UNIVERSITY OF CALGARY

Myoviridae as Part of Cattle Fecal Microbiome

by

Mónica Rincón García

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN VETERINARY MEDICAL SCIENCES

CALGARY, ALBERTA

MARCH, 2019

Abstract

Double-stranded deoxyribonucleic acid (dsDNA) phages from the order of *Caudovirales* are the most abundant viruses in the microbiome of the mammalian digestive tract. There is increasing awareness of the role of phages in modulating bacterial composition in digestive compartments. Mammalian digestive microbiomes are generally dominated by members of the *Bacteroidetes* and *Firmicutes* phyla. Moreover, phages from the *Siphoviridae* family (order *Caudovirales)* are the most abundant members in the rumen of cattle and large intestine of humans, horses and pigs. However, there is a knowledge gap regarding composition of the virome in the large intestine of cattle. To facilitate these studies, suitable methods to analyze bovine fecal sample for virus content are needed. Therefore, the objective of this thesis was to establish appropriate methodologies and evaluate fecal samples from various sources. Two virus particle purification (VPP) procedures, namely filtration and cesium chloride (CsCl) ultracentrifugation were compared, using electron microscopy (EM), polymerase chain reaction (PCR) and metagenomic analyses. In total, 39 cattle fecal samples were processed. Metagenomic analysis was the most suitable methodology for measuring relative abundance (RA) and diversity of *Myoviridae* in fecal samples from two beef and two dairy cattle. Overall, numbers of phage members of the order *Caudovirales* was higher compared to other dsDNA viruses, with *Myoviridae* being the most abundant family within this order. Furthermore, a functional procedure to estimate total viral dsDNA virome in feces was developed. In conclusion, this thesis includes methods to detect and characterize phages in cattle feces, with generation of novel data that improve understanding of viral diversity in fecal microbiomes.

Acknowledgements

First of all and foremost, thanks to Jesus Christ, to my husband Camilo and to our daughters Naomi, Zuri and Gabriella for giving me direction, freedom and propose in life. I express my unconditional gratitude for the opportunity that Dr. Frank van der Meer, Dr. Karin Orsel and Dr. Edouard Timsit gave me. It has been a learning experience through all these years and a privilege to work in the research laboratories at the University of Calgary. I am also very grateful to Dr. Claudia Bachofen, Dr. Matthew Workentine, Dr. John Kastelic and Dr. Jacob Thundathil for their support, guidance and expertise in their fields. Special thanks to my lab teams, veterinarians, dairy farmers and feedlot operators who provided samples for this project. I acknowledge and appreciate the funding provided throughout the program.

Dedication

To the Immaculate Heart of Mary

Table of Contents

List of Tables

List of Figures

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
dsDNA	Double stranded deoxyribonucleic acid
VPP	Virus particle purification
CsCl	Cesium chloride
MAP	*Mycobacterium avium* subsp. *paratuberculosis*
EM	Electron microscopy
GI	Gastrointestinal
PCR	Polymerase chain reaction
IBD	Inflammatory bowel disease
Gp20	Gene 20 from the Major capsid protein
ICTV	International Committee on Taxonomy of Viruses
NGS	Next Generation Sequencing
ssRNA	Single stranded ribonucleic acid
RNA	Ribonucleic acid
BLAST	Basic Local Alignment Search Tool
VFA	Volatile fatty acid
CH_4	Methane
H_2	Hydrogen
NA	Nucleic acid
CO_2	Carbon dioxide
BWT	Burrows-Wheeler transform
OTU	Operational taxonomic units
FM	Ferragina-Manzini index
LB	Lysogenic broth
F-CS_3	Forward primer CPS_3
R-CS_8	Reverse primer CPS_8
HT	Heat-treated
E. coli	*Escherichia coli*
STEC	Shiga toxin-producing *Escherichia coli*

Epigraph

"O Lord, you make out peace to us, for it is you who have accomplished all we have done"
Is 26:12

Chapter 1: Introduction

The microbiome is the total collection or community of microorganisms that inhabit an ecosystem or body site. Specific relationships exist among and between microorganisms and their host and/or ecosystem. Typically, a microbiome includes bacteria, viruses, fungi, protozoa and archaea (Figure 1.1) (1) . Microbiomes, an integral part of every body system, have critical roles in body functions (2, 3), including digestion, metabolism and immunity (4). Constituents of a microbiome vary among individuals (5), locations and body systems (6). An organism that is widespread in one microbiome can be rare or absent in another that is in close proximity; for example, although human teeth and tongue are both in the oral cavity, they have distinctly different microbiota (7). Many factors affect microbiome composition, including age (8), diet (9, 10), environment, host genetics, (5), gender (11), antibiotic treatments (8) and disease state (12).

A microbiome has a primary characteristic called diversity, a way of assessing species richness and evenness (10, 13). Richness refers to the number of species within a sample derived from a community of microorganisms, whereas evenness or relative abundance refers to the relative proportion of each species within that community (10, 13). Diversity indices can be primary (only number of species or frequency) or sub-indexes (or composite) such as Shannon, Simpson and Nei indexes (14). The Shannon index alpha diversity (*H'*), the most commonly used index, combines richness and evenness as a single characteristic of a microbiome (13, 14). When two or more microbiomes are compared, the estimated index is called beta diversity (ß), the difference between two or more samples (13).

Although several microbiomes in cattle have been described for various body systems and locations, including the respiratory tract (15-17), rumen (18), vagina (19) and feces (9, 12), this introduction only discusses the microbiome in the gastrointestinal system.

1.1 Digestive microbiomes in cattle

Microbiomes in the bovine digestive tract and the host have a symbiotic relationship. Consequently, both have critical roles in animal performance. Understanding the metabolic capacity of rumen microbiota, characterizing fecal microbiota and host phenotype should help

to develop feeding strategies to increase efficiency of body weight gain and may decrease digestive pathologies (20). Recent evidence suggests that the upper and lower gut portions of the digestive system act upon or communicate with each other by means of intestinal factors such as hormones (21, 22) produced by enteroendocrine cells of intestinal villi (22). Factors like these may be affected by microbiota, diets and feeding strategies (21). Conversely, specific host-microbiome mechanisms (i.e., Pattern Recognition Receptors) of intestinal communication have been analyzed in humans (3), but not yet in ruminants. Each of the principal microbial components of the rumen microbiome has been thoroughly characterized. However, characterization of the bovine intestinal microbiome has only focused on the bacterial component (9, 12). However, a comprehensive assessment of viruses in the bovine GI tract has not been done.

1.1.1 Rumen microbiome

The rumen microbiome metabolizes feedstuffs into volatile fatty acids (VFA), microbial biomass, vitamins and various other compounds, including formic acid, H_2, CO_2 and methane (CH_4) (23, 24). Anaerobic microorganisms constitute the microbiome of the rumen (24). The bacterial component accounts for 50 to 75% of the rumen microbial population (24). The bacterial component of the rumen's microbiome in dairy cows (mid lactation) is dominated by *Bacteroidetes* (71.36%), complemented with *Firmicutes* (21.16%) and other bacteria (7.48%) (25). The dominant archaeal compartment of the rumen microbiome is composed of methanogens which account for 1% of the total microbial population (26). Predominant methanogens, belonging to the order *Methanobacteriales,* utilize both H_2 and CO_2 to produce CH_4, energy (ATP) and other compounds, including methanol, methylamines and formate from protozoa, bacteria and fungi (26). The dominant core mycobiome of the rumen, composed of anaerobic fungi, accounts for 20% of the rumen microbial mass and is responsible for degradation of lignocellulose (24). The rumen also harbors a complex protozoal community, where ciliates are the most abundant ruminal protozoa (27). The majority of viruses in the rumen are phages infecting prokaryotes, whereby bacteriophages infect bacteria and archaeal viruses infect archaea (28, 29).

1.1.2 The large intestinal microbiome

The bovine ruminal and large intestinal microbiomes have distinct microbial communities (Figure 1.2) (9, 30, 31); in particular, there are differences in proportions of the two main phyla (*Firmicutes* and *Bacteroidetes*). The fecal microbiome is dominated by *Firmicutes* (55.2%), complemented with *Bacteroidetes* (25.4%) and other bacteria (19.4%) (9). These proportional shifts imply that different functional groups of bacteria dominate each compartment and can be affected by diet (10). The human intestinal microbiome has been studied using fecal samples (32, 33), they can be easily obtained and represent the microbial composition of the terminal part of the digestive tract. Furthermore, there was greater variation in microbiota amongst humans than there was variation between stool and large intestinal mucosa microbiota within an individual (32); therefore, the stool sample microbiome was relatively representative of microbiota in the large intestine, at least in humans.

Of specific interest in ruminants are so called methanogens (*Methanobrevibacter* spp.); they are also part of the large intestinal microbiome, but with a much lower density than in the rumen (31).

1.2 Virome as part of the cattle digestive microbiome

A virome is a community of viruses that infect bacteria (bacteriophages), archaea (archaeal viruses), other microorganisms that inhabit the microbiome, (e.g., protists) and host cells (viruses of eukaryotes or eukaryotic viruses) (Figure 1.3) (34, 35). Regarding the latter, viruses of eukaryotes can cause acute, transient, persistent or latent infections in mammals (34). Based on studies in humans, virome composition is affected by several factors, including age, diet and presence of other microbiome components (36, 37).

Bacteriophages are the main component of the rumen virome (28, 38) as well as the large intestine virome of monogastric species (36, 37). The vital aspect of bacteriophages is their interaction with bacteria of the microbiome (28, 37). However, in ruminants, the virome of the digestive tract distal to the rumen, including the rectum, has not been described.

Based on studies of the human gut, roles of bacteriophages in the digestive system include: a) phage-induced bacterial lysis (promoting nutrient recycling); b) maintenance of bacterial population diversity; and c) phage-mediated transfer of genes for antimicrobial resistance or toxin production (virulence) in bacteria (39, 40) . Therefore, the gut virome can have effects that are both beneficial and harmful to the host. Characterizing phages and other eukaryotic viruses in the bovine digestive system could be of interest to the cattle industry. For example, archaeal phages could reduce methane emissions by lysing methanogens (41, 42), or bacteriophages could be used to manipulate bacterial populations, thereby increasing feed efficiency (similar to ionophores) (43-45).

1.2.1 Rumen virome

The rumen virome includes phages infecting prokaryotes plus eukaryotic viruses infecting other microorganisms, e.g. protozoa, in addition to host cells. Phages mainly belong to the order *Caudovirales*: *Siphoviridae* (36 ± 3% of total viral families), *Myoviridae* (28 ± 4%) and *Podoviridae* (14 ± 2%). The remaining fraction of the rumen virome (13 ± 9%) consists of other phages and eukaryotic viruses (28) .

The first report characterizing phages in rumen used electron microscopy (28, 46). Rumen fluid has a dense and morphologically diverse population of viruses, with up to 40 distinct morphotypes of phages detected, mainly tailed phages of the *Caudovirales* order (Figure 1.4) (28). *Siphoviridae* are small phages with long contractible tails, very similar to one other, except for minor morphological details (Figures 1.4 - 1.6). *Myoviridae* include contractile-tailed phages (Figures 1.4 – 1.6) and *Podoviridae,* which are very small, short-tailed phages (Figure 1.4-1.6). Phages not belonging to the *Caudovirales* are filamentous phages belonging to the family *Inoviridae*, polyhedral phages such as *Microviridae*, *Tectiviridae*, *Leviviridae, Cystoviridae* and, *Corticoviridae* and pleomorphic phages such as *Plasmaviridae* (47). In addition, some archaeal phages in the rumen include *Methanobacterium* ψ M1, *Methanobacterium* phage ψ 10, *Methanothermobacter* ψ M100 and *Methanobacterium* phage ψ M2 (24). Their importance is related to control of methanogen populations in the rumen (e.g., *Methanobrevibacter ruminicatum* M1 phage ϕ -mru (48).

1.2.2 Large intestinal virome

Diversity of viruses in the bovine large intestine has not been characterized. The human enteric virome is the most extensively studied (3); based on those findings, it would be interesting to determine if there are similar links between health status of the host (12) and virome composition in cattle, as described for IBD in humans (49).

Bacteriophages, belonging to the order *Caudovirales* are the main component of the virome in the large intestine of humans (36). In a metagenomic analysis of human feces, approximately 2.7% of total mapped sequences corresponded to viral sequences, although only 15% of those sequences could be classified to the *Caudovirales* or *Microviridae* taxa (3, 50). In horse fecal samples, metagenomic analysis identified predominantly bacteriophages; although a wide variety of phage species were detected, even the most abundant phages did not represent more than 5-10% of the total population (39). Eukaryotic viruses, including ssRNA, ssDNA, retroviruses and dsDNA viruses, considered pathogens or opportunistic pathogens, are resident in the human intestine and constantly shed in fecal samples of humans that are healthy and asymptomatic (3). During early life, the human fecal virome contains only a small fraction of eukaryotic viruses, whereby at least 16 DNA viral families and 10 RNA families were identified in fecal samples from adults (36, 51).

In healthy individuals a group of bacteriophages, collectively called the core phageome, is associated with a healthy microbiome (52). Imbalances in bacterial composition, for example during dysbacteriosis (humans with inflammatory bowel disease or metabolic diseases) are associated with greater abundance of bacteriophages of the order *Caudovirales* (53, 54). Perhaps this increased abundance of *Caudovirales* is similar in cattle with decreased bacterial diversity, e.g. those with Johne's disease (12).

In conclusion, the influence of bacteriophages on bacterial populations in fecal microbiomes of cattle is unknown. However, there are two possible outcomes: increasing the number of bacteriophages results in a lower diversity of bacteria or an increase in bacterial diversity will promote a change in bacteriophage composition (3, 36, 46).

Similar to the human fecal virome, there is a high probability that there is a 'core' group of phages in cattle feces. Furthermore, it can be expected that fecal phage diversity will be affected by ecological or environmental factors, e.g. diet, age and host genetics, as reported in humans (36, 52, 53).

1.2.3 Bacteriophages as part of the microbiome

Bacteriophages were discovered by Edward Twort (1915) and Felix d´Herelle (1917), who identified a filterable agent able to kill bacteria (39). The International Committee on Taxonomy of Viruses (ICTV) classifies bacteriophages according to morphology and type of nucleic acids (55). Morphologically, there are three categories of bacteriophages: icosahedral, filamentous and complex. Based on nucleic acid, there are four categories: single or double-stranded, DNA or RNA bacteriophages (40).

Bacteriophages can be either lytic (often virulent) or lysogenic (temperate) phages (55). The lytic stage can be induced by exposing bacteria that are infected with a lysogenic phage to ultraviolet light or certain chemicals (56). A lytic phage infects a bacterial cell and causes lysis, with release of ~50 to 1000 viral particles that can infect other bacteria (40). However, lytic phages do not always cause lysis, but rather they may form quasi-stable, pseudolysogenic associations with their hosts (39). A lysogenic bacteriophage (prophage) does not immediately initiate a lytic cycle (40). For these, their DNA remains integrated into a bacterial cell chromosome for numerous generations or exists as a plasmid (39). The prophage may contain genes that confer virulence (toxin production) or antimicrobial resistance, thereby altering bacterial phenotype (such as *V. cholerae* phage) (40).

Most studies of bacteriophages in the large intestines of feedlot cattle were designed to identify and characterize phages that could potentially target Shiga toxin-producing *Escherichia coli* (STEC) O157:H7 (57, 58), with a long-term goal of mitigating these bacteria in feedlot cattle (59), thereby reducing the impact of these organisms in the food chain. T1-like bacteriophages (lytic phages) belonging to the *Siphoviridae* family, have been isolated from feedlot cattle in southern Alberta. These phages lysed *Escherichia coli* O157:H7, but not non-O157 *E. coli* (59).

1.2.4 Myoviridae as part of the microbiome

The *Myoviridae* family consists of tailed bacteriophages with double-strand DNA genomes belonging to the order *Caudovirales,* the most abundant viruses on earth (60). Based on tail morphology, Myoviruses are classified by the ICTV (2017) into six subfamilies *Peduoviridae*, *Spounavirinae, Eucampyvirinae, Vequintavirinae*, *Tevenviridae* and *Ounavirinae*. Also, there are 42 other genera within the *Myoviridae* family recognized by ICTV, but not assigned to subfamilies (61, 62).

Contractile tails (80-455 x 16-20 nm) are distinguishing attributes of the family *Myoviridae* (Figure 1.4-1.6). These tails are composed of a pivotal core of stacked rings in six subunits, encircled by a helical contractile sheath. The neck partitions the sheath from the head. When infections occur, contraction of sheath subunits allows the tail core to link to the bacterial plasma membrane. Compared to other tailed phages, Myovirus heads are bigger, heavier and have more DNA. Genome organization and mechanisms of DNA replication and packaging characterize the genera. The presence or absence of unusual bases and DNA polymerases can be used to differentiate genera within this family (61, 63).

T4-like viruses infect (as lytic phages) various bacteria, including members of the family *Enterobacteriaceae* in the human gut and the family *Cyanobacteriaceae* in water. Specifically, phage *Cyanophage* S-PM2 and coliphage T4 have homologous genes (64), suggesting that these *Myoviridae* have analogous roles as modulators of bacteria populations in various environments (39). Presence and persistence of *Myoviridae* in a variety of aquatic environments is well characterized (65). Relative abundance of *Myoviridae* aquatic environments in wastewater (~26%), freshwater (30%) and marine water (41%) of all phages detected in these environments belonged to this family (66).

Based on findings described above, including strong similarities among *Myoviridae* derived from various ecosystems, we can speculate that *Myoviridae* modulate composition of bacterial flora in the bovine intestinal tract. Due to their interactions with bacteria, the population of *Myoviridae* in cattle feces could reflect microbiome composition at the end of the digestive process.

Potential to use *Myoviridae* to influence the microbiome in the gastro-intestinal tract in humans has been studied. Phages from the order *Caudovirales* passed through the human upper digestive tract with a high rate of survival (>90%) in a dairy food matrix (67). Consequently, it appears that *Myoviridae* will not be limited to the upper gastrointestinal tract, but they may also be useful in the large intestines. Encapsulated bacteriophages (e.g. *Lactococcus lactis* phage P008) were reported to alter the human gut microbiome (67, 68).

Characterizing *Myoviridae* in bovine fecal samples could be important for several reasons. For example, they could be used to: a) induce changes in bovine microbiomes over the lifespan of the animal (perhaps specific phages could be used as an alternative to ionophores to enhance feed efficiency); or b) reduce or eliminate the population of *Proteobacteria* in the fecal microbiome, potentially including *E.coli* O157.

1.3 Factors affecting the microbiome

1.3.1 Age

In humans, microbiome establishment start in the fetus (4). There are apparently no reports of similar studies in cattle. However, in 1 to 3 day old calves, colonization of the gastrointestinal tract takes to cccur, with organisms derived from maternal vaginal, fecal, skin and saliva microbiomes (21). Some genera identified in the first days of life in calves are still present in mature cattle (69).

In the bovine fecal microbiome, there are profound changes in number of *Proteobacteria* from birth to weaning. Newborn dairy calves have a relative abundance of 52.9% of Proteobacteria at day 1, which decreases to <20% by day 7 (70) to a low of ~4.1% pre-weaning and 5.4% post weaning (71). It is noteworthy that the latter was similar to the relative abundance of *Proteobacteria* in bovine (2.5 and 9.6%) (9, 72) and feedlot cattle ~1 year old (12.36%) (30).

Under typical management, beef calves remain their dams for approximately 30 weeks, whereas newborn dairy calves are often immediately separated from their dam or within a few hours after birth. It was speculated that contact between a dam and her calf is necessary to establish a balanced microbiome (21). This is based on indications that maternal contact and

exchange of gut microbiomes promote health and development in neonatal humans (73). Whether this is also applicable to cattle remains to be established.

Most studies regarding the viral component of the microbiome and its variations were performed in humans. The human large intestine harbors at least 16 DNA virus and 10 RNA eukaryotic virus families (36). Their diversity is low during the neonatal period, but increases with age (36). Conversely, gut phage communities in newborn children had high richness and diversity, but low stability during the first days of life, with a gradual decrease in diversity until 2 years of age, whereas bacterial diversity of the microbiome increased during the same interval (74). Specifically, there is a shift in phage communities from *Caudovirales* to *Microviridae* in the first 2 years of life in humans (51).

1.3.2 Diet

Diet is the most studied factor affecting abundance and diversity in microbiomes of the rumen and large intestine (18). Dietary changes in 10-month old beef steers are usually accompanied by a shift in the rumen microbiome that enhances digestion (75). After weaning, most beef cattle eat hay or other forms of forage and varying amounts of grain. Rumen microbiome diversity was similar to the microbiome of the large intestine when cattle were fed high-forage diets, but this changed when high-grain diets were fed (10). Cattle fed a high-forage diet and then switched to a high-grain diet increased production of VFAs (76), which decreased rumen pH and created a new environment for microbiota. Cattle changed from a forage-based diet to a diet of unprocessed grain had *2.4 times* higher relative abundance of *Bacteriodetes* OTUs, attributed largely to a >10-fold increase in abundance of *Prevotellaceae* OTUs (9), a family within the order Bacteriodetes (77). Therefore, in addition to diet composition, perhaps feed processing can also affect abundance of specific bacterial OTUs.

Another example of dietary influence of microbiome development is giving newborn dairy calves heat-treated versus fresh colostrum. Feeding heat-treated colostrum facilitated colonization of the young calf gut by *Bacteroidetes (Bifidobacterium spp.).* Furthermore, calves fed heat-treated colostrum also had reduced colonization by *Escherichia coli* when compared to

calves fed fresh colostrum, implying that the type of colostrum provided influenced establishment of the microbiome in neonatal calves (78).

1.3.3 Infections and Antibiotics

Bacteria that cause enteric infections in neonatal calves usually induce changes in the microbiome and can cause calf death (79). Antimicrobials given to pre-weaned dairy calves with enteric infections may be effective if the pathogen is a bacterium (80); however, they usually alter the microbiome, perhaps making the calf more susceptible to various diseases (81). There are non-antimicrobial, alternative approaches to prevent or treat disease. In that regard, gut health depends on many factors (20). For example, manipulation of the gut microbiome with probiotics can improve calf gut health and prevent diarrhea (80, 81).

During infection of the bovine small intestine, such as with *Mycobacterium Avium* subsp. *paratuberculosis* (MAP), the relative abundance of *Proteobacteria* increased and the relative abundance of *Bacteroidetes* and *Firmicutes* decreased in adult dairy cows (12). Similar changes in the microbiome diversity in humans were accompanied by changes in the fecal virome. A relative increase in *Caudovirales* compared to *Microviridae* occurred in human patients with IBD (82). Perhaps the *Caudovirales* population within the phageome*,* more specifically the *Myoviridae*, are a good indicator of bacterial dynamics in the microbiome.

1.3.4 Different locations and variations

As previously described, microbiomes of ruminal liquid and feces have distinct bacterial and archaeal compositions. Relative abundance of the two main phyla, *Bacteroidetes* and *Firmicutes,* in the fecal microbiome is inverted compared to the rumen (9, 25). Methanogens, a subcomponent of archaea in the rumen, are not only more diverse but also more abundant in the rumen compared to the lower intestinal tract of adult ruminants (26). It is noteworthy that methanogens are already present in the GI-tract of milk-fed newborn dairy calves (83). Bacteriophages populations in cattle rumen are not stable, but vary over time, usually in diurnal variations (84). When cattle were fed once a day, the total bacteriophage population was lowest 2 hours post-feeding; soon thereafter, their numbers increased, peaking 8-10 hours post

feeding. Approximately 8-10 hours after feeding, rumen digestion was at its peak, whereby the number of bacteria available to infect was increasing and new phages were released to the rumen content. Finally, their numbers decreased to a stable population over the next 4 to 6 hours (84). This observation also suggests that when animals haven't been fed for a long time, phage numbers will be at a low level, since they may be in a temperate state with their hosts, depending on availability of prey (bacteria) to infect (39, 85).
Regarding bacteriophages in the large intestine of cattle, there is no information about their variation in time or their survival when they leave the rumen to pass through the large intestine.

Bacteriophages in feces from healthy adult humans, are highly personalized (even differ between housemates (36) and maintained over intervals of 1 and 2.5 years, with retention rates of 95 and 80% respectively (36, 86). Furthermore, phage to bacterium ratios in feces are low (1:10 or 1:1) when compared to intestinal mucosal surfaces (21:1-87:1) (36). Fecal human bacteriophages appear to be integrated in lysogenic state rather than lytic, in contrast to other microbiomes such as aquatic environments, where a higher relative abundance of lytic phages is present (36, 39).

1.4 Laboratory techniques to characterize the microbiome

There is a variety of culture-independent approaches, which can be broadly described as sequence-dependent and sequence-independent. The first approach requires knowledge of conserved portions of the genome to develop primers to amplify nucleic acid present in the sample such as the 16S ribosomal RNA subunit for bacteria (87), 18S or the internal transcribed spacer (ITS) regions of rRNA genes for fungi and protozoal communities (10) and family-specific genes for viruses (65). In contrast, sequence-independent techniques rely on sequencing all nucleic acids (NA), firstly by cutting the NA into small fragments, followed by random amplification of these fragments (so called deep sequencing) and lastly applying bioinformatics (88). This technique does not require *a priori* knowledge of genomic sequences and thus identifies microbiome diversity more objectively. Initial attempts to characterize microbiomes

were based on culturing bacteria (initially aerobic, then anaerobic and eventually more specialized approaches), with increasing numbers of organisms identified as culture techniques were refined (89). It is noteworthy that only some bacteria are easily cultured and can be studied with conventional methods (90); therefore, metagenomic analysis to characterize microbiomes provides much more complete information regarding the bacterial community.

1.4.1 Laboratory techniques for identification of bacteriophages

Both traditional and molecular methods can be combined to characterize bacteriophages. Transmission electron microscopy (TEM) is a conventional method to morphologically characterize bacteriophages (47) (Figures 1.4 - 1.6). The population of bacteriophages infecting all bacterial hosts in a sample can be estimated with epifluorescent microscopy (91) or flow cytometry (92). Bacteriophages infecting specific hosts can be determined through isolation using a plaque purification method with specific bacteria (93). All these methods result in a rough estimation of the abundance of bacteriophages.

Regarding the use of molecular approaches, there are no genes conserved among all phages. To overcome this lack of universal markers, specific gene-markers (sequence-dependent approach) detected through degenerate primers can be used to assess phage diversity and abundance (65). As not all identified markers are available in phages in various, it is likely that only known bacteriophages will be identified. The most appropriate approach to assess phage diversity is through the use of metagenomics, a sequence-independent approach (52). One difficulty to overcome in metagenomic analysis is the availability of sufficient amounts of viral nucleic acids prior to deep sequencing (94, 95). Various enrichment protocols are required to reduce host and bacterial genomes and to ensure that viral DNA or RNA are detectable. The most common procedures, herein described as Virus Particle Purification Procedures, are homogenization, centrifugation, filtration, density-gradient ultracentrifugation (e.g. CsCl) and nuclease treatments (96-98). In nuclease treatments, DNase removes host and bacterial DNA, whereas RNase removes ribosomal RNA (94-98). Following this enrichment step, nucleic acids are extracted (94, 96, 98). Extraction of viral DNA or viral RNA (which can be converted into cDNA)

must yield nucleic acids of sufficient purity and concentration for downstream library preparation and deep sequencing (94, 96).

Bioinformatic tools are needed to interpret data from viral metagenomes (99). Substantial computational language knowledge is required (100), or sometimes web-based applications can be used (101). Alignment of a viral sequence to a reference sequence requires three steps: i) sequence alignment (raw data or contigs); ii) assembly of contigs with a chosen algorithm; and iii) submission of data to a database for viral reference sequences. Furthermore, some sequences or contigs can be assembled in putative complete genomes, referred to as scaffolds (102-104). Viral metagenomic studies usually up to 90% of viral sequences unalignable (105), due to a lack of viral reference libraries. These unknown sequences can be further characterized using a 'K-mer' based classification (104, 105). This approach is based on variability in genomic composition and therefore frequency of nucleotide oligomers (K-mers) in various taxa. For example viral genomes have specific identifiable K-mers that can be used for taxonomical classification (105). Gene prediction, an essential tool for functional annotation, can be done using assembled contigs and applying a variety of algorithms (106). Predicted genes can be annotated against public databases (i.e. a homology-based approach for protein family analysis) to gain insights into gene function (106).

