio fibrisolvens* OB156 and its use for tracking an OB156-derived recombinant. *FEMS Microbiology Letters* **188**(2): 185-190.

Koike S, Kobayashi Y. 2009. Fibrolytic rumen bacteria: Their ecology and functions. *Asian-Aust J Anim Sci* **22**(1): 131 - 138.

Krupovic M, Prangishvili D, Hendrix RW, Bamford DH. 2011. Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. *Microbiology and Molecular Biology Reviews* **75**(4): 610-635.

Laird PW. 2003. Early detection: The power and the promise of DNA methylation markers. *Nature Reviews Cancer* **3**(4): 253.

Laird PW. 2010. Principles and challenges of genome-wide DNA methylation analysis. *Nat Rev Genet* **11**(3): 191-203.

Lee JJ, Hutner SH, Bovee EC. 1985. *An Illustrated Guide to the Protozoa.*, Society of Protozoologists, Lawrence, Kansas, USA.

Legay-Carmier F, Bauchart. 1989. Distribution of bacteria in the rumen contents of dairy cows given a diet supplemented with soya-bean oil. *Br J Nutr* **61**: 725-740.

Li J, Heath IB. 1992. The phylogenetic relationships of the anaerobic chytridiomycetous gut fungi (Neocallimasticaceae) and the Chytridiomycota. I. Cladistic analysis of rRNA sequences. *Canadian Journal of Botany* **70**(9): 1738-1746.

Liebetanz E. 1910. Die parasitischen protozoen des widenkauermagens. *Archiv fur Protistenkunde* **19**: 19-80.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-M et al. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**(7271): 315-322.

López-Bueno A, Tamames J, Velázquez D, Moya A, Quesada A, Alcamí A. 2009. High diversity of the viral community from an antarctic lake. *Science* **326**(5954): 858-861.

Lorenzi HA, Hoover J, Inman J, Safford T, Murphy S, Kagan L, Williamson SJ. 2011. The viral metagenome annotation pipeline (VMGAP): An automated tool for the functional annotation of viral metagenomic shotgun sequencing data. *Stand Genomic Sci* **4**(3): 418-429.

Lowe SE, Theodorou MK, Trinci AP. 1987. Cellulases and xylanase of an anaerobic rumen fungus grown on wheat straw, wheat straw holocellulose, cellulose, and xylan. *Appl Environ Microbiol* **53**(6): 1216-1223.

Lubinsky G. 1957a. Studies on the evolution of the Ophryoscolecidae (Ciliata : Oligotricha). II. On the origin of the higher Ophryoscolecidae. *Canadian Journal of Zoology* **35**(1): 135-140.

Lubinsky G. 1957b. Studies on the evolution of the Ophryoscolecidae (Ciliata: Oligotricha): III. Phylogeny of the Ophryoscolecidae based on their comparative morphology. *Canadian Journal of Zoology* **35**(1): 141-159.

Lund HG. 2007. Accounting for the world's rangelands. *Rangelands* **29**(1): 3-10.

Maranger R, Bird DF. 1996. High concentrations of viruses in the sediments of lac gilbert, québec. *Microbial Ecology* **31**(2): 141-151.

Marinus MG, Morris NR. 1973. Isolation of deoxyribonucleic acid methylase mutants of *Escherichia coli* K-12. *Journal of Bacteriology* **114**(3): 1143-1150.

Marounek M, Fliegrova K, Bartos S. 1989. Metabolism and some characteristics of ruminal strains of *Megasphaera elsdenii*. *Applied and Environmental Microbiology* **55**(6): 1570-1573.

Maxam AM, Gilbert W. 1977. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences* **74**(2): 560-564.

May MS, Hattman S. 1975. Analysis of bacteriophage deoxyribonucleic acid sequences methylated by host- and R-factor-controlled enzymes. *Journal of Bacteriology* **123**(2): 768-770.

McAllister TA, Bae HD, Jones GA, Cheng KJ. 1994. Microbial attachment and feed digestion in the rumen. *Journal of Animal Science* **72**(11): 3004-3018.

McCrady MH. 1915. The numerical interpretation of fermentation-tube results. *The Journal of Infectious Diseases* **17**(1): 183-212.

McElroy LW, Goss H. 1940. A quantitative study of vitamins in the rumen contents of sheep and cows fed vitaminlow diets. 1. Riboflavin and vitamin K. *Journal of Nutrition* **20**: 527-540.

McSweeney CS, Dulieu A, Katayama Y, Lowry JB. 1994. Solubilization of lignin by the ruminal anaerobic fungus *Neocallimastix patriciarum*. *Applied and Environmental Microbiology* **60**(8): 2985-2989.

Miller OJ, Schnedl W, Allen J, Erlanger BF. 1974. 5-Methylcytosine localised in mammalian constitutive heterochromatin. *Nature* **251**(5476): 636-637.

Minarovits J, Minarovits-Kormuta S, Ehlin-Henriksson B, Falk K, Klein G, Ernberg I. 1991. Host cell phenotype-dependent methylation patterns of Epstein—Barr virus DNA. *Journal of General Virology* **72**(7): 1591-1599.

Minato H, Misumori M, Cheng KJ. 1993. Attachment of microorganisms to solid substrates in the rumen. In *MIE Bioforum on Genetic, Biochemistry and Ecology of Lignocellulose Degradation* pp. 139-145, Institut Pasteur, Paris, France.

Miron J, Ben-Ghedalia D, Morrison M. 2001. Invited review: Adhesion mechanisms of rumen cellulolytic bacteria. *J Dairy Sci* **84**(6): 1294-1309.

Moir RJ, Masson MJ. 1952. An illustrated scheme for the microscopic identification of the rumen microorganisms of sheep. *The Journal of Pathology and Bacteriology* **64**(2): 343-350.

Morand-Fehr P, Boyazoglu J. 1999. Present state and future outlook of the small ruminant sector. *Small Ruminant Res* **34**(3): 175-188.

Morrison M, Miron J. 2000. Adhesion to cellulose by *Ruminococcus albus*: a combination of cellulosomes and Pil-proteins? *FEMS Microbiology Letters* **185**(2): 109-115.

Mountfort DO, Orpin CG. 1994. *Anaerobic Fungi: Biology, Ecology and Function*. Marcel Dekker, Inc., New York.

Mueller RE, Asplund JM, Iannotti EL. 1984. Successive changes in the epimural bacterial community of young lambs as revealed by scanning electron microscopy. *Applied and Environmental Microbiology* **47**(4): 715-723.

Munn EA, Orpin CG, Greenwood CA. 1988. The ultrastructure and possible relationships of four obligate anaerobic chytridiomycete fungi from the rumen of sheep. *Biosystems* **22**(1): 67-81.

Nakamura K, Kanegasaki S. 1969. Densities of ruminal protozoa of sheep established under different dietary conditions. *J Dairy Sci* **52**(2): 250-255.

Nan X, Ng H-H, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A. 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**(6683): 386-389.

New England Biolabs. Accessed Mar 2012. Restriction endonucleases. New England Biolabs Inc. , Ipswich, MA. (http://www.neb.com/nebecomm/products/category1.asp)

Ogimoto K, Imai S. 1981. *Atlas of Rumen Microbiology*. Japan Scientific Societies Press, Tokyo, Japan.

Orpin CG. 1975. Studies on the rumen flagellate *Neocallimastix frontalis*. *J Gen Microbiol* **91**(2): 249-262.

Orpin CG. 1976. Studies on the rumen flagellate *Sphaeromonas communis*. *J Gen Microbiol* **94**(2): 270-280.

Orpin CG. 1977. Invasion of plant tissue in the rumen by the flagellate *Neocallimastix frontalis*. *J Gen Microbiol* **98**(2): 423-430.

Orpin CG. 1994. Anaerobic fungi: Taxonomy, biology, and distribution in nature. In *Anaerobic Fungi: Biology, Ecology and Function*, Vol 12 (ed. DO Mountfort, CG Orpin). Marcel Dekker, Inc., New York.

Orpin CG, Joblin KN. 1997. The rumen anaerobic fungi. In *The Rumen Microbial Ecosystem* (ed. PN Hobson, CS Stewart). Chapman & Hall, New York.

Orpin CG, Munn EA. 1986. *Neocallimastix patriciarum* sp.nov., a new member of the Neocallimasticaceae inhabiting the rumen of sheep. *Transactions of the British Mycological Society* **86**(1): 178-181.

Owens FN, Goetsch AL. 1988. Ruminal fermentation. In *The ruminant animal digestive physiology and nutrition*, (ed. DC Church). Prentice Hall, Englewood Cliffs, NJ.

Pace NR. 1997. A molecular view of microbial diversity and the biosphere. *Science* **276**(5313): 734-740.

Pace NR, Olsen GJ, Woese CR. 1986. Ribosomal RNA phylogeny and the primary lines of evolutionary descent. *Cell* **45**(3): 325-326.

Pell AN, Schofield P. 1993. Microbial adhesion and degradation of plant cell walls. In *Forage Cell Wall Structure and Digestibility*, (ed. RD Hatfield, HG Jung, J Ralph, DR Buxton, DR Mertens, PJ Weimer), pp. 397-423. ASA-CSSA-SSSA, Madison, WI.

Pomraning KR, Smith KM, Freitag M. 2009. Genome-wide high throughput analysis of DNA methylation in eukaryotes. *Methods* **47**(3): 142-150.

Porter V. 1993. *Pigs: A handbook to the breeds of the world*. Cornell University Press, Ithaca, New York.

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T et al. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**(7285): 59-65.

Rachel R, Bettstetter M, Hedlund BP, Häring M, Kessler A, Stetter KO, Prangishvili D. 2002. Remarkable morphological diversity of viruses and virus-like particles in hot terrestrial environments. *Arch Virol* **147**(12): 2419-2429.

Raskin L, Capman WC, Sharp R, Poulson LK, Stahl DA. 1997. Molecular ecology of gastrointestinal systems. In *Gastrointestinal Microbiology II*, (ed. RI Mackie, BA White, R Isaacson), pp. 243–298. Chapman and Hall, New York.

Razin A, Cedar H. 1977. Distribution of 5-methylcytosine in chromatin. *Proceedings of the National Academy of Sciences* **74**(7): 2725-2728.

Razin A, Riggs AD. 1980. DNA methylation and gene function. *Science* **210**(4470): 604-610.

Rebecca Vega T. 2009. Current insights into phage biodiversity and biogeography. *Current Opinion in Microbiology* **12**(5): 582-587.

Redder P, Peng X, Brügger K, Shah SA, Roesch F, Greve B, She Q, Schleper C, Forterre P, Garrett RA et al. 2009. Four newly isolated fuselloviruses from extreme geothermal environments reveal unusual morphologies and a possible interviral recombination mechanism. *Environmental Microbiology* **11**(11): 2849-2862.

Regenmortel MHVv, Fauquet CM, Bishop DHL, Carstens EB, Estes MK, Lemon SM, Maniloff J, Mayo MA, McGeoch DJ, Pringle CR et al. 2000. Virus Taxonomy. Classification and nomenclature of viruses. Seventh report of the International Committee on Taxonomy of Viruses, San Diego.

Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, Gordon JI. 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**(7304): 334-338.

Rezaeian M, Beakes GW, Parker DS. 2004. Distribution and estimation of anaerobic zoosporic fungi along the digestive tracts of sheep. *Mycological Research* **108**(10): 1227-1233.

Riggs AD. 1975. X inactivation, differentiation, and DNA methylation. *Cytogenetics Cell Genetics* **14**(1): 9-25.

Roberts AP, Chandler M, Courvalin P, Guédon G, Mullany P, Pembroke T, Rood JI, Jeffery Smith C, Summers AO, Tsuda M et al. 2008. Revised nomenclature for transposable genetic elements. *Plasmid* **60**(3): 167-173.

Rohwer F, Prangishvili D, Lindell D. 2009. Roles of viruses in the environment. *Environmental Microbiology* **11**(11): 2771-2774.

Roossinck M. 2011. changes in population dynamics in mutualistic versus pathogenic viruses. *Viruses* **3**(1): 12-19.

Ruckebusch Y. 1988. Motility of the grastro-intestinal tract. In *The ruminant animal Digestive physiology and nutrition*, (ed. DC Church). Prentice Hall, Englewood Cliffs, NJ.

Russell JB, Bottje WG, Cotta MA. 1981. Degradation of protein by mixed cultures of rumen bacteria: Identification of *Streptococcus bovis* as an actively proteolytic rumen bacterium. *Journal of Animal Science* **53**(1): 242-252.

Russell JB, Rychlik JL. 2001. Factors that alter rumen microbial ecology. *Science* **292**(5519): 1119-1122.

Russell JB, Wilson DB. 1988. Potential opportunities and problems for genetically altered rumen microorganisms. *The Journal of Nutrition* **118**(2): 271-279.

Salser WA, Cummings I, Liu A, Strommer J, Padayatty J, Clarke P. 1979. Analysis of chicken globin cDNA clones: discovery of a novel chicken alpha globin gene induced by stress in young chickens. In *Cellular and molecular regulation of hemoglobin switching*, (ed. G Stamatoyannopoulos, AW Nienhius). Grune and Stratton, New York.

Scheppach W, Weiler F. 2004. The butyrate story: old wine in new bottles? *Curr Opin Clin Nutr Metab Care* **7**(5): 563-567.

Short HL. 1964. Postnatal stomach development of white-tailed deer. *The Journal of Wildlife Management* **28**(3): 445-458.

Sidney J, J.R.Lyford. 1988. RGrowth and development of the ruminant digestive system. In *The ruminant animal digestive physiology and nutrition*, (ed. DC Church). Prentice Hall, Englewood Cliffs, NJ.

Sijpesteijn AK. 1951. On *Ruminococcus flavefaciens*, a cellulose-decomposing bacterium from the rumen of sheep and cattle. *Journal of General Microbiology* **5**(5): 869-879.

Singer J, Schnute Jr WC, Shively JE, Todd CW, Riggs AD. 1979. Sensitive detection of 5-methylcytosine and quantitation of the 5-methylcytosine/cytosine ratio in DNA by gas chromatography-mass spectrometry using multiple specific ion monitoring. *Analytical Biochemistry* **94**(2): 297-301.

Singer J, Stellwagen RH, Roberts-Ems J, Riggs AD. 1977. 5-Methylcytosine content of rat hepatoma DNA substituted with bromodeoxyuridine. *J Biol Chem* **252**(15): 5509-5513.

Sisson S. 1910. *A text book of Veterinary Anatomy*. W.B. Saunders and Company, Philadelphia adn London.

Smith HO. 1979. Nucleotide sequence specificity of restriction endonucleases. *Science* **205**(4405): 455-462.

Sonnenburg JL, Angenent LT, Gordon JI. 2004. Getting a grip on things: how do communities of bacterial symbionts become established in our intestine? *Nat Immunol* **5**(6): 569-573.

Stahl DA, Flesher B, Mansfield HR, Montgomery L. 1988. Use of phylogenetically based hybridization probes for studies of ruminal microbial ecology. *Applied and Environmental Microbiology* **54**(5): 1079-1084.