With respect to exploring bacteriophages using metagenomic analysis, two K-mer based , namely Kraken (105) and Centrifuge (107), can both be used to taxonomically classify sequences that are likely derived from viruses.

1.5 Thesis aims and overview

There is abundant evidence that microbiomes are complex, affected by many factors and have essential roles in health and disease in mammalian digestive systems. For cattle, the ruminal microbiome has been studied in detail; in contrast, only a few publications characterized the microbiome of the bovine large intestine or their modulation by bacteriophages such as *Caudovirales*.

There is mounting evidence that bacteriophages have key roles in shaping the rumen microbiome by influencing its diversity. However, at this moment, there are apparently no

publications that describe diversity of bacteriophages in bovine feces. Although TEM was used to morphologically characterize bacteriophages in bovine feces, molecular approaches could provide more detailed descriptions of this part of the microbiome. Metagenomic analysis can assess diversity and taxonomy of all bacteriophages present and perhaps provide insights into how a virome functions.

The research question addressed in this thesis is as follows:

What is the diversity and abundance of *Myoviridae* in the feces of beef and dairy cattle?

This question will be answered by addressing the following specific objectives:

1) Develop and evaluate methods to concentrate *Myoviridae* from cattle feces;

2) Visualize *Myoviridae* using electron microscopy;

3) Evaluate the use of PCR and metagenomic approaches to assess diversity of *Myoviridae*; and

4) Describe diversity of *Myoviridae* in fecal samples derived from two dairy and two beef cattle.

There is one hypothesis to be tested: Metagenomic analysis is a more suitable tool to estimate diversity and relative abundance of *Myoviridae* in cattle feces compared to a PCR-based method.

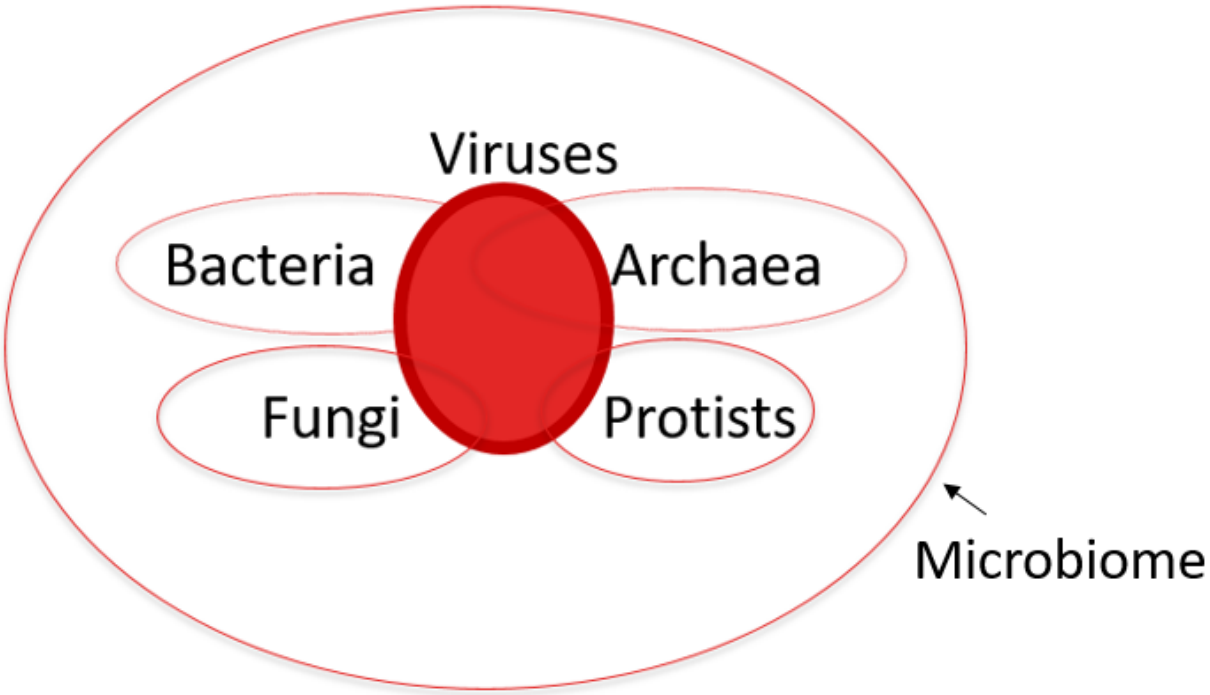

Figure 1.1 Composition and structure of a microbiome.

A microbiome is the community of microorganisms that inhabit a site or ecosystem, or a system of the body (e.g. digestive system), including bacteria, archaea, fungi, protists and viruses (1). Viruses infect all groups of microorganisms in a microbiome; their relationship with their hosts are represented with a red circle. In this thesis, we discuss prokaryotes (bacteria and archaea) and the viral component (phages infecting prokaryotes and viruses infecting eukaryotes) of the microbiome. The virus component that infects prokaryotes is called the phageome and consists of phages infecting archaea (archaeal viruses) and bacteria (bacteriophages).

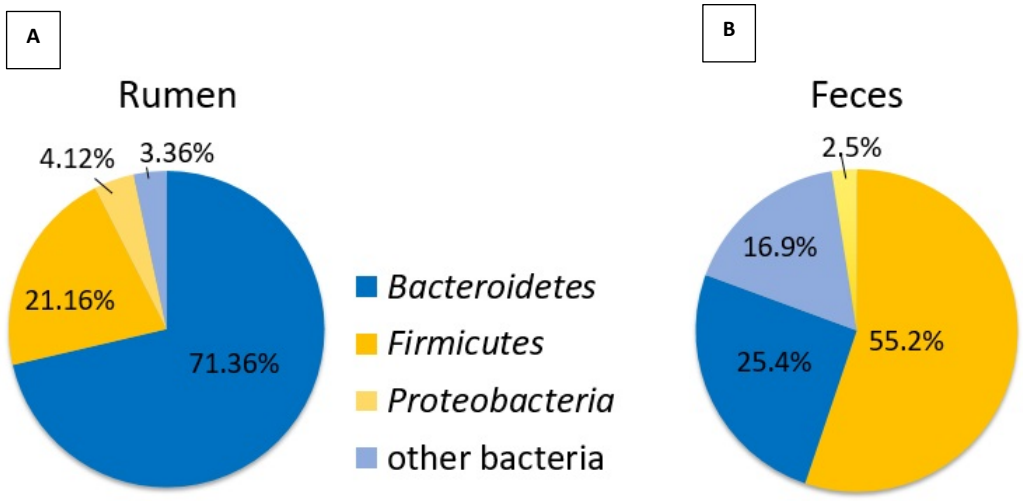

Figure 1.2 Comparison of ruminal and fecal microbiomes, based on bacterial composition

A. Bacterial composition of bovine ruminal microbiome (25).
B. Bacterial composition of bovine fecal microbiome (9).

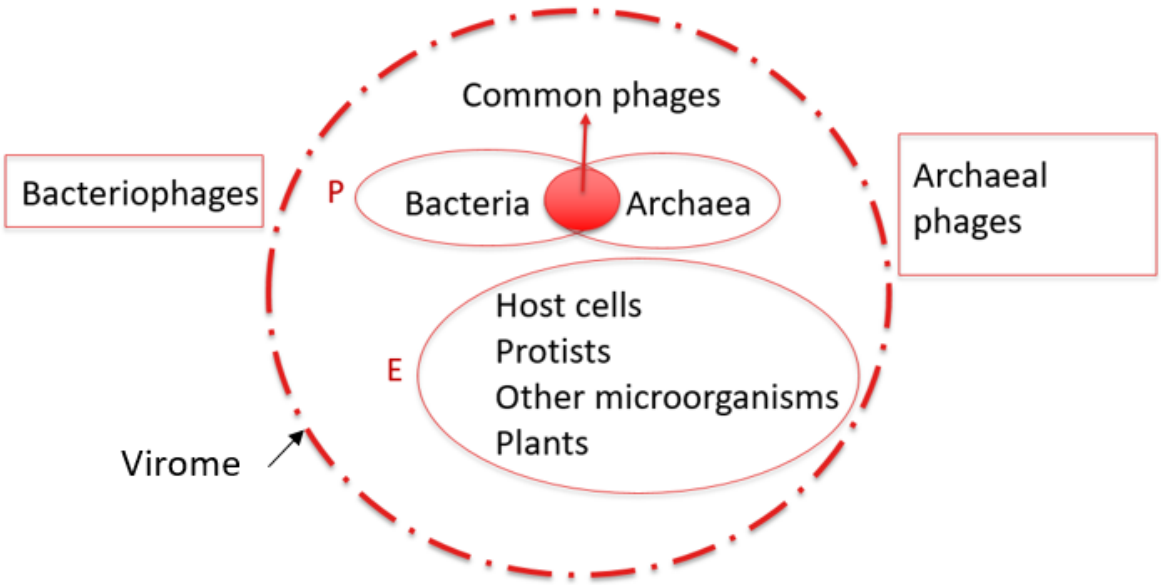

Figure 1.3 Composition and structure of a virome

A virome is the total community of viruses that inhabit the microbiome of an ecosystem or body site. The community is classified as viruses infecting Prokaryotes (P) called phages and viruses infecting eukaryotes (E) such as mammals and other microorganisms in the microbiome, e.g. protists and fungi (36). Phages belonging to families *Myoviridae* and *Siphoviridae* infect archaea and bacteria (47, 108).

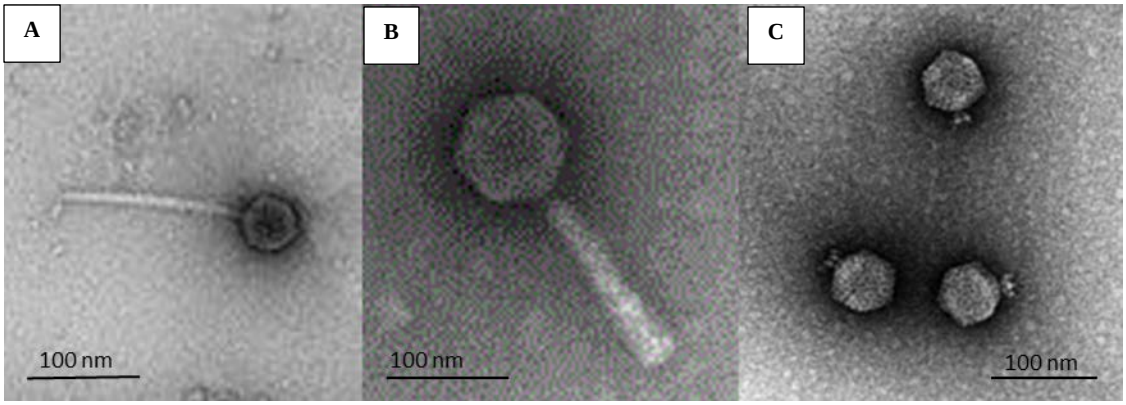

Figure 1.4 Composition of the order *Caudovirales*

Siphoviridae (A) (109), *Myoviridae* (B) (110) and *Podoviridae* (C) (111) are the main families of viruses belonging to the order *Caudovirales*. They are known as tailed phages. *Siphoviridae* have flexible tails, *Myoviridae* rigid tails and *Podoviridae* very short tails.

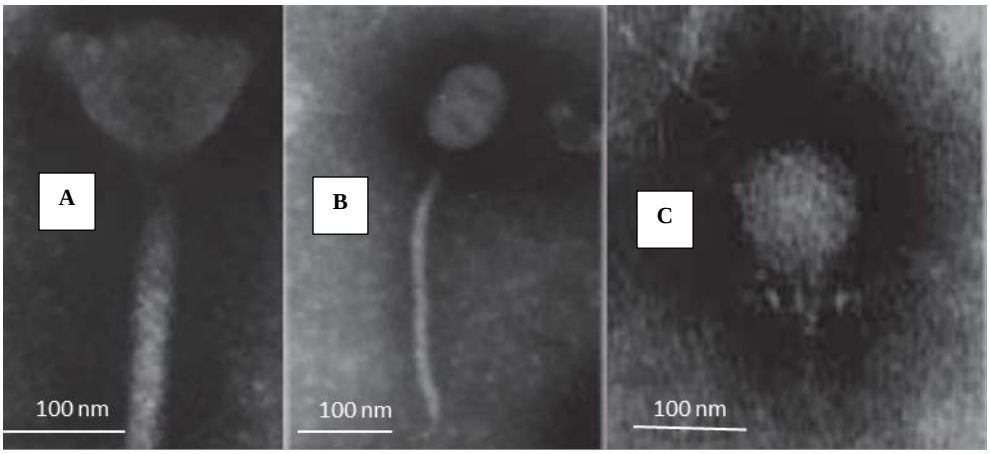

Figure 1.5 Morphotypes of ruminal bacteriophages, identified by EM

Myoviridae (A), *Siphoviridae* (B) and *Podoviridae* (C) are the main families of viruses belonging to the order *Caudovirales* (28).

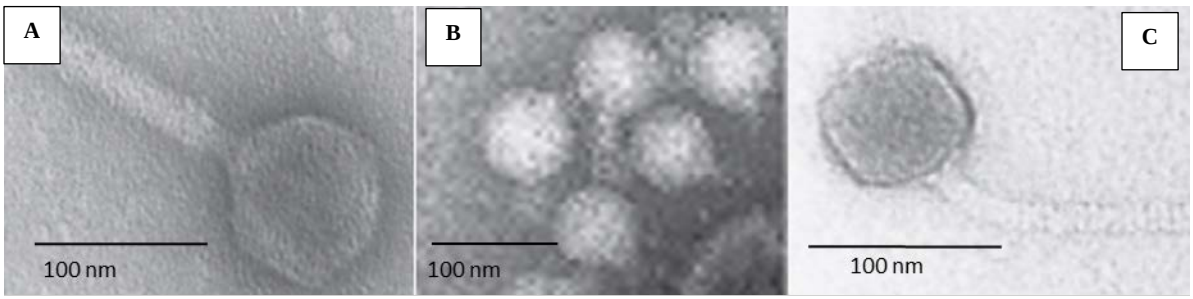

Figure 1.6 Morphotypes of bacteriophages from horse feces, identified by EM

Myoviridae (A), *Podoviridae* (B) and *Siphoviridae* (C) are the main families of viruses belonging to the order *Caudovirales* (39).

Chapter 2: Materials and Methods

2.1 Sample collection

Our goal was to characterize and compare fecal viromes of young, healthy cattle, both dairy and beef. During visits to dairy farms and beef feedlots between 2015 and 2018, 39 fecal samples were collected < 1 minute after defecation (Table 2.1). Samples were stored at 4°C, shipped in coolers to the University of Calgary, aliquoted and stored at -80°C until processed. Fecal samples were frozen without cryopreservatives, as dsDNA phages (*Myoviridae*) were our target and infectivity was not characterized (112). Sampling was based on age and location of cattle, with convenience samples collected from cattle that had defecated (Table 2.1).

2.2 Virus particle purification methods

Samples were sequentially processed using various virus particle purification (VPP) methods (Figure 2.2). All samples were homogenized, centrifuged, filtered and ultracentrifuged, followed by additional procedures, as detailed below.

VPP1: Homogenization and centrifugation

Each fecal sample was thawed for 2 hours at 4°C. Then, 1 g of feces was suspended in 2 mL sulphate-magnesium (SM) buffer (100 mM NaCl, 50 mM Tris·HCl, 8 mM $MgSO_4 \cdot 7H_2O$, 5 mM $CaCl_2 \cdot H_2O$, pH 7.4) in a 3 mL, 12 x 75 mm culture tube (VWR, Radnor PA, USA). The sample was homogenized at 3,000 rpm for 30 seconds in a PRO200 tissue homogenizer (Pro Scientific, Oxford CT, USA). Thereafter, the sample was transferred to a sterile conical propylene tube (Superclear™, VWR) that was centrifuged for 5 minutes at 2500 *g* (Allegra X-15 R, Beckman Coulter, Brea CA, USA). The supernatant was transferred to a sterile tube and centrifuged at 5000 *g* for 15 minutes. Two aliquots of the last supernatant were set apart for PCR and metagenomic analysis (MiSeq). The remainder of the supernatant was used for additional procedures.

VPP2: Filtration

Approximately 500 µL of the supernatant (from VPP1) was filtered through a 0.8 µm centrifugal filter (Sartorius, Göttingen, Germany), selected based on results reported in a modular

approach to customise sample (96). In that study, more viral reads were sequenced (88.8-96.7%), herpes viral reads increased (0.9%) as well as mimivirus reads dramatically increased (0.91%) compared to viral reads obtained from the mock virome (nine highly diverse viruses/phages and bacterial species common in the human gut) control using smaller filters (0.45 or 0.22 µm) (96). The filtrate was divided into aliquots for PCR and further metagenomic analysis (MiSeq and NextSeq).

VPP 3: CsCl gradient centrifugation

Approximately 500 µL of the filtrate (from VPP2) was layered on top of a discontinuous CsCl density gradient consisting of 3 x 1 mL of decreasing CsCl densities of 1.7, 1.5 and 1.35 g/mL in a 3.3 mL propylene tube (OptiSeal, Beckman-Coulter) and centrifuged at 66,000 *g* for 16 hours (MLS50; Beckman Coulter) at 4°C. The 1.35 and 1.5 g/mL density regions were collected together and divided in proportional volumes for the following two procedures to remove CsCl and re-suspend viruses in SM buffer.

VPP 3A-1: Amicon Filtration (no DNase treatment)

Approximately half (250 µl) of the volume from the fractions collected (from VPP 3) was washed with three volumes of SM buffer (4 mL) using an Amicon Ultra Centrifugal Filter 50,000 Molecular Weight Cut Off (MWCO; Beckman Coulter), following manufacture's protocol. This filtrate was used for EM and metagenomic analysis (MiSeq)

VPP 3A-2: Dialysis (no DNase treatment)

The other half (250 µl) of the volume from the fractions (from VPP 3) was directly injected into a 3500 MWCO Slide-A-Lyzer™ Dialysis Cassette (VWR), following manufacturer's protocol. This filtrate was used for EM and metagenomic analysis, using MiSeq and NextSeq platforms.

VPP 4: CsCl gradient centrifugation (from DNase treatment) and Amicon purification

The filtrate (from VPP2) was treated with 100 U/ml of DNAse (Promega Corp., Madison, WI, USA) at 37°C for 1 hour; thereafter, the filtrate was loaded onto a CsCl gradient, as described under VPP 3. The two fractions (1.5 and 1.35 g/mL) from the CsCl gradient were collected and washed with three volumes of SM buffer (4 mL) using an Amicon Ultra Centrifugal Filter 50,000 Molecular Weight Cut Off (MWCO; Beckman Coulter), following manufacture's protocol. This filtrate was used only for sequencing with the MiSeq platform.

2.3 Negative staining for EM

Nine samples were visualized by EM (Table 2.2); these were from 15-month old dairy heifers (i.e., 4, 38 and 39), 15-month old beef steers (36 and 37), 7-month old beef steers (34 and 35) and one-month old dairy heifers (31 and 32). Selection was based on positive results by PCR or metagenomic analyses (Table 2.3). Goals were to identify phages prior to metagenomic analysis using 1-butanol as a screening method and to confirm, in the fractions from CsCl (VPP 3), that these were phages from the family *Myoviridae* (Figure 2.3). The 1-butanol protocol for detecting viruses in feces was provided by Dr. Claudia Bachofen (Group Leader, Head of Diagnostic Unit, Institute of Virology, University of Zurich).

1- Butanol

For 1-butanol, each fecal sample (Table 2.2) was thawed for 2 hours at 4°C. Then, 1 g of feces was suspended in 6 mL sulphate-magnesium (SM) buffer in a 15-mL sterile conical propylene tube and 1 volume of 1-butanol was added. The sample was homogenized at 3,000 rpm for 3 min on a vortex. Thereafter, the sample was centrifuged at 700 *g* (Allegra X-15 R, Beckman Coulter) for 30 minutes, followed by incubation at 4°C for 5 hours or overnight until a clear separation into layers was observed. The clear aqueous phase was removed and transferred into a new 15 mL tube. The final volume was adjusted to 6 mL using SM buffer. Then, 6 mL of 1-butanol was added, mixed for 30 min using a vortex, followed by centrifugation for 30 min at 700 *g*. The tube was incubated at 4°C for 5 hours or overnight, until a clear separation into layers was observed. The clear aqueous phase was removed and transferred into a sterile tube suitable for mid-speed centrifugation at 10,000 *g* (Avanti®J-20 XP, Beckman Coulter) for 20 min. The supernatant was transferred to a 3.3 mL propylene tube and centrifuged at 40,000 *g*, for 2 hours in a MLS50 (Beckman-Coulter) at 4°C. The pellet was removed and re-suspended in 20 µL of SM buffer. Then, viruses were fixed and inactivated by adding formaldehyde (final concentration, 2%) and a 10-µL aliquot removed and processed (negative stain) in the Electron Microscopy facility (University of Calgary). This aliquot (10 µL) was mixed with an equal volume (10 µL) of 1% aqueous uranyl acetate. A carbon-coated TEM copper grid was floated with mixed sample for several minutes until it dried. The grid was floated on a drop of uranyl acetate 1% previously dropped on the parafilm solution for 30 seconds and then dried with filter paper.

One aliquot (2 µL) from each sample was examined under a Hitachi H7650 TEM (Hitachi High Technologies GLOBAL, Tokyo, Japan) at 80 kV and digital images acquired with an AMT 16000 CCD camera (Advanced Microscopy Techniques, Corp., Woburn, MA, USA) mounted on the microscope.

VPP 3A-1: Amicon Filtration

An aliquot (10 µL) of the filtered-CsCl fractions, (from the 1.5 and 1.35 g/mL CsCl fractions), was mixed with 10 µL of 1% aqueous uranyl acetate. Thereafter, examination by EM was done exactly as described above for 1-butanol.

VPP 3A-2: Dialysis Filtration

An aliquot (10 µL) of the dialyzed-CsCl fraction, containing both fractions, was mixed with 10 µL of 1% aqueous uranyl acetate. Thereafter, examination by EM was done exactly as described above for 1-butanol.

2.4 DNA extraction

Two methods of DNA extraction were used, based on previous results for purifying viral nucleic acids, specifically phages, in various biological samples (52). Mag-Bind Viral DNA/RNA kit (Omega Bio-Tek Inc., Norcross, GA, USA) is recommended for dsDNA viruses such as herpesvirus (113), whereas a PureLink Viral DNA/RNA extraction kit (Invitrogen-Thermo Fischer Scientific, Waltham MA, USA) is recommended for virus such as Hepatitis B virus, with a rigid protein coat. Mag-Bind Viral DNA/RNA kit was used (phage size was similar to Herpesvirus) for extraction of DNA prior to PCR. Conversely, PureLink Viral DNA/RNA extraction kit was used for extraction of DNA prior to metagenomics analysis, as reported (52). These methods of DNA extraction were used following manufacturer's protocol for polymerase chain reaction (PCR) and following manufacturer's protocol for metagenomic analysis. The concentration of DNA extracted from Mag-Bind Viral DNA/RNA kit was determined with a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific), whereas concentration of DNA extracted from the PureLink Viral DNA/RNA extraction kit was determined with a Qubit dsDNA HS kit in a Qubit machine (Thermo Fischer Scientific).

2.5 Polymerase Chain Reaction (PCR)

Four sets of degenerate primers (Table 2.5), previously used to explore cyanophages diversity belonging to *Myoviridae* in various habitats (65), were mapped using T4 phage NC_00086.4 (Figure 2.1) and tested using a reference control, *Escherichia coli* T4 phage ATCC 11303-B4 (kindly provided by Dr. Manuel Kleiner, Geoscience Institute, University of Calgary). Conditions for the Gp 20 gene-PCR were identical using various set of primers (Table 2.5). The final total reaction volume was 25 µL, comprised of: 150 ng of DNA of phage T4 as a template, 1.25 units of AccuTaq™ Lamba DNA Polymerase (Sigma-Aldrich), 3.75 pmol of dNTP, 5 pmol of forward primer CPS_3 (F-CPS_3) and reverse primer CPS_8 (R-CPS_8) and 1X PCR buffer (50 mM Tris-HCL, 100 mM NaCl, 0.15 mM $MgCl_2$). Amplification conditions were as follows: initial denaturation for 1.5 minutes at 94°C, followed by 35 cycles of 94°C for 45 seconds, 50°C for 1 minute, 72°C for 1 minute and a final extension step of 5 minutes at 72°C. A negative control was included; it consisted of all reagents plus sterile water in lieu of DNA template. The PCR products were separated on 1% agarose in 1X Tris-acetate–EDTA buffer (pH 8.0) at 100 V for 45 minutes.
One set of primers, F-CPS_3 and R-CPS_8 was selected for screening of specific phage using gene Gp20 (Major capsid protein) in cattle feces. PCR was performed using nucleic acids derived from 32 fecal sample filtrates (1-32), comparing two VPP methods (VPP1 and VPP2). The final total reaction volume was 25 µL, comprised of: 75-150 ng of DNA as a template, 1.25 units of AccuTaq™ Lamba DNA Polymerase (Sigma-Aldrich), 3.75 pmol of dNTP, 5 pmol of forward primer CPS_3 and reverse primer CPS_8. Amplification conditions were as follows: initial denaturation for 1.5 minutes at 94°C, followed by 35 cycles of 94°C for 45 seconds, 50°C for 1 minute, 72°C for 1 minute and a final extension step of 5 minutes at 72°C. A negative control was included; it consisted of all reagents plus sterile water in lieu of DNA template. The PCR products were separated on 1% agarose in 1X Tris-acetate–EDTA buffer (pH 8.0) at 100 V for 45 minutes.

2.6 Gel extraction and cloning

Amplicons ranging from 480 bp to 2 Kb were excised from the gel. The DNA bands were purified from the agarose gel, using the E.Z.N.A Gel extraction Kit (Omega Bio-Tek), following the manufacturer's protocol. Elution of DNA from the Hi-Bind column step was done twice, using 25 µL of heated elution buffer (70°C) to increase yield. Purified DNA from amplification of the Gp-like gene from reference strain T4 was sent to the sequencing facility at the University of Calgary for sequence analysis. For each VPP method, positive bands corresponding to the fecal sample were ligated into pGEM-T Easy Vector (Promega), following the manufacturer's protocol and transformed into One Shot® TOP10 Chemically Competent *E. coli* (Invitrogen Corp., Carlsbad, CA, USA). Transformed bacteria were streaked onto Lysogenic Broth (LB) agar plates with 100 µg/mL ampicillin and spread with 1.6 mg X-Gal (Promega) for blue-white screening. From each plate of transformed *E. coli*, five white colonies were selected for overnight culture on LB broth with 100 µg/mL ampicillin (113). Plasmids were extracted and purified from the *E. coli* culture using the E.Z.N.A Plasmid Mini Kit (Omega Bio-Tek) following the manufacturer's protocol and sent for sequencing using T7 and SP6 primers.

2.7 Cleaning Sequences

The DNA sequences obtained from Sanger sequencing were trimmed (Geneious 9, Biomatters Ltd. Auckland, New Zealand) to remove plasmid and primer sequences. Sequences were searched against published sequences in GenBank using the Basic Local Alignment Search Tool (BLAST(N); NCBI 2013) to confirm the identity of the Gp-like sequences. Identity of the query sequence was determined by comparing the closest related sequence hits with the greatest query coverage closest to the original length of the amplified sequence, the highest percentage identical sites (closest to 100%) and a cut-off value of 10^{-5}. In addition, sequences were inspected for quality by inspecting chromatograms.

2.8 Metagenomic analysis

2.8.1 Samples and VPP methods

2.8.1.1 MiSeq run

Three fecal samples from dairy cows (30, 31 and 32) and three fecal samples from beef cows (33, 34 and 35) were used for testing the four VPP methods (Figure 2.1). Five amplicons from the Gp-like gene PCR from fecal samples (11 and 28), including two amplicons of the phage T4 reference Gp-like gene controls, were included for metagenomic analysis (Table 2.3). These samples were sequenced with MiSeq (Illumina, San Diego, CA, USA).

2.8.1.2 NextSeq run

Another two dairy fecal samples (38 and 39) and two beef fecal samples (36 and 37) were used for testing two VPP methods (VPP2: Filtration and VPP3A: Dialysis) (Table 2.4). These samples were sequenced using NextSeq (Illumina).

2.8.2 Library preparation and sequencing

2.8.2.1 MiSeq run

Library preparation was performed with a Nextera XT v2 Kit (Illumina Co., San Diego CA, USA) following manufacturer's instructions, including a PhiX control (spiked at 30%) and a modified hybridization buffer for subnanomolar libraries (114). Phage PhiX control is an internal control that must be included in the run to measure fidelity of the DNA cluster generation. Due to a limited amount of DNA in all samples, normalization of the library required them to be spiked with PhiX control (30%) and used a modified hybridization buffer (114). Automatic cluster generation and paired-end sequencing with dual-index reads were performed on a MiSeq analyzer, in a single run, resulting in 250 bp paired-end reads in 600 cycles. Raw reads were stored in the BaseSpace cloud (Illumina).

2.8.2.2 NextSeq run

Total viral DNA from the fecal samples were sheared to approximately 350 bp using a Covaris S220 Sonicator (Woburn, MA, USA). Libraries were prepared using the NEBNext Ultra II Library

Preparation Kit (Illumina®) according to their user manual; adaptors used were from the NEBNext Multiplex Oligos for Illumina Set 1 and Set 2. After adaptor ligation, average library sizes ranged from 322-419 bp. Libraries were quantitated using KAPA qPCR Library Quantification Kit for Illumina platforms, then pooled and sequenced using paired-end reads (2 x 75 bp) with a 150 cycle mid-output cartridge on Illumina NextSeq 500 sequencer, according to Illumina guidelines.

Fecal samples processed with VPP2 required RNAses treatment (to eliminate RNA), prior to library preparation using a NEBNext Ultra II Library Preparation kit.

2.8.3 Sequencing processing and virus identification

All sequences from the run were verified to have passed quality control, using the FastQC tool (Babraham Informatics, Babraham Institute, Cambridge, UK). Data generated were processed with the following software: Kraken Taxonomic Sequence Classification System (Johns Hopkins University, Center for Computational Biology, Baltimore, MD, USA) and Centrifuge Microbial Taxonomical Classification (Johns Hopkins University, Center for Computational Biology).

Table 2. 1 Characteristics of cattle from which fecal samples were collected.

Sample	Type		Age	Sex		Housing			Procedure		
	Dairy	Beef	(mo)	Heifer	Steer	Pen	Feedlot	Hutches	EM	PCR	NGS
1	●		15	●		●				●	
2	●		15	●		●				●	
3	●		15	●		●				●	
4	●		15	●		●			●	●	
5	●		15	●		●				●	
6		●	7		●		●			●	
7		●	7		●		●			●	
8		●	7		●		●			●	
9		●	7		●		●			●	
10		●	7		●		●			●	
11	●		15	●		●				●	
12	●		15	●		●				●	
13	●		15	●		●				●	
14	●		15	●		●				●	
15	●		15	●		●				●	
16	●		15	●		●				●	
17	●		15	●		●				●	
18	●		15	●		●				●	
19	●		15	●		●				●	
20	●		15	●		●				●	
21	●		15	●		●				●	
22	●		15	●		●				●	
23	●		15	●		●				●	
24	●		15	●		●				●	
25	●		15	●		●				●	
26	●		1	●				●		●	
27	●		1	●				●		●	
28	●		1	●				●		●	
29	●		1	●				●		●	
30	●		1	●				●		●	●
31	●		1	●				●	●	●	●
32	●		1	●				●	●	●	●
33		●	7		●		●				●
34		●	7		●		●		●		●
35		●	7		●		●		●		●
36		●	15		●		●		●		●
37		●	15		●		●		●		●
38	●		15	●		●			●		●
39	●		15	●		●			●		●

Table 2. 2 Samples processed for EM.

Fecal sample	1-butanol	VPP 3	
		3 A-1 (Amicon)	3 A-2 (Dialysis)
4		●	
31	●		
32	●		
34	●		
35	●		
36	●		●
37	●		●
38	●		●
39	●		●

VPP: Virus particle purification procedure
3A-1 = CsCl ultracentrifugation and Amicon filtration
3A-2 = CsCl ultracentrifugation and dialysis

Table 2. 3 Fecal samples and amplicons sequenced with MiSeq.

Sample		**VPP**			
ID	Type	1	2	3 A-1 (- DNase)	4 (+ DNase)
28	Feces	●	●	●	●
29		●	●	●	●
30		●	●	●	●
33		●	●	●	●
34		●	●	●	●
35		●	●	●	●
T4	Amplicon	NA			
T4-1					
12		●			
28		●	●		

VPP: Virus particle purification;
1 = Homogenization and centrifugation
2 = Filtration
3A-1 = No DNase treatment prior to CsCl centrifugation, followed by Amicon filtration
3A-2 = No DNase treatment prior to CsCl centrifugation, followed by dialysis
4 = DNase treatment prior to CsCl centrifugation, followed with Amicon filtration

Amplicons:
Samples T4 and T4-1: As a positive control, Gp-like gene was amplified (two replicates) from phage T4 virus.
12 and 28: These samples were subjected to VPP2 prior to PCR. Gp-like genes were amplified and bands purified prior to library prep and deep sequencing.

Table 2. 4 Bovine fecal samples subjected to VPP and sequenced with NextSeq.

Fecal sample	1-butanol	VPP 3	
		3 A-1 (Amicon)	3 A-2 (Dialysis)
4		●	
31	●		
32	●		
34	●		
35	●		
36	●		●
37	●		●
38	●		●
39	●		●

VPP: Virus particle purification
2 = Filtration
3A-2 = CsCl ultracentrifugation and dialysis

Table 2. 5 Primers used to assess environmental diversity of *Myoviridae* in bovine feces.

Phage gene	Gene product	Target virus group	Primer direction-name	Sequence (5' – 3')	Reference
Gp20	Portal protein	Cyanophages: belong to *Myoviridae* family	F CPS_1	GTAG(T/A)ATTTTCTACATTGA(C/T)GTTGG	(65, 115)
			R CPS_2	GGTA(G/A)CCAGAAATC(C/T)TC(C/A)AGCAT	
			F CPS_3	TGGTA(T/C)GT(T/C)GATGG(A/C)AGA	(65, 116)
			R CPS_8	AAATA(C/T)TT(G/A/T)CCAACA(A/T)ATGGA	
			R CPS_4	CAT(A/T)TC(A/T)TCCCA(A/T/C)TCTTC	(65, 117)
			F $G_{20\text{-}2}$	G/C)(A/T)(A/G)AAATA(C/T)TTICC(A/G)AC(A/G)(A/T)A(G/T)GGATC	
Gp23	Major capsid protein	T4-related members of *Myoviridae*	F MZIAbis 1	GATATTTGIGGIGTTCAGCCIATGA	(65, 118)
			R MZIA 6	CGCGGTTGATTTCCAGCATGATTTC	

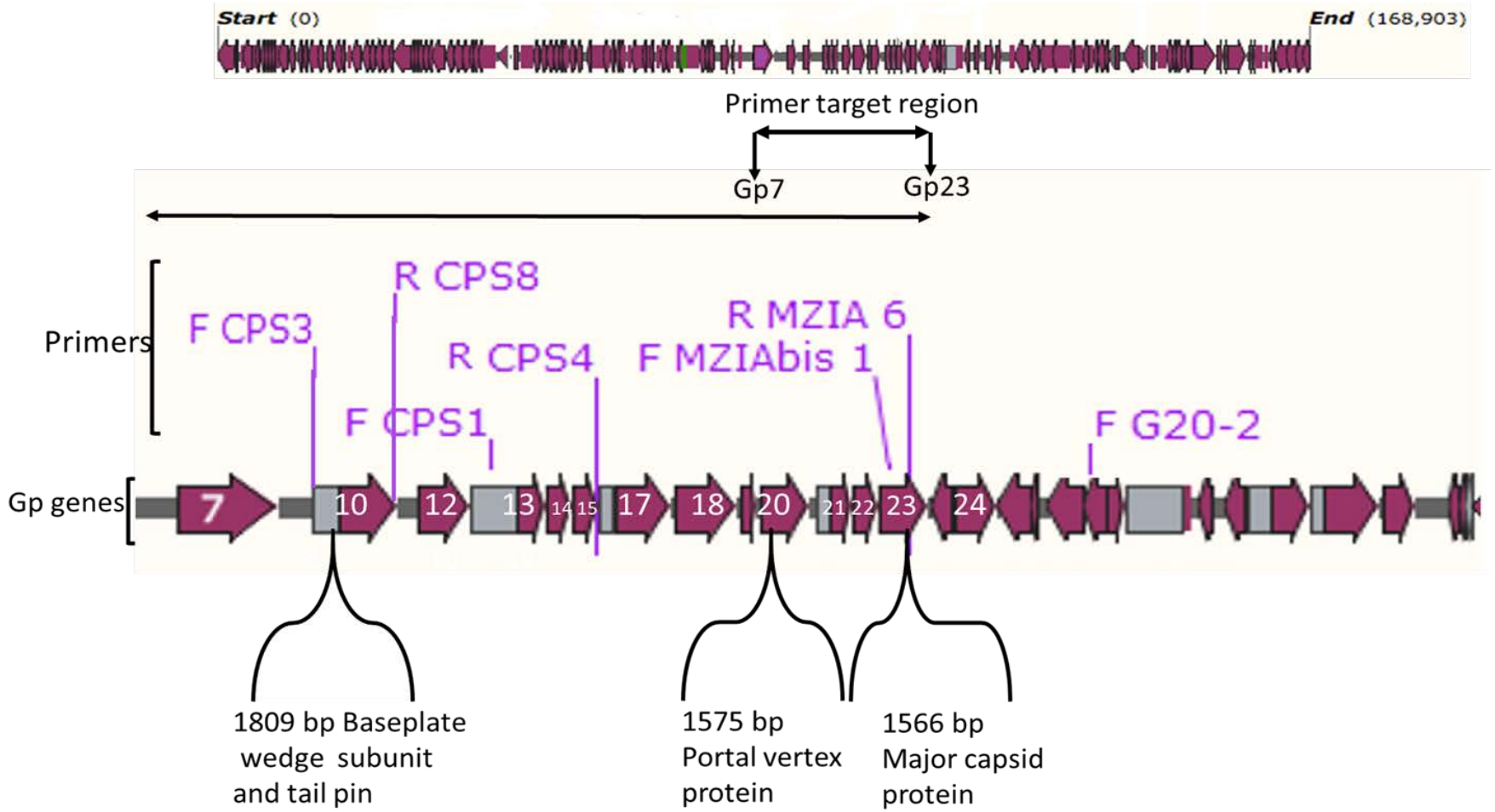

Figure 2. 1 Genetic Map of T4 phage NC_000866.4 (GenBank) (168,903 bp).

Primers were aligned on various Gp protein complex genes (Gp7, 10, 12, 13, 14, 15, 17, 18, 20, 21, 21 and 23); these genes encode proteins of the head, base plate and tail components of the phage. The baseplate consists of a hub surrounded by six wedges (119).

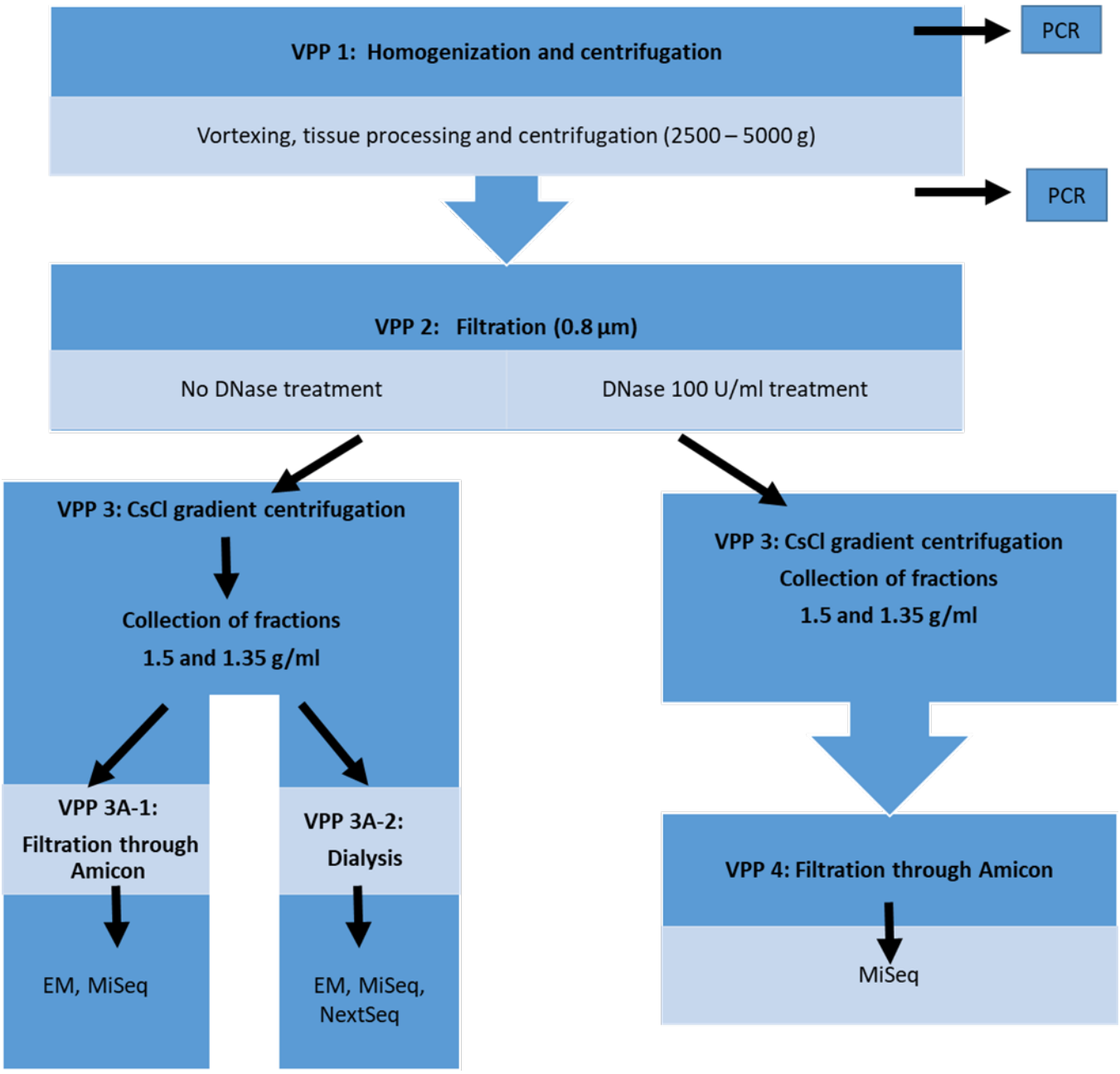

Figure 2. 2 Virus particle purification procedures.

VPP: Virus particle purification
1= Homogenization and centrifugation
2= Filtration
3A-1 = Filtration through Amicon (No DNase treatment)
3A-2 = Dialysis (No DNase treatment)
4 = Filtration through Amicon (DNase treatment)

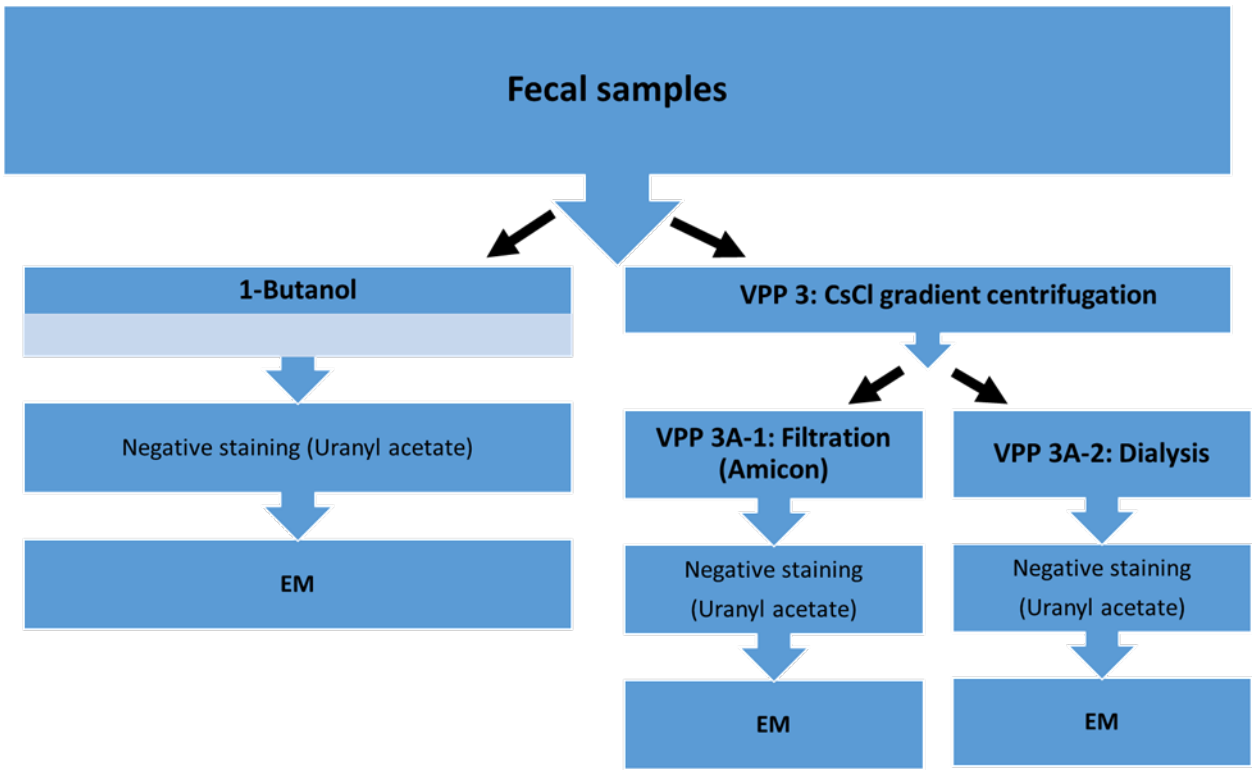

Figure 2. 3 Flow chart for fecal samples processed for EM.

Eight fecal samples (Table 2.1 and Figure 2.2) were processed using 1-butanol.
Fecal sample 4 (Table 2.5) was processed using CsCl, as described for VPP 3A-1 (Fig 2.2).
Samples 36 to 39 were prepared using CsCl, as described for VPP 3A-2 (Fig 2.2).

Chapter 3: Results

This chapter addresses specific objectives listed in Chapter 1. Three methods to assess diversity of *Myoviridae* developed in this study were used to compare methods to concentrate *Myoviridae* in feces.

3.1 EM for visualizing *Myoviridae*

Phages stained by uranyl acetate, from the family *Myoviridae,* appeared to be tailed and had heads of variable size, with a deep black color, due to affinity of staining for dsDNA**.** As mentioned in Section 2.3, eight fecal samples (i.e., 31, 32, 34, 35, 36, 37, 38 and 39) were selected, as they were positive for Gp like genes using PCR, or using next generation sequencing, *Myoviridae* reads were detected. This preparation, based on 1-Butanol, enabled virus visualization. The final preparation was not suitable for other applications, e.g. sequencing. However, it was used as a confirmatory test to detect phages in feces using various VPP. Pictures of phages, resembling members of the family *Myoviridae* from the order *Caudovirales* were obtained; selected images were captured and magnified for classification purposes (Figure 3.1). These phages were tailed phages with thick tails and big heads. (All photo credits: Wei-Xiang Dong, Microscopy and Imaging Facility, University of Calgary and Monica Rincon).

Five fecal samples (i.e., 4, 36, 37, 38 and 39) were processed by VPP type 3A and preparations were negatively stained (Figure 3.2). Selection of these five fecal samples was based on positive results for Gp-like genes by PCR (Fecal sample 4, Table 3.2) or to confirm presence of phages by visualizing them before DNA extraction for metagenomic analysis. Fecal samples (i.e., 36, 37, 38 and 39) were processed by 1-Butanol to observe morphotypes in 1 g of feces and by CsCl to confirm phages had been extracted before preparing the sample for deep sequencing. Pictures of phages, resembling members of the family *Myoviridae* from the order *Caudovirales,* were obtained through CsCl gradient ultracentrifugation and were selected and magnified (Figure 3.2). These were tailed phages with thick tails and big heads. (All photo credits: Wei-Xiang Dong, Microscopy and Imaging Facility, University of Calgary and Monica Rincon).

3.2 PCR for characterizing diversity of *Myoviridae* in bovine feces

3.2.1 Standardization of PCR using T4 phage

A PCR product (480 bp) was amplified with two sets of primers (F CPS_3 - R CPS_4 and F CPS_3 - R CPS_8; Figure 3.3). A clear and a defined band was obtained with primers F CPS_3 and R CPS_8 (116) using T4 phage as a control. This PCR product from phage T4 using primers F CPS_3 and R CPS_8, was Sanger sequenced. The sequence was identified by BLAST(N) as bacteriophage T4 (NCBI accession number KM607003.1) genes 22 (partial) and 23 (complete cds), with an identity of 94%, with a maximum score and query cover of 100% (E value = 10^{-5}). It is noteworthy that gene gp23 encodes for the major portal capsid protein.

3.2.2 Amplification of Gp-like gene from fecal samples

Based on results previously described (Section 3.2.1), primers F CPS_3 and R CPS_8 were used to assess samples 1-32. Purification with VPP 2 yielded more distinct bands compared to purification with VPP1. For both purification methods, PCR products of varying size were obtained (Table 3.1).

From the 32 collected fecal samples, Gp like genes were amplified in only eight samples (i.e., 4, 9, 10, 11, 12, 17, 28 and 29), representing only 40% of the total number (n=32) of fecal samples processed (Figures 3.4 - 3.9).

3.2.3 Identification of amplified sequences

For eight samples (4, 9, 10, 11, 12, 17, 28 and 29), positive bands ranging from 480 bp to 2 Kb were purified from the gel (Table 3.2). Clones (n=130) were sent to Eurofins Genomics (Louisville, KY, USA) for bi-directional sequencing using T7 and SP6 primers, compatible with the vector system used. Of these 130 clones, only 21 (16.2%) were sequenced and only five clones (Table 3.2) met the quality control standard using BLAST(N), with a cut-off value of 10^{-5}. Of the five clones identified, two were classified as belonging to *Myoviridae*.

However, they only accounted for only a small number of all clones obtained from the gels (n=130). It was noteworthy that most clones sequenced (n=21) did not have a significant match with any genome within GenBank (Table 3.2).

3.2.4 Metagenomic analysis

3.2.4.1 MiSeq analysis

In total, 1,563,458 raw sequences were generated (Table 3.3). Quality evaluation of reads was performed using FASTC Version 0.11.5. The average sequence quality score was $\geq$ 30 for most raw sequences generated during the run. Consequently, there was no need to remove low-quality sequence reads.

The average relative abundance of bacterial reads was 40.84% (± 22.35) in relation to total number of classified reads. Furthermore, relative abundance of viral reads ranged from 97.3 to 0.01% of total number of reads classified by the Kraken program (Table 3.3). Viral reads with a relative abundance of ~97% in relation to the total number of classified reads were attributed to amplicons included in the run (i.e., T4, T4-A, 12-A, 28-B and 28-C). Due to extreme variation of viral reads, sequences were compared and analyzed as sample sets: 1st sample set included six fecal samples (i.e., 28, 29, 30, 33, 24 and 35) with four types of VPP (1-4), whereas the 2nd sample set included amplicons (Table 3.5).

In the 1st sample set, an outlier was detected (Sample 34, VPP3; Table 3.3 and Figure 3.10). Comparing this outlier to the 2nd sample set, sample T4-A (positive control) and 34, VPP3 had similar total viral read numbers (492,204 and 422,861 respectively). Therefore, 34 VPP3 was removed from subsequent analysis, as it likely represented a contamination with sample T4-A, the amplicons derived from T4-phage used as a positive control. Furthermore, the Kraken analysis was consistent with assigning all viral reads to a single type of phage (*E. coli* phage T4) for both sample T4A and 34-3. Following this removal, number of viral reads from each fecal sample for each VPP (1-4) was still not comparable, due to great variability in number of classified reads between and within samples (Figure 3.11).

Total viral reads derived from fecal samples according to each VPP for MiSeq run, based on Kraken, were classified into the three families of the order *Caudovirales* (Table 3.4). Relative

abundance of *Caudovirales* in relation to total number of viral reads, corresponded to 64.3%, of which *Myoviridae* corresponded to 42.27%, *Siphoviridae* to 17.85% and *Podoviridae* to 4.18%. When the RA was calculated in relation to total reads of *Caudovirales*, RA of *Myoviridae* was 65.78%*, Siphoviridae* 28.89% and *Podoviridae* 5.33% (Table 3.4)*.* Although relative abundance and diversity of *Myoviridae* was our research question, the other two families in the same viral order were also characterized to determine their abundance in relation to *Myoviridae*, as they formed a main component of the fecal phageome. Nevertheless*,* only family *Myoviridae* was further analyzed into genera and species as a component of the diversity and as part of the specific objectives. Subsequently, the only genus detected was *T4-like virus* (Table 3.6) with a RA of 69.56% in relation to total reads in the family *Myoviridae.* The only species detected with a reliable number of reads (> 10) in the genus *T4-like virus* corresponded to *Escherichia coli* T4 phage (Table 3.7) with an RA of 70.76%. Based on these results (MiSeq), *Myoviridae* was the most predominant family of phages in assessed cattle feces, despite great variability in number of reads between and within samples (Figure 3.12). Similarly, when analysis of the RA of *Myoviridae* in relation to the *Caudovirales* reads were compared between VPP in each fecal sample, there was great variability in each VPP within each sample and between samples (Figure 3.13).

Analysis of amplicons T4 and T4-A (Table 3.5), consisting of amplicons derived from the PCR described above (Sections 3.2.1 and 3.2.2), confirmed that *E. coli* phage T4 reference strain was the only virus present in samples T4 and T4-A and no reads belonged to *Siphoviridae* or *Podoviridae*. Additionally, amplicons from fecal samples 12 and 28 (two replicates: 28-B and 28-C) had RAs of *Myoviridae* (61.76%), *Siphoviridae* (35.51%) and *Podoviridae* (2.72%), (Table 3.5 and Figure 3.14) were included in the analysis. Further classification of these amplicons (Table 3.7) confirmed that the amplified products corresponded to *T4-like virus* (genus) and *E.coli* phage T4 (species) with RAs of 93.6% in sample T4 and 96.6 in T4-A amplicons. In addition, amplicons from fecal samples 12 and 28 (for this sample, we used two replicates: 28-B and 28-C) yielded RAs of 61.15% (Sample 12-A), 25% (Sample 28-B) and 63.15% (Sample 28-C) for “T4 like virus“ (genus level).

3.2.4.2 NextSeq analysis

A total number of 171,371,746 sequences were generated, corresponding to an average of 21,421,468 reads per sample (± 3,175,122) (Table 3.8). FASTC version 0.11.5 was used to evaluate sequence quality (average score was ≥ 30). Consequently, there was no need to remove low-quality sequence reads. *Bos taurus* sequences were removed and bacterial reads were only 8.72% (± 3.25) across samples. Percentage of total reads that could be classified was 17.62% (± 3.05), with 0.1% (± 0.039) of viral origin (Table 3.8). Viral read numbers were consistent between and within samples and between VPP methods (Figure 3.15).

In an analysis of the *Caudovirales* component, *Caudovirales* represented 35.43% (± 5.14) in relation to the total number of viral reads (Table 3.9 and Figure 3.16) and 93.33% in relation to total dsDNA virus reads (Table 3.12 and Figure 3.19). Additionally, the RA in relation to the total virus reads of the virus families *Myoviridae, Siphoviridae and Podoviridae* was 16.45% (± 2.70), 13.80% (± 5.0) and 4.84% (± 1.5) respectively (Table 3.9 and Figure 3.16). The RA in relation to total *Caudovirales* reads obtained was 46.19, 37.28 and 13.95% for *Myoviridae*, *Siphoviridae* and *Podoviridae,* respectively, whereas the remaining 2.58% were unclassified (Table 3.9 and Figure 3.21). Although RA and diversity of *Myoviridae* was our research question, 38 genera in the same family *Myoviridae* were additionally classified (Table 3.11), to confirm their diversity by number of genera within the family *Myoviridae* in cattle feces. Based on these results, the most predominant genera corresponded to *T4-like virus*, *Schizot4virus, Kayvirus, Phikzvirus and Cp220virus* (Table 3.10). These genera were also characterized to the species level (Appendix A, Tables 1 to 5). Genus T4-like viruses were classified into 72 species (Appendix A, Table 1).