Stewart CS, Fonty G, Gouet P. 1988. The establishment of rumen microbial communities. *Animal Feed Science and Technology* **21**(2–4): 69-97.

Suttle CA. 1994. The significance of viruses to mortality in aquatic microbial communities. *Microbial Ecology* **28**(2): 237-243.

Tajima K, Nagamine T, Matsui H, Nakamura M, Aminov RI. 2001. Phylogenetic analysis of archaeal 16S rRNA libraries from the rumen suggests the existence of a novel group of archaea not associated with known methanogens. *FEMS Microbiology Letters* **200**(1): 67-72.

Tamate H, McGilliard AD, Jacobson NL, Getty R. 1962. Effect of various dietaries on the anatomical development of the stomach in the calf. *J Dairy Sci* **45**(3): 408-420.

Tap J, Mondot S, Levenez F, Pelletier E, Caron C, Furet J-P, Ugarte E, Muñoz-Tamayo R, Paslier DLE, Nalin R et al. 2009. Towards the human intestinal microbiota phylogenetic core. *Environmental Microbiology* **11**(10): 2574-2584.

Theodorou MK, Gill M, King-Spooner C, Beever DE. 1990. Enumeration of anaerobic Chytridiomycetes as thallus-forming units: novel method for quantification of fibrolytic fungal populations from the digestive tract ecosystem. *Applied and Environmental Microbiology* **56**(4): 1073-1078.

Theodorou MK, Lowe SE, Trinci APJ. 1992. Anaerobic fungi and the rumen ecosystem. In *The Fungal Community: Its Organisation and Role in the Ecosystem*, (ed. GC Carroll, DT Wicklow). Marcel Dekker, New York

Trotter DM, Lumb JW, ed. 1958. *Bovine Anatomy*. Burgess Publishing Company.

Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**(7122): 1027-1131.

Vanyushin BF, Tkacheva SG, Belozersky AN. 1970. Rare bases in animal DNA. *Nature* **225**(5236): 948-949.

Vavraj J, Joyon L. 1966. Etude sur la morphologie, le cycle evolutif et la position systematique de *Callimastix cyclopis*, Weissenburg I 91 2. *Protistologia* **2**: 5-16.

Waalwijk C, Flavell RA. 1978. DNA methylation at a CCGG sequence in the large intron of the rabbit β-globin gene: tissue-specific variations. *Nucleic Acids Res* **5**(12): 4631-4642.

Wang RY-H, Gehrke CW, Ehrlich M. 1980. Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues. *Nucleic Acids Res* **8**(20): 4777-4790.

Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, Wood TK. 2010. Cryptic prophages help bacteria cope with adverse environments. *Nat Commun* **1**: 147.

Warner ACI. 1966. Diurnal changes in the concentrations of microorganisms in the rumens of sheep fed limited diets once daily. *Journal of General Microbiology* **45**(2): 213-235.

Warner ACI, Stacy BD. 1965. Solutes in the rumen of the sheep. *Exp Physiol* **50**(2): 169-184.

Warren RA. 1980. Modified bases in bacteriophage DNAs. *Annu Rev Microbiol* **34**: 137-158.

Webb J, Theodorou MK. 1991. *Neocallimastix hurleyensis* sp.nov., an anaerobic fungus from the ovine rumen. *Canadian Journal of Botany* **69**(6): 1220–1224.

Wen L, Ley RE, Volchkov PY, Stranges PB, Avanesyan L, Stonebraker AC, Hu C, Wong FS, Szot GL, Bluestone JA et al. 2008. Innate immunity and intestinal microbiota in the development of Type 1 diabetes. *Nature* **455**(7216): 1109-1113.

Weynberg KD, Allen MJ, Ashelford K, Scanlan DJ, Wilson WH. 2009. From small hosts come big viruses: the complete genome of a second *Ostreococcus tauri* virus, OtV-1. *Environmental Microbiology* **11**(11): 2821-2839.

Whitford M, Teather R, Forster R. 2001. Phylogenetic analysis of methanogens from the bovine rumen. *Bmc Microbiol* **1**(1): 5.

Whitford MF, Forster RJ, Beard CE, Gong J, Teather RM. 1998. Phylogenetic analysis of rumen bacteria by comparative sequence analysis of cloned 16S rRNA genes. *Anaerobe* **4**(3): 153-163.

Whitman WB, Coleman DC, Wiebe WJ. 1998. Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences* **95**(12): 6578-6583.

Wilhelm SW, Suttle CA. 1999. Viruses and nutrient cycles in the sea. *BioScience* **49**(10): 781.

Woese CR, Kandler O, Wheelis ML. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences* **87**(12): 4576-4579.

Wright A-DG, Dehority BA, Lynn DH. 1997. Phylogeny of the rumen ciliates *Entodinium, Epidinium and Polyplastron* (Litostomatea: Entodiniomorphida) inferred from small subunit ribosomal RNA sequences. *Journal of Eukaryotic Microbiology* **44**(1): 61-67.

Yarlett N, Orpin CG, Munn EA. 1986. Evidence for hydrogenosomes in the rumen fungus *Neocallimastix patriciarum. Biochem J* **236**: 729-739.

Yokoyama T, Johnson A. 1988. Microbiology of the rumen and intestine. In *The ruminant animal digestive physiology and nutrition*, (ed. DC Church). Prentice Hall, Englewood Cliffs, NJ.

Youssoufian H, Hammer SM, Hirsch MS, Mulder C. 1982. Methylation of the viral genome in an in vitro model of herpes simplex virus latency. *Proceedings of the National Academy of Sciences* **79**(7): 2207-2210.

Yuan R, Meselson M. 1970. A specific complex between a restriction endonuclease and its DNA substrate. *Proceedings of the National Academy of Sciences* **65**(2): 357-362.

Ziolecki A, Briggs CAE. 1961. The microflora of the rumen of the young calf. II. Source, nature and development. *Journal of Applied Microbiology* **24**(2): 148-163.

Zuckerkandl E, Pauling L. 1965. Molecules as documents of evolutionary history. *Journal of Theoretical Biology* **8**(2): 357-366.

CHAPTER II

DOUBLE STRANDED DNA VIROME OF THE BOVINE RUMEN

INTRODUCTION

Viruses are the most abundant form of life on earth. Bacteriophages are viruses that infect bacteria and account for the largest number and diversity of the global virome (Regenmortel et al. 2000). The tailed bacteriophages are the largest group of bacterial viruses, and account for 96% of the reported bacteriophages. They are classified under the order of Caudovirales, which consists of three families; Myoviridae (24%), Siphoviridae (62%) and Podoviridae (14%) (King et al. 2012).It has been estimated that there are 10^{31} tailed bacteriophage particles present on earth, and if laid end to end would span up to 200 million light years (Krupovic et al. 2011). Ocean and lake waters contain an estimated number of 10^{6} phage particles per milliliter and 10^{9} phage particles per gram of sediment (Maranger and Bird 1996; Danovaro et al. 2001; Hewson et al. 2001). Bacteriophages are natural predators of bacteria and have an impact on bacterial population densities and community structure (Wilhelm and Suttle 1999; Rohwer et al. 2009). There are about 10^{24} phage infections taking place every second maintaining the vast phage population inhabiting the global biosphere (Krupovic 2011). An estimated number 100 million metric tons (10^{25} microbial cells) are destroyed by viruses every minute (Rohwer et al. 2009). They play an important role in maintaining the other

microbial populations under the carrying capacity of the global biosphere (Rohwer et al. 2009). Hence, phages have a profound effect on nutrient and biogeochemical cycling, horizontal gene transfer, antimicrobial resistance and bacterial virulence (Wilhelm and Suttle 1999; Casas and Rohwer 2007; Rohwer et al. 2009). Horizontal gene transfer refers to the process by which exchange of DNA occurs among different species. Transduction is one of the three main mechanism of horizontal gene transfer in which the genes move from one prokaryotic species to another via bacteriophages. Transduction plays an important role in evolution of microbial genomes (Jain et al. 2002).

The viruses inhabiting the rumen of cattle are equivalent to any of the above mentioned environments in terms of number and diversity. It has been estimated that the bacteriophage count in a milliliter of rumen fluid is between 3 x 10^9 and 1.6 x 10^{10} particles (Klieve and Swain 1993). Rumen bacteriophages play an important role in population dynamics of rumen bacteria, and they cause a specific and considerable lysis of bacteria (Klieve and Swain 1993). Lysis of crucial bacteria involved in metabolic pathways inside the rumen could alter the foregut fermentation and the total animal performance. However, very little is known about the diversity of bacteriophages inside the rumen and their effects on bacteria. Only one previous study reports on rumen virome and its diversity based on high throughput sequencing of virus enriched samples (Berg Miller et al. 2012).

In general, a majority of the characterized bacteriophages contain double stranded DNA in their genomes, except a few families carrying single stranded DNA, single

stranded RNA, or double stranded RNA (Regenmortel et al. 2000). This study was geared to isolate the double stranded (ds) component of the rumen virome (this includes all tailed bacteriophages) and to study their diversity and role inside the rumen by high throughput sequencing methods.

MATERIALS AND METHODS

Rumen fluid was collected from canulated steers fed with a high protein, hay-based diet and filtered with 4 layers of cheesecloth to remove bigger particles. The resulting filtrate was then blended with a kitchen blender to detach virus particles associated with larger particles and bacteria. Next, the rumen fluid was centrifuged at 8,000 rpm (Beckman Coulter Avanti J-E centrifuge with a JA-14 rotor) at $4^{o}C$ for 20 minutes to remove the smaller particles. The supernatant was pipetted out without disturbing the sediment and filtered through a 0.45µm filter (Surfactant Free Cellulose Acetate filter, Nalgene, Thermo Scientific) at $4^{o}C$. Then, the sample was further filtered through a 0.2µm fine filter (Polyethersulfone filter, Nalgene, Thermo Scientific). Subsequently, the filtrate was centrifuged with 15% polyethylene glycol (PEG) 8,000 and 2% NaCl at 15,000 rpm (Beckman Coulter Avanti J-E centrifuge with a JA-14 rotor) for 2 hours at $4^{o}C$. The sediment was saved and dissolved in a virus isolation buffer (10mM Tris-HCl pH 7.5 with 10mM $MgCl_2$). A CsCl block gradient and was prepared in an ultrahigh centrifugation tube (Beckman 16X76mm) by adding a 1.22g/ml CsCl solution over a 1.47g/ml solution. The interface between the two CsCl solutions was marked on the tube. Then, the virus isolation buffer from the previous step was carefully poured on top of the 1.22g/ml CsCl solution and spun at 40,000 rpm (Beckman L8-70M ultracentrifuge with a NVT-65 rotor) overnight (approximately 16 hours) at $4^{o}C$. The tube was pierced with a needle at the interface and the contents of the interface as well as material adhering to the wall of the tube were aspirated into a 1.5 ml tube. The aspirate was digested with DNase-I to remove any DNA floating outside the protein capsid of virus particles. DNase-I digestion mix contained 135µl above virus suspension, 15µl of

10X buffer, and 4U of DNase. This mix was then incubated at 37°C for one hour followed by inactivation of DNase by further incubating at 75°C for 15 minutes and 95°C for 5 minutes. The viral DNA was extracted with QIAamp DNA Mini kit (Qiagen, Valencia, CA, USA) following manufacturer's protocol. The viral DNA was then subjected to restriction digestion with three enzymes *Sau3AI*, *MseI,* and *FatI*. Linker adapters were ligated to the digested viral DNA (Johnson 1990). Two modifed linker adapters were used with *MseI* and *FatI.*

RESTRICTION DIGESTION OF VIRAL DNA

A 25µl digestion mix contained 1X NEBuffer and 1U of the enzyme (New England Biolabs, Ipswich, MA), and 200ng of viral DNA. For the enzymes *MseI* and *Sau3AI,* the above mix was supplemented with BSA at a final concentration of 100µg/ml and incubated at 37°C for 1 hour for digestion. The digestion mix for *FatI* was incubated 55°C for 1 hour. All three digested products were incubated at 65°C for 20 min to heat inactivate the enzymes.

LINKER ADAPTER LIGATION

1µM linker adapter solution was prepared by mixing the two primers described in Fig. 4. In order to make the linker adapters, this primer mixture was heated to 95°C and then ramped to 4°C by cooling the sample at 0.1^{o}C per second in a MJ dyad thermocycler (MJ research, Watertown, MA). Next, the linker adapters were ligated to the restriction digested DNA. A 12µl ligation mix contained 1X T4 ligase buffer, 2 picomoles of linker

adapter prepared above, 5U of T4 ligase (Fermentas Inc. Glen Burnie, MD) and 30ng of digested viral DNA. The ligation mix was incubated at room temperature for 30 min.

A single primer based on the linkers were used to PCR amplify the DNA. A 25µl PCR reaction contained ~ 3ng viral DNA, 1X PCR buffer (Promega, Madison, WI, USA), 1.5mM MgCl2, 15 picomoles of the primer (GGATTTGCTGGTGCAGTACA), 200µM dNTPs, and 1.5U Taq DNA polymerase (Promega, Madison, WI, USA). The cycling conditions were: one cycle of 2 min at 95ºC, 30 sec at 60ºC, 60 sec at 72ºC followed by 34 more cycles of 30 sec at 95ºC, 30 sec at 60ºC, 60 sec at 72ºC and a final extension of 2 min at 72ºC. This whole process is schematically shown in Fig. 4. Parallel DNA extractions were carried out with and without proteinase K digestion during the viral DNA extraction process to confirm that the resulting DNA was encapsulated by a protein coat. Transmission electron microscopy was carried out by negative staining of CsCl purified material to confirm that the samples did not contain any cellular organisms such as bacteria and organelles such as mitochondria or chloroplast.

The above PCR products were sequenced using tag-encoded FLX amplicon pyrosequencing (bTEFAP) using a Roche 454 FLX Genome Sequencer system. Following sequencing, tags, primers, low-quality sequence ends, and chimeric sequences were removed (Bailey et al. 2010). The putative viral sequences were annotated using blastn, tblastx and blastp searches with the BLAST2GO program (http://www.blast2go.com/b2ghome) against the NCBI non-redundant database. An e-value cutoff of 10^{-6}, high scoring segment pair (HSP) cutoff length of 33, and a maximum of 20 reported hits were used. This step was followed by mapping and annotation (e-

value hit filter 1e-6, annotation cutoff 55and GO weight 5) to assign the respective functions associated with the Gene Ontology (GO) terms. The resulting xml files were corrected for missing and duplicate sequences. Then, the xml files were manually formatted and imported to the metagenomic analyzing software MEGAN version 4.62.5 to assign sequences to taxonomic classes, and to perform SEED, and KEGG analysis. A bit score cutoff value of 50 was used to classify sequences in MEGAN. Gene Ontology analyses were conducted in the Blast2GO program with an alpha score of 0.6.