It was evident that *Caudovirales* was the most abundant taxon in the group of dsDNA phages (Table 3.12). Furthermore, it was noteworthy that archaeal phages belonging to the order *Ligamenvirales* were also detected in all samples, with an average RA of 0.59% (Table 3.12 and Figure 3.19).

Due to the small number of samples of cattle feces included in both metagenomic analyses, neither statistical analyses nor diversity indexes could be used in these analyses.

In conclusion, this last metagenomic analysis (NextSeq) revealed the composition of the dsDNA virome of feces from two dairy and two beef cattle (Table 3.12). All samples analyzed contained

a similar phageome structure in all cattle feces, independent of type of VPP procedure. Furthermore, when comparing RA of the three major families of phages that constituted the phageome, *Myoviridae* was the most abundant family in this group of samples of cattle feces and order *Caudovirales* the most abundant taxa order in these cattle feces.

Table 3. 1 Cattle fecal samples and PCR products.

Sample ID	VPP 1					VPP 2					Figure
1					NP						3.4
2	480					480		800			
3					NP						
4			800					800			3.5
5	480		800				640				3.4
6			800				640				3.5
7			800				640				
8	480		800	1000		480	640				3.4
9	480			1000				800	1000		
10	480		800								3.5
11					NP			800	1000		3.6
12					NP			800	1000		
13					NP					NP	
14					NP					NP	3.7
15					NP					NP	
16					NP					NP	
17					NP	480	640			NP	
18					NP	480				NP	3.8
19					NP					NP	
20					NP					NP	
21					NP					NP	
22					NP					NP	
23					NP					NP	
24					NP					NP	
25					NP					NP	
26					NP					NP	3.9
27					NP					NP	
28	480	480		1000		480	640		1000		
29		480		1000		480	640	800			
30				1000					1000		
31					NP					NP	
32		480				480			1000		

VPP: Virus particle purification procedure
1 = Homogenization and centrifugation
2 = Filtration
NP: no PCR product detected

Table 3. 2 Taxonomical Identification of amplified sequences using BLAST(N).

Fecal sample ID	Clone ID	Query size sequence (bp)	Score	Query Cover (%)	E score	Identification
4	4-A	189	183	83	$1e^{-45}$	*E. coli* Phage slur07
9	9-A	163	71	52	$1e^{-12}$	*Shigella* Phage SHBML
	9-B	472	NSM			
	9-C	301	NSM			
	9-D	301	NSM			
	9-E	343	NSM			
	9-F	458	NSM			
	9-G	472	NSM			
10	10-A	157	NSM			
11	11-A	94	NSM			
12	12-A	818	226	38	$2e^{-89}$	Uncultured bacterium partial 16S rRNA: *Provotella ruminicola* (100%)
17	17-A	521	NSM			
	17-A	870	NSM			
28	28-A	337	NSM			
	28-B	271	NSM			
	28-C	376	NSM			
	28-D	828	556	51	2^{e-154}	Uncultured bacterium from a study in anaerobic microbiome in poultry manure
	28-E	490	NSM			
	28-F	480	NSM			
29	29-A	591	NSM			
	29-B	596	540	36%	2^{e-149}	Uncultured bacterium from an aquatic study.

Note: Query size sequence: Size of the sequence that aligned with reference genome sequence in NCBI database. Score is a numerical value that describes the overall quality of an alignment; higher scores indicate greater similarity.
E = number of hits assigned to the sequence aligned; lower E-values indicate a better match.
NSM = No significant match.

Table 3. 3 Metagenomic analysis using MiSeq

Sample ID		No. reads		Relative abundance (%)				No. reads		
	VPP	Raw	Classified	Bacteria (*)	Virus (*)	All viruses	Caudovirales	Myoviridae	Siphoviridae	Podoviridae
28	1	13,176	13,147	41.52	0.74	98	55	36	11	8
	2	7,568	7,556	35.21	3.03	229	179	175	3	1
	3A-1	9,737	9,725	63.57	0.37	36	22	15	5	2
	4	25,059	25,039	53.65	0.36	90	31	17	14	0
29	1	4,783	4,767	64.65	0.71	34	10	3	0	7
	2	1,349	1,349	54.95	1.93	26	6	6	0	0
	3A-1	23,860	23,843	58.0	1.00	239	80	64	16	0
	4	111,034	45,947	59.65	0.90	415	45	17	22	6
30	1	11,734	8,154	18.56	0.11	9	3	2	1	0
	2	26,434	19,672	27.5	0.08	15	2	0	2	0
	3A-1	43,813	33,196	68.7	0.14	48	26	19	7	0
	4	179,741	127,997	66.8	0.11	142	29	13	14	2
33	1	14,781	2,795	52.33	1.64	46	20	20	0	0
	2	7,577	1,502	43.17	1.2	18	10	8	2	0
	3A-1	8,141	1,457	49.9	4.31	63	55	54	1	0
	4	10,724	7,136	66.68	0.56	40	29	24	5	0
34	1	87,060	16,546	52.33	0.68	112	44	25	14	5
	2	72,095	44,692	7.59	0.13	57	20	12	6	2
	3A-1	437,904	431,711	0.44	97.1	422,861	NA	NA	NA	NA
	4	89,716	60,636	66.12	0.01	58	22	13	8	1
35	1	68,189	12,901	51.48	1.11	144	73	70	1	2
	2	89,716	4,944	48.45	1.11	55	33	27	5	1
	3A-1	27,540	1,069	53.17	9.22	99	91	91	0	0
	4	12,672	8,138	13.21	0.11	9	3	3	0	0
T4	Amplicons	32,503	31,767	1.40	95.5	30,730	29,871	29,871	0	0
T4-A		511,201	503,644	0.14	97.3	492,204	480,075	480,074	1	0
12-A		826,895	46,290	39.24	2.29	1,060	98	55	35	8
28-B		252,864	16,987	11.53	1.07	182	8	4	4	0
28-C		886,853	50,851	14.55	1.01	516	24	19	5	0
			Average	**40.84 (± 22.35)**						

VPP: Virus particle purification procedure

1 = Homogenization and centrifugation; 2 = Filtration; 3A-1 = No DNase treatment, CsCl ultracentrifugation and Amicon filtration; 4 = DNase treatment, CsCl ultracentrifugation and Amicon filtration.*Relative abundance (%) was calculated in relation to number of raw reads.

Table 3. 4 Relative abundance of *Caudovirales* (and its families) in bovine feces (MiSeq).

		Relative abundance (%)						
Sample ID	**VPP**	***Caudovirales* (*)**	***Myoviridae* (*)**	***Myoviridae* (**)**	***Siphoviridae* (*)**	***Siphoviridae* (**)**	***Podoviridae* (*)**	***Podoviridae* (**)**
28	1	56.1	36.7	65.45	11.22	20	8.16	14.54
	2	78.17	20.42	97.76	1.31	1.68	0.43	0.56
	3A-1	61.11	55.5	68.11	13.89	22.7	5.55	9.09
	4	34.44	43.3	54.84	15.55	45.16	0	0
29	1	29.41	11.1	30	0	0	20.58	70
	2	23.01	23	100	0	0	0	0
	3A-1	33.47	36.8	80	6.98	20	0	0
	4	10.84	10.37	37.77	5.3	48.88	1.44	13.35
30	1	33.33	33.33	66.66	11.11	33.33	0	00
	2	13.33	13.33	13.33	13.33	86.67	0	0
	3A-1	54.16	56.25	73.07	14.58	26.92	0	6.89
	4	20.42	20.42	44.82	9.86	48.27	1.40	0
33	1	43.48	43.48	100	0	0	0	0
	2	55.55	55.5	80	11.11	20	0	0
	3A-1	87.30	87.3	98.19	1.59	1.81	0	0
	4	72.5	72.5	82.75	12.5	17.24	0	0
34	1	39.28	39.29	56.81	12.5	31.81	4.46	11.36
	2	35.09	35.08	60	3.51	30	3.5	10
	3A-1	NA	NA	NA	NA	NA	NA	NA
	4	37.93	37.93	59.09	13.79	36.36	1.72	4.54
35	1	50.70	59.02	95.89	0.70	1.37	1.38	3.03
	2	60	56.36	81.81	49.1	15.15	1.8	3.04
	3A-1	91.91	91.9	100	91.9	0	0	0
	4	33.33	33.33	100	33.3	0	0	0
Average		**64.3**	**42.27**	**65.78**	**17.85**	**28.89**	**4.18**	**5.33**

VPP: Virus particle purification procedure

1 = Homogenization and centrifugation; 2 = Filtration; 3A-1 = No DNase treatment, CsCl ultracentrifugation and Amicon filtration; 4 = DNase treatment, CsCl ultracentrifugation and Amicon filtration.

*Relative abundance (%) was calculated in relation to total viral reads.

**RA of each family in relation to total *Caudovirales* reads.

Table 3.5 Relative abundance of *Caudovirales* (and its families) from amplicons.

Amplicons	No. reads				Relative abundance (%)			
	Caudovirales	***Myoviridae***	***Siphoviridae***	***Podoviridae***	***Caudovirales* (*)**	***Myoviridae* (**)**	***Siphoviridae* (**)**	***Podoviridae* (**)**
T4	29,871	29,871	0	0	60.68	100	0	0
T4-A	480,075	480,075	0	0	97.35	100	0	0
12-A	98	55	35	8	9.24	56.12	35.71	8.17
28-B	8	4	4	0	4.39	50	50	0
28-C	24	19	5	0	4.6	79.17	20.83	0
				Average	**6.07**	**61.76**	**35.51**	**2.72**

T4 and T4A correspond to amplicons from T4 phage with primers F-CPS_3 and F-CPS_8.
Samples 12-A, 28-B and 28-C correspond to amplicons from fecal samples (12 and 28) using primers F-CPS_3 and R-CPS_8.
*Relative abundance (%) was calculated in relation to total viral reads.
**RA of each family in relation to total *Caudovirales* reads

Table 3. 6 Relative abundance of T4 virus and phage T4 within *Myoviridae* in fecal samples from MiSeq.

Sample ID	VPP	No. reads		Relative abundance (%)	
		"T4 virus" (Genus)	***E. coli* phage T4 (Species)**	**"T4 virus" (Genus) (*)**	***E. coli* phage T4 (Species) (*)**
28	1	24	22	66.67	66.67
	2	157	147	89.71	89.71
	3A-1	12	12	80	80
	4	9	8	52.94	52.94
29	1	3	2	100	100
	2	6	5	100	100
	3A-1	53	47	82.81	82.81
	4	5	3	29.41	29.41
30	1	1	0	50	50
	2	0	0	0	0
	3A-1	13	8	68.42	68.42
	4	1	0	7.59	7.59
33	1	15	15	75	75
	2	5	5	62.5	62.5
	3A-1	52	48	96.29	96.29
	4	22	21	91.67	91.67
34	1	10	5	40	40
	2	10	3	83.3	83.3
	3A-1	NA	NA	NA	NA
	4	11	7	84.61	84.61
35	1	50	44	71.42	71.42
	2	20	16	74.07	74.07
	3A-1	85	75	93.40	93.40
	4	3	3	100	100
			Average	**69.56**	**70.76**

VPP: Virus particle purification

1 = Homogenization and centrifugation; 2 = Filtration; 3A-1 = No DNase treatment, CsCl ultracentrifugation and dialysis; 4 = DNase treatment, CsCl ultracentrifugation and Amicon filtration.

*Relative abundance (%) was calculated in relation to *Myoviridae* reads.

Table 3. 7 Relative abundance of "*T4 virus*" and phage T4 within *Myoviridae* in amplicons from fecal samples (MiSeq).

Amplicon ID	Target phage gene (by PCR)	BLAST(N)			NGS (MiSeq)			
		Gene identified	**E value**	**Query Cover (%)**	**No. reads**		**Relative abundance (%)**	
					"T4 like viruses"	*E. coli* phage T4	"T4 like viruses" (*)	*E.coli* phage T4 (*)
T4	Gp 10	Gp 22 and Gp 23	$3.78e^{-150}$	85	28,781	26,244	93.6	91.18
T4-A	Gp 10	Gp 22 and Gp 23	$3.78e^{-150}$	85	463,977	448,202	96.64	96.6
12-A	Unknown gene from *Provotella ruminicola*	NSM	$2e^{-89}$	38	34	5	61.2	14.70
28-B	Unknown	NSM			1	1	25	4
28-C	Unknown	NSM			12	9	63.15	75

T4 and T4A correspond to amplicons from T4 phage with primers F-CPS_3 and R-CPS_8.
Samples 12-A, 28-B and 28-correspond to amplicons from fecal samples 12 and 28, using primers F-CPS_3 and R-CPS_8.
*Relative abundance (%) was calculated in relation to total *Myoviridae* reads.

Table 3. 8 Metagenomic analysis using NextSeq.

Sample ID	VPP	No. reads								Relative abundance (%)		
		Total	**Virus**	**dsDNA**	***Caudovirales***	***Myoviridae***	***Siphoviridae***	***Podoviridae***	**Unclassified *Caudovirales***	**Classified (*)**	**Bacteria (*)**	**Virus (*)**
36	2	24,166,902	18,997	6,589	6,134	3,230	1,935	841	128	23.1	15.1	0.08
	3A-2	17,043,821	31,813	8,857	8,503	3,720	2,116	2,436	231	18.6	6.3	0.19
37	2	25,804,978	18,886	7,413	6,821	3,210	2,698	748	165	19.1	12.1	0.07
	3A-2	20,134,394	16,944	6,824	6,418	3,272	2,170	857	119	16.5	7.08	0.08
38	2	22,894,213	22,363	8,048	7,451	4,187	2,350	704	210	17	8.3	0.09
	3A-2	17,043,821	20,157	9,662	9,020	3,333	4,514	918	255	14.2.	5.9	0.11
39	2	21,582,188	16,685	6,678	6,164	2,889	2,463	621	191	15.4	8.61	0.08
	3A-2	22,701,429	21,149	9,099	8,507	2,980	3,984	1,313	230	13.6	6.38	0.09
Average		**21,421,468**								**17.62**	**8.72**	**0.1**

VPP: Virus particle purification
2 = Filtration;
3A-2 = No DNase treatment, CsCl ultracentrifugation and dialysis.
*Relative abundance (%) was calculated in relation to total number of reads.

Table 3. 9 Relative abundance of viral families in the order *Caudovirales* from NextSeq analysis of fecal samples from two dairy and two beef cattle.

Sample ID	VPP	Relative abundance (%)							
		As a proportion of total viral reads				As a proportion of *Caudovirales* reads			
		Caudovirales	*Myoviridae*	*Siphoviridae*	*Podoviridae*	*Myoviridae*	*Siphoviridae*	*Podoviridae*	Unclassified
36	**2**	32.29	17	10.18	4.44	52.66	31.54	13.71	2.09
	3A-2	26.74	11.69	6.65	7.66	43.75	24.88	28.64	2.73
37	**2**	36.11	16.99	14.28	3.96	47.06	39.55	10.97	2.42
	3A-2	37.88	19.31	12.81	5.06	50.98	33.81	13.35	1.86
38	**2**	33.31	18.72	10.51	3.14	56.19	31.54	9.4	2.87
	3A-2	44.74	16.53	22.39	4.55	36.95	50.04	10.17	2.84
39	**2**	36.94	17.31	14.76	3.72	46.87	39.95	10.07	3.11
	3A-2	40.22	14.09	18.83	6.21	35.03	46.83	15.43	2.71
	Average	**35.46**	**16.45**	**13.80**	**4.84**	**46.19**	**37.28**	**13.95**	**2.58**

VPP: Virus particle purification
2 = Filtration;
3A-2 = No DNase treatment, CsCl ultracentrifugation and dialysis.

Table 3. 10 Relative abundance of the most predominant genera within *Myoviridae* (NextSeq).

Sample ID	VPP	Relative abundance (%) in relation to *Myoviridae* reads				
		T4-like virus	***Schizot4virus***	***Kayvirus***	***Phikzvirus***	***Cp220virus***
36	**2**	11.30	0.83	2.85	1.2	0.86
	3A-2	4.45	0.48	6.64	0.91	2.5
37	**2**	9.87	1.71	0.62	1.71	0.56
	3A-2	5.35	0.03	1.0	1.87	0.40
38	**2**	9.86	2.12	0.23	1.34	1.60
	3A-2	8.16	2.37	0.33	2.1	1.60
39	**2**	9.93	2.11	0.21	2.36	0.97
	3A-2	9.26	1.51	0.17	1.6	0.10

VPP: Virus particle purification
2 = Filtration;
3A-2 = No DNase treatment, CsCl ultracentrifugation and dialysis.

Table 3. 11 Classification of *Myoviridae* family into genera (number of reads), based on NextSeq sequencing.

	Genus within *Myoviridae*	Sample ID							
		36		37		38		39	
		VPP		VPP		VPP		VPP	
		2	3A-2	2	3A-2	2	3A-2	2	3A-2
1	*T4virus*	365	167	317	175	413	272	287	276
2	*Sp18virus*	32	25	18	7	70	13	11	15
3	*Schizot4virus*	27	18	55	43	89	79	61	45
4	*Js98virus*	20	2	8	7	10	11	6	6
5	*Jd18virus*	17	26	12	3	21	7	7	9
6	*Rb49virus*	12	9	8	7	10	6	3	3
7	*Rb69virus*	2	2	4	1	8	7	5	10
8	*Kayvirus*	92	247	20	33	10	11	6	5
9	*Silviavirus*	34	16	23	15	55	33	21	20
10	*Twortvirus*	13	12	12	3	27	9	16	18
11	*Tsarbombavirus*	9	10	9	1	24	13	22	5
12	*Phikzvirus*	39	34	55	61	56	70	68	48
13	*Cp220virus*	28	94	18	13	66	53	28	35
14	*Cp8virus*	12	15	22	11	68	18	9	18
15	*Secunda5virus*	49	9	41	14	48	11	43	21
16	*Sep1virus*	40	30	19	10	51	21	33	17
17	*Agrican357virus*	21	23	63	47	33	51	58	89
18	*Felixo1virus*	5	1	2	4	9	10	15	10
19	*Ea214virus*	5	2	5	4	13	6	13	7
20	*Se1virus*	12	24	36	19	19	18	30	21
21	*Cr3virus*	4	6	14	14	23	13	35	31
22	*Rsl2virus*	14	1	7	7	4	10	8	8
23	*Pbunavirus*	13	2	5	5	20	11	8	5
24	*Agatevirus*	13	9	5	5	9	11	7	4
25	*Bc431virus*	13	15	12	12	22	29	19	22
26	*B4virus*	13	19	11	14	10	10	18	15
27	*V1virus*	12	2	6	4	20	7	19	7
28	*Elvirus*	12	59	9	66	2	3	2	2
29	*P2virus*	11	3	5	19	2	5	4	5
30	*Cvm10virus*	10	0	4	3	1	4	1	2
31	*M12virus*	10	2	10	1	1	3	2	3
32	*Cp51virus*	9	7	8	7	12	2	9	10
33	*Tg1virus*	9	4	7	6	11	8	3	8
34	*Wphvirus*	8	11	4	2	14	4	6	11
35	*Pakpunavirus*	6	28	1	5	0	8	0	0
36	*Nit1virus*	3	0	0	1	7	17	6	10
37	*Bcep78virus*	1	2	8	10	21	34	3	42
38	*Rslunavirus*	1	2	6	3	3	8	10	13
	Total	**996**	**938**	**869**	**662**	**1482**	**906**	**902**	**876**

VPP: Virus particle purification
2 = Filtration
3A-2 = No DNase treatment, CsCl ultracentrifugation and dialysis.

Table 3. 12 Overview of dsDNA virus orders.

Sample ID	VPP	No. reads			Relative abundance (%)		
		Caudovirales	***Herpesvirales***	***Ligamenvirales***	***Caudovirales***	***Herpesvirales***	***Ligamenvirales***
36	2	6,134	415	40	93.1	6.29	0.61
	3A-2	8,503	340	14	96.0	3.8	0.2
37	2	6,821	562	30	91.8	7.58	0.62
	3A-2	6,418	396	10	94	5.8	0.2
38	2	7,451	464	133	92.6	5.76	1.64
	3A-2	9,020	584	68	93.35	6.0	0.62
39	2	6,164	491	23	92.3	7.35	0.35
	3A-2	8,507	574	18	93.49	6.3	0.21
				Average	**93.33**	**6.11**	**0.59**

VPP: Virus particle purification
2 = Filtration
3A-2 = No DNase treatment, CsCl ultracentrifugation and dialysis.

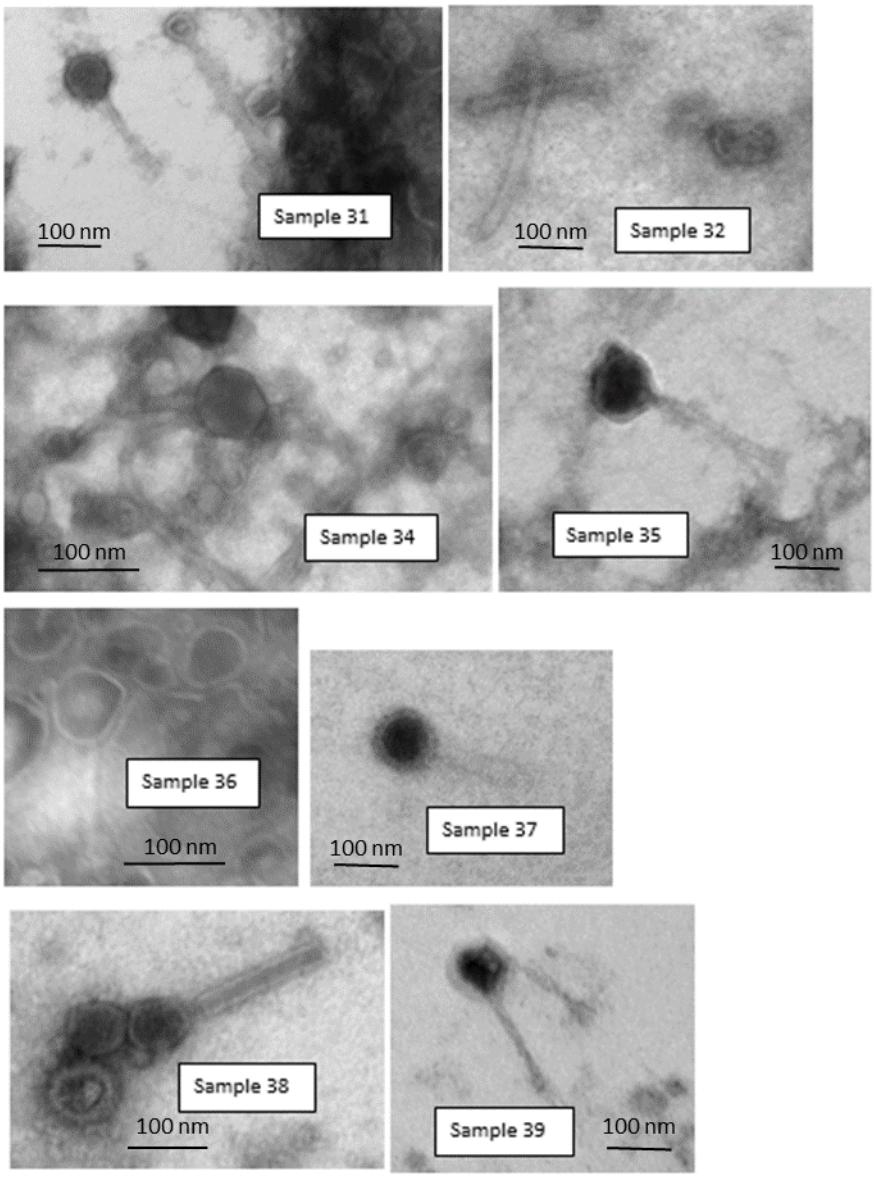

Figure 3. 1 Morphotypes of selected Myoviridae virus like particles from samples processed using 1-Butanol.

Morphology of *Myoviridae*-like particles: tailed phages with rigid tails and a neck.

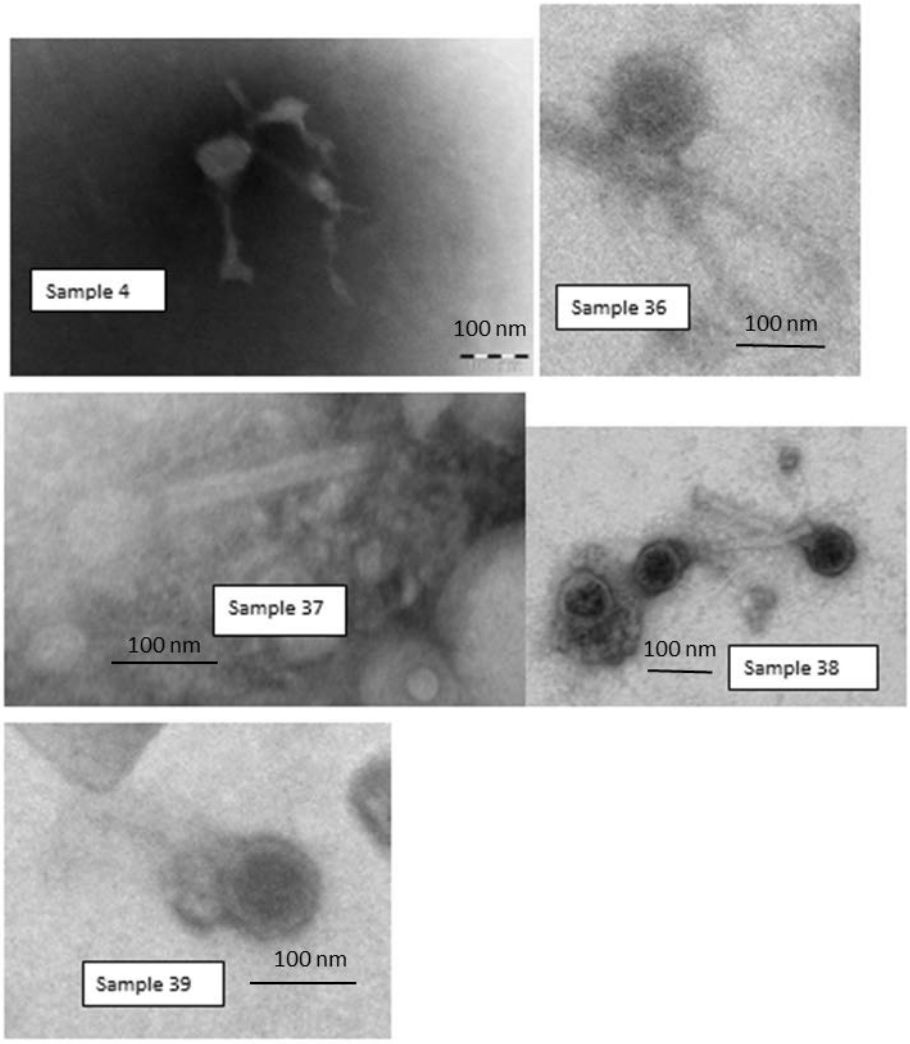

Figure 3. 2 Morphotypes of selected Myoviridae virus like particles from samples processed using CsCl.

Morphology of *Myoviridae*-like particles: tailed phages with rigid tails and a neck.

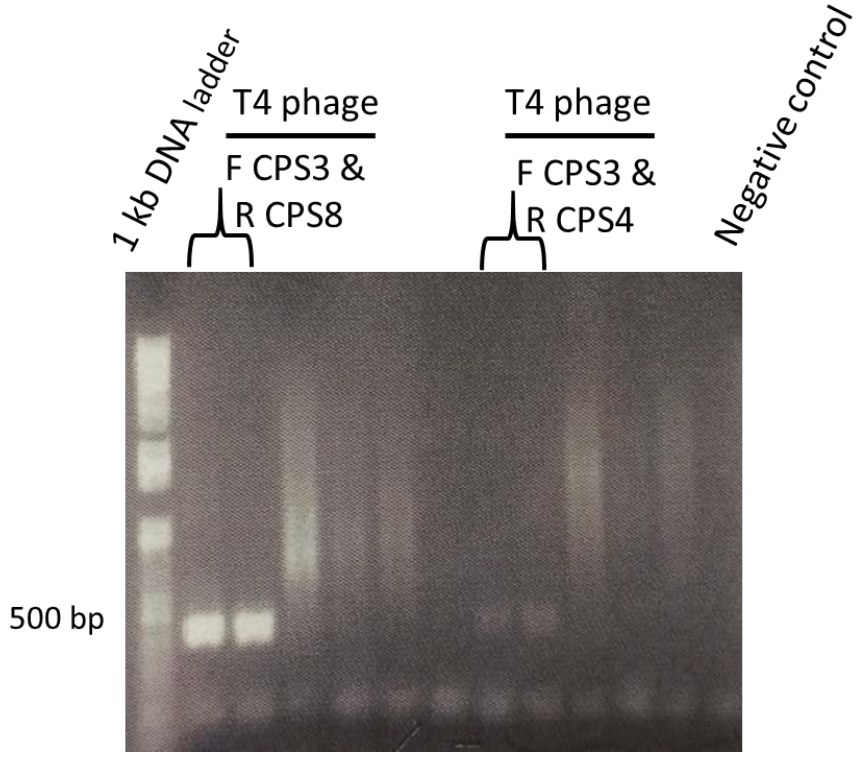

Figure 3. 3 Evaluation of primer sets for PCR amplification of Gp-like gene.

Sybr Green-stained agarose gel of DNA fragments produced by PCR amplification of Gp-like gene from *Escherichia coli* T4 phage ATCC®11303-B4™. T4 phage (*Myoviridae*) amplified with two sets of primers. Note the 500-bp from the 1 Kb plus DNA ladder (Invitrogen). A band of 480 bp was amplified by both sets of primers. The negative control was water.

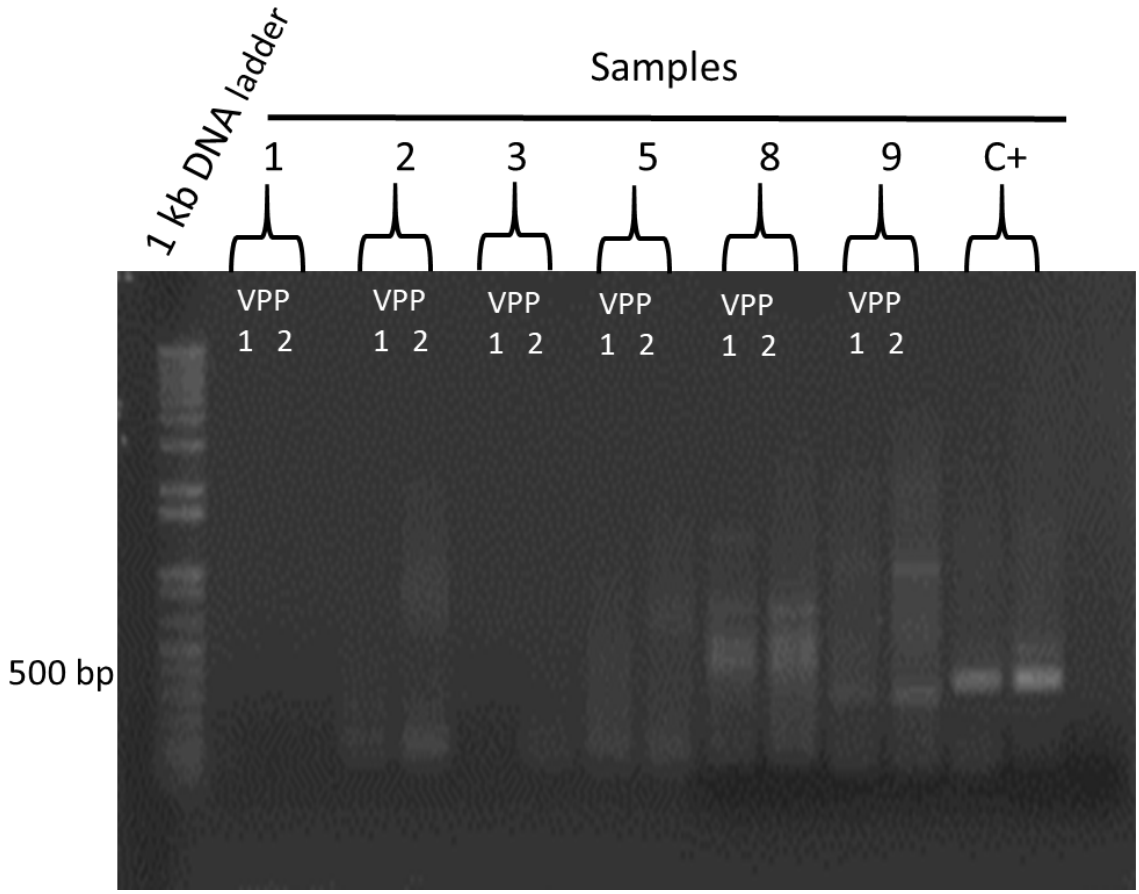

Figure 3. 4 PCR amplification of Gp-like gene.

Sybr Green-stained agarose gel of DNA fragments produced by PCR amplification of Gp-like gene from six fecal samples. Note the 500-bp from 1 Kb plus DNA ladder. Each fecal sample was subjected to two VPP procedures prior to Gp-like gene amplification and various size bands amplified (Table 3.1). Positive control (C+) was *Escherichia coli* T4 phage ATCC®11303-B4™ and negative control was water.

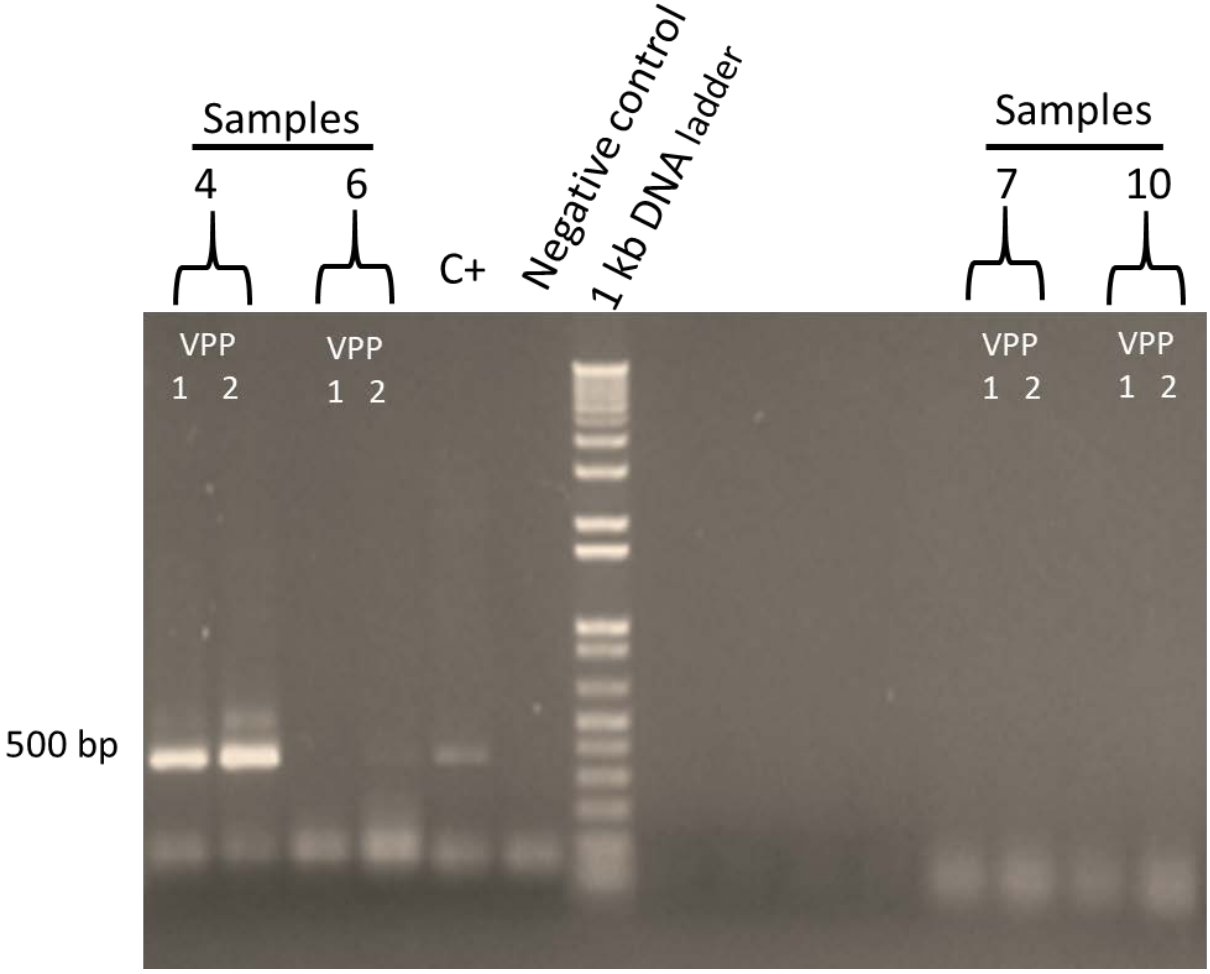

Figure 3. 5 PCR amplification of Gp-like gene.

Sybr Green-stained agarose gel of DNA fragments produced by PCR amplification of a Gp-like gene from four fecal samples. Note the 500-bp from 1 Kb plus DNA ladder. Each fecal sample was subjected to two VPP procedures prior to Gp-like gene amplification and various size bands amplified (Table 3.1). Positive control (C+) was *Escherichia coli* T4 phage ATCC®11303-B4™ and the negative control was water.

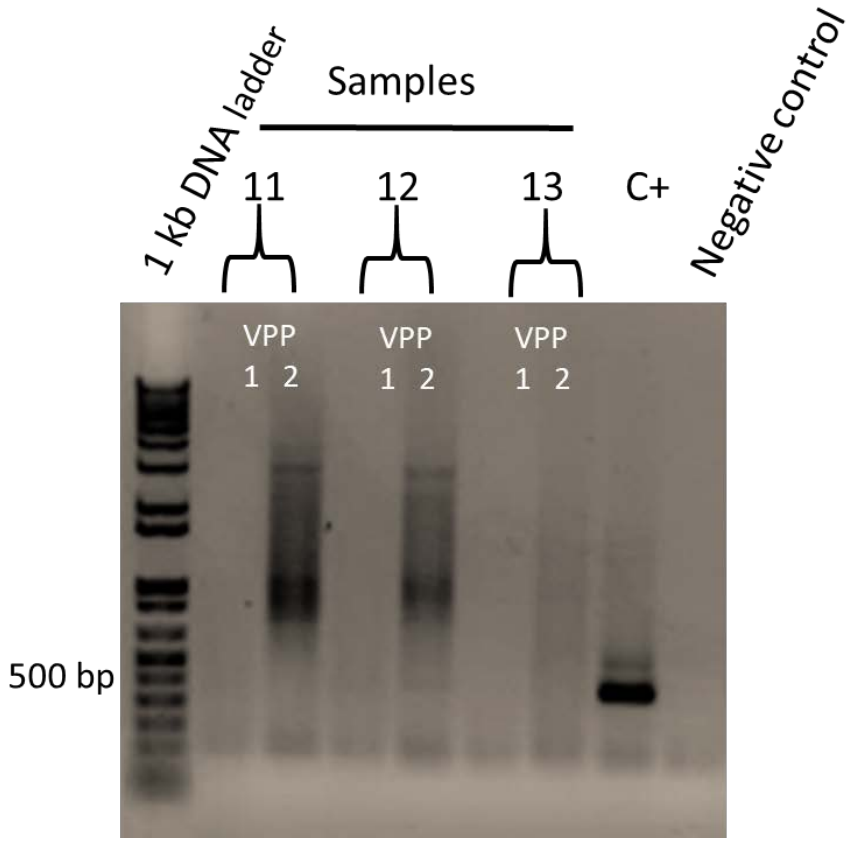

Figure 3. 6 PCR amplification of gene Gp-like gene.

Sybr Green-stained agarose gel with DNA fragments produced by PCR amplification of a Gp- like gene from three fecal samples. Note the 500-bp from 1 Kb plus DNA ladder. Each fecal sample was subjected to two VPP procedures prior to Gp-like gene amplification and variable size bands were amplified (Table 3.1). Positive control (C+) was *Escherichia coli* T4 phage ATCC®11303-B4™ and negative control was water.

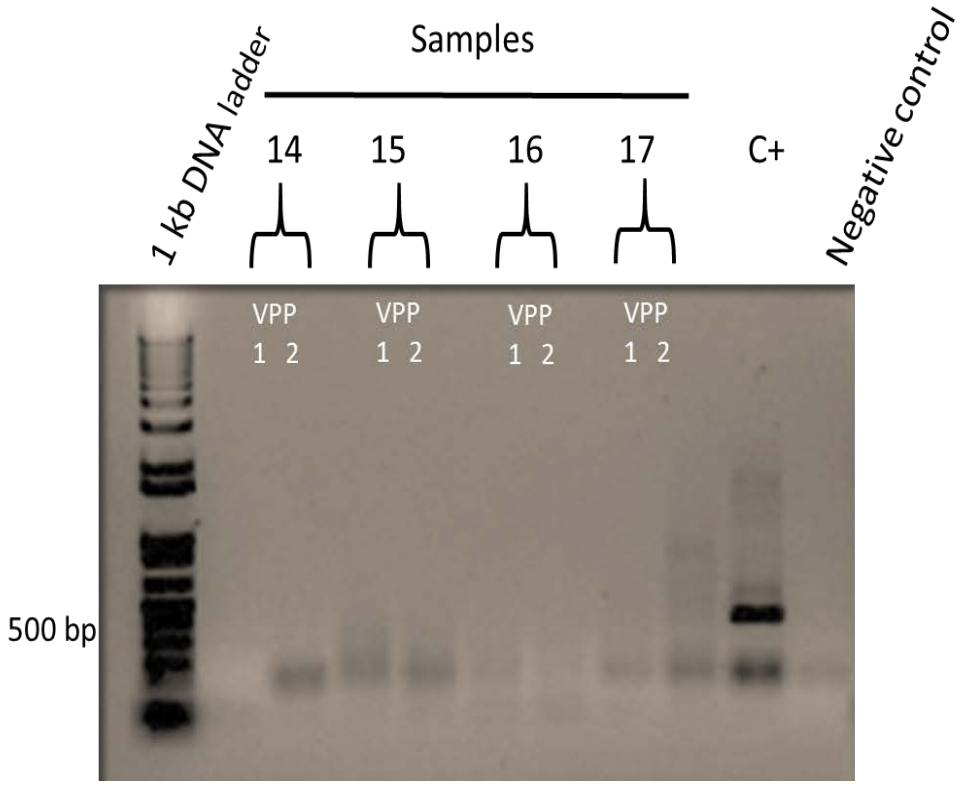

Figure 3. 7 PCR amplification of gene Gp-like gene.

A Sybr Green-stained agarose gel with DNA fragments produced by PCR amplification of Gp-like gene from four fecal samples. Note the 500-bp from 1 Kb plus DNA ladder. Each fecal sample was subjected to two VPP procedures prior to Gp-like gene amplification and variable size bands amplified (Table 3.1). Positive control (C+) was *Escherichia coli* T4 phage ATCC®11303-B4™ and negative control was water.

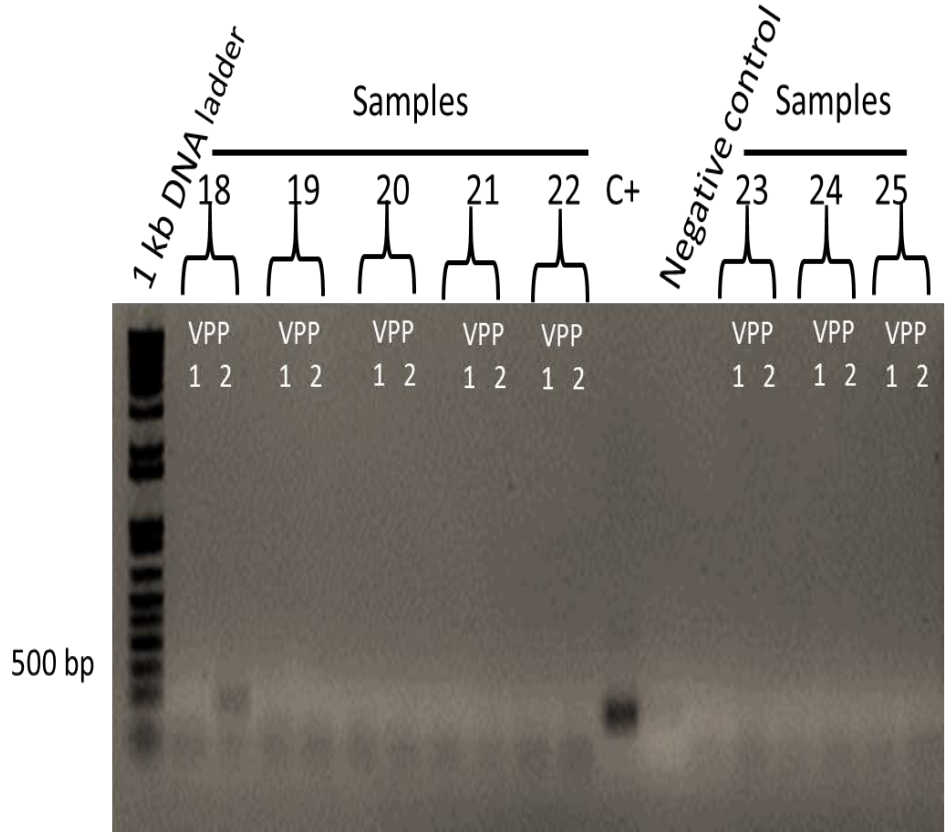

Figure 3. 8 PCR amplification of gene Gp-like gene.

Sybr Green-stained agarose gel of DNA fragments produced by PCR amplification of Gp-like gene from eight fecal samples. Note the 500-bp from 1 Kb plus DNA ladder. Each fecal sample was subjected to two VPP procedures prior to Gp-like gene amplification and variable size bands were amplified (Table 3.1). Positive control (C+) was *Escherichia coli* T4 phage ATCC®11303-B4™ and the negative control was water.

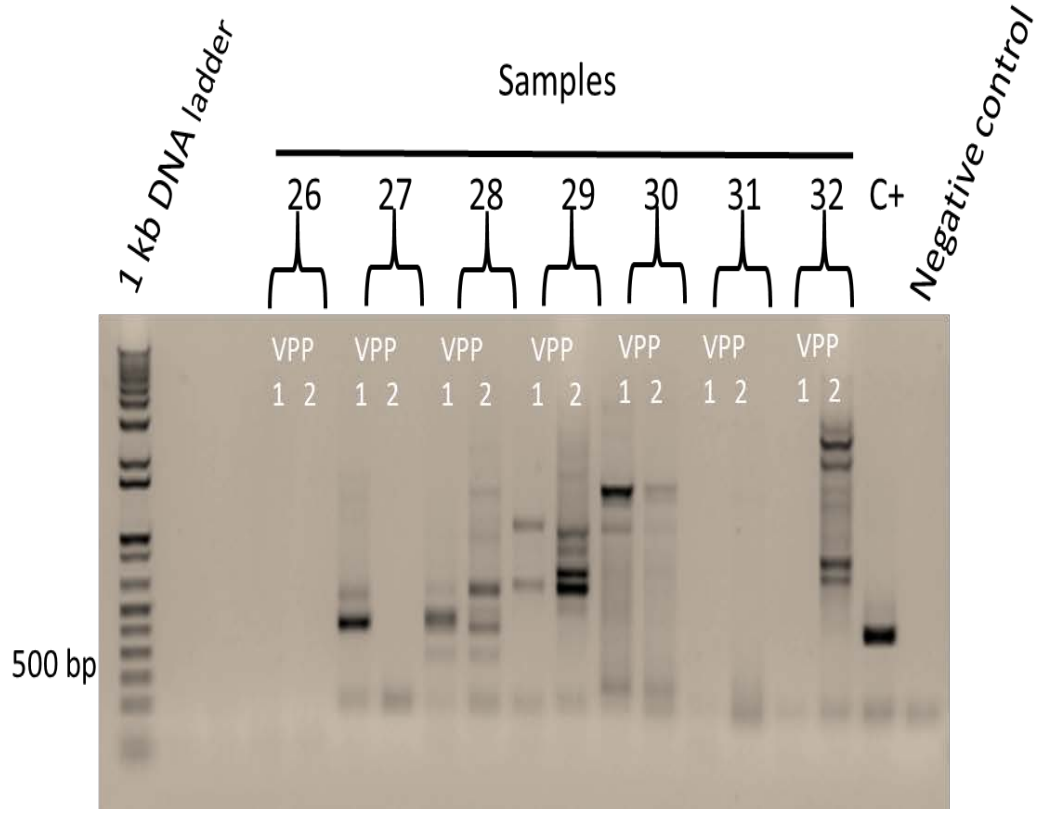

Figure 3. 9 PCR amplification of Gp-like gene.

Sybr Green-stained agarose gel of DNA fragments produced by PCR amplification of gene Gp-like gene from seven fecal samples. Note the 500-bp from 1 Kb plus DNA ladder. Each fecal sample was subjected to two VPPs prior to Gp-like gene amplification and variable size bands were amplified (Table 3.1). Positive control (C+) was *Escherichia coli* T4 phage ATCC®11303-B4™ and negative control was water.

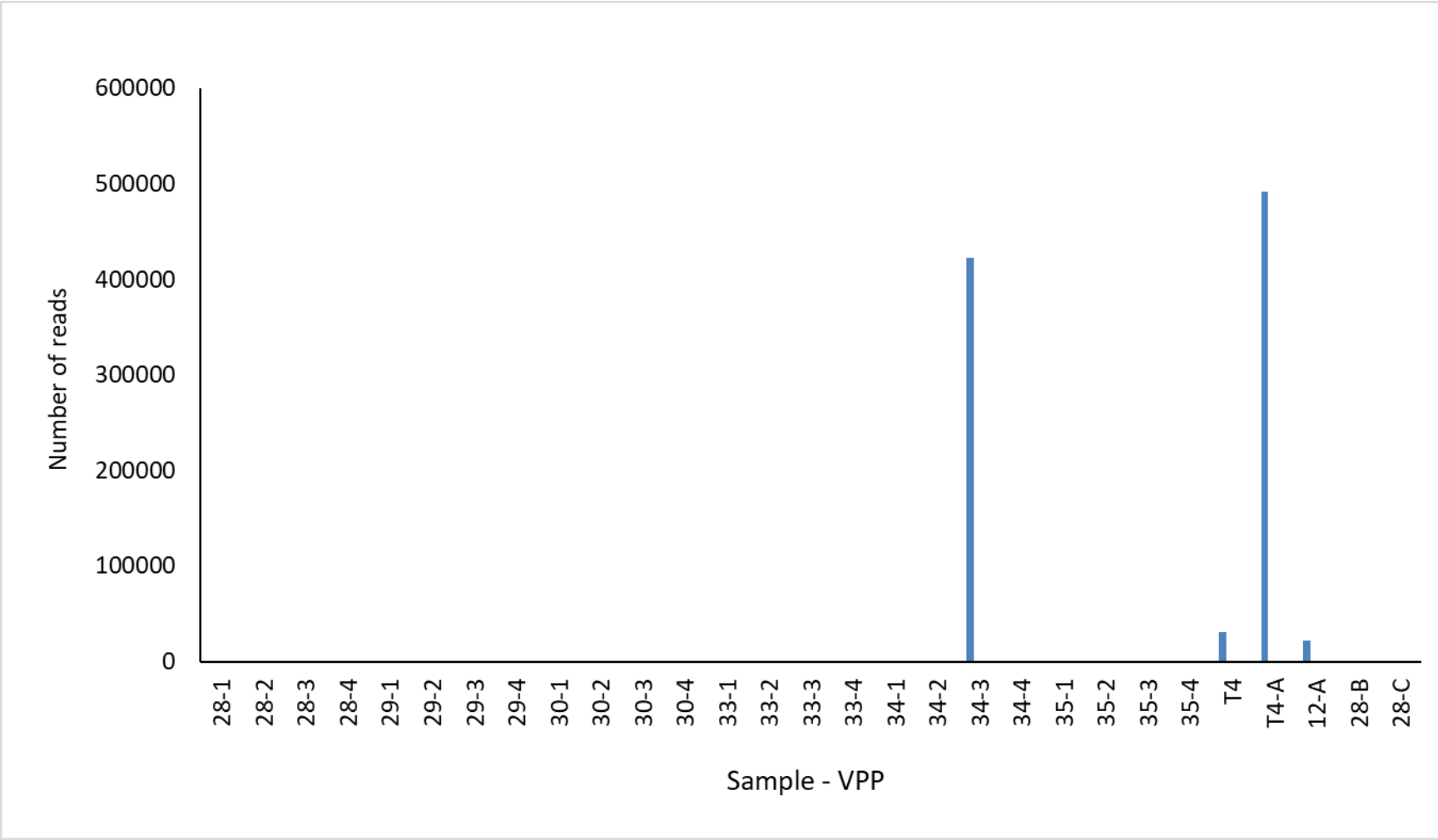

Figure 3. 10 Number of virus reads derived with MiSeq sequencing.

Total number of reads with sequence identities to domain viruses of six fecal samples using four VPP procedures (1-4). All PCR derived amplicons from Gp-like gene from T4 phage (Positive control, labelled T4 and T4-A), Samples 12 (labelled 12-A) and 28 (labelled 28-B and 28-C) are shown (Table 3.3).

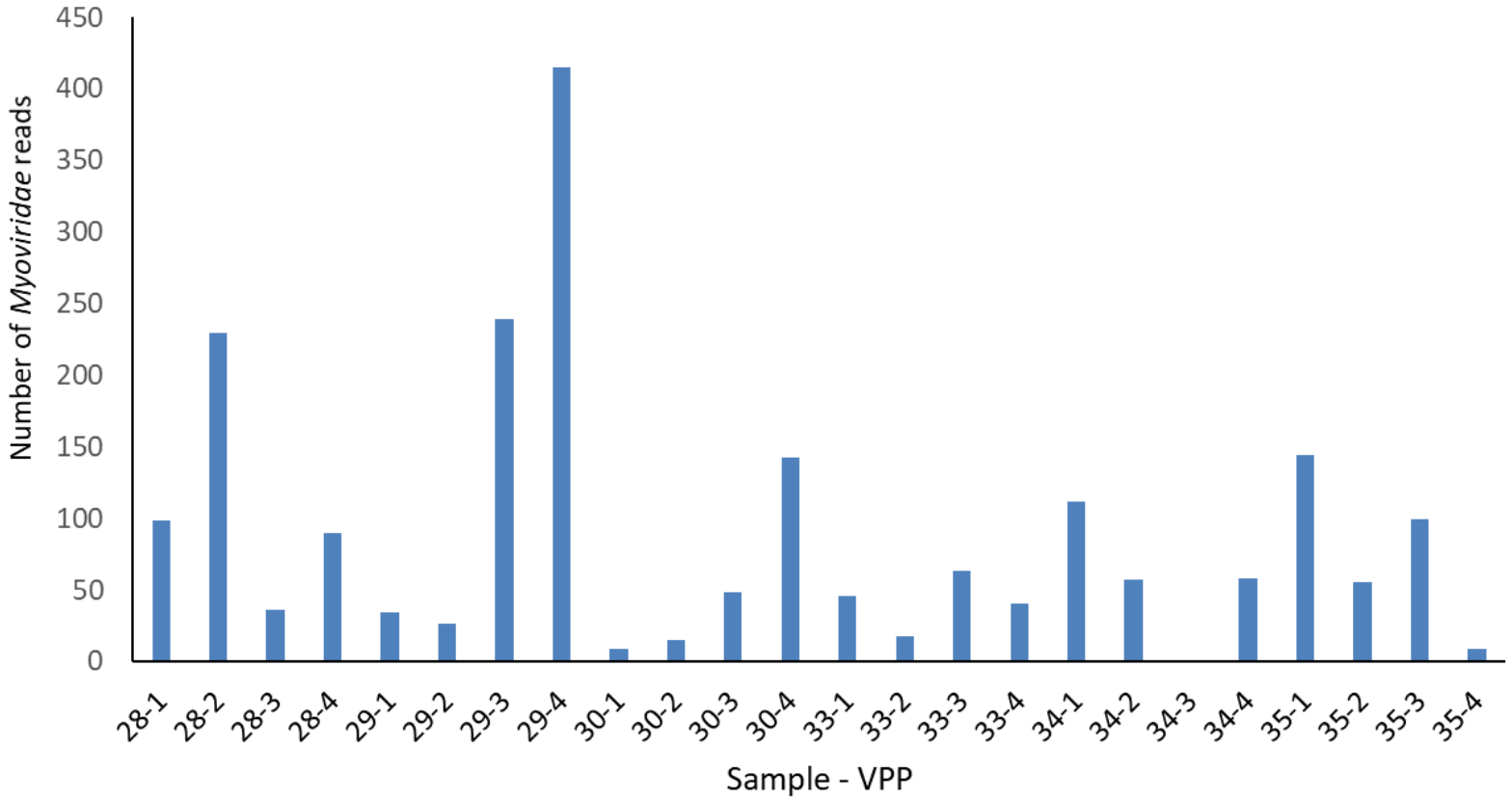

Figure 3. 11 Distribution of viral reads through fecal samples without outliers (Sample 34-3) and amplicons.