TRANSMISSION ELECTRON MICROSCOPY

Negatively stained virus particles were observed under a transmission electron microscope. The sample, CsCl gradient purified viral particles dissolved in virus isolation buffer, was placed on a carbon coated grid, and allowed to sit for one minute. Then, the excess liquid was wicked off with a piece of filter paper. Next, a drop of 2% phosphotungstic acid was added onto the gird, allowed to sit for one minute, and wicked off with a filter paper. Finally, grid was allowed to dry and observed under a JEOL JEM-2100 transmission electron microscope.

RESULTS

Transmission electron microscopy revealed that the samples only contain viral particles after filtering and enriching for viruses. The most commonly observed viral particle type was tailed bacteriophages (Fig. 6). Parallel DNA extractions with and without proteinase K proved that there was no free floating DNA in the extraction and all recovered DNA was covered in a protein The sample subjected to proteinase K released protein encapsulated DNA, while the sample without proteinase K did not release the DNA inside the viral capsid and did not produce any PCR amplification (Fig.5). Also, this further proved that the sample did not have any floating double stranded DNA outside the viral particle after DNase-I treatment. We further confirmed the absence of bacterial DNA by performing a 16S rRNA gene-based PCR assay (data not shown).

Three restriction enzymes, with four nucleotide recognition sites were used to digest viral DNA before the linker mediated PCR amplification as described in the method section. This was designed to generate overlapping sequence reads, which later allows contig generation. The three enzymes used were carefully selected to maximize the resulting sequence coverage, to avoid star activity, and for being heat labile for easy inactivation. Further, this method yields restriction digested DNA fragments with overhang sequences. The DNA fragments with overhang sequences are extremely efficient in binding to the linker adapters compared to the blunt end DNA fragments. Subsequent improved sensitivity of the DNA amplification process is very useful when working with very small quantities of DNA. After high throughput sequencing of the amplified DNA, viral contigs were generated using GS *De Novo* Newbler Assembler (Roche 454 Life Sciences, Branford, CT). It was found that these sequence assembling

software have over assembled the sequences into false contigs even at a high level of stringency (40bp overlap with a maximum of 10% mismatches) reducing the viral diversity. There are other viral metagenomics studies that have reported similar observations (Breitbart et al. 2002). Therefore, the putative viral sequences were analyzed as separate and individual reads.

A total of 184,104 sequences were obtained by Roche 454 pyrosequencing, and after quality trimming 147,734 reads of highest quality were used for downstream analysis. The blastn information processed by MEGAN (Huson et al. 2011) gave an overview of the distribution of sequences among different taxonomic classes (Fig. 7A). However, a majority of the sequences (96.6%) did not have any homologous sequences in public databases. The rest of the putative viral sequences showed a similarity to bacteria (3.4%), known viruses (0.6%), eukaryotes (0.2%), and archaea (~0.0%) above a bit score of 50. With the tblastx analysis, the number of unidentified putative viral sequences decreased to 80.8% (Fig. 7B) with the same bit score in MEGAN. As a result the amount of putative viral sequences showing a similarity to bacterial (13.5%), known viral (4.7%), eukaryotic (0.2%), and archaeal (0.1%) sequences turned out to be comparatively high.

Fig. 8 shows a detailed overview of the viral sequence classification in MEGAN based on the tblastx annotations. Out of the putative viral sequences annotated as viruses, 85% belonged to the order Caudovirales. Other sequences belonged to all three families of the tailed bacteriophages; Myoviridae (13.8%), Podoviridae (1.5%), and Siphoviridae (69.8%). 14.4% of the sequences were identified as unclassified phages and viruses.

Blastx search gave the highest number of annotations (26.6%) for the putative viral sequences compared to blastn (4.4%) and tbalstx (19.2%). Fig. 7C depicts the taxonomic diversity of the putative viral sequences in the bovine rumen based on the blastx search. This improvement is predominantly due to a high number of annotations obtained for viral sequences resulting in similar proteins already described in bacteria (22%) compared to the tblastx search (13.5%). The functions of these viral genes were studied based on GO annotations. An overview of the GO terms related to biological processes that are associated with putative viral genes is shown in Fig. 9A. A majority of the viral genes were related to metabolic processes (45%), cellular processes (38%), biological regulation (6%), and cellular component organization and biogenesis (5%).

Specific biological processes associated with the annotated viral genes are given in Fig. 9B. There were many viral genes related to peptidoglycan (a component in the cell wall of bacteria) catabolism, DNA methylation, viral genome packaging, regulation of transcription, nucleotide biosynthesis, protein splicing, DNA replication, and tRNA metabolism. The molecular functions of the annotated viral genes are presented in Fig. 10A. Out of the annotated viral genes, 32% belonged to hydrolases, 27% binds to nucleic acids, and 25% had transferase activity. Figure 10B provides the specific molecular functions of the most abundant putative viral genes. These molecular functions include: nucleoside triphosphatase, ATP binding, DNA and RNA polymerase, DNA methyltransferase, and zinc ion binding activities.

DISCUSSION

This study was designed to isolate the double stranded (ds) DNA component of the rumen virome and to study its diversity and role inside the rumen by high throughput sequencing methods. Transmission electron microscopy in combination with several other quality control steps ensured the presence of only viral DNA in the samples. This is particularly important in studying the viral metagenomes of complex environments such as the bovine rumen or human gut with a complex assortment of microorganisms. These microorganisms have nucleotide sequences with little or no similarity to the sequences present in public databases. With the increasingly available bacterial sequence data in pubic databases, novel bacteriophage and other viral sequences are often misidentified as many viral sequences are present in bacteria as proviral sequences. During high throughput shotgun sequencing even a slight contamination of viral metagenomic DNA by any extraneous DNA such as bacterial DNA could make a relatively high contribution to the viral DNA pool due to larger genome sizes of bacteria compared to that of viruses. For an example *E.coli* K-12 genome size is about 4.6 million base pairs (Blattner et al. 1997) and the average phage genome size range is 15,600-27,200 base pairs (Angly et al. 2009). A contamination resulted from a single bacterial genome could contribute up to 200 times the DNA of a bacteriophage. Out of the putative viral sequences obtained, only about 26% could be annotated using the information present in the public databases. This indicates the lack of information available for the rumen virome and the future potential in exploring the dsDNA virome of the rumen. It is a well-accepted that about 50-90% of the viral metagenomes present in complex environments such as animal gut do not code for similar proteins found in already known viruses or cellular organisms (Edwards and

Rohwer 2005; Kristensen et al. 2010). In this study, a higher level of similarity was observed with viral sequences at protein level than at nucleotide level. This is probably due to the degeneracy of the genetic code and selective pressure driving to conserve viral proteins. Only 4.7% of the putative viral sequences identified in this study had a significant similarity to the viral sequences in GenBank. Out of these sequences 85% belong to the order Caudovirales. This is not entirely surprising since this group is comprised of tailed bacteriophages. Tailed bacteriophages are the most abundant type of viruses present in the rumen (Klieve and Swain 1993). During sample preparation, rumen samples were enriched for bacteriophages initially and then, selected only the double stranded component of the rumen virome by a selective DNA amplification process. A vast majority of known bacterial viruses are tailed bacteriophages and have dsDNA genomes, which is the target for our linker mediated amplification process. There are only a few examples of phages that lack tails or have single stranded (ss) DNA or RNA genomes. The rumen ecosystem has a diverse population of bacteria and the estimated bacterial cell count in a milliliter of rumen fluid is 10^9 cells (Hungate 1966; Miron et al. 2001). This makes the rumen an ideal environment to study the dsDNA containing bacteriophage diversity. Viral sequences belonging to all three families of the order Caudovirales were found in this study. Among the annotated viral sequences the members of the family Siphoviridae were the most abundant in the rumen virome.

A study conducted by Berg Miller et al. (2012) reported that the prophages outnumber the lytic phages approximately by a ratio of 2:1 in the rumen ecosystem. However, it is not clear in this report whether prophage refers to the bacterial prophage elements or lytic phages having similar sequences to bacterial prophages since they have

failed to completely isolate viral DNA. Also, the authors report that these prophages or viral sequences similar to known prophage elements in bacteria were mainly coming from Firmicutes (68±1) and Proteobacteria (18±1). The Bacteroidetes related prophages were relatively low (8±1) in abundance. Further, this paper reports that approximately 2% of the sequences had a significant similarity to known viral sequences deposited in GenBank and 72%±1 did not have any significant hits (based on tblastx search with an e-value cutoff of 10^{-3}). In the present study, 3.7% of the sequences could be annotated to viral sequences in GenBank (with an e-value cutoff of $10^{-6}$and a bit score cutoff of 50), and 73% of the sequences did not have any significant hits. These sequence annotation statistics were similar between the two studies beside the differences in methodology and samples. Many lytic phage particles sharing sequence similarity with bacteria (previously identified prophages and unidentified prophage regions) were found in the present study. These lytic phages had sequence similarity to Firmicutes (56%), Bacteroidetes (13%), Proteobacteria (3%), Actinobacteria (3%), Spirochetes (0.4) and other rumen bacteria and prophages found in them (25%). Firmicutes and Bacteroidetes are considered as the most abundant rumen bacterial phyla under the both high concentrate and high forage diets (Fernando et al. 2010). This information provides an insight to the rumen bacteriophage host range. It has been estimated that 60% of the sequenced bacterial genomes contain a minimum of one prophage sequence and on average about 3% of the bacterial DNA is of viral origin (Edwards and Rohwer 2005). The relative abundances of the sequences belonging to Siphoviridae (70%), Myoviridae (13.77%), and Podoviridae (1.54%) families found in the our study was significantly different from what is described by Berg Miller et al (2012), which were 36±3%, 28±4% and 14±2%, respectively.

Viruses carry minimum amount of genetic material essential for their propagation since viruses are under a continuous pressure to minimize their genome size to sustain and produce virus particles within a host replication cycle. As the genomes get bigger, the capsids have to be larger in size requiring more time and energy to be synthesized. Therefore, the conventional thinking does not support the presence of certain genes in the viral genomes that are redundant with the host genome. With recent advancement of metagenomic techniques scientist have found that viruses do carry certain host genes when it is advantageous. (Lindell et al. 2005). Results of the present study revealed that the rumen viruses carry nucleoside triphosphatase, DNA and RNA polymerase, DNA methyltransferase and many other redundant genes with their host genomes. Several previous studies have reported similar findings with respect to bacteriophages (Breitbart et al. 2002; Reyes et al. 2010).

In order to investigate the possible role of rumen viruses, a detailed analysis of the annotated viral genes was conducted. It was found that the sequences associated with the GO term nucleoside triphosphatase were DNA helicases with nucleoside triphosphatase activity. DNA helicases are motor proteins that separate double stranded nucleic acids using energy derived from hydrolysis of nucleoside triphosphates and play a key role in DNA replication. DNA helicases with nucleoside triphosphatase activity has been described in viral genomes previously (Xu et al. 2005; Xiang et al. 2012). The sequences annotated as DNA helicase had a very low sequence homology to the previously described viral DNA helicases at DNA as well as protein levels. This hindered the process of sequence alignment and subsequent meaningful prediction of phylogenetic relationships.

This study sheds light on the diversity of the dsDNA virome of the rumen. An improved understanding of viral diversity and viral gene pool in the gut microbiome would definitely alter the perception of future animal scientists when feeding livestock to maximize feed efficiency toward an increased animal performance by promoting a beneficial and healthier gut microbiome.

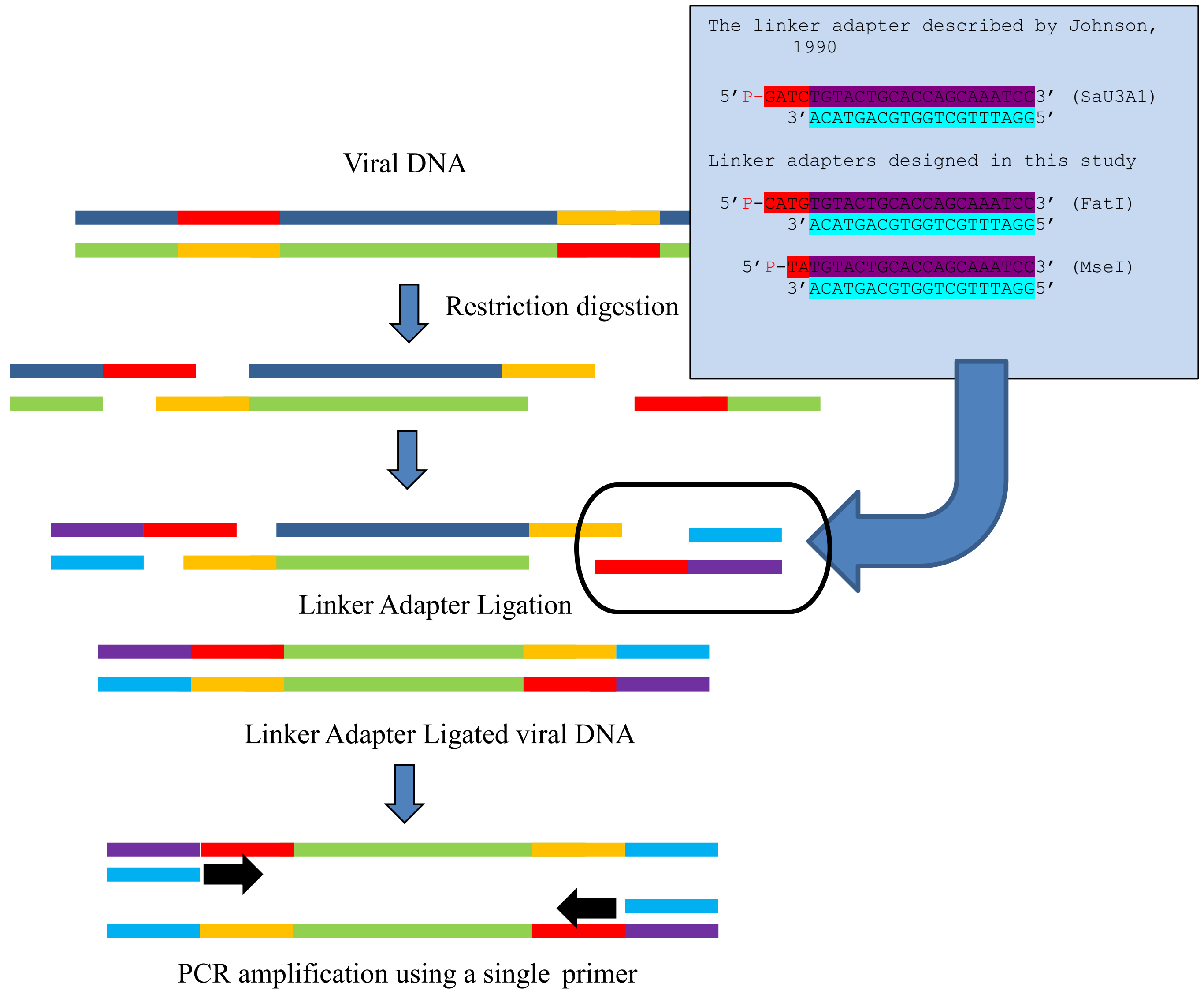

Figure 4. Schematic diagram showing the linker mediated PCR amplification process

Figure 4. Schematic diagram showing the linker mediated PCR amplification process. Blue and green lines represent the double stranded viral DNA. Red and yellow lines represent the restriction enzyme recognition sequences. Purple and blue lines show the linker adapters with the red colored overhang sequence, complementary to the opposite strand restriction enzyme recognition site. The short blue lines indicate the single PCR primer bound to denatured DNA and black colored arrows show the direction of PCR amplification.

Fig. 5A)

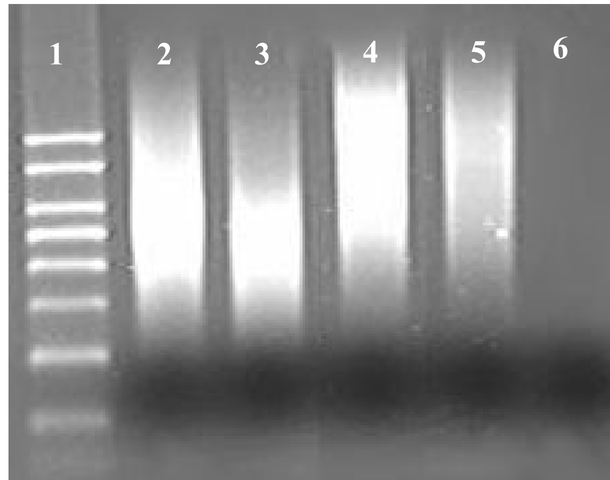

Fig. 5B)

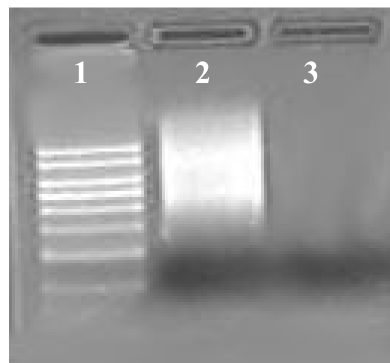

Figure 5. Agarose gel electropherograms with viral DNA amplicons

Figure 5. Agarose gel electropherograms with viral DNA amplicons. Fig. 5A. Lane 1 DNA size standard. Lanes 2, 3, and 4 - *Sau3AI*, *MseI* and *FatI* digested DNA respectively. Lanes 5 and 6 were positive control (rumen DNA) and negative control (water) respectively. Fig. 5B. Shows PCR products of the material through parallel DNA extractions with proteinase K (lane 2) and without digestion (lane 3).

Figure 6. Transmission electron micrographs of negatively stained virus particles

Figure 6. Transmission electron micrographs of negatively stained virus particles. In transmission election microscopic studies, the most commonly observed viral type was tailed bacteriophages. These images show a fraction of the various head and tail morphologies encountered in the rumen samples (magnification of 30,000 X).

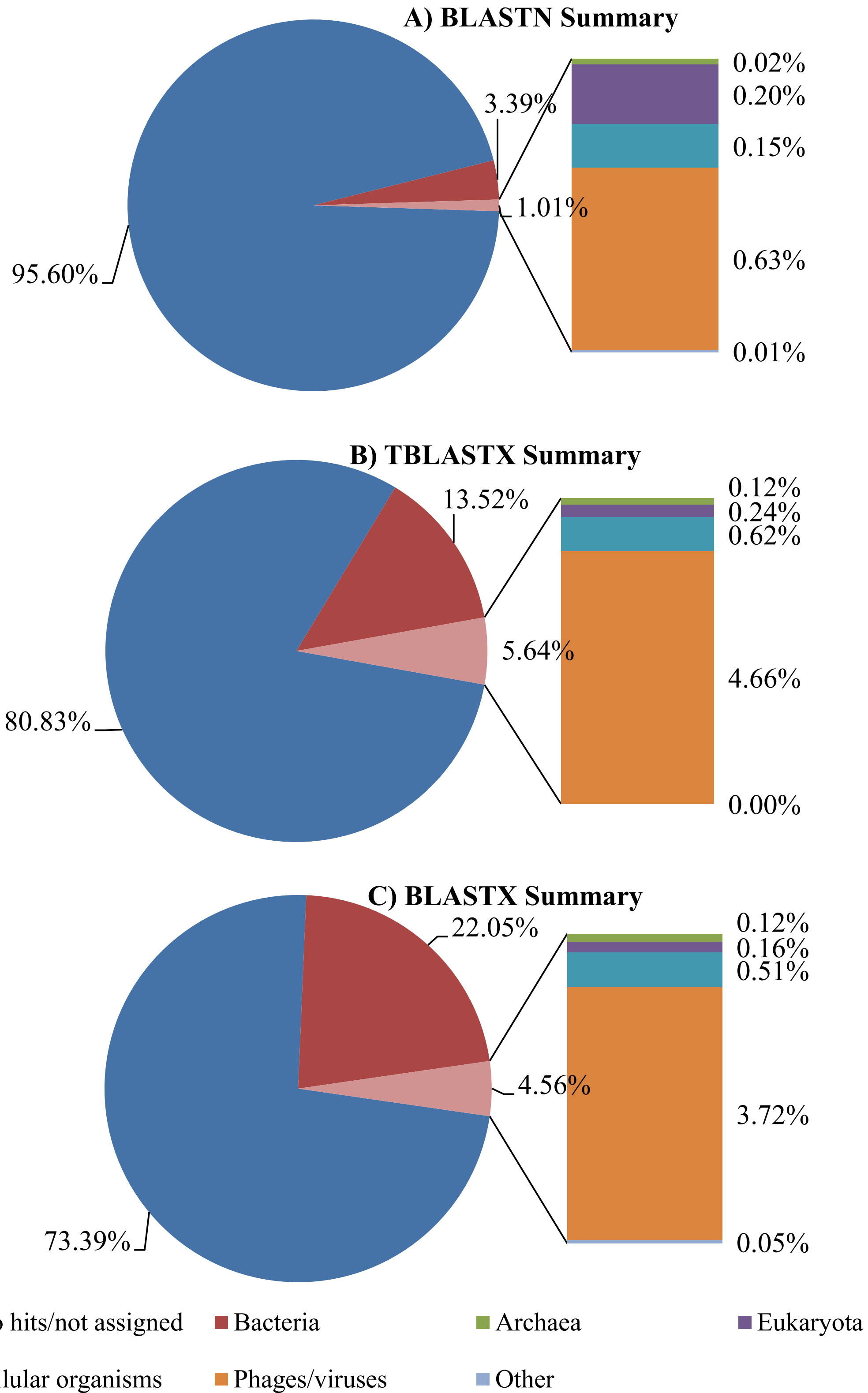

Figure 7. Similarity of the rumen viral sequences to the domains of life and viruses

Figure 7. Similarity of the isolated rumen viral sequences to the domains of life and viruses. The summary was generated using MEGAN metagenomic analyzer based on the tblastx and blastx searches against GenBank non-redundant databases. The e-value was set at 10E-6 in blast searches, and the bit score cut-off was set at 50 to assign taxa in MEGAN. For any sequence having more than one hit above the threshold values from two different taxa were pushed to the root of the two taxa in MEGAN. i.e. cellular organisms represent sequences with having at least two hits from two groups of cellular organisms.

Fig. 8A) Signifcant hits to known viruses

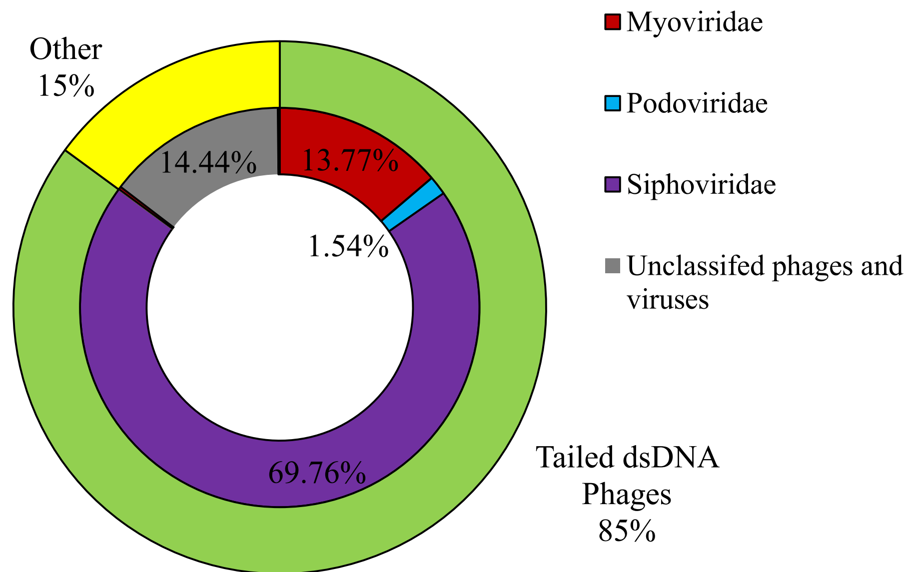

Fig. 8B) Signifcant hits to bacteria

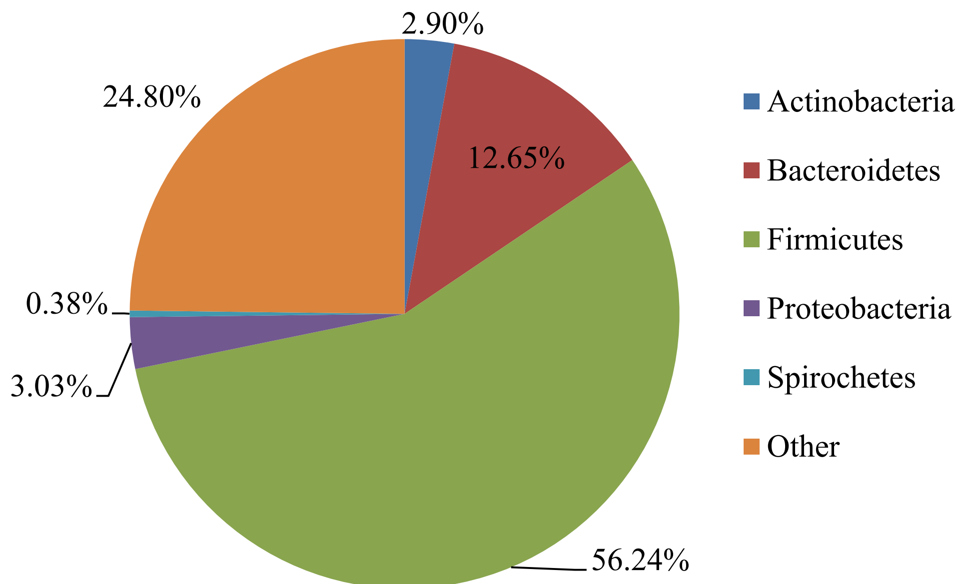

Figure 8. Taxonomic classification of the putative rumen viral sequences

Figure 8. Taxonomic classification of the putative rumen viral sequences. The outer circle of Fig. 8A shows the proportions of tblastx hits classified under the order Caudovirales and other viral sequences in MEGAN above a bit score value of 50. The order Caudovirales is represented by three tailed bacteriophage families namely Siphoviridae, Myoviridae and Podoviridae. The inner circle shows the distribution of Caudovirales hits among its three families. There were few other viral families encountered in the study, apart from the unclassified viruses, and are unable to see in Fig.8A corresponding to "other" in the outer circle due to their low numbers. Fig. 8B shows the distribution of tblastx hits among bacterial phyla. This group is represented by lytic phages having high sequence similarity to bacterial genomic regions such as prophage elements.

Fig. 9A)

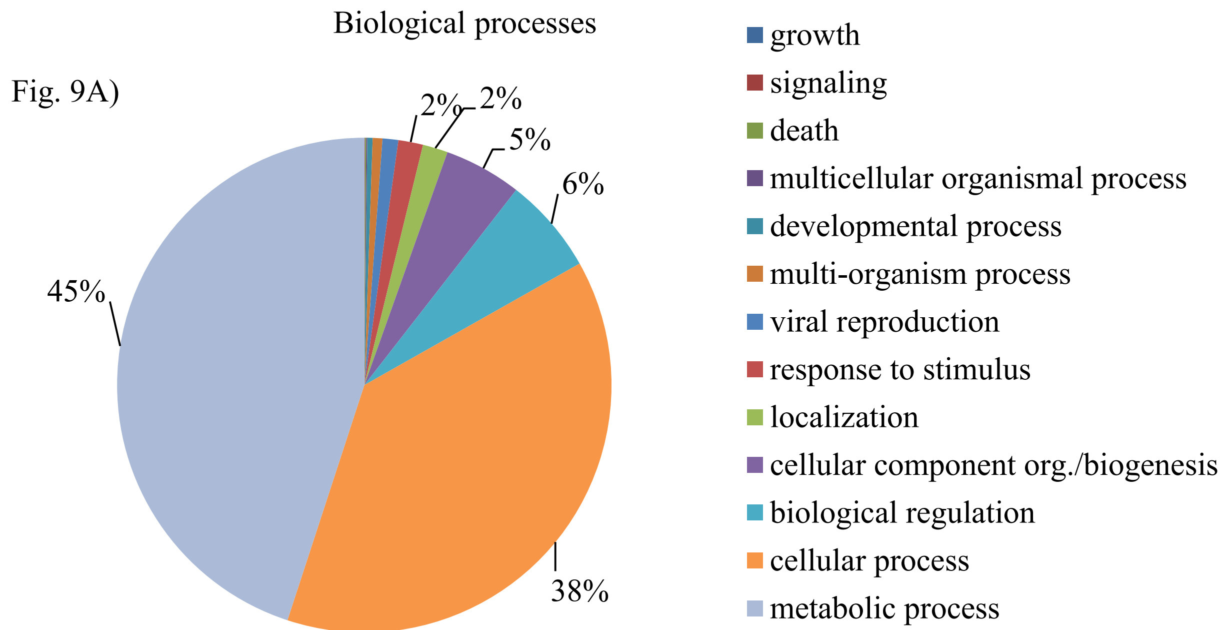

Fig. 9B)

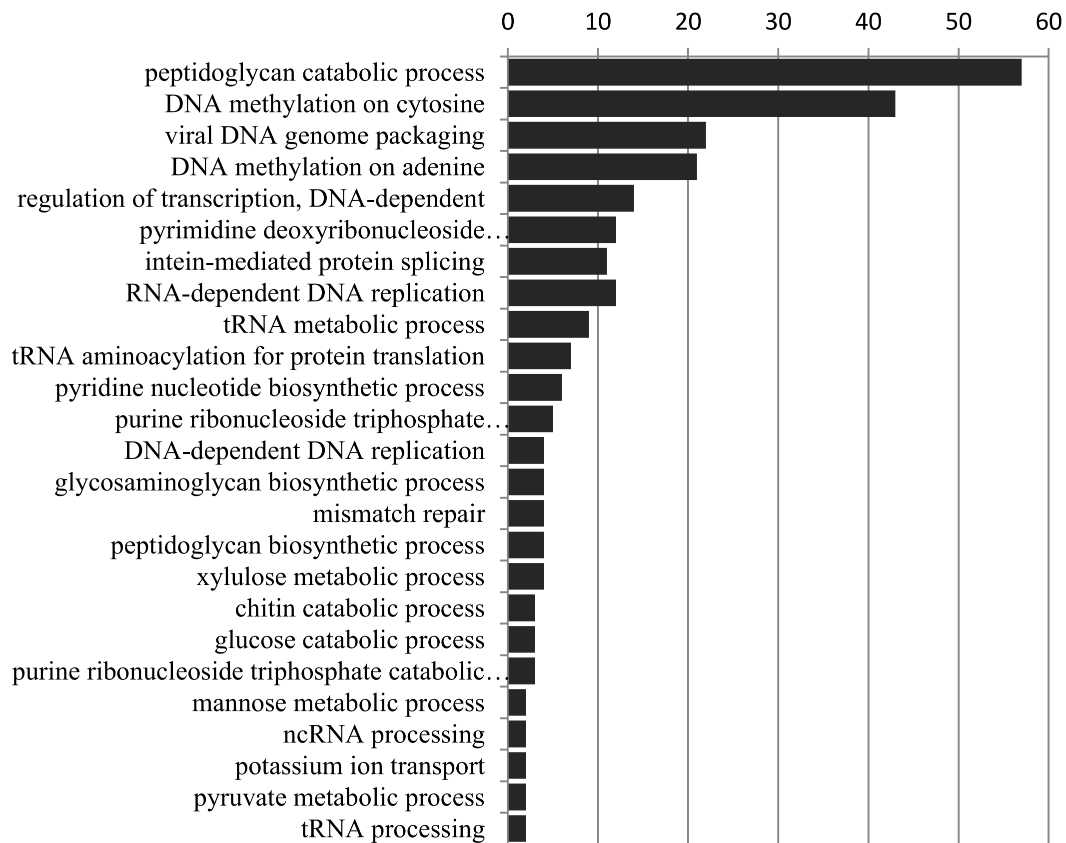

Figure 9. Biological processes associated with the bovine rumen virome

Figure 9. Biological processes associated with the bovine rumen virome. The abundance and distribution of GO terms related to biological processes associated with the viral sequences retrieved by the BLAST2GO program (based on blastx) is shown in Fig. 9. Fig. 9A gives a general overview of the spectrum of biological processes while Fig. 9B provides more specific biological processes associated with the putative viral sequences.

Fig. 10A) GO terms of molecular function (level-3)

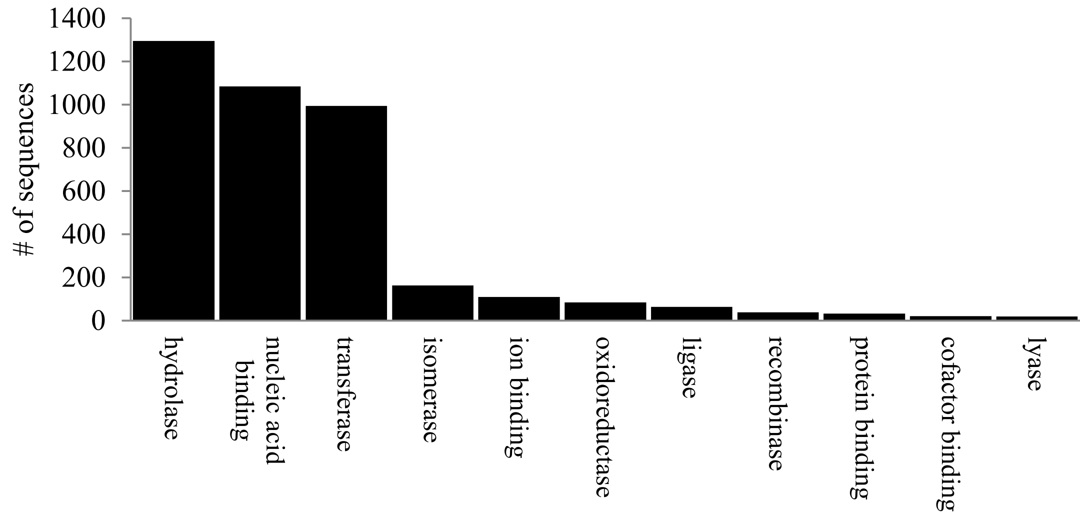

Fig. 10B)

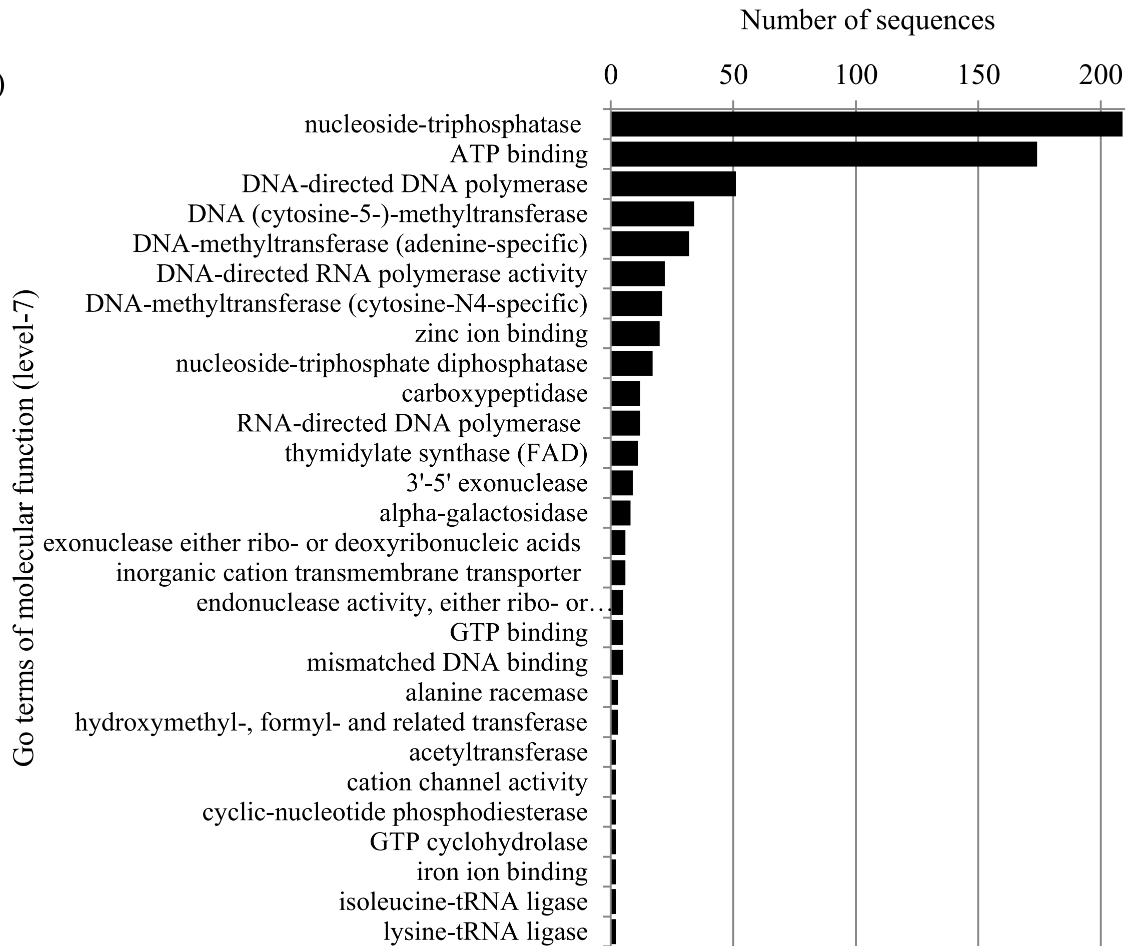

Figure 10. Molecular functions associated with the bovine rumen virome

Figure 10. Molecular functions associated with the bovine rumen virome. Fig. 10A shows an overview of the GO terms related to molecular function associated with the viral genes. A majority of the genes found to be hydrolases and transferases with nucleic acid binding capability. Fig. 10B provides a more detail description of the functions summarized in Fig. 10A.

REFERENCES

Angly FE, Willner D, Prieto-Davó A, Edwards RA, Schmieder R, Vega-Thurber R, Antonopoulos DA, Barott K, Cottrell MT, Desnues C et al. 2009. The GAAS metagenomic tool and its estimations of viral and microbial average genome size in four major biomes. *PLoS Comput Biol* **5**(12): e1000593.

Bailey MT, Dowd SE, Parry NMA, Galley JD, Schauer DB, Lyte M. 2010. Stressor exposure disrupts commensal microbial populations in the intestines and leads to increased colonization by *Citrobacter rodentium*. *Infection and Immunity* **78**(4): 1509-1519.

Berg Miller ME, Yeoman CJ, Chia N, Tringe SG, Angly FE, Edwards RA, Flint HJ, Lamed R, Bayer EA, White BA. 2012. Phage–bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome. *Environmental Microbiology* **14**(1): 207-227.

Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF et al. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**(5331): 1453-1462.

Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, Azam F, Rohwer F. 2002. Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences of the United States of America* **99**(22): 14250-14255.

Casas V, Rohwer F. 2007. Phage metagenomics. *Advanced Bacterial Genetics: Use of Transposons and Phage for Genomic Engieering* **421**: 259-268.

Danovaro R, Dell'Anno A, Trucco A, Serresi M, Vanucci S. 2001. Determination of virus abundance in marine sediments. *Appl Environ Microbiol* **67**(3): 1384-1387.

Edwards RA, Rohwer F. 2005. Viral metagenomics. *Nat Rev Micro* **3**(6): 504-510.

Fernando SC, Purvis HT, II, Najar FZ, Sukharnikov LO, Krehbiel CR, Nagaraja TG, Roe BA, DeSilva U. 2010. Rumen microbial population dynamics during adaptation to a high grain diet. *Appl Environ Microbiol*: **76**(22): 7482-7490.

Hewson I, O'Neil JM, Fuhrman JA, Dennison WC. 2001. Virus-like particle distribution and abundance in sediments and overlying waters along eutrophication gradients in two subtropical estuaries. *Limnology and Oceanography* **46**(7): 1734-1746.

Hungate RE, ed. 1966. The rumen and its microbes. Academic Press, Inc., New York.

Huson DH, Mitra S, Ruscheweyh H-J, Weber N, Schuster SC. 2011. Integrative analysis of environmental sequences using MEGAN4. Genome Res 21(9): 1552-1560.

Jain R, Rivera MC, Moore JE, Lake JA. 2002. Horizontal gene transfer in microbial genome evolution. Theoretical Population Biology 61(4): 489-495.

Johnson DH. 1990. Molecular cloning of DNA from specific chromosomal regions by microdissection and sequence-independent amplification of DNA. *Genomics* **6**(2): 243-251.

King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ. 2012. Virus Taxonomy. Ninth report of the International Committee on Taxonomy of Viruses, San Diego.

Klieve AV, Swain RA. 1993. Estimation of ruminal bacteriophage numbers by pulsed-field gel-electrophoresis and laser densitometry. *Applied and Environmental Microbiology* **59**(7): 2299-2303.

Kristensen DM, Mushegian AR, Dolja VV, Koonin EV. 2010. New dimensions of the virus world discovered through metagenomics. *Trends in Microbiology* **18**(1): 11-19.

Krupovic M, Prangishvili D, Hendrix RW, Bamford DH. 2011. Genomics of bacterial and archaeal viruses: Dynamics within the prokaryotic virosphere. *Microbiology and Molecular Biology Reviews* **75**(4): 610-635.

Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW. 2005. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**(7064): 86-89.

Maranger R, Bird DF. 1996. High concentrations of viruses in the sediments of lac gilbert, québec. *Microbial Ecology* **31**(2): 141-151.

Miron J, Ben-Ghedalia D, Morrison M. 2001. Invited review: Adhesion mechanisms of rumen cellulolytic bacteria. J Dairy Sci 84(6): 1294-1309.

Regenmortel MHVv, Fauquet CM, Bishop DHL, Carstens EB, Estes MK, Lemon SM, Maniloff J, Mayo MA, McGeoch DJ, Pringle CR et al. 2000. Virus Taxonomy. Classification and nomenclature of viruses. Seventh report of the International Committee on Taxonomy of Viruses, San Diego.

Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, Gordon JI. 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**(7304): 334-338.

Rohwer F, Prangishvili D, Lindell D. 2009. Roles of viruses in the environment. *Environmental Microbiology* **11**(11): 2771-2774.

Wilhelm SW, Suttle CA. 1999. Viruses and nutrient cycles in the sea. *BioScience* **49**(10): 781.

Xiang H, Ishibashi K, Nishikiori M, Jaudal MC, Ishikawa M, Katoh E. 2012. Expression, purification, and functional characterization of a stable helicase domain from a tomato mosaic virus replication protein. *Protein Expression and Purification* **81**(1): 89-95.

Xu T, Sampath A, Chao A, Wen D, Nanao M, Chene P, Vasudevan SG, Lescar J. 2005. Structure of the Dengue virus helicase/nucleoside triphosphatase catalytic domain at a resolution of 2.4 A. *J Virol* **79**(16): 10278-10288.

CHAPTER III

DOUBLE STRANDED VIRAL DNA METHYLOME OF THE BOVINE RUMEN

INTRODUCTION

Epigenetics provides an additional stability and diversity to the phenotype of a cell by modifying the process of transcription through methylating nucleotides that are maintained or regenerated during cell division (Laird 2010). In eukaryotic organisms methylation of a cytosine base at the fifth carbon atom is a common epigenetic landmark. Numerous methyltransferases have been found in higher eukaryotes, prokaryotes, and viruses that methylate the fifth carbon of the pyrimidine ring of a cytosine base (Kumar et al. 1994). In prokaryotes, DNA adenine methyltransferase (Dam) recognizes a specific tetranucleotide sequence, 'GATC', and methylates the second adenine base. Similarly, DNA cytosine methyltransferase (Dcm) methylates the fifth carbon atom of the second cytosine bases in specific sequences identified as 'CCAGG' and 'CCTGG' (Marinus and Morris 1973; May and Hattman 1975) in prokaryotes. There are many other methyltransferases described in prokaryotes (Kumar et al. 1994). These methylated residues in bacteria are mainly important in restriction-modification systems and mismatch repair strand discrimination (Laird 2010). Bacteria and archaea are known to carry 5-methylcytosine residues, N-4-methylcytosine and N-6-methyladenine.