Total number of reads with sequence identities to *Myoviridae* in relation to all virus reads of six fecal samples with four VPP procedures (1-4) per sample (Table 3.3).

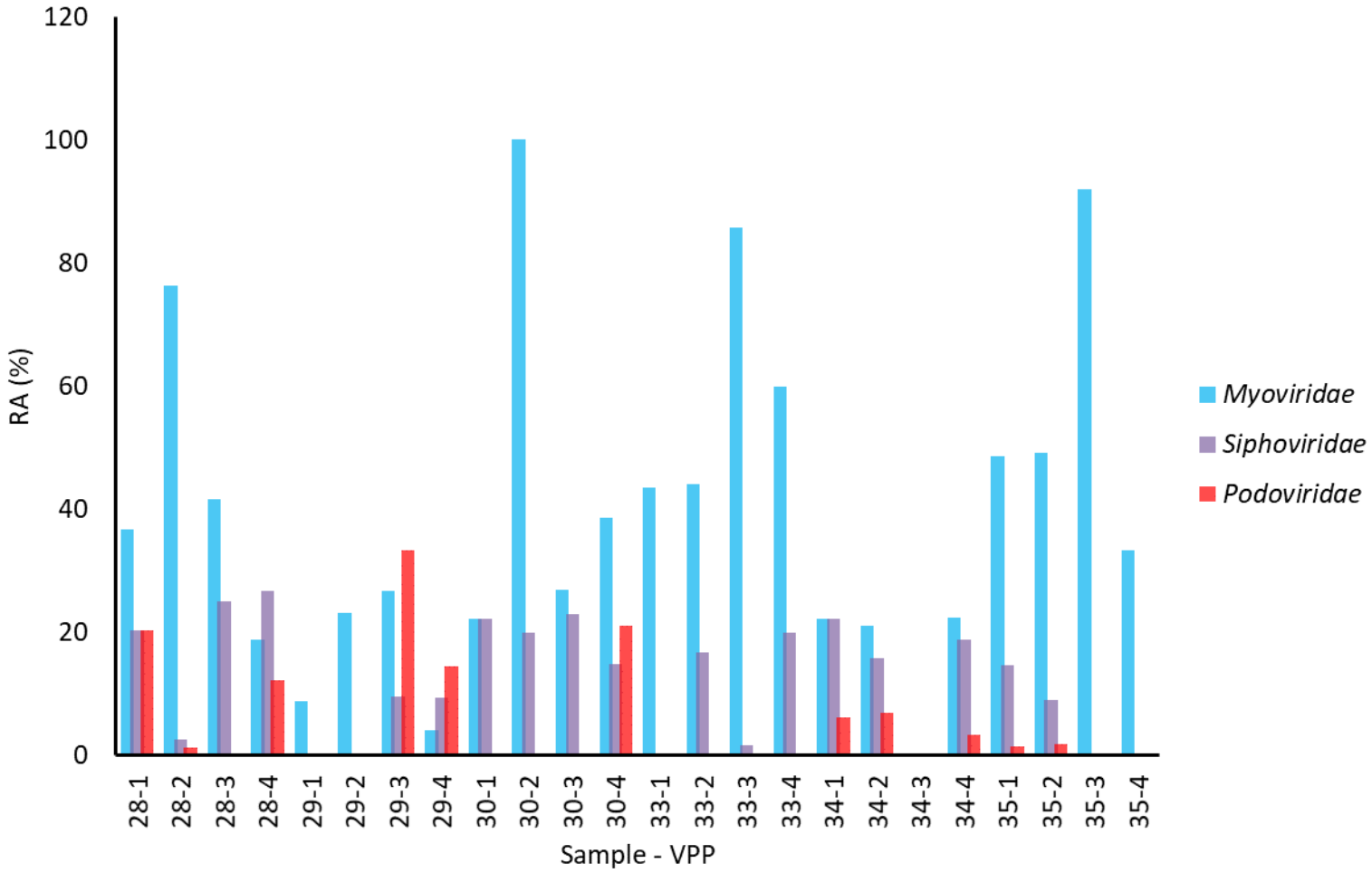

Figure 3. 12 Relative abundance of *Myoviridae*, *Siphoviridae* and *Podoviridae*.

Percentage of reads with sequence identities to *Myoviridae*, *Siphoviridae* and *Podoviridae* families in relation to total *Caudovirales* reads (Table 3.4).

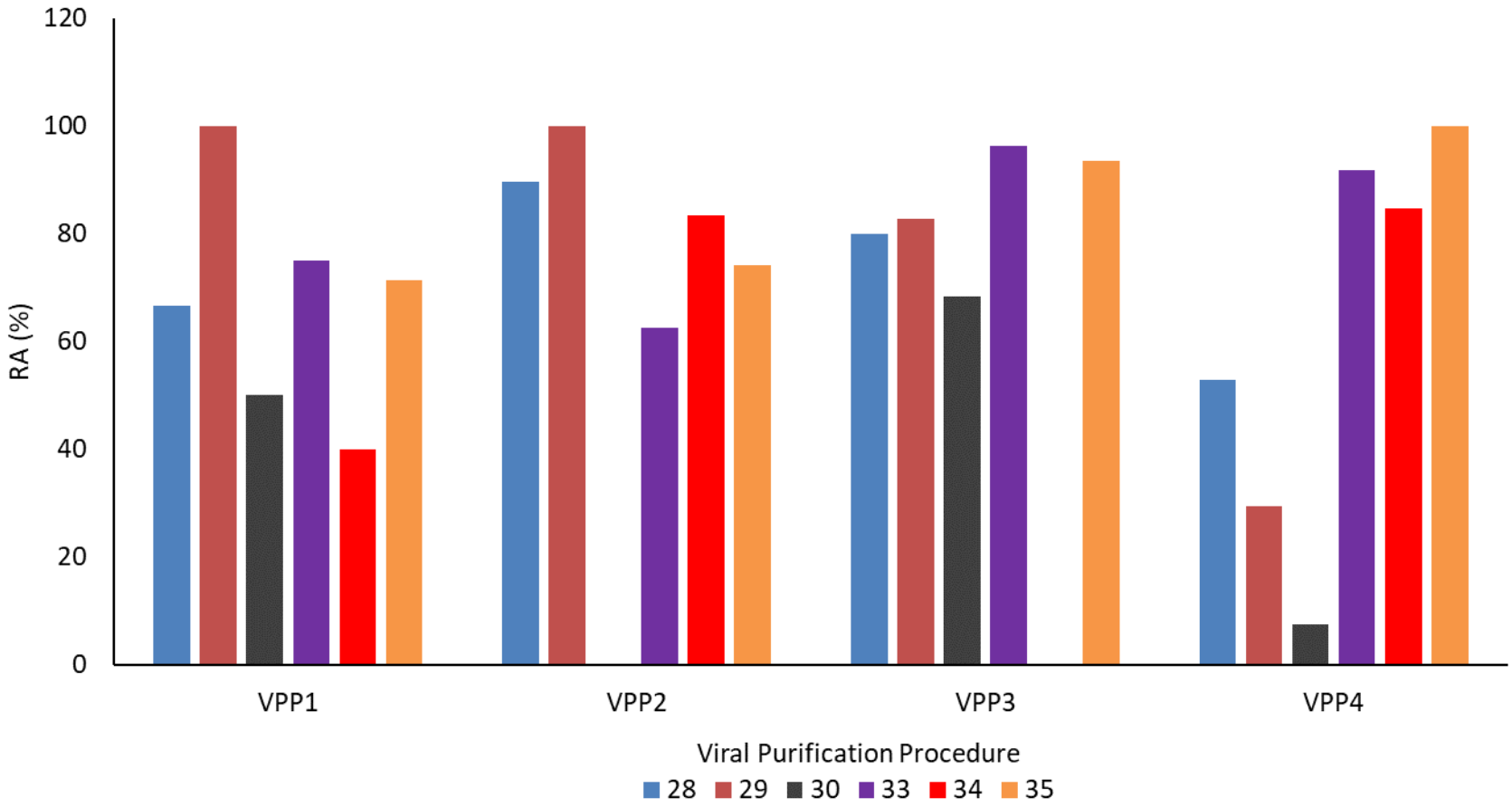

Figure 3. 13 Relative abundance of *Myoviridae* in MiSeq run in samples comparing four procedures.

Percentage of reads of *Myoviridae* family in relation to *Caudovirales* reads (Tables 3.3 and 3.4).
Each color represents one fecal sample.

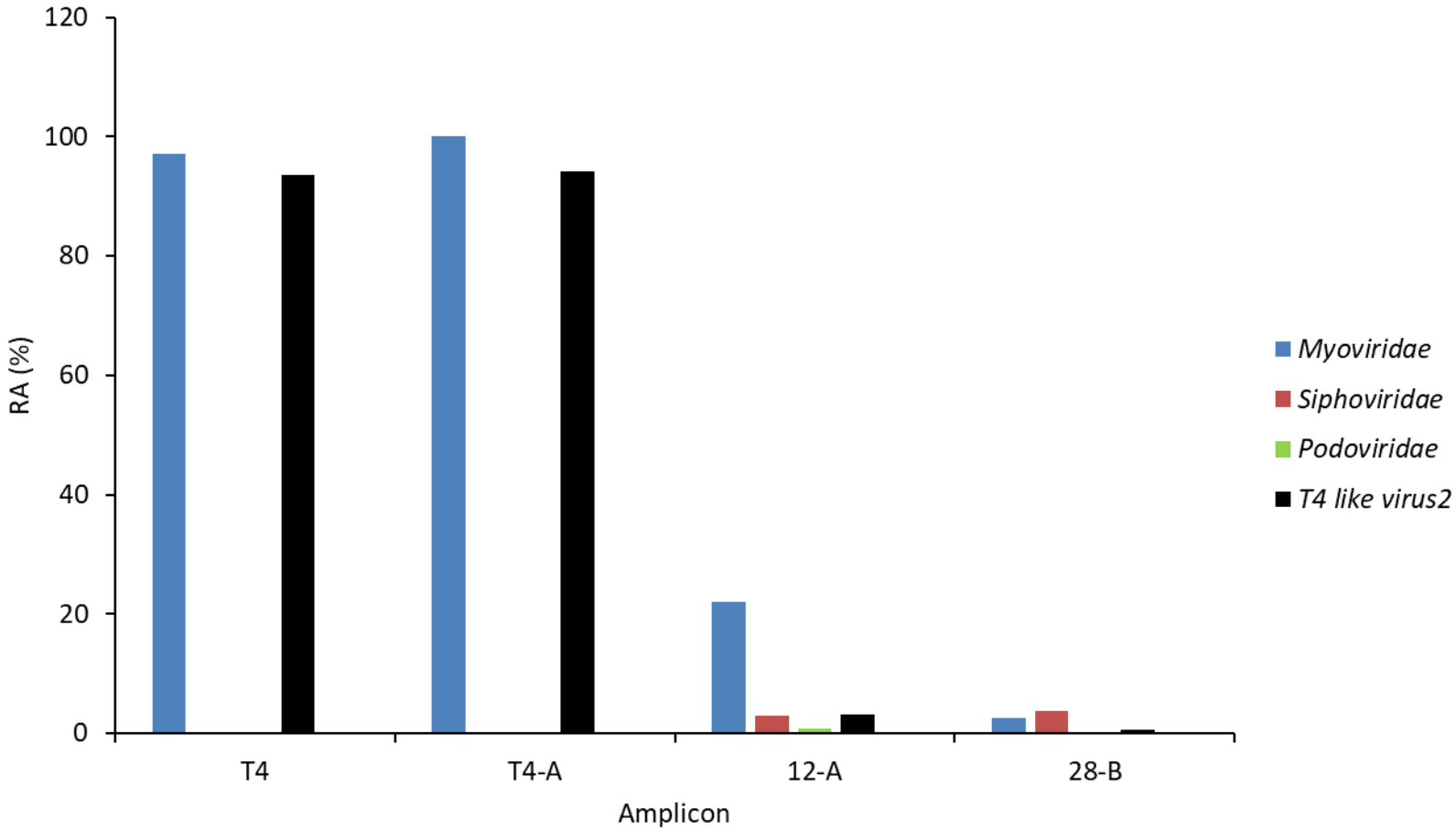

Figure 3. 14 Relative abundance of families of the order *Caudovirales* and genus *"T4 like virus"* from amplicons from gene Gp-like gene.

Percentage of reads with sequence identities to the three families in *Caudovirales* order and *"T4 like virus"* as the main genus within *Myoviridae* family. Note that in positive controls (T4 and T4A), no viral reads to *Siphoviridae* and *Podoviridae* were detected (Table 3.5 and Table 3.7).

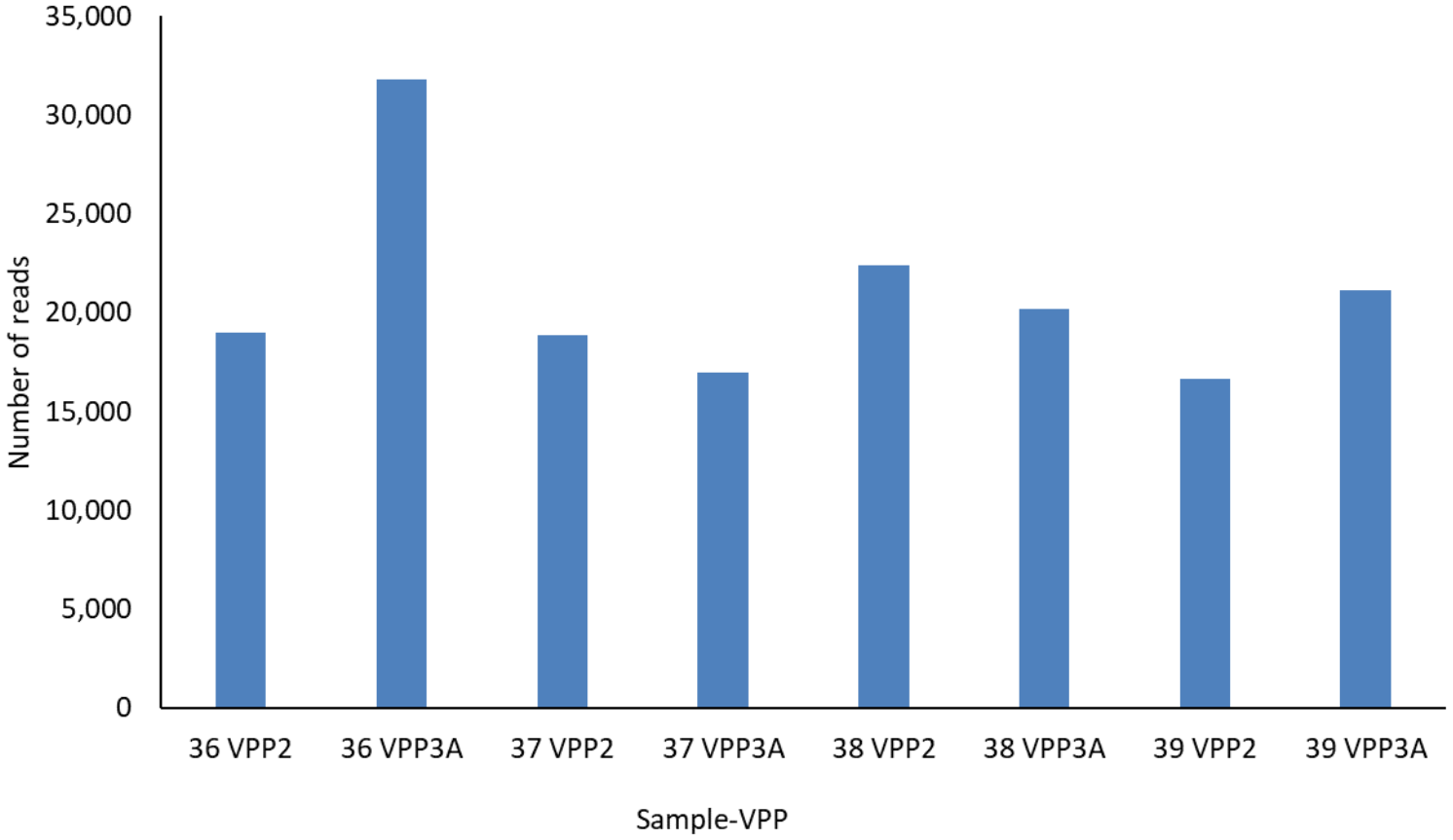

Figure 3. 15 Number of virus reads derived using NextSeq.

Total number of reads with sequences identities to the domain viruses in four fecal samples with two VPP procedures (2 and 3A) per sample (Table 3.8).

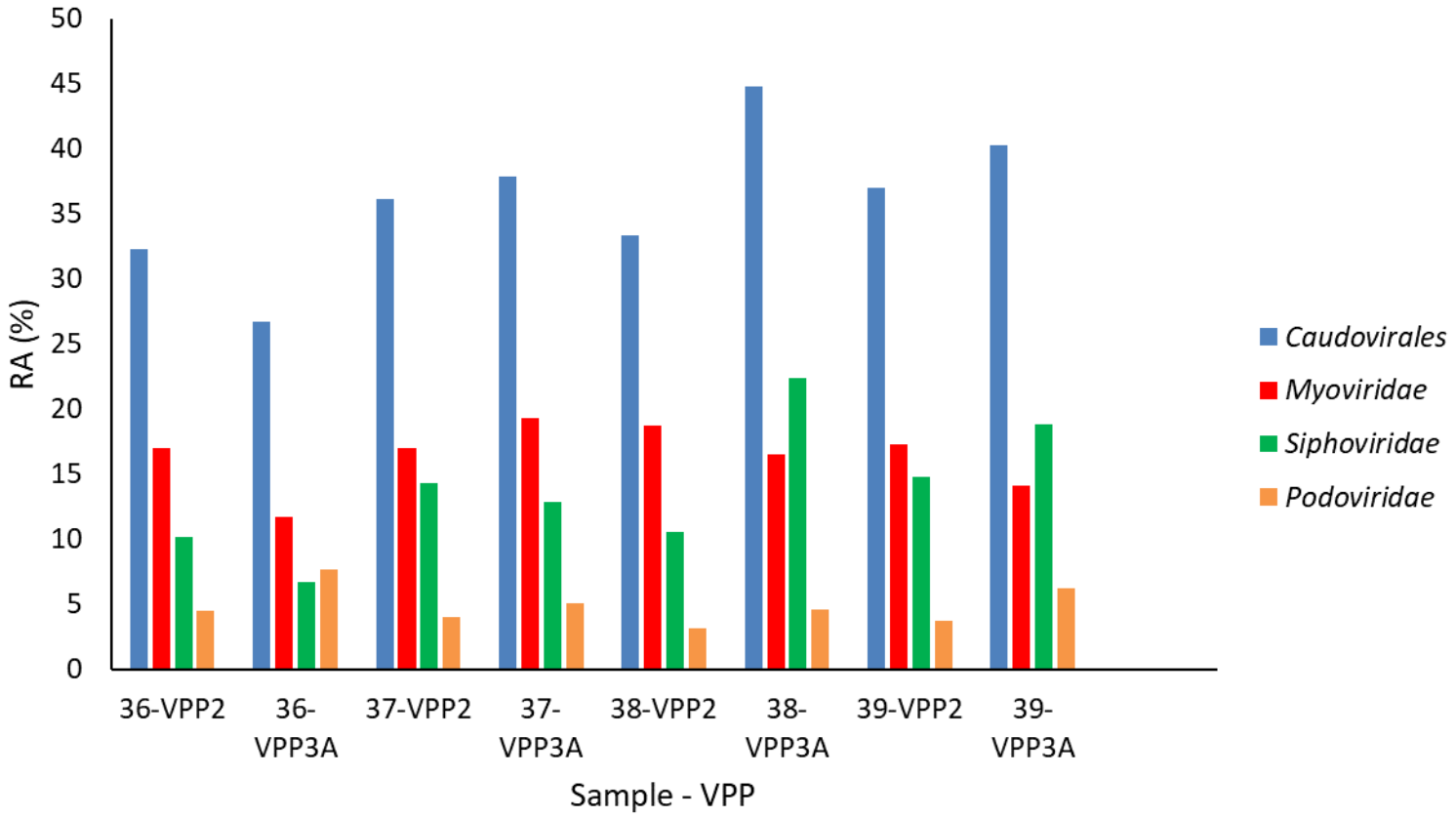

Figure 3. 16 Relative abundance (across all fecal samples) of *Caudovirales*, *Myoviridae*, *Siphoviridae* and *Podoviridae*, based on metagenomic analysis using NextSeq.

Percentage of reads with sequence identities to *Caudovirales*, namely families *Myoviridae*, *Siphoviridae* and *Podoviridae,* in relation to total virus reads (Table 3.9).

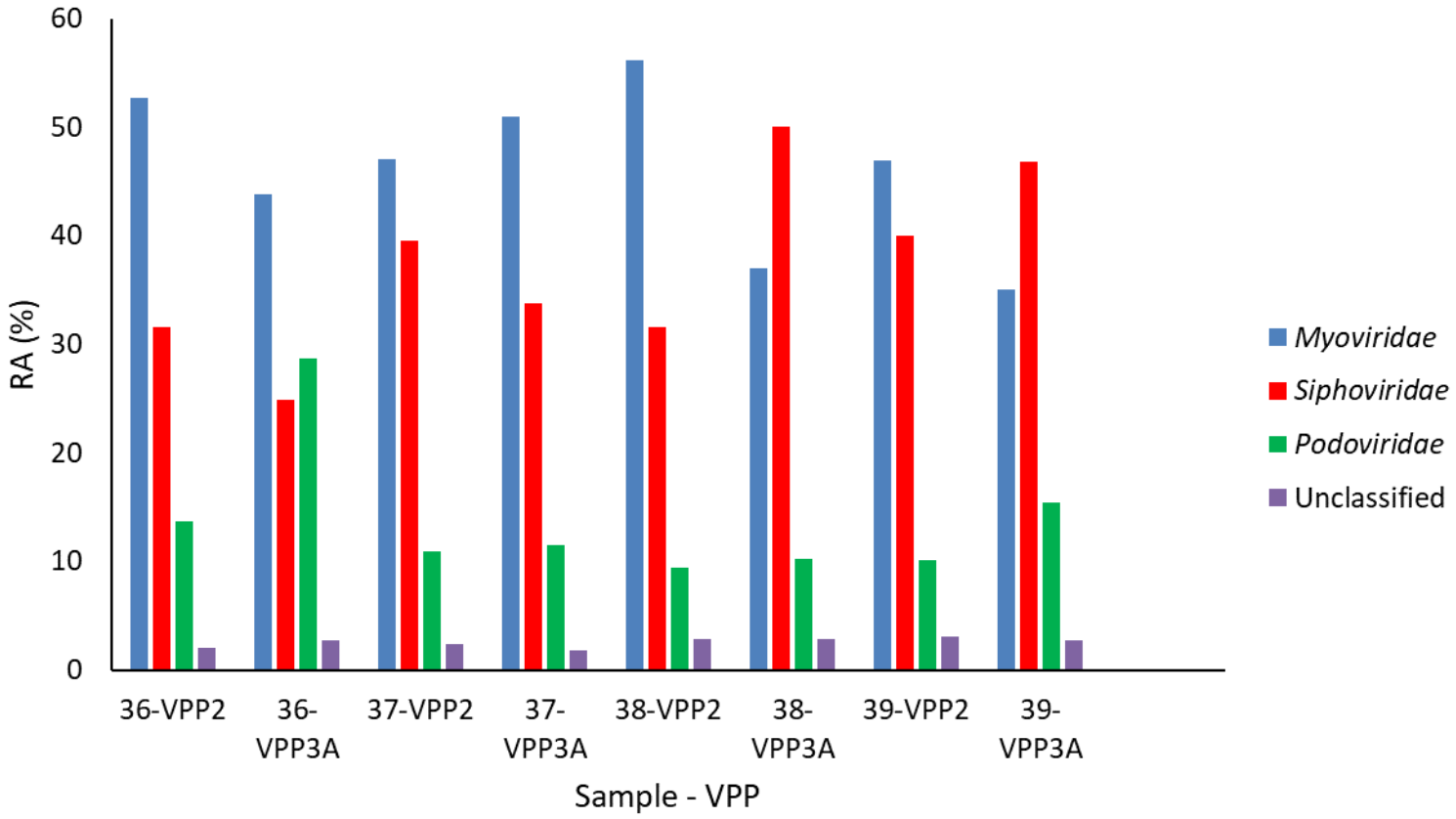

Figure 3. 17 Relative abundance (across all fecal samples) of *Myoviridae*, *Siphoviridae* and *Podoviridae,* based on metagenomic analysis using NextSeq.

Percentage of reads with sequence identities to *Myoviridae*, *Siphoviridae* and *Podoviridae,* in relation to total *Caudovirales* reads (Table 3.9).

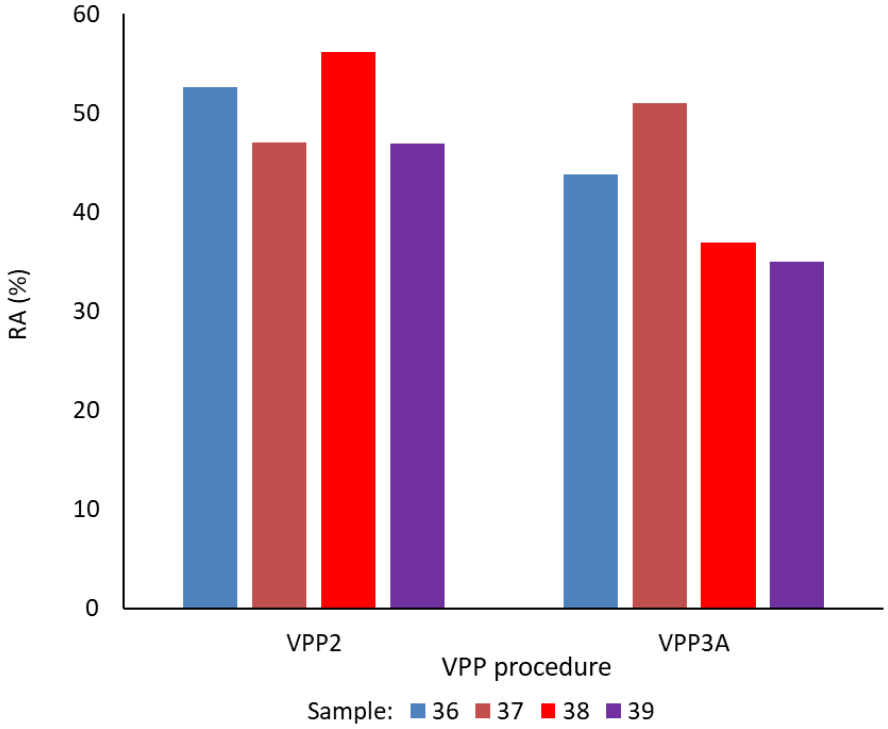

Figure 3. 18 Relative abundance of *Myoviridae* in VPP2 and 3A.

Percentage of reads with sequence identities to *Myoviridae* family, grouped by VPP procedure, in relation to total *Caudovirales* reads (Table 3.9). Each color represents one fecal sample.

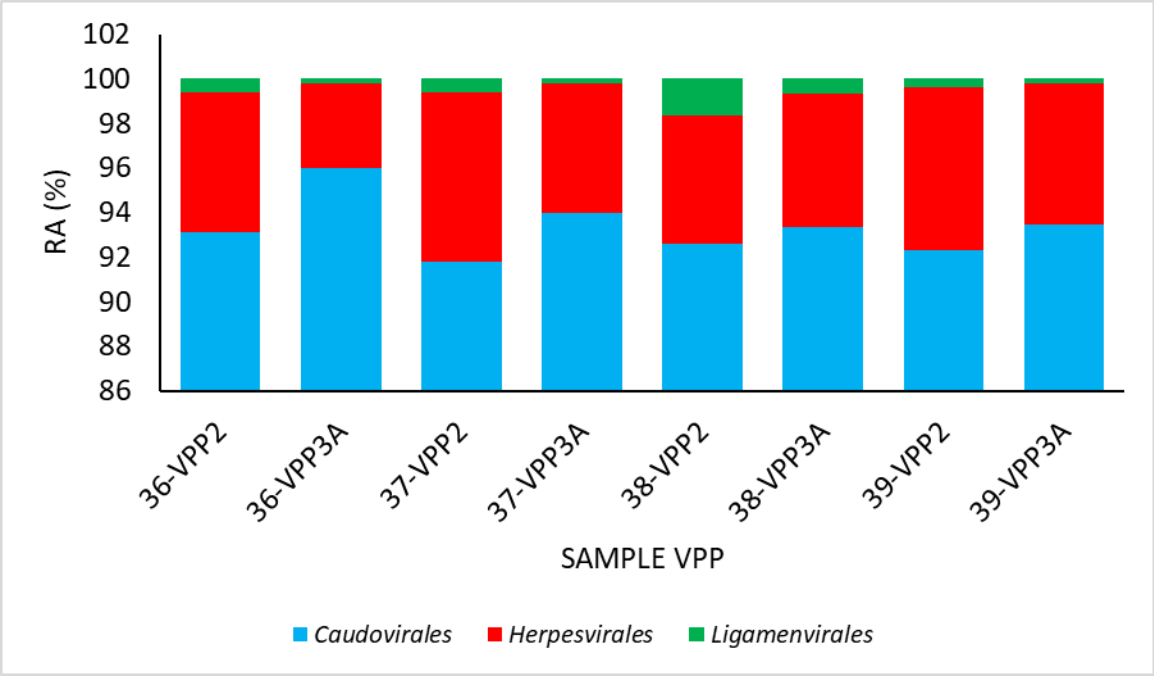

Figure 3. 19 Sequence reads for various viral orders, classified based on dsDNA.

Total number of reads, classified to orders of dsDNA viruses, based on analysis of four fecal samples with two VPP.

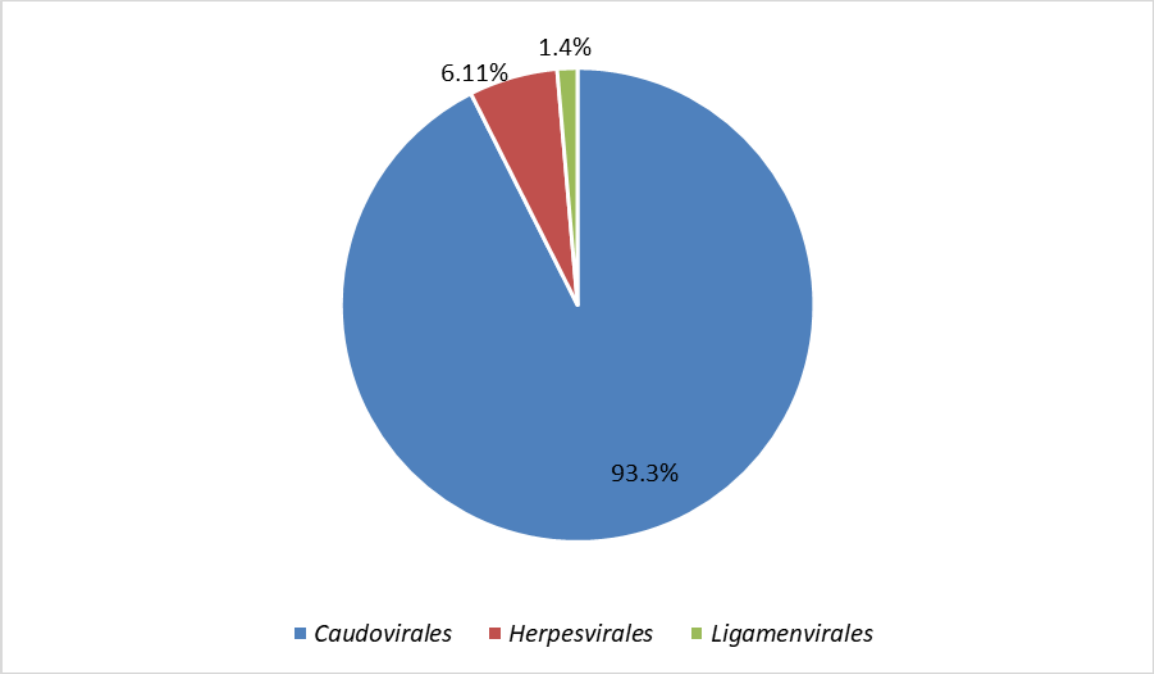

Figure 3. 20 Relative abundance (classified by viral orders) of assigned viral sequence reads in samples of cattle feces.

Average percentage of reads with sequence identities to all orders of dsDNA viruses present in four fecal samples.

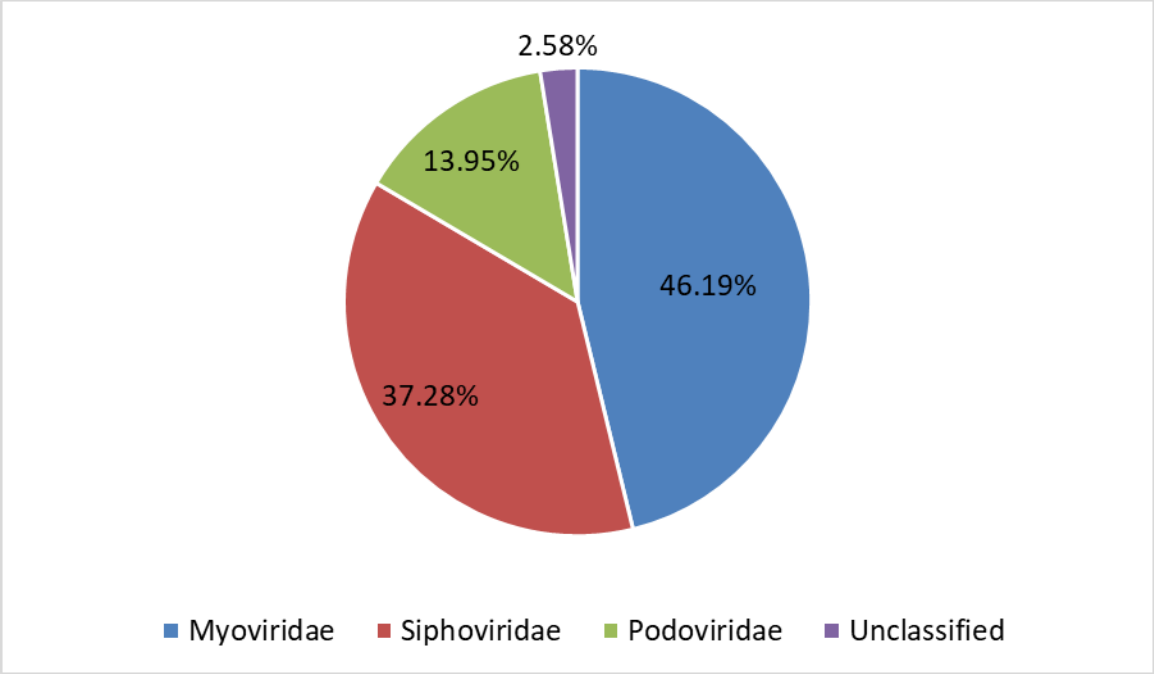

Figure 3. 21 Relative abundance of families within *Caudovirales* in cattle feces.

Average percentage of reads of the three viral families in *Caudovirales,* plus unclassified reads (Table 3.9).

Chapter 4: Discussion

In the current study, metagenomic analysis was a very useful methodology to investigate relative abundance (RA) and diversity of phages in cattle feces. However, EM and PCR were of less value to determine diversity or abundance of *Myoviridae* as part of the microbiome. Based on our final metagenomic analysis using the Illumina NextSeq platform, *Myoviridae* RA and diversity were determined, using fecal samples derived from two beef and two dairy cattle. Although metagenomic analysis is considered the best way to characterize RA and diversity of viruses (120), it is still a relatively expensive and time-consuming procedure.

Diversity of morphotypes in *Myoviridae* was demonstrated by EM. The two approaches used (1-butanol and CsCl) efficiently concentrated phages of the family *Myoviridae.* However, 1-butanol inactivates viruses; therefore, it was only used as a confirmatory test to detect and identify phages present in feces. On the contrary, CsCl maintains phage infectivity; therefore this approach was chosen for DNA extraction and other downstream analyses. Electron microscopy was not suitable to estimate RA, as it cannot be used to determine number of species and their abundance in a sample. Furthermore, EM can only be used to identify morphotypes to the family level, but not to the species level (47). We were unable to measure size of heads and tails, which would have been useful to classify viruses into genera. Although, 1-butanol is the optimal procedure for visualizing phages for classification into order and families (47), more detailed taxonomical classification depends on determining genetic sequence or other molecular data (47).

In the current study, the use of PCR with degenerate primers (targeting Gp20 and Gp23, as described for cyanophages (65), had limited success to characterize diversity of *Myoviridae* in cattle feces. By design, degenerate primers lack specificity, as they can easily adhere to similar regions. In the present study, primers designed for the phage T4 Gp-23 gene did not produce an amplification product, whereas those that produced an amplification product were aligned to the Gp10 gene, according to the genetic map from Genbank. Sanger sequencing of bands from phage T4 amplification products corresponded to Gp22 (partially) and Gp23 genes.

PCR products from positive fecal samples closely matched other published results, with values of 480 bp (118), 600 bp, 640 bp (117), 800 bp and 2 kb (115), as described for various cyanophages that were derived from mainly aquatic habitats (65). However, similar phage sequences were not reported from mammalian fecal samples. Sequences of non-significant match clones (n=16/21) from fecal samples represented only nucleotide similarity to cyanophages < 50% with the Genbank database. As degenerate primers have low specificity, it is certainly possible that not all sequences identified were actually viral sequences. To better characterize the derived amplicons, three of the amplicons were sequenced using MiSeq technology. The presence of *Myoviridae* (T4-like viruses) was confirmed in two fecal samples, providing evidence that these phages were present in bovine fecal samples (Table 3.7).

A high percentage of the PCR on fecal samples (60%) did not result in an amplification product may have been caused by inhibitors in these samples, as experienced in other laboratories working with bovine fecal samples (personal communication, Dr. Bachofen). Therefore, failure of amplification during PCR may not be exclusively a primer sensitivity problem, but may also due to inhibition (121).

Composition of the bovine fecal virome requires the taxonomical classification of raw reads. We used two classifying programs, namely Kraken and Centrifuge, which assign taxonomical labels and determines abundance of microbial genomes for metagenomic analyses (122, 123). Kraken was chosen as the standard, for our first metagenomic analysis (MiSeq), as it has superior accuracy, speed compared to several other programs and is accessible for free online (124). Kraken has higher specificity than Centrifuge (99.0 versus 96.9%, respectively), although Centrifuge has higher sensitivity than Kraken (76.9 versus 73.9%) (123). Kraken was published in 2014 (124) and has not been updated, whereas Centrifuge was published in 2016, with updated genomes (125). Although both classifiers are based on K-mers algorithms, Kraken uses a k-mer of 31 and Centrifuge uses two, namely 22 and 53 (124). The smaller the k-mer, the more false positives expected. Conversely, the larger the k-mer, the higher specificity, as k-mers are used as signatures in a microbial genome. Centrifuge yields reads that are unclassified at the order

level, whereas that does not occur with Kraken. Based on all metagenomic classifiers used in this study, *Caudovirales* and its family of *Myoviridae* were the most predominant phages in the cattle fecal microbiome.

Unfortunately, our MiSeq run lacked standardized methods of preparation, extraction and concentration of DNA from each VPP prior to sequence analysis. Samples were not normalized, specifically during library preparation. Consequently, there was substantial variation in reads among samples and VPP. Another limitation of our MiSeq run was inclusion of amplicons and fecal samples. This resulted in evident cross-contamination between pure T4 phage amplicons and fecal Sample 34 (VPP 3A preparation). Another important limitation was disparity in age of cattle sampled, ranging from 1 to 7 months (dairy and steer calves, respectively).

One noteworthy advantage of deep sequencing of amplicons was the opportunity to confirm that amplicons from Samples 12 and 28 (two replicates) contained Gp-like genes, as the classifier kraken identified them as *T4-like virus* and *E.coli* T4 phage. It was evident that control phage T4 amplicons contained ~96% of *E.coli* T4 phage, whereas fecal samples contained other members of *Caudovirales*.

Based on results from MiSeq (Kraken analysis), in the NextSeq analysis, many of these obstacles were corrected (normalization of the DNA samples prior to deep sequencing and comparable fecal samples from 15 month old dairy and beef cattle). Based on NextSeq run and Centrifuge classifier, our findings indicated that *Myoviridae* were more abundant in feces than what was described in literature for the rumen (28). Their presence seems to be associated with the bacterial component of the cattle fecal microbiome, where *Firmicutes* is the leading phylum (9). Other mammalian (non-ruminant) fecal viromes are more similar to the rumen virome; in that regard, the proportion of *Caudovirales* in the bovine rumen is comparable to that in the fecal virome of pigs (126-128), horses (127, 129) and humans (130).

On the basis of dsDNA viral reads classified using Centrifuge, bacteriophages from the Order *Caudovirales* constituted 93.33% of the overall number of ds DNA viruses in fecal samples sequenced by NextSeq (Figure 3.20). In a study characterizing the rumen virome (28),

Caudovirales represented approximately 78% of total dsDNA; the most abundant phage families were *Siphoviridae*, followed by *Myoviridae* This contrasts with the current study of the bovine fecal virome, whereby *Myoviridae* was the most dominant phage family, followed by *Siphoviridae* (Figure 3.21*)*. This is understandable, as *Bacteroidetes* is the most dominant phylum in the bovine rumen, followed by *Firmicutes* (18), whereas in feces, *Firmicutes* dominate (9, 30).

As mentioned before, *Caudovirales* were also the predominant viral component of equine fecal viromes, constituting 73% of total viral sequences in horse feces. That study used a sequence-independent cloning approach (129). The difference (~20%) in the RA of total *Caudovirales* in cattle fecal samples from our study versus horse feces may be due to differences in digestive tract anatomy and physiology. Horses are monogastrics, similar to pigs and humans (131). The stool of a healthy human had an RA of *Siphoviridae* of approximately 60% (130). Furthermore, fecal microbiomes of healthy humans have a high percentage of *Bacteroidetes,* with the greatest abundance in people eating fiber-rich diets (132). The equine fecal bacterial microbiome is also highly abundant in *Bacteroidetes* (133, 134), consistent with *Siphoviridae* (52%) being the leading phage family in horse feces (129). In contrast, as *Firmicutes* dominated the bacterial component of the bovine fecal microbiome (9, 30) it was expected that *Myoviridae* would be the most abundant group of phages, consistent with our findings.

The RA and number of species within *Myoviridae* were determined in this study. However, it was not reasonable to use diversity indices (e.g. Shannon index) because we selected a proportion of the total viral population (only dsDNA) in the sample. Furthermore, a lack of scientific literature providing details of *Myoviridae* populations in domesticated animals limited our ability to interpret and discuss diversity (number of genera and species) in cattle feces.

An interesting finding in the NextSeq run using Centrifuge classifier, was presence of viral reads belonging to linear viruses with dsDNA genome such as *Ligamenvirales*. Members of this order infect hyperthermophile and hypersaline members of the domain Archaea, present in extreme environments (108, 135). However, these types of archaeal phages have apparently not been

reported in cattle digestive systems. The two families in this order *Lipothrixviridae* (enveloped virions) and *Rubiviridae* (non-enveloped virions), share a core of orthologous genes, suggesting a common evolutionary ancestry (108). Based on the RA of this order (1.4%) in fecal samples in our study, we inferred that these archaeal phages may infect archaea members present in cattle digestive microbiomes.

The human fecal virome consistently contains a core phageome, a certain group of bacteriophages (52). In our study, a core of dsDNA phages with relatively constant RA of *Myoviridae* was present, regardless of the virus particle purification (VPP) used (Figure 3.18). There were small differences among methods used. CsCl concentrated free virus particles (130), whereas filtration may have retained bacterial DNA (96), perhaps including some with incorporated prophage genomes. It is noteworthy that filtration is a very rapid method to prepare fecal samples for metagenomic analysis, although it requires subsequent removal of RNA prior to further processing. Although CsCl centrifugation is a tedious procedure, in this study, CsCl centrifugation provided the best results. In that regard, it was the most sensitive method to concentrate phage particles and dsDNA viruses. There was an average of 7,980 *Caudovirales* reads using CsCl centrifugation, compared to 6,643 reads from filtration (Table 3.12). As CsCl centrifugation yielded a sample with high purity, no further purification or treatment was necessary before library preparation. Furthermore, EM visualization was possible after CsCl centrifugation, as the negative stained preparation facilitated morphological classification of phages.

In conclusion, although metagenomic analysis is currently the best way to characterize abundance and diversity of viruses, including *Myoviridae*, it remains a relatively expensive and time-consuming procedure.

Chapter 5: Limitations and Future Work

Due to limited technical experience with EM, it was not possible to measure head and tails of members of the family *Myoviridae.* Therefore, we were unable to validate EM as a method to determine diversity or abundance of these phages in cattle feces. Additionally, PCR with degenerate primers was not very successful for assessing cattle feces, most likely due to the presence of PCR inhibitory components in fecal material. We detected several bands in a single fecal sample and within a band, various viral sequences were detected. Unfortunately, due to limited genomic information of mammalian *Myoviridae* in the NCBI database, we were unable to validate PCR as a suitable method to determine diversity and abundance. In future studies, perhaps the use of other degenerate primers more suitable for T4 like viruses for PCR (65), should be explored. Next-generation sequencing could be more successful to uncover unknown diversity of phages in cattle feces.

There are limited data on the virome of cattle feces. Although there are some reports of analysis of fecal material of other domesticated (127) and wild animals (136, 137), most are focused on eukaryotic viruses, not necessarily on bacteriophages. However, there are many sequences from *Caudovirales* published in GenBank (65) as a consequence of intensive research in aquatic systems (39). These data should be used in future analysis and can be used to compare unclassified sequences to viral protein databases and use *de novo* assembly to identify unique sequences of *Caudovirales* in cattle feces. Further work is required to generate more data regarding diversity and abundance of *Myoviridae* in cattle feces. As only 17.2% of the total reads obtained from the NextSeq run were classified and only 0.1% of the total reads were classified to the virus domain, there is also a need for improved bioinformatic analysis strategies specifically focussed on viromes.

In our analysis, *Caudovirales* reads were classified into three families, due to reference genomes included in the database used by Centrifuge. However, *Caudovirales* has been updated (ICTV- 2018 release) and divided into four families, including the recently discovered

family: *Ackermannviridae* (61), infecting *Dickeya solani*, a phytopathogen (138) and pathogenic species of *Salmonella* (139). Perhaps an updated version of bioinformatics software containing all four families would result in a more complete classification, as there were some unclassified reads in the order *Caudovirales.*

That we used only two beef and two dairy fecal samples prevented us from conducting any statistical analyses. Although cost of NGS and bioinformatic analyses limited the number of samples analyzed, future work should use a larger sample size and also aim to characterize ssDNA and RNA bacteriophages and eukaryotic RNA viruses.

It was noteworthy that our analyses identified reads for *Ligamenvirales* archaeal phages in cattle feces. This was a novel discovery, as these phages have apparently not been reported in mammalian feces. Therefore, further studies to confirm the presence of these phages and describe them in more detail are indicated.

Future work could include comparisons between feces from healthy versus sick animals with acute or chronic bacterial diseases of the large intestinal tract, especially where dysbacteriosis can be expected, e.g. Johne's disease (12) and hemorrhagic diarrhea (30). Phages contribute to maintenance of bacterial diversity of the microbiome in the digestive tract of mammals (39), based on the predator-prey relationship between bacteria and phages. As a consequence of these interactions, numbers of bacteria fluctuate, directed mainly by lytic phages such as *Myoviridae* and *Siphoviridae*. Perhaps assessment of the virome and bacterial population in cattle with Johne's disease may also detect a specific shift in the phage population. It is important to relate these microbiome changes with viruses that infect bacteria, by measuring shifts in RA of phage species. Analysis of the phageome of cattle with and without Johne's disease could contribute to a better understanding of this disease.

In conclusion, despite its limitations, the current study provided detailed methodology to characterize dsDNA viromes in cattle feces, including phages from the order *Caudovirales*. It was clear that *Myoviridae* are an important component of the fecal microbiome in cattle. In

addition, future studies could help to elucidate their role in healthy and sick cattle, especially those with intestinal disorders.

References

1. Parker MT. An ecological framework of the human virome provides classification of current knowledge and identifies areas of forthcoming discovery. Yale Journal of Biology and Medicine. 2016;89(3):339-51.
2. Cani PD, Van Hul M, Lefort C, Depommier C, Rastelli M, Everard A. Microbial regulation of organismal energy homeostasis. Nature Metabolism. 2019;1(1):34-46.
3. Metzger RN, Krug AB, Eisenächer K. Enteric virome sensing—its role in intestinal homeostasis and immunity. Viruses. 2018;10(4).
4. Jenmalm MC. The mother–offspring dyad: microbial transmission, immune interactions and allergy development. Journal of Internal Medicine. 2017;282(6):484-95.
5. Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, et al. Genetic Determinants of the Gut Microbiome in UK Twins. Cell Host and Microbe. 2016;19(5):731-43.
6. Robinson C, Pfeiffer JK. Viruses and The Microbiota. Annual Review Virology. 2014;1:55-69.
7. Woo M. Inner Workings: Mapping the microbiome location helps elucidate its role. Proceedings of the National Academy of Sciences. 2018;115(48):12078.
8. Vandenplas Y, Huysentruyt K. The microbiome in healthy children. In: Green R, editor. Viral Infections in Children. 1. New York City: Springer; 2017. p. 183-91.
9. Shanks OC, Kelty CA, Archibeque S, Jenkins M, Newton RJ, McLellan SL, et al. Community structures of fecal bacteria in cattle from different animal feeding operations. Applied and Environmental Microbiology. 2011;77(9):2992-3001.
10. Khafipour E, Li S, Tun H, Derakhshani H, Moossavi S, Plaizier K. Effects of grain feeding on microbiota in the digestive tract of cattle. Animal Frontiers. 2016;6(2):13-9.
11. Santos-Marcos JA, Rangel-Zuñiga OA, Jimenez-Lucena R, Quintana-Navarro GM, Garcia-Carpintero S, Malagon MM, et al. Influence of gender and menopausal status on gut microbiota. Maturitas. 2018;116:43-53.
12. Fecteau ME, Pitta DW, Vecchiarelli B, Indugu N, Kumar S, Gallagher SC, et al. Dysbiosis of the fecal microbiota in cattle infected with Mycobacterium avium subsp. paratuberculosis. PLoS One. 2016;11(8).
13. Segal LN, Rom WN, Weiden MD. Lung microbiome for clinicians: New discoveries about bugs in healthy and diseased lungs. Annals of the American Thoracic Society. 2014;11(1):108-16.
14. Norman P, Duran R, Le Roux X, Morris C, Poggiale J-C. Biodeversity and Microbial Ecosystems Functioning. In: Bertrand JC, Caumette P, Lebaron P, Matheron R, Normand P, Sime-Ngando T, editors. Environmental Microbiology: Fundamentals and Applications. Dordrecht: Springer; 2015. p. 275-6.

15. Lima SF, Teixeira AGV, Higgins CH, Lima FS, Bicalho RC. The upper respiratory tract microbiome and its potential role in bovine respiratory disease and otitis media. Scientific Reports. 2016;6.
16. Timsit E, D.B. H, Hallewell J, Alexander TW. The nasopharyngeal microbiota in feedlot cattle and its role in respiratory disease. Animal Frontiers. 2016;6(2):44-50.
17. Gaeta NC, Lima SF, Teixeira AG, Ganda EK, Oikonomou G, Gregory L, et al. Deciphering upper respiratory tract microbiota complexity in healthy calves and calves that develop respiratory disease using shotgun metagenomics. Journal of Dairy Science. 2017;100(2):1445-58.
18. Henderson G, Cox F, Ganesh S, Jonker A, Young W, Janssen PH, et al. Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. Scientific Reports. 2015;5.
19. Laguardia-Nascimento M, Branco KMGR, Gasparini MR, Giannattasio-Ferraz S, Leite LR, Araujo FMG, et al. Vaginal microbiome characterization of Nellore cattle using metagenomic analysis. PLoS One. 2015;10(11).
20. Malmuthuge N, Guan LL. Gut microbiome and omics: a new definition to ruminant production and health. Animal Frontiers. 2016;6(2):8-12.
21. Steele MA, Penner GB, Chaucheyras-Durand F, Guan LL. Development and physiology of the rumen and the lower gut: Targets for improving gut health. Journal of Dairy Science. 2016;99(6):4955-66.
22. Peterson LW, Artis D. Intestinal epithelial cells: Regulators of barrier function and immune homeostasis. Nature Reviews Immunology. 2014;14(3):141-53.
23. Seshadri R, Leahy SC, Attwood GT, Teh KH, Lambie SC, Cookson AL, et al. Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. Nature Biotechnology. 2018;36(4):359-67.
24. Choudhury PK, Salem AZM, Jena R, Kumar S, Singh R, Puniya AK. Rumen Microbiology: An Overview. In: Puniya AK, Singh R, Kamra DN, editors. Rumen Microbiology: From Evolution to Revolution. New Delhi: Springer; 2015. p. 3-16.
25. Wu S, Baldwin Vi RL, Li W, Connor E, Li RW, . The Bacterial Community Composition of the Bovine Rumen Detected Using Pyrosequencing of 16S rRNA Genes. Metagenomics. 2012;1:1-11.
26. Cersosimo L, Wright A-D. Rumen Methanogens. In: Puniya AK, Singh R, Kamra DN, editors. Rumen Microbiology: From Evolution to Revolution. New Delhi: Springer; 2015. p. 143-50.
27. Wright A-DG. Rumen Protozoa. In: Puniya AK, Singh R, Kamra DN, editors. Rumen Microbiology: From Evolution to Revolution. New Delhi: Springer; 2015. p. 113-9.
28. Gilbert RA, Klieve AV. Ruminal Viruses (Bacteriophages, Archaeaphages). In: Puniya AK, Singh R, Kamra DN, editors. Rumen Microbiology: From Evolution to Revolution. New Delhi: Springer; 2015. p. 122-41.
29. Prangishvili D, Forterre P, Garrett RA. Viruses of the Archaea: a unifying view. Nat Rev Microbiol. 2006;4:837.

30. Zeineldin M, Aldridge B, Lowe J. Dysbiosis of the fecal microbiota in feedlot cattle with hemorrhagic diarrhea. Microbial Pathogenesis. 2018;115:123-30.
31. Frey JC, Pell AN, Berthiaume R, Lapierre H, Lee S, Ha JK, et al. Comparative studies of microbial populations in the rumen, duodenum, ileum and faeces of lactating dairy cows. Journal of Applied Microbiology. 2010;108(6):1982-93.
32. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, et al. Microbiology: Diversity of the human intestinal microbial flora. Science. 2005;308(5728):1635-8.
33. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. Nature. 2012;486(7402):222-7.
34. Wylie KM, Weinstock GM, Storch GA. Emerging view of the human virome. Translational Research. 2012;160(4):283-90.
35. Ross EM, Petrovski S, Moate PJ, Hayes BJ. Metagenomics of rumen bacteriophage from thirteen lactating dairy cattle. BioMed Central microbiology. 2013;13.
36. Zárate S, Taboada B, Yocupicio-Monroy M, Arias CF. Human Virome. Archives of Medical Research. 2017;48(8):701-16.
37. Hillman ET, Lu H, Yao T, Nakatsu CH. Microbial ecology along the gastrointestinal tract. Microbes and Environments. 2017;32(4):300-13.
38. Berg Miller ME, Yeoman CJ, Chia N, Tringe SG, Angly FE, Edwards RA, et al. Phage-bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome. Environmental Microbiology. 2012;14(1):207-27.
39. Clokie MRJ, Millard AD, Letarov AV, Heaphy S. Phages in nature. Bacteriophage. 2011;1(1):31-45.
40. Engelkirk PG, Duben-Engelkirk J. Microbial Diversity. Burton's Microbiology for the Health Sciences. 10th ed. Philadelphia: Wolters Kluwer Health; 2015. p. 45-72.
41. Malik PK, Singhal KK, Deshpande SB, Siddique RA. Mitigation strategies for enteric methane emission with special emphasis on biological approaches: A review. Indian Journal of Animal Sciences. 2012;82(8):794-804.
42. Klieve AV, Hegarty RS. Opportunities for biological control of ruminal methanogenesis. Australian Journal of Agricultural Research. 1999;50(8):1315-9.
43. Callaway TR, Edrington TS, Rychlik JL, Genovese KJ, Poole TL, Jung YS, et al. Ionophores: Their use as ruminant growth promotants and impact on food safety. Current Issues in Intestinal Microbiology. 2003;4(2):43-51.
44. Witzig M, Zeder M, Rodehutscord M. Effect of the ionophore monensin and tannin extracts supplemented to grass silage on populations of ruminal cellulolytics and methanogens in vitro. Anaerobe. 2018;50:44-54.
45. Novilla ME. Veterinary Toxicology: Basic and Clinical Principles. In: Gupta RC, editor. 3 ed. Philadelphia: Elsevier 2018. p. 1073-92.
46. Adams JC, Gazaway, J.A., Brailsford, M.D. et al. Isolation of bacteriophages of bovine rumen. Experientia. 1966;22(1).