Unmethylated restriction sites have been shown to be vulnerable to bacterial restriction enzymes, while methylated restriction sites either have a reduced affinity or no affinity to restriction enzymes (Yuan and Meselson 1970; Smith 1979). There are reports of bacteriophages carrying methylated bases (May and Hattman 1975). For an example one study has found that nearly 50% of the bacteriophage lambda DNA is Dam methylated (Pirotta 1976), and the level of methylation is probably decided by the time period between DNA replication and packaging into the phage head. DNA sequences coding for several methyltransferases have been found in bacteriophage metagenomes (Breitbart et al. 2008).

In mammals, two to seven percent of the total cytosine is converted to 5-methylcytosine after new DNA is synthesized by the action of DNA methyltransferases (Vanyushin et al. 1970). Methylation of cytosine residues plays an important role in gene expression and differentiation in higher eukaryotes (Razin and Riggs 1980). Eukaryotic DNA methyltransferases (DMTs) can be broadly categorized into five families: 1) DNMT1 or maintenance DMT family, 2) DNMT2 family, 3) DNMT3 or *de novo* DMT family, 4) plant-specific chromomethylase (CMT) family and 5) fungal-specific DMT-like family (Goll and Bestor 2005). In eukaryotes, methylation of DNA takes place at DNA replication forks in a cell cycle dependent manner (Liu et al. 1998). Eukaryotic DNA methyltransferases have shown to methylate CpG islands of prokaryotic and viral DNA (Simon et al. 1978). It is not clear why a dsDNA viral genome present in the nucleus should not be methylated (Karlin et al. 1994). It is possible to think that somehow small viral genomes can escape methylation due to their short replication times.

There are reports of presence of exclusive CpG methyltransferases in bacteria as well (Renbaum et al. 1990). There are reports of epigenetic regulation of viral genomes associated with latent infections. In neurotropic herpesviruses, precursors to infectious progeny are not methylated. The latent genomes of lymphotropic herpesviruses have found to be extensively methylated. Further, the immediate early locus of the latent genomes of cytomegaloviruses has been shown to be methylated (Honess et al. 1989).

Some of the DNA methyltransferase coding sequences isolated in the present study had a significant similarity to cytosine-5-methyltransferases identified from mouse ESTs (AF071754-unpublished), higher plants (e.g. accessions XM_002534542, NM_001247819-unpublished), fungi (DmtA) and CpG methyltransferases found in bacteria. (Renbaum et al. 1990; Machida et al. 2005). DNA methyltransferase A (DmtA), which was isolated form *Aspergillus nidulans,* was shown to be important in formation of viable ascospores (Lee et al. 2008). Therefore, DmtA may have an important epigenetic role in regulation of gene expression in fungi. The authors of the above mouse EST study have observed some in-frame stop codons in the sequence. And further, the authors have not been able to detect any hybridization signal, with respect to this particular sequence AF071754, on a *Mus musculus* DNA blot. These evidences suggest that this particular EST probably may have resulted from transcription of viral DNA. The incidental finding of cytosine-5-methyltransferse coding sequences in viral genomes was further investigated by examining the sequences obtained by digesting with *Sau3AI* restriction enzyme prior to linker mediated PCR amplification. In these sequences, the presence of CpG methylation should be evident as undigested 5'GATC 3' sites in between the sequences with an overlapping CpG dinucleotide (i.e. the sequences were screened for

CGATC pentanucleotide sequences). *Sau3AI* restriction enzyme is blocked when a methylated CpG site overlaps with their recognition sequences (New England Biolabs Inc., Ipswich, MA). The presence of such undigested CGATC pentanucleotide sequences were observed in the viral sequences. However, this observation could also have resulted from partial digestion of viral DNA with *Sau3AI* enzyme. The above observations suggest the presence of CpG methylation in the viral metagenome isolated from the bovine rumen with a certain degree of confidence. Interestingly, neither the CpG methylation of bacteriophage DNA nor methylation of metagenomes isolated from environmental samples have been described previously to the best of our knowledge.

MATERIALS AND METHODS

RESTRICTION DIGESTION OF VIRAL DNA

Double stranded viral DNA isolated, according to the procedure detailed in chapter two, was digested with three restriction enzymes *Sau3AI, MseI* and *FatI.* A 25µl digestion mix contained 1X NEBuffer and 1 U of the enzyme (New England Biolabs, Ipswich, MA) and 200ng viral DNA. For *MseI* and *Sau3AI* above mix was supplemented with BSA at a final concentration of 100µg/ml and incubated at 37°C for 1 hour for digestion. The digestion mix for *FatI* was incubated 55°C for 1 hour. All three digested products were incubated at 65°C for 20 min to heat inactivate the restriction enzymes.

ISOLATING CPG METHYLATED VIRAL DNA

In order to separate and enrich methylated viral DNA, a commercial kit was used (Methylation Promoter PCR Kit, Panomics Inc., Redwood city, California). Since this kit has been optimized to work with larger amounts of eukaryotic DNA, certain modification to the manufacturer's recommended protocol were introduced to make the kit compatible with small quantities of viral DNA available. Four basic steps are involved in this procedure: 1) Restriction digestion of DNA, 2) The resulting DNA fragments are then incubated with MeCP2, which is a methylation binding protein, 3) isolation methylated DNA fragment using a separation column 4) PCR amplification of isolated DNA fragments. In the modified protocol, the DNA purification step after restriction digestion before incubating with the MeCP2 was excluded and an additional cycle of column washing was carried out to separate the MeCP2-DNA complexes from the rest of the DNA.

PCR AMPLIFICATION OF METHYLATED DNA

Double stranded methylated viral DNA isolated in the previous step was PCR amplified using a sequence independent linker adapter mediated process described in chapter two of this document.

RESULTS

DNA without methylated CpG dinucleotides will not bind to the methylation binding protein (MeCP2) and subsequently to the separation column provided with the Methylation Promoter PCR kit (Cross et al. 1994). During washing of the separation column such unbound DNA gets washed away. Only the DNA bound to MeCP2, i.e., DNA with CpG methylation sites, is retained in the separation column after the washing step. Then, the separation column bound DNA can be eluted and presence of CpG methylated DNA can be confirmed by PCR. An absence of a PCR product in the final step indicates that the DNA in the sample was not CpG methylated. Based on the findings of this experiment it is quite conclusive that the viral DNA isolated from the rumen contains methylated CpG dinucleotides (Fig. 11).

Several controls were used in the CpG methylation detection experiment to assure that even a minute quantity of methylated viral DNA results in a positive PCR reaction while unmethylated DNA is totally removed from the separation columns during washing avoiding any false positive reactions. Adding CpG methylated template DNA at two different concentrations helps to overcome false negative PCRs if the methylated viral DNA is minute and becomes a limiting factor (Fig. 11). After the restriction digestion of viral DNA, and control DNA provided with kit, a purification of the digested DNA is recommended by the manufacturer. This step was omitted in our experiment to avoid any losses of DNA since the starting viral DNA amount was low compared the recommended DNA amount. Therefore, additional controls were included in the experiment to monitor whether any PCR inhibitors are coming through prior steps. The third well of gel electropherogram shown in Fig. 11 contained a PCR reaction carried out with template

viral DNA isolated by the kit (methylated DNA) which has been spiked with some extraneous DNA known to give a positive PCR reaction. If this control fails to amplify DNA during PCR there is an impurity in the template DNA. The commercial kit used to separate methylated DNA provides CpG methylated DNA to be used as a control. This DNA was diluted to a similar concentration as the viral DNA and then, used in the experiment.

DISCUSSION

The findings of this experiment confirmed the presence of CpG methylated DNA in rumen virome. This is the first time a CpG methylation is being reported by a metagenomics study. To the best of our knowledge, there are no previous reports of CpG methylated DNA found in bacteriophages. It was experimentally demonstrated that viral DNA subjected to CpG methylation in the presence of a cytosine-5-methyltransferase results in an altered gene expression, though many viral genomes do not carry the genetic information to code for cytosine-5-methyltransferases (Youssoufian et al. 1982; Hu et al. 1991; Minarovits et al. 1991). The exception with bacteriophages is that their hosts (bacteria) generally are not known to carry the methyltransferases required for CpG methylation. However, there are few reports of bacterial methyltransferase capable of producing CpG methylation (Renbaum et al. 1990). Hence, it is not impossible to assume that phage DNA could undergo CpG methylation although it has not been reported previously. There is a fine balance between bacteriophages and their hosts in all studied environments (Rohwer et al. 2009; Reyes et al. 2010). Not all bacteria of a particular species or strain are completely eliminated by their parasitic phages. Besides from the protection they get from the restriction enzymes, CpG methylation is shown be important in keeping this fine balance between bacteria and their phages (Burkhart et al. 1992; Kashlev et al. 1993), which needs to be further investigated. DNA methylation patterns of environmental samples could provide valuable information about how microbial gene expression is epigenetically regulated, if such mechanism exists. Further, we believe that CpG methylation status of environmental metagenomes could potentially provide

valuable information on the fine balance between microbial hosts and their predators in the biosphere.

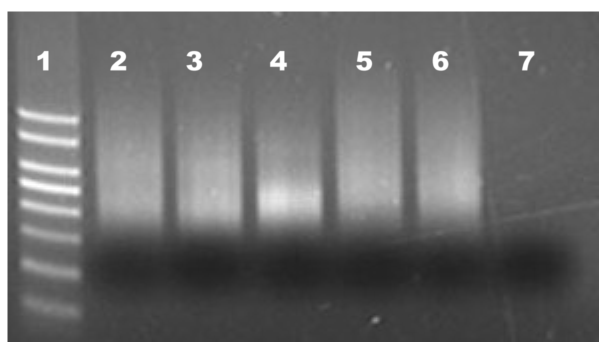

Figure 11. Linker mediated PCR amplification of CpG methylated DNA.

First lane contained the DNA size standard (100-1000 bps). The second and third lanes were loaded with PCR products resulted from MeCP2 bound template DNA at 1X and 2X concentrations, respectively. Fourth lane consists of MeCP2 bound viral DNA, which has been spiked with a control DNA sample known to result in a positive PCR reaction. The Fifth lane contains a PCR product obtained using the control methylated DNA provided with the kit. The seventh lane contains a PCR reaction from column purified unmethylated viral DNA (negative control). The gel was made with 1.2% agarose.

.

REFERENCES

Breitbart M, Haynes M, Kelley S, Angly F, Edwards RA, Felts B, Mahaffy JM, Mueller J, Nulton J, Rayhawk S et al. 2008. Viral diversity and dynamics in an infant gut. *Research in Microbiology* **159**(5): 367-373.

Burkhart JG, Burkhart BA, Sampson K, Malling HV. 1992. Evidence for a previously undetected CpG methyl-directed restriction system in E. coli. *Nucleic Acids Res* **20**(16): 4368.

Cross SH, Charlton JA, Nan X, Bird AP. 1994. Purification of CpG islands using a methylated DNA binding column. *Nat Genet* **6**(3): 236-244.

Goll MG, Bestor TH. 2005. Eukaryotic cytosine methyltransferases. *Annual Review of Biochemistry* **74**(1): 481-514.

Honess RW, Gompels UA, Barrell BG, Craxton M, Cameron KR, Staden R, Chang Y-N, Hayward GS. 1989. Deviations from expected frequencies of CpG dinucleotides in herpesvirus DNAs may be diagnostic of differences in the states of their latent genomes. *Journal of General Virology* **70**(4): 837-855.

Hu LF, Minarovits J, Cao SL, Contreras-Salazar B, Rymo L, Falk K, Klein G, Ernberg I. 1991. Variable expression of latent membrane protein in nasopharyngeal carcinoma can be related to methylation status of the Epstein-Barr virus BNLF-1 5'-flanking region. *Journal of Virology* **65**(3): 1558-1567.

Karlin S, Doerfler W, Cardon LR. 1994. Why is CpG suppressed in the genomes of virtually all small eukaryotic viruses but not in those of large eukaryotic viruses? *Journal of Virology* **68**(5): 2889-2897.

Kashlev M, Nudler E, Goldfarb A, White T, Kutter E. 1993. Bacteriophage T4 Alc protein: A transcription termination factor sensing local modification of DNA. *Cell* **75**(1): 147-154.

Kumar S, Cheng X, Klimasauskas S, Mi S, Posfai J, Roberts RJ, Wilson GG. 1994. The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res* **22**(1): 1-10.

Laird PW. 2010. Principles and challenges of genome-wide DNA methylation analysis. *Nat Rev Genet* **11**(3): 191-203.

Lee DW, Freitag M, Selker EU, Aramayo R. 2008. A cytosine methyltransferase homologue is essential for sexual development in *Aspergillus nidulans*. *Plos One* **3**(6): e2531.

Liu Y, Oakeley EJ, Sun L, Jost J-P. 1998. Multiple domains are involved in the targeting of the mouse DNA methyltransferase to the DNA replication foci. *Nucleic Acids Res* **26**(4): 1038-1045.

Machida M, Asai K, Sano M, Tanaka T, Kumagai T, Terai G, Kusumoto K-I, Arima T, Akita O, Kashiwagi Y et al. 2005. Genome sequencing and analysis of *Aspergillus oryzae*. *Nature* **438**(7071): 1157-1161.

Marinus MG, Morris NR. 1973. Isolation of deoxyribonucleic acid methylase mutants of *Escherichia coli* K-12. *Journal of Bacteriology* **114**(3): 1143-1150.

May MS, Hattman S. 1975. Analysis of bacteriophage deoxyribonucleic acid sequences methylated by host- and R-factor-controlled enzymes. *Journal of Bacteriology* **123**(2): 768-770.

Minarovits J, Minarovits-Kormuta S, Ehlin-Henriksson B, Falk K, Klein G, Ernberg I. 1991. Host cell phenotype-dependent methylation patterns of Epstein—Barr virus DNA. *Journal of General Virology* **72**(7): 1591-1599.

Pirotta V. 1976. Two restriction endonucleases from *Bacillus globiggi*. *Nucleic Acids Res* **3**(11): 3253.

Razin A, Riggs AD. 1980. DNA methylation and gene function. *Science* **210**(4470): 604-610.

Renbaum P, Abrahamove D, Fainsod A, Wilson GG, Rottem S, Razin A. 1990. Cloning, characterization, and expression in *Escherichia coli* of the gene coding for the CpG DNA methylase from *Spiroplasma* sp. strain MQ1(M Sssl). *Nucleic Acids Res* **18**(5): 1145-1152.

Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, Gordon JI. 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**(7304): 334-338.

Rohwer F, Prangishvili D, Lindell D. 2009. Roles of viruses in the environment. *Environmental Microbiology* **11**(11): 2771-2774.

Simon D, Grunert F, von Acken U, Doring HP, Kroger H. 1978. DNA-methylase from regenerating rat liver: purification and characterisation. *Nucleic Acids Res* **5**(6): 2153-2167.

Smith HO. 1979. Nucleotide sequence specificity of restriction endonucleases. *Science* **205**(4405): 455-462.

Vanyushin BF, Tkacheva SG, Belozersky AN. 1970. Rare bases in animal DNA. *Nature* **225**(5236): 948-949.

Youssoufian H, Hammer SM, Hirsch MS, Mulder C. 1982. Methylation of the viral genome in an in vitro model of herpes simplex virus latency. *Proceedings of the National Academy of Sciences* **79**(7): 2207-2210.

Yuan R, Meselson M. 1970. A specific complex between a restriction endonuclease and its DNA substrate. *Proceedings of the National Academy of Sciences* **65**(2): 357-362.

.

CHAPTER IV

METAGENOMIC ANALYSIS OF ANAEROBIC FUNGAL COMMUNITIES IN THE BOVINE RUMEN

INTRODUCTION

Rumen microbiome is a complex ecosystem consisting of bacteria, fungi, protozoa and viruses. Anaerobic fungi (AF) are an integral part of this ecosystem. Rumen AF number has been estimated to be in-between 10^3-10^5 per milliliter of rumen fluid based on the zoospore count (Mountfort and Orpin 1994). AF play a significant role in plant material digestion inside the rumen (Lowe et al. 1987; McSweeney et al. 1994; Orpin and Joblin 1997). However, very little information is available on the role and diversity of the fungal component in the rumen ecosystem. AF were first discovered in 1975 (Orpin 1975) and they were originally classified as Chytridiomycetes based on their thallus morphology and the presence of chitin in their cell walls (Orpin 1976; Orpin 1977). AF have been now classified under a new phylum called Neocallimastigomycota based on extensive molecular phylogenetic studies (Hibbett et al. 2007). It is believed that have evolved from free-living chytridiomycetes to survive the anaerobic conditions in the digestive tract of their herbivorous hosts (Heath 1988).

However, they have not been isolated from any other anaerobic, terrestrial or aquatic habitat so far. AF have been formally classified into six genera; *Neocallimastix*, *Piromyces, Caecomyces*, *Orpinomyces, Cyllamyces* and *Anaeromyces,* based on the morphology of their thalli, number of flagella present in zoospores and development. So far, over 20 species have been named and described under these six genera (Griffith et al. 2009). These species have been recognized according to the taxonomic convention by the ultrastructure of zoospores (Heath et al. 1983; Orpin and Munn 1986; Webb and Theodorou 1991). The validity of this classification has been debated due to occurrence of natural morphological variants within the same species of AF (Ho and Barr 1995). Age, method and quality of preparation for identification are some of the factors that could adversely affect the morphology based classification schemes.

The study of AF was hindered for many years since their discovery due to unavailability of convenient and reliable methods for accurate identification, differentiation and enumeration (Orpin 1994; Brookman et al. 2000). Over the past few decades, various methods have been used to identify and quantify rumen AF with varying degrees of success. Most Probable Number (MPN) technique, a commonly used enumeration technique in microbiological work (McCrady 1915), has been adapted to enumerate AF (Theodorou et al. 1990). MPN technique is based on the culture of AF and has the tendency to underestimate the true population. In order to overcome the limitations associated with culturing of fastidious AF, a chitin assay capable of quantifying rumen biomass has been introduced (Gordon and Phillips 1989; Gay 1991). However this assay is reported to have a limited usage in *in vivo* experiments (Rezaeian

et al. 2004) and lower extraction efficiencies with certain substrates (Davies 1991). Therefore, the greatest promise in identifying and quantifying AF remains with metagenomic techniques (Griffith et al. 2009).

The controversy of gut fungal classification using classical methods is apparent throughout the literature. *Neocallimastix frontalis* (Heath et al. 1983) is the first AF species that have been formally classified. The very first gut fungal isolate described by Orpin (1975) differed in various aspects to *N. frontalis* isolate described by Heath et al. and, as a result reclassified later as *N. patriciarum* (Orpin and Munn 1986). So far, three species of *Neocallimastix* have been described: *Neocallimastix frontalis* (Heath et al. 1983), *N. patriciarum* (Orpin and Munn 1986) and *N. hurleyensis* (Webb and Theodorou 1991). A reclassification of *N. frontalis* and *N. patriciarum* into a single species has been proposed by Ho and Barr in 1995. Some of the polycentric fungal species also have been renamed since they were first described. The genera *Sphaeromonas* and *Piromonas* were renamed as *Caecomyces* and *Piromyces,* respectively (Gold et al. 1988). *Ruminomyces elegans* was renamed as *Anaeromyces elegans* (Ho et al. 1990; Ho et al. 1993).

After the invention of PCR (Mullis et al. 1986) the field of molecular genetics developed rapidly. Dore and Stahl (1991) and later Bowman et al. (1992) have used 18S ribosomal RNA (rRNA) gene sequences to classify AF. These studies support the fact that AF evolved from free living Chytridiomycetes fungi. However, the inter-relationships of AF, at genera or species level, have not been discussed in the above studies, probably due to limited variability in the 18S rRNA genes. In another study, the internal transcribed spacer-1 (ITS-1) region flanked by the 18S and 5.8S rRNA gene

regions was used in a comparative sequence analysis (Li and Heath 1992). Unlike the 18S rRNA genes, ITS-1 region (and also ITS-2 region) is not well conserved among species and strains of AF due to less evolutionary pressure. This particular study based on ITS-1 region reports that the genera *Orpinomyces, Neocallimastix,* and *Piromyces* are closely related to each other and are distantly related to the genera *Anaeromyces* and *Caecomyces*. In a later study conducted by Li et al. (1993), made an attempt to determine the phylogenetic relationships of the AF using a cladistic analysis of 42 morphological, ultrastructural and mitotic characters. This method has improved the resolution, but the ambiguity of interrelationships among AF remains.

Brookman et al. (2000) have attempted to identify the phylogenetic relationships among *Neocallimastix* and *Piromyces* isolates, together with a smaller number of polycentric isolates using DNA sequence information. According to this study *Neocallimastix* has formed a single clade. However, *N.hurleyensis* and *N. frontalis* have clearly separated from *N. patriciarum* and the other *Neocallimastix* isolates in the clade. Therefore, this study reports two broad groups of *Neocallimastix*; *N. patriciarum*-like and *N. frontalis*-like. These findings suggest at least there are two *Neocallimastix* species as exemplified by *N. patriciarum* and *N. frontalis*. Brookman *et al.* further reports that the *Piromyces* and *Anaeromyces* sequences groups together and is separated from *Neocallimastix* consistently in all dendrograms.

Recently, a wide scale study on AF communities in thirty different herbivore species has been conducted using ITS-1 region as the phylogenetic marker (Liggenstoffer et al. 2010). This study was able to deposit a large number of sequences (267, 287) in the

public databases compared to what was present prior to this study (Only 236 AF sequences as of Oct 2009). Yet, this large number of sequence were not annotated at genera or species level, probably due to the selection of a poorly conserved ITS-I region.

Besides the ambiguity in classification of AF based on conventional methods, the laborious nature and requirement of a high level of expertise have compelled the search for better molecular-based techniques to classify AF. AF are also extremely resistant to culture. Therefore, culture independent molecular-based techniques such as metagenomics appear to be the state of the art in studying rumen fungal communities. Having said that, there are several challenges that needs to be overcome when studying AF using metagenomics. Rumen contains many sources of rRNA genes such as plant material, protozoa, sloughing off host animal cells etc. and it is quite challenging to design very specific primers for AF due to the high conservation of rRNA genes. Additionally, such primers should amplify a region with enough variability to differentiate between species of AF. In this study, we were able to study the AF diversity in bovine rumen using metagenomics overcoming most of the challenges described above.

MATERIALS AND METHODS

Rumen contents were collected from Angus steers fed on a high protein and hay diet. The samples were immediately frozen in liquid nitrogen and then at -80°C until DNA was extracted. Samples were ground in liquid nitrogen and, total DNA was extracted following a protocol that uses CTAB buffer. (2.5 g of CTAB + 10.33 g NaCl + 5 ml of 0.5 M EDTA at pH 8.0 + 12.5 ml 1 M Tris at pH 8.0 + water up to 125 ml). To 0.6 g of ground rumen contents 10 ml of CTAB buffer, containing 0.3 mg/ml proteinase K, was added and homogenized briefly. Then, the samples were incubated at 65°C for 30 min. Once the samples have reached room temperature, they were extracted with an equal volume (10ml) of chloroform and isoamyl alcohol (24:1). An equal volume of isopropanol was added to the supernatant saved in the previous step and DNA was pelleted by spinning at 8,000 g for 10 min at room temperature. The resulting DNA pellet was resuspended in 200 μl of distilled water and 5 μl of RNase (50mg/ml) was added. The samples were then incubated at room temperature for 20 min. Samples were then extracted with an equal amount of phenol, chloroform, and: isoamyl alcohol mixture (25:24:1) at pH 7.9. Afterwards, DNA was precipitated using ethanol to remove any impurities remaining that could inhibit subsequent PCR amplification. Finally, the precipitated DNA was resuspended in 100 μl of distilled water.

A novel set of degenerate PCR primers were designed, after aligning all 236 rRNA gene sequences that were available in public databases as of Oct 2009 (For – 5' AAAAGTTGGDCAAACTTGGTC 3' and Rev – 5' AGATCCRTTGTYAAAAGTTG 3'). This PCR primer set amplifies ITS-1 region with small segments of flanking 18S and

5.8S rRNA genes. The length of this amplicon is about 400 base pairs based on multiple sequence alignment used to design the primers. A 25 µl PCR contained about 100 ng of total DNA, 1X PCR buffer (Promega, Madison, WI, USA) 1.5 mM $MgCl_2$, 15 picomoles of each primer, 200 µM dNTPs and 1.5 U Taq DNA polymerase (Promega, Madison, WI, USA). The cycling conditions were: one cycle of 4 min at 95º C, 30 sec at 55º C and 40 sec at 72º C followed by 34 more cycles of 30 sec at 95º C, 30 sec at 55º C, 40sec at 72º C and a final extension of 2 min at 72º C. Next, approximately an equal amount of PCR product resulting from 4 separate samples were pooled together and gel purified. The resulting amplicons were in the range of 250-400 bps (Fig. 12). Later, the gel purified PCR product was sequenced using tag-encoded FLX amplicon pyrosequencing (bTEFAP) using a Roche 454 FLX Genome Sequencer system. Following sequencing, tags, primers, low quality sequence ends and chimeric sequences were removed (Bailey et al. 2010).Then, the sequences were annotated using a blastn search. Two putative sequence divergence cutoff values were used to classify sequences at species (4.80%) and genus (16.95%) levels as described by Liggenstoffer et al. (2010).

RESULTS

The protocol that has been developed to isolate rumen DNA and subsequent cleaning up steps were effective in isolating PCR-able intact DNA from complex rumen material. PCR amplification of total rumen DNA with AF specific primers resulted in an amplicon in the range of 250-400 bps in length (Fig. 13). The diversity of the length of the ITS-1 regions in AF was quite high.

Roche 454 pyrosequencing resulted in a total of 40,501 AF sequences. After quality trimming, the downstream analysis was conducted with 10,029 full length sequences having both forward and reverses primer sequences to avoid any potential incorrect annotations. Fig. 14 depicts the community structure and diversity of AF in the bovine rumen. Interestingly, none of the sequences belonged to species other than AF. *Orpinomyces* (41.6%) was the most abundant genus of AF found in the rumen of the animals sampled. In general, 23.7% of the sequences belonged to previously described AF species and 74.9% found to be coming from new Operational Taxonomic Units, based on the sequence divergence cutoff value (4.80%) used. About 1.5% of the sequences had a higher sequence divergence beyond 16.95% to the most similar sequence in the BLAST database. Therefore, they are believed to be resulting from new genera of AF, so far those have not been formally described.

DISCUSSION

Orpinomyces (41.6%) was found to be the most abundant genus of AF present in the bovine rumen of animals fed a high protein and hay-based diet. About 43% of the *Orpinomyces* sequences belonged to new species/OTUs that have not been formally described before, according to the sequence divergence cutoff values described by Liggenstoffer et al. (2010). This observation of a high relative abundance of *Orpinomyces* species in the rumen is contrary to the findings of Liggenstoffer et al. (2010) where the authors reported that the genus *Orpinomyces* related sequences were found to be low in abundance (with an average of about 3% of the AF community composition). The genus *Anaeromyces* was the second highest and accounted for about 30% of AF in the rumen. Interestingly 99% of the *Anaeromyces* sequences belonged to novel OTUs. A similar study conducted recently have found that the relative abundance of *Anaeromyces* species may occasionally go over 30% in pygmy hippopotamus, American bison, and in domestic goat all of which are foregut fermenters (Liggenstoffer et al. 2010). Only 0.1% of the sequences were classified as *Neocallimastix frontalis* and the rest unassigned *Neocallimastix* sequences accounted for 13.8% of the total sequences. The relative abundance of *Piromyces* was 10.2% and none of these sequences had a sufficient similarity to any of the sequence in the BLAST database to classify them at species level. The least abundant genera fond in the bovine rumen were *Caecomyces* (2.9%) and *Cyllamyces* (0.1%). *Cyllamyces* was the very last genus of AF discovered from the alimentary tracts of herbivorous animals, probably due to its extremely low abundance making it difficult to isolate from the rest of the rumen microbiome. *Cyllamyces* was the most infrequently encountered genus of AF according to Liggenstoffer et al. (2010) as

well. Remarkably, a vast majority of the sequences (74.9%) could not be classified at species level based on the sequence similarity to the existing sequences in public databases. About 1.5% (about 150 individual sequence reads) of the sequences belonged to genera that have not been previously described. This highlights the point that the conventional methods of identifying AF have not been able to explore the AF diversity in the rumen adequately. Further, there is high potential of metagenomics applications in investigating the AF diversity inside the rumen.

The PCR primers designed in this study did not amplify any known plant or other fungal sequence, not belonging to the phylum Neocallimastigomycota, which has been a problem with many previously published primers.

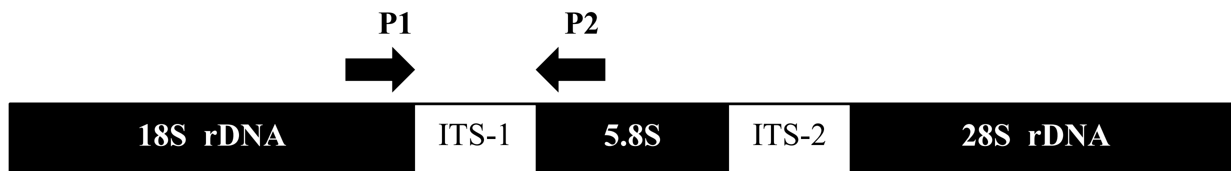

Figure 12. Anaerobic fungi specific degenerate PCR primers.

The arrow P1 corresponds to the forward primer 5’ AAAAGTTGGDCAAACTTGGTC 3’ and the arrow P2 corresponds to the reverse primer 5’ AGATCCRTTGTYAAAAGTTG 3’. These PCR primers amplify ITS-1 region with small segments of flanking 18S and 5.8S ribosomal RNA genes. The estimated length of the PCR amplicon resulting with this primer set is about 400 base pairs.

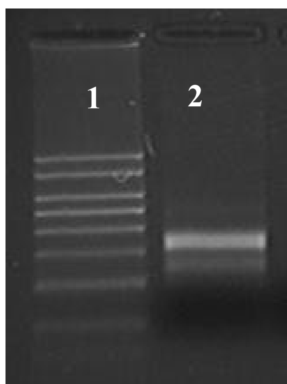

Figure 13. Genomic diversity of the ITS-1 region of anaerobic fungi.

Lane 1 of the 1.2% agarose gel electropherogram shows the DNA marker with the fragment sizes, from top to bottom, 1000, 800, 600, 500, 400, 300, 200, and 100 base pairs in length. Lane 2 shows the PCR amplicons resulted from rumen anaerobic fungi ITS-1 region ranging from 250-400 bps in length.

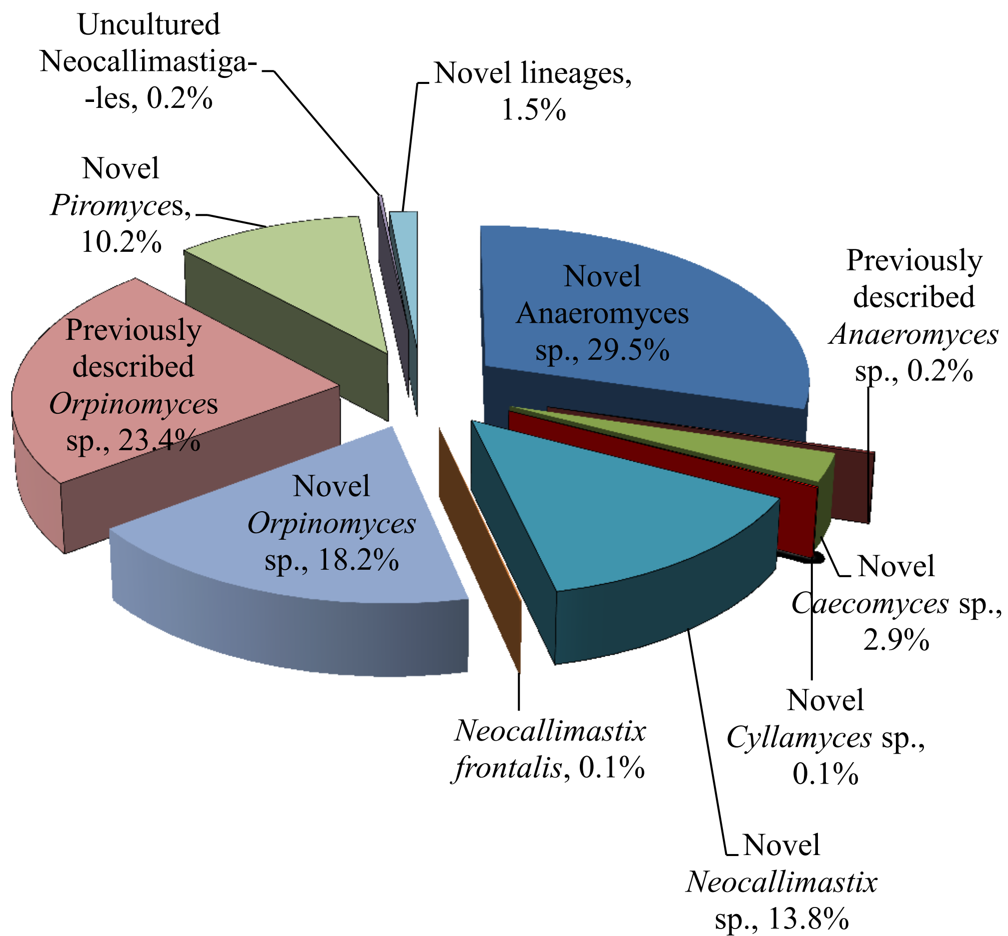

Figure 14. Community structure and diversity of anaerobic fungi present in the rumen. The sequences were classified into species and genera based on the two putative sequence divergence cutoff values 4.80% and 16.95%, respectively. Novel lineages represent the sequences that had a sequence divergence over 16.95% to the closest sequence present in the BLAST database.

REFERENCES

Bailey MT, Dowd SE, Parry NMA, Galley JD, Schauer DB, Lyte M. 2010. Stressor exposure disrupts commensal microbial populations in the intestines and leads to increased colonization by *Citrobacter rodentium*. *Infection and Immunity* **78**(4): 1509-1519.

Bowman BH, Taylor JW, Brownlee AG, Lee J, Lu SD, White TJ. 1992. Molecular evolution of the fungi: relationship of the Basidiomycetes, Ascomycetes, and Chytridiomycetes. *Mol Biol Evol* **9**(2): 285-296.

Brookman JL, Mennim G, Trinci APJ, Theodorou MK, Tuckwell DS. 2000. Identification and characterization of anaerobic gut fungi using molecular methodologies based on ribosomal ITS1 and 18S rRNA. *Microbiology-Uk* **146**: 393-403.

Davies DR. 1991. *Growth and survival of anaerobic fungi in batch culture and in the digestive tract of ruminants*. University of Manchester, Manchester.

Dore J, Stahl DA. 1991. Phylogeny of anaerobic rumen Chytridiomycetes inferred from small subunit ribosomal RNA sequence comparisons. *Canadian Journal of Botany* **69**(9): 1964–1971.

Gay L. 1991. Chitin content and chitin synthase activity as indicators of the growth of three different anaerobic rumen fungi. *FEMS Microbiology Letters* **80**(1): 99-102.

Gold JJ, Brent Heath I, Bauchop T. 1988. Ultrastructural description of a new chytrid genus of caecum anaerobe, *Caecomyces equi* gen. nov., sp. nov., assigned to the Neocallimasticaceae. *Biosystems* **21**(3–4): 403-415.

Gordon GL, Phillips MW. 1989. Degradation and utilization of cellulose and straw by three different anaerobic fungi from the ovine rumen. *Applied and Environmental Microbiology* **55**(7): 1703-1710.

Griffith GW, Ozkose E, Theodorou MK, Davies DR. 2009. Diversity of anaerobic fungal populations in cattle revealed by selective enrichment culture using different carbon sources. *Fungal Ecol* **2**(2): 87-97.

Heath B, Bauchop T, Skipp RA. 1983. Assignment of the rumen anaerobe *Neocallimastix frontalis* to the Spizellomycetales (Chytridiomycetes) on the basis of its polyflagellate zoospore ultrastructure. *Canadian Journal of Botany* **61**(1): 295–307.

Heath IB. 1988. Gut fungi. *Trends in Ecology and Evolution* **3**(7): 167-171.

Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF, Eriksson OE, Huhndorf S, James T, Kirk PM, Lücking R et al. 2007. A higher-level phylogenetic classification of the Fungi. *Mycological Research* **111**(5): 509-547.

Ho YW, Barr DJS. 1995. Classification of anaerobic gut fungi from herbivores with emphasis on rumen fungi from malaysia. *Mycologia* **87**(5): 655-677.

Ho YW, Barr DJS, Abdullah N, Jalaludin S. 1993. *Anaeromyces*, an earlier name for *Ruminomyces*. *Mycotaxon* **47**: 283-284.

Ho YW, Bauchop T, N A, S.Jalaludin. 1990. *Ruminomyces elegans* gen. et sp. nov., a polycentric anaerobic rumen fungus from cattle. *Mycotaxon* **38**: 397-405.

Jinliang LI, Heath I B, Packer L. 1993. The phylogenetic relationships of the anaerobic chytridiomycetous gut fungi (Neocallimasticaceae) and the Chytridiomycota. II: Cladistic analysis of structural data and description of Neocallimasticales ord.nov. *Canadian Journal of Botany* **71**(3): 393-407.

Li J, Heath IB. 1992. The phylogenetic relationships of the anaerobic chytridiomycetous gut fungi (Neocallimasticaceae) and the Chytridiomycota. I. Cladistic analysis of rRNA sequences. *Canadian Journal of Botany* **70**(9): 1738-1746.

Liggenstoffer AS, Youssef NH, Couger MB, Elshahed MS. 2010. Phylogenetic diversity and community structure of anaerobic gut fungi (phylum Neocallimastigomycota) in ruminant and non-ruminant herbivores. *ISME J*.

Lowe SE, Theodorou MK, Trinci AP. 1987. Cellulases and xylanase of an anaerobic rumen fungus grown on wheat straw, wheat straw holocellulose, cellulose, and xylan. *Appl Environ Microbiol* **53**(6): 1216-1223.

McCrady MH. 1915. The numerical interpretation of fermentation-tube results. *The Journal of Infectious Diseases* **17**(1): 183-212.

McSweeney CS, Dulieu A, Katayama Y, Lowry JB. 1994. Solubilization of lignin by the ruminal anaerobic fungus *Neocallimastix patriciarum*. *Applied and Environmental Microbiology* **60**(8): 2985-2989.

Mountfort DO, Orpin CG. 1994. *Anaerobic Fungi: Biology, Ecology and Function*. Marcel Dekker, Inc., New York.

Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. 1986. Specific enzymatic amlification of DNA in vitro – The polymerase chain reaction. *Cold Spring Harbor Symp Quant Biol* **51**: 263-273.

Orpin CG. 1975. Studies on the rumen flagellate *Neocallimastix frontalis*. *J Gen Microbiol* **91**(2): 249-262.

Orpin CG. 1976. Studies on the rumen flagellate *Sphaeromonas communis*. *J Gen Microbiol* **94**(2): 270-280.

Orpin CG. 1977. Invasion of plant tissue in the rumen by the flagellate *Neocallimastix frontalis*. *J Gen Microbiol* **98**(2): 423-430.

Orpin CG. 1994. Anaerobic fungi: Taxonomy, biology, and distribution in nature. In *Anaerobic Fungi: Biology, Ecology and Function*, Vol 12 (ed. DO Mountfort, CG Orpin). Marcel Dekker, Inc., New York.

Orpin CG, Joblin KN. 1997. The rumen anaerobic fungi. In *The Rumen Microbial Ecosystem* (ed. PN Hobson, CS Stewart). Chapman & Hall, New York.

Orpin CG, Munn EA. 1986. *Neocallimastix patriciarum* sp.nov., a new member of the Neocallimasticaceae inhabiting the rumen of sheep. *Transactions of the British Mycological Society* **86**(1): 178-181.

Rezaeian M, Beakes GW, Parker DS. 2004. Distribution and estimation of anaerobic zoosporic fungi along the digestive tracts of sheep. *Mycological Research* **108**(10): 1227-1233.

Theodorou MK, Gill M, King-Spooner C, Beever DE. 1990. Enumeration of anaerobic Chytridiomycetes as thallus-forming units: novel method for quantification of fibrolytic fungal populations from the digestive tract ecosystem. *Applied and Environmental Microbiology* **56**(4): 1073-1078.

Webb J, Theodorou MK. 1991. *Neocallimastix hurleyensis* sp.nov., an anaerobic fungus from the ovine rumen *Canadian Journal of Botany* **69**(6): 1220–1224.

VITA

Biyagamage Ruchika Fernando

Candidate for the Degree of

Doctor of Philosophy

Dissertation: METAGENOMIC ANALYSIS OF MICROBIAL COMMUNITIES IN THE BOVINE RUMEN

Major Field: Animal Breeding and Reproduction

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Animal Breeding and Reproduction at Oklahoma State University, Stillwater, Oklahoma in July, 2012.

Completed the requirements for the Bachelor of Veterinary Medicine and Animal Science at the University of Peradeniya, Peradeniya, Sri Lanka in 2003.

Experience:

I worked as a veterinarian and a lecturer attached to the Department of Veterinary Public Health and Pharmacology at the University of Peradeniya, Sri Lanka for four years before I came to the US to pursue my doctoral studies. During my doctoral studies, I have gained a sound knowledge of classical, population, and molecular genetics with hands-on experience using state-of-the-art molecular genetic techniques. I have nine years of teaching experience at the college level both in the US and internationally. I was involved in teaching animal genetics, veterinary pharmacology, and veterinary public health to undergraduates. Also, I have mentored numerous undergraduates who conducted research projects at the Animal Genomics Laboratory of Oklahoma State University and the Veterinary Public Health Laboratory of the University of Peradeniya. Furthermore, I worked on international collaborations in both Denmark and Japan related to veterinary practices.

Professional Memberships:

American Association for the Advancement of Science
Sri Lanka Veterinary Association
Sri Lanka Veterinary Council

Name: Biyagamage Ruchika Fernando Date of Degree: December 2012

Institution: Oklahoma State University Location: Stillwater, Oklahoma

METAGENOMIC ANALYSIS OF MICROBIAL COMMUNITIES
IN THE BOVINE RUMEN

Pages in Study: 117 Candidate for the Degree of Doctor of Philosophy

Major Field: Animal Breeding and Reproduction

Scope and Method of Study:

The mammalian gut ecosystem consists of bacteria, fungi, protozoa and viruses. The role of gut microbes in the performance and well-being of the host animal is well recognized. This study has analyzed the diversity and genomic complexity of the double stranded DNA virome of the bovine rumen. A parallel study was conducted to investigate the diversity of anaerobic fungi within the bovine rumen. Viral particles were separated from other rumen microorganisms and their DNA was amplified using a sequence-independent protocol and sequence analyzed using Roche 454 pyrosequencing technology. A PCR-based assay was developed using conserved genomic regions specific to anaerobic fungal communities and amplified PCR products were sequence analyzed using next-generation sequencing technology.

Findings and Conclusions:

Only 26.6% of the putative viral sequences could be annotated using the information present in the public databases. Out of this about 14% of the sequences had homology to previously described viruses and a vast majority of the rest were classified as bacterial sequences. These bacterial sequences probably represent proviral sequences found in already sequenced bacterial genomes or some of the lytic phage particles isolated in this study that carry genetic information acquired from their hosts. Interestingly, the distribution of bacterial hits was similar to the relative abundances of these bacteria described in the rumen environment. About 85% of the viral hits corresponded to tailed bacteriophages belonging to Siphoviridae (69.8%), Myoviridae (13.8%), and Podoviridae (1.5%). The gene ontology studies of the putative viral sequences revealed that these viruses carry a diverse array of genes related to a range of virus specific and non-virus specific functions. Further, the findings of this study suggest the presence of CpG methylation in the rumen virome which needs to be further investigated.

A high level diversity was observed among anaerobic fungi of the bovine rumen. Only 23.7% of the sequences could be annotated to previously described anaerobic fungi and 74.9% were found to be from new species-level operational taxonomic units. Further, we found evidences for the presence of novel genera (~1.5%) of anaerobic fungi in the rumen. In animals fed with a high protein, hay-based diet, the genus *Orpinomyces* (41.6%) was the most abundant followed by *Anaeromyces* (29.7%), *Neocallimastix* (13.9%), *Piromyces* (10.2%), *Caecomyces* (2.9%) and *Cyllamyces* (0.1%). The PCR primers designed in this study did not amplify any plant or other fungal DNA which has been a problem with many previously published primers.

ADVISER'S APPROVAL: Dr. Udaya DeSilva