47. Ackermann H-W. Bacteriophage electron microscopy. In: Łobocka M, Szybalski WT, editors. Advances in Virus Research. 82. New York City: Academic Press; 2012. p. 1-32.
48. Leahy SC, Kelly WJ, Altermann E, Ronimus RS, Yeoman CJ, Pacheco DM, et al. The Genome Sequence of the Rumen Methanogen Methanobrevibacter ruminantium Reveals New Possibilities for Controlling Ruminant Methane Emissions. PLoS One. 2013;5(1).
49. Ni J, Wu GD, Albenberg L, Tomov VT. Gut microbiota and IBD: Causation or correlation? Nature Reviews Gastroenterology and Hepatology. 2017;14(10):573-84.
50. Norman JM, Handley SA, Baldridge MT, Droit L, Liu CY, Keller BC, et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. Cell. 2015;160(3):447-60.
51. Lim ES, Zhou Y, Zhao G, Bauer IK, Droit L, Ndao IM, et al. Early life dynamics of the human gut virome and bacterial microbiome in infants. Nature Medicine. 2015;21(10):1228-34.
52. Manrique P, Bolduc B, Walk ST, Van Oost JD, De Vos WM, Young MJ. Healthy human gut phageome. Proceedings of the National Academy of Sciences of the United States of America. 2016;113(37):10400-5.
53. Broecker F, Russo G, Klumpp J, Moelling K. Stable core virome despite variable microbiome after fecal transfer. Gut Microbes. 2017;8(3):214-20.
54. Zuo T, Kamm MA, Colombel JF, Ng SC. Urbanization and the gut microbiota in health and inflammatory bowel disease. Nature Reviews Gastroenterology and Hepatology. 2018:1-13.
55. Dafale NA, Hathi ZJ, Bit S, Purohit HJ. Bacteriophage Diversity in Different Habitats and Their Role in Pathogen Control. In: Kalia VC, editor. Microbial Factories: Biodiversity, Biopolymers, Bioactive Molecules: Volume 2. New Delhi: Springer 2015. p. 259-80.
56. Bauman WR. Microbiology. With diseases by Taxonomy. 5th ed. New York City: Pearson; 2017.
57. Wang J, Niu YD, Chen J, Anany H, Ackermann HW, Johnson RP, et al. Feces of feedlot cattle contain a diversity of bacteriophages that lyse non-o157 shiga toxin-producing escherichia coli. Canadian Journal of Microbiology. 2015;61(7):467-75.
58. Hallewell J, Niu YD, Munns K, McAllister TA, Johnson RP, Ackermann HW, et al. Differing populations of endemic bacteriophages in cattle shedding high and low numbers of escherichia coli O157:H7 bacteria in feces. Applied and Environmental Microbiology. 2014;80(13):3819-25.
59. Niu YD, McAllister TA, Nash JHE, Kropinski AM, Stanford K. Four Escherichia coli O157:H7 phages: A new bacteriophage genus and taxonomic classification of T1-like phages. PLoS One. 2014;9(6).
60. Ackermann HW. Phage classification and characterization. Methods in molecular biology 2009;501:127-40.

61. International Committee on Taxonomy of Viruses: ICTV [Internet]. 2011 [cited 2018 July 31]. Available from: https://talk.ictvonline.org/ictv-reports/ictv_9th_report/dsdna-viruses-2011/w/dsdna_viruses/68/myoviridae.
62. International Committee on Taxonomy of Viruses: ICTV [Internet]. 2018 [cited 2018 Dec 1]. Available from: https://talk.ictvonline.org/files/master-species-lists/m/msl/7992.
63. King AMQ, Lefkowitz E, Adams MJ, Carstens EB. Family Myoviridae. Virus Taxonomy. San Diego: Elsevier; 2012. p. 46-62.
64. Hambly E, Tétart F, Desplats C, Wilson WH, Krisch HM, Mann NH. A conserved genetic module that encodes the major virion components in both the coliphage T4 and the marine cyanophage S-PM2. Proceedings of the National Academy of Sciences of the United States of America. 2001;98(20):11411-6.
65. Adriaenssens EM, Cowan DA. Using signature genes as tools to assess environmental viral ecology and diversity. Applied and Environmental Microbiology. 2014;80(15):4470-80.
66. Parmar K, Dafale N, Pal R, Tikariha H, Purohit H. An Insight into Phage Diversity at Environmental Habitats using Comparative Metagenomics Approach. Current Microbiology. 2018;75(2):132-41.
67. Samtlebe M, Denis S, Chalancon S, Atamer Z, Wagner N, Neve H, et al. Bacteriophages as modulator for the human gut microbiota: Release from dairy food systems and survival in a dynamic human gastrointestinal model. LWT - Food Science and Technology. 2018;91:235-41.
68. Samtlebe M, Ergin F, Wagner N, Neve H, Küçükçetin A, Franz CMAP, et al. Carrier systems for bacteriophages to supplement food systems: Encapsulation and controlled release to modulate the human gut microbiota. LWT - Food Science and Technology. 2016;68:334-40.
69. Jami E, Israel A, Kotser A, Mizrahi I. Exploring the bovine rumen bacterial community from birth to adulthood. Isme J. 2013;7:1069-79.
70. Takino T, Kato-Mori Y, Motooka D, Nakamura S, Iida T, Hagiwara K. Postnatal changes in the relative abundance of intestinal Lactobacillus spp. In newborn calves. Journal of Veterinary Medical Science. 2017;79(3):452-5.
71. Meale SJ, Li S, Azevedo P, Derakhshani H, Plaizier JC, Khafipour E, et al. Development of ruminal and fecal microbiomes are affected by weaning but not weaning strategy in dairy calves. Frontiers in Microbiology. 2016;7(5).
72. Rudi K, Moen B, Sekelja M, Frisli T, Lee MRF. An eight-year investigation of bovine livestock fecal microbiota. Veterinary Microbiology. 2012;160(3):369-77.
73. Penders J, Stobberingh EE, Brandt PAVD, Thijs C. The role of the intestinal microbiota in the development of atopic disorders. Allergy: European Journal of Allergy and Clinical Immunology. 2007;62(11):1223-36.
74. Langmead B. Aligning short sequencing reads with Bowtie. Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al]. 2010;CHAPTER:Unit-11.7.

75. Hernandez-Sanabria E, Goonewardene LA, Wang Z, Durunna ON, Moore SS, Guan LL. Impact of feed efficiency and diet on adaptive variations in the bacterial community in the rumen fluid of cattle. Applied and Environmental Microbiology. 2012;78(4):1203-14.
76. Taniguchi M, Penner GB, Beauchemin KA, Oba M, Guan LL. Comparative analysis of gene expression profiles in ruminal tissue from Holstein dairy cows fed high or low concentrate diets. Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics. 2010;5(4):274-9.
77. Krieg NR, Ludwig W, Euzéby J, Whitman WB. Phylum XIV. Bacteroidetes phyl. nov. In: al. KNRe, editor. Bergey's Manual® of Systematic Bacteriology. 4. 2nd ed. New York City: Springer; 2010. p. 85-101.
78. Malmuthuge N, Chen Y, Liang G, Goonewardene LA, Guan LL. Heat-treated colostrum feeding promotes beneficial bacteria colonization in the small intestine of neonatal calves. Journal of Dairy Science. 2015;98(11):8044-53.
79. Uetake K. Newborn calf welfare: A review focusing on mortality rates. Animal Science Journal. 2013;84(2):101-5.
80. House J, Smith GF, McGuirk S, Gunn A, Izzo M. Manifestations and Management of Disease in Neonatal Ruminants. In: Smith BP, editor. Large Animal Internal Medicine. 5th ed. St. Louis: Elsevier 2015. p. 302-38.
81. Kim MK, Lee HG, Park JA, Kang SK, Choi YJ. Effect of feeding direct-fed microbial as an alternative to antibiotics for the prophylaxis of calf diarrhea in holstein calves. Asian-Australasian Journal of Animal Sciences. 2011;24(5):643-9.
82. Babickova J, Gardlik R. Pathological and therapeutic interactions between bacteriophages, microbes and the host in inflammatory bowel disease. World Journal of Gastroenterology. 2015;21(40):11321-30.
83. Zhou M, Chen Y, Griebel PJ, Guan LL. Methanogen prevalence throughout the gastrointestinal tract of pre-weaned dairy calves. Gut Microbes. 2015;5(5):628-38.
84. Swain RA, Nolan JV, Klieve AV. Natural variability and diurnal fluctuations within the bacteriophage population of the rumen. Applied and Environmental Microbiology. 1996;62(3):994-7.
85. Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature. 2010;466(7304):334-8.
86. Minot S, Bryson A, Chehoud C, Wu GD, Lewis JD, Bushman FD. Rapid evolution of the human gut virome. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(30):12450-5.
87. Sarethy IP, Pan, S. and Danquah, M. Modern taxonomy of Microbial Diversity [Internet]. Intech 2014 [cited 2018 April 29]. Available from: https://www.intechopen.com/books/biodiversity-the-dynamic-balance-of-the-planet/modern-taxonomy-for-microbial-diversity.
88. Illumina. Nextera DNA Library Preperation Kits [Internet]. San Diego: Illumina; 2013 [cited 2018 Jun 20]. Available from: https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_nextera_dna_sample_prep.pdf.

89. Zhou M, Chen Y, Guan LL. Rumen Bacteria. In: Puniya AK, Singh R, Kamra DN, editors. Rumen Microbiology: From Evolution to Revolution. New Delhi: Springer; 2015.
90. Stewart EJ. Growing unculturable bacteria. Journal of Bacteriology. 2012;194(16):4151-60.
91. Budinoff CR, Loar SN, Lecleir GR, Wilhelm SW, Buchan A. A protocol for enumeration of aquatic viruses by epifluorescence microscopy using Anodisc 13 membranes. BioMed Central Microbiology. 2011;11.
92. Brussaard CPD, Marie D, Bratbak G. Flow cytometric detection of viruses. Journal of Virological Methods. 2000;85(1):175-82.
93. Shende RK, Hirpurkar SD, Sannat C, Rawat N, Pandey V. Isolation and characterization of bacteriophages with lytic activity against common bacterial pathogens. Veterinary World. 2017;10(8):973-8.
94. Haramoto E, Kitajima M, Hata A, Torrey JR, Masago Y, Sano D, et al. A review on recent progress in the detection methods and prevalence of human enteric viruses in water. Water Research. 2018;135:168-86.
95. Fancello L, Raoult D, Desnues C. Computational tools for viral metagenomics and their application in clinical research. Virology. 2012;434(2):162-74.
96. Conceição-Neto N, Zeller M, Lefrère H, De Bruyn P, Beller L, Deboutte W, et al. Modular approach to customise sample preparation procedures for viral metagenomics: A reproducible protocol for virome analysis. Scientific Reports. 2015;5.
97. Hurwitz BL, Deng L, Poulos BT, Sullivan MB. Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. Environmental Microbiology. 2013;15(5):1428-40.
98. Kleiner M, Hooper LV, Duerkop BA. Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. BioMed Central Genomics. 2015;16(1).
99. Hayes S, Mahony J, Nauta A, Van Sinderen D. Metagenomic approaches to assess bacteriophages in various environmental niches. Viruses. 2017;9(6).
100. Buffalo V. Bioinformatics Data Skills. 1st ed. Sebastopol, CA: O'Reilly Media, Inc; 2015.
101. Lin J, Kramna L, Autio R, Hyöty H, Nykter M, Cinek O. Vipie: Web pipeline for parallel characterization of viral populations from multiple NGS samples. BioMed Central Genomics. 2017;18(1).
102. Yamashita A, Sekizuka T, Kuroda M. VirusTAP: Viral genome-targeted assembly pipeline. Frontiers in Microbiology. 2016;7.
103. Li P-E, Lo C-C, Anderson JJ, Davenport KW, Bishop-Lilly KA, Xu Y, et al. Enabling the democratization of the genomics revolution with a fully integrated web-based bioinformatics platform. Nucleic Acids Research. 2017;45(1):67-80.
104. Roux S, Tournayre J, Mahul A, Debroas D, Enault F. Metavir 2: New tools for viral metagenome comparison and assembled virome analysis. BioMed Central Bioinformatics. 2014;15(1).

105. Krishnamurthy SR, Wang D. Origins and challenges of viral dark matter. Virus Research. 2017;239:136-42.
106. Li RW. Rumen Metagenomics. In: Puniya AK, Singh R, Kamra DN, editors. Rumen Microbiology: From Evolution to Revolution. New Delhi: Springer 2015. p. 223-45.
107. Center for Computational Biology. Centrifuge [Internet]. Johns Hopskins University; 2016 [cited 2018 Jun 13]. Available from: http://www.ccb.jhu.edu/software/centrifuge.
108. Pina M, Bize A, Forterre P, Prangishvili D. The archeoviruses. FEMS Microbiology Reviews. 2011;35(6):1035-54.
109. Ncumby. Bacteriophage HK97 *Siphoviridae* [Internet]. Commoms wikimedia2011 [cited 2018 Jul 12]. Available from: https://commons.wikimedia.org/wiki/File:BacteriophageHK97.jpg.
110. Mostafa F. *Myoviridae* [Internet]. Commons wikimedia2004 [cited 2018 Jul 12]. Available from: https://commons.wikimedia.org/wiki/File:PhageP2.jpg.
111. Casjens S, Lenk l. P22 bacteriophage from the family *Podoviridae* [Internet]. Commons wikimedia 2011 [cited 2018 Jul 12]. Available from: https://commons.wikimedia.org/wiki/File:Bacteriophage_P22_Casjens_Lenk.png.
112. Gould EA. Virus cryopreservation and storage. Methods in molecular biology (Clifton, NJ). 1995;38:7-20.
113. Dalton CS, van de Rakt K, Fahlman Å, Ruckstuhl K, Neuhaus P, Popko R, et al. Discovery of herpesviruses in Canadian wildlife. Archives of Virology. 2017;162(2):449-56.
114. Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, et al. A large genome center's improvements to the Illumina sequencing system. Nature Methods. 2008;5:1005.
115. Fuller NJ, Wilson WH, Joint IR, Mann NH. Occurrence of a sequence in marine cyanophages similar to that of T4 g20 and its application to PCR-based detection and quantification techniques. Applied and Environmental Microbiology. 1998;64(6):2051-60.
116. Zhong Y, Chen F, Wilhelm SW, Poorvin L, Hodson RE. Phylogenetic diversity of marine cyanophage isolates and natural virus communities as revealed by sequences of viral capsid assembly protein gene g20. Applied and Environmental Microbiology. 2002;68(4):1576-84.
117. Short CM, Suttle CA. Nearly identical bacteriophage structural gene sequences are widely distributed in both marine and freshwater environments. Applied and Environmental Microbiology. 2005;71(1):480-6.
118. Filée J, Tétart F, Suttle CA, Krisch HM. Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(35):12471-6.
119. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Rüger W. Bacteriophage T4 genome. Microbiology and Molecular Biology Reviews. 2003;67(1):86-156.

120. Herath D, Jayasundara D, Ackland D, Saeed I, Tang SL, Halgamuge S. Assessing Species Diversity Using Metavirome Data: Methods and Challenges. Computational and Structural Biotechnology Journal. 2017;15:447-55.
121. Schrader C, Schielke A, Ellerbroek L, Johne R. PCR inhibitors - occurrence, properties and removal. Journal of Applied Microbiology. 2012;113(5):1014-26.
122. Wood DE, Salzberg SL. Kraken: Ultrafast metagenomic sequence classification using exact alignments [Internet]. 2014 [cited 2017 Jun 20]. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84899090573&doi=10.1186%2fgb-2014-15-3-r46&partnerID=40&md5=fb87d9947996810736d38056e021a665.
123. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: Rapid and sensitive classification of metagenomic sequences. Genome Research. 2016;26(12):1721-9.
124. Lindgreen S, Adair KL, Gardner PP. An evaluation of the accuracy and speed of metagenome analysis tools. Scientific Reports. 2016;6.
125. Wood DE, Salzberg SL. Centrifuge [Internet]. 2016 [cited 2018 Nov 13]. Available from: https://ccb.jhu.edu/software/centrifuge/.
126. Amimo JO, El Zowalaty ME, Githae D, Wamalwa M, Djikeng A, Nasrallah GK. Metagenomic analysis demonstrates the diversity of the fecal virome in asymptomatic pigs in East Africa. Archives of Virology. 2016;161(4):887-97.
127. Mihalov-Kovács E, Fehér E, Martella V, Bányai K, Farkas SL. The fecal virome of domesticated animals. VirusDisease. 2014;25(2):150-7.
128. Sachsenröder J, Twardziok SO, Scheuch M, Johne R. The general composition of the faecal virome of pigs depends on age, but not on feeding with a probiotic bacterium. PLoS One. 2014;9(2).
129. Cann AJ, Fandrich SE, Heaphy S. Analysis of the virus population present in equine faeces indicates the presence of hundreds of uncharacterized virus genomes. Virus Genes. 2005;30(2):151-6.
130. Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P, et al. Metagenomic analyses of an uncultured viral community from human feces. Journal of Bacteriology. 2003;185(20):6220-3.
131. Krause D, E K. The Fecal Environment, The Gut. In: Sadowsky MJ, Withman RL, editors. The Fecal bacteria. USA: ASM Press; 2001. p. 14.
132. Sharma S, Tripathi P. Gut microbiome and type 2 diabetes: where we are and where to go? Journal of Nutritional Biochemistry. 2019;63:101-8.
133. Beckers KF, Schulz CJ, Childers GW. Rapid regrowth and detection of microbial contaminants in equine fecal microbiome samples. PLoS One. 2017;12(11).
134. Dougal K, Harris PA, Girdwood SE, Creevey CJ, Curtis GC, Barfoot CF, et al. Changes in the total fecal bacterial population in individual horses maintained on a restricted diet over 6 weeks. Frontiers in Microbiology. 2017;8(8).
135. Prangishvili D, Krupovic M. A new proposed taxon for double-stranded DNA viruses, the order "Ligamenvirales". Archives of Virology. 2012;157(4):791-5.

136. Anthony SJ, Epstein JH, Murray KA, Navarrete-Macias I, Zambrana-Torrelio CM, Solovyov A, et al. A strategy to estimate unknown viral diversity in mammals. mBio. 2013;4(5).
137. Lojkić I, Biđin M, Prpić J, Šimić I, Krešić N, Bedeković T. Faecal virome of red foxes from peri-urban areas. Comparative Immunology, Microbiology and Infectious Diseases. 2016;45:10-5.
138. Day A, Ahn J, Salmond GPC. Jumbo bacteriophages are represented within an increasing diversity of environmental viruses infecting the emerging phytopathogen, Dickeya solani. Frontiers in Microbiology. 2018;9.
139. Newase S, Kapadnis BP, Shashidhar R. Isolation and Genome Sequence Characterization of Bacteriophage vB_SalM_PM10, a Cba120virus, Concurrently Infecting Salmonella enterica Serovars Typhimurium, Typhi, and Enteritidis. Current Microbiology. 2019;76(1):86-94.

Appendix A: Supplementary Tables, Chapter 3

Virus species within the *Myoviridae* whereby > 10 reads were detected in any fecal sample. Only five genera are included in this appendix (Total genera within *Myoviridae* in this study corresponds to 38): *T4-like virus*, *Schizot4virus*, *Kayvirus*, *Phikzvirus* and *Cp220viru*s. Note than within *T4-like viruses*, the number of species is 71.

Table 1. Species within genus *T4-like virus* of the family Myoviridae.

		Number of reads /sample by VPP							
		36 VPP2	**36 VPP3A**	**37 VPP2**	**37 VPP3A**	**38 VPP2**	**38 VPP3A**	**39 VPP2**	**39 VPP3A**
1	Acinetobacter phage Acj9	42	31	1	3	11	3	6	1
2	Stenotrophomonas phage IME-SM1	35	6	3	2	7	5	7	6
3	Prochlorococcus phage P-SSM4	30	1	5	6	5	9	7	8
4	Cyanophage P-RSM3	21		3	4		5	5	2
5	Prochlorococcus phage P-SSM2	25	20	45	19	38	22	15	30
6	Prochlorococcus phage P-SSM5	11	3	25	12	20	13	7	14
7	Aeromonas phage PX29	14	2	11	1	10	7	8	21
8	Synechococcus phage ACG-2014f	14	24	17	2	27	26	18	13
9	Synechococcus phage S-PM2	13	2	7	2	15	12	11	18
10	Aeromonas phage phiAS5	8	1	14	4	11	5	14	15
11	Proteus phage PM2	8	12	7	2	18	0	4	5
12	Acinetobacter phage Acj61	7	11	1	2	10	2	5	3
13	Acinetobacter phage Ac42	7	4	7	1	18	6	12	6
14	Synechococcus phage ACG-2014b	7	3	11	2	5	3	8	
15	Synechococcus phage ACG-2014d	7		16	16	36	11	1	9
16	Synechococcus phage ACG-2014h	6	7	4		12	1	4	2
17	Synechococcus phage S-RSM4	5		5	5	8	9	9	7
18	Escherichia phage slur07	5	1	3		3	1		2
19	Escherichia phage slur08	5	3		1	3	4	4	
20	Escherichia phage slur14	3	1					2	
21	Synechococcus phage syn9	4	5	9	4	9	11	17	2
22	Cyanophage Syn10	2	4	4	2	6	6	15	2
23	Acinetobacter phage ZZ1	4	4	4	1	23	15	15	5
24	Synechococcus phage ACG-2014c	4		16	3	1	6	6	4
25	Aeromonas phage CC2	4			1	6	12	8	32
26	Synechococcus phage ACG-2014i	4	1	3	2	2	1	3	3
27	Proteus phage vB_PmiM_Pm5461	4	1	3		7	3	3	1

Continuation Table 1. Species within genus *T4-like virus* of the family *Myoviridae.*

		Number of reads/sample by VPP							
		36 VPP2	**36 VPP3A**	**37 VPP2**	**37 VPP3A**	**38 VPP2**	**38 VPP3A**	**39 VPP2**	**39 VPP3A**
28	Synechococcus phage ACG-2014g	3		2	1	2	4	3	3
29	Synechococcus phage metaG-MbCM1	2	2	39	33	6	7	5	7
30	Synechococcus phage S-MbCM100	2	3	4	1	11	2	6	1
31	Serratia phage PS2	2		2	2	10	2	3	5
32	Synechococcus phage ACG-2014e	2		1	2		4	2	4
33	Citrobacter phage IME-CF2	2		12	1	11	10	10	2
34	Salmonella phage vB_SnwM_CGG4-1	2		1		11	18	3	16
35	Enterobacteria phage JS	1	3	1	1	3		2	1
36	Escherichia phage JS98	1		1	1	3		2	1
37	Citrobacter phage Margaery	1	1	6	23	2	3	1	1
38	Escherichia virus T4	1	4	1	4	4	2	5	5
39	Shigella phage Shfl2	2						9	
40	Yersinia phage PST	2						2	2
41	Shigella phage SHFML-11		2						
42	Enterobacteria phage RB70		1						
43	Aeromonas phage Aes516		1	1	1	2		1	
44	Escherichia phage Lw1		1	7	3	3		1	3
45	Escherichia phage MX01		1	1	1	2	1	2	
46	Citrobacter phage vB_CfrM_CfP1		1			1		1	1
47	Synechococcus phage ACG-2014j			3		4	3	5	3
48	Enterobacteria phage Ac3			1		1			
49	Aeromonas phage Aes120			1	1				2
50	Escherichia phage vB_EcoM_PhAPEC2			1			1	2	
51	Escherichia phage slur02			1		2			
52	Shigella phage SHSML-52-1			1		1			
53	Escherichia phage YUEEL01			1			1		
54	Escherichia virus HY01			2	1	6	10		

Continuation: Species within genus *T4-like virus* of the family *Myoviridae*

		Number of reads/sample by VPP							
		36 VPP2	**36 VPP3A**	**37 VPP2**	**37 VPP3A**	**38 VPP2**	**38 VPP3A**	**39 VPP2**	**39 VPP3A**
55	Escherichia virus ECML134			1			1		
56	Escherichia virus C40			1		3			
57	Shigella virus Pss1			1		5	1	2	2
58	Cronobacter phage vB_CsaM_GAP161				1		2	1	1
59	Escherichia virus E112				1	1			
60	Escherichia phage WG01					2			
61	Enterobacteria phage RB51					1			
62	Salmonella phage STP4-a					1			
63	Shigella phage SHBML-50-1					1			
64	Shigella phage SHFML-26					1			
65	Staphylococcus phage SAJK-IND					1			
66	Escherichia virus Ime09					1		1	1
67	Yersinia virus D1						1	2	1
68	Aeromonas phage AS-zj						1		
69	Escherichia virus AR1							2	
70	Escherichia phage LZ1								
71	Escherichia virus RB3								1
72	Enterobacteria phage GEC-3S	1				6		5	3
	Total number of reads	**365**	**167**	**317**	**175**	**413**	**272**	**287**	**276**

Table 2. Species within genus *Schizot4virus* of the family *Myoviridae*

		Number of reads/sample by VPP							
		36 VPP2	36 VPP3A	37 VPP2	37 VPP3A	38 VPP2	38 VPP3A	39 VPP2	39 VPP3A
1	Vibrio phage VH7D	9	2	18	6	20	13	25	14
2	Vibrio phage nt1	7	5	17	1	36	42	15	14
3	Vibrio phage ValKK3	6	3	9	14	17	10	8	3
4	Vibrio phage KVP40	5	8	11	22	16	14	13	14
	Total	27	18	55	43	89	79	61	45

Table 3. Species within genus *Kayvirus* of the family *Myoviridae*

		Number of reads/sample by VPP							
		36 VPP2	36 VPP3A	37 VPP2	37 VPP3A	38 VPP2	38 VPP3A	39 VPP2	39 VPP3A
1	Staphylococcus virus JD7	39	9	8	4	6	6	1	5
2	Staphylococcus virus Rodi	31	223	12	26	0	4		
3	Staphylococcus virus S253	12	2		1			3	
4	Staphylococcus virus SA12	10	13		2	4	1	2	
	Total	92	247	20	33	10	11	6	5

Table 4. Species within genus *Phikzvirus* of the family *Myoviridae*

		Number of reads/sample by VPP							
		36 VPP2	36 VPP3A	37 VPP2	37 VPP3A	38 VPP2	38 VPP3A	39 VPP2	39 VPP3A
1	Pseudomonas virus phiKZ	28	7	28	17	33	29	23	20
2	Pseudomonas phage Phabio	7	18	14	20	7	14	10	2
3	Pseudomonas phage PhiPA3	2	4	10	4	8	17	25	19
4	Pseudomonas phage PhiPA3	2	5	6	3	8	10	10	7
	Total	39	34	55	61	56	70	68	48

Table 5. Species within genus *Cp220virus* of the family *Myoviridae*

		Number of reads/sample by VPP							
		36 VPP2	**36 VPP3A**	**37 VPP2**	**37 VPP3A**	**38 VPP2**	**38 VPP3A**	**39 VPP2**	**39 VPP3A**
1	Campylobacter virus CP220	12	75	2	3	36	22	6	6
2	Campylobacter virus CP21	9	11	6	4	14	10	12	13
3	Campylobacter virus IBB35	7	8	10	6	16	21	10	16
	Total	**28**	**94**	**18**	**13**	**66**	**53**	**28**	**35